European Radiology

EUROPEAN SOCIETY OF RADIOLOGY

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Distinguishing common renal cell carcinomas from benign renal tumors based on machine learning: comparing various CT imaging phases, slices, tumor sizes, and ROI segmentation strategies

Tao Zhou [1,2] · Jian Guan [2] · Bao Feng [1,3] · Huimin Xue [1,3] · Jin Cui [1] · Qionglian Kuang [1,2] · Yehang Chen [4] · Kuncai Xu [4] · Fan Lin [5] · Enming Cui [1,3] · Wansheng Long [1,3]

Abstract

Objectives To determine whether a CT-based machine learning (ML) can differentiate benign renal tumors from renal cell carcinomas (RCCs) and improve radiologists' diagnostic performance, and evaluate the impact of variable CT imaging phases, slices, tumor sizes, and region of interest (ROI) segmentation strategies.

Methods Patients with pathologically proven RCCs and benign renal tumors from our institution between 2008 and 2020 were included as the training dataset for ML model development and internal validation (including 418 RCCs and 78 benign tumors), and patients from two independent institutions and a public database (TCIA) were included as the external dataset for individual testing (including 262 RCCs and 47 benign tumors). Features were extracted from three-phase CT images. CatBoost was used for feature selection and ML model establishment. The area under the receiver operating characteristic curve (AUC) was used to assess the performance of the ML model.

Results The ML model based on 3D images performed better than that based on 2D images, with the highest AUC of 0.81 and accuracy (ACC) of 0.86. All three radiologists achieved better performance by referring to the classifier's decision, with accuracies increasing from 0.82 to 0.87, 0.82 to 0.88, and 0.76 to 0.87. The ML model achieved higher negative predictive values (NPV, 0.82–0.99), and the radiologists achieved higher positive predictive values (PPV, 0.91–0.95).

Conclusions A ML classifier based on whole-tumor three-phase CT images can be a useful and promising tool for differentiating RCCs from benign renal tumors. The ML model also perfectly complements radiologist interpretations.

Key Points
• *A machine learning classifier based on CT images could be a reliable way to differentiate RCCs from benign renal tumors.*
• *The machine learning model perfectly complemented the radiologists' interpretations.*
• *Subtle variances in ROI delineation had little effect on the performance of the ML classifier.*

Keywords Renal cell carcinoma · Angiomyolipoma · Oncocytoma · Machine learning · Artificial intelligence

---

Tao Zhou and Jian Guan contributed equally to this study.

✉ Fan Lin
foxetfoxet@gmail.com

✉ Enming Cui
cem2008@163.com

✉ Wansheng Long
jmlws2@163.com

1 Department of Radiology, Jiangmen Central Hospital, Guangdong Medical University, Zunyi Medical University, 23 Beijie Haibang Street, Jiangmen 529030, People's Republic of China

2 Department of Radiology, The First Affiliated Hospital of Sun Yat-Sen University, 58 Zhongshan Second Road, Guangzhou 510000, People's Republic of China

3 Guangzhou Key Laboratory of Molecular and Functional Imaging for Clinical Translation, Guangzhou, People's Republic of China

4 School of Electronic Information and Automation, Guilin University of Aerospace Technology, 2 Jinji Road, Guilin 541000, People's Republic of China

5 Department of Radiology, The First Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen Second People's Hospital, 3002 SunGangXi Road, Shenzhen 518035, People's Republic of China

Abbreviations

ACC	Accuracy
AML	Angiomyolipoma
ASL	Arterial spin labeling
CECT	Contrast-enhanced computed tomography
CMP	Corticomedullary phase
DWI	Diffusion-weighted imaging
GLCM	Gray-level cooccurrence matrix
GLDM	Gray-level dependence matrix
GLRLM	Gray-level run-length matrix
GLSZM	Gray-level size-zone matrix
ML	Machine learning
NGTDM	Neighborhood gray-tone difference matrix
NP	Nephrographic phase
PCP	Precontrast phase
RCC	Renal cell carcinoma
ROI	Region of interest

Introduction

In recent years, the number of identified solid renal masses has significantly increased due to the rapid development of advanced imaging techniques [1], especially in the detection of small renal masses (measuring 4 cm or less) [2, 3]. However, among surgically resected renal masses, approximately 20% are reported to be benign [4], leading to an overall increase in health costs and risks to the patients resulting from overtreatment.

In clinical practice, percutaneous biopsy is the gold standard for differentiating renal masses. However, as it is an invasive method, concern about the high risk of hemorrhage or infection remains [3, 5]. Conventional computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound are used as noninvasive preoperative imaging methods due to their safety and availability. Recently, the functional assessments provided by diffusion-weighted imaging (DWI), arterial spin labeling (ASL), and dynamic contrast enhancement (DCE) imaging have also been considered in the analysis of renal masses [6]. However, these techniques have limited accuracy in the characterization of some benign tumors, such as oncocytomas and fat-poor angiomyolipomas (AMLs), due to their similarity with malignant renal masses in the resulting imaging, limiting their ability to reliably differentiate benign from malignant tumors [7]. In addition, the diagnostic performance is significantly influenced by the experience of the radiologist, which is a major challenge for nonspecialized medical centers.

Machine learning (ML) algorithms have been shown to be valuable in the evaluation of the histopathological characteristics of the disease [8, 9]. Recently, a growing number of studies have shown that ML models have promising performance in predicting the grade of renal tumors and the outcome after renal tumor resection and in diagnosing incidental renal lesions [10–13], However, these studies are confined either by a small population or a lack of external testing data and contrast-enhanced CT (CECT) examinations. In addition, previous studies have focused little on the influence of the ML model itself on the diagnostic decision-making of the radiologist.

