npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-02034-x

A multimodal AI model for precision prognosis in clear cell renal cell carcinoma: A multicenter study

Check for updates

Xinyi Zang[1,16], Yujia Xia[2,3,16], Haibing Xiao[4,16], Haolun Luo[5,16], Minggui Si[6], Naiqiao Hou[1], Axide Haoni[1], Tianyi Chen[1], Ziyi Liu[1], Xinyuan Pu[1], Xiangyu Zi[1], Lijun Xu[7], Jin Zhu[7], Zhipeng Xu[8], Jianning Wang[8], Zaoyu Wang[9], Jun Xia[9], Dengfeng Cao[9], Yu Yin[10], Jieying Wang[11,12], Xiaorong Wu[1], Wen Kong[1], Jiwei Huang[1], Jin Zhang[1], Yonghui Chen[1], Yiran Huang[1], David Ka-Wai Leung[13], Jeremy Yuen-Chun Teoh[13], Keliang Wang[6] ✉, Chaozhao Liang[4] ✉, Junhua Zheng[1] ✉, Zhangsheng Yu[2,3,14] ✉ & Wei Zhai[1,15] ✉

Patients with clear cell renal cell carcinoma (ccRCC) face a high risk of recurrence after surgery, but existing clinical tools based on clinicopathological factors or costly molecular profiling often lack precision and clinical feasibility. We developed the multimodal predictive recurrence score (MPRS), a multimodal prognostic model using clinical features, CT images, and histopathological whole-slide images (WSIs) from 1648 patients across six centers and the TCGA database. MPRS outperformed unimodal models and clinical tools (Leibovich and UISS scores, KEYNOTE-564 risk classification), achieving C-index values of 0.886 and 0.838 in the internal and external validation cohorts, respectively. Importantly, MPRS correctly reclassified 83.3% (50/60) of KEYNOTE-564-defined low-risk recurrence patients as high-risk, avoiding inadequate adjuvant therapy, while reclassifying 57.7% (15/26) of KEYNOTE-564-defined intermediate/high-risk non-recurrence patients as low-risk, preventing excessive adjuvant therapy. By leveraging routinely available data, MPRS provides a cost-effective and accurate approach for recurrence risk stratification, optimizing personalized ccRCC management and therapeutic decision-making.

Renal cell carcinoma (RCC), which is one of the most prevalent malignancies of the urinary system, accounts for more than 430,000 new cases worldwide and resulted in more than 150,000 deaths in 2022[1]. Clear cell renal cell carcinoma (ccRCC) is the most common subtype of RCC and accounts for approximately 70% of renal cancer cases[2]. Although surgical intervention is the primary initial treatment, approximately 20–30% of ccRCC patients experience a recurrence and metastasis after surgery, and the disease is typically incurable[3]. Adjuvant therapy can reduce the risk of recurrence and metastasis but is accompanied by certain toxicities[4]. Therefore, current clinical guidelines recommend the use of adjuvant

[1]Department of Urology, Renji Hospital, School of Medicine in Shanghai Jiao Tong University, Shanghai, China. [2]Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [3]SJTU-Yale Joint Center for Biostatistics and Data Science, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [4]Department of Urology, The First Affiliated Hospital of Anhui Medical University, Anhui Medical University, Hefei, Anhui, China. [5]Department of Urology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan. [6]Department of Urology, The Fourth Affiliated Hospital of Harbin Medical University, Heilongjiang Key Laboratory of Scientific Research in Urology, Harbin, China. [7]Department of Urology, The Second Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China. [8]Department of Urology, The First Affiliated Hospital of Shandong First Medical University, Jinan, China. [9]Department of Pathology, Renji Hospital, School of Medicine in Shanghai Jiao Tong University, Shanghai, China. [10]Department of Pathology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China. [11]Department of Nephrology, Molecular Cell Laboratory for Kidney Disease, Shanghai Peritoneal Dialysis Research Center, Ren Ji Hospital, Uremia Diagnosis and Treatment Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [12]Clinical Research Center, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [13]S.H. Ho Urology Centre, Department of Surgery, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. [14]Clinical Research Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [15]Shanghai Immune Therapy Institute State, Key Laboratory of Oncogenes and Related Genes, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. [16]These authors contributed equally: Xinyi Zang, Yujia Xia, Haibing Xiao, Haolun Luo. ✉e-mail: wangkeliang@hrbmu.edu.cn; liang_chaozhao@ahmu.edu.cn; zhengjh0471@sina.com; yuzhangsheng@sjtu.edu.cn; jacky_zw2002@hotmail.com

therapy for treating patients with intermediate risk, high risk, or M1 no evidence of disease (NED) status on the basis of the risk classification of the KEYNOTE-564 clinical trial, which was constructed with clinicopathological risk factors[3,5,6]. Additionally, clinical tools, such as the Leibovich score and UISS score, are widely recognized as effective prognostic assessment indicators[7,8]. In recent years, models incorporating molecular factors have also been developed[9,10], but in clinical practice, the need to integrate multiple complex factors, the high heterogeneity that often occurs in the assessment of medical image data, such as radiological and histopathological images, and the long turnaround time and high cost of molecular testing remain challenges that cannot be overlooked. Furthermore, medical images involve a wealth of underutilized visual information. These details are often difficult to capture with the naked eye, which may lead radiologists and pathologists to overlook critical information.

Routine disease diagnostic processes generate a wealth of medical information, including clinical features, laboratory test results, radiological images, and histological analyses, each of which provides valuable insights into the diagnosis and prognosis of diseases. In recent years, deep learning models have shown immense potential in utilizing medical data for tumor classification, molecular subtype identification, prognosis prediction, and treatment response evaluation, often outperforming the capabilities of experts in these fields[11–14]. In particular, multimodal deep learning models, which integrate diverse data modalities, can fully explore the complexity and heterogeneity of data and are expected to overcome the limitations of single-modality approaches. In some studies, multimodal deep learning methods have been used to analyze tumor prognosis; however, these methods often rely on the analysis of more complex tumor components, such as genomic data, and remain challenging to implement widely in clinical practice[15,16]. For the specific clinical scenario of ccRCC, a method is needed to construct a multimodal predictive model on the basis of routine clinical data in order to more accurately assess the risk of recurrence and inform clinical trial stratification; this is a critical issue that must be addressed.

In this study, we developed a multimodal predictive recurrence score (MPRS) by integrating clinical features, contrast-enhanced computed tomography (CT) images, and histopathological whole-slide images (WSIs) of tumor sections into the model. The MPRS is a promising method for improving the prognostic assessment and prolonging survival outcomes for ccRCC patients by assisting clinicians in developing personalized follow-up strategies and screening suitable candidates for adjuvant therapy.

Results

Study design and patient baseline characteristics

As shown in Fig. S1, following the application of the inclusion and exclusion criteria, 1145 patients from six centers in China were enrolled. The discovery cohort included 788 patients from three centers, namely, Renji Hospital, the Fourth Affiliated Hospital of Harbin Medical University, and the Second Affiliated Hospital of Soochow University. Among these patients, 134 patients with complete multimodal information were randomly selected from the discovery cohort for inclusion in the internal validation cohort, and the remaining 654 patients formed the training cohort. An external validation cohort comprising 357 patients from the First Affiliated Hospital of Anhui Medical University, Kaohsiung Chang Gung Memorial Hospital, and the First Affiliated Hospital of Shandong First Medical University was subsequently established. Critically, all patients in both the internal validation cohort ($n = 134$) and external validation cohort ($n = 357$) had complete clinical, radiological, and histopathological data.

In the training cohort ($n = 654$), patients who were missing one or two modalities were utilized to develop the corresponding unimodal prognostic models: patients whose clinical data contributed to the clinical predictive recurrence score (CPRS), those whose radiological images contributed to the radiological predictive recurrence score (RPRS), and those whose WSIs contributed to the histopathological predictive recurrence score (HPRS). In addition, to augment the histopathology-specific training data, the WSIs from 503 ccRCC patients in The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) were incorporated into only the HPRS model training. To develop the MPRS, which integrates features from all three modalities (HPRS, RPRS, and CPRS), only the subset of patients in the training cohort with data for all three modalities ($n = 550$) was utilized. Consequently, while the WSIs from TCGA strengthened the unimodal HPRS model, they were not directly used in the multimodal MPRS training, which relied solely on the in-house patients with all three modalities. The final MPRS model benefited indirectly from the enhanced HPRS features that were obtained from the larger combined histopathology dataset.

