scientific reports

Check for updates

OPEN

Multimodal deep learning framework integrating multiphase CT and histopathological whole slide imaging for predicting recurrence in ccRCC

Changyi Ma[1,6], Bao Feng[2,3,6], Yan Lei[1], Zhaole Yu[2], Yu Liu[2], Jin Cui[1], Rong-gang Li[4], Xiufang Huang[4], Baiming Wu[1,5], Zhixin Luo[1,5] & Enming Cui[1,3,5]✉

ccRCC is an aggressive, heterogeneous tumor with a poor prognosis. Prognostic assessments need multi-modal data. Radiological images have limits, while pathological images offer micro-level details. Integrating these for ccRCC outcome prediction is important. Our study aimed to develop and validate a DL fusion model using multiphase CT images and WSI for postoperative risk stratification in ccRCC patients. This retrospective study included 274 ccRCC patients who underwent multiphase CT scans (Jan 2008-Mar 2021), with diagnoses confirmed by histopathology post-surgery. The patient cohort was divided into a training cohort of 164 patients for model development and a test cohort of 110 patients for model validation. The primary outcome was local recurrence or metastasis versus non-recurrence (NR) with a minimum follow-up of 3 years. DL models based on multiphase CT images and histopathological WSIs were developed and validated. Performance comparisons among models were made through accuracy (ACC) and receiver operating characteristic (ROC) curve analyses, with integrated discrimination improvement (IDI) analysis and the DeLong test assessing diagnostic performance. Decision curve analysis (DCA) evaluated clinical utility, and Kaplan-Meier analysis assessed variable-survival correlations. The CT and Pathology Mutual Guidance Fusion Diagnostic Network (CPNet) exhibited superior performance in predicting postoperative disease-free survival (DFS) in ccRCC patients. Among the models, the PCP-Pathology Fuse model achieved the highest AUC of 0.8363 and accuracy of 75.45%, outperforming the CMP-Pathology Fuse (AUC 0.7965, ACC 69.09%) and NP-Pathology Fuse (AUC 0.798, ACC 69.09%) models. Its performance was comparable to the Three-phase-Pathology Fuse model (AUC 0.8341, ACC 70.00%, $P > 0.05$). IDI and DCA confirmed significant net benefits (0.01–0.95) for the PCP-based model. The PCP-based CPNet model shows promise for predicting postoperative DFS in ccRCC patients, with performance comparable to three-phase CT-pathology models. It may serve as a potential bioimaging prognostic marker, pending external validation to support clinical integration.

Keywords Renal cell carcinoma, Prognosis, X-ray computed tomography, Pathology, Artificial intelligence

Abbreviations

PCP	Precontrast phase
CMP	Corticomedullary phase
NP	Nephrogenic phase
DL	Deep learning

[1]Department of Radiology, Jiangmen Central Hospital, 23 Beijie Haibang Street, Jiangmen, People's Republic of China. [2]School of Electronic Information and Automation, Guilin University of Aerospace Technology, 2 Jinji Road, Guilin, People's Republic of China. [3]Jiangmen Key Laboratory of Artificial Intelligence in Medical Image Computation and Application, 23 Beijie Haibang Street, Jiangmen, People's Republic of China. [4]Department of Pathology, Jiangmen Central Hospital, 23 Beijie Haibang Street, Jiangmen, People's Republic of China. [5]Guangdong Medical University, 2 Wenming East Road, Zhanjiang, People's Republic of China. [6]These authors contributed equally: Changyi Ma and Bao Feng. ✉email: cem2008@163.com

WSI	Whole slide image
ccRCC	Clear cell renal cell carcinoma
ACC	Accuracy
PPV	Positive predictive value
NPV	Negative predictive value
ROC	Receiver operating characteristic
AUC	The area under the ROC curve
IDI	Integrated discrimination improvement
DCA	Decision curve analysis
CPNet	CT and Pathology mutual guidance fusion diagnostic Network
DFS	Disease-free survival
AUC	Area under the ROC curve
CNN	Convolutional neural network
TME	Tumor microenvironment
CAM	Class activation mapping

Clear cell renal cell carcinoma (ccRCC) is a highly heterogenous tumor known for its high aggressiveness and propensity for metastasis, both of which make it predisposed toward an unfavorable prognosis[1–4] and the potential need for prompt aggressive treatment and close follow-up. The local postoperative recurrence rate in patients with ccRCC is 20–30%, with most recurrences occurring within 3 years of the operation[5]. In the management of patients with ccRCC, prognostic assessments typically require a comprehensive analysis of medical data from multiple modalities[6]. Radiological images can provide macrolevel information such as lesion location, morphological texture, and the presence of metastases, which are useful for predicting patient outcomes[7]. However, owing to its relatively low sensitivity, radiology is not considered the gold standard for diagnosing the disease[8]. On the other hand, pathological images can offer microlevel insights into tumor cells and the tumor microenvironment (TME)[9], revealing potential microscopic characteristics, biological behaviors and pathological phenotypes of the tumor[10]. Thus, developing a method that integrates multimodal information for outcome prediction in ccRCC patients is highly important.

In recent years, deep learning (DL) methods have garnered widespread attention in the medical research field because of their strong feature learning capabilities, allowing them to extract deep, task-relevant features from medical data[11–14]. Previous studies have made progress in the use of DL methods for predicting the prognosis of ccRCC[15–17], but there are some limitations. For example, the model developed by Wessels et al.[18] only uses features from a single modality for predicting disease prognosis, preventing a comprehensive analysis of the state of the tumor from multiple perspectives. Schulz et al.'s[19] multimodal fusion model, meanwhile, does not fully consider the potential correlations and interactions among the modalities, which may lead to the underuse of some task-relevant features, thereby affecting the ability of the model to handle complex data and reducing its predictive performance.

To address these issues, this study proposes a deep learning-based CT and Pathology Mutual Guidance Fusion Diagnostic Network (CPNet). The model uses deep learning methods to extract deep features from lesion areas on both CT and pathology images, which are then used to construct a multimodal fusion module that establishes interactions between and aligns cross-modal information. This study aimed to investigate the effectiveness of the CPNet model in predicting the outcomes of patients with ccRCC.

Materials and methods

Patients

The institutional pathology database was searched to identify patients with pathologically confirmed ccRCC following surgical resection in our hospital between January 2008 and March 2021. The pathological diagnosis was confirmed by a senior genitourinary pathologist. This retrospective study was approved by the institutional review board, and the requirement for informed consent was waived.

The inclusion criteria were as follows: (a) the presence of preoperative CT scans with three-phase imaging and (b) pathological diagnosis of ccRCC without multiple or other types of malignant tumors.

The exclusion criteria were as follows: (a) preoperative systemic therapy or chemoradiotherapy; (b) images with significant noise or artifacts; (c) significant lesion rupture with abundant hemorrhage leading to obscuration of tumor features; (d) lesions completely composed of cystic components; and (e) contaminated or missing whole-slide images (WSIs). The patient inclusion and exclusion flowchart are presented in Fig. 1. The workflow of the study design is shown in Fig. 2.

The primary endpoints were local recurrence or metastasis and nonrecurrence (NR), with a minimum follow-up time of 3 years for all patients. Detailed follow-up strategies are provided in Supplementary Material S1. Imaging assessments were jointly performed by urologists and radiologists. The dataset was randomly divided into a training cohort and a test cohort at a 6:4 ratio, ensuring that the proportion of negative and positive samples is similar in both cohorts, and the basic patient information is listed in Table 1.