This study aimed to evaluate the performance of a CT-based ML model in discriminating benign renal tumors (including AMLs without visible fat (AMLwvf) and oncocytomas) from common renal cell carcinomas (including clear cell RCCs (ccRCCs), chromophobe RCCs (chrRCCs), and papillary RCCs (pRCCs)) in a large population along with various factor analysis, and to discuss the role that the model plays in radiologists' diagnostic decision-making in routine clinical practice.

Materials and methods

This retrospective study was approved by the ethics committee of our institution, and the requirement for informed consent was waived because the data were obtained from preexisting institutional and public databases without additional burdens to the patients.

Patients

The institutional pathology database was queried to identify pathologically confirmed renal masses obtained via biopsy or surgical resection in our hospital between 2008 and 2020. The pathological diagnosis was reconfirmed by a pathologist with 10 years of genitourinary experience. The inclusion criteria were as follows: (1) preoperative CT scans with three-phase imaging and (2) a primary lesion that was pathologically confirmed. The exclusion criteria were as follows: (1) significant lesion rupture with abundant hemorrhage leading to obscured tumor features; (2) incorrect delay times after contrast injection on CT study; (3) lesions completely composed of cystic components; and (4) lesions with visible fat on precontrast-phase (PCP) CT. The patient inclusion and exclusion flowchart is shown in Fig. 1.

Finally, 798 patients with 680 RCCs (533 ccRCCs, 69 chrRCCs, and 78 pRCCs) and 125 benign renal lesions (83 AMLwvf and 42 oncocytomas) were included in this study. Among these lesions, 418 RCCs and 78 benign lesions from our institution were included as the training dataset for ML model development and internal validation, and 262 RCCs and 47 benign lesions from two independent institutions and a public database (The Cancer Imaging Archive, TCIA) were included as the external dataset for individual testing.

Fig. 1 Flow chart of the study population (*7 patients with more than one lesion)
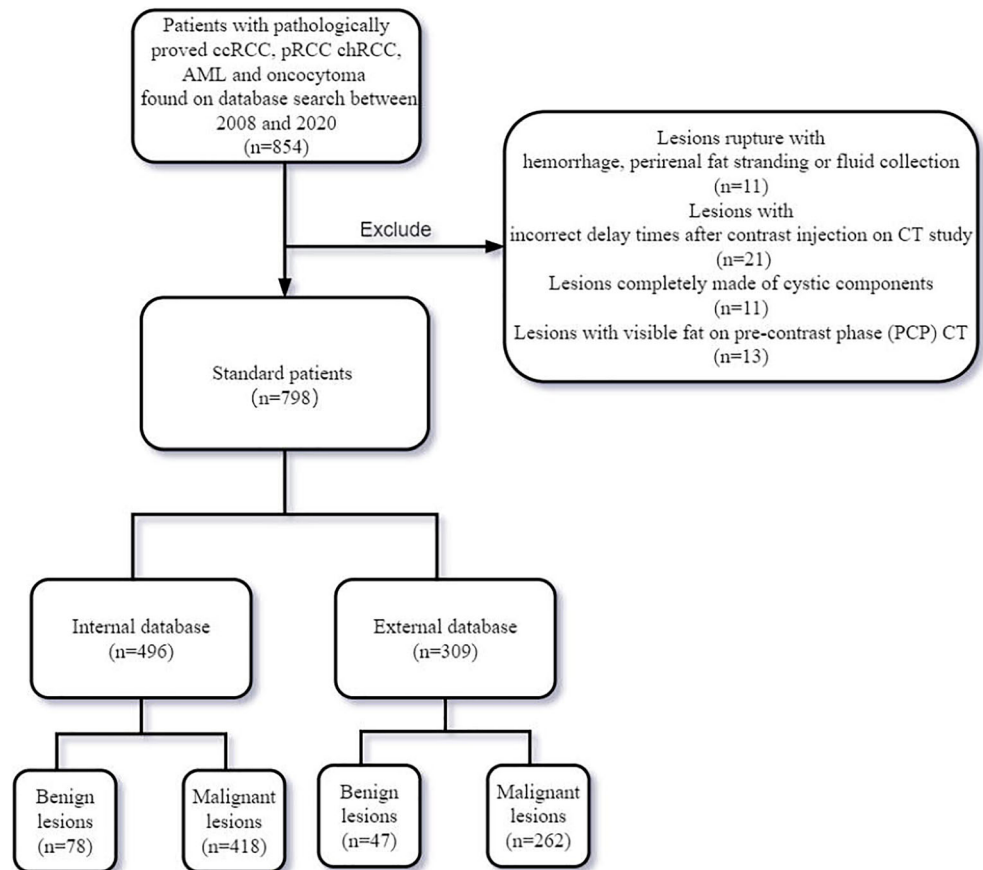
CT examination

CT examinations were performed in our institution using a 16-detector CT (SOMATOM Sensation 16, Siemens Healthineers), a 64-detector CT (Aquilion 64, Canon Medical Systems), or a dual-source CT (SOMATOM Force, Siemens Healthineers) with the same examination protocol before the patient underwent surgical treatment. The CT parameters and scanning protocol were as follows: tube voltage, 120 kVp; effective tube current-exposure time product, 200–350 mAs; matrix, 512 × 512; and slice thickness, 1.0 or 3.0 mm. Three phases were scanned: the PCP, corticomedullary phase (CMP: 30-s delay after contrast injection), and nephrographic phase (NP: 90-s delay after contrast injection). A total volume of 70–100 mL of contrast medium (Iopamidol, BRACCO) was intravenously injected at a rate of 3.0 mL/s, followed by flushing with 20 mL saline.

