BMC Cancer

RESEARCH

Open Access

Multi-sequence MRI-based radiomics model to preoperatively predict the WHO/ISUP grade of clear Cell Renal Cell Carcinoma: a two-center study

Ruihong Chen[1,3], Qiaona Su[2,3], Yangyang Li[1,3], Pengxin Shen[1,3], Jianxin Zhang[2*] and Yan Tan[1,4*]

Abstract

Objectives To develop radiomics models based on multi-sequence MRI from two centers for the preoperative prediction of the WHO/ISUP grade of Clear Cell Renal Cell Carcinoma (ccRCC).

Methods This retrospective study included 334 ccRCC patients from two centers. Significant clinical factors were identified through univariate and multivariate analyses. MRI sequences included Dynamic contrast-enhanced MRI, axial fat-suppressed T2-weighted imaging, diffusion-weighted imaging, and in-phase/out-of-phase images. Feature selection methods and logistic regression (LR) were used to construct clinical and radiomics models, and a combined model was developed using the Rad-score and significant clinical factors. Additionally, seven classifiers were used to construct the combined model and different folds LR was used to construct the combined model to evaluate its performance. Models were evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), and decision curve analysis (DCA). The Delong test compared ROC performance, with $p < 0.050$ considered significant.

Results Multivariate analysis identified intra-tumoral vessels as an independent predictor of high-grade ccRCC. In the external validation set, the radiomics model (AUC = 0.834) outperformed the clinical model (AUC = 0.762), with the combined model achieving the highest AUC (0.855) and significantly outperforming the clinical model ($p = 0.003$). DCA showed that the combined model had a higher net benefit within the 0.04–0.54 risk threshold range than clinical model. Additionally, the combined model constructed using logistic regression has a higher priority compared to other classifiers. Additionally, 10-fold cross-validation with LR for the combined model showed consistent AUC values (0.849–0.856) across different folds.

Conclusion The radiomics models based on multi-sequence MRI might be a noninvasive and effective tool, demonstrating good efficacy in preoperatively predicting the WHO/ISUP grade of ccRCC.

Keywords Clear cell renal cell carcinoma, WHO/ISUP, Pathological grade, MRI, Radiomics

*Correspondence:
Jianxin Zhang
zjx2012032@163.com
Yan Tan
tanyan123456@sina.com

Full list of author information is available at the end of the article

Introduction

Renal cell carcinoma (RCC) is a malignant tumor originating from the epithelial cells of the renal tubules. Clear cell renal cell carcinoma (ccRCC) is the most common subtype, accounting for approximately 70-80% [1]. Pathological grading of ccRCC is crucial for determining prognosis and stratifying risk, with higher grades being associated with greater invasiveness, increased metastasis risk, and lower five-year survival rates [2–5]. In 2015, the World Health Organization (WHO) recommended the International Society of Urological Pathology (ISUP) grading system, which uses nucleolar prominence to differentiate grades 1 to 3 and nuclear pleomorphism for grade 4, with nuclear grades 1–2 considered low-grade, while grades 3–4 are considered high-grade [6–8].

Accurate pathological grading is essential for developing personalized treatment strategies. For patients who are not surgical candidates, pathological grading serves as a key risk stratification factor, guiding the selection of non-surgical treatment approaches, such as drug selection or radiofrequency ablation. In surgical patients, high-grade ccRCC patients with positive surgical margins face higher recurrence rates [9–11]. Xu et al. suggest that radical resection offers better outcomes for high-grade pT1-ccRCC compared to partial nephrectomy, emphasizing the importance of preoperative grading [12]. Preoperative pathological grading is typically determined through percutaneous renal biopsy, but it can result in sampling bias and complications, potentially underestimating the actual pathological grade. Studies indicate that the accuracy in distinguishing between high and low grades through percutaneous biopsy ranges from 69 to 76% [13–15]. Consequently, it is often overlooked, delaying the acquisition of pathological grading information. Therefore, the formulation of surgical plans relies on preoperatively accessible staging and tumor size [1, 9]. Developing non-invasive tools for the preoperative assessments of ccRCC grades could refine treatment standards.

MRI is an important tool for preoperative diagnosis, staging, detection, and treatment planning of ccRCC [16]. Radiomics, based on MRI, further explores intrinsic imaging information, utilizing machine learning and data analysis to enhance the precision of disease diagnosis, prognosis assessment, and treatment decision-making [17, 18]. Numerous studies have used radiomics to predict the pathological grading of malignant tumors, such as liver cancer and gliomas [19, 20]. Currently, research on radiomics for predicting the WHO/ISUP grade of ccRCC is limited to single sequence and lacks multi-center validation [18]. Our research aims to develop a multi-sequence MRI-based radiomics model and evaluate its ability to preoperatively predict the WHO/ISUP grade of ccRCC through external validation.

Materials and methods

Patients

This retrospective study was approved by the Institutional Review Board of the Hospital 1 and Hospital 2, with informed consent waived. Data were reviewed from patients diagnosed with ccRCC at the Hospital 1 (Center 1) between January 2019 and January 2024 and at Hospital 2 (Center 2) between January 2019 and January 2023. The inclusion criteria were: (1) Complete preoperative MRI images; (2) Surgery within 20 days post-examination; (3) Postoperative pathology diagnosed with ccRCC; and (4) No other malignant diseases. The exclusion criteria were: (1) Incomplete target sequence images; (2) Poor image quality; (3) Other malignant tumors or a history of kidney surgery, radiation, or chemotherapy; and (4) No clear WHO/ISUP grade.

A total of 203 patients from Center 1 were randomly divided into training ($n=142$) and testing ($n=61$) sets in a 7:3 ratio, while 131 patients from Center 2 formed the external validation set. Grades 1 and 2 were classified as low-grade, and grades 3 and 4 were classified as high-grade. The selection process is shown in Fig. 1.

Data collection

MRI image acquisition

This study used a 3.0T MRI scanner (Signa HDxt, GE Healthcare, USA) with an 8-channel phased-array coil. Scanning sequences included dynamic contrast-enhanced MRI (DCE-MRI) with cortical phases (CP), medullary phases (MP), and excretory phases (EP); axial fat-suppressed T2-weighted imaging (FST2WI); diffusion-weighted imaging (DWI) with echo planar imaging (EPI) for apparent diffusion coefficient (ADC) calculation; and in-phase (IP) and out-of-phase (OP) images for subtraction (SUB) calculation. The DCE-MRI was performed by intravenous injection of a gadolinium-based contrast agent at a dose of 0.2 ml/kg, with images acquired at 20–30 s, 70–80 s, and 180 s post-injection. Specific parameters are detailed in Supplementary Material Table 1.

Clinical data collection

Clinical data included age, gender, height, weight, T stage, history of hypertension, pathological results, examination time, methods, and surgical time. The pathology results for all patients were derived from postoperative biopsy. Nuclear grade was assessed according to the WHO/ISUP grading system (Supplementary Material Table 2).

Data Analysis

Acquisition and assessment of radiological features

Two experienced abdominal radiologists, blinded to clinical and pathological results, independently reviewed MRI