WSI acquisition and preprocessing

A senior genitourinary pathologist evaluates all HE-stained slides for each patient and selects the most representative slide with key criteria (high-quality sections with clear structures/cell details, ISUP grade by highest nuclear grade, and priority to minimal artifacts) for digital scanning using a Precpoint M8 Microscope and Slide Scanner (Precpoint, Freising, Germany) with a 40X objective to generate WSIs for analysis.

The preprocessing steps of WSI images include tissue segmentation and image patches extraction. For detailed steps, please refer to Supplementary Material S2.

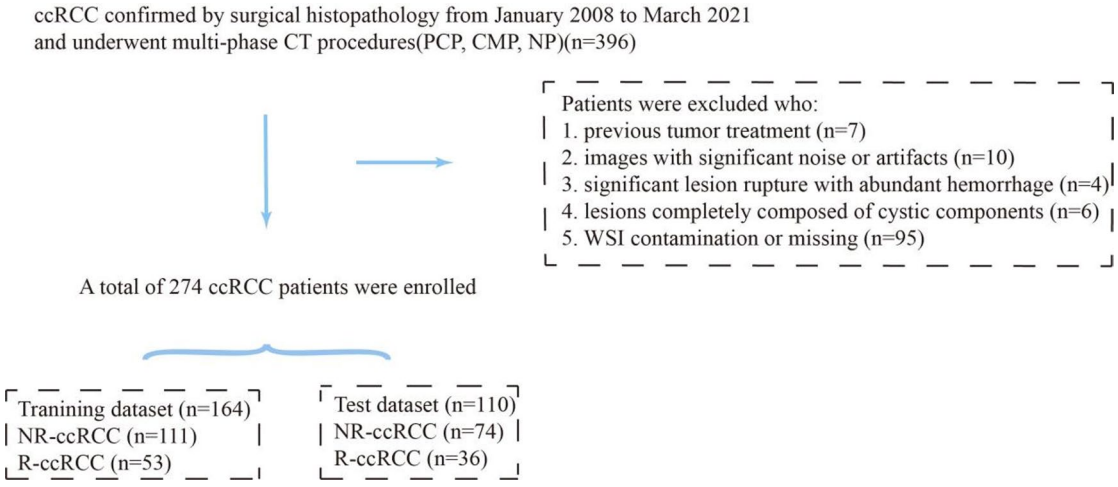

Fig. 1. The patient inclusion and exclusion flowchart. *PCP* precontrast phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *R* recurrence, *NR* nonrecurrence.

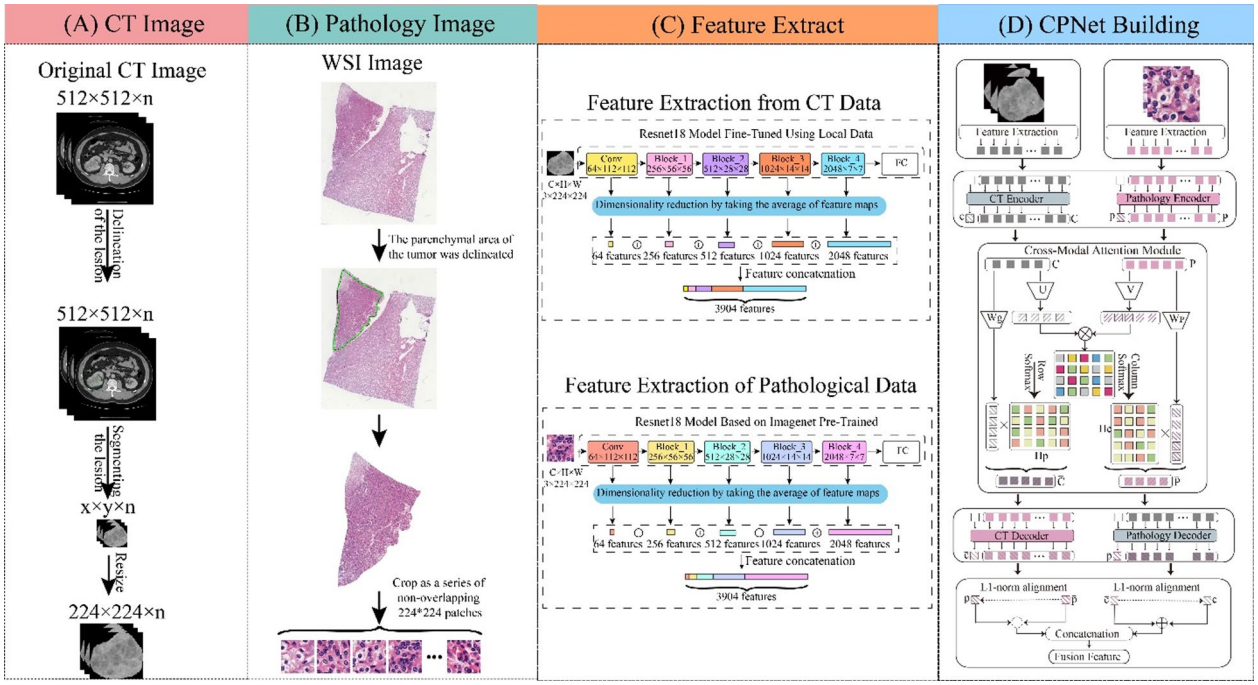

Fig. 2. Workflow of the study design. *WSI* whole slide image, *CPNet* CT and pathology mutual guidance fusion diagnostic Network.

CT examination and imaging preprocessing

CT examinations were performed with Emotion 16-row, Somatom Force 192-row (Siemens Healthineers), Aquilion One 64-row (Canon Medical Systems), all using the same examination protocol, prior to surgical treatment.

After retrieving and acquiring the images of all patients from our institutional picture archiving and communication system (PACS), we loaded the images into ITK-SNAP[20] (version 3.6.0), anonymized them and stored them in NIfTI format. The tumor images were then spatially matched and segmented by three radiologists with 6, 8, and 11 years of experience in abdominal radiology, and the effective margins of the ROIs were confirmed by two other senior radiologists with 13 and 15 years of abdominal radiologic experience. Details of CT examinations and image preprocessing are provided in Supplementary Material S3.

		Training cohort ($n=164$)				Test cohort ($n=110$)				
		Overall	NR-ccRCC ($n=111$)	R-ccRCC ($n=53$)	*P*-value	Overall	NR-ccRCC ($n=74$)	R-ccRCC ($n=36$)	*P*-value	*P*-value (overall)
Age*		56.06 ± 11.29	56.44 ± 11.59	55.26 ± 10.71	0.53	53.98 ± 13.05	53.88 ± 13.2	54.19 ± 12.91	0.91	0.17
Sex#	Male	96	68(61.3)	28(52.8)	0.31	68	44(59.5)	24(66.7)	0.47	0.59
	Female	68	43(38.7)	25(47.2)		42	30(40.5)	12(33.3)		
Position#	Left	76	51(45.9)	25(47.2)	0.88	58	40(54.1)	18(50.0)	0.69	0.30
	Right	88	60(54.1)	28(52.8)		52	34(45.9)	18(50.0)		
Size(cm)*		5.54 ± 2.70	4.67 ± 2.24	7.35 ± 2.70	< 0.05	5.53 ± 2.90	4.83 ± 2.9	6.97 ± 2.36	< 0.05	0.97
T State#	1a	78	7	71	< 0.05	59	51	8	< 0.05	0.083
	1b	61	33	28		25	10	15		
	2a	9	2	7		7	2	5		
	2b	2	1	1		6	4	2		
	3a	12	3	9		11	7	4		
	4	2	1	1		2	0	2		
N State#	0	111	52	163	0.32	105	73	32	0.069	0.078
	1	0	1	1		5	1	4		
M State#	0	160	109	51	0.46	103	71	32	0.31	0.19
	1	4	2	2		7	3	4		
Clinical Stages#	Ⅰ	137	34	103	< 0.05	80	59	21	0.12	0.166
	Ⅱ	10	3	7		12	6	6		
	Ⅲ	12	2	10		11	6	5		
	Ⅳ	5	3	2		7	3	4		

Table 1. Baseline characteristics of patients. Note: R = recurrence, NR = nonrecurrence; Unless otherwise specified, data in parentheses are percentages. *Data are means ± standard deviation and statistic by independent sample T test. #Data statistic by Pearson's chi-squared test. $P < 0.05$ was considered statistically significant.