Image preprocessing

The image preprocessing steps, which included normalization, pixel resampling, and discretization, were performed during feature extraction for all data. Normalization aims to manage data weight inconsistency. Image normalization was conducted using the z score. The voxel size was defined as $0.8 \times 0.8 \times 0.8$ mm in resampling. Pixel resampling can improve the accuracy and population parameter estimation of the mode, including upsampling and downsampling. Discretization is the classification of continuous features into discrete feature values; a typical application of discretization is the binarization of gray images. A bin width of 25 was used for discretization in our study.

Tumor segmentation

After retrieving and acquiring the images of all patients from our institutional picture archiving and communication system (PACS), we loaded the images into ITK-SNAP [14] (Version 3.6.0) and then anonymized and stored the images in NIfTI format. Spatial matching and segmentation were then performed on the tumor images. Through subtle spatial adjustment on the three-phase images, a preliminarily defined region of interest (ROI) was carefully delineated on each selected slice to cover the whole tumor, by three radiologists with 5, 7, and 10 years of experience in abdominal radiology. The effective margins of the ROIs were reconfirmed by another two senior radiologists with 13 and 15 years of abdominal radiologic experience.

Feature extraction

We chose the PCP, CMP, and NP images for feature analysis, and the corresponding ROIs were determined automatically using the Python (version 3.6.5) package "PyRadiomics" [15]. The extracted texture features included the following: (i) first-order features; (ii) shape features; (iii) gray-level cooccurrence matrix (GLCM) features; (iv) gray-level size-zone matrix (GLSZM) features; (v) gray-level run-length matrix (GLRLM) features; (vi) neighborhood gray-tone difference matrix (NGTDM) features; and (vii) gray-level dependence matrix (GLDM) features [16–19]. The definition and mathematical formulas for these features have been described previously [20]. By using different filters (e.g., wavelet, Laplacian of Gaussian, square, square root, logarithm, and exponential filters), the final images were obtained. All features apart from the shape features, which are independent descriptors extracted from the label mask, were calculated on both the original and derived images [15]. For the three-dimensional tumor segmentations, we also extracted the texture features mentioned above.

Feature extraction was performed based on the two groups of ROIs by two independent senior radiologists. The selected features that had good to excellent reliability (ICC ≥ 0.75) were included for model development.

ML model

A gradient boosting decision library based on decision trees, CatBoost [21, 22], was used for feature selection and predictive model establishment based on the single and all-phase images. Fivefold cross-validation was performed to evaluate the average value and standard deviation of each performance indicator. Texture analysis and ML were performed using Python (version 3.6.5, www.python.org). For 2D model development, only the largest tumor slice was used. In contrast, every renal tumor slice was included for 3D model development. The tuning parameters were as follows: the

learning rate was set to 0.05, and the loss function was logloss. Given the imbalance of the data, the weights of the class were 5:1.

To explore the influence of the segmentation strategy on the ability of the ML model to discriminate malignant from benign tumors, the ROI was expanded or shrunk 1 mm or 3 mm based on the delineated tumor contour. In addition, we tested the performance of the ML model with tumors smaller than 3 cm from the external database as an independent validation group to confirm the practical value of the ML model in identifying small tumors. Figure 2 shows the process of ML model development and validation in our study.

Subjective radiologist assessment

The radiological analysis was performed by three radiologists in our institution who have interpreted images from more than 2000 urologic cases, all of whom were blinded to the histopathologic data and clinical information.

By using a 10-point scale scoring system, the readers assessed the likelihood for each renal lesion using the following scoring points as described previously: shape (regular = 0, irregular = 1), lesion heterogeneity (homogeneous = 0, heterogeneous = 1), internal septa (absent = 0, present = 1), extrarenal extension (absent = 0, ≤ 50% extension of renal contour = 1, ≥ 50% extension of renal contour = 2), internal calcification (absent = 0, present = 1), internal hemorrhage (absent = 0, present = 1), internal necrosis (absent = 0, present = 1), internal arteries (absent = 0, present = 1), and pseudo-capsule (absent = 0, present = 1), which appeared as a hypointense ring around the lesion contour consisting of a fibrous structure that was formed by compression of the distended growth of the renal mass to the surrounding renal parenchyma. Higher scores (≥ 6) indicated a greater possibility of malignancy, and all images were interpreted and scored individually [23–25].

Three months later, the readers reviewed all the images alongside the decision of the ML model for each renal lesion.
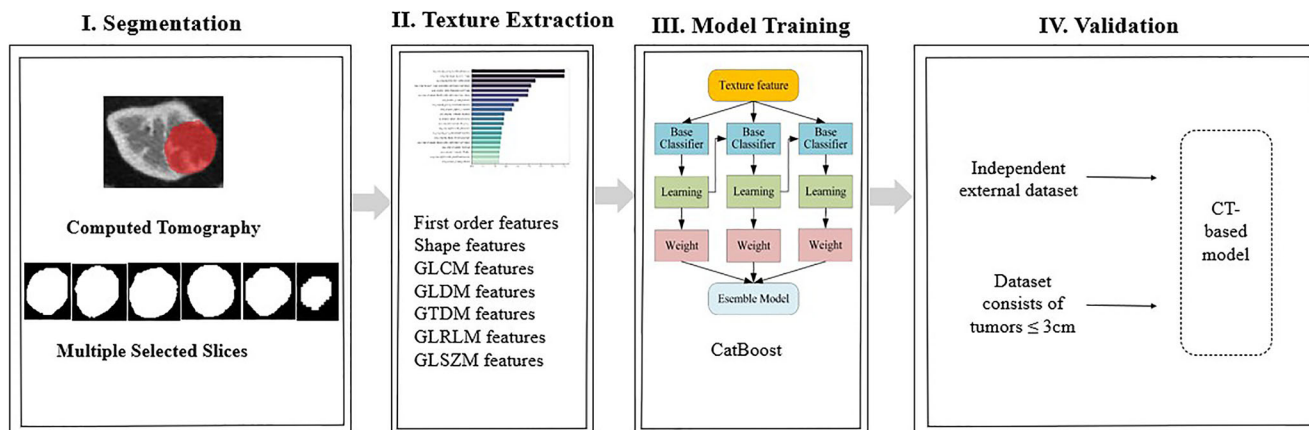
Fig. 2 Process flow diagram for developing and validating the CT-based ML model

The radiologist assessed the agreement between themselves and that of the model and then made a final decision regarding the related renal masses.