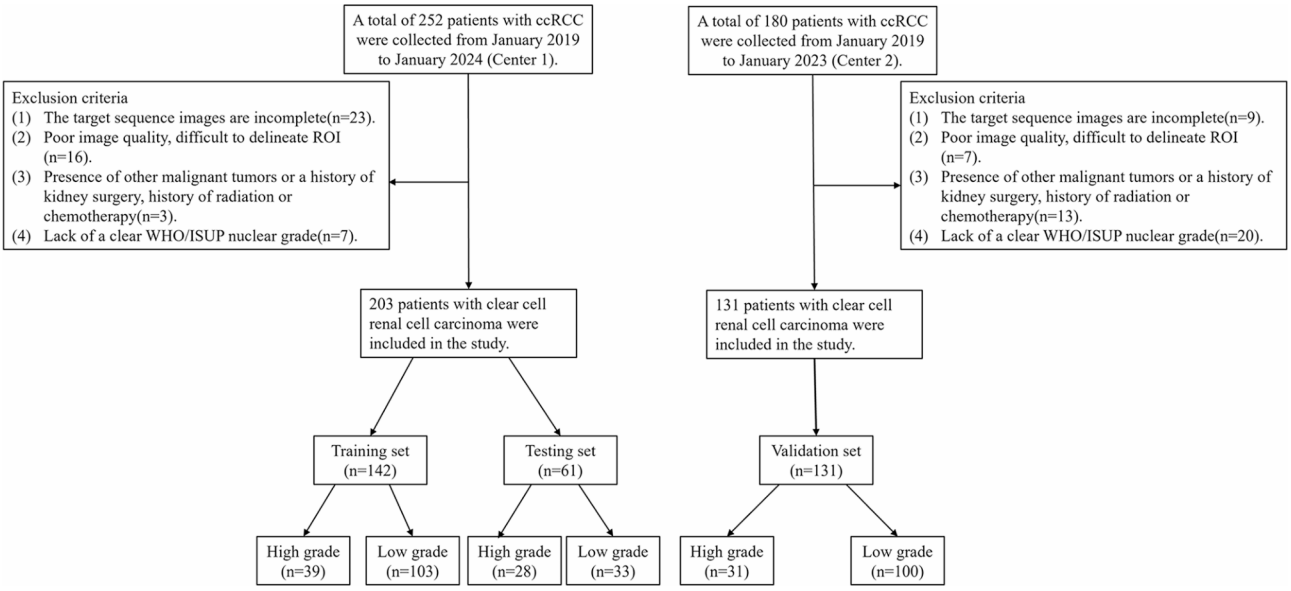

Fig. 1 Flowchart of the study population. ROI = Region of Interest, ccRCC = clear cell Renal Cell Carcinoma

images for the following features: tumor location (left/right), maximum diameter, intra-tumoral vessels (present/absent), enhancement degree (mild to moderate/intense), cystic necrosis (present/absent), lipid composition (present/absent), pseudocapsule (present/absent), venous tumor thrombus (present/absent), T stage, renal sinus invasion (present/absent), and tumor boundary (clear/unclear). T stage classification followed the 8th edition TNM staging system by the American Joint Committee on Cancer (AJCC) [21] (detailed criteria are in Supplementary Material Table 3). For discrepancies, a senior abdominal radiologist made the final decision.

Establishment and evaluation of models

ROI segmentation

An experienced radiologist delineated tumor regions of interest (ROI) for six target sequence images (Dynamic Contrast-enhanced MRI (DCE-MRI) with cortical phases (CP), medullary phases (MP), and excretory phases (EP); axial fat-suppressed T2-weighted imaging (FST2WI); apparent diffusion coefficient (ADC) and Chemical Shift Imaging Subtracted Images (SUB)) individually using ITK-SNAP software (version 4.0.0, http://www.itksnap.org/). The ROI segmentation of the MRI target sequences is shown in Supplementary Material Fig. 1. Confirmation was provided by an abdominal imaging physician with ten years of experience, and the software generated the entire volume of interest (VOI) of the tumor. One month later, 20 patients were randomly selected for a second segmentation to calculate the intraclass correlation coefficient (ICC) for intra-observer agreement.

Radiomics signature development

This study utilized FeAture Explorer V.0.5.8 [22] in Python to extract radiomics features. Images were normalized and resampled to a pixel size of $1.0\times1.0\times1.0$ mm^3, and gray levels were discretized into 80 bins. Features included first-order, shape, GLCM, GLRLM, GLSZM, GLDM, and NGTDM categories. Nine filters were applied, including Wavelet Transform (WT), Square, Square root, Laplacian of Gaussian (LoG), Logarithm, Exponential, Gradient, and Local Binary Pattern (LBP) 2D/3D, and each target sequence extracted 1,688 radiomics features.

Development of radiomics models

Radiomics features were constructed on the training set. Features with Intra-class Correlation Coefficient (ICC) for intra-observer agreement ≥ 0.75 were retained, followed by z-score normalization. The normality of the radiomics features was assessed using the Shapiro-Wilk test. For features that followed a normal distribution, an independent samples t-test was employed. For features not conforming to a normal distribution, the Mann-Whitney U test was used. Features with a p-value less than 0.05 were retained. Additionally, the Spearman correlation matrix of the significant features was calculated, and one of each pair of features with a correlation greater than 0.9 was removed. Least Absolute Shrinkage and Selection Operator (LASSO) with ten-fold cross-validation selected valuable features for single-sequence models. Correlation analysis, Analysis of Variance (ANOVA), Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE) were used to select the optimal features from single-sequence models and construct

the multisequence radiomics model. The Rad-score was calculated, and radiomics models were evaluated using ROC curves and AUC values. The DeLong test was used to compare predictive performance among models.

Development of a clinical model

Univariate logistic regression analyzed clinical and routine imaging features (e.g., age, gender, maximum diameter, intra-tumoral vessels, enhancement degree, T stage). Features with *p*-values less than 0.05 underwent multivariate logistic regression to construct the clinical model, which was evaluated using ROC curves and AUC.

Development of a combined model

The most relevant clinical factors were combined with the Rad-score to create a combined model using LR, and a nomogram was generated. The DeLong test compared AUC values between the clinical model, radiomics model, and combined model. DCA evaluated clinical utility, and calibration curves assessed model calibration. Various machine learning algorithms (LR, Support Vector Machine [SVM], Multilayer Perceptron [MLP], Linear Discriminant Analysis [LDA], K-Nearest Neighbors [KNN], Extreme Gradient Boosting [XGBoost], Random Forest [RF], and Naive Bayes [NB]) combined with bootstrapping were used to evaluate the generalization and stability of the combined model. The process is shown in Fig. 2. Additionally, different folds LR was used to construct the combined model to evaluate its performance.

Statistical analysis

Statistical analysis was performed using R 4.2.3 software (http://www.R-project.org) and IBM SPSS version 25.0. Descriptive statistics (mean ± standard deviation) were used for clinical factor measurement data. An independent samples t-test was applied for normally distributed data, while the Mann-Whitney U test was used for non-normally distributed data. Clinical factor count data were presented as frequencies, and between-group comparisons were conducted using the chi-square test. The Intra-class Correlation Coefficient (ICC) was calculated to assess intra-reader agreement for radiomic feature extraction, with an ICC value ≥ 0.75 indicating good consistency. The DeLong test was employed to compare the AUC among different models. A difference was considered statistically significant when $p < 0.050$.

Results

Clinical data analysis

The study included 203 ccRCC patients from Center 1, divided into a training set (142 cases) and a validation set (61 cases), and 131 patients from Center 2 as the external validation set. There were no significant differences in 15 clinical and imaging features between the training and validation sets, except for gender ($p = 0.018$), as shown in Table 1.

Predictive performance of the clinical model

Univariate and multivariate logistic regression analyses between high-grade and low-grade groups in the training set are shown in Table 2. Intra-tumoral vessels, showing