The overall design of the study is presented in Fig. 1. We collected rich and comprehensive multimodal data that were generated during routine ccRCC diagnosis and treatment, including data on clinical features, preoperative contrast-enhanced CT, and postoperative histopathological WSIs of tumor sections (Fig. S2). In total, 4001 WSIs and 1098 contrast-enhanced CT images were included. For both the radiological and histopathological modalities, deep learning and radiomic or pathomic features were extracted to construct unimodal models. Finally, the clinical, radiological, and histopathological modalities were integrated to construct the MPRS to accurately predict recurrence and stratify patients according to risk, and the MPRS was validated in an external cohort.

The demographics and baseline characteristics of the enrolled patients are detailed in Table 1. The baseline characteristics were evenly distributed between the cohorts, with no significant differences ($p > 0.05$). Overall, the majority of patients were male (67.4%), with a median age of 60 years (interquartile range [IQR], 52–67 years), and nearly one-third had an Eastern Cooperative Oncology Group performance status (ECOG-PS) of ≥1. In terms of clinicopathological characteristics, most patients were diagnosed with TNM stage I (79.0%) and histopathological grade 2 (63.5%) disease. A proportion of patients exhibited adverse prognostic features, such as tumor necrosis (14.0%) or sarcomatoid differentiation (1.7%). In the different risk stratification models, most patients were classified as being at low or intermediate risk. During a median follow-up of 49 months (IQR, 33–67 months), 197 patients (17.2%) experienced ccRCC recurrence.

Predictive performance and comparison of the MPRS with unimodal and clinical prognostic tools

Compared with all unimodal models (HPRS, RPRS, and CPRS) and established clinical tools (Leibovich score, UISS score, and KEYNOTE-564 risk classification), the MPRS showed superior predictive performance across all cohorts. The MPRS achieved significantly higher C-index values in the training (0.924), internal validation (0.886), and external validation (0.838) cohorts (Fig. 2a–c). For time-specific recurrence prediction, the MPRS consistently yielded the highest AUC values, ranging from 0.835 to 0.944 at 3 years and from 0.829 to 0.960 at 5 years (Fig. S3). Compared with traditional machine learning models, such as random survival forest (RSF), gradient boosting for survival (GBS), and survival support vector machine (SSVM), we found that the MPRS had better accuracy and robustness (Fig. S4). When evaluated at each individual center, MPRS demonstrated high discrimination with C-index values of 0.909–0.957 in the training cohort, 0.833–0.896 in the internal validation cohort, and 0.803-0.847 in the external validation cohort (Fig. S5). Additionally, the calibration curves for the MPRS at 3 years and 5 years in each cohort closely aligned with the 45° line, indicating strong concordance between the MPRS predictions and observed outcomes (Fig. 2d–f).

Moreover, the optimal threshold for the MPRS was determined to be −1.955 at the maximum Youden's index in the ROC analysis in the training cohort; thus, the patients were divided into high- and low-risk groups. This stratification revealed marked differences in DFS in the training cohort ($p < 0.001$). The prognostic value of the MPRS was consistently validated in both the internal and external validation cohorts, with the difference in DFS remaining highly significant between the risk groups ($p < 0.001$) (Fig. 2g-i). Notably, the distribution of the MPRS showed the most distinct separation between patients with and without recurrence compared with the unimodal scores across all cohorts, demonstrating a clear trend toward higher scores indicating greater recurrence risk (Fig. S6a–d). A prognostic landscape plot further illustrated the associations between the ascending MPRS scores, patient survival status, and key clinical prognostic factors (Fig. S6e).

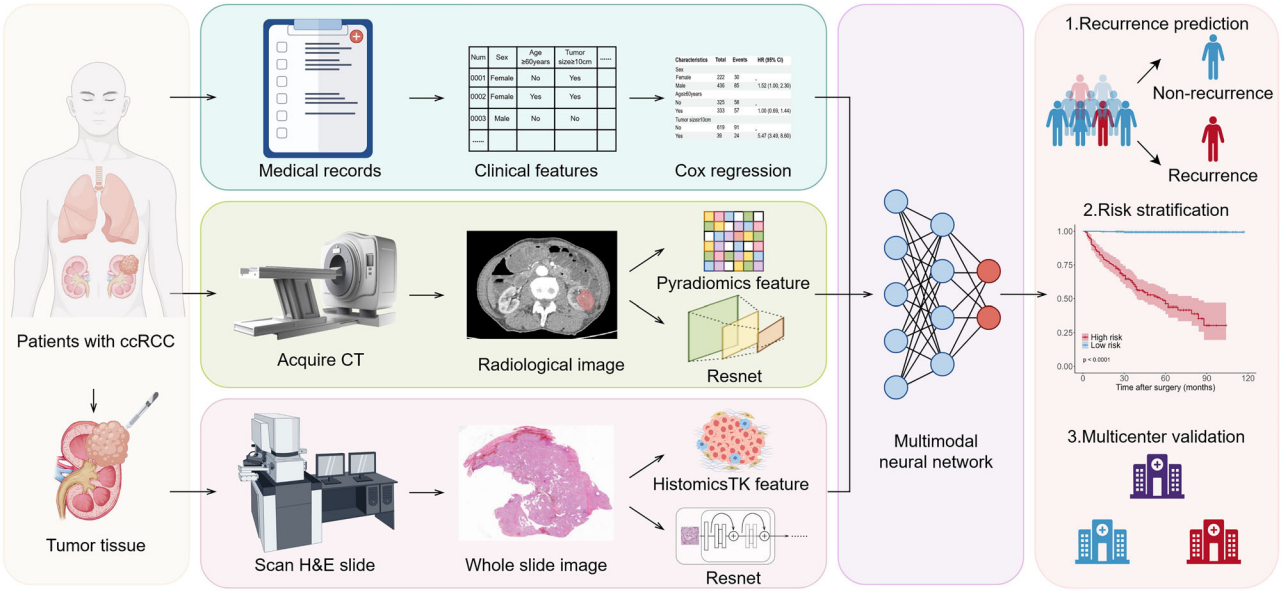


Fig. 1 | Study workflow. Multimodal data, including clinical features, preoperative contrast-enhanced CT images, and histopathological whole-slide images (WSIs) of tumor sections, were acquired through conventional diagnostic procedures. The multimodal data were analyzed by means of numerous distinct approaches, including Cox regression, PyRadiomics features, ResNet, and HistomicsTK features. The clinical, radiological and histopathological models were fused to construct a multimodal predictive recurrence score (MPRS) that was used for the postoperative risk stratification of ccRCC patients, which was validated in multicenter cohorts. ccRCC=clear cell renal cell carcinoma. CT = computed tomography. WSI=whole slide image.

To assess the independent prognostic value of the MPRS, we conducted Cox regression analyses incorporating basic clinical features, including sex, age, and TNM stage. Across all cohorts, high risk according to the MPRS was significantly associated with poor DFS (training cohort: HR = 106.07, 95% CI: 25.68–438.21, $p < 0.001$; internal validation cohort: HR = 13.59, 95% CI: 3.63–50.83, $p < 0.001$; external validation cohort: HR = 10.70, 95% CI: 5.07–22.58, $p < 0.001$) (Table 2). These findings indicate that the MPRS is a robust and independent prognostic factor for predicting DFS in patients with ccRCC.