Statistical analysis

Statistical analysis for the performance of the CT-based model was conducted in Python (version 3.6.5, www.python.org), and Sklearn was adopted for index analysis [26]. The evaluation indicators included the true positive rate (TPR), specificity (SPC), positive predictive value (PPV), negative predictive value (NPV), accuracy (ACC), and area under the receiver operating characteristic (ROC) curve (AUC). The mean age was compared between the patients with RCCs and the patients with benign renal tumors using Student's t test. Chi-square tests were used to compare the male to female ratio between the patients with RCCs and the patients with benign renal tumors. A p value of less than 0.05 was considered statistically significant.

Results

Demographics

Of all 805 renal lesions included in this study, 680 (84.4%) were RCCs, and 125 (15.5%) were benign renal tumors. The median age of patients with RCCs was higher than that of patients with benign renal tumors. The gender structure difference was not statistically significant. Baseline characteristics for each of the groups are presented in Table 1.

Three-phase vs. single-phase image models

As demonstrated in Table 2, the machine learning classifier based on all-phase images achieved a higher AUC than the classifiers constructed from each set of single-phase images. In addition, the ML model based on 3D images was slightly superior to that based on 2D images.

In the model development based on 3D images, 107 features were extracted from PCP, CMP, and NP CT images, and 321 features were extracted from all-phase CT images.

However, 30, 29, 37, and 96 features, respectively, were excluded due to low ICC values (< 0.75). Finally, 77, 78, 70, and 225 features from PCP, CMP, NP, and all-phase CT images, respectively, were selected for ML model development. The top 20 ranking features and the ROC curve of the best model are shown in Figs. 3 and 4.

All renal tumors vs. renal tumors < 3 cm

The machine learning model was externally validated, and the classifier performed better when assessing all tumors than when assessing only tumors measuring < 3 cm; however, acceptable performance was achieved in the latter case (3D image–based model: AUC 0.81 vs. 0.79, ACC 0.86 vs. 0.77; 2D image–based model: AUC 0.75 vs. 0.76, ACC 0.86 vs. 0.75). The performance of the ML model in differentiating renal tumors < 3 cm is shown in Table 3.

Contour vs. noncontour focus

To explore the influence of the segmentation margin, the performances of the ML models based on contour focus, expansions of 1 mm and 3 mm, and shrinkages of 1 mm and 3 mm were compared (Table 4). The performance of the model was not significantly different when the tumor margin was shrunk/ expanded by 1 mm or 3 mm than when the margin was unaltered, achieving AUCs of 0.79 (tumor margin shrunk by 1 mm), 0.77 (tumor margin expanded by 1 mm), 0.77 (tumor margin shrunk by 3 mm), and 0.74 (tumor margin expanded by 3 mm).

Radiological interpretation with and without the machine learning algorithm

As shown in Table 5, the radiologists had relatively poor performance compared with the ML classifier, especially in terms of the NPV. Notably, all three radiologists achieved better performance when referring to the machine learning classifier's decision, especially in terms of the NPV. The AUC of the three readers with and without ML model assistance are shown in Fig. 5.

Table 1 Baseline characteristics of patients

Patient characteristic	External cohort			Internal cohort		
	Benign tumor ($n = 47$)	RCC ($n = 262$)	p	Benign tumor ($n = 78$)	RCC ($n = 418$)	p
Age	53.88 ± 12.24	56.35 ± 12.66	0.006	50.68 ± 13.30	55.58 ± 12.15	< 0.001
Gender (male:female)	98:393	62:245	0.935	26:97	134:541	0.743

**Table 2** Performance of the ML classifier in differentiating benign renal tumors from malignant tumors

| Phase	2D						3D					
	TPR	SPC	PPV	NPV	ACC	AUC	TPR	SPC	PPV	NPV	ACC	AUC
NP	0.58	0.9	0.4	0.94	0.86	0.75	0.38	0.91	0.58	0.82	0.78	0.78
PCP	0.44	0.9	0.46	0.89	0.82	0.73	0.37	0.9	0.5	0.83	0.78	0.74
CMP	0.51	0.88	0.31	0.93	0.83	0.69	0.41	0.89	0.43	0.88	0.81	0.75
ALL	0.76	0.87	0.2	0.99	0.87	0.74	0.57	0.91	0.51	0.92	0.86	0.81

Discussion

Our study indicates that machine learning could be a reliable and reproducible method for helping distinguish RCCs from benign renal masses. Compared to previous studies, we considered several potential factors, including the tumor size and delineation of the tumor margins, and investigated the influence of these factors on the performance of the ML classifier. Moreover, the ML model may be an optimal complementary assistant for radiologists in differentiating common RCCs from benign renal tumors and may be especially helpful in the detection of benign renal tumors.

