



Interpretable multiphasic CT-based radiomic analysis for preoperatively differentiating benign and malignant solid renal tumors: a multicenter study

Yaohai Wu¹ · Fei Cao¹ · Hanqi Lei¹ · Shiqiang Zhang¹ · Hongbing Mei² · Liangchao Ni³ · Jun Pang¹

Received: 8 March 2024 / Revised: 16 April 2024 / Accepted: 17 April 2024
© The Author(s) 2024

Abstract

Background To develop and compare machine learning models based on triphasic contrast-enhanced CT (CECT) for distinguishing between benign and malignant renal tumors.

Materials and Methods In total, 427 patients were enrolled from two medical centers: Center 1 (serving as the training set) and Center 2 (serving as the external validation set). First, 1781 radiomic features were individually extracted from corticomedullary phase (CP), nephrographic phase (NP), and excretory phase (EP) CECT images, after which 10 features were selected by the minimum redundancy maximum relevance method. Second, random forest (RF) models were constructed from single-phase features (CP, NP, and EP) as well as from the combination of features from all three phases (TP). Third, the RF models were assessed in the training and external validation sets. Finally, the internal prediction mechanisms of the models were explained by the SHapley Additive exPlanations (SHAP) approach.

Results A total of 266 patients with renal tumors from Center 1 and 161 patients from Center 2 were included. In the training set, the AUCs of the RF models constructed from the CP, NP, EP, and TP features were 0.886, 0.912, 0.930, and 0.944, respectively. In the external validation set, the models achieved AUCs of 0.860, 0.821, 0.921, and 0.908, respectively. The “original_shape_Flatness” feature played the most important role in the prediction outcome for the RF model based on EP features according to the SHAP method.

Conclusions The four RF models efficiently differentiated benign from malignant solid renal tumors, with the EP feature-based RF model displaying the best performance.

✉ Jun Pang
pangjun2@mail.sysu.edu.cn

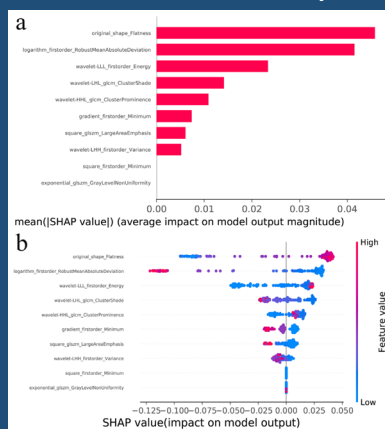
¹ Department of Urology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

² Department of Urology, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China

³ Department of Urology, Guangdong and Shenzhen Key Laboratory of Reproductive Medicine and Genetics, Peking University Shenzhen Hospital, Shenzhen, China

Graphical Abstract

Interpretable multiphase CT-based radiomic analysis for preoperatively differentiating benign and malignant solid renal tumors: a multicenter study



Yaohai Wu et al; 2024

SHAP values for the interpretation of the RF models. Bar plot (a) and beeswarm plot (b) showing the SHAP values of the RF model based on EP features in the external validation set.

Abdominal Radiology
The Official Journal of the Society of Abdominal Radiology www.abdominalradiology.org

Keywords Computed tomography · Radiomics · Random forest · Renal tumors · SHapley Additive exPlanations

Introduction

Renal cancer is the third leading cause of genitourinary cancer death, and its incidence has been increasing over the past decade [1]. Renal cell carcinoma (RCC) constitutes approximately 70% of all renal cancers; its most common histopathological subtypes include clear cell RCC (ccRCC), papillary RCC (pRCC), and chromophobe RCC (chRCC) [2]. Patients diagnosed with renal cancer are recommended to undergo radical or partial nephrectomy as soon as possible [3]. However, in clinical practice, for some patients with benign renal tumors, including renal oncocytoma (RO) and angioliomyolipoma (AML), cannot be accurately distinguished from malignant renal tumors, including ccRCC, pRCC and chRCC, with imaging before surgery [4, 5]. RO, a benign tumor constituting 3–5% of renal neoplasms in adults, can be misclassified as RCC on imaging; this misclassification is responsible for 4–10% of surgeries performed for suspected RCC [4]. Additionally, as many as 5% of renal AML exhibit insufficient fat content, making differentiation from RCC using conventional computed tomography (CT) and magnetic resonance imaging (MRI) techniques challenging [5]. Consequently, patients with benign renal tumors can be misdiagnosed with malignant renal tumors, potentially resulting in surgical treatment that may impose surgical complications and a heavy economic burden. Therefore, it is highly important to find a simple, noninvasive and

accurate method for preoperatively differentiating benign from malignant solid renal tumors.

Developments in medical examination techniques have led to CT, ultrasound and MRI playing increasingly important roles in the differential diagnosis of tumors [6, 7]. In clinical practice, CT is an important method for the differential diagnosis of renal tumors because it is more precise than ultrasound and more convenient than MRI [8, 9]. While the value of CT imaging in the differential diagnosis of renal tumors continues to increase [10, 11], importantly, its performance in differentially diagnosing renal tumors remains less than ideal. Renal tumor biopsy is another important tool used for histological diagnosis, with a diagnostic rate of 78–97% for malignancies [12]. However, despite this importance, renal biopsy presents with inherent risks such as bleeding, seeding and pain [13]. Moreover, due to the limited tissue samples obtained during biopsy, there is a risk of incomplete tumor representation, leading to a potentially inaccurate analysis. Additionally, both renal imaging reports and biopsy results heavily rely on the proficiency of the medical practitioner or operator, impacting their accuracy. Therefore, the pursuit of a less invasive yet highly accurate method for renal tumor differentiation is imperative.

Radiomics can very efficiently extract large-scale imaging features from medical images, including first-order features and transformed filter-based features that cannot be identified by the naked eye, turning imaging data into

a high-resolution mineable resource for guiding clinical practice [14–17]. Many studies have reported that radiomics plays an important role in solid tumor diagnosis, biological characterization and prognosis prediction, and decision-making assistance [18, 19]. Moreover, several studies have shown that CT-based radiomics combined with machine learning (ML) has important advantages for renal tumor differentiation and risk and prognosis prediction [20–22]. Thus, it is important to investigate the utility of radiomics for improving and optimizing the CT-based differential diagnosis of renal tumors.

