BMC Medical Imaging

RESEARCH

Open Access

Multiphase CT-based deep learning radiomics nomogram models for preoperative WHO/ISUP grading of clear cell renal cell carcinoma: a two-center validation study

Chunsen Yang[1], Zhiling Zhang[2,3], Buwei Wu[1], Hongcheng Zhao[1], Tianhao Ma[1], Yunhan Luo[2,3], Wenfeng Liao[1*] and Xin Yao[1*]

Abstract

Background Preoperative determination of clear cell renal cell carcinoma (ccRCC) World Health Organization/International Society of Urological Pathology (WHO/ISUP) nuclear grade is crucial for surgical planning, yet current invasive biopsy approaches carry significant risks and diagnostic limitations. Existing radiomics and deep learning studies predominantly utilize single-phase imaging or isolated methodologies. We developed and validated a Deep Learning Radiomics Nomogram (DLRN) that integrates multiphase computed tomography (CT) imaging with machine learning for WHO/ISUP grading.

Methods This two-center study analyzed 1499 histologically confirmed ccRCC patients, allocated to training ($n=929$), internal validation ($n=398$), and external validation ($n=172$) cohorts. Our DLRN model integrates three complementary data streams: radiomics features extracted from non-contrast, corticomedullary, and nephrographic CT phases; deep learning features from a multi-channel DenseNet201 architecture; and clinical variables. Model performance was evaluated using the area under the curve (AUC) and calibration analysis. Interpretability was enhanced through Shapley Additive Explanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) analyses.

Results The DLRN model achieved favorable discriminative performance with AUC values of 0.935, 0.901, and 0.911 in training, internal validation, and external validation cohorts, respectively. This significantly outperformed individual component models in external validation: clinical (AUC = 0.730), radiomics (AUC = 0.845), and deep learning (AUC = 0.868) models ($p<0.05$). Calibration analysis demonstrated excellent agreement between predicted probabilities and observed outcomes. SHAP analysis revealed deep learning features as dominant predictive contributors, while Grad-CAM visualization consistently focused on tumor heterogeneity patterns characteristic of different grades.

*Correspondence:
Wenfeng Liao
liaowf2011@163.com
Xin Yao
yaoxin@tjmuch.com

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

Conclusion The DLRN model based on multiphase CT imaging provides an accurate, non-invasive tool for preoperative prediction of ccRCC nuclear grade.

Keywords Clear cell renal cell carcinoma, Computed tomography, Deep learning, Radiomics, WHO/ISUP grade

Introduction

Renal cell carcinoma ranks among the most prevalent malignancies of the urinary system, with clear cell renal cell carcinoma (ccRCC) as the predominant subtype, accounting for 70–75% of renal cell carcinoma (RCC) cases and representing the primary cause of RCC-related mortality [1]. In recent years, the detection rate of ccRCC has increased annually, largely due to the advancement of imaging technology and the improvement of diagnostic methods [2]. Surgery remains the principal treatment for localized ccRCC; nevertheless, approximately 30% of patients experience tumor recurrence and mortality following radical surgery [3]. The prognosis of ccRCC is influenced by various factors, with pathological nuclear grading playing a pivotal role in determining outcomes [4]. The WHO/ISUP grading system categorizes ccRCC into grades I to IV, with high-grade (grade III/IV) tumors exhibiting more aggressive characteristics and a higher risk of metastasis and poorer clinical outcomes in comparison to low-grade (grade I/II) [5, 6]. Furthermore, the heterogeneity of high-grade and low-grade ccRCC in terms of tumor progression affects the available treatment options [7]. A retrospective study suggests that opting for radical nephrectomy (RN), as opposed to partial nephrectomy (PN), is more advantageous for local recurrence-free survival and distant metastases-free survival in patients with high nuclear grade pT1-ccRCC [8]. Consequently, accurate preoperative prediction of pathologic nuclear grading is imperative for optimally adjusting treatment options. Currently, the sole method for acquiring preoperative ccRCC pathologic nuclear grading is through puncture biopsy. However, this invasive procedure puts patients at risk of complications such as hemorrhage, infection, and implantation metastasis [9, 10]. Moreover, percutaneous biopsies fail to provide definitive diagnostic results in approximately 20% of cases, with about 10% of biopsy results being inconsistent with the final pathological diagnosis of resected tissues [11]. This discrepancy is likely due to the significant spatial heterogeneity of ccRCC tumors, where biopsy samples represent only localized features of the tumor, failing to fully reflect the overall pathological grade and thereby reducing diagnostic accuracy [12]. Consequently, there is an urgent need to develop a non-invasive, cost-effective, and efficient method for preoperative prediction of pathologic nuclear grading of ccRCC.

Computed tomography (CT) is the most commonly used imaging modality for the diagnosis of ccRCC, as it is capable of providing rich anatomical information about the kidney and tumor [13]. Nevertheless, conventional CT features are limited in their capacity to assess tumor staging and nuclear grading. Recently, radiomics and deep learning (DL) techniques have been extensively applied in medical imaging due to their capacity to comprehensively characterize tumor heterogeneity [14]. The use of radiomics and DL methodologies, based on CT imaging, has been demonstrated to facilitate the assessment of the clinical progression of renal cancer by distinguishing between benign and malignant neoplasms, predicting tumor pathologic subtypes and patient prognosis [15–17]. However, most previous studies have predominantly relied on radiomic features from one or two CT phases to predict ccRCC nuclear grading, failing to fully exploit the potential of multiphase CT imaging and three-dimensional data, resulting in models with limited performance and generalizability [18]. Moreover, these previous reports only focus on radiomics features and clinical factors, excluding the deep learning features, which contain valuable information. [19, 20]. So far, there is no literature available on combined multiphase CT radiomic features and multi-channel deep learning for preoperative prediction of ccRCC pathologic nuclear grading.

In this study, we develop and validate a noninvasive preoperative prediction model for ccRCC nuclear grading based on radiomic features and DL of patients' preoperative CT. This model aims to serve as an effective adjunctive diagnostic tool for clinicians, enabling a more accurate assessment of renal cancer malignancy before surgery. Consequently, it holds the potential to optimize treatment options and improve patient prognosis.

Materials and methods

Patients

This study initially enrolled 1919 histologically confirmed ccRCC patients who underwent partial or radical nephrectomy at Center 1 from January 2014 to December 2023. Exclusion criteria removed 128 patients with incomplete clinical/pathological data, 363 lacking analyzable contrast-enhanced CT scans within one preoperative month, 16 receiving neoadjuvant/adjuvant therapy, and 85 with multifocal or bilateral tumors, yielding a final cohort of 1327 patients. Pathologic nuclear grading stratified these into low-grade (I-II, $n = 965$) and high-grade (III-IV, $n = 475$) subgroups, which were randomly split into training ($n = 929$) and internal validation ($n = 398$) sets at a 7:3 ratio. A separate cohort from Center 2 ($n = 172$) meeting identical criteria served as

the external validation set. Figure 1 details the enrollment workflow, while Table 1 summarizes baseline characteristics including age, sex, tumor size, and stage. The institutional review board of Center 1 approved this retrospective study using de-identified data, waiving individual consent per Helsinki Declaration guidelines. All participating centers provided institutional consent, and the study complied with transparent reporting standards for multivariate predictive models.

Histopathological assessment

Renal clear cell carcinoma specimens were evaluated according to WHO Classification of Tumors of the Urinary System guidelines. All cases underwent systematic re-examination by an experienced genitourinary