In previous studies, Kunapuli et al [27] and Erdim et al [7] compared different ML algorithms in building a prediction model for renal mass differentiation. Sun et al [12] compared the performance of radiomic-radiologic ML models and expert-level radiologists in classifying solid renal masses and found that their optimal model achieved an AUC ranging from 0.83 to 0.92. Xi et al [28] investigated the diagnostic value of a deep learning (DL) model based on MRI data, and this model achieved higher performance values than four expert radiologists. All these studies indicate that radiomics could be a promising method for differentiating common RCCs from benign renal tumors, which is consistent with the results of our study, but the performance of these ML or DL approaches is questionable given the loss of information from entire volumetric lesion images or the lack of consideration for certain pieces of radiologic identification such as size and shape margin. Therefore, in our study, we trained our ML model using features derived from both 2D and 3D images. We also
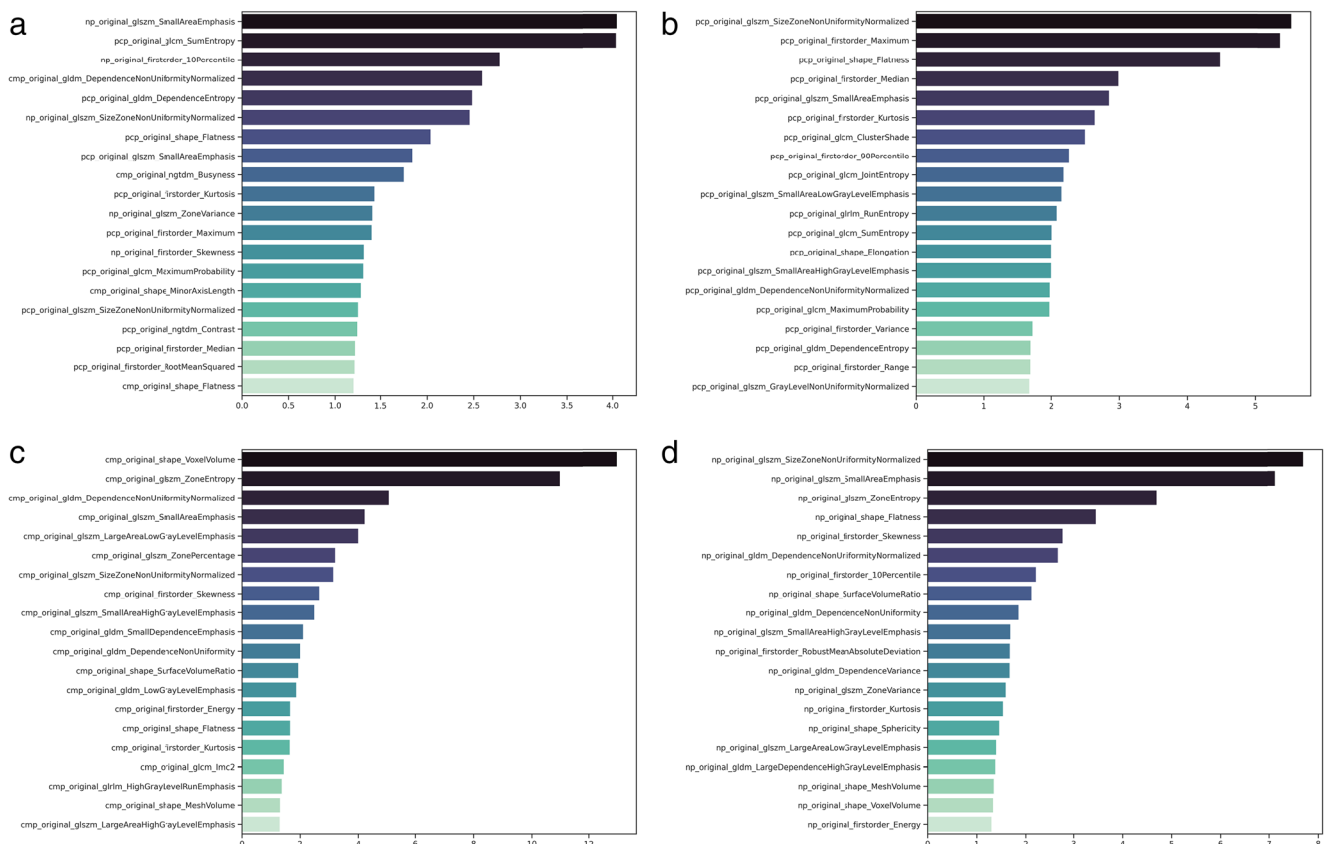
Fig. 3 The top 20 ranking feature scores in the differentiation of RCCs from benign renal tumors (**a** three-phase; **b** PCP; **c** CMP; **d** NP)