Development of the CT and pathology mutual guidance fusion diagnostic network (CPNet)

The construction of CPNet can be divided into two main parts: deep learning feature extraction and multimodal fusion framework construction.

DL feature extraction

To address potential issues of vanishing and exploding gradients caused by the use of deep model layers, the ResNet18[21] network was employed as the feature extraction model for both the CT and pathology images. Additionally, to prevent overfitting, transfer learning was applied by initializing the ResNet18 model, pretrained on ImageNet data with rich prior knowledge, for both the CT and pathology data. The initialized model was then fine-tuned with local CT data to adapt the model to the target task and achieve better generalization. Given the large amount of pathology data, fine-tuning of the initialized model still required significant computational power. Since the initialized model already possessed rich prior knowledge, the initialized model was directly used for extracting deep features from the pathology data. Detailed information for the feature extraction process of the CT and pathology data is provided in Supplementary Material S4.

Construction of the multimodal fusion framework

Pathology is the "gold standard" for diagnosing tumors, as it provides direct evidence of tumor cell morphological characteristics[22]. In contrast, radiology-based diagnosis enables rapid, noninvasive evaluation of overall tumor status, offering crucial information for clinical decision-making[23], but there is significant heterogeneity between these two modalities, and direct fusion may overlook potential correlations between them. Therefore, a multimodal fusion framework is proposed to explore the intrinsic connections between the modalities. Specifically, during the modality interaction process, CT features, while attending to information within their own modality, are guided by pathology features to compute their associations with the microscopic tumor microenvironment in pathology images, thereby highlighting macro-structural regions in CT that are relevant to pathology. Conversely, pathology features, while focusing on their own modality-specific information, are guided by CT features to emphasize microscopic tissue regions that are associated with radiological characteristics. In this way, each modality selectively extracts the task-relevant components from the other, aligns them, and integrates them to form a more complementary joint representation, which can be leveraged for more accurate prediction of patient prognosis. The multimodal fusion framework is illustrated in Fig. 3. Detailed information for the multimodal fusion framework is provided in Supplementary Material S5.

Visualization analysis

To demonstrate the interpretability of the model, we selected the patches and slices with the highest correlation scores from CPNet when guided by the pathology and CT modalities, respectively. This was done to explore

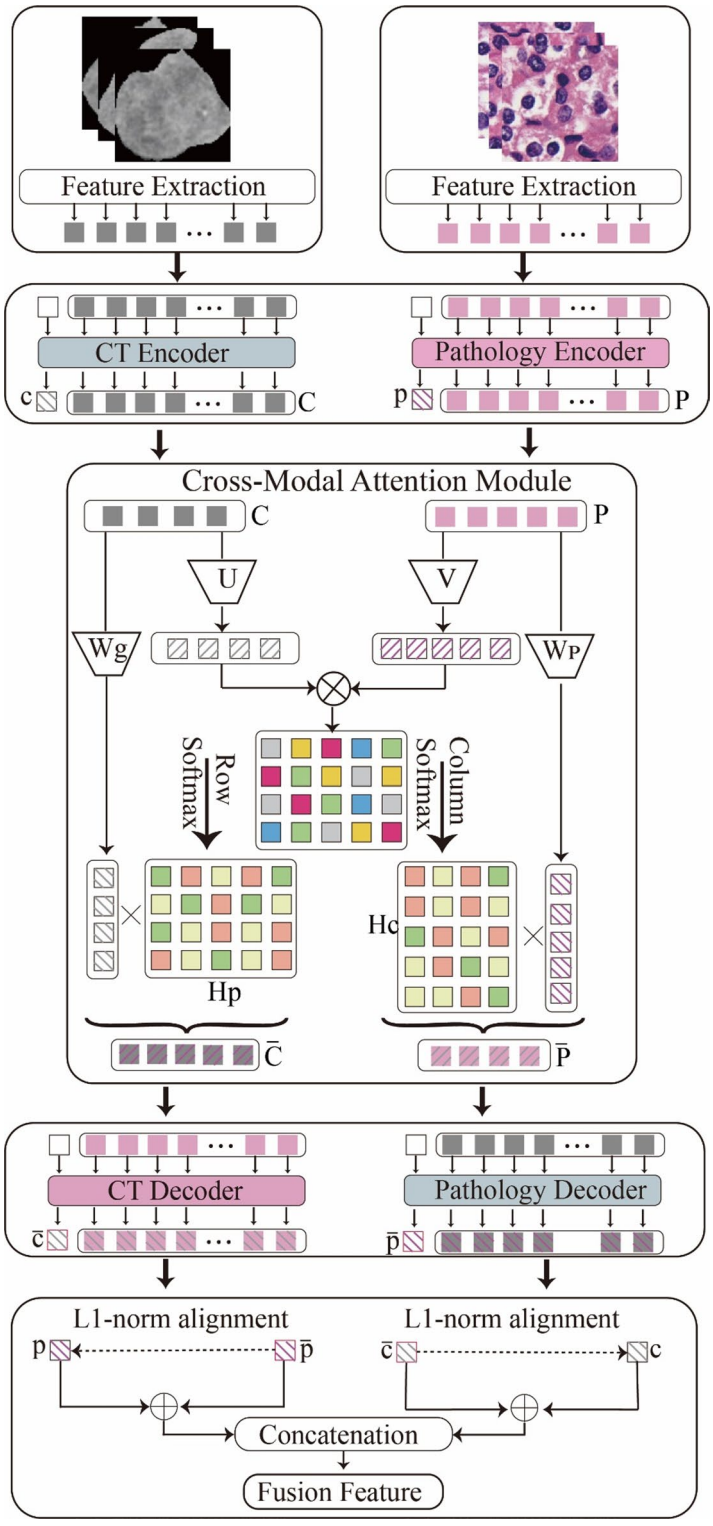

Fig. 3. Multi-modal fusion framework.

whether the model based on the guided modality could focus on important yet easily overlooked regions under the guidance of the other modality. Additionally, we visualized the corresponding patches and slices with class activation maps (CAMs)[24] to investigate the ROIs in the feature extraction model.

Statistical analysis

In this study, model performance was comprehensively evaluated with metrics including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver

operating characteristic curve (AUC). The model's performance improvements were assessed using the Integrated Discrimination Improvement (IDI) analysis and Delong's test. Additionally, decision curve analysis (DCA) was employed to evaluate the clinical benefits of the model in predicting the outcomes of patients with ccRCC. We used the Kaplan–Meier method to analyze the correlations between variables and survival.