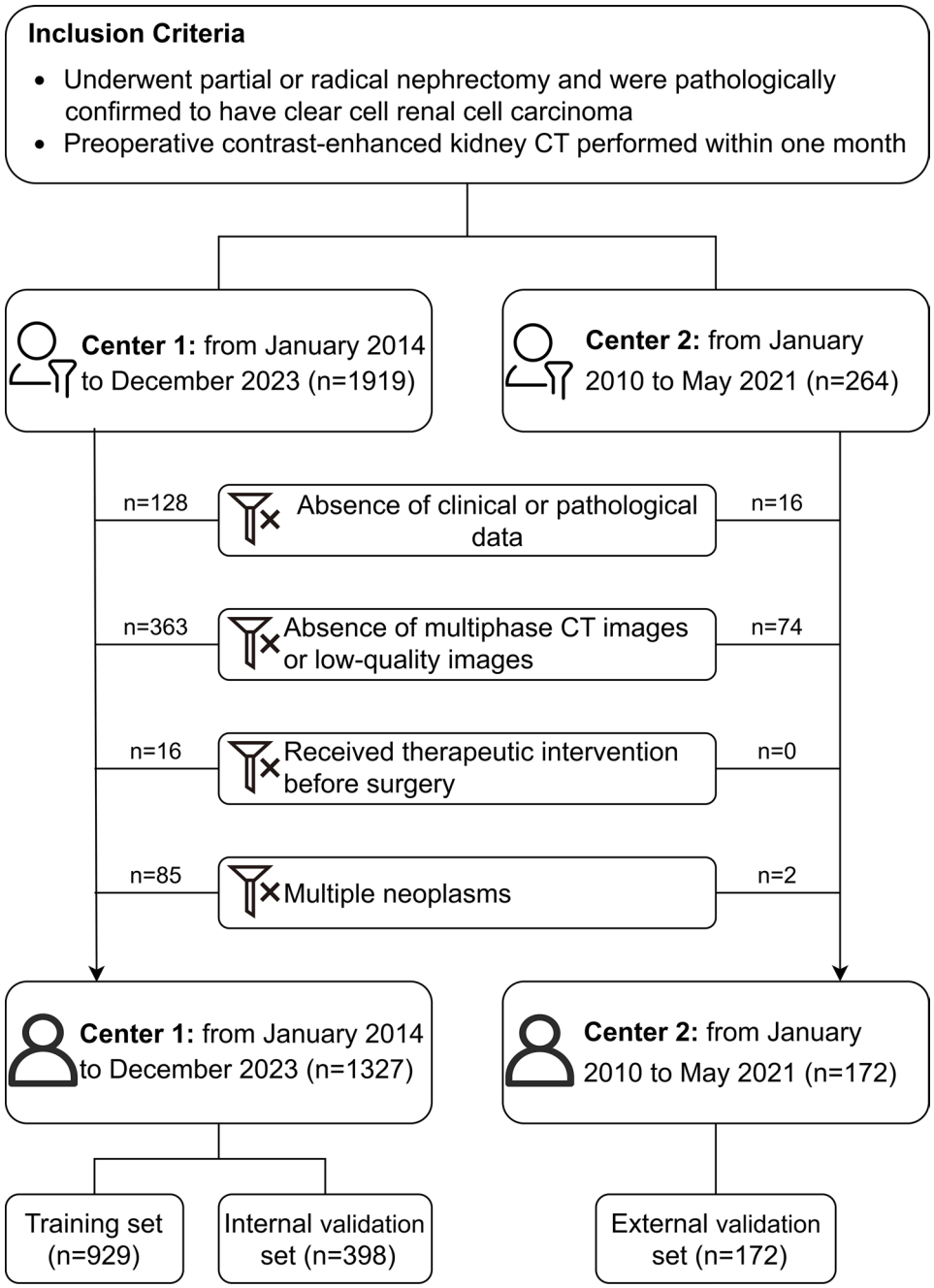

Fig. 1 Flowchart of the study. Two-center study enrolled histologically confirmed clear cell renal cell carcinoma patients who underwent nephrectomy. Inclusion criteria required preoperative contrast-enhanced kidney CT within one month. Exclusion criteria included incomplete data, missing multiphase CT, preoperative therapy, and multiple neoplasms. Center 1 (2014–2023) provided 1327 patients for training ($n=929$) and internal validation ($n=398$). Center 2 (2010–2021) provided 172 patients for external validation

Table 1 Baseline characteristics of patients with clear cell renal cell carcinoma

Variables	All patients ($n=1499$)	Training set ($n=929$)	Internal validation set ($n=398$)	External validation set ($n=172$)
nuclear grade, n (%)				
Low-grade	1025(68.38)	619(66.63)	270(67.84)	136(79.07)
High-grade	474(31.62)	310(33.37)	128(32.16)	36(20.93)
Age (years), mean ± SD	56.58 ± 10.48	57.07 ± 10.19	57.19 ± 10.08	52.55 ± 11.97
Tumor size (mm)	55.10 ± 25.93	53.41 ± 25.68	51.92 ± 24.05	71.59 ± 25.61
Sex, n (%)				
Male	962(64.18)	601(64.69)	256(64.32)	105(61.05)
Female	537(35.82)	328(35.31)	142(35.68)	67(38.95)
Surgery procedure, n (%)				
Partial nephrectomy	441(29.42)	259(27.88)	118(29.65)	64(37.21)
Radical nephrectomy	1058(70.58)	670(72.12)	280(70.35)	108(62.79)
cT stage, n (%)				
T1a	451(30.09)	299(32.19)	141(35.43)	11(6.40)
T1b	672(44.83)	422(45.43)	169(42.46)	81(47.09)
T2	261(17.41)	137(14.75)	57(14.32)	67(38.95)
T3/T4	115(7.67)	71(7.64)	31(7.79)	13(7.56)
cN stage, n (%)				
N0/Nx	1377(91.86)	865(93.11)	372(93.47)	140(81.40)
N1	122(8.14)	64(6.89)	26(6.53)	32(18.60)
cTNM stage, n (%)				
I	1071(71.45)	693(74.60)	299(75.13)	79(45.93)
II	209(13.94)	119(12.81)	47(11.81)	43(25.00)
III	216(14.41)	115(12.38)	52(13.07)	49(28.49)
IV	3(0.20)	2(0.22)	0	1(0.58)
Tumor thrombus, n (%)				
Negative	1385(92.39)	858(92.36)	367(92.21)	160(93.02)
Positive	114(7.61)	71(7.64)	31(7.79)	12(6.98)
pT stage, n (%)				
T1a	491(32.76)	292(31.43)	140(35.18)	59(34.30)
T1b	617(41.16)	402(43.27)	151(37.94)	64(37.21)
T2	183(12.21)	117(12.59)	45(11.31)	21(12.21)
T3/T4	208(13.88)	118(12.70)	62(15.58)	28(16.28)
pN stage, n (%)				
N0/Nx	1492(99.53)	926(99.68)	395(99.25)	171(99.42)
N1	7(0.47)	3(0.32)	3(0.75)	1(0.58)
pTNM stage, n (%)				
I	1143(76.25)	718(77.29)	300(75.38)	125(72.67)
II	150(10.01)	93(10.01)	36(9.05)	21(12.21)
III	205(13.68)	118(12.70)	62(15.58)	25(14.53)
IV	1(0.07)	0	0	1(0.58)
Pathological necrosis				
Negative	1311(87.46)	815(87.73)	349(87.69)	147(85.47)
Positive	188(12.54)	114(12.27)	49(12.31)	25(14.53)

pathologist (Y-H. Z) with 10 years of specialized expertise in renal neoplasm diagnostics. WHO/ISUP grading (1–4) was determined through comprehensive microscopic analysis of nuclear morphology, including nuclear size, contour irregularity, and nucleolar prominence. Each case received independent review of original diagnoses alongside clinical and radiological correlation to ensure accurate grade assignment. For multi-institutional validation, histological and clinical data were collected from collaborating centers under designated pathologist supervision. Standardized grading criteria were uniformly applied across all sites, with centralized review confirming final WHO/ISUP grade assignments to maintain diagnostic consistency.

CT examination and image segmentation

To maintain image quality, a junior radiologist independently verified that all CT scans fully captured the lesion

across all phases. To minimize inter-phase registration errors, images from all phases were first resampled to a uniform voxel size of $1 \times 1 \times 1\,mm^3$. Subsequently, using the "SlicerElastix" module in 3D Slicer software (version 5.2.2; https://www.slicer.org), the Non-Contrast (NCP) and Nephrographic (NP) phases were rigidly registered to the Corticomedullary Phase (CMP) using B-spline interpolation. The alignment accuracy was independently verified by a radiologist to ensure spatial consistency across phases. Renal mass segmentation employed the nnU-Net network pretrained on the KiTS23 challenge dataset. To ensure the model's generalization capability was assessed fairly, we used the pretrained weights directly for inference without fine-tuning on our internal dataset [21]. Following automated segmentation,an abdominal radiologist with over five years of experience reviewed all segmentations; cases with failed automated segmentation were manually annotated by this specialist and validated by a second radiologist with two decades of abdominal imaging expertise. Additional scanning and segmentation protocols are documented in Supplementary Material S1 and Supplementary Table S1.

Features selection and radiomics models construction

The PyRadiomics package in Python (version 3.7.12) was utilized for the extraction of radiomics features. Seven types of image features were extracted: (1) first-order features, (2) shape features, (3) gray-level co-occurrence matrix features, (4) gray-level dependence matrix features, (5) gray-level run-length matrix features, (6) gray-level size-zone matrix features, and (7) neighborhood gray-tone difference matrix features. Data were standardized using the StandardScaler algorithm of scikit-learn packages for subsequent feature selection and model training. The reliability of these radiomics features was then assessed by calculating inter- and intra-reader intraclass correlation coefficient (ICC) analysis, where features with an ICC value greater than 0.85 were considered to be reliable, as detailed in the methodology described in Supplementary Material S2. The Mann–Whitney U-test, Pearson correlation analysis and maximum relevance minimum redundancy (mRMR) algorithm were used to eliminate irrelevant and redundant features, and then the least absolute shrinkage and selection operator (LASSO) logistic regression model was used to select optimal radiomic (Rad)-related features with nonzero coefficients. Four subgroups of Rad-related features were developed for the construction of four corresponding Rad prediction models, including three basic models (Rad-NCP, Rad-CMP, and Rad-NP) and one composite model (Rad-Triphasic).