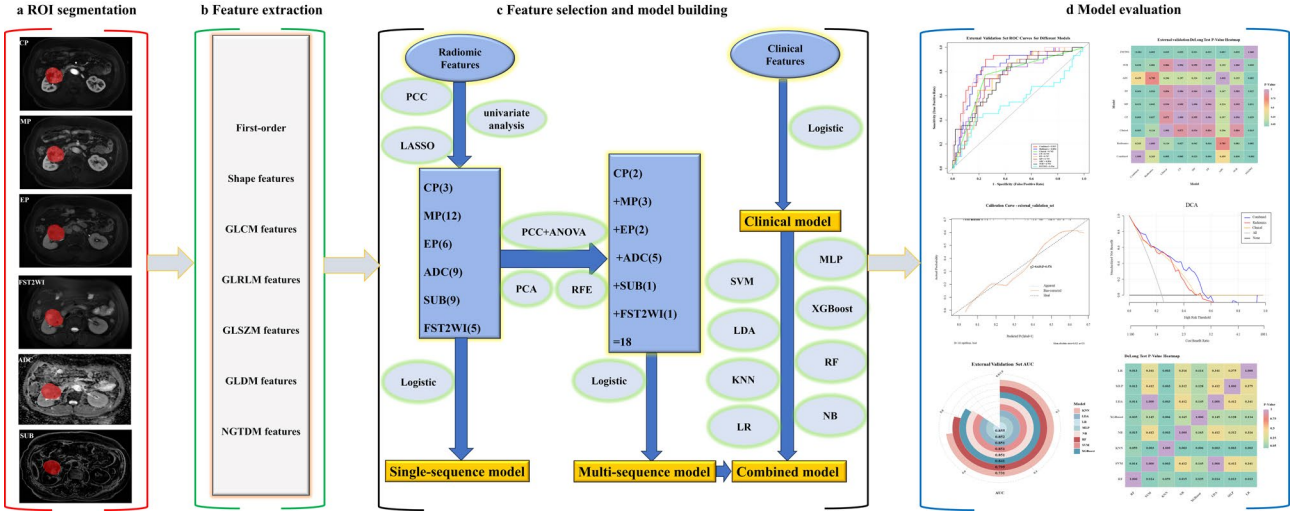

Fig. 2 Radiomics analysis workflow. (**a**) Tumor ROI segmentation by radiologists. (**b**) Feature extraction from tumor region. (**c**)Feature selection and model constructed. (**d**) Model evaluation. GLCM: Gray Level Co-occurrence Matrix. GLRLM: Gray Level Run Length Matrix. GLSZM: Gray Level Size Zone Matrix. GLDM: Gray Level Dependence Matrix. NGTDM: Neighborhood Gray Tone Difference Matrix. PCC: Pearson Correlation Coefficient. LASSO: Least Absolute Shrinkage and Selection Operator. PCA: Principal Component Analysis. ANOVA: Analysis of Variance. RFE: Recursive Feature Elimination. LR: Logistic Regression. SVM: Support Vector Machine. MLP: Multilayer Perceptron. LDA: Linear Discriminant Analysis. KNN: K-Nearest Neighbors. XGBoost: Extreme Gradient Boosting. RF: Random Forest. NB: Naive Bayes. Radiomics = multi-sequence radiomics

Table 1 Comparison of clinical factors between different sets

Characteristics	Training set (*N* = 142)	Testing set(*N* = 61)	*p*-value	Validation set (*N* = 131)	*p*-value
Age(y)	56.915 ± 11.978	57.426 ± 11.317	0.773	58.580 ± 9.503	0.203
Maximum diameter(cm)	4.695 ± 2.530	4.957 ± 2.922	0.543	5.665 ± 3.335	0.007*
BMI(Kg/m^2)	25.527 ± 5.003	25.148 ± 2.817	0.494	24.826 ± 3.199	0.166
T stage	107(75.4%)	48(78.7%)	0.520	91(69.5%)	0.094
1	6(4.2%)	2(3.3%)		14(10.7%)	
2	20(14.1%)	10(16.4%)		13(9.9%)	
3	9 (6.3%)	1(1.6%)		13(9.9%)	
4					
Renal sinus invasion	41(28.9%)	15(24.6%)	0.632	22(16.8%)	0.027*
present	101(71.1%)	46(75.4%)		109(83.2%)	
absent					
Gender	94(66.2%)	46(75.4%)	0.248	56(42.8%)	0.129
male	48(33.8%)	15(24.6%)		75(57.2%)	
female					
Location	77(54.2%)	21(34.4%)	0.018*	72(55.0%)	0.916
right	65(45.8%)	40(65.6%)		59(45.0%)	
left					
Intra-tumoral vessels	55(38.7%)	26(42.6%)	0.638	49(37.4%)	0.916
present	87(61.3%)	35(57.4%)		82(62.6%)	
absent					
Enhancement degree	8(5.63%)	3(4.92%)	>0.999	8(6.1%)	>0.999
mild to moderate	134(94.37%)	58(95.08%)		123(93.9%)	
intense					
Cystic necrosis	120(84.5%)	52(85.3%)	>0.999	102(77.9%)	0.169
present	22(15.5%)	9(14.7%)		29(22.1%)	
absent					
Lipid composition	74(52.1%)	30(49.2%)	0.789	87(66.4%)	0.023*
present	68(47.9%)	31(50.8%)		44(33.6%)	
absent					
Pseudocapsule	85(59.9%)	31(50.8%)	0.280	63(48.1%)	0.057*
present	57(40.1%)	30(49.2%)		68(51.9%)	
absent					
Venous tumor thrombus	7(4.9%)	4(6.6%)	0.716	18(13.7%)	0.011*
present	135(95.1%)	57(93.4%)		113(86.3%)	
absent					
Hypertension	51(35.9%)	26(42.6%)	0.424	58(44.3%)	0.181
yes	91(64.1%)	35(57.4%)		73(55.7%)	
no					
Tumor boundary	108(76.1%)	48(78.7%)	0.729	99 (75.6%)	>0.999
clear	34(23.9%)	13(21.3%)		32 (24.4%)	
unclear					

BMI = body mass index

* Signifcance at $p < 0.050$

Measurement data are presented as mean ± standard deviation, and count data as frequencies

statistical significance in both analyses, was used to construct the clinical model. The AUCs of the clinical model were 0.763, 0.766, and 0.762 in the training, testing, and external validation sets, respectively. ROC curves were plotted in Fig. 3a-c.

Developing and evaluating radiomics models

We extracted 1,688 features from each single-sequence image. Finally, 3, 12, 6, 9, 9, and 5 features were retained from the CP, MP, EP, SUB, ADC, and FST2WI sequences, respectively. Single-sequence radiomics models were established using LR. ROC curves were plotted in Fig. 3a-c, and the AUC, sensitivity, and specificity for each model in the training, testing, and external validation sets are presented in Table 3. In the external validation set, the ADC model had the highest AUC (0.824). The DeLong test (Fig. 3d-f) indicated that the performance of the FST2WI model was significantly lower than that of the other single-sequence models.

The 44 features from the six single sequences were further reduced to 18 features for the multi-sequence radiomics model. The process flow is shown in Fig. 4.

Table 2 The clinic-radiological features analysis in the training set

Characteristics	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)	*p*-value	Odds ratio (95% CI)	*p*-value
Age	1.031 (0.998–1.067)	0.073		
Gender	0.88 (0.410–1.937)	0.745		
BMI	1.023 (0.949–1.104)	0.543		
Hypertension	0.619 (0.269–1.353)	0.241		
Location	0.737 (0.350–1.545)	0.418		
Maximum diameter	1.274 (1.102–1.489)	0.001*	0.885 (0.671–1.151)	0.368
T stage	1.948 (1.351–2.858)	<0.001*	1.157 (0.613–2.200)	0.652
Renal sinus invasion	2.17 (0.988–4.746)	0.052		
Intra-tumoral vessels	10.4 (4.510–26.073)	<0.001*	11.138 (3.700–36.364)	<0.001*
Enhancement degree	2.771 (0.471–52.701)	0.348		
Cystic necrosis	0.779 (0.299–2.198)	0.619		
Lipid composition	1.101 (0.526–2.320)	0.799		
Pseudocapsule	0.822 (0.39–1.748)	0.606		
Venous tumor thrombus	3.81 (0.803–20.145)	0.090		
Tumor boundary	0.256 (0.111–0.579)	0.001*	0.634 (0.216–1.897)	0.407