Fig. 4 ROC curves of the best machine learning classifier based on various phase images
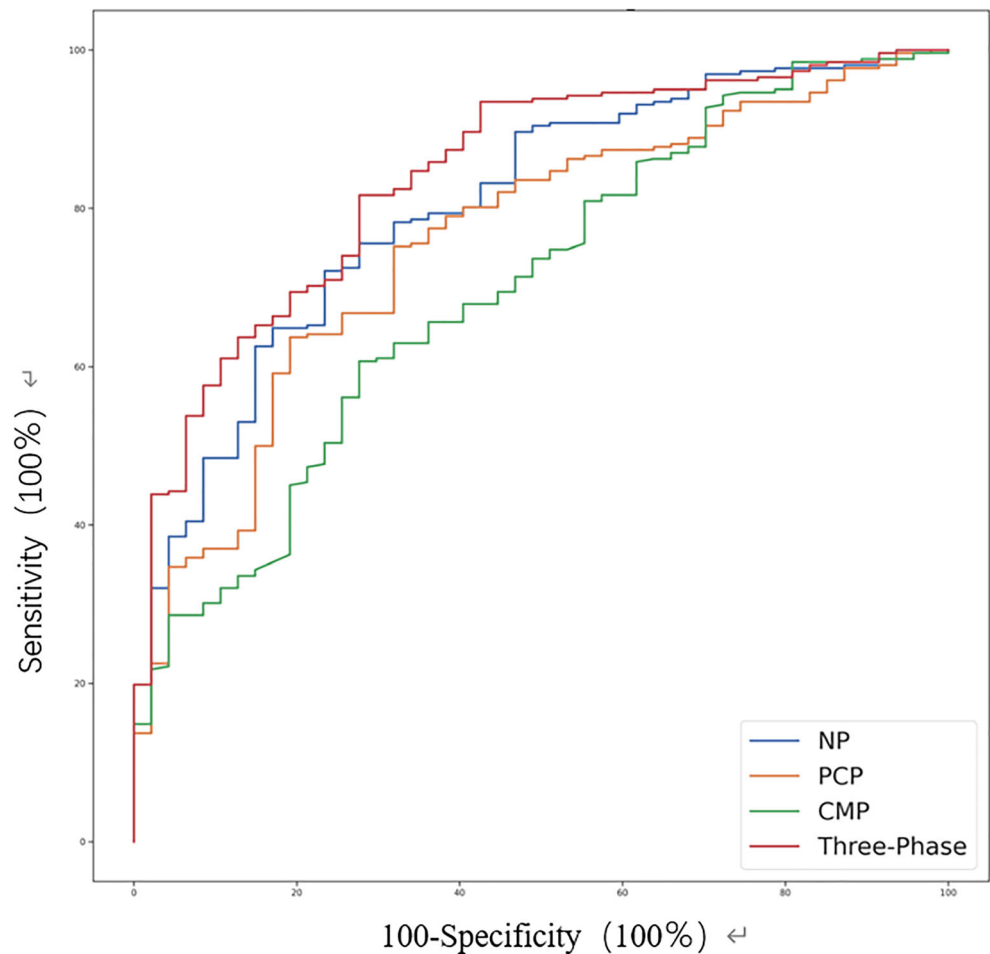
assessed the changes in the ML model performance during clinical application with regard to tumor size and segmentation strategy.

Comparing the performance of the model based on features from different CT phase images, we found that the model based on whole-tumor features derived from the three-phase CT images had better performance in renal lesion classification. This may be because the all-phase images and the whole tumor pattern provided more information about the lesion to the model.

Regarding the lesion volume, Tanaka et al [29] investigated the performance of a CNN-based deep learning model in classifying small renal tumors (less than 4 cm), and they found that the deep learning model based on CMP CT images performed better, achieving an ACC of 0.88. In our study, the size of renal tumors varied, and tumors less than 3 cm were selected from the external database to test the ML model. The limited information provided by tumors measuring less than 3 cm reduced the renal mass differentiation accuracy of the model, especially for the single-CMP-image model. As

Table 3 The performance of the ML model in differentiating renal tumors < 3 cm

Phase	2D						3D					
	TPR	SPC	PPV	NPV	ACC	AUC	TPR	SPC	PPV	NPV	ACC	AUC
NP	0.59	0.84	0.65	0.79	0.75	0.76	0.45	0.91	0.85	0.54	0.63	0.76
PCP	0.55	0.83	0.66	0.74	0.71	0.71	0.48	0.85	0.75	0.63	0.67	0.72
CMP	0.51	0.77	0.46	0.77	0.68	0.6	0.43	0.81	0.69	0.57	0.61	0.64
ALL	0.77	0.78	0.39	0.95	0.78	0.74	0.61	0.9	0.8	0.77	0.77	0.79

Table 4 The performance of the ML model based on various segmentation strategies

Groups	Phase	2D						3D					
		TPR	SPC	PPV	NPV	ACC	AUC	TPR	SPC	PPV	NPV	ACC	AUC
Shrink 1 mm	NP	0.59	0.89	0.39	0.93	0.85	0.75	0.46	0.9	0.5	0.88	0.82	0.79
	PCP	0.54	0.89	0.39	0.92	0.84	0.74	0.51	0.89	0.4	0.92	0.84	0.73
	CMP	0.57	0.86	0.16	0.96	0.84	0.66	0.38	0.89	0.42	0.85	0.79	0.76
	ALL	0.75	0.87	0.23	0.98	0.87	0.75	0.62	0.89	0.34	0.96	0.86	0.79
Shrink 3 mm	NP	0.56	0.89	0.39	0.94	0.85	0.77	0.48	0.9	0.46	0.9	0.83	0.77
	PCP	0.51	0.89	0.41	0.92	0.84	0.73	0.49	0.89	0.37	0.92	0.84	0.73
	CMP	0.62	0.86	0.17	0.97	0.85	0.73	0.56	0.88	0.29	0.95	0.85	0.73
	ALL	0.69	0.88	0.28	0.98	0.87	0.75	0.68	0.88	0.29	0.96	0.86	0.77
Expand 1 mm	NP	0.4	0.9	0.47	0.87	0.81	0.72	0.34	0.91	0.6	0.78	0.75	0.76
	PCP	0.41	0.89	0.44	0.87	0.81	0.71	0.36	0.89	0.45	0.85	0.79	0.72
	CMP	0.44	0.88	0.3	0.93	0.83	0.7	0.48	0.88	0.34	0.92	0.83	0.74
	ALL	0.6	0.87	0.24	0.97	0.86	0.75	0.53	0.89	0.37	0.93	0.84	0.77
Expand 3 mm	NP	0.39	0.9	0.45	0.87	0.8	0.72	0.32	0.93	0.62	0.68	0.67	0.69
	PCP	0.36	0.88	0.37	0.88	0.8	0.7	0.41	0.88	0.35	0.9	0.82	0.73
	CMP	0.43	0.88	0.33	0.89	0.81	0.67	0.37	0.89	0.46	0.83	0.77	0.7
	ALL	0.62	0.87	0.24	0.97	0.86	0.74	0.49	0.89	0.36	0.93	0.84	0.74

expected, fewer characteristic features can be extracted from smaller lesions, making differentiation more difficult. However, we found that the ML model performance was still acceptable. In routine image interpretation, making an accurate diagnosis for incidental renal tumors measuring less than 3 cm is more important and more difficult than it is for larger renal tumors. Our study showed that the ML model is a practical technique for aiding radiologists in clinical practice, especially for identifying smaller renal tumors.