The eXtreme Gradient Boosting (XGBoost) classifier was applied based on the best subset of features selected by LASSO to construct the radiomic models. The Synthetic Minority Oversampling Technique (SMOTE) oversampling technique was used to resolve data imbalance between the high and low-grade groups during model fitting. The predictive power of the radiomic models was quantified by receiver operating characteristic (ROC) analysis, and the area under the curve (AUC) was calculated concurrently. To identify the model with the best performance, their AUCs were compared using the DeLong test in the training cohort and the internal and external validation cohorts.

Establishment of deep learning models

We adjusted CT images to a soft tissue window (width: 350; level: 50) and normalized grayscale values to the range [0,255]. For each patient, the axial slice with the largest tumor cross-section served as the reference image, with two additional slices selected 2 mm above and below this plane. We adopted this "2.5D" multi-channel input strategy to effectively capture volumetric tumor heterogeneity and inter-slice spatial correlations. This approach offers a balance between feature richness and computational efficiency, avoiding the excessive computational resources and overfitting risks often associated with full 3D neural networks on datasets of this magnitude. Each image was cropped to the minimal bounding box enclosing the region of interest to improve computational efficiency. The five selected images from each phase (NCP, CMP, NP) were resampled to 224×224 pixels and concatenated into a 15-channel input tensor, which we normalized using Z-scores. During training, we applied real-time augmentation via random cropping and axial flipping, whereas validation used only normalized images.

We conducted a comparative performance analysis of several deep learning models initialized with ImageNet pre-trained weights. Transfer learning with these pre-trained weights enhanced clinical applicability across heterogeneous populations. Model optimization utilized stochastic gradient descent (SGD) with softmax cross-entropy loss and a cosine-decay learning rate schedule (initial learning rate: 0.01) over 200 epochs to achieve optimal convergence while maintaining generalization capability. Complete hyperparameter specifications are detailed in Supplementary Material S3.

Deep learning radiomics nomogram (DLRN) model construction and validation

Features extracted from the penultimate layer of the deep learning model underwent dimensionality reduction to 256 components using Principal Component Analysis (PCA), necessitated by the inherent complexity of neural networks. These resulting deep learning features were subsequently integrated with radiomics features via a feature fusion algorithm, yielding deep learning

radiomics (DLR) features. Following standard radiomics protocols for feature selection and model construction, the area under the receiver operating characteristic curve (AUC) served as the primary metric for quantitatively comparing the performance of different feature fusion methods. Univariate and stepwise multivariate analyses were employed to screen all features, aiming to identify those with clinical significance. Finally, the Deep Learning Radiomics Nomogram (DLRN) was constructed via multivariate logistic regression by integrating the predictive probability scores generated by the Clinical, Rad, DL, and DLR models. Unlike the component models which employed specific feature selection techniques, the DLRN serves as an ensemble decision tool that utilizes the probability outputs of these sub-models as inputs without requiring additional feature selection at the nomogram level. The resulting integrated model was visualized as a nomogram.

Visualization and interpretability of DLRN models

To clarify the model's decision-making process, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM) with Shapley Additive Explanations (SHAP). Grad-CAM identified critical regions in target images by extracting gradient information from the convolutional neural network's (CNN's) final convolutional layer, generating class activation maps that highlight features relevant to classification. SHAP analysis quantified each feature's global contribution to predictions, with absolute values indicating importance and signs reflecting directional effects. Together, these methods provide both spatial visualization of salient features and systematic assessment of feature importance, enhancing model interpretability. This combined approach strengthens the analytical framework while preserving methodological precision.

Statistical analysis

Continuous variables were reported as mean ± standard deviation and analyzed using Student's t-test or Mann-Whitney U test, while categorical variables were assessed with chi-square or Fisher's exact tests. Statistical significance was defined as $p < 0.05$ (two-tailed). All statistical analyses were conducted in R (v4.3.2) and SPSS (v26.0). Radiomic features were extracted using PyRadiomics v3.0.1, and machine learning algorithms were implemented through Scikit-learn v1.0.2. The deep learning architecture employed PyTorch v1.11.0 with CUDA v11.3.1 and cuDNN v8.2.1 acceleration. We evaluated model stability using 5-fold cross-validation during training. Model performance was measured by the area under the curve (AUC) and further examined through calibration curves. Figure 2 illustrates the overall study design.

Results

Clinical characteristics

This study included a total of 1499 patients with postoperative pathologically confirmed ccRCC. The average age of the cohort was 56.58 ± 10.48 years, with males representing 64.1% of the population. Radical nephrectomy

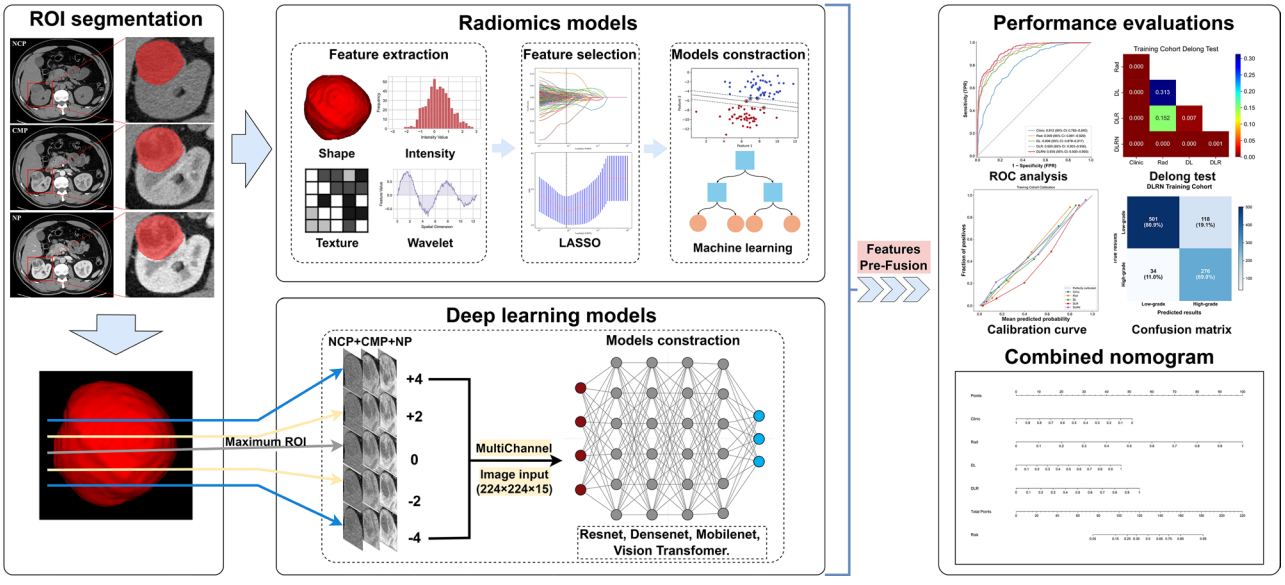

Fig. 2 Workflow of deep learning radiomics nomogram model development. Pipeline includes: (1) ROI segmentation from multiphase CT; (2) radiomics feature extraction with LASSO selection; (3) deep learning using 15-channel input (5 slices × 3 phases) with multiple architectures; (4) performance evaluation using ROC analysis, calibration curves, and confusion matrices. Feature fusion combined radiomics and deep learning features into final DLRN model for WHO/ISUP grading prediction. ROI, region of interest; NCP, non-contrast phase; CMP, corticomedullary phase; NP, nephrographic phase; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating curve; DLRN, deep learning radiomics nomogram

was the predominant surgical approach, performed in 70.6% of the cases. The clinical T stage distribution showed that the majority of patients were in T1a and T1b stages, comprising 30.1% and 44.8% of the cohort, respectively, while 7.7% were classified as clinical T3a or higher. Most patients were in clinical TNM stage I, accounting for 71.5% of the cases. Detailed clinicopathological characteristics are presented in Table 1, and preoperative clinical features used for modeling are listed in Supplementary Table S5. Table 2 outlines the differences in clinical variables between low-grade and high-grade ccRCC patients across the training, internal validation, and external validation sets. In the training cohort, high-grade ccRCC patients exhibited larger tumor sizes ($p < 0.001$), a higher incidence of venous thromboembolism ($p < 0.001$), and more advanced clinical and pathological TNM stages ($p < 0.001$) compared to their low-grade counterparts.