BMI = body mass index, CI = confidence interval

* Signifcance at $p < 0.050$

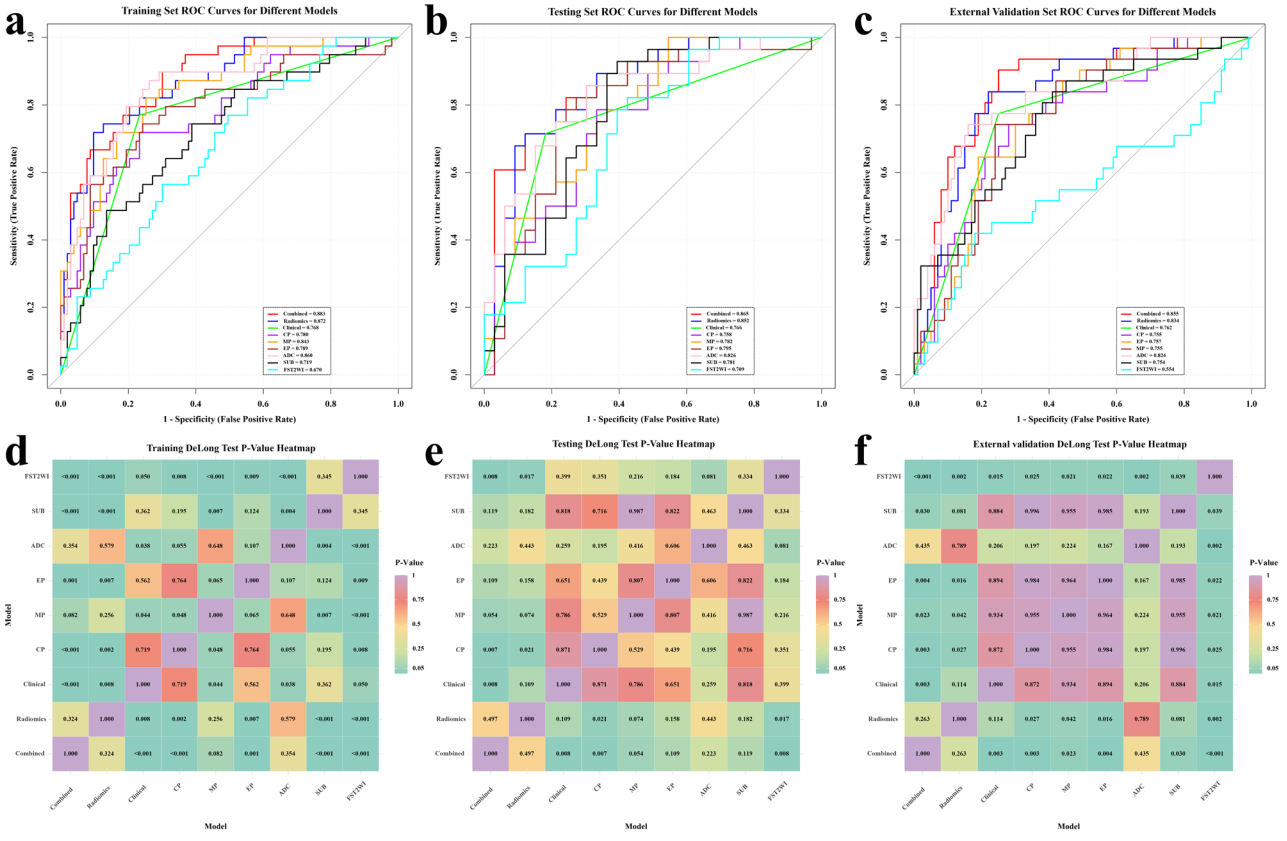

Fig. 3 The ROC curves and AUC values for the models in the Training (**a**), Testing (**b**) and External Validation sets (**c**). The DeLong Test Results for the models in the Training (**d**), Testing (**e**) and External Validation sets (**f**). CP = cortical phases, MP = medullary phases, EP = excretory phases, FST2WI = fat-suppressed T2-weighted imaging, ADC = apparent diffusion coefficient, SUB = chemical shift imaging subtraction map, Radiomics = multi-sequence radiomics

Table 3 Predictive performance of each model for ccRCC grading in the training, testing, and external validation sets

Model	Set	AUC (95% CI)	SEN	SPE	Accuracy	PPV	NPV	F1-Score
CP	Training	0.780(0.693–0.867)	0.717	0.767	0.753	0.538	0.877	0.615
	Testing	0.758 (0.637–0.878)	0.536	0.727	0.639	0.625	0.649	0.577
	Validation	0.755 (0.658–0.851)	0.581	0.790	0.741	0.462	0.859	0.515
MP	Training	0.843 (0.772–0.914)	0.821	0.748	0.768	0.552	0.917	0.660
	Testing	0.782 (0.669–0.896)	0.607	0.697	0.656	0.630	0.676	0.618
	Validation	0.757 (0.668–0.847)	0.645	0.810	0.771	0.513	0.880	0.571
EP	Training	0.789 (0.689–0.880)	0.744	0.757	0.753	0.537	0.886	0.624
	Testing	0.795 (0.678–0.913)	0.750	0.758	0.754	0.724	0.781	0.737
	Validation	0.755 (0.664–0.846)	0.806	0.580	0.633	0.373	0.906	0.510
ADC	Training	0.860 (0.793–0.926)	0.872	0.738	0.775	0.558	0.938	0.680
	Testing	0.826 (0.720–0.931)	0.679	0.818	0.754	0.760	0.750	0.717
	Validation	0.824 (0.738–0.910)	0.774	0.720	0.648	0.461	0.911	0.578
SUB	Training	0.719 (0.624–0.814)	0.744	0.612	0.648	0.421	0.863	0.537
	Testing	0.781 (0.665–0.898)	0.857	0.636	0.737	0.666	0.840	0.750
	Validation	0.754 (0.656–0.852)	0.871	0.510	0.595	0.355	0.927	0.505
FST2WI	Training	0.670 (0.574–0.766)	0.769	0.505	0.578	0.370	0.852	0.500
	Testing	0.709 (0.579–0.839)	0.643	0.636	0.639	0.600	0.677	0.621
	Validation	0.554 (0.422–0.685)	0.452	0.430	0.435	0.197	0.717	0.275
Radiomics	Training	0.872 (0.809–0.935)	0.745	0.854	0.824	0.659	0.898	0.699
	Testing	0.852 (0.755–0.949)	0.714	0.879	0.803	0.834	0.784	0.769
	Validation	0.834 (0.757–0.912)	0.774	0.810	0.801	0.558	0.920	0.649
Clinical	Training	0.763 (0.684–0.842)	0.769	0.767	0.768	0.556	0.898	0.645
	Testing	0.766 (0.658–0.875)	0.714	0.818	0.770	0.769	0.771	0.741
	Validation	0.762 (0.676–0.848)	0.774	0.750	0.756	0.490	0.915	0.600
Combined	Training	0.885 (0.828–0.942)	0.897	0.699	0.753	0.530	0.947	0.666
	Testing	0.865 (0.772–0.957)	0.857	0.697	0.770	0.706	0.852	0.774
	Validation	0.855 (0.781–0.929)	0.903	0.750	0.786	0.528	0.962	0.667

CP=cortical phases, MP=medullary phases, EP=excretory phases, FST2WI=fat-suppressed T2-weighted imaging, ADC=apparent diffusion coefficient, SUB=chemical shift imaging subtraction map, Radiomics=multi-sequence radiomics, AUC=area under the curve; ccRCC=clear cell renal cell carcinoma; CI=confidence interval; SEN=Sensitivity; SPN=Specificity, Negative Predictive Value=NPV, Positive Predictive Value=PPV

The external validation set AUC of the multi-sequence radiomics model reached 0.834, significantly outperforming the CP, EP, MP, and FST2WI models (Fig. 3d-f). ROC curves were plotted in Fig. 3a-c, and the AUC, sensitivity, and specificity for each model on the training, testing, and external validation sets are presented in Table 3.

Developing and evaluating combined model

The combined model, constructed using LR, incorporated intra-tumoral vessels and the Rad-score. The nomogram is shown in Supplementary Material Fig. 2, with its visualization schematic is shown in Fig. 5a-b. ROC curves are displayed in Fig. 3a-c. The AUC, sensitivity, and specificity for training, testing, and external validation sets are summarized in Table 3, with the combined model achieving 78.6% accuracy in distinguishing high and low grades. The DeLong test showed the combined model (AUC=0.855) performed significantly better than the clinical model (AUC=0.763, p=0.003) but not significantly different from the radiomics model (AUC=0.834, p=0.263) in the external validation set. Calibration curves indicated a good fit (Fig. 5c), and DCA (Fig. 5d) showed the combined model had a higher net benefit within a risk threshold range of 0.04–0.54 in the external validation set. Eight classifiers were used to evaluate the stability and reliability of the combined model. The DeLong test and the bootstrapping method with 1000 resamples were applied to compare the AUC differences between different classifiers (Fig. 6). In the external validation set, the LR model had the highest predictive performance, significantly outperforming the KNN and RF models (Fig. 6e). Additionally, 10-fold cross-validation with LR for the combined model showed consistent AUC values (0.849–0.856) across different folds (Fig. 6f).