Subgroup risk stratification analyses for the MPRS

Given that current guidelines recognize TNM stage, histopathological grade, the Leibovich score, the UISS score, and the KEYNOTE-564 risk classification as key prognostic factors in ccRCC, we evaluated whether the MPRS could provide further prognostic refinement within these subgroups, with a particular focus on the KEYNOTE-564 risk classification subgroup. According to the KEYNOTE-564 risk classification, in the entire cohort, 930 patients (low risk) did not require adjuvant therapy, and 111 patients (106 at intermediate risk and five at high risk) did require adjuvant therapy, while the MPRS classified 699 patients as being at low risk and 342 as being at high risk (Fig. 3a). Kaplan–Meier analyses revealed that the MPRS markedly improved risk stratification in both the KEYNOTE-564 low-risk group and the intermediate/high-risk group ($p < 0.001$, Fig. 3b–d). Furthermore, patients in the KEYNOTE-564 intermediate/high-risk group who were reclassified as being at low risk by the MPRS had significantly better survival than those in the KEYNOTE-564 low-risk group who were reclassified as being at high risk by the MPRS (Fig. 3e). Consistent prognostic refinement by the MPRS was also observed across subgroups defined by the TNM stage, histopathological grade, the Leibovich score, and the UISS score (Fig. S7–S10).

Interpretability with clinical insights

To explore the consistency between our model's predictions and the established clinical knowledge, we performed explainability analyses on multimodal information. First, we calculated the SHapley Additive exPlanations (SHAP) values for clinical features to interpret their contributions to the model's output. The results revealed that histopathological grade, TNM stage, tumor size, and necrosis had significant effects on the prediction of tumor recurrence, whereas the influence of age and ECOG-PS was minimal (Fig. 4a). Additionally, we utilized gradient-weighted class activation mapping (Grad-CAM) for visual qualitative analyses of radiological and histopathological images to identify regions of interest focused on by the model. In terms of radiological modality, the model focused particularly on irregular tumor margins[17] and regions with heterogeneous enhancement[18], these features are typically associated with poor prognosis in ccRCC patients (Fig. 4b, c). In terms of histopathological modality, the model effectively identified known histopathological features indicating poor prognosis in ccRCC, such as sarcomatoid differentiation[19,20] and other high-grade tumor regions (Fig. 4d, e). Moreover, for the extracted set of 365-dimensional clinical, radiomic, and pathomic features, we performed SHAP analysis and visualized the SHAP values for the top 15 most important features. The results revealed that these key features consisted of five radiomic features, six pathomic features, and four clinical features, demonstrating that radiomic and pathomic features can provide additional predictive information beyond baseline clinical features such as grade and TNM stage (Fig. S11). These analyses demonstrate that the MPRS can extract pertinent knowledge from patients' multimodal information to accurately predict ccRCC recurrence. The identification of these critical decision variables aligns with current clinical findings, confirming the reliability and clinical relevance of the MPRS.

Error analysis of misclassifications by the MPRS compared to clinical risk stratification tools

To evaluate the performance of the MPRS and established clinical risk stratification tools in identifying subgroups at high and low risk of recurrence, we computed confusion matrices by combining the data from the internal and external validation cohorts. The MPRS showed a balanced predictive profile with 85.4% sensitivity and 78.2% specificity (Fig. 5a). In contrast, all the clinical tools showed notable imbalances: the KEYNOTE-564 risk classification offered high specificity (93.6%) but critically low sensitivity (26.8%); the Leibovich score showed moderate specificity (83.1%) with low sensitivity (54.9%); and the UISS score achieved high sensitivity (81.7%) but poor specificity (50.4%) (Fig. 5b–d). Similar results were observed when the validation cohorts were analyzed separately (Fig.

Table 1 | Patient demographics and baseline characteristics

Characteristic	Group				p value
	Overall, N = 1,145[1]	Training cohort, N = 654[1]	Internal validation cohort, N = 134[1]	External validation cohort, N = 357[1]	
Sex					0.277[2]
Female	373 (32.6%)	222 (33.9%)	36 (26.9%)	115 (32.2%)	
Male	772 (67.4%)	432 (66.1%)	98 (73.1%)	242 (67.8%)	
Age, years	60.00 (52.00,67.00)	60.00 (52.00,67.75)	61.00 (52.00,67.75)	59.00 (52.00,66.00)	0.462[4]
ECOG PS ≥ 1					0.903[2]
No	771 (67.3%)	441 (67.4%)	88 (65.7%)	242 (67.8%)	
Yes	374 (32.7%)	213 (32.6%)	46 (34.3%)	115 (32.2%)	
Tumor size ≥ 10 cm					0.240[2]
No	1,087 (94.9%)	615 (94.0%)	130 (97.0%)	342 (95.8%)	
Yes	58 (5.1%)	39 (6.0%)	4 (3.0%)	15 (4.2%)	
pT stage					0.233[2]
T1a	299 (26.1%)	186 (28.4%)	31 (23.1%)	82 (23.0%)	
T1b	609 (53.2%)	325 (49.7%)	77 (57.5%)	207 (58.0%)	
T2	124 (10.8%)	73 (11.2%)	15 (11.2%)	36 (10.1%)	
T3	113 (9.9%)	70 (10.7%)	11 (8.2%)	32 (9.0%)	
pN stage					0.368[3]
N0	1,136 (99.2%)	647 (98.9%)	133 (99.3%)	356 (99.7%)	
N1	9 (0.8%)	7 (1.1%)	1 (0.7%)	1 (0.3%)	
AJCC TNM stage					0.724[2]
I	905 (79.0%)	509 (77.8%)	108 (80.6%)	288 (80.7%)	
II	122 (10.7%)	71 (10.9%)	15 (11.2%)	36 (10.1%)	
III	118 (10.3%)	74 (11.3%)	11 (8.2%)	33 (9.2%)	
Grade					0.506[2]
1	119 (10.4%)	68 (10.4%)	16 (11.9%)	35 (9.8%)	
2	727 (63.5%)	421 (64.4%)	82 (61.2%)	224 (62.7%)	
3	250 (21.8%)	133 (20.3%)	29 (21.6%)	88 (24.6%)	
4	49 (4.3%)	32 (4.9%)	7 (5.2%)	10 (2.8%)	
Necrosis					0.406[2]
No	985 (86.0%)	557 (85.2%)	120 (89.6%)	308 (86.3%)	
Yes	160 (14.0%)	97 (14.8%)	14 (10.4%)	49 (13.7%)	
Sarcomatoid RCC					0.311[3]
No	1,126 (98.3%)	641 (98.0%)	131 (97.8%)	354 (99.2%)	
Yes	19 (1.7%)	13 (2.0%)	3 (2.2%)	3 (0.8%)	
Leibovich group					0.668[2]
Low risk	861 (75.2%)	484 (74.0%)	104 (77.6%)	273 (76.5%)	
Intermediate risk	217 (19.0%)	129 (19.7%)	21 (15.7%)	67 (18.8%)	
High risk	67 (5.9%)	41 (6.3%)	9 (6.7%)	17 (4.8%)	
UISS group					0.813[2]
Low risk	493 (43.1%)	272 (41.6%)	61 (45.5%)	160 (44.8%)	
Intermediate risk	589 (51.4%)	345 (52.8%)	67 (50.0%)	177 (49.6%)	
High risk	63 (5.5%)	37 (5.7%)	6 (4.5%)	20 (5.6%)	
KEYNOTE-564 group					0.563[3]
Low risk	1016 (88.7%)	573 (87.6%)	121 (90.3%)	322 (90.2%)	
Intermediate-high risk	120 (10.5%)	74 (11.3%)	12 (9.0%)	34 (9.5%)	
High risk	9 (0.8%)	7 (1.1%)	1 (0.7%)	1 (0.3%)	
Recurrence					0.614[2]
No	948 (82.8%)	539 (82.4%)	115 (85.8%)	294 (82.4%)	
Yes	197 (17.2%)	115 (17.6%)	19 (14.2%)	63 (17.6%)	

Table 1 (continued) | Patient demographics and baseline characteristics

Characteristic	Group				p value
	Overall, N = 1,145[1]	Training cohort, N = 654[1]	Internal validation cohort, N = 134[1]	External validation cohort, N = 357[1]	
DFS, month	49.00 (33.00,67.00)	50.50 (35.00, 67.00)	47.50 (31.25, 67.75)	49.00 (31.00, 66.00)	0.547[4]

[1]n (%); Median (IQR).
[2]Pearson's chi-squared test.
[3]Fisher's exact test.
[4]Kruskal-Wallis rank-sum test.