Statistical tests were conducted in the R language environment (version 3.4.0, http://www.Rproject.org).

Deep learning experiments were performed on an NVIDIA RTX A6000 GPU with 48 GB of memory, running CUDA version 11.5. The deep learning framework used was PyTorch 2.1.0 GPU version, running on Python 3.10.13.

Results

Characteristics of the study population

A total of 274 ccRCC patients were ultimately included in this study; 164 patients (111 patients in the nonrecurrence group and 53 patients in the recurrence group) composed the training cohort, and 110 patients (74 patients in the nonrecurrence group and 36 patients in the recurrence group) composed the test cohort. There was no overlap among the patients in the cohorts.

There were no statistically significant differences in classification performance across different devices. Details of Stratified analysis by imaging equipment are provided in Supplementary Material S6.

There was a significant difference in tumor size among the training set and the test set. There was no statistically significant difference in sex, position, age and TNM staging among the datasets. The characteristics of the study population are shown in Table 1.

Development and performance comparison of the DL-based CPNet models

We developed models for predicting the outcomes of ccRCC patients with single modality images (CMP, NP, PCP, Three-phase, and Pathology), simple multimodal fusion images (CMP-Pathology Concat, NP-Pathology Concat, PCP-Pathology Concat, Three-phase-Pathology Concat), and multimodal interaction fusion images (CMP-Pathology Fuse, NP-Pathology Fuse, PCP-Pathology Fuse, and Three-phase-Pathology Fuse). Here, among the single modalities, "Three-phase" refers to the concatenation of the CMP, NP, and PCP images but is still considered a single CT modality. Simple multimodal fusion refers to the concatenation of the individual and combined CT phase images and the pathology images, whereas multimodal interaction fusion uses CPNet to integrate the images from the two modalities. The diagnostic performance of these models in the test cohort and the training cohort are shown in Table 2, and Table S2 in the Supplementary material, respectively.

ROC curve analysis demonstrated that the DL-based CPNet models exhibited superior predictive performance for postoperative DFS in ccRCC patients than the non-CPNet–based models. Specifically, the AUCs for the CMP-Pathology Fuse, NP-Pathology Fuse, PCP-Pathology Fuse, and Three-phase-Pathology Fuse models were 0.7965, 0.798, 0.8363, and 0.8341, respectively, and the corresponding ACCs were 69.09%, 69.09%, 75.45%, and 70.00%, respectively.

Compared with those of the WSI-based single-modality model, the AUC of the CPNet models was greater by 15.85% on average, and the misdiagnosis rate was lower by 39.59% on average. Furthermore, compared with those of the CT-based single-modality models, the AUC and ACC of the CPNet models were greater by 15.71% and 18.41% on average, respectively, and the misdiagnosis rate was lower by 22.97% on average. Compared with that of the simple multimodal fusion models, the AUC of the CPNet models was greater by 11.83% on average, and the misdiagnosis rate was lower by 22.22% on average.

Modal	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
CMP	0.5893	0.6944 (25/36)	0.4189 (31/74)	0.5091 (56/110)	0.3676 (25/68)	0.7381 (31/42)
NP	0.6528	0.6389 (23/36)	0.5676 (42/74)	0.5909 (65/110)	0.4182 (23/55)	0.7636 (42/55)
PCP	0.7485	0.7778 (28/36)	0.6216 (46/74)	0.6727 (74/110)	0.5000 (28/56)	0.8519 (46/54)
All_CT	0.7110	0.5278 (19/36)	0.7568 (56/74)	0.6818 (75/110)	0.5135 (19/37)	0.7671 (56/73)
Pathology	0.6577	0.4722 (17/36)	0.7162 (53/74)	0.6364 (70/110)	0.4474 (17/38)	0.7361 (53/72)
CMP and pathology concat	0.6010	0.6111 (22/36)	0.5541 (41/74)	0.5727 (63/110)	0.4000 (22/55)	0.7455 (41/55)
NP and pathology concat	0.6952	0.6389 (23/36)	0.7027 (52/74)	0.6818 (75/110)	0.5111 (23/45)	0.8000 (52/65)
PCP and pathology concat	0.7755	0.6389 (23/36)	0.7297 (54/74)	0.7000 (77/110)	0.5349 (23/43)	0.8060 (54/67)
All_CT and pathology concat	0.7736	0.6944 (25/36)	0.7297 (54/74)	0.7182 (79/110)	0.5556 (25/45)	0.8308 (54/65)
CMP and pathology fuse	0.7965	0.8611 (31/36)	0.6081 (45/74)	0.6909 (76/110)	0.5167 (31/60)	0.9000 (45/50)
NP and pathology fuse	0.7980	0.8611 (31/36)	0.6081 (45/74)	0.6909 (76/110)	0.5167 (31/60)	0.9000 (45/50)
PCP and pathology fuse	0.8363	0.8611 (31/36)	0.7027 (52/74)	0.7545 (83/110)	0.5849 (31/53)	0.9123 (52/57)
All_CT and pathology fuse	0.8341	0.8889 (32/36)	0.6081 (45/74)	0.7000 (77/110)	0.5246 (32/61)	0.9184 (45/49)

Table 2. Diagnostic performance of all models in the test set. *PCP* precontrast phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *PPV* positive predictive value, *NPV* negative predictive value, *AUC* the area under the ROC curve.

The PCP-pathology fuse model predicts disease-free survival (DFS) in CcRCC patients

ROC curve analysis also demonstrated that the PCP-Pathology Fuse model exhibited greater predictive performance for the postoperative risk of recurrence or metastasis in ccRCC patients than the other CT modality-based CPNet models, with an AUC, sensitivity, and specificity of 0.8363, 0.8611, and 0.7027, and an accuracy of 75.45%. However, its performance was not significantly different from that of the CPNet model based on the three-phase CT images (AUC, sensitivity, specificity and accuracy of 0.8341, 0.8889, and 0.6081, and 70.00%, respectively).

Specifically, as shown in Fig. 4A, the All_CT model based on PCP, CMP, and NP data did not outperform the model relying solely on PCP data. Figure 4B illustrates that incorporating pathological image information further enhances prognostic performance. Figure 4C demonstrates that the PCP-Pathology Fuse model exhibited superior predictive ability for postoperative recurrence or metastasis risk in ccRCC patients compared with other CT modality–based CPNet models, while its performance was comparable to the CPNet model constructed from three-phase CT images. Figure 4D summarizes ROC curves of all models. Furthermore, the DCA results in Fig. 4E indicate that the PCP-Pathology Fuse model achieved a substantial net benefit across a wide threshold range (0.01–0.95), closely resembling the Three-phase-Pathology Fuse model and surpassing other models.In addition, as Fig. 4F,G radar plots illustrated, the CPNet-based fusion models—including CMP and Pathology Fuse, NP and Pathology Fuse, PCP and Pathology Fuse, and All_CT and Pathology Fuse—consistently demonstrated balanced performance across multiple metrics.The ROC curves of the above models in the training cohort are provided in Figure S5 in Supplementary Material.