A multivariate logistic regression analysis was conducted using preoperative clinical information. Initially, candidate variables including gender, age, body mass index (BMI), tumor diameter, clinical T stage, clinical N stage, and clinical TNM stage were screened using univariate analysis. Only those showing statistical significance ($p < 0.05$), specifically gender, age, and tumor diameter, were retained. The analysis identified gender (odds ratio [OR]: 0.51, 95% CI: 0.39–0.67), age (OR: 0.95, 95% CI: 0.95–0.96), and tumor diameter (OR: 1.03, 95% CI: 1.02–1.03) as significant factors incorporated into the clinical model (Supplementary Table S5).

Development and validation of radiomic models

As shown in Supplementary Table S2, a total of 5502 radiomic features were extracted from the NCP, CMP, and NP regions of interest (ROIs). With the XGBoost classifier, three basic and one composite prediction models were developed, and their AUC values were calculated for predicting WHO/ISUP grade in the training, internal validation, and external validation cohorts. Supplementary Table S3 displays the AUC, accuracy, sensitivity, specificity, and positive and negative predictive values of all four models for predicting WHO/ISUP nuclear grading in the three cohorts. Radiomic feature selection results for different models and feature correlation coefficients are shown in Supplementary Fig. S1.

Of note, the Rad-Triphasic model showed favorable predictive efficacy in the training cohort (AUC = 0.909, 95% CI: 0.890–0.928), internal validation cohort (AUC = 0.863, 95% CI: 0.826–0.901), and external validation cohort (AUC = 0.845, 95% CI: 0.763–0.927) (Supplementary Fig. S2A-C). The calibration curves of the four models (Supplementary Fig. S2D–F) provide further evidence of the superior calibration performance of the Rad-Triphasic model. Specifically, its predicted probabilities demonstrated the closest alignment with the ideal calibration line across all three cohorts, suggesting robust probability estimation and excellent generalizability.

Compared with the other three models, although the AUCs of the Rad-Triphasic model performed comparably to the other models in the internal validation cohort ($p > 0.05$), they were significantly higher in the training cohort and external validation cohort ($p < 0.05$) for all comparisons according to the DeLong test (Supplementary Fig. S2G-I). The Rad-Triphasic model was finally determined to be the best performing radiomic prediction model. It was composed of five NCP features, eight CMP features and seven NP features that were selected with the LASSO logistic regression model (Supplementary Fig. S1).

Deep learning models performance

We developed and evaluated nine deep learning models, including ResNet, DenseNet, MobileNet, and Vision Transformer (ViT) architectures. Supplementary Table S4 summarizes their performance across training, internal validation, and external validation datasets. The DenseNet201 model demonstrated superior performance on the external validation set and was therefore chosen as the base architecture for our fusion model. This model achieved AUC values of 0.898, 0.907, and 0.868 on the training, internal validation, and external validation sets, respectively.

DLRN model Development, comparison, and evaluation

To enhance the predictive capability for ccRCC nuclear grading, we developed an integrated Deep Learning Radiomics Nomogram (DLRN) model that systematically combines clinical variables, radiomic features, and deep learning features. Initially, 256 deep learning features were extracted from the penultimate layer of the optimal DenseNet201 architecture and subsequently integrated with radiomic features through a feature fusion algorithm to construct the deep learning radiomics (DLR) model. The final DLRN model was developed by incorporating the predictive probabilities derived from the clinical model (based on 3 clinical factors), Rad model (based on 20 radiomics features), DL model (based on 256 deep learning features), and DLR model through logistic regression, with the integrated model visualized as a comprehensive nomogram (Fig. 5A).

The DLRN model demonstrated superior discriminative performance across all validation cohorts through ROC analysis (Fig. 3A–C). In the training cohort, the DLRN achieved the highest AUC of 0.935 (95% CI: 0.920–0.951), significantly outperforming the clinical model (AUC = 0.812), radiomic model (AUC = 0.900), deep learning model (AUC = 0.895), and DLR model (AUC = 0.920). Internal validation confirmed robust

Table 2 Comparison of clinical and pathologic factors between low-grade and high-grade ccRccs in the training, internal validation, and external validation sets

Variables	Training set (*n* = 929)			Internal validation set (*n* = 398)			External validation set (*n* = 172)		
	Low-grade (*n* = 619)	High-grade (*n* = 310)	P-value	Low-grade (*n* = 270)	High-grade (*n* = 128)	P-value	Low-grade (*n* = 136)	High-grade (*n* = 36)	P-value
Age (years), mean ± SD	56.19 ± 10.51	58.84 ± 9.29	0.002	56.34 ± 10.29	58.97 ± 9.41	0.019	52.39 ± 12.68	53.14 ± 8.91	0.739
Tumor size (mm)	45.40 ± 20.59	69.40 ± 27.34	<0.001	42.75 ± 17.56	71.27 ± 24.45	<0.001	66.22 ± 21.46	91.86 ± 29.90	<0.001
Sex, n (%)			0.030			<0.001			0.841
Male	385(62.20)	216(69.68)		157(58.15)	99(77.34)		82(60.29)	23(63.89)	
Female	234(37.80)	94(30.32)		113(41.85)	29(22.66)		54(39.71)	13(36.11)	
Surgery procedure, n (%)			<0.001			<0.001			0.008
Partial nephrectomy	236(38.13)	23(7.42)		111(41.11)	7(5.47)		58(42.65)	6(16.67)	
Radical nephrectomy	383(61.87)	287(92.58)		159(58.89)	121(94.53)		78(57.35)	30(83.33)	
cT stage, n (%)			<0.001			<0.001			<0.001
T1a	274(44.26)	25(8.06)		131(48.52)	10(7.81)		9(6.62)	2(5.56)	
T1b	283(45.72)	139(44.84)		121(44.81)	48(37.50)		73(53.68)	8(22.22)	
T2	44(7.11)	93(30.00)		14(5.19)	43(33.59)		48(35.29)	19(52.78)	
T3/T4	18(2.91)	53(17.10)		4(1.48)	27(21.09)		6(4.41)	7(19.44)	
cN stage, n (%)			<0.001			<0.001			<0.001
N0/Nx	596(96.28)	269(86.77)		261(96.67)	111(86.72)		119(87.50)	21(58.33)	
N1	23(3.72)	41(13.23)		9(3.33)	17(13.28)		17(12.50)	15(41.67)	
cTNM stage, n (%)			<0.001			<0.001			<0.001
I	539(87.08)	154(49.68)		247(91.48)	52(40.62)		71(52.21)	8(22.22)	
II	41(6.62)	78(25.16)		11(4.07)	36(28.12)		36(26.47)	7(19.44)	
III	39(6.30)	76(24.52)		12(4.44)	40(31.25)		28(20.59)	21(58.33)	
IV	0	2(0.65)		0	0		1(0.74)	0	
Tumor thrombus, n (%)			<0.001			<0.001			0.003
Negative	601(97.09)	257(82.90)		266(98.52)	101(78.91)		131(96.32)	29(80.56)	
Positive	18(2.91)	53(17.10)		4(1.48)	27(21.09)		5(3.68)	7(19.44)	
pT stage, n (%)			<0.001			<0.001			<0.001
T1a	269(43.46)	23(7.42)		130(48.15)	10(7.81)		55(40.44)	4(11.11)	
T1b	275(44.43)	127(40.97)		114(42.22)	37(28.91)		54(39.71)	10(27.78)	
T2	43(6.95)	74(23.87)		13(4.81)	32(25.00)		11(8.09)	10(27.78)	
T3/T4	32(5.17)	86(27.74)		13(4.81)	49(38.28)		16(11.76)	12(33.33)	
pN stage, n (%)			0.066			0.507			0.474
N0/Nx	619(100.00)	307(99.03)		269(99.63)	126(98.44)		136(100.00)	35(97.22)	
N1	0	3(0.97)		1(0.37)	2(1.56)		0	1(2.78)	
pTNM stage, n (%)			<0.001			<0.001			<0.001
I	555(89.66)	163(52.58)		249(92.22)	51(39.84)		110(80.88)	15(41.67)	
II	32(5.17)	61(19.68)		8(2.96)	28(21.88)		11(8.09)	10(27.78)	
III	32(5.17)	86(27.74)		13(4.81)	49(38.28)		14(10.29)	11(30.56)	
IV	0	0		0	0		1(0.74)	0	
Pathological necrosis			<0.001			<0.001			<0.001
Negative	608(98.22)	207(66.77)		266(98.52)	83(64.84)		126(92.65)	21(58.33)	
Positive	11(1.78)	103(33.23)		4(1.48)	45(35.16)		10(7.35)	15(41.67)	