Discussion

In recent years, advances in imaging technology and radiomics algorithms have significantly improved the early diagnosis and treatment planning for ccRCC, enhancing clinical management and precision medicine [17, 23]. However, many studies are limited by single-center and reliance on single sequence, highlighting the need for improved predictive efficacy [19]. Our study combined multi-sequence MRI and clinical factors to

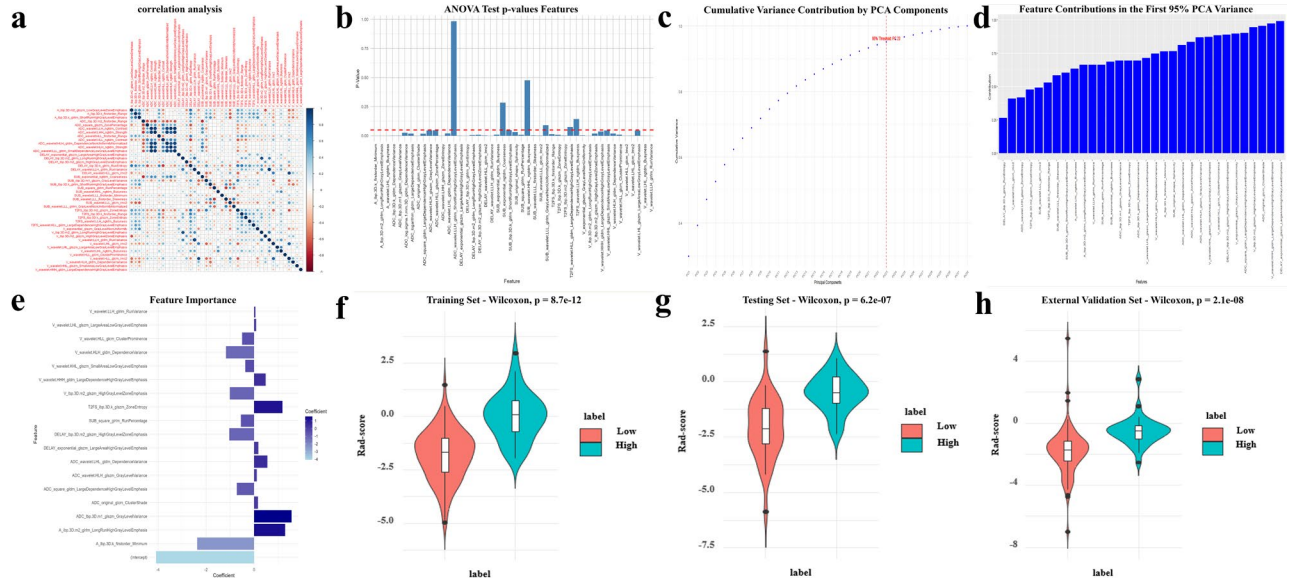

Fig. 4 Visualization of feature selection and analysis methods. (**a**) This heatmap shows the correlation coefficients among 44 features derived from various sequences. The color intensity and size of the circles represent the strength and direction of these correlations. The feature names are prefixed with the sequence sources, where A represents CP, V represents MP, and DELAY represents EP. (**b**) ANOVA Test p-values for Features: This bar plot shows the p-values from the ANOVA test for each feature. Features with p-values below the red dashed line (0.05 threshold) are considered statistically significant and were retained for further analysis. We calculated the correlation matrix of the features. Among features with a correlation coefficient greater than 0.8, only the feature with the smaller p-value was retained. (**c**) Cumulative variance contribution by PCA Components: This plot shows the cumulative variance explained by the principal components. The red dashed line indicates the point where the first 23 principal components collectively explain more than 95% of the variance. (**d**) Feature Contributions in the First 95% PCA Variance: This bar plot illustrates the contribution of each feature to the first 23 principal components. Only the top 23 features with the highest contributions are retained. (**e**) Coefficients and thresholds of the final features selected by RFE. (**f**-**h**) Difference in the Rad-score between high-grade and low-grade group in the training (**f**), testing (**g**), and external validation sets (**h**). The p-value is shown at the top of each image. The Rad-score in patients with high grade was higher than in those with low grade

construct a combined model, which demonstrated good predictive performance in the external validation set (AUC=0.855), providing valuable information for treatment selection.

Predictive value of the clinical model

Analysis of clinical factors and conventional imaging features showed that intra-tumoral vessels are most associated with the pathological grading of ccRCC. Some studies suggest that observing intra-tumoral vasculature through MRI could unveil tumor coagulative necrosis, which is attributed to excessive blood supply, immature vessels, and hypoxia-related bleeding and coagulation reactions associated with vascular remodeling within the tumor. This implies stronger invasiveness, possibly correlating with the pathological grading of ccRCC [24, 25]. In addition, univariate analysis revealed correlations between T staging, maximum tumor diameter, clarity of the tumor margin, and pathological grading. Furthermore, some scholars have found that age, shape, vein thrombosis, lymphadenopathy, necrosis, and perinephric invasion may also be factors correlated with pathological grading [25–27]. The clinical model demonstrated moderate performance in predicting the pathological grading of ccRCC, similar to previous studies [28]. This suggests that non-specific imaging features assessed through manual visual evaluation have limitations.

Predictive value of radiomics models

Most radiomics studies on ccRCC grading use CT imaging and the Fuhrman grading system, with fewer focusing on MRI and WHO/ISUP grades [29]. Studies have indicated that MRI radiomics outperform CT radiomics in predicting the pathological grading of ccRCC [30]. Our study, based on MRI radiomics and WHO/ISUP grading, evaluated predictive efficacy through external validation.

We integrated six target sequences, each serve distinct roles in ccRCC research. For instance, DCE-MRI captures tumor blood perfusion dynamics, vascular permeability, and micro-vessel density over time [31]. FST2WI provides clearer depictions of tumor morphology and anatomical structure, ADC indicates the degree of diffusion restriction within the tumor [32], while SUB reveals information about the tumor's lipid composition [33]. Among them, the ADC model demonstrated optimal efficacy, consistent with findings from previous studies [34]. Furthermore, each single-sequence model exhibits significantly higher efficacy than the FST2WI model. The deposition of lipids in ccRCC cytoplasm is significant in its occurrence [34, 35]. MRI chemical shift imaging,