Although the aforementioned studies successfully constructed adequate radiomic-based models, they did not elucidate the influence of internal features on model performance. Nevertheless, it is imperative for a robust prediction model to possess the capacity for interpreting its inherent “black box” nature, a consideration of great significance. To address the challenge posed by the “black box” nature, the SHapley Additive exPlanations (SHAP) method was introduced to help clinicians understand the results of constructed models [23]. The sign of the SHAP value of a feature signifies the direction in which that feature exerts its influence, while the magnitude of the SHAP value denotes the “weight” or “significance” of the respective feature. Consequently, a number of researchers have embraced the SHAP method as a means to clarify the influence of radiomic features used for model construction on both overall and individual prediction outcomes [24–27].

Consequently, the primary objective of this study was to establish and evaluate radiomic models incorporating ML algorithms to differentiate between benign and malignant solid renal tumors via contrast-enhanced CT (CECT) scans. Furthermore, the SHAP model was employed to interpret our model results.

Materials and methods

Patients

This retrospective study received approval from the institutional review board (KY-2023–083-01), who waived the need for written informed consent.

The training set for this retrospective study was compiled from Center 1, spanning from May 2016 to September 2022, while the external validation set was sourced from Center 2, encompassing the period from August 2013 to November 2021. The inclusion criteria were as follows: (1) renal tumors diagnosed via complete pathology results and (2) available three-phase CECT imaging performed before surgery. The exclusion criteria were as follows: (1) unsatisfactory image quality for volume of interest (VOI) segmentation and (2) a history of treatments before CECT examination, such as

nephrectomy and biopsy. A schematic representation of the study population is provided in Fig. 1. Given the greater ease with which T3 RCC can be differentiated from any benign renal mass than T1a RCC can, we also selected patients with renal tumors measuring ≤ 4 cm in diameter for subanalysis.

Image acquisition and tumor segmentation

Axial three-phase scanning, encompassing the corticomedullary phase (CP), nephrographic phase (NP), and excretory phase (EP), was conducted for each patient using multislice spiral CT scanners. Supplementary Table 1 displays the CECT protocols used in this study. Renal CECT images were downloaded from the local medical image management system of the respective hospitals. The three-dimensional (3D) VOI of the renal tumor was manually delineated on the axial CP, NP, and EP images using ITK-SNAP 3.8.0 by a junior urologist (YH W, with 5 years of experience) and a senior urologist (SQ Z, with more than 10 years of experience). All VOIs were subsequently examined by a senior radiology professor (with more than 20 years of experience) and a senior urology professor (J P, with more than 20 years of experience). The VOI was manually outlined along the outermost boundaries of the tumor, with care taken to avoid adjacent normal tissue. If the patient had multiple renal tumors on one side, only the one with the largest diameter was outlined.

Feature extraction

A total of 1781 radiomic features were extracted from the 3D VOIs of the renal tumors on each set of CECT phase images using PyRadiomics (version 3.0.1) in Python 3.7.6. Voxel resampling and gray discretization were conducted to make the different pixel spacings of the CECT images from different CT scanners uniform between the patients in the two centers before feature extraction [28–30]. The images were resampled to a $1 \times 1 \times 1$ mm³ voxel size using nearest neighbor interpolation, and the bin width of gray discretization was set to 25. The extracted radiomic features were normalized utilizing the Z score method after feature extraction [31].

Feature selection

The minimum redundancy maximum relevance (mRMR) method was used to select the 10 most representative features. This algorithm can ensure the selection of highly relevant features with respect to real categories while simultaneously eliminating redundancy among the selected features [32–34]. Agglomerative hierarchical clustering was employed to generate a cluster heatmap illustrating disparities in the radiomic features selected from the CP, NP, and

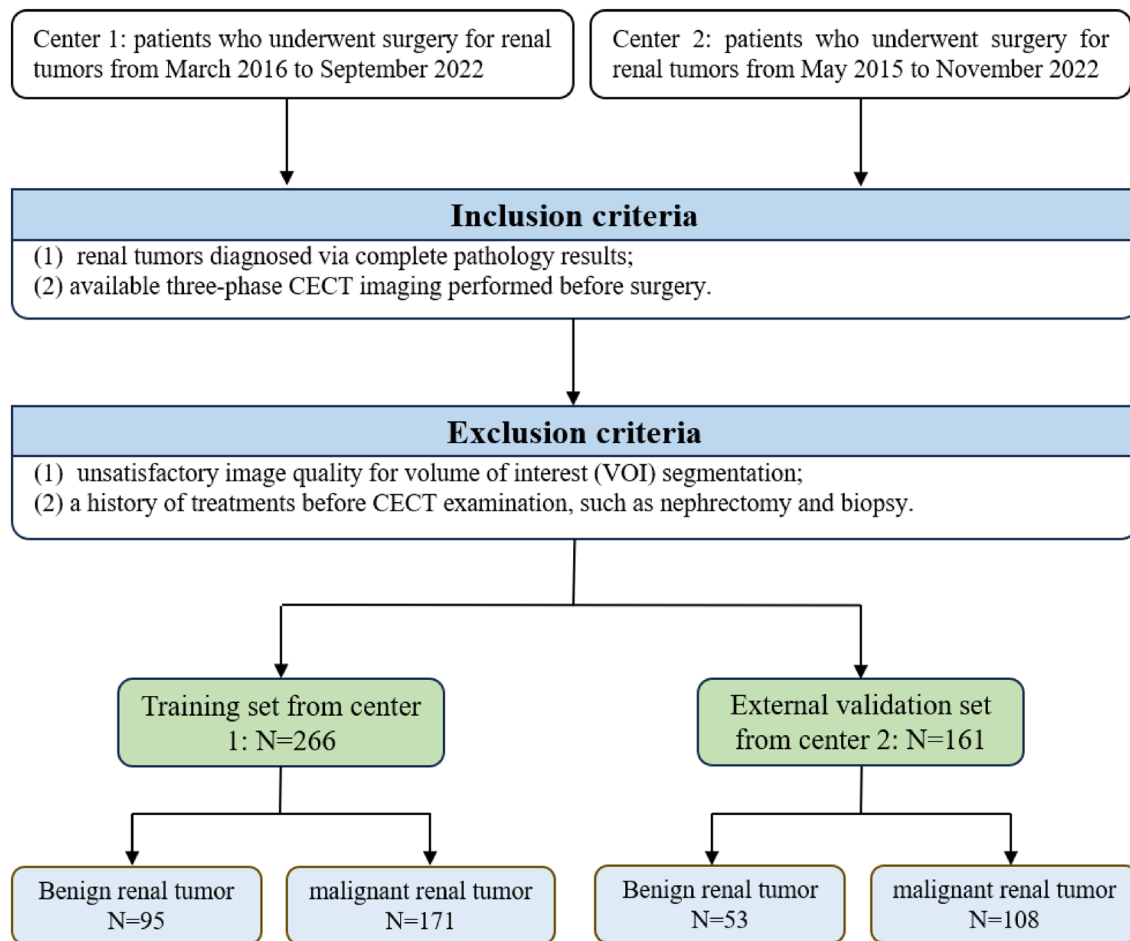


Fig. 1 Flow diagram for selection of the study population

EP images between the malignant and benign solid renal tumor groups and to visualize the relationships between subject clusters and these selected radiomic features.