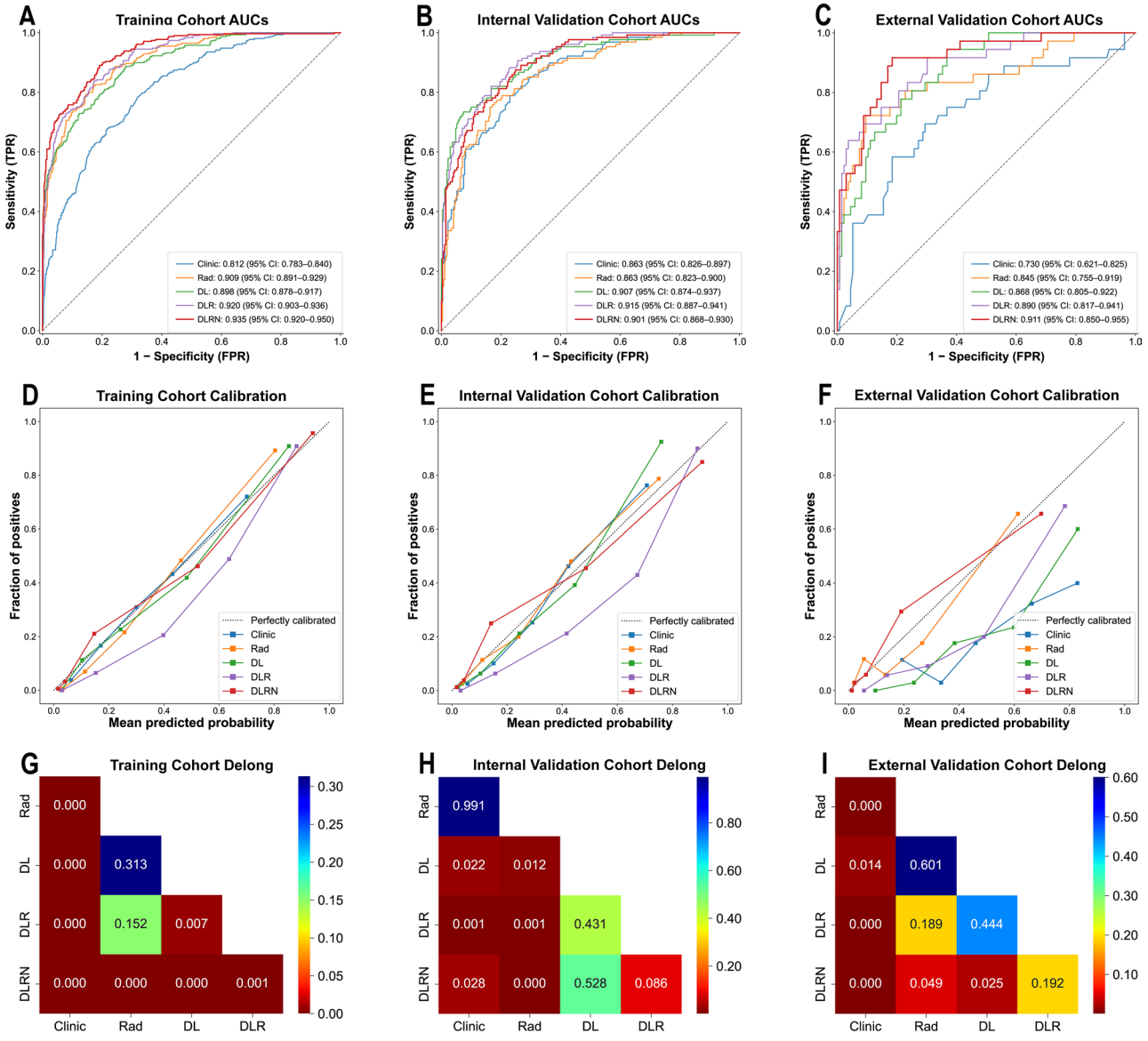

Fig. 3 Performance comparison of integrated models. **A–C**) ROC curves for training, internal validation, and external validation cohorts. **D–F**) Calibration curves. **G–I**) DeLong test heat maps. DLRN model achieved highest performance with AUC values of 0.935 (95% CI: 0.920–0.951), 0.901 (95% CI: 0.866–0.930), and 0.911 (95% CI: 0.850–0.955) in training, internal validation, and external validation cohorts, respectively. Clinic, clinical variables model; radiomics, radiomics features model; DL, deep learning features model; DLR, deep learning radiomics model; DLRN, deep learning radiomics nomogram; AUC, area under the curve; CI, confidence interval

generalizability with an AUC of 0.901 (95% CI: 0.866–0.936), maintaining superiority over component models. Critically, external validation demonstrated cross-institutional reliability, achieving an AUC of 0.911 (95% CI: 0.890–0.925), substantially outperforming individual models: clinical (AUC = 0.730), radiomic (AUC = 0.845), deep learning (AUC = 0.868), and DLR (AUC = 0.890).

DeLong's test revealed distinct patterns of statistical significance across validation cohorts (Fig. 3G–I). In the training cohort, the DLRN demonstrated statistically significant superiority over all component models ($p < 0.001$ for all comparisons). However, in the internal validation cohort, while the DLRN maintained significant advantages over clinical ($p = 0.028$) and radiomic ($p < 0.001$) models, differences with deep learning ($p = 0.528$) and DLR ($p = 0.086$) models were not statistically significant. In the external validation cohort, the DLRN showed significant improvements over clinical ($p < 0.001$), radiomic ($p = 0.049$), and deep learning ($p = 0.025$) models, while the comparison with the DLR model approached but did not reach statistical significance ($p = 0.192$).

Calibration analysis revealed excellent agreement between predicted probabilities and observed outcomes across all cohorts (Fig. 3D–F), with the DLRN model

exhibiting superior calibration performance. Comprehensive performance metrics substantiated the DLRN model's clinical utility (Table 3). In the external validation cohort, the DLRN achieved balanced performance with an accuracy of 83.1%, sensitivity of 75.0%, and specificity of 85.3%. The positive predictive value of 57.4% and negative predictive value of 92.8% indicate particular strength in accurately identifying low-grade ccRCC cases.

Confusion matrix analysis revealed excellent predictive accuracy across all cohorts (Fig. 4B–D). In the training cohort, the DLRN correctly classified 501 of 619 low-grade cases (80.9%) and 276 of 310 high-grade cases (89.0%). Internal and external validation maintained consistent performance, demonstrating remarkable cross-institutional stability. The superior performance can be attributed to synergistic integration of complementary information sources, with the nomogram visualization facilitating clinical interpretation through intuitive scoring systems. The consistent performance advantages, combined with excellent calibration and robust cross-institutional validation, establish the DLRN model as a reliable, non-invasive tool for preoperative ccRCC nuclear grading prediction.

DLRN model interpretation and visualization

To enhance model interpretability, we employed SHAP analysis and Grad-CAM visualization to elucidate the DLRN model's decision-making process. SHAP analysis revealed that deep learning features dominated the top predictive contributors, with DL_2 exhibiting the highest mean SHAP value (1.51), followed by square_glcm_Idn_CMP (1.01) (Fig. 5A–B).

Grad-CAM visualization demonstrated that the model consistently focused on heterogeneous regions within tumors across all CT phases (Fig. 5C–D). In high-grade ccRCC cases, the model particularly emphasized areas of irregular enhancement and necrosis, while in low-grade cases, attention concentrated on more homogeneous tumor regions. This spatial attention pattern aligned with established pathological characteristics, where high-grade tumors typically exhibit greater heterogeneity and necrotic components. The integration of SHAP and Grad-CAM analyses not only validated the model's clinical relevance but also provided transparent insights into feature contributions, facilitating clinical acceptance and trust in AI-assisted diagnosis.

Discussion

This study developed and externally validated a multiphase CT-based deep learning radiomics nomogram (DLRN) that integrates radiomic features from multiple CT phases with deep learning features and clinical characteristics for preoperative prediction of the WHO/ISUP nuclear grading of ccRCC. Including 1499 cases from two institutions, this study had a larger sample size compared to many previous similar studies, thereby enhancing the model's generalization performance and effectively reducing the risk of overfitting. The results demonstrated that the DLRN model, which combines multiphase deep learning radiomic features with clinical characteristics, exhibited superior predictive performance. Specifically, the model achieved AUC values of 0.935, 0.901, and 0.911 in the training, internal validation, and external validation sets, respectively, significantly outperforming standalone clinical, radiomic, or deep learning models. The large sample size and two-center design contributed to the high consistency and stability of the model across data from different institutions. This approach provides a potentially useful, non-invasive tool for preoperative pathological grading assessment in ccRCC patients, aiding in the evaluation of renal cancer malignancy before

Table 3 Predictive performance metrics of different models in training, internal validation, and external validation cohorts