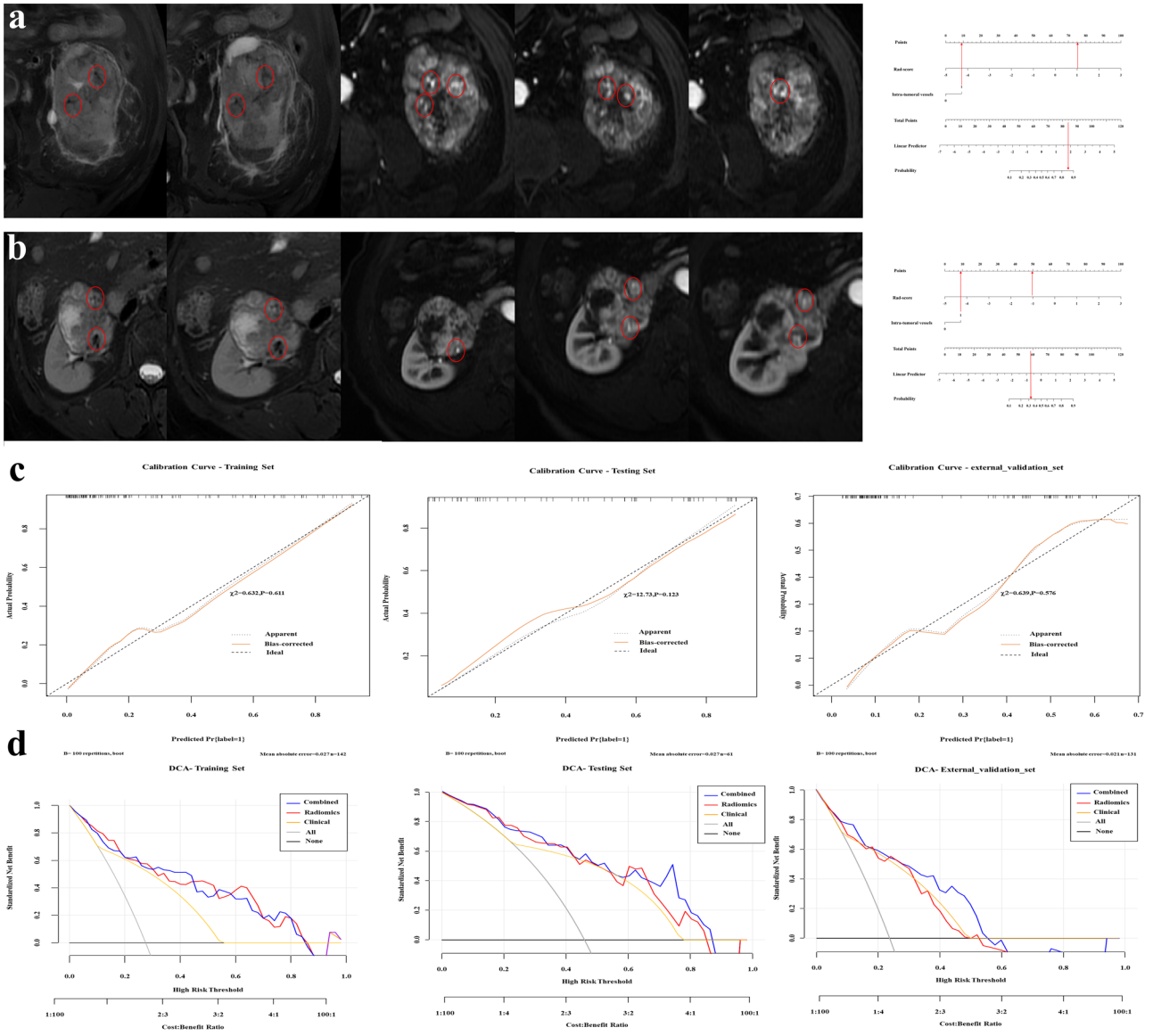

Fig. 5 The nomogram visualization schematic(**a**-**b**). (**a**), intra-tumoral vessels (As indicated by the red circle, there is a clear enhancement of blood vessels with continuous cortical courses), Rad-score = 1.058, calculates a probability of 0.872. The actual pathological grade is high. (**b**), with intra-tumoral vessel, Rad-score = -1.029, calculates a probability of 0.332. The actual pathological grade is low. Calibration Curves(**c**) and DCA(**d**) of combined model in training, testing and external validation sets. Radiomics = multi-sequence radiomics model

based on the frequency difference in proton motion between water and fat, allows lipid composition imaging in ccRCC [20, 36]. We developed a radiomics model using SUB to explore the correlation between lipid composition and pathological grading. Bardia hypothesized that higher ccRCC grades might have lower lipid composition, but this hypothesis was not confirmed [37]. We manually assessed SUB to define the presence of lipid composition and showed no clear correlation with grading. In contrast, radiomics analysis suggested a correlation, aligning with histological studies [34]. Pan et al. predicted WHO/ISUP grades using texture features from chemical shift imaging, but our radiomics approach achieved higher predictive performance (AUC = 0.75>0.64) [38]. It may be that SUB imaging reduces water influence, amplifying lipid information. Zhang et al. found lower lipid accumulation in grade 4 than grade 3 ccRCC using chemical shift imaging [39]. However, our study lacked integrated pathological analysis, limiting a deeper exploration of the specific correlation between lipid composition and pathological grading.

We removed highly correlated features through variance and correlation analysis, and selected the best radiomics features through principal component analysis and recursive feature elimination. Finally, we retained 18 valuable features out of 44, mostly high-order, reflecting

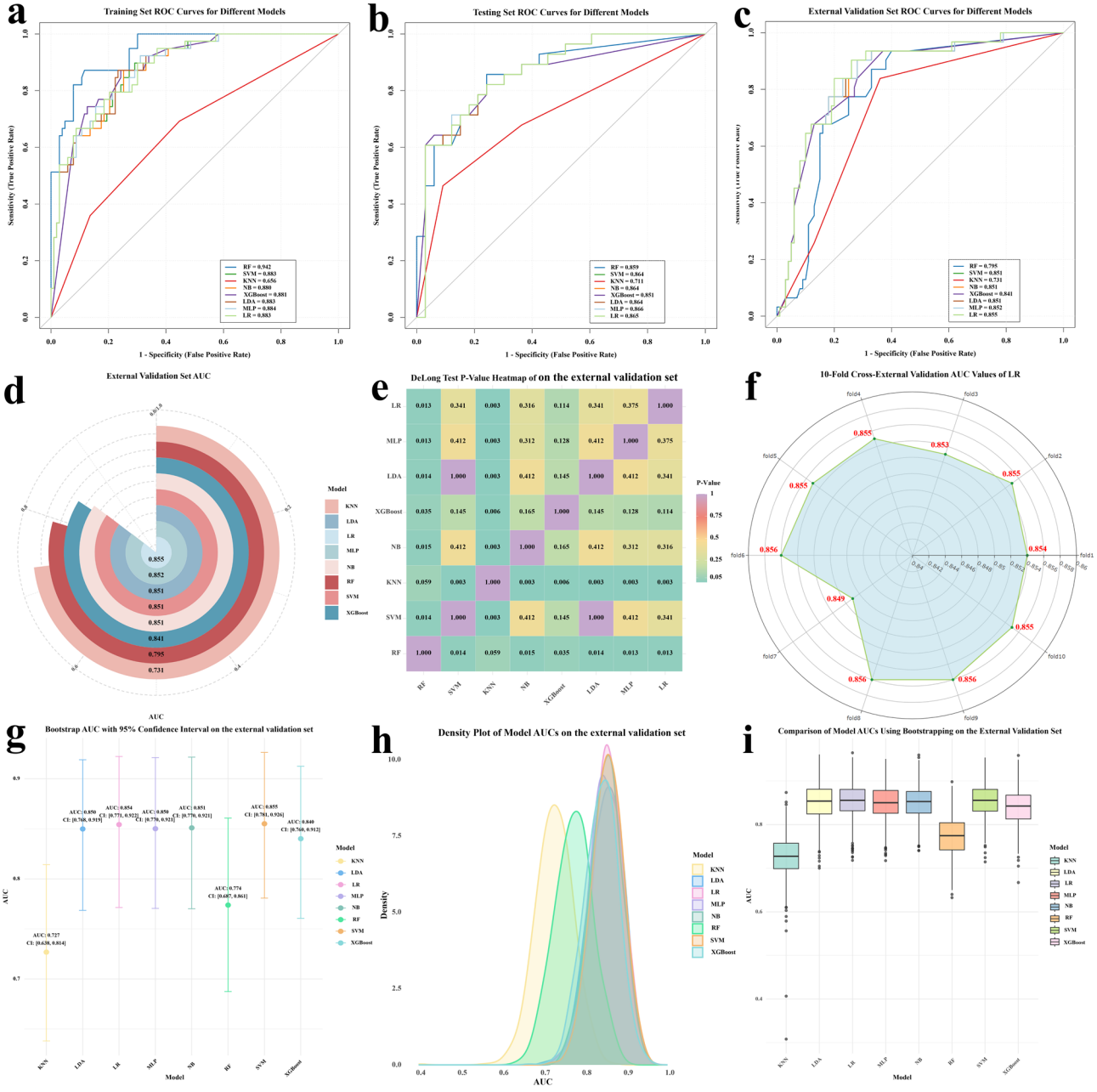

Fig. 6 Performance evaluation of eight classifiers. (**a**-**c**) ROC curves of eight classifiers. (**d**) AUC of the external validation set for eight classifiers. (**e**) DeLong test heatmap of the external validation set for eight classifiers. (**f**) AUC values of the combined models constructed using LR across different folds. (**g**) Bootstrap mean AUC with 95% confidence interval on the external validation set. (**h**) Density plot of Bootstrap AUCs on the external validation set. (**i**) Comparison of model AUCs with Bootstrapping on the external validation set. These plots demonstrate the predictive performance of different models on the external validation set, with LR showing the best performance

voxel relationships and intricate structural information crucial for grading ccRCC [40]. The multi-sequence radiomics model outperformed the single-sequence models by integrating multiple sequences, providing richer information and enhanced robustness. Li et al. found no significant performance differences between multi-sequence and single-sequence models, but our study showed the multi-sequence model significantly outperformed all single-sequence models except for ADC and SUB [26]. We attempted to exclude the low-performing FST2WI sequence from the multi-sequence model, but this reduced predictive efficacy, suggesting that even lower-performing sequences may contain features that can enhance overall model performance.