Model construction and evaluation

Subsequently, random forest (RF) models were constructed using the Scikit-learn library version 1.0.2, employing the selected 10 features. The RandomizedSearchCV method was individually utilized to fine-tune the hyperparameters of the constructed models. A summary of the ML model development workflow is provided in Fig. 2.

Four RF models were constructed utilizing selected representative features derived from both the single phases (CP, NP, and EP) and the combination of all three phases (TP). The RF model based on TP features was constructed using the combination of the 10 features selected from the CP images, the 10 features selected from the NP images, and the 10 features selected from the EP images. These models were subsequently applied to the external validation set for evaluation. Model performance was evaluated through receiver

operating characteristic (ROC) curve analysis. Additionally, metrics including the area under the ROC curve (AUC), accuracy, sensitivity, and specificity were computed for each model.

Model interpretation

SHAP was employed to interpret the importance of the radiomic features incorporated within the best RF model. SHAP enables the visualization of feature importance within intricate ML-based models, helping elucidate how individual features within the model impact the likelihood of a particular output either positively or negatively.

Statistical analysis

Differences in demographic characteristics between the training set and the external validation set were evaluated using R 4.3.0. The χ^2 test was employed for comparing categorical variables, while the Mann–Whitney U test was applied for comparing nonnormally distributed continuous

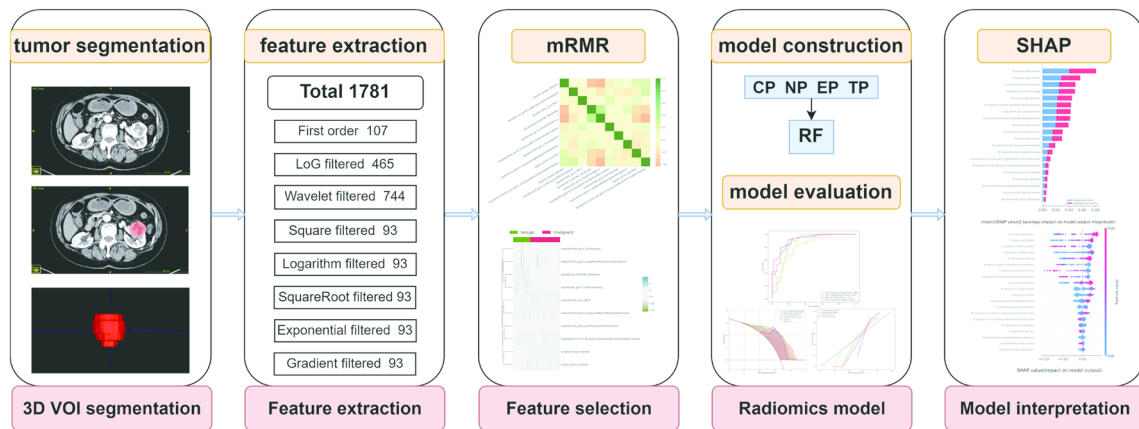


Fig. 2 Flow diagram of model construction and interpretation in this study

variables. A significance level of $P < 0.05$ indicated statistical significance [35].

Results

Clinical characteristics of the study population

In this study, 427 patients were enrolled. A total of 266 patients from Center 1 were included in the training set, and 161 patients from Center 2 were included in the external validation set. No significant differences were observed in the clinical characteristics (age, sex, laterality, tumor classification, or tumor size) between the training set and the external validation set. The detailed clinical characteristics of the study population are provided in Table 1. The detailed histologic distribution of renal tumors is described in Supplementary Table 2. The detailed stages and grades of RCC tumors are provided in Supplementary Table 3. The detailed

distribution of benign and malignant renal tumors in patients with tumors measuring ≤ 4 cm in diameter is shown in Supplementary Table 4.

Features selected by mRMR

The details of the 10 features selected by the mRMR method from the CP, NP, and EP images are provided in Supplementary Table 5. A cluster heatmap illustrating differences in radiomics features selected from CP, NP, and EP images between the malignant and benign renal tumor groups is shown in Fig. 3. The three most distinct features between the malignant and benign renal tumor groups are summarized in Table 2.

RF model construction and evaluation

The RF models were constructed according to the hyperparameters derived with the training set. Detailed

Table 1 Demographic characteristics of the study population

Characteristics		Training set (N = 266)	External validation set (N = 161)	P
Age (median [IQR])		49.5 [40.2, 59.0]	51.0 [42.0, 61.0]	0.263*
Sex (%)	Female	137 (51.50)	77 (47.83)	0.524**
	Male	129 (48.50)	84 (52.17)	
Laterality (%)	Left	146 (54.89)	86 (53.42)	0.845**
	Right	120 (45.11)	75 (46.58)	
Tumor classification (%)	Benign	95 (35.71)	53 (32.92)	0.629**
	Malignant	171 (64.29)	108 (67.08)	
Size (median [IQR])		4.2 [3.1, 6.0]	4.0 [2.8, 5.5]	0.199*

Note. The median (interquartile range, IQR) is shown for nonnormally distributed continuous variables. The number of patients (percentage, %) is shown for categorical variables