In addition, the segmentation margin remains a challenge in ML model development. Although manual delineation was considered the standard reference, the tumor contour might be ill-defined in some cases and lead to inaccurate delineations. Few studies have discussed the influence of the segmentation margin on model performance. In our study, the performance

of the model was not influenced substantially by shrinking/expanding the margin of the tumor by 1 mm or 3 mm. We assumed that the features subsequently lost or added by the slight changes in the tumor margin comprised a small proportion of all features extracted from the whole tumor; in other words, the features that we actually acquired were sufficient for estimating the homogeneity or heterogeneity of the tumor. Thus, in similar studies, we may not have to focus excessively on the subtle variance in ROI delineation.

In our study, the accuracy of the machine learning classifier was slightly higher than that of the three radiologists, while after considering the decision of the machine learning model, the three radiologists had a higher accuracy than the model, which shows the usefulness of the model in the identification of renal masses. In addition, the NPV (0.82–0.99) of the ML model complemented the PPV (0.91–0.95) of the radiologists, potentially leading to perfect complementary assistance.

One limitation of our study that should be noted is the imbalanced nature of the dataset (RCCs:benign renal tumors = 680:125). To address the adverse impact of the imbalanced dataset on the performance of the classifiers, a synthetic minority oversampling technique (SMOTE) was adopted for the sample generation of the minority group from the joint weighting of the optimal features. As a result, the representation of the minority benign renal tumor group (AMLwvf and oncocytoma) increased, and the balance of the dataset improved, followed by improvement of the performance indicators (with an AUC of 0.97, an ACC of

Table 5 Performance of the three radiologists and their performance with the ML classifier as a reference in differentiating benign renal tumors from malignant tumors

Reader	TPR	SPC	PPV	NPV	ACC	AUC
Radiologist A	0.85	0.63	0.93	0.44	0.82	0.74
Radiologist B	0.87	0.55	0.91	0.44	0.82	0.71
Radiologist C	0.75	0.8	0.95	0.36	0.76	0.78
Radiologist A + radiomics	0.91	0.68	0.94	0.58	0.87	0.79
Radiologist B + radiomics	0.91	0.70	0.94	0.58	0.88	0.80
Radiologist C + radiomics	0.87	0.70	0.94	0.49	0.87	0.78

Fig. 5 ROC curves of subjective interpretations of radiologists with and without the ML classifier for renal mass discrimination
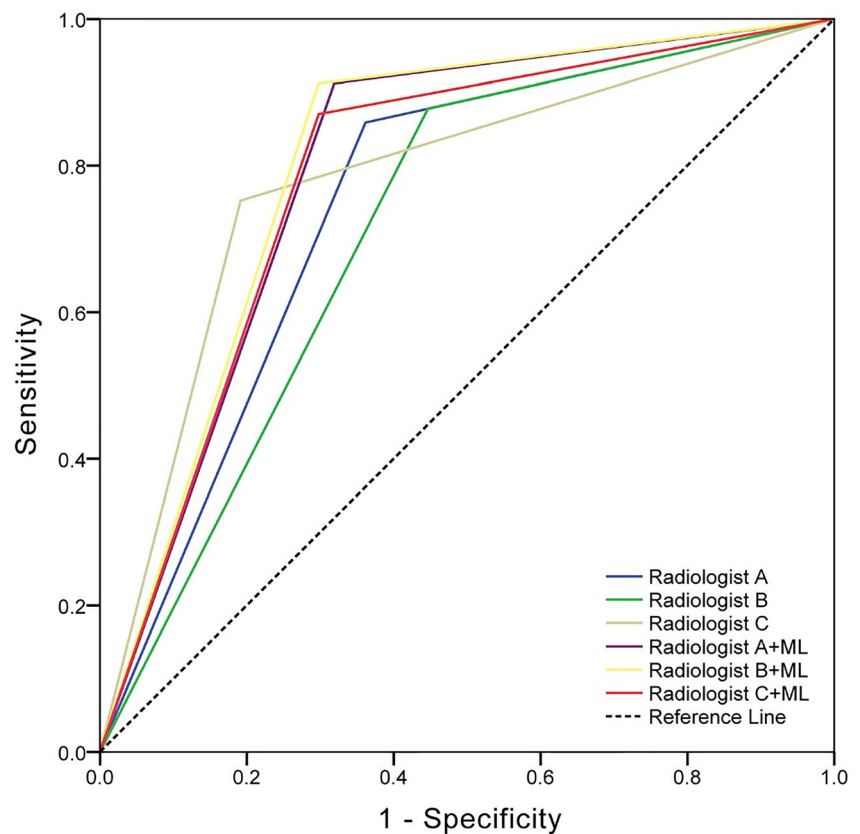
0.93). However, the statistics generated by SMOTE were less authentic due to the inevitable overfitting of these algorithms. To some extent, the imbalanced data reflect the real incidence of malignant and benign renal tumors. In addition, only one algorithm was used for the establishment of ML in our study, which is also a potential weakness. Thus, a large-scale, multicenter and multialgorithm study is necessary to further validate our study.

Conclusion

Our study shows that a machine learning classifier based on texture features derived from whole-tumor three-phase CT images can be a useful and promising technique for differentiating RCCs from benign renal tumors, which also contributes to the identification of small renal tumors. Furthermore, the ML model perfectly complemented the radiologists' interpretations and could be useful in improving performance, especially in the precise diagnosis of benign renal tumors.

Declarations

Guarantor The scientific guarantor of this publication is Wansheng Long.