S12a–h). Overall, these results suggest that the MPRS provides the most balanced predictive performance, optimizing sensitivity without compromising specificity; this is a critical advantage for clinical risk stratification where both under- and overtreatment have serious consequences.

Moreover, comparison heatmaps were generated to assess predictions between the MPRS and established clinical risk stratification tools in the combined validation cohorts. Given the clinical importance of the KEYNOTE-564 classification in guiding adjuvant immunotherapy, its comparison with the MPRS is highlighted (Fig. 5e). The two methods correctly classified 66.2% of the patients ($n = 325$), whereas only 4.2% ($n = 21$) were misclassified by both methods. Importantly, among the 60 recurrence patients misclassified as KEYNOTE-564 low-risk, the MPRS correctly reclassified 83.3% (50/60) as high-risk. Among the 26 non-recurrence patients misclassified as KEYNOTE-564 intermediate/high-risk, the MPRS correctly reclassified 57.7% (15/26) as low-risk. Similar analyses for the Leibovich score (high in false negatives) and UISS score (high in false positives) demonstrated that the MPRS correctly reclassified 78.4% (29/37) of the Leibovich false negatives and 69.0% (140/203) of the UISS false positives (Fig. 5f, g). These trends were consistent with those of the individual validation cohort analyses (Figure. S12i-n). In summary, the MPRS significantly rectified critical misclassifications by existing clinical tools, highlighting its practical value for refining clinical risk assessment.

Subgroup analyses revealed distinct patterns in the false-negative rate (FNR) and false-positive rate (FPR) of the MPRS across clinical characteristics (Table 3). The FNR was significantly elevated only in necrosis-negative tumors. The FPR was increased in male patients, patients with advanced TNM stages (II/III), patients with high-grade tumors (G3/4), and patients with necrosis. Although it was associated with a high FPR, sarcomatoid differentiation was excluded from further analysis because of the limited sample size ($n = 3$).

To gain deeper insights into the specific patterns of MPRS failure that were identified in the subgroup analyses, we performed targeted case reviews and analyzed radiological and histopathological images using the Grad-CAM method to identify error sources, resulting in detailed examinations of representative FN and FP cases. For the FNR analysis, we specifically selected a representative patient with recurrence who was correctly identified as being at intermediate to high risk by KEYNOTE-564 but was misclassified as being at low risk by the MPRS without tumor necrosis. The analysis revealed that on radiological images, the nnU-Net-based segmentation focused mainly on the kidney and tumor mass and failed to capture the renal vein tumor thrombus located outside the tumor, which is a critical adverse prognostic factor in ccRCC[21] (Fig. 6a). On the histopathological image, the model's attention was concentrated on normal renal tissue rather than tumor regions, resulting in limited utilization of key information in tumor areas (Fig. 6c). The omission of the tumor thrombus and insufficient focus on tumor regions contributed to this false negative result. For the FPR analysis, we selected another representative nonrecurrent patient who was correctly stratified as being at low risk by KEYNOTE-564 but who was misclassified as being at high risk by the MPRS; this patient was a male patient with advanced TNM stage, high tumor grade, and necrosis. On the radiological images, the model focused excessive attention on central necrotic areas and failed to achieve a balance between the necrotic areas and the enhancement of peripheral tumor regions (Fig. 6b). On the histopathological image, the model could not fully leverage the features of necrotic tissue, providing only limited information related to necrosis (Fig. 6d). The overemphasis on necrosis in the radiological information, combined with the inadequate incorporation of histopathological features, led the model to assign excessively high-risk predictions to patients with necrosis, thereby causing the false positive result.

Discussion

In this study, we developed the MPRS, with a focus on both accuracy and clinical utility, using data from 1648 patients who were diagnosed with stage I-III ccRCC across six medical centers in China, as well as data from the TCGA database. The MPRS integrates three modalities: clinical features, radiological CT images, and histopathological WSIs of tumor sections. Notably, external validation in cohorts from other medical centers further underscores the potential generalizability and adaptability of the MPRS for use across different healthcare settings.

Although unimodal models based on clinical features, CT images, and histopathological WSIs have been reported to be useful for predicting ccRCC recurrence[22–24], the potential enhancement of performance due to the use of integrated information from multiple modalities remains underexplored. In a recent study, a multimodal recurrence model for ccRCC was developed by integrating clinical characteristics, single-nucleotide polymorphisms (SNPs), and WSIs. Although the performance of this model was exceptional, its high detection cost may hinder its feasibility for widespread clinical translation[25]. Furthermore, although another multimodal model that combined clinical data, CT images, and WSIs from 414 ccRCC patients achieved satisfactory results, concerns were raised about the potential risk of overfitting because of the limited sample size and lack of multicenter validation[26].

Therefore, we utilized easily accessible multimodal data from a sufficiently large cohort of ccRCC patients to construct the MPRS. Crucially, our design prioritized not only predictive power but also clinical deployability. This focus on practicality extended to our technical implementation: for pathological patch feature extraction, our selection of ResNet-34 aims to balance performance with engineering feasibility. While current state-of-the-art pathological foundation models exhibit powerful feature representation capabilities, these models (e.g., Virchow2[27], MUSK[28], UNI[29]) can have hundreds of millions to billions of parameters, hindering their usage for analysis on edge devices. In contrast, ResNet's lightweight architecture reduces inference costs by orders of magnitude while maintaining clinically viable feature extraction quality (as evidenced by our prognostic validation performance), thereby better aligning with future clinical deployment requirements.

The MPRS integrates information from different modalities through a deep survival network. The multimodal model was trained separately for each modality, effectively utilizing all available data and addressing missing modality data in some patients. Moreover, compared to most late-fusion models that employ a decision-level fusion approach, which applies simple weighting to the outputs of each modality, our model integrates the feature representations of each modality into a deep survival network, enabling a more efficient representation of information from different modalities. This