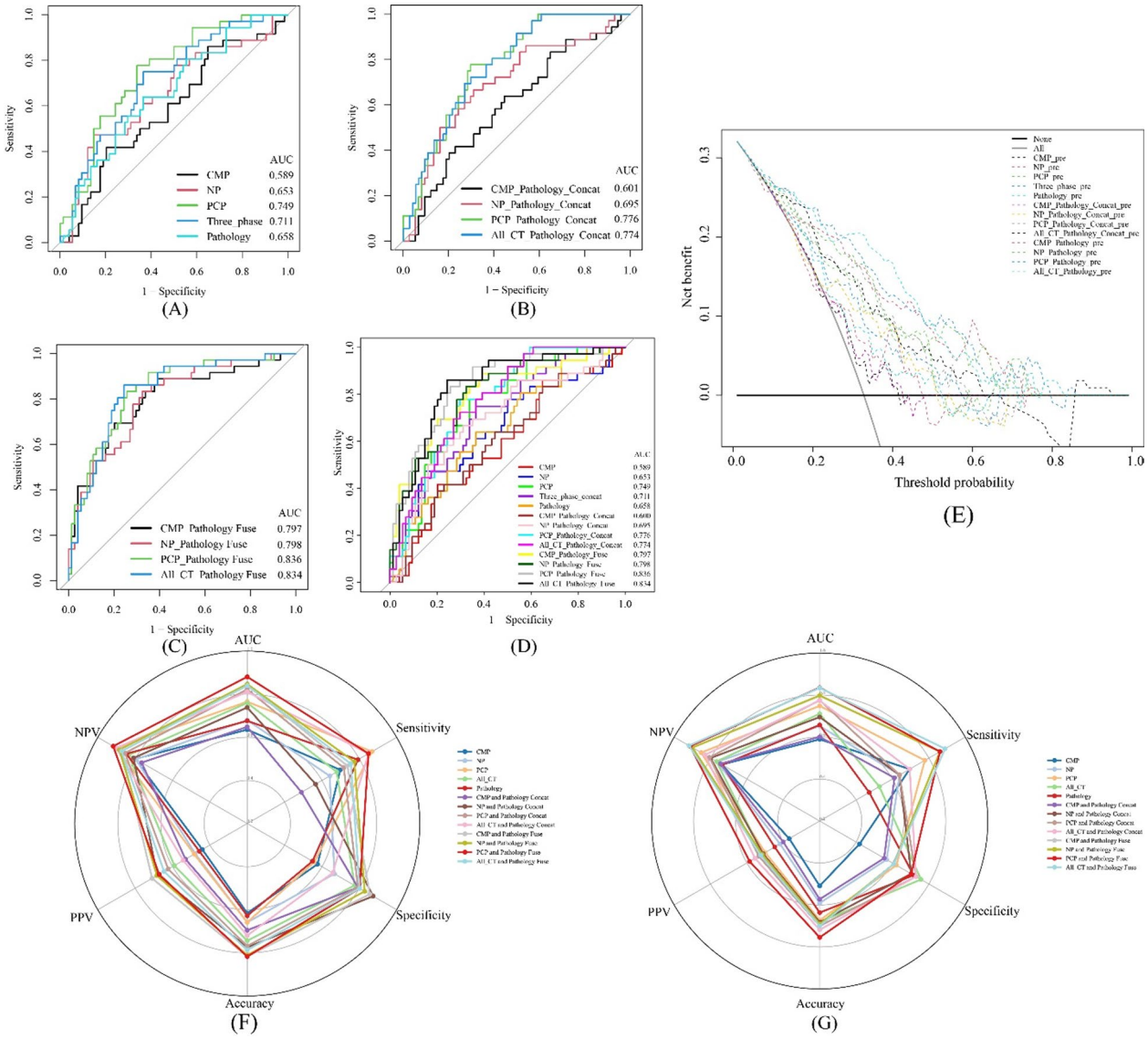

Fig. 4. ROC curves of the unimodal model (**A**), simple fusion model (**B**), cross-modal fusion model (**C**), all models (**D**), DCA curves of all models (**E**) in the test cohort; and radar charts of the training cohort (**F**) and the test cohort (**G**). *PCP* precontrast phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *AUC* area under the ROC curve, *DCA* decision curve analysis.

IDI and Kaplan–Meier analysis of DFS further confirmed that the PCP-based CPNet model significantly outperformed the other models in terms of prediction performance, except for the Three-phase-Pathology Fuse model, with which the difference in prediction performance was not significant ($p > 0.05$) (Table 3). The results of the Kaplan–Meier analysis of DFS for the multimodal interaction fusion models are presented in Fig. 5, and those of the other models are provided in Figure S6 in Supplementary Material.

Furthermore, the Delong's test of the training and test cohorts showed that PCP-Pathology Fuse model had statistical differences compared with the single modality model and simple multimode concat model ($p < 0.05$, except for the All CT and Pathology concat model), but no statistical differences compared with other multimodal interaction fusion models ($p > 0.05$). Although the p-value between the PCP-Pathology Fuse model and the All CT and Pathology concat model was 0.05704 in the training set, it reached 0.01622 in the test set, indicating a significant difference. Detailed results are provided in Table S3 in the Supplementary Material.

Visualization of CcRCC classes

In the pathological patches, the model focused on most areas, including the tumor nuclei, cytoplasm, and the microenvironment surrounding the tumor cells. In the categories of recurrence or metastasis, the model assigned higher correlation scores to the regions with a higher degree of malignancy within these patches; simultaneously, in contrast to the nonrecurrence category, the model also focused on a broader range of tumor CT image information.

To qualitatively analyze the effect of model-enhanced and recalibrated modality representations under guided modalities, we selected patches based on inter-modality correlation scores. When guiding the pathology modality with the CT modality, the top three pathology patches with the highest correlation to the CT modality were selected for display. Conversely, when guiding the CT modality with the pathology modality, the top three CT ROIs with the highest correlation to the pathology modality were chosen. Class activation mapping (CAM) was applied to visualize the selected images, as shown in Fig. 6A. We have observed that, within the pathological patches, the model focused on most areas including the tumor nuclei, cytoplasm, and the microenvironment surrounding the tumor cells. In the categories of recurrence or metastasis, the model assigned higher correlation scores to the regions with a higher degree of malignancy in the pathological patches; simultaneously, in contrast to the non-recurrence category, the model also focused on a broader range of tumor CT image information.

Furthermore, to quantitatively assess the model's discriminative ability between recurrence and non-recurrence classes—and thereby indirectly demonstrate its capacity to capture pathological microscopic information as well as CT macroscopic features—we performed a t-test on the model's predicted values.