Model	Study cohort	AUC (95% CI)	ACC	SEN	SPE	PPV	NPV	Recall	F1
DLRN	Training	0.935 (0.920–0.951)	0.836	0.890	0.809	0.701	0.936	0.890	0.784
	Internal validation	0.901 (0.870–0.931)	0.791	0.875	0.752	0.626	0.927	0.875	0.730
	External validation	0.911 (0.859–0.963)	0.837	0.917	0.816	0.569	0.974	0.917	0.702
DLR	Training	0.920 (0.903–0.938)	0.831	0.839	0.827	0.708	0.911	0.839	0.768
	Internal validation	0.915 (0.887–0.942)	0.829	0.820	0.833	0.700	0.907	0.820	0.755
	External validation	0.890 (0.829–0.950)	0.744	0.917	0.699	0.446	0.969	0.917	0.600
DL	Training	0.898 (0.878–0.919)	0.776	0.887	0.721	0.614	0.927	0.887	0.726
	Internal validation	0.907 (0.875–0.939)	0.864	0.734	0.926	0.825	0.880	0.734	0.777
	External validation	0.868 (0.810–0.927)	0.698	0.944	0.632	0.405	0.977	0.944	0.567
Rad	Training	0.909 (0.890–0.928)	0.826	0.823	0.827	0.704	0.903	0.823	0.759
	Internal validation	0.863 (0.826–0.901)	0.802	0.773	0.815	0.664	0.884	0.773	0.715
	External validation	0.845 (0.763–0.927)	0.866	0.722	0.904	0.667	0.925	0.722	0.693
Clinic	Training	0.812 (0.784–0.841)	0.716	0.784	0.682	0.552	0.863	0.784	0.648
	Internal validation	0.863 (0.826–0.900)	0.776	0.789	0.770	0.620	0.885	0.789	0.694
	External validation	0.730 (0.631–0.829)	0.703	0.694	0.706	0.385	0.897	0.694	0.495

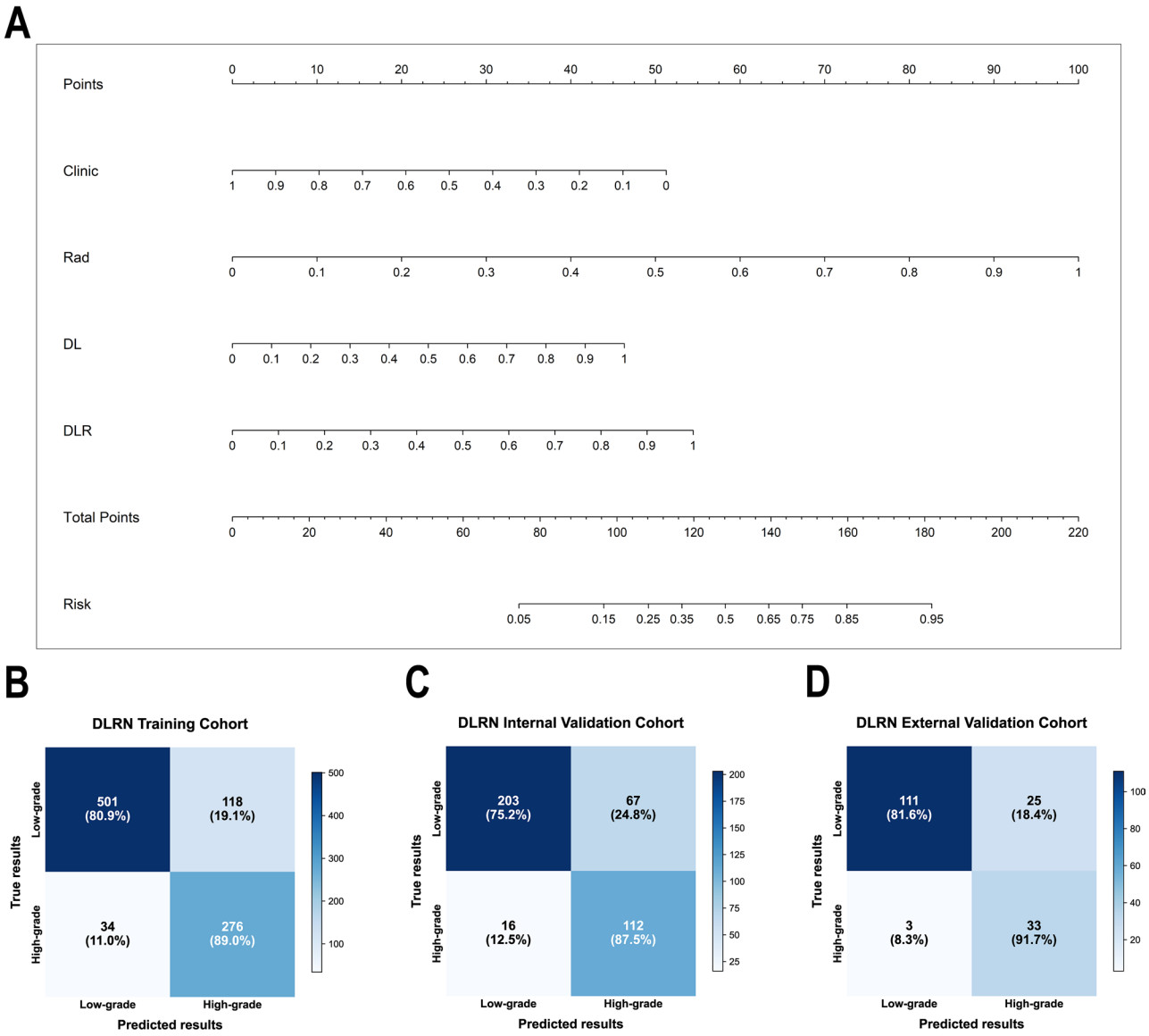

Fig. 4 Clinical nomogram and model performance evaluation. **A**) Clinical nomogram incorporating clinical variables, radiomics, deep learning, and DLR features with point scales (0–100) and risk probabilities (0.05–0.95). **B–D**) Confusion matrices for DLRN model showing classification accuracy. Model demonstrated excellent performance with sensitivity (89.0%, 87.5%, 91.7%) and specificity (80.9%, 75.2%, 81.6%) across all cohorts. Clinic, clinical variables model; radiomics, radiomics features model; DL, deep learning features model; DLR, deep learning radiomics model; DLRN, deep learning radiomics nomogram

surgery, optimizing treatment plans, and improving patient prognosis.

Radiomics refers to the comprehensive quantitative analysis of disease phenotypes through the application of imaging features [22]. In recent years, radiomics, as a non-invasive method, has been widely used for preoperative prediction of ccRCC nuclear grading [19, 20, 23–25]. Previous studies have used machine learning to develop radiomic models integrating traditional radiological features and compared the performance of different machine learning algorithms for predicting ccRCC patients' WHO/ISUP nuclear grading. However, these studies demonstrated variable performance with significant limitations. Yi et al. [19] achieved AUC values of 0.924 and 0.910 for training and validation cohorts using a combination model integrating traditional radiological and radiomic texture features, yet this approach was limited to single-center data without external validation. Ma et al. [23] developed an exploratory CT radiomic model based on both tumor and peritumoral regions, achieving AUC values of 0.812 and 0.804 in training and validation sets, respectively, substantially lower than our external validation performance of 0.911. Similarly, Yang et al. [24] reported AUC values of 0.964 and 0.933 using a