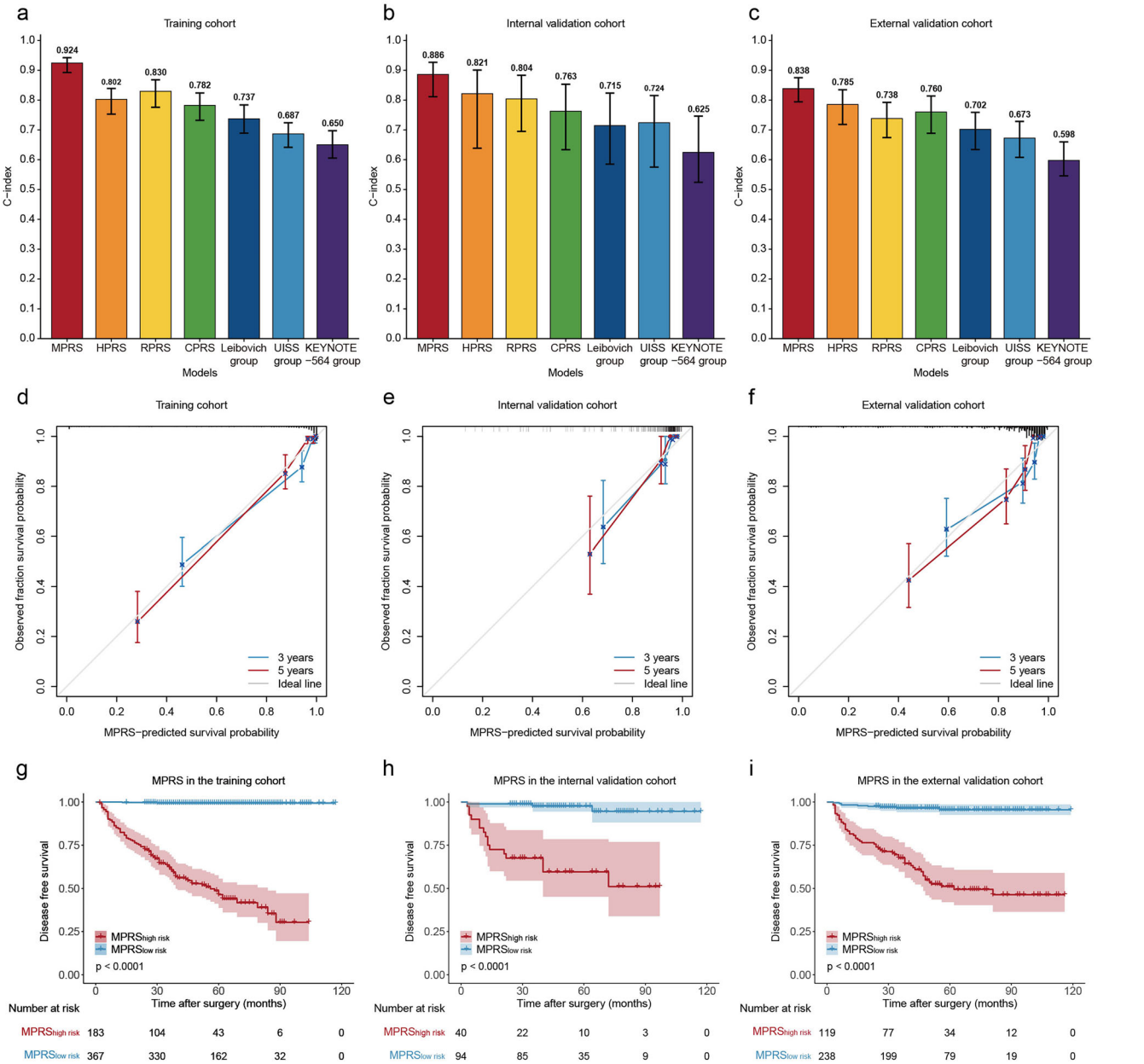


Fig. 2 | Model performance. a–c Bar plots depict the C-Index values for the MPRS, HPRS, RPRS, CPRS, Leibovich score, UISS score, and KEYNOTE-564 risk classification in patients with ccRCC in the training cohort, internal validation cohort, and external validation cohort. The C-index and its associated 95% CI were calculated using the bootstrap method with 1000 resamples. In each plot, the height of the bar represents the C-index value, and the error bars represent the 95% CI. **d–f** Calibration curves for the MPRS on the basis of the concordance between the predicted and observed DFS rates for ccRCC patients at 3 and 5 years. The diagonal reference line represents perfect calibration; deviations above this line indicate underestimation of risk, while deviations below indicate overestimation. The blue and red curves with error bars (95% CI) represent bootstrapped 3-year and 5-year observed (Y-axis) versus predicted probabilities (X-axis), respectively. **g–i** Kaplan–Meier analysis of the MPRS for DFS in patients with ccRCC, divided into low-risk and high-risk groups according to the risk score. p values were calculated with the log-rank test. ROC = receiver operating characteristic. AUC = area under the curve. MPRS = multimodal predictive recurrence score. HPRS = histopathological predictive recurrence score. RPRS = radiological predictive recurrence score. CPRS = clinical predictive recurrence score. KEYNOTE-564 = KEYNOTE-564 group. ccRCC = clear cell renal cell carcinoma. DFS = disease-free survival.

innovative approach enables the MPRS to more effectively manage complexities across different modalities, thereby improving its predictive performance. Compared with unimodal scores (such as the HPRS, RPRS, and CPRS) and clinical prognostic tools (such as the Leibovich score, UISS score, and KEYNOTE-564 risk classification), the MPRS has notable advantages in terms of robustness and generalizability that were further confirmed in our analysis. Furthermore, in contrast to tree-based ensembles, which rely primarily on hierarchical splits along single features or coarse interactions, the deep survival network uses a multilayer perceptron architecture, thereby enabling it to automatically learn complex, subtle, nonlinear feature interactions.

In the clinical decision-making process, significant variations in survival outcomes are often observed among patients in the same risk stratum, indicating that the heterogeneity of ccRCC may not be completely captured by traditional clinical features alone[30,31]. A prospective study based on the ASSURE clinical trial cohort revealed a notable decline in the predictive accuracy of the eight commonly used prognostic models, further highlighting the limitations of current RCC risk assessment methods[32]. We

Table 2 | Univariable and multivariable analyses for MPRS and other clinical factors for disease-free survival in different cohorts

Characteristic	Univariable			Multivariable		
	HR[1]	95% CI[1]	p value	HR[1]	95% CI[1]	p value
Training cohort						
MPRS						
Low risk	Ref.			Ref.		
High risk	132.20	32.53, 537.17	**<0.001**	106.07	25.68, 438.21	**<0.001**
Sex						
Female	Ref.			Ref.		
Male	1.54	1.01, 2.33	**0.043**	0.95	0.59, 1.52	0.832
Age ≥ 60 years						
No	Ref.			Ref.		
Yes	1.00	0.69, 1.44	0.985	0.85	0.56, 1.29	0.442
TNM stage						
I	Ref.			Ref.		
II	5.05	3.16, 8.07	**<0.001**	1.35	0.81, 2.25	0.248
III	7.85	5.14, 11.97	**<0.001**	2.39	1.45, 3.93	**<0.001**
Internal validation cohort						
MPRS						
Low risk	Ref.			Ref.		
High risk	15.94	4.63, 54.85	**<0.001**	13.59	3.63, 50.83	**<0.001**
Sex						
Female	Ref.			Ref.		
Male	1.99	0.58, 6.82	0.276	1.33	0.38, 4.73	0.658
Age ≥ 60 years						
No	Ref.			Ref.		
Yes	1.20	0.48, 2.99	0.690	0.85	0.31, 2.37	0.756
TNM stage						
I	Ref.			Ref.		
II	4.07	1.39, 11.92	**0.011**	1.20	0.38, 3.79	0.759
III	5.25	1.64, 16.82	**0.005**	1.83	0.48, 6.93	0.374
External validation cohort						
MPRS						
Low risk	Ref.			Ref.		
High risk	14.54	7.18, 29.46	**<0.001**	10.70	5.07, 22.58	**<0.001**
Sex						
Female	Ref.			Ref.		
Male	2.69	1.37, 5.28	**0.004**	1.33	0.65, 2.72	0.439
Age ≥ 60 years						
No	Ref.			Ref.		
Yes	2.11	1.25, 3.54	**0.005**	1.55	0.90, 2.66	0.111
TNM stage						
I	Ref.			Ref.		
II	4.12	2.25, 7.57	**<0.001**	1.92	1.03, 3.59	**0.040**
III	5.25	2.80, 9.83	**<0.001**	2.60	1.36, 4.97	**0.004**

[1]*HR* Hazard Ratio, *CI* Confidence Interval

Bold values indicates a significant p value ($p < 0.05$).

observed that the MPRS can effectively differentiate survival outcomes among different patient subgroups in the same stratum on the basis of factors, such as TNM stage, histopathological grade, Leibovich score, UISS score, and KEYNOTE-564 risk classification, demonstrating its potential to further refine risk stratification and enhance prognostic value.

The KEYNOTE-564 clinical trial revealed notable improvements in overall survival (OS) and disease-free survival (DFS) among high-risk ccRCC patients who received adjuvant therapy with pembrolizumab, and a higher incidence of adverse events was observed among patients who were treated with immune checkpoint inhibitors (ICIs) than among those in the placebo group[5,6]. However, controversy about the risk classification in the KEYNOTE-564 trial has arisen, as certain patients with a low risk of recurrence may not benefit from adjuvant therapy owing to tumor heterogeneity[33–35]. Inaccurate risk stratification can be a critical factor that impedes the widespread implementation of adjuvant therapy for ccRCC. Patients who receive excessive adjuvant therapy inappropriately when they are in a disease-free state after surgery may experience negative consequences in terms of their health, rational healthcare resource allocation, and social equity. Conversely, if high-risk patients are unjustly denied the opportunity to access adjuvant therapy, their survival times may be diminished, and the potential clinical benefits of ICIs may not be fully realized.

Therefore, the design of the MPRS holds considerable promise for clinical implementation. With appropriate validation in prospective clinical trials, the MPRS may offer a feasible approach for balancing the risks of overtreatment and undertreatment in adjuvant therapy while assisting clinicians in the personalized risk stratification of ccRCC patients, including patients with risk of recurrence ranging from low to high. On this basis, subsequent treatment decisions can be guided accordingly: for patients who are predicted to be at low risk by the MPRS, a more conservative active surveillance strategy may be adopted; for high-risk patients, adjuvant systemic therapy could be recommended. Since the MPRS integrates three routine and low-cost modalities, namely, clinical features, CT images, and H&E-stained tumor slides, the system not only effectively overcomes the technical challenges associated with molecular testing and expert histopathological review in resource-limited settings but also significantly reduces the associated testing costs.