Predictive value of the combined model

Combining intra-tumoral vessels with the Rad-score to construct the combined model further enhanced predictive performance, offering higher clinical net benefits across most risk thresholds, consistent with previous research findings [26–28, 40]. Due to ccRCC heterogeneity, percutaneous biopsy has low sensitivity (66-75%) for high-grade cases and overall accuracy (69-76%), leading some scholars to believe that nuclear grading determined from biopsy is not suitable for preoperative risk stratification [13–15]. Our combined model demonstrated higher sensitivity (90.3%), and an accuracy of 78.6%. Considering the imbalance between high and low-grade samples, the positive predictive value is not high, being 0.756 in a validation set with more balanced cases. We attempted to use resampling methods to balance the sample sizes, but the results were unsatisfactory. Therefore, we tested the stability and credibility of the combined model using eight machine learning algorithms. The KNN model performed the worst, likely due to sample imbalance, while the random forest model exhibited overfitting. Other models showed similar performance, with AUCs ranging from 0.841 to 0.855 in the external validation, indicating good robustness. In this context, LR was preferred for its lower computational cost and better interpretability. Additionally, 10-fold cross-validation with LR confirmed the stability and generalizability of the combined model, with consistent AUC values across folds.

We found that the predictive performance of the ADC model, Radiomics model, and Combined model did not show significant differences in the external validation set. However, in the actual clinical environment, accurately identifying high-grade cases of ccRCC is crucial. Due to the varying incidence rates between high- and low-grade cases, F1-score can provide a more comprehensive evaluation of model performance. Additionally, clinicians place more emphasis on conventional imaging features, making the Combined model more practical for real-world applications, which had the highest sensitivity and F1-score in the testing set and external validation set. The Radiomics model excels in predicting low-grade cases and complements the strengths of the Combined model, further enhancing clinical decision-making. Although overall predictive capability of the ADC model is inferior to the other two, its simple structure and ease of use make it a viable and cost-effective option in scenarios with limited resources or equipment that does not support multi-sequence scanning.

Clinical application

As aforementioned, our study can help overcome the limitations of percutaneous renal biopsy, which often suffers from sampling errors and invasiveness, and may not accurately reflect the pathological grading of ccRCC [13, 14]. Notably, while the clinical model achieved an AUC of 0.762 in the external validation set, our combined model significantly outperformed it with an AUC of 0.855 ($p=0.003$). This marked improvement in discrimination ability highlights the added value of our combined model in predicting high-grade ccRCC cases more accurately, providing more reliable information for clinical decision-making, has the potential to enhance surgical planning and decision-making preoperatively. For example, our model can guide surgical planning by preoperatively predicting high-grade ccRCC cases, helping to determine the appropriate surgical margin or opting for radical resection, thereby reducing the potential recurrence rate and improving patient outcomes. Furthermore, it provides valuable information for the selection of medical therapies and non-surgical treatment options, which is particularly important for managing patients with comorbidities or high surgical risks. For instance, in the context of active surveillance, particularly for elderly or severely ill patients with small tumors (<4 cm) and high mortality risk, our model can distinguish between low- and high-grade tumors non-invasively. For low-risk patients, it helps avoid unnecessary surgeries and their associated risks, preserving renal function and improving quality of life. For high-risk patients, follow-up intensity can be customized based on the individual circumstances, ensuring timely intervention if any progression is detected.

Limitations

Firstly, despite including data from two centers, the overall sample size remains relatively small and imbalanced due to differences in the incidence rates of high and low-grade cases. The small sample size may have caused significant differences in certain predictive factors between the training and testing sets, increasing the risk of confounding. This imbalance limits our ability to adequately adjust for potential confounders, potentially resulting in overestimation or underestimation of specific associations within the models. Additionally, as a retrospective study, it inherently suffers from selection bias and lacks prospective validation. Finally, our research focused solely on the tumor itself and did not investigate potential information from the surrounding areas.

Conclusions

The radiomics models based on multi-sequence MRI might be a noninvasive and effective tool, demonstrating good efficacy in preoperatively predicting the WHO/ISUP grade of ccRCC, which may be a reasonable supplementary tool for percutaneous biopsy.

Abbreviations

WHO	World health organization
ISUP	International society of urological pathology

ccRCC	Clear cell renal cell carcinoma
DCE-MRI	Dynamic contrast-enhanced MRI
DWI	Diffusion-weighted imaging
FST2WI	Fat-suppression T2-weighted imaging
SUB	Chemical shift imaging subtracted images
LR	logistic regression
ROC	Receiver operating characteristic
AUC	Area under the curve
DCA	Decision curver analysis
CP	Cortical phases
MP	Medullary phases
EP	Excretory phases
EPI	Echo planar imaging
AJCC	American joint committee on cancer
ROI	Region of interest
VOI	Volume of interest
ICC	Intraclass correlation coefficient
GLCM	Gray Level co-occurrence matrix
GLRLM	Gray level run length matrix
GLSZM	Gray level size zone matrix
GLDM	Gray level dependence matrix
NGTDM	Neighborhood gray tone difference matrix
LASSO	Least absolute shrinkage and selection operator
ANOVA	Analysis of variance
PCA	Principal component analysis
RFE	Recursive feature elimination
SVM	Support vector machine
MLP	Multilayer perceptron
LDA	Linear discriminant analysis
KNN	K-nearest neighbors
XGBoost	Extreme gradient boosting
RF	Random forest
NB	Naive bayes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-12930-2.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
Conceptualization: R.H.C, Y.Y.L, P.X.S and Y.T. Data curation: R.H.C, Q.N.S, J.X.Z and Y.T. Formal analysis: R.H.C, Y.Y.L, P.X.S. and Q.N.S. Funding acquisition: Y.T. Investigation: R.H.C, Y.Y.L, P.S.X. Methodology: R.H.C, Y.T. Project administration: Y.T. and J.X.Z. Resources: Y.T. and J.X.Z. Software: R.H.C. and Y.Y.L. Supervision: Y.T. and J.X.Z. Validation: R.H.C. Visualization: Q.N.S. Writing-original draft: R.H.C. Writing-review & editing: Y.T. Final approval of the manuscript: All authors.All authors reviewed the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China [grant numbers 82071893, 82371941 to Yan Tan]; the Research Project Supported by Shanxi Scholarship Council of China [grant number 2023 – 186 to Yan Tan]; Shanxi Province Higher Education " BillionProject " Science and Technology Guidance Project [grant number BYJL017].

Data availability
The datasets generated or analyzed during the study are not publicly available due to institutional regulations but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Academic Ethics Committee of the First Hospital of Shanxi Medical University and Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences. Patient consent was waived due to the retrospective nature of the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, First Hospital of Shanxi Medical University, No. 85 Jiefang South Road, Taiyuan, Shanxi Province 030001, P.R. China
[2]Department of Radiology, Shanxi Province Cancer Hospital/ Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/ Cancer Hospital Affiliated to Shanxi Medical University, No. 3 Workers' New Street, Taiyuan, Shanxi Province 030013, P.R. China
[3]Department of College of Medical Imaging, Shanxi Medical University, Taiyuan, Shanxi Province 030001, P.R. China
[4]Department of Shanxi Key Laboratory of Intelligent Imaging and Nanomedicine, First Hospital of Shanxi Medical University, Taiyuan, Shanxi Province 030001, P.R. China

Received: 11 August 2024 / Accepted: 10 September 2024
Published online: 27 September 2024