Dataset	Model2	Model1	*p*-value
Training cohort	PCP and Pathology fuse	PCP	IDI [95% CI]: 0.2337 [0.1641–0.3034]; *p*-value: 0 < 0.01
		CMP	IDI [95% CI]: 0.3759 [0.2982–0.4536]; *p*-value: 0 < 0.01
		NP	IDI [95% CI]: 0.3386 [0.2649–0.4123]; *p*-value: 0 < 0.01
		Three-phase	IDI [95% CI]: 0.2554 [0.1786–0.3322]; *p*-value: 0 < 0.01
		Pathology	IDI [95% CI]: 0.3337 [0.2416–0.4257]; *p*-value: 0 < 0.01
		PCP and pathology concat	IDI [95% CI]: 0.1288 [0.0673–0.1903]; *p*-value: 4e-5 < 0.01
		CMP and pathology concat	IDI [95% CI]: 0.3679 [0.2908–0.445]; *p*-value: 0 < 0.01
		NP and pathology concat	IDI [95% CI]: 0.2576 [0.1886–0.3266]; *p*-value: 0 < 0.01
		All CT and pathology concat	IDI [95% CI]: 0.1785 [0.088–0.2689]; *p*-value: 0 < 0.01
		CMP and pathology fuse	IDI [95% CI]: 0.0772 [0.0099–0.1445]; *p*-value:0.02462 < 0.05
		NP and pathology fuse	IDI [95% CI]: 0.0825 [0.0274–0.1376]; *p*-value: 0.00333 < 0.01
		All CT and pathology fuse	IDI [95% CI]: 0.1031 [0.048–0.1582]; *p*-value: 0.00025 < 0.01
Test cohort	PCP and Pathology fuse	PCP	IDI [95% CI]: 0.1473 [0.0992–0.1954]; *p*-value: 0 < 0.01
		CMP	IDI [95% CI]: 0.2843 [0.2038–0.3647]; *p*-value: 0 < 0.01
		NP	IDI [95% CI]: 0.2566 [0.1819–0.3313]; *p*-value: 0 < 0.01
		Three-phase	IDI [95% CI]: 0.1883 [0.1278–0.2489]; *p*-value: 0 < 0.01
		Pathology	IDI [95% CI]: 0.2367 [0.1294–0.3441]; *p*-value: 2e-05 < 0.01
		PCP and pathology concat	IDI [95% CI]: 0.1121 [0.0728–0.1513]; *p*-value: 0 < 0.01
		CMP and pathology concat	IDI [95% CI]: 0.2794 [0.2000–0.3588]; *p*-value: 0 < 0.01
		NP and pathology concat	IDI [95% CI]: 0.2205 [0.1500–0.2911]; *p*-value: 0 < 0.01
		All CT and pathology concat	IDI [95% CI]: 0.1286 [0.0745–01828]; *p*-value: 0 < 0.01
		CMP and pathology fuse	IDI [95% CI]: 0.0629 [0.0018–0.124]; *p*-value: 0.04365 < 0.05
		NP and pathology fuse	IDI [95% CI]: 0.0649 [0.014–0.1157]; *p*-value: 0.01244 < 0.05
		All CT and pathology fuse	IDI [95% CI]: -0.0024 [-0.0396 -0.0348]; *p*-value: 0.89885 > 0.05

Table 3. IDI test between the fused model based on PCP and pathology and other models in the training cohort and test cohort. *PCP* precontrast phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *PPV* positive predictive value, *NPV* negative predictive value, *AUC* the area under the ROC curve, *IDI* integrated discrimination improvement.

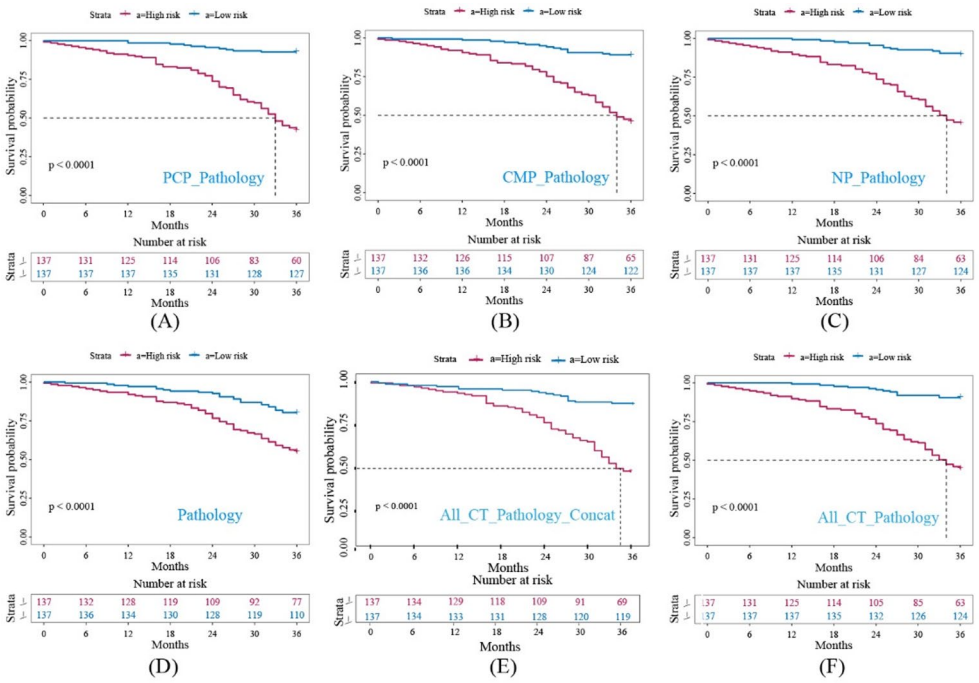

Fig. 5. Kaplan-Meier analysis of DFS in the PCP, CMP, NP-Pathology fuse models (**A**–**C**), Pathology model (**D**), All CT and pathology concat model (**E**), and all CT and pathology fuse model (**F**). *PCP* precontrast phase, *CMP* arterial phase, *NP* venous phase.

Additionally, we generated violin plots (Fig. 6B) to visually and quantitatively illustrate the distribution differences of the predicted values between the recurrence and non-recurrence groups. As shown in Fig. 6, there is a clear distinction in the predicted value distributions between the two classes, indicating that the model effectively captures discriminative features associated with recurrence risk.

Discussion

In this study, we developed DL-based CPNet models that effectively predicted the risk of local recurrence or metastasis in ccRCC patients after surgery. Our findings indicate that the CPNet model utilizing PCP and histopathological images outperformed other single-phase CT-based and pathology-based CPNet models. Its diagnostic performance was comparable to models relying on three-phase CT and pathology images, suggesting that the former could assist radiologists and urologists in stratifying and managing the risk of postoperative tumor recurrence or metastasis in ccRCC, as well as in developing individualized patient management strategies.

A core advantage of our framework lies in its bidirectional attention mechanism, which enables cross-modal information highlighting rather than simple feature concatenation. This design preserves prognostically critical relationships between macro-structural (CT) and microstructural (pathology) patterns, avoiding information dilution. Consequently, the CPNet models showed consistent performance gains: compared with WSI-based unimodal models, they achieved an average 15.85% higher AUC and 39.59% lower misdiagnosis rate; relative to CT-based unimodal models, average improvements were 15.71% in AUC, 18.41% in ACC, and 22.97% in misdiagnosis rate; and compared with simple concatenation-based fusion models, they demonstrated an average 11.83% higher AUC and 22.22% lower misdiagnosis rate.

Notably, among all models tested, PCP-based models (including unimodal PCP, PCP-Pathology Concat, and PCP-Pathology Fuse) exhibited the best performance in their respective categories, consistent with findings from previous PCP-based studies[25,26]. The superiority of PCP can be attributed to three key factors: (1)PCP-CT captures comprehensive whole-tumor information, fully reflecting internal tissue heterogeneity; (2)PCP images are more stable across individuals, with reduced variability from scanning protocols or contrast agent differences[27], enhancing generalizability; (3)clinically, PCP avoids repeated scanning, reducing radiation exposure and contrast agent use—particularly valuable for patients with contrast allergies or renal impairment, improving accessibility.