* P values were calculated utilizing the Mann–Whitney U test

** P values were computed using χ^2 tests

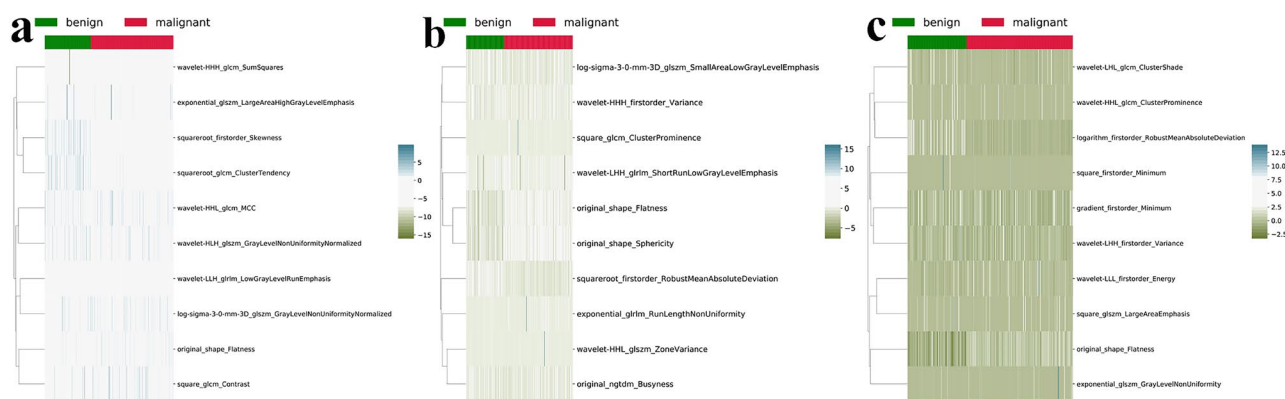


Fig. 3 Cluster heatmap of the 10 most representative features extracted from CP (a), NP (b) and EP features (c)

Table 2 The three most distinct features between the malignant and benign renal tumor groups for each phase of images

Phase	Feature name	Change
CP	original_shape_Flatness	higher
	squareroot_firstorder_Skewness	lower
	squareroot_glcM_ClusterTendency	lower
NP	squareroot_firstorder_RobustMeanAbsoluteDeviation	lower
	original_shape_Sphericity	higher
	original_shape_Flatness	higher
EP	original_shape_Flatness	higher
	logarithm_firstorder_RobustMeanAbsoluteDeviation	lower
	wavelet-LLL_firstorder_Energy	higher

Note. The term ‘Change’ denotes the change in the feature value in the malignant group with respect to the benign group

hyperparameter information is available in Supplementary Table 6.

The AUC, accuracy, specificity and sensitivity of the RF model based on CP were 0.886 (0.848–0.924), 0.838 (0.794–0.883), 0.832 (0.787–0.877) and 0.801 (0.753–0.849), respectively, in the training set. The AUC, accuracy, specificity and sensitivity of the RF model based on NP images were 0.912 (0.878–0.946), 0.842 (0.798–0.886), 0.737 (0.684–0.790) and 0.930 (0.899–0.961), respectively, in the training set. The AUC, accuracy, specificity and sensitivity of the RF model based on EP images were 0.930 (0.900–0.961), 0.786 (0.736–0.835), 0.821 (0.775–0.867) and 0.883 (0.844–0.922), respectively, in the training set. The AUC, accuracy, specificity and sensitivity of the RF model based on TP were 0.944 (0.916–0.972), 0.868 (0.828–0.909), 0.821 (0.775–0.867) and 0.906 (0.871–0.941), respectively, in the training set.

The constructed RF models based on CP, NP, EP, and TP features yielded AUC values of 0.860 (0.806–0.914), 0.821 (0.762–0.880), 0.921 (0.879–0.963), and 0.908

(0.864–0.953), respectively, in the external validation set. For the RF model based on CP features, the sensitivity, specificity, and accuracy were 0.889 (0.840–0.937), 0.774 (0.709–0.838), and 0.857 (0.803–0.911), respectively. The RF model based on NP features achieved sensitivity, specificity, and accuracy values of 0.843 (0.786–0.899), 0.717 (0.647–0.787), and 0.789 (0.726–0.852), respectively. Similarly, the RF model based on EP features exhibited good sensitivity (0.926 (0.885–0.966)), specificity (0.774 (0.709–0.838)), and accuracy (0.764 (0.698–0.830)). The RF model based on TP features achieved a sensitivity, specificity, and accuracy of 0.917 (0.874–0.959), 0.811 (0.751–0.872), and 0.826 (0.768–0.885), respectively.

Finally, the AUC of the RF model based on EP features (0.921) exceeded that based on CP (0.860), NP (0.821), and TP features (0.908) in the external validation set. The predictive performance of the four RF models is summarized in Table 3. The ROC curves, visualization of the results of decision curve analysis, and calibration curves of the RF models based on CP, NP, EP, and TP features are shown in Fig. 4 and Supplementary Fig. 1. The predictive performance of the four RF models for renal tumor patients with tumors measuring ≤ 4 cm in diameter is summarized in Supplementary Table 7. The ROC curves of the RF models based on CP, NP, EP, and TP features for renal tumor patients with tumors measuring ≤ 4 cm in diameter are also shown in Fig. 4. Our findings indicate that the performance of the best model based on EP features for distinguishing T1a RCC has the best AUC of 0.864 in the external validation set.

Enhancing model interpretability using SHAP values

The SHAP bar plot, which lists the most crucial features influencing the model output in descending order, and the SHAP beeswarm plot, illustrating the impacts of the features on RF model decisions and the interactions between these

Table 3 Performance of RF models in the training set and external validation set

Metric		AUC	ACC	Spe	Sen	PPV	NPV
Training set	CP	0.886 (0.848–0.924)	0.838 (0.794–0.883)	0.832 (0.787–0.877)	0.801 (0.753–0.849)	0.895 (0.859–0.932)	0.699 (0.644–0.754)
	NP	0.912 (0.878–0.946)	0.842 (0.798–0.886)	0.737 (0.684–0.790)	0.930 (0.899–0.961)	0.864 (0.823–0.905)	0.854 (0.811–0.896)
	EP	0.930 (0.900–0.961)	0.786 (0.736–0.835)	0.821 (0.775–0.867)	0.883 (0.844–0.922)	0.899 (0.863–0.935)	0.796 (0.747–0.844)
	TP	0.944 (0.916–0.972)	0.868 (0.828–0.909)	0.821 (0.775–0.867)	0.906 (0.871–0.941)	0.901 (0.865–0.937)	0.830 (0.785–0.875)
External validation set	CP	0.860 (0.806–0.914)	0.857 (0.803–0.911)	0.774 (0.709–0.838)	0.889 (0.840–0.937)	0.889 (0.840–0.937)	0.774 (0.709–0.838)
	NP	0.821 (0.762–0.880)	0.789 (0.726–0.852)	0.717 (0.647–0.787)	0.843 (0.786–0.899)	0.858 (0.805–0.912)	0.691 (0.620–0.762)
	EP	0.921 (0.879–0.963)	0.764 (0.698–0.830)	0.774 (0.709–0.838)	0.926 (0.885–0.966)	0.893 (0.845–0.941)	0.837 (0.780–0.894)
	TP	0.908 (0.864–0.953)	0.826 (0.768–0.885)	0.811 (0.751–0.872)	0.917 (0.874–0.959)	0.908 (0.864–0.953)	0.827 (0.768–0.885)