References

1. Motzer RJ, Jonasch E, Agarwal N, Alva A, Baine M, Beckermann K, et al. Kidney Cancer, Version 3.2022, NCCN Clinical Practice guidelines in Oncology. J Natl Compr Canc Netw. 2022;20(1):71–90.
2. Acosta-Jiménez E, Jerónimo-Guerrero D, Macías-Clavijo M, Rivera-Diez D, Hernández-Briseño L, Beltrán-Suárez E, et al. [Renal cell carcinoma: pathological prognostic factors, staging and histopathological classification of 355 cases]. Rev Med Inst Mex Seguro Soc. 2015;53(4):454–65.
3. Andreiana BC, Stepan AE, Mărgăritescu C, Al Khatib AM, Florescu MM, Ciurea RN, et al. Histopathological prognostic factors in Clear Cell Renal Cell Carcinoma. Curr Health Sci J. 2018;44(3):201–5.
4. Gudbjartsson T, Hardarson S, Petursdottir V, Thoroddsen A, Magnusson J, Einarsson GV. Histological subtyping and nuclear grading of renal cell carcinoma and their implications for survival: a retrospective nation-wide study of 629 patients. Eur Urol. 2005;48(4):593–600.
5. Patard JJ, Kim HL, Lam JS, Dorey FJ, Pantuck AJ, Zisman A, et al. Use of the University of California Los Angeles integrated staging system to predict survival in renal cell carcinoma: an international multicenter study. J Clin Oncol. 2004;22(16):3316–22.
6. Novara G, Martignoni G, Artibani W, Ficarra V. Grading systems in renal cell carcinoma. J Urol. 2007;177(2):430–6.
7. 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.
8. Xiao Q, Yi X, Guan X, Yin H, Wang C, Zhang L, et al. Validation of the World Health Organization/International Society of Urological Pathology grading for Chinese patients with clear cell renal cell carcinoma. Transl Androl Urol. 2020;9(6):2665–74.
9. Kanesvaran R, Porta C, Wong A, Powles T, Ng QS, Schmidinger M, et al. Pan-asian adapted ESMO Clinical Practice guidelines for the diagnosis, treatment and follow-up of patients with renal cell carcinoma. ESMO Open. 2021;6(6):100304.
10. Oza B, Eisen T, Frangou E, Stewart GD, Bex A, Ritchie AWS, et al. External validation of the 2003 Leibovich Prognostic score in patients randomly assigned to SORCE, an International Phase III Trial of Adjuvant Sorafenib in Renal Cell Cancer. J Clin Oncol. 2022;40(16):1772–82.
11. Shah PH, Moreira DM, Okhunov Z, Patel VR, Chopra S, Razmaria AA, et al. Positive Surgical margins increase risk of recurrence after partial nephrectomy for high risk renal tumors. J Urol. 2016;196(2):327–34.
12. Xu H, Xing Z, Ai K, Wang J, Lv Z, Deng H, et al. Patients with high nuclear grade pT1-ccRCC are more suitable for radical nephrectomy than partial nephrectomy: a multicenter retrospective study using propensity score. World J Surg Oncol. 2024;22(1):24.
13. Blumenfeld AJ, Guru K, Fuchs GJ, Kim HL. Percutaneous biopsy of renal cell carcinoma underestimates nuclear grade. Urology. 2010;76(3):610–3.

14. Ficarra V, Brunelli M, Novara G, D'Elia C, Segala D, Gardiman M, et al. Accuracy of on-bench biopsies in the evaluation of the histological subtype, grade, and necrosis of renal tumours. Pathology. 2011;43(2):149–55.
15. Lebret T, Poulain JE, Molinie V, Herve JM, Denoux Y, Guth A, et al. Percutaneous core biopsy for renal masses: indications, accuracy and results. J Urol. 2007;178:1184–8.
16. Sasaguri K, Takahashi N. CT and MR imaging for solid renal mass characterization. Eur J Radiol. 2018;99:40–54.
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77.
18. Raman AG, Fisher D, Yap F, Oberai A, Duddalwar VA. Radiomics and Artificial Intelligence: renal cell carcinoma. Urol Clin North Am. 2024;51(1):35–45.
19. Gong XQ, Tao YY, Wu YK, Liu N, Yu X, Wang R, et al. Progress of MRI Radiomics in Hepatocellular Carcinoma. Front Oncol. 2021;11:698373.
20. Niu J, Tan Q, Zou X, Jin S. Accurate prediction of glioma grades from radiomics using a multi-filter and multi-objective-based method. Math Biosci Eng. 2023;20(2):2890–907.
21. Paner GP, Stadler WM, Hansel DE, Montironi R, Lin DW, Amin MB. Updates in the Eighth Edition of the Tumor-Node-Metastasis staging classification for urologic cancers. Eur Urol. 2018;73(4):560–9.
22. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, Zhou M, Yao YF, Yang G. FeAture Explorer (FAE): a tool for developing and comparing radiomics models. PLoS ONE. 2020;15(8):e0237587.
23. Motzer RJ, Jonasch E, Agarwal N, Alva A, Bagshaw H, Baine M, et al. NCCN Guidelines® insights: kidney Cancer, Version 2.2024. J Natl Compr Canc Netw. 2024;22(1):4–16.
24. Jiang Y, Li W, Huang C, Tian C, Chen Q, Zeng X, et al. A computed tomography-based Radiomics Nomogram to Preoperatively Predict Tumor necrosis in patients with Clear Cell Renal Cell Carcinoma. Front Oncol. 2020;10:592.
25. Li Q, Liu YJ, Dong D, Bai X, Huang QB, Guo AT, et al. Multiparametric MRI Radiomic Model for Preoperative Predicting WHO/ISUP Nuclear Grade of Clear Cell Renal Cell Carcinoma. J Magn Reson Imaging. 2020;52(5):1557–66.
26. Moldovanu CG, Boca B, Lebovici A, Tamas-Szora A, Feier DS, Crisan N, et al. Preoperative Predicting the WHO/ISUP Nuclear Grade of Clear Cell Renal Cell Carcinoma by Computed Tomography-based Radiomics features. J Pers Med. 2020;11(1):8.
27. Zheng Z, Chen Z, Xie Y, Zhong Q, Xie W. Development and validation of a CT-based nomogram for preoperative prediction of clear cell renal cell carcinoma grades. Eur Radiol. 2021;31(8):6078–86.
28. Zhou H, Mao H, Dong D, Fang M, Gu D, Liu X, et al. Development and External Validation of Radiomics Approach for Nuclear Grading in Clear Cell Renal Cell Carcinoma. Ann Surg Oncol. 2020;27(10):4057–65.
29. Luo S, Wei R, Lu S, Lai S, Wu J, Wu Z, et al. 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. Eur Radiol. 2022;32(4):2340–50.
30. Cui E, Li Z, Ma C, Li Q, Lei Y, Lan Y, et al. Predicting the ISUP grade of clear cell renal cell carcinoma with multiparametric MR and multiphase CT radiomics. Eur Radiol. 2020;30(5):2912–21.
31. Rosen MA, Schnall MD. Dynamic contrast-enhanced magnetic resonance imaging for assessing tumor vascularity and vascular effects of targeted therapies in renal cell carcinoma. Clin Cancer Res. 2007;13(2):770–6.
32. Hötker AM, Mazaheri Y, Wibmer A, Zheng J, Moskowitz CS, Tickoo SK, et al. Use of DWI in the differentiation of renal cortical tumors. AJR Am J Roentgenol. 2016;206(1):100–5.
33. Chen LS, Zhu ZQ, Wang ZT, Li J, Liang LF, Jin JY, et al. Chemical shift magnetic resonance imaging for distinguishing minimal-fat renal angiomyolipoma from renal cell carcinoma: a meta-analysis. Eur Radiol. 2018;28(5):1854–61.
34. Reuter VE, Tickoo SK. Differential diagnosis of renal tumours with clear cell histology. Pathology. 2010;42(4):374–83.
35. Stepanovska Tanturovska B, Manaila R, Fabbro D, Huwiler A. Lipids as targets for renal cell carcinoma therapy. Int J Mol Sci. 2023;24(4):3272.
36. Jhaveri KS, Elmi A, Hosseini-Nik H, Hedgire S, Evans A, Jewett M, Harisinghani M. Predictive value of Chemical-Shift MRI in distinguishing Clear Cell Renal Cell Carcinoma from Non-clear Cell Renal Cell Carcinoma and Minimal-Fat Angiomyolipoma. AJR Am J Roentgenol. 2015;205(1):W79–86.
37. Moosavi B, Shabana WM, El-Khodary M, van der Pol CB, Flood TA, McInnes MD, et al. Intracellular lipid in clear cell renal cell carcinoma tumor thrombus and metastases detected by chemical shift (in and opposed phase) MRI: radiologic-pathologic correlation. Acta Radiol. 2016;57(2):241–8.
38. Pan L, Chen M, Sun J, Jin P, Ding J, Cai P, et al. Prediction of Fuhrman grade of renal clear cell carcinoma by multimodal MRI radiomics: a retrospective study. Clin Radiol. 2024;79(2):e273–81.
39. Zhang YD, Wu CJ, Wang Q, Zhang J, Wang XN, Liu XS, et al. Comparison of utility of Histogram Apparent Diffusion Coefficient and R2* for differentiation of low-Grade from High-Grade Clear Cell Renal Cell Carcinoma. AJR Am J Roentgenol. 2015;205(2):W193–201.
40. Amini B, Chenglei L, Duran-Sierra E, Wang WL, Canjirathinkal MA, Moradi H, et al. Role of Apparent Diffusion Coefficient Map-Based First- and high-order Radiomic features for the discrimination of Sacral Chordomas and Chondrosarcomas with overlapping conventional imaging features. JCO Precis Oncol. 2023;7:e2300243.

Publisher's note