Integrating macro-CT and micro-pathology data is critical for capturing multi-scale tumor characteristics, as supported by relevant studies. Specifically, Song et al.[28] proposed a multimodal fusion method combining CT and pathology images using a Swin Transformer and attention mechanisms to enable multi-region and multi-scale fusion, effectively capturing interactions across different tumor regions and image scales. Similarly, Li et al.[29] developed a hierarchical multimodal fusion framework that captures interactions between pathological patches and radiological images at multiple scales through a layered architecture. These additions support the advantages of multimodal fusion in medical AI applications. Besides, as Li et al.[30] indicated, studies integrating CT and whole-slide imaging (WSI) for ccRCC prognosis remain scarce. Chen et al.[31] developed a multimodel prediction

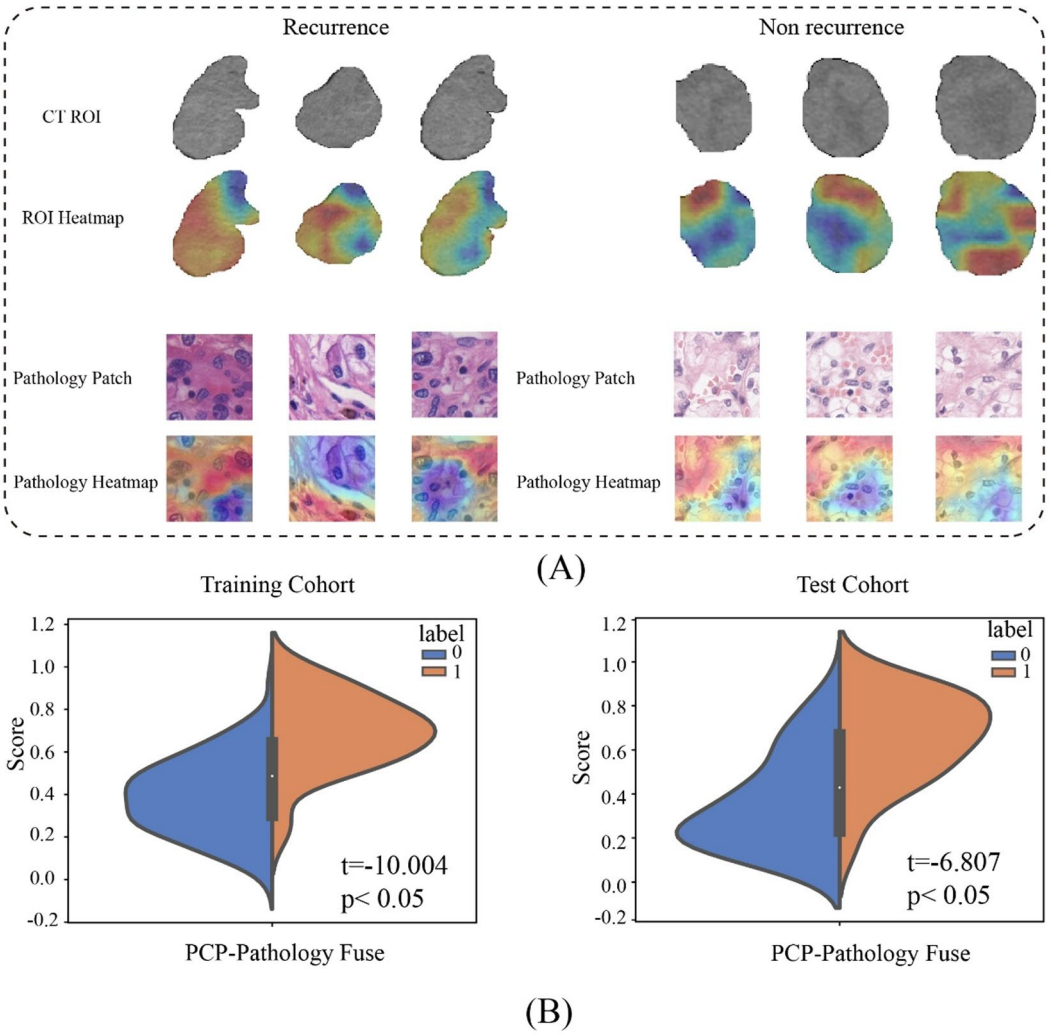

Fig. 6. Visualization of ccRCC classes. (**A**) Visualization images of top three pathological patches and CT ROIs with the highest inter-modality correlation by class activation mapping (CAM), (**B**) Violin plots of predicted values for recurrence and non-recurrence groups. The t-test shows significant distribution differences.

signature (MMPS) combining WSI, cell segmentation, CT, and clinical data, achieving AUCs of 0.742, 0.917, and 0.900 across cohorts. However, it relies on multiple data sources, increasing complexity. Our model focuses on fusing CT (especially pre-contrast phase) and WSI, avoiding redundancy from extra data. Aligning with Li et al.[30]'s note that interactive fusion better captures cross-modal correlations, our CPNet achieves competitive performance (AUC 0.8363) with greater clinical feasibility, reveals the potential of CT-WSI fusion for ccRCC prognosis.

Due to significant heterogeneity between WSI and CT data, simple feature concatenation (a common fusion method) may overlook inherent inter-modal correlations (e.g., the biological correspondence between histopathological patterns such as nuclear atypia or stromal composition, and imaging phenotypes such as tumor enhancement patterns or spatial heterogeneity), resulting in suboptimal model performance. Our CPNet model advances this by enhancing cross-modal alignment: under mutual guidance, it recalibrates modality-specific representations, strengthening correlations between pathological features (e.g., nuclear atypia) and imaging phenotypes (e.g., enhancement patterns). IDI and DCA analyses confirmed that the PCP-Pathology Fuse model had superior diagnostic capability and net clinical benefit, supporting its potential for predicting postoperative DFS in ccRCC.

We also observed a trend where model performance improved with effective feature integration but was constrained by excessive redundancy. Simple fusion models outperformed unimodal models, and CPNet (interactive fusion) outperformed simple fusion, but the Three-phase-Pathology Fuse model showed limited gains due to increased inter-modal heterogeneity. Visualization of top-correlated pathological patches and CT ROIs further indicated CPNet's ability to identify heterogeneous regions across modalities.

In clinical workflow, this model could potentially complement existing tools (e.g., TNM staging) by refining risk stratification: high-risk patients identified preoperatively might benefit from more aggressive resection or adjuvant therapy, while low-risk patients could undergo less frequent surveillance. When considering

practical clinical application, this model has the potential to strike a balance between meeting requirements and practicality: it only requires an initial investment of computing resources and can be scaled according to institutional needs. Moreover, by fine-tuning with local data, it can generate predictive results within minutes without imposing an additional burden on the clinical workflow, thereby enabling real-time decision support. This personalized management approach has the potential to enhance patient treatment outcomes while also optimizing the utilization of medical resources.

Therefore, the PCP-based CPNet model shows promise to predict the risk of postoperative recurrence or metastasis in patients with ccRCC and for aiding in developing plans to improve the management of these patients. Moreover, this approach may reduce patients' economic burden and lessen imaging-related radiation doses and contrast agent use—potential benefits with clinical and economic value. Yet such potential (including associated costs) needs further validation in cost-effectiveness studies, though PCP-based protocols align with established strategies to minimize patient exposure.