Note. The data in brackets are the 95% CIs of the corresponding values

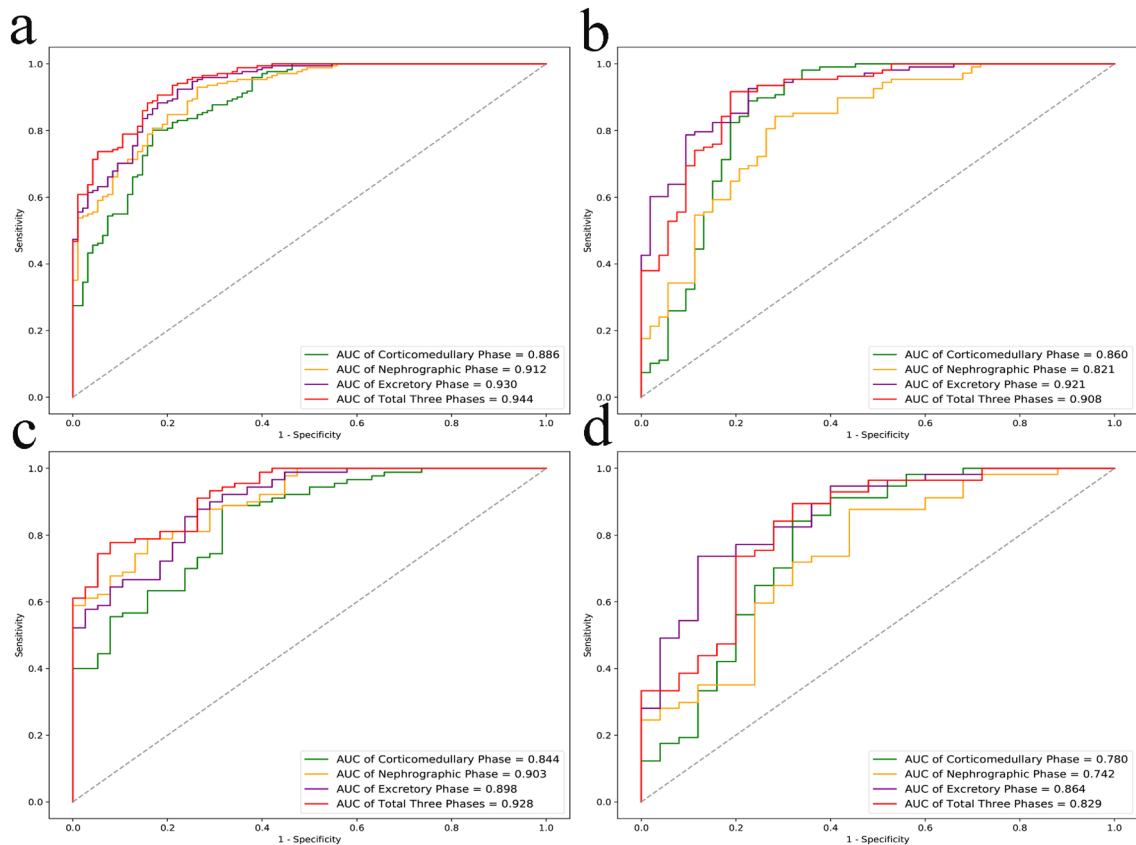
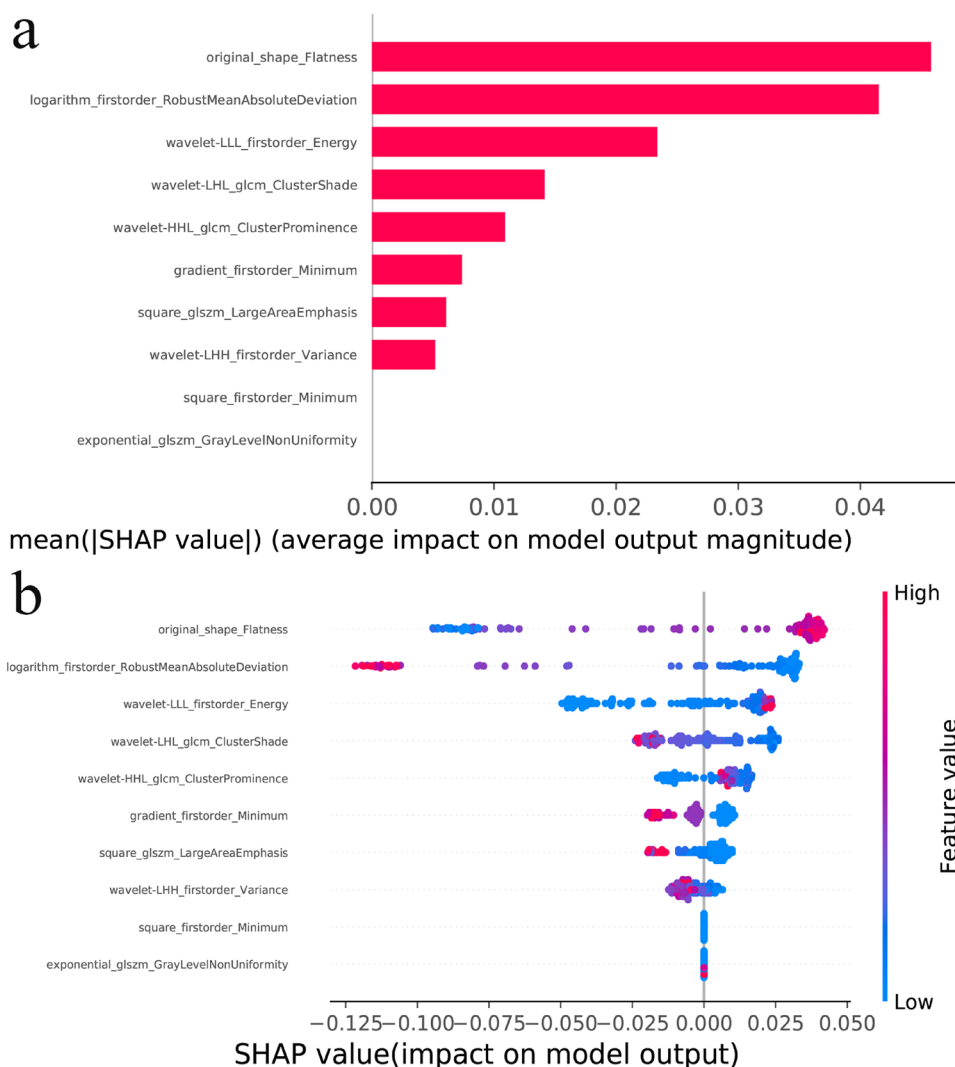