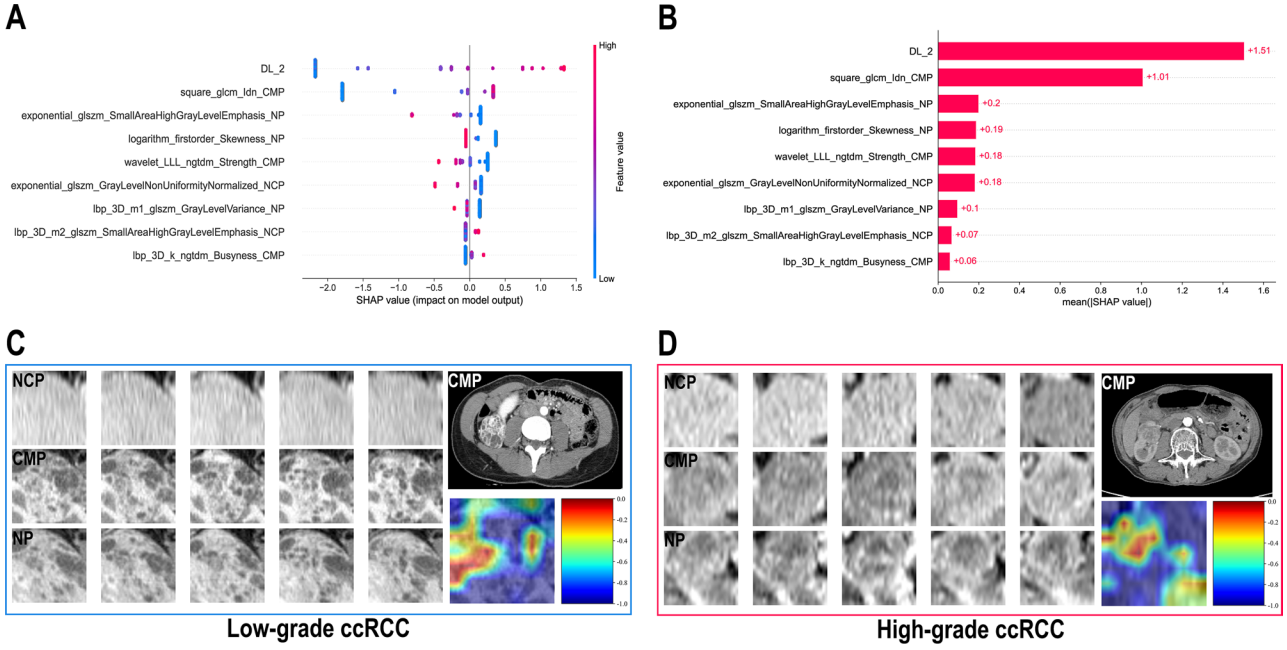

Fig. 5 Model interpretability analysis and clinical cases. **A**) SHAP summary plot showing feature impact on predictions. **B**) SHAP feature importance ranking with DL_2 (mean SHAP = +1.51) and square_glcm_idn_cmp (+1.01) as top features. **C, D**) Representative low-grade and high-grade ccRCC cases with multiphase CT and Grad-CAM heatmaps. Low-grade tumors showed homogeneous patterns while high-grade displayed heterogeneous enhancement. SHAP, Shapley Additive Explanations; Grad-CAM, Gradient-weighted class Activation Mapping; ccRCC, clear cell renal cell carcinoma

multimodal MRI-based nomogram, but the small sample size and lack of external validation limited generalizability. Zheng et al. [20] developed and validated a nomogram based on arterial-phase CT and clinical features for preoperative prediction of ccRCC nuclear grading, achieving AUC values of 0.929 and 0.876 in the training and test groups, respectively. Another study showed that the support vector machine (SVM)-ReliefF algorithm model based on nephrographic-phase CT images had high diagnostic efficacy, with AUC values of 0.887, 0.859, and 0.828 in the training, validation, and test sets, respectively [25]. Critically, these studies predominantly utilized single-phase imaging, failing to harness the comprehensive information available in multiphase CT acquisitions.

The evolution toward deep learning methodologies in renal cancer imaging represents a paradigm shift in diagnostic capabilities. Deep learning algorithms have demonstrated exceptional accuracy in recognizing complex medical image features, occasionally surpassing clinical expert performance [15, 26]. These algorithms have also been shown to assist in developing more personalized treatment plans and predicting treatment outcomes, providing significant support in renal cancer clinical management [17, 27, 28]. In this study, we proposed and implemented the use of multiphase CT data (NCP, CMP, and NP) for multilayer image fusion into a multi-channel deep learning model for ccRCC nuclear grading prediction. In previous studies, Xu et al. [29] developed a deep learning model based on single-phase images, with an AUC value of 0.882 for the integrated model. However, they did not fully consider the possible contribution of multiphase CT imaging information to predictive performance. Another study based on MRI developed a deep learning model that showed substantial performance in distinguishing low-grade from high-grade ccRCC, but the study only used two conventional MRI sequences [30]. In contrast, our multichannel fusion strategy fully exploits the complementary information across different contrast phases. This approach more accurately reflects clinical practice, where radiologists routinely evaluate multiple CT phases to assess tumor characteristics and make diagnostic decisions.

The synergistic integration of deep learning and radiomics represents a key innovation of this study. Wang et al. [31] previously developed a multimodal approach combining radiomics, deep learning, and transcriptomics, achieving AUC values of 0.946 and 0.864 in training and test cohorts. However, the requirement for transcriptomic data severely limits clinical applicability, as such molecular information is rarely available preoperatively. Our DLRN model addresses this limitation by achieving comparable or superior performance (AUC = 0.911 in external validation) using only imaging and clinical data readily available in routine practice.

In this study, feature importance analysis through SHAP revealed the hierarchical contribution of different data modalities to model performance [32]. Deep learning features dominated the predictive framework,

with DL_2 demonstrating the highest impact (mean SHAP value = 1.51), followed by radiomic features such as square_glcm_Idn_CMP (mean SHAP value = 1.01). This finding suggests that deep learning effectively captures high-level tumor characteristics that are particularly discriminative for WHO/ISUP grading, while radiomic features provide complementary quantitative texture information. The integration of these modalities through our fusion algorithm created a synergistic effect that significantly outperformed individual approaches. Grad-CAM visualization provided crucial insights into model decision-making processes, demonstrating clinical relevance and interpretability [33]. The model consistently focused on tumor heterogeneity patterns across all CT phases, with distinct attention patterns for different grades. In high-grade ccRCC cases, the model emphasized regions of irregular enhancement and necrosis, consistent with the aggressive biological behavior of these tumors. Conversely, low-grade cases showed more uniform attention patterns, reflecting the homogeneous nature of these lesions. This spatial attention alignment with established pathological characteristics validates the model's clinical logic and enhances physician trust in AI-assisted diagnosis.

From a clinical perspective, the DLRN model demonstrates substantial utility by enabling a non-invasive, preoperative prediction of nuclear grade. Accurate grading is pivotal for surgical decision-making, particularly for cT1 renal masses. While partial nephrectomy remains the standard of care for small masses, the preoperative identification of high-grade tumors could necessitate a shift towards radical nephrectomy or wider resection margins to mitigate recurrence risks. Furthermore, for elderly or comorbid patients eligible for active surveillance or thermal ablation, excluding high-grade pathology is a prerequisite for safety. By providing a quantified risk assessment, our nomogram functions as a vital decision-support tool to personalize treatment strategies.

To further delineate the model's reliability and facilitate clinical translation, we scrutinized the misclassified cases to understand potential limitations. False predictions were predominantly observed in two scenarios: (1) Small tumors (<2 cm), where the paucity of voxels hindered the extraction of robust texture patterns; and (2) Tumors with extensive necrosis. Necrotic regions often mimic cystic components or exhibit heterogeneous enhancement that can confound the deep learning attention mechanism, leading to grade overestimation. Recognizing these "diagnostic blind spots" is crucial for clinicians when interpreting the model's output in complex cases.

To the best of our knowledge, this is the first study to combine multi-channel deep learning with traditional radiomics techniques for predicting the WHO/ISUP nuclear grading of ccRCC. However, several limitations remained. Firstly, this study adopted a retrospective design, and despite including a large sample from two medical centers to enhance generalizability, patient selection bias and potential information collection biases are inevitable. Future prospective, multi-center clinical trials are needed to further validate the findings. Secondly, ccRCC is a subtype of malignant renal tumor. Despite its high occurrence, other renal cancer subtypes could have similar radiological features, and therefore, should be evaluated in future studies. Thirdly, variations in CT scanning protocols remain an inherent challenge in multi-center radiomics studies. Specifically, differences in contrast agent injection rates and scan delay times between the two institutions may influence pixel intensity values, representing a potential source of variation that future standardization protocols must address. Additionally, while our model showed favorable performance in this two-center validation, the clinical impact on patient outcomes and treatment decisions requires prospective validation studies with broader patient populations. Beyond these limitations, it is essential to highlight the enormous potential of artificial intelligence (AI) technology in optimizing clinical decision-making. The predictive model proposed in this study integrates advanced AI and radiomics methodologies, offering clinicians a refined tool for more accurate prediction of ccRCC nuclear grading. This enhanced predictive capability is critical in guiding patient management and treatment decisions. By improving the precision of tumor pathological grading predictions, the model aids clinicians in selecting the most appropriate surgical and therapeutic interventions, thereby contributing to improved clinical outcomes for patients. Nonetheless, several pivotal challenges remain unresolved, including the non-adherence to established machine learning best practices, the absence of standardized imaging analysis protocols, and the existing "clinical translation gap" that hinders the seamless transition of models from research to real-world clinical applications. Addressing these challenges is crucial for maximizing the model's utility and impact in clinical settings.

In conclusion, the DLRN model demonstrates excellent performance and generalization ability in preoperatively predicting ccRCC nuclear grading. This innovative fusion strategy overcomes the limitations of traditional single-method approaches, significantly improving the ability to capture and analyze spatial heterogeneity information within tumors. Through further validation and improvement in prospective clinical studies, this method is expected to become an important tool for non-invasive preoperative assessment of ccRCC, providing more accurate and efficient technical support for clinical personalized treatment decisions and ultimately improving patient prognosis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-026-02268-5.

Supplementary material 1

Acknowledgements
We thank the Information, Imaging and Pathology departments of our hospital for supporting this study. Some of our experiments were carried out on the OnekeyAI platform. We thank OnekeyAI and its developers for supporting this research.

Author contributions
C.Y., W.L., and X.Y. conceived and designed the study. C.Y., Z.Z., B.W., H.Z., T.M., and Y.L. collected and curated the data. C.Y. developed the methodology and deep learning models. C.Y. wrote the original draft of the manuscript. W.L. and X.Y. supervised the project and reviewed and edited the manuscript. X.Y. obtained funding and provided resources. W.L. administered the project. All authors reviewed the manuscript.