An error analysis of misclassified cases revealed clinically informative patterns of failure in the performance of the MPRS while clearly highlighting its advantages over established clinical risk stratification tools. The MPRS achieved balanced sensitivity (85.4%) and specificity (78.2%) overall; this balanced accuracy directly enhanced clinical utility: it reclassified 83.3% of KEYNOTE-564 false-negative results to prevent therapeutic inadequacy and 57.7% of false-positive results to avoid unnecessary overtreatment. Further subgroup analyses of the cases that were misclassified by the MPRS revealed specific contexts in which uncertainty existed: false-negative results occurred mainly in recurrent patients without tumor necrosis, suggesting limitations in detecting adverse features independent of necrosis. Conversely, false-positive results occurred more frequently in nonrecurrent patients with established high-risk features, such as male sex, advanced TNM stage (II-III), high grade (3/4), or necrosis, resulting in challenges in refining prognosis for clinically high-risk subgroups. To conduct a more in-depth exploration of these patterns, visual analysis of representative misclassified cases using Grad-CAM revealed the following potential reasons for model failure: false-negative results occurred due to missing extratumoral prognostic features (such as tumor thrombus) and an inadequate histopathological focus on tumor regions, and false-positive results arose due to an overemphasis on adverse radiological signs such as necrosis and insufficient utilization of histopathological features, ultimately leading to imbalanced weighting of multimodal information. These findings highlight the efficacy of the MPRS in addressing a key drawback of current risk stratification; this is the inherent difficulty in maintaining both high sensitivity and high specificity. Although residual errors remain restricted to specific patient subgroups, they provide clear directions for model optimization.

To address these specific failure patterns and further enhance the performance of the MPRS, targeted optimization strategies are proposed. First, enriching the training dataset with more representative samples from the subgroups that were identified as challenging (e.g., cases of recurrence lacking necrosis and cases of nonrecurrence with multiple high-risk

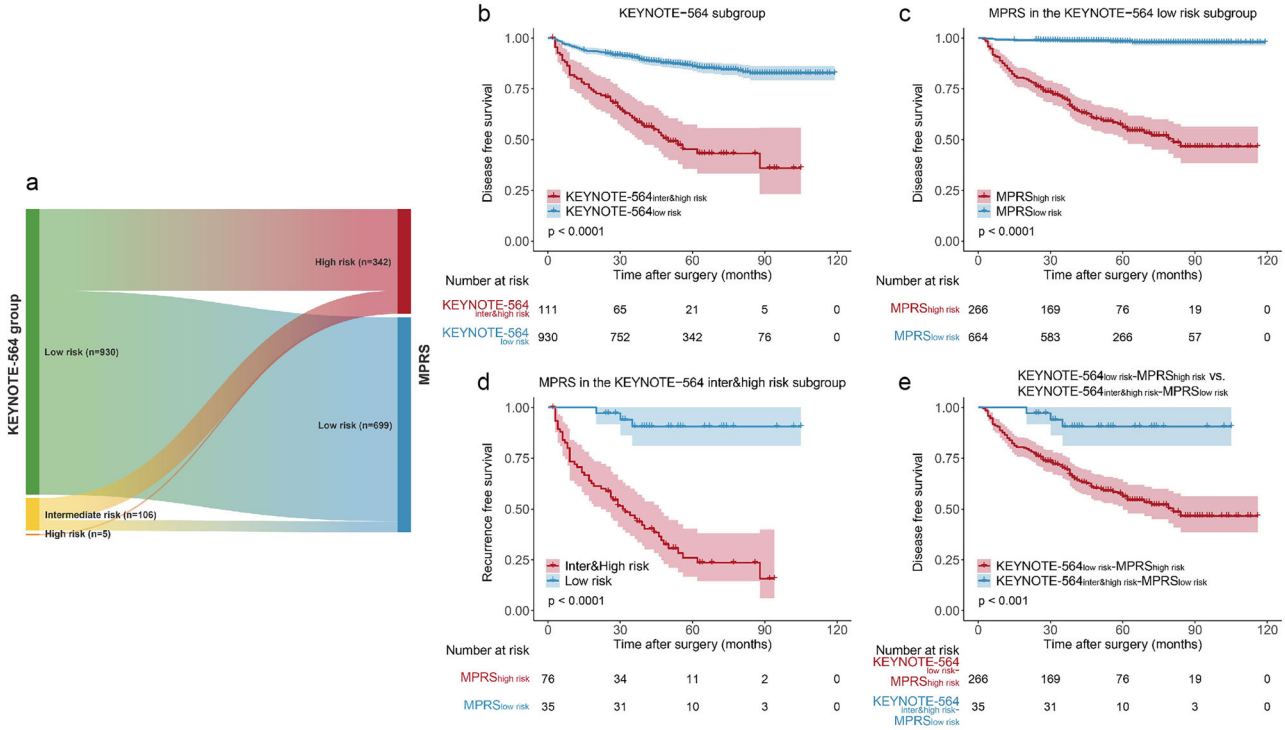


Fig. 3 | Risk stratification analysis for DFS of the MPRS on the basis of the KEYNOTE-564 risk classification subgroups. a Sankey diagram of reclassification from the KEYNOTE-564 risk classification to the MPRS. **b** Kaplan–Meier analysis of the KEYNOTE-564 risk classification for DFS in all enrolled patients. **c**, **d** Kaplan–Meier analysis of the MPRS for DFS of patients in the KEYNOTE-564 low-risk, inter and high-risk groups. **e** Kaplan–Meier analysis for DFS between patients classified as inter&high-risk by the KEYNOTE-564 risk classification but as low-risk by the MPRS, and patients classified as low-risk by the KEYNOTE-564 risk classification but as high-risk by the MPRS. p values were calculated with the log-rank test. MPRS=multimodal predictive recurrence score. DFS = disease-free survival.

features) is crucial to improve the model's learning capacity and generalizability for these complex scenarios. Second, refining the model architecture to better balance multimodal information integration is essential. This could involve implementing more sophisticated attention mechanisms or adaptive feature weighting schemes in the fusion layers, which should be specifically designed to reduce the observed overreliance on adverse radiological signs (such as necrosis) while amplifying the contribution of underutilized histopathological features and critical extra-tumoral prognostic factors (such as tumor thrombus). Third, exploring other advanced techniques, such as graph neural networks, to explicitly model relationships between different tumor regions and prognostic features, or incorporating uncertainty estimation modules to identify predictions requiring clinician review, represents promising technical approaches to reduce residual errors and enhance clinical trust.

Admittedly, this study has several limitations. First, as a retrospective study, potential confounders may not have been adequately controlled, necessitating further validation in prospective cohorts. Second, although we used TCGA data as the training cohort, both the internal and external validation cohorts consisted only of Chinese individuals. The racial homogeneity in this study highlights the need for further validation of the model's generalizability in other regional populations. Third, despite sample size disparities across centers, our MPRS maintained clinically relevant performance (C-index > 0.8), suggesting broad applicability. Future studies should include expanded cohorts at underrepresented centers. Finally, our imaging model relied exclusively on CT images, despite MRI being a widely available imaging method with richer biological information. Future studies will explore the combination of CT and MRI data to enhance the radiological component of our model.

In conclusion, the MPRS model is a promising approach for improving the prediction of the risk of recurrence and the stratification of ccRCC patients through the integration of readily available multimodal data. The MPRS enables clinicians to identify patients who are most likely to benefit from adjuvant therapy, facilitating individualized management and optimizing therapeutic decisions.