Fig. 4 ROC curves of the RF models for all benign and malignant renal tumors in the training set (a) and external validation set (b). ROC curves of the RF models for benign and malignant renal tumors measuring ≤ 4 cm in diameter in the training set (c) and external validation set (d)

features in the model, are shown in Fig. 5. As depicted in the bar plots, the top feature, “original_shape_Flatness”, exhibited the greatest contribution to the model output

and possessed stronger predictive capabilities than did the lower-ranked features. As demonstrated in the beeswarm plots, a positive SHAP value for “original_shape_Flatness”

Fig. 5 SHAP values for the interpretation of the RF models. Bar plot (a) and beeswarm plot (b) showing the SHAP values of the RF model based on EP features in the external validation set



corresponds to an increased likelihood of a malignant renal tumor for each prediction, while a negative value indicates the opposite. The magnitude of the value is directly proportional to the (increased or decreased) risk of malignancy. The top three feature importance rankings of the RF model based on EP features are summarized in Table 4.

Discussion

In this study, four RF models were developed based on CP, NP, EP, and TP features to preoperatively distinguish between benign and malignant renal tumors in patients who underwent CECT scans for solid renal tumor assessment. The RF models were constructed with the features from the training set, and their discriminative capabilities were demonstrated in an external validation set, achieving AUC values ranging from 0.821 to 0.921 and accuracies ranging from 76.4% to 85.7%. Among these models, the

Table 4 The most influential features according to the SHAP method

Phase	SHAP	Feature name	Change	Significance
EP	top 1	original_shape_Flatness	higher	increased risk
	top 2	logarithm_firstorder_RobustMeanAbsoluteDeviation	lower	decreased risk
	top 3	wavelet-LLL_firstorder_Energy	higher	increased risk

Note. The column 'Change' denotes the change in SHAP value in the malignant group with respect to the benign group. The column 'Significance' denotes that a change in the feature direction results in an increased or decreased risk of malignancy

RF model based on EP image features exhibited the highest diagnostic performance in the external validation set, with an AUC of 0.921. These results indicate that ML-based radiomic models could offer a straightforward, non-invasive, and highly accurate approach for preoperatively

predicting whether a solid renal tumor is benign or malignant with CECT images.

Previous investigations have assessed the potential of radiomic models based on CECT, MRI, and ultrasound for distinguishing benign from malignant renal tumors [36–40]. Nevertheless, these studies exhibited limitations that we aimed to address in this study. Alhussaini et al. concluded that ML-based radiomic analysis holds substantial promise for effectively distinguishing chRCC from RO when employing CECT images (37 chRCC and 41 RO) [38]. Their model yielded the best AUC of 1.00, but the dataset was small, and external validation was lacking. Wang et al. randomly divided 190 RCC patients (147 ccRCC patients and 43 non-ccRCC patients) from one center into two groups (a training set and a testing set at a ratio of 7:3) and combined three ML algorithms to develop radiomic models, which demonstrated promising potential in differentiating ccRCC patients from nonccRCC patients [39]. The best AUC achieved by their models was 0.909, but again, the dataset was small, and external validation was lacking. Moreover, when patients with renal tumors seek medical attention, their primary concern is whether the tumor is benign or malignant. Consequently, the most critical aspect is the pre-operative differential diagnosis of their renal tumors. Wentland et al. identified 148 solid renal tumors (50 benign: 23 AML, 27 RO; 98 malignant: 23 ccRCC, 44 pRCC, 31 chRCC) and parsed them into two sets (training/testing at a ratio of 7:3) to construct an RF model based on CECT images, which achieved an AUC of 0.80 (with a superior accuracy of 0.82) in distinguishing benign from malignant renal tumors, outperforming 3 radiologists (accuracies from 0.67 to 0.75) [36]. Although the topic of the study was the differentiation of benign and malignant renal tumors, the dataset was still small and lacked an external validation set, and the constructed model lacked interpretability. Massa' a et al. included 182 renal tumors in 160 patients and divided them into training (70%) and testing (30%) sets to assess the performance of ML models trained with MRI-based radiomic features in distinguishing benign from malignant solid renal masses. The model, based on T2-weighted MR images, achieved a high level of accuracy (0.80, with an AUC of 0.79); however, the dataset was still small and lacked an external validation set, and the constructed model lacked interpretability. [37]. Erdim et al. included 79 patients with 84 solid renal tumors and combined 8 ML algorithms to distinguish between benign and malignant renal tumors [40]. In that study, the best model had an AUC of 0.916 and an accuracy of 0.917, which suggested that CT radiomics, in conjunction with ML algorithms, has utility in the noninvasive discrimination of benign and malignant renal tumors. However, the dataset was still small and lacked an external validation set, and the constructed model lacked interpretability. In summary, the abovementioned studies only

involved relatively small sample sizes, typically involving fewer than 160 patients, and often lacked an independent validation set and model interpretation. Consequently, the generalizability of the aforementioned research findings may be limited; additional efforts to enhance generalizability would necessitate larger sample sizes, including an independent external validation set, and model interpretation.