Limitations

This study is subject to several limitations.First, its retrospective single-center design yielded a cohort with high geographic homogeneity. This may restrict the model's generalizability to patients from other regions or with distinct demographic traits (e.g., ethnic background, dietary habits), as regional differences in diagnostic protocols and patient populations could impair model performance.Second, the current CPNet model only integrates CT and WSI data, and excludes clinically relevant variables—such as patient comorbidities (e.g., hypertension, diabetes), smoking history, and molecular biomarkers (e.g., VHL mutation status). Furthermore, the selection of slides may introduce potential bias, given that this subjective process could influence the results. Third, patient characteristics, imaging equipment, and diagnostic standards can all impact the model. As such, its generalizability cannot be fully evaluated in the present study and necessitates further assessment in future research. To address these limitations, we recommend conducting a multicenter prospective study focused on patients with ccRCC. Specifically, data from diverse CT scanners should be utilized to validate the model's generalizability across different institutions. Furthermore, integrating electronic health records and TCGA ccRCC genomic data will enrich the model's input, thereby facilitating the development of a comprehensive multi-dimensional framework.

Conclusion

Overall, this study demonstrates that the PCP-based CPNet model shows promise for predicting postoperative disease-free survival (DFS) in ccRCC patients and warrants external validation across multi-institutional cohorts before potential clinical integration.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 7 April 2025; Accepted: 17 October 2025
Published online: 21 November 2025

References

1. Rathmell, W. K. et al. Management of metastatic clear cell renal cell carcinoma: ASCO guideline. *JCO* **40** (25), 2957–2995 (2022).
2. Hu, J. et al. Tumor heterogeneity in VHL drives metastasis in clear cell renal cell carcinoma. *Sig Transduct. Target. Ther.* **8** (1), 155 (2023).
3. Ke, J., Chen, J. & Liu, X. Analyzing and validating the prognostic value and immune microenvironment of clear cell renal cell carcinoma. *Anim. Cells Syst.* **26** (2), 52–61 (2022).
4. Wang, D. et al. Integrated analysis of the roles of oxidative stress related genes and prognostic value in clear cell renal cell carcinoma. *J. Cancer Res. Clin. Oncol.* **149** (13), 11057–11071 (2023).
5. Zhang, H. et al. Research on molecular characteristics of ADME-related genes in kidney renal clear cell carcinoma. *Sci. Rep.* **14** (1), 16834 (2024).
6. Zhou, F. & Chen, H. Cross-Modal Translation and Alignment for Survival Analysis. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 21428–21437 (2023).
7. Ning, Z. et al. Integrative analysis of cross-modal features for the prognosis prediction of clear cell renal cell carcinoma. *Bioinformatics.* **36**(9), 2888–2895 (2020).
8. Yang, Z. et al. Imaging features and clinic value of mri and Ct in diagnosis of clear cell renal cell carcinoma. *Food Sci. Technol.* **42**, e40520 (2022).
9. Nyman, J. et al. Spatially aware deep learning reveals tumor heterogeneity patterns that encode distinct kidney cancer States. *Cell. Rep. Med.* **4** (9), 101189 (2023).
10. Synnott, N. C. et al. Characterizing the tumor microenvironment in rare renal cancer histological types. *J. Pathol. CR.* **8** (1), 88–98 (2022).
11. Fu, Y. et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer*. **1** (8), 800–810 (2020).
12. Byun, S-S. et al. Deep learning based prediction of prognosis in nonmetastatic clear cell renal cell carcinoma. *Sci. Rep.* **11** (1), 1242 (2021).
13. Xi, I. L. et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. *Clin. Cancer Res.* **26** (8), 1944–1952 (2020).
14. Suarez-Ibarrola, R. 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. Urol.* **38** (10), 2329–2347 (2020).
15. Ursprung, S. et al. Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis. *Eur. Radiol.* **30** (6), 3558–3566 (2020).
16. Nie, P. et al. A CT-based deep learning radiomics nomogram outperforms the existing prognostic models for outcome prediction in clear cell renal cell carcinoma: a multicenter study. *Eur. Radiol.* **33** (12), 8858–8868 (2023).

17. Ohe, C. et al. Deep learning-based predictions of clear and eosinophilic phenotypes in clear cell renal cell carcinoma. *Hum Pathol.* **131**, 68-78.2022 (2023).
18. Wessels, F. et al. Deep learning can predict survival directly from histology in clear cell renal cell carcinoma. *PLOS ONE*. **17** (8), e0272656 (2022).
19. Schulz, S. et al. Multimodal deep learning for prognosis prediction in renal cancer. *Front. Oncol.* **11**, 788740 (2021).
20. Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage* **31** (3), 1116–1128 (2006).
21. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778 (2016).
22. Kar, A. et al. Clear cell lesions in pathology: histomorphologic approach to diagnosis. *Indian J. Pathol. Microbiol.* **63** (2), 177–187 (2020).
23. Ahmad, Z. et al. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, Obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagn. Pathol.* **16** (1), 24 (2021).
24. Zhou, B. et al. Learning Deep Features for Discriminative Localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2921–2929 (2016).
25. Feliciani, G. et al. Radiomics in the characterization of lipid-poor adrenal adenomas at unenhanced CT: time to look beyond usual density metrics. *Eur. Radiol.* **34** (1), 422–432 (2023).
26. Yi, X. et al. Radiomics improves efficiency for differentiating subclinical pheochromocytoma from lipid-poor adenoma: a predictive, preventive and personalized medical approach in adrenal incidentalomas. *EPMA J.* **9** (4), 421–429 (2018).
27. Torresan, F. et al. Radiomics: a new tool to differentiate adrenocortical adenoma from carcinoma. *BJS Open.* **5** (1), zraa061 (2021).
28. Song, B. et al. Deep learning informed multimodal fusion of radiology and pathology to predict outcomes in HPV-associated oropharyngeal squamous cell carcinoma. *eBioMedicine* **114**, 105663 (2025).
29. Li, Z. et al. Survival prediction via hierarchical multimodal Co-Attention transformer: A computational Histology-Radiology solution. *IEEE Trans. Med. Imaging.* **42** (9), 2678–2689 (2023).
30. Li, M-Y. et al. Digital pathology and artificial intelligence in renal cell carcinoma focusing on feature extraction: a literature review. *Front. Oncol.* **15**, 1516264 (2025).
31. Chen, S. et al. Deep learning-based multimodel prediction for disease-free survival status of patients with clear cell renal cell carcinoma after surgery: a multicenter cohort study. *Int. J. Surg.* **110** (5), 2970–2977 (2024).

Author contributions

C.M., B.F.: Researching, Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing - Original Draft; Y.L., Z.Y.: Data Curation, Writing - Original Draft; Z.Y., Y.L.: Software, Visualization, Investigation; R.L., X.H.:Resources, Supervision; B.F., Y.L.: Validation; B.W., Z.L.: Researching, Data Curation, Resources; E.C. (Corresponding Author): Conceptualization, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.Among the listed authors, Changyi Ma and Bao Feng jointly take charge of Researching，Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing Original Draft and made the same contribution to this paper. After consultation, all the authors also agreed to the addition of Bao Feng as co-first author in this paper, and all the authors agreed with the arrangement of the names. All authors reviewed the manuscript.

Funding

This study has received fundings by Guangdong Basic and Applied Basic Research Foundation (Grant Number: 2021A1515220080) and Jiangmen Science and Technology Plan Program (Grant Number: 2024YL01033).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This retrospective study complied with the Declaration of Helsinki, and the study was approved by the ethics review board of Jiangmen Central Hospital, and the requirement for informed consent was waived.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-25109-x.

Correspondence and requests for materials should be addressed to E.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2025