Methods

Study participants and patient cohorts

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Renji Hospital (IRB No: LY2024-057-B), the Ethics Committee of the Fourth Affiliated Hospital of Harbin Medical University (IRB No: 2025-ER-60), the Ethics Committee of the Second Affiliated Hospital of Soochow University (IRB No: JD-HG-2025-101), the Clinical Research Ethics Committee of the First Affiliated Hospital of Anhui Medical University (IRB No: PJ 2024-11-96), the Human Trial Ethics Committee of Changgung Medical (IRB No: 202402309B0), and the Ethics Committee of the First Affiliated Hospital of Shandong First Medical University (IRB No: YXLL-KY-2024-175). Given its retrospective and observational nature, the need for informed consent was waived. The study is registered on ClinicalTrials.gov (NCT06656039).

We retrospectively collected multimodal data from 1145 patients who underwent surgery and received postoperative histopathological diagnoses of stage I–III ccRCC between January 1, 2014, and December 31, 2021, across six centers in China. The inclusion criteria included comprehensive clinical features, accessible preoperative arterial-phase contrast-enhanced CT images, available postoperative hematoxylin and eosin (HE)-stained sections of tumor tissues, and complete follow-up information. The exclusion criteria were the presence of other malignancies and treatment with neoadjuvant or adjuvant therapies.

The primary outcome measure was DFS, which was defined as the time from surgery to the first detection of ccRCC recurrence (including local recurrence and distant metastases) or until the end of the follow-up period. Detailed scoring criteria and application methods for clinical prognostic

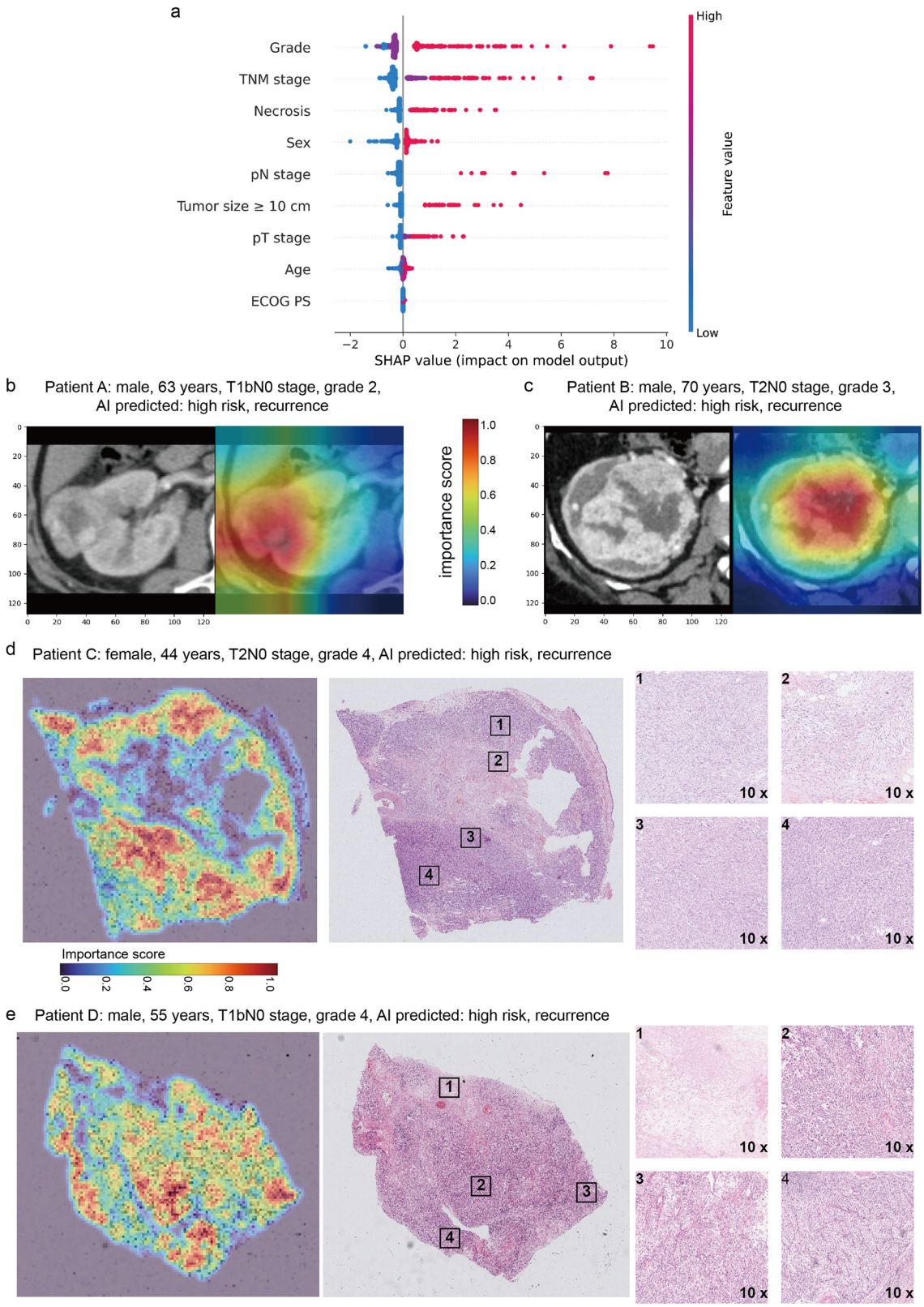


Fig. 4 | Interpretability analysis of the model. a SHAP values for clinical features that drive the prediction of recurrence in the model for ccRCC. **b–e** Grad-CAM analysis of images obtained with the radiological and histopathological modalities. The clinical information of representative patients is included above the figure: sex, age, TNM stage, tumor grade, risk group defined by our model, and clinical outcome (recurrence status). On the basis of the importance scores, the deep red areas suggest greater contributions to the prediction of tumor recurrence, and the deep blue areas indicate lower contributions. Radiological modality: High-attention regions (deep red) localized predominantly in the irregular border area of the tumor (**b**) and the heterogeneous area of enhancement with a rich blood supply and necrosis (**c**). Histopathological modality: The model focused more on sarcomatoid differentiated tissue (Regions 1 and 3, deep red) than on well-differentiated tissue (Region 2, light blue) and high-grade tissue (Region 4, yellow–green) (**d**); comparatively, it placed greater emphasis on poorly differentiated, higher-grade tissue (Regions 2 and 3, deep red) than on stroma-rich tissue (Region 1, deep blue) and well-differentiated, lower-grade tissue (Region 4, yellow–green) (**e**).