It is challenging for clinicians to determine how models draw conclusions and to identify radiomic features that are critical in decision making. In our research, we found that `original_shape_Flatness`, a 3D shape feature, was the most important feature in informing the prediction outcome of the model according to the SHAP method. The term "flatness" is employed to describe the geometric characteristics of an object, specifically referring to the ratio between its width and length. More precisely, it can be defined as the ratio between the maximum diameter of the target and the minimum diameter perpendicular to it. The values for flatness range from 0 (indicating a perfectly flat shape) to 1 (representing a nonflat, spherical shape). The `original_shape_Flatness` eigenvalues are computed based on the morphology of the original VOI. In this study, we observed that malignant tumors exhibited greater values for `original_shape_Flatness` than did benign tumors. This finding aligns with clinical practice, where irregularly shaped tumors are typically considered malignant based on CECT imaging results.

Our research exhibits several strengths to previous studies. First, our study had a larger sample size ($n=427$) than the investigations described above. Moreover, we incorporated an external validation dataset ($n=161$) to validate the models developed from our training set ($n=266$). Second, we harnessed radiomic features extracted from three distinct phases of CECT imaging—CP, NP, and EP imaging—to construct three RF models. Additionally, a fourth RF model was formulated by amalgamating the radiomic features across all phases, denoted as the TP model. The most significant advantage of this study lies in the utilization of the mRMR method for feature selection, coupled with the use of SHAP values to improve model interpretability. The utilization of the mRMR method ensures that the chosen features exhibit the highest correlations with the actual classifications while minimizing redundancy. Notably, the most distinctive features selected by the mRMR method align with the top features identified by SHAP as exerting the greatest influence on the prediction outcomes. As a result of these aforementioned strengths, our research findings possess greater generalizability and practicality.

Although the findings presented in this study provide promising insights into distinguishing renal tumors, several limitations must be acknowledged. First, the retrospective nature of this study introduced inherent selection bias, and the sample size was insufficiently large. Consequently, future research should consider conducting a multicenter

study with a larger population to further assess the proposed model. Additionally, prospective studies could be conducted to confirm the robustness of the proposed model. Second, all VOIs in this study were manually segmented, which is a time-intensive process. The use of automatic segmentation techniques is warranted to enhance segmentation efficiency and reduce labor costs in the future.

In summary, all four RF models efficiently differentiated benign from malignant solid renal tumors, and the RF model based on EP features displayed the best performance. The feature “original_shape_Flatness” played the greatest role in predicting the outcome of RF model based on EP image features.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00261-024-04351-3>.

Funding This study received financial support from the National Natural Science Foundation of China (Grant No. 82272689), Sanming Project of Medicine in Shenzhen (Grant No. SZSM202011011), and Research Start-up Fund of the Seven Affiliated Hospital, Sun Yat-sen University (Grant No. ZSQYJZPI202003).

Declarations

Conflicts of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A: **Cancer statistics, 2023**. *CA: a cancer journal for clinicians* 2023, **73**(1):17-48.
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM: **The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours**. *Eur Urol* 2016, **70**(1):93-105.
3. Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, Fernández-Pello S, Giles RH, Hofmann F, Hora M *et al*: **European Association of Urology Guidelines on Renal Cell Carcinoma: The 2022 Update**. *Eur Urol* 2022, **82**(4):399-410.
4. Young JR, Margolis D, Sauk S, Pantuck AJ, Sayre J, Raman SS: **Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT**. *Radiology* 2013, **267**(2):444-453.
5. Hodgdon T, McInnes MDF, Schieda N, Flood TA, Lamb L, Thornhill RE: **Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images?** *Radiology* 2015, **276**(3):787-796.
6. Vogel C, Ziegelmüller B, Ljungberg B, Bensalah K, Bex A, Canfield S, Giles RH, Hora M, Kuczyk MA, Merseburger AS *et al*: **Imaging in Suspected Renal-Cell Carcinoma: Systematic Review**. *Clinical genitourinary cancer* 2019, **17**(2):e345-e355.
7. Rossi SH, Prezzi D, Kelly-Morland C, Goh V: **Imaging for the diagnosis and response assessment of renal tumours**. *World journal of urology* 2018, **36**(12):1927-1942.
8. Zhang J, Tehrani YM, Wang L, Ishill NM, Schwartz LH, Hricak H: **Renal masses: characterization with diffusion-weighted MR imaging--a preliminary experience**. *Radiology* 2008, **247**(2):458-464.
9. Hecht EM, Israel GM, Krinsky GA, Hahn WY, Kim DC, Belit-skaya-Levy I, Lee VS: **Renal masses: quantitative analysis of enhancement with signal intensity measurements versus qualitative analysis of enhancement with image subtraction for diagnosing malignancy at MR imaging**. *Radiology* 2004, **232**(2):373-378.
10. Zhang J, Lefkowitz RA, Ishill NM, Wang L, Moskowitz CS, Russo P, Eisenberg H, Hricak H: **Solid renal cortical tumors: differentiation with CT**. *Radiology* 2007, **244**(2):494-504.
11. Dyer R, DiSantis DJ, McClennan BL: **Simplified imaging approach for evaluation of the solid renal mass in adults**. *Radiology* 2008, **247**(2):331-343.
12. Marconi L, Dabestani S, Lam TB, Hofmann F, Stewart F, Norrie J, Bex A, Bensalah K, Canfield SE, Hora M *et al*: **Systematic Review and Meta-analysis of Diagnostic Accuracy of Percutaneous Renal Tumour Biopsy**. *Eur Urol* 2016, **69**(4):660-673.
13. Leveridge MJ, Finelli A, Kachura JR, Evans A, Chung H, Shift DA, Fernandes K, Jewett MA: **Outcomes of small renal mass needle core biopsy, nondiagnostic percutaneous biopsy, and the role of repeat biopsy**. *Eur Urol* 2011, **60**(3):578-584.
14. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B: **Radiomics in medical imaging--“how-to” guide and critical reflection**. *Insights into imaging* 2020, **11**(1):91.
15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A *et al*: **Radiomics: extracting more information from medical images using advanced feature analysis**. *Eur J Cancer* 2012, **48**(4):441-446.
16. Gillies RJ, Kinahan PE, Hricak H: **Radiomics: Images Are More than Pictures, They Are Data**. *Radiology* 2016, **278**(2):563-577.
17. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A *et al*: **Radiomics: the bridge between medical imaging and personalized medicine**. *Nat Rev Clin Oncol* 2017, **14**(12):749-762.
18. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A: **Predicting cancer outcomes with radiomics and artificial intelligence in radiology**. *Nature Reviews Clinical Oncology* 2021, **19**(2):132-146.
19. Zhang X, Ruan S, Xiao W, Shao J, Tian W, Liu W, Zhang Z, Wan D, Huang J, Huang Q *et al*: **Contrast-enhanced CT radiomics for preoperative evaluation of microvascular invasion in hepatocellular carcinoma: A two-center study**. *Clinical and translational medicine* 2020, **10**(2):e111.
20. Uhlig J, Leha A, Delonge LM, Haack AM, Shuch B, Kim HS, Bremmer F, Trojan L, Lotz J, Uhlig A: **Radiomic Features and Machine Learning for the Discrimination of Renal Tumor Histological Subtypes: A Pragmatic Study Using Clinical-Routine Computed Tomography**. *Cancers (Basel)* 2020, **12**(10):3010.
21. Nazari M, Shiri I, Zaidi H: **Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients**. *Computers in biology and medicine* 2021, **129**:104135.