Funding
This study was supported by the Clinical Trial Fund of the 358 Program of Tianjin Medical University Cancer Institute and Hospital (358-2023-02) and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-009A).

Data availability
The datasets generated and analyzed during the current study are available from the corresponding author Xin Yao upon reasonable request. The data are not publicly available due to privacy and ethical restrictions related to patient confidentiality and institutional policies.

Declarations

Ethical approval
The study was conducted in strict accordance with the Declaration of Helsinki and received approval from the Ethical Review Committee of the Tianjin Medical University Cancer Institute & Hospital (No: bc20251107). Due to the retrospective and deidentified data utilized in the study, the informed consent was waived.

Consent for publication
Not applicable. This study used retrospective, de-identified data and no individual patient details, images, or videos are presented that would compromise anonymity.

Competing interest
The authors declare no competing interests.

Author details
[1]Department of Urologic Oncology, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Huanhuxi Road, Hexi Distinct, Tianjin 300060, China
[2]Department of Urology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, P. R. China
[3]State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-Sen University Cancer Center, Guangzhou 510060, P. R. China

Received: 7 August 2025 / Accepted: 3 March 2026
Published online: 09 March 2026

References

1. Bukavina L, Bensalah K, Bray F, et al. Epidemiology of renal cell carcinoma: 2022 update. Eur Urology. 2022;82(5):529–42. https://doi.org/10.1016/j.eururo.2022.08.019.
2. Padala SA, Barsouk A, Thandra KC, et al. Epidemiology of renal cell carcinoma. World J Oncol. 2020;11(3):79–87. https://doi.org/10.14740/wjon1279.
3. Abu-Ghanem Y, Powles T, Capitanio U, et al. The impact of histological subtype on the incidence, timing, and patterns of recurrence in patients with renal cell carcinoma after surgery-results from RECUR consortium. Eur Urology Oncol. 2021;4(3):473–82. https://doi.org/10.1016/j.euo.2020.09.005.
4. Mattila KE, Vainio P, Jaakkola PM. Prognostic factors for localized clear cell renal cell carcinoma and their application in adjuvant therapy. Cancers. 2022;14(1). https://doi.org/10.3390/cancers14010239.
5. Delahunt B, Eble JN, Egevad L, et al. Grading of renal cell carcinoma. Histopathology. 2019;74(1):4–17. https://doi.org/10.1111/his.13735.
6. Paner GP, Stadler WM, Hansel DE, et al. Updates in the eighth edition of the tumor-node-metastasis staging classification for Urologic Cancers. Eur Urology. 2018;73(4):560–69. https://doi.org/10.1016/j.eururo.2017.12.018.
7. Escudier B, Porta C, Schmidinger M, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up [J]. Ann Oncol: Off J The Eur Soc Med Oncol. 2014;25Suppl(3):iii49–56. https://doi.org/10.1093/annonc/mdu259.
8. Xu H, Xing Z, Ai K, 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. https://doi.org/10.1186/s12957-024-03302-y.
9. Macklin PS, Sullivan ME, Tapping CR, et al. Tumour seeding in the tract of percutaneous renal Tumour biopsy: a report on seven cases from a UK tertiary referral centre. Eur Urology. 2019;75(5):861–67. https://doi.org/10.1016/j.eururo.2018.12.011.
10. Andrulli S, Rossini M, Gigliotti G, et al. The risks associated with percutaneous native kidney biopsies: a prospective study. Nephrol Dial Transplant. 2023;38(3):655–63. https://doi.org/10.1093/ndt/gfac177.
11. Silverman SG, Gan YU, Mortele KJ, et al. Renal masses in the adult patient: the role of percutaneous biopsy. Radiology. 2006;240(1):6–22. https://doi.org/10.1148/radiol.2401050061.
12. Patel HD, Johnson MH, Pierorazio PM, et al. Diagnostic accuracy and risks of biopsy in the diagnosis of a renal mass suspicious for localized renal cell carcinoma: systematic review of the literature. J Urol. 2016;195(5):1340–47. https://doi.org/10.1016/j.juro.2015.11.029.
13. Shen J, Zou Y. Diagnostic value of contrast-enhanced CT in clear cell renal cell carcinoma: a systematic review and meta-analysis. BMC Urol. 2024;24(1):189. https://doi.org/10.1186/s12894-024-01574-w.
14. Suarez-Ibarrola R, Hein S, Reis G, et al. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. World J Urology. 2020;38(10):2329–47. https://doi.org/10.1007/s00345-019-03000-5.
15. Dai C, Xiong Y, Zhu P, et al. Deep learning assessment of Small renal masses at contrast-enhanced multiphase CT. Radiology. 2024;311(2):e232178. https://doi.org/10.1148/radiol.232178.
16. Nie P, Liu S, Zhou R, et al. A preoperative CT-based deep learning radiomics model in predicting the stage, size, grade and necrosis score and outcome in localized clear cell renal cell carcinoma: a multicenter study. Eur J Radiol. 2023;166:111018. https://doi.org/10.1016/j.ejrad.2023.111018.
17. Gui CP, Chen YH, Zhao HW, et al. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. The Lancet Digit Health. 2023;5(8):e515–24. doi: https://doi.org/10.1016/s2589-7500(23)00095-x.
18. Shu J, Wen D, Xi Y, et al. Clear cell renal cell carcinoma: machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade. Eur J Radiol. 2019;121:108738. https://doi.org/10.1016/j.ejrad.2019.108738.
19. Yi X, Xiao Q, Zeng F, et al. Computed tomography radiomics for predicting pathological grade of renal cell carcinoma. Front Oncol. 2020;10:570396. https://doi.org/10.3389/fonc.2020.570396.
20. Zheng Z, Chen Z, Xie Y, et al. 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. https://doi.org/10.1007/s00330-020-07667-y.
21. Isensee F, Jaeger PF, Kohl SA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11. https://doi.org/10.1038/s41592-020-01008-z.
22. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77. https://doi.org/10.1148/radiol.2015151169.

23. Ma Y, Guan Z, Liang H, et al. Predicting the WHO/ISUP grade of clear cell renal cell carcinoma through CT-Based tumoral and peritumoral radiomics. Front Oncol. 2022;12:831112. https://doi.org/10.3389/fonc.2022.831112.
24. Yang Y, Zhang Z, Zhang H, et al. Machine learning-based multiparametric MRI radiomics nomogram for predicting WHO/ISUP nuclear grading of clear cell renal cell carcinoma. Front Oncol. 2024;14:1467775. https://doi.org/10.3389/fonc.2024.1467775.
25. Xv Y, Lv F, Guo H, et al. Machine learning-based CT radiomics approach for predicting WHO/ISUP nuclear grade of clear cell renal cell carcinoma: an exploratory and comparative study. Insights Imag. 2021;12(1):170. https://doi.org/10.1186/s13244-021-01107-1.
26. Xiong Y, Yao L, Lin J, et al. Artificial intelligence links CT images to pathologic features and survival outcomes of renal masses. Nat Commun. 2025;16(1):1425. https://doi.org/10.1038/s41467-025-56784-z.
27. Chen S, Gao F, Guo T, et al. Deep learning-based multi-model prediction for disease-free survival status of patients with clear cell renal cell carcinoma after surgery: a multicenter cohort study. Int J Surg (Lond, Engl). 2024;110(5):2970–77. https://doi.org/10.1097/js9.0000000000001222.
28. Toda N, Hashimoto M, Arita Y, et al. Deep learning algorithm for fully automated detection of Small (≤4 cm) renal cell carcinoma in contrast-enhanced computed tomography using a multicenter database. Invest Radiol. 2022;57(5):327–33. https://doi.org/10.1097/rli.0000000000000842.
29. Xu L, Yang C, Zhang F, et al. Deep learning using CT images to grade clear cell renal cell carcinoma: development and validation of a prediction model. Cancers. 2022;14(11). https://doi.org/10.3390/cancers14112574.
30. Zhao Y, Chang M, Wang R, et al. Deep learning based on MRI for differentiation of low- and high-grade in low-stage renal cell carcinoma. J Magn Reson Imag. 2020;52(5):1542–49. https://doi.org/10.1002/jmri.27153.
31. Wang S, Zhu C, Jin Y, et al. A multi-model based on radiogenomics and deep learning techniques associated with histological grade and survival in clear cell renal cell carcinoma. Insights Imag. 2023;14(1):207. https://doi.org/10.1186/s13244-023-01557-9.
32. Lundberg SM, S-I L. A unified approach to interpreting model predictions [Z]. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc; 2017: 4768–77.
33. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual Explanations from deep networks via Gradient-based localization. Int J Comput Vision. 2020;128(2):336–59. https://doi.org/10.1007/s11263-019-01228-7.

Publisher's Note