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
- diagnostic or prognostic study
- multicenter study

References

1. Vendrami CL, Villavicencio CP, Dejulio TJ et al (2017) Differentiation of solid renal tumors with multiparametric MR imaging. Radiographics 37(7):2026–2042. https://doi.org/10.1148/rg.2017170039

2. Volpe A, Panzarella T, Rendon RA, Haider MA, Kondylis FI, Jewett MA (2004) The natural history of incidentally detected small renal masses. Cancer 100(4):738–745. https://doi.org/10.1002/cncr.20025

3. Gill IS, Aron M, Gervais DA, Jewett MAS (2010) Clinical practice. Small renal mass. N Engl J Med 362(7):624–634. https://doi.org/10.1056/NEJMcp0910041

4. Lane BR, Babineau D, Kattan MW et al (2007) A preoperative prognostic nomogram for solid enhancing renal tumors 7 cm or less amenable to partial nephrectomy. J Urol 178(2):429–434. https://doi.org/10.1016/j.juro.2007.03.106

5. Lane BR, Samplaski MK, Herts BR, Zhou M, Novick AC, Campbell SC (2008) Renal mass biopsy–a renaissance? J Urol 179(1):20–27. https://doi.org/10.1016/j.juro.2007.08.124

6. de Leon AD, Kapur P, Pedrosa I (2019) Radiomics in kidney cancer: MR imaging. Magn Reson Imaging Clin N Am 27(1):1–13. https://doi.org/10.1016/j.mric.2018.08.005

7. Erdim C, Yardimci AH, Bektas CT et al (2020) Prediction of benign and malignant solid renal masses: machine learning-based CT texture analysis. Acad Radiol 27(10):1422–1429. https://doi.org/10.1016/j.acra.2019.12.015

8. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278(2):563–577. https://doi.org/10.1148/radiol.2015151169

9. Davnall F, Yip CSP, Ljungqvist G et al (2012) Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging 3(6):573–589. https://doi.org/10.1007/s13244-012-0196-6

10. Suarez-Ibarrola R, Basulto-Martinez M, Heinze A, Gratzke C, Miernik A (2020) Radiomics applications in renal tumor assessment: a comprehensive review of the literature. Cancers (Basel) 12(6):1–25. https://doi.org/10.3390/cancers12061387

11. Tabibu S, Vinod PK, Jawahar CV (2019) Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. Sci Rep 9(1):1–9. https://doi.org/10.1038/s41598-019-46718-3

12. Sun XY, Feng QX, Xu X et al (2020) Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists. AJR Am J Roentgenol 214(1):W44–W54. https://doi.org/10.2214/AJR.19.21617

13. 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(5):2912–2921. https://doi.org/10.1007/s00330-019-06601-1

14. Yushkevich PA, Piven J, Hazlett HC et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015

15. Cui EM, Lin F, Li Q et al (2019) Differentiation of renal angiomyolipoma without visible fat from renal cell carcinoma by machine learning based on whole-tumor computed tomography texture features. Acta Radiol 60(11):1543–1552. https://doi.org/10.1177/0284185119830282

16. Chen F, Gulati M, Hwang D et al (2017) Voxel-based whole-lesion enhancement parameters: a study of its clinical value in differentiating clear cell renal cell carcinoma from renal oncocytoma. Abdom Radiol (NY) 42(2):552–560. https://doi.org/10.1007/s00261-016-0891-8

17. Mazzei FG, Mazzei MA, Cioffi Squitieri N et al (2014) CT perfusion in the characterisation of renal lesions: an added value to multiphasic CT. Biomed Res Int 2014:135013. https://doi.org/10.1155/2014/135013

18. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. Stud Media Commun SMC 3(6):610–621. 10.1109/TSMC.1973.4309314.

19. Varghese BA, Chen F, Hwang DH (2018) Differentiation of predominantly solid enhancing lipid-poor renal cell masses by use of contrast-enhanced CT: evaluating the role of texture in tumor subtyping. AJR Am J Roentgenol 211(6):W288–W296. https://doi.org/10.2214/AJR.18.19551

20. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77(21):e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339

21. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) Catboost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018-Decem:6638–6648. https://doi.org/10.48550/arXiv.1706.09516

22. Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support. 1–7. https://doi.org/10.48550/arXiv.1810.11363

23. Dyer R, DiSantis DJ, McClennan BL (2008) Simplified imaging approach for evaluation of the solid renal mass in adults. Radiology 247(2):331–343. https://doi.org/10.1148/radiol.2472061846

24. Leveridge MJ, Bostrom PJ, Koulouris G, Finelli A, Lawrentschuk N (2010) Imaging renal cell carcinoma with ultrasonography, CT and MRI. Nat Rev Urol 7(6):311–325. https://doi.org/10.1038/nrurol.2010.63

25. Kutikov A, Smaldone MC, Egleston BL et al (2011) Anatomic features of enhancing renal masses predict malignant and high-grade pathology: a preoperative nomogram using the RENAL Nephrometry score. Eur Urol 60(2):241–248. https://doi.org/10.1016/j.eururo.2011.03.029

26. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(3):90–95. https://doi.org/10.1109/MCSE.2007.55

27. Kunapuli G, Varghese BA, Ganapathy P et al (2018) A decision-support tool for renal mass classification. J Digit Imaging 31(6):929–939. https://doi.org/10.1007/s10278-018-0100-0

28. Xi IL, Zhao Y, Wang R et al (2020) Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. Clin Cancer Res 26(8):1944–1952. https://doi.org/10.1158/1078-0432.CCR-19-0374

29. Tanaka T, Huang Y, Marukawa Y et al (2020) Differentiation of small (≤ 4 cm) renal masses on multiphase contrast-enhanced CT by deep learning. AJR Am J Roentgenol 214(3):605–612. https://doi.org/10.2214/AJR.19.22074