22. Nazari M, Shiri I, Hajianfar G, Oveisi N, Abdollahi H, Deevband MR, Oveisi M, Zaidi H: **Noninvasive Fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomic features and machine learning.** *La Radiologia medica* 2020, **125**(8):754-762.
23. Hung PS, Lin PR, Hsu HH, Huang YC, Wu SH, Kor CT: **Explainable Machine Learning-Based Risk Prediction Model for In-Hospital Mortality after Continuous Renal Replacement Therapy Initiation.** *Diagnostics (Basel, Switzerland)* 2022, **12**(6):1496.
24. Li C, He Z, Lv F, Liu Y, Hu Y, Zhang J, Liu H, Ma S, Xiao Z: **An interpretable MRI-based radiomics model predicting the prognosis of high-intensity focused ultrasound ablation of uterine fibroids.** *Insights into imaging* 2023, **14**(1):129.
25. Bang M, Park YW, Eom J, Ahn SS, Kim J, Lee SK, Lee SH: **An interpretable radiomics model for the diagnosis of panic disorder with or without agoraphobia using magnetic resonance imaging.** *Journal of affective disorders* 2022, **305**:47-54.
26. Yang H, Wu K, Liu H, Wu P, Yuan Y, Wang L, Liu Y, Zeng H, Li J, Liu W *et al*: **An automated surgical decision-making framework for partial or radical nephrectomy based on 3D-CT multi-level anatomical features in renal cell carcinoma.** *Eur Radiol* 2023, **33**(11):7532-7541.
27. Ye JY, Fang P, Peng ZP, Huang XT, Xie JZ, Yin XY: **A radiomics-based interpretable model to predict the pathological grade of pancreatic neuroendocrine tumors.** *Eur Radiol* 2024, **34**(3):1994-2005.
28. Hong JH, Jung JY, Jo A, Nam Y, Pak S, Lee SY, Park H, Lee SE, Kim S: **Development and Validation of a Radiomics Model for Differentiating Bone Islands and Osteoblastic Bone Metastases at Abdominal CT.** *Radiology* 2021, **299**(3):626-632.
29. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao D, Lancia A, Ammari S *et al*: **A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study.** *Lancet Oncol* 2018, **19**(9):1180-1191.
30. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D *et al*: **Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels.** *Med Phys* 2017, **44**(3):1050-1062.
31. Jiang YW, Xu XJ, Wang R, Chen CM: **Radiomics analysis based on lumbar spine CT to detect osteoporosis.** *Eur Radiol* 2022, **32**(11):8019-8026.
32. Yi Z, Hu S, Lin X, Zou Q, Zou M, Zhang Z, Xu L, Jiang N, Zhang Y: **Machine learning-based prediction of invisible intraprostatic prostate cancer lesions on 68 Ga-PSMA-11 PET/CT in patients with primary prostate cancer.** *Eur J Nucl Med Mol Imaging* 2021, **49**:1523-1534.
33. Zamboglou C, Carles M, Fechter T, Kiefer S, Reichel K, Fassbender TF, Bronsert P, Koeber G, Schilling O, Ruf J *et al*: **Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate- and high-risk prostate cancer - a comparison study with histology reference.** *Theranostics* 2019, **9**(9):2595-2605.
34. Gong L, Xu M, Fang M, Zou J, Yang S, Yu X, Xu D, Zhou L, Li H, He B *et al*: **Noninvasive Prediction of High-Grade Prostate Cancer via Biparametric MRI Radiomics.** *Journal of magnetic resonance imaging : JMIR* 2020, **52**(4):1102-1109.
35. Chia CS, Wong LCK, Henedige TP, Ong WS, Zhu HY, Tan GHC, Kwek JW, Seo CJ, Wong JSM, Ong CJ *et al*: **Prospective Comparison of the Performance of MRI Versus CT in the Detection and Evaluation of Peritoneal Surface Malignancies.** *Cancers (Basel)* 2022, **14**(13):3179.
36. Wentland AL, Yamashita R, Kino A, Pandit P, Shen L, Brooke Jeffrey R, Rubin D, Kamaya A: **Differentiation of benign from malignant solid renal lesions using CT-based radiomics and machine learning: comparison with radiologist interpretation.** *Abdominal radiology (New York)* 2023, **48**(2):642-648.
37. Massa'a RN, Stoeckl EM, Lubner MG, Smith D, Mao L, Shapiro DD, Abel EJ, Wentland AL: **Differentiation of benign from malignant solid renal lesions with MRI-based radiomics and machine learning.** *Abdominal Radiology* 2022, **47**(8):2896-2904.
38. Alhussaini AJ, Steele JD, Nabi G: **Comparative Analysis for the Distinction of Chromophobe Renal Cell Carcinoma from Renal Oncocytoma in Computed Tomography Imaging Using Machine Learning Radiomics Analysis.** *Cancers (Basel)* 2022, **14**(15):3609.
39. Wang P, Pei X, Yin XP, Ren JL, Wang Y, Ma LY, Du XG, Gao BL: **Radiomics models based on enhanced computed tomography to distinguish clear cell from non-clear cell renal cell carcinomas.** *Sci Rep* 2021, **11**(1):13729.
40. Erdim C, Yardimci AH, Bektas CT, Kocak B, Koca SB, Demir H, Kilickesmez O: **Prediction of Benign and Malignant Solid Renal Masses: Machine Learning-Based CT Texture Analysis.** *Academic radiology* 2020, **27**(10):1422-1429.

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