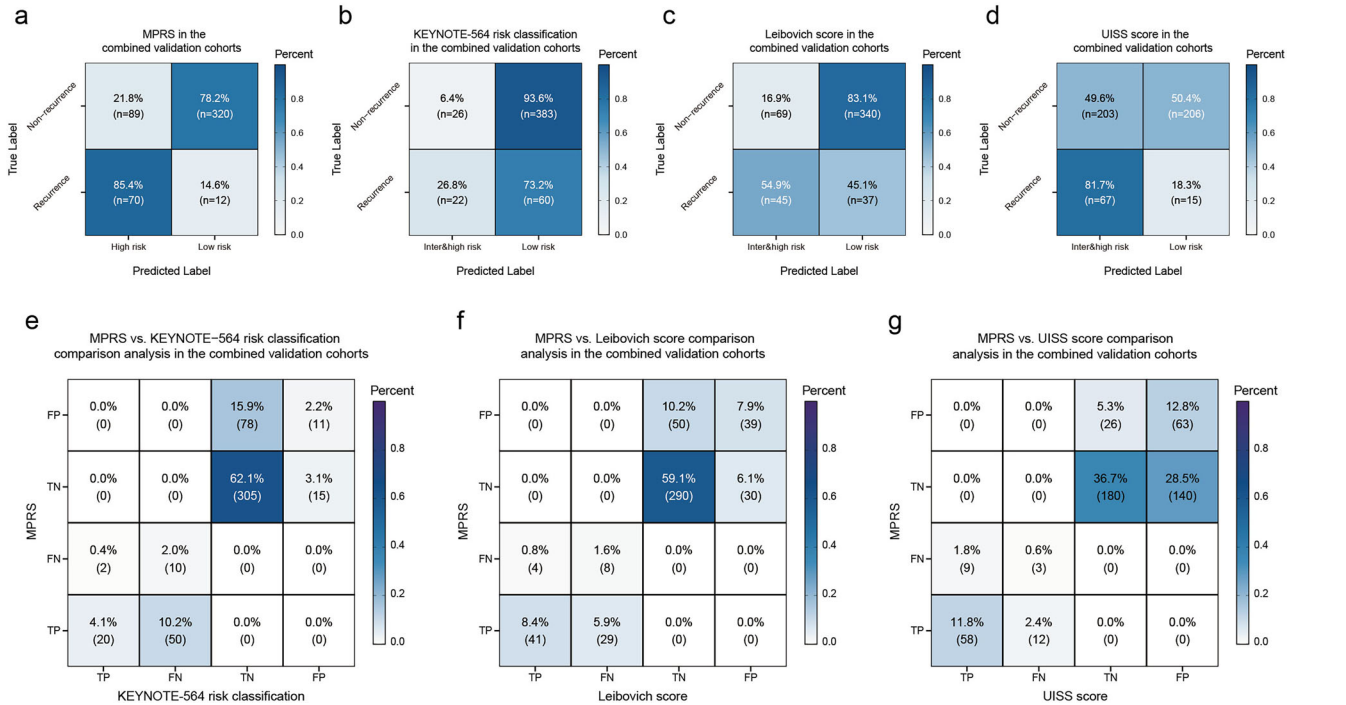


Fig. 5 | Comparative analyses of MPRS and clinical risk stratification tools in the combined internal and external validation cohorts. a–d Confusion matrices for MPRS, KEYNOTE-564 risk classification, Leibovich score, and UISS score in prediction of recurrence risk stratification. **e**, **f** Heatmaps for comparative analysis of recurrence risk prediction between MPRS and KEYNOTE-564 risk classification, Leibovich score, and UISS score. TP = true positive. FN = false negative. TN = true negative. FP = false positive.

tools including the Leibovich score, UISS score, and KEYNOTE-564 risk classification were presented (Tables S1–S3). This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting and Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) guidelines[36,37] (Tables S5–S6).

Image preprocessing

For H&E-stained WSIS, we used Otsu's thresholding method[38] to extract the foreground containing histopathological tissue and cut it into patches measuring 768 × 768 in size at 40× magnification (0.25 μm/pixel). Patches with a foreground over 30% were retained. For the CT sequences, we extracted the arterial phase images from the multiphase contrast-enhanced series and resampled them to a resolution of 0.8 × 0.8 × 1 mm. For region of interest extraction, we used nnU-Net, which is a self-adaptive encoder-decoder model based on U-Net, to automatically segment the tumor regions from CT images. The model load weights were pretrained on the public kidney segmentation dataset, which we downloaded from https://zenodo.org/record/3734294#.YvXiWnZBw2z. Then, all the automatically segmented masks were reviewed and refined by two senior urologists (W.K. and Y.C.) specializing in genitourinary oncology images (14 and 24 years of experience, respectively). Any discrepancies were resolved through consensus.

Model development

We propose a multimodal prognostic prediction model that integrates information from various modalities through a deep survival network to assess the risk of recurrence in patients with renal clear cell carcinoma. For the histopathological modality, we concatenated the features from ResNet34 pretrained on ImageNet with the average nuclear features extracted by HistomicsTK, as well as the dominant cell type in the patch, to form the patch information. We then used multi-instance learning to aggregate the patch information to present histopathological features. For the radiological modality, we first extracted features by means of PyRadiomics and ImageNet-pretrained ResNet34 and then used principal component analysis (PCA) to retain components that achieved an explanation ratio of 0.8 to serve as the radiological feature representation. Finally, we integrated the histopathological, radiological, and clinical information by means of the deepSurv network[39] to generate a multimodal prognostic risk score. The deep learning model was implemented via PyTorch 1.13.1 in Python. Heatmaps were generated by means of gradient-weighted class activation mapping (Grad-CAM)[40] to visualize the decision process of the developed model. The following sections introduce each part in detail.

HPRS construction

After histopathological image preprocessing, we obtained 768×768-pixel patches at 40× magnification from all the slides. Afterward, all the patches underwent automated nucleus segmentation using HoverNet. HoverNet can segment and classify cells into five types according to their nuclei: neoplastic, inflammatory, soft tissue, dead, and epithelial cells. Patches containing <10 nuclei were discarded as noninformative, and those with ≥10 nuclei were retained for downstream feature extraction. This ensures unbiased sampling across all tissues, enabling our models to learn from spatially diverse microenvironmental contexts while also balancing computational efficiency. Afterward, for all the retained patches, the histomic features, cell type features, and ResNet-based embeddings were extracted.

HistomicsTK was used to extract nuclear features from individual cells. HistomicsTK can provide nuclear morphology features, including shape, texture, and intensity characteristics, for nuclei, enabling quantitative analysis of cellular structures. The histomic features of a patch were then computed as the average of the features from the most abundant cell type in that patch (e.g., if a patch was dominated by tumor cells, the average nuclear features of all tumor cells were used). Additionally, cell type features were represented as one-hot encoding vectors (e.g., (1, 0, 0, 0, 0) for a tumor-rich patch).

Table 3 | Analysis of false negative and false positive rates of MPRS across clinical characteristic subgroups

Characteristic	Recurrence	FNR(number)	p value	Non-recurrence	FPR(number)	p value
Sex			0.092			**0.002**
Female	13	30.8% (4)		138	13.0% (18)	
Male	69	11.6% (8)		271	26.2% (71)	
Age ≥ 65 years			0.341			0.189
No	30	20.0% (6)		214	19.2% (41)	
Yes	52	11.5% (6)		195	24.6% (48)	
ECOG PS ≥ 1			0.357			0.516
No	43	18.6% (8)		287	20.9% (60)	
Yes	39	10.3% (4)		122	23.8% (29)	
Tumor size ≥ 10 cm			0.556			0.218
No	77	14.3% (11)		408	21.6% (88)	
Yes	5	20.0% (1)		1	100.0% (1)	
pT stage			0.472			**<0.001**
T1a	3	33.3% (1)		110	5.5% (6)	
T1b	42	16.7% (7)		242	21.5% (52)	
T2	20	15.0% (3)		31	64.5% (20)	
T3	17	5.9% (1)		26	42.3% (11)	
pN stage			1			
N0	80	15.0% (12)		409	21.8% (89)	
N1	2	0.0% (0)		0	NA% (0)	
TNM stage			0.491			**<0.001**
I	44	18.2% (8)		352	16.5% (58)	
II	20	15.0% (3)		31	64.5% (20)	
III	18	5.6% (1)		26	42.3% (11)	
Grade			0.923			**<0.001**
1	1	0.0% (0)		50	14.0% (7)	
2	34	17.6% (6)		272	17.3% (47)	
3	37	13.5% (5)		80	36.2% (29)	
4	10	10.0% (1)		7	85.7% (6)	
Necrosis			**0.001**			**<0.001**
No	48	25.0% (12)		380	18.2% (69)	
Yes	34	0.0% (0)		29	69.0% (20)	
Sarcomatoid RCC			1			**0.034**
No	80	15.0% (12)		405	21.2% (86)	
Yes	2	0.0% (0)		4	75.0% (3)	

Fisher's exact test was used to calculate statistical differences between subgroups.
Bold values indicate a significant p value ($p < 0.05$).

In addition, we adopted a ResNet-34 model pretrained on ImageNet to extract the morphological embeddings of the previously retained 768 × 768-pixel patches at 40× magnification. These ResNet-based features are expected to represent the overall tissue organization and spatial relationships within each patch. The final feature vector for each patch was constructed as the concatenation of the histomicsTK features, the cell type encodings, and the ResNet-based embeddings.

After the patch features were constructed, attention-based multi-instance learning was applied to aggregate them into 256-dimensional slide-level histopathological features. Attention-based multi-instance learning[41] can learn the most relevant instance within all patches that contribute to prognosis prediction. For the histopathological unimodal model that calculates the HPRS, a multilayer perceptron layer was used to derive the risk score on the basis of the aggregated 256-dimensional slide-level histopathological features.

RPRS construction

For the radiological model, we first performed PCA dimensionality reduction on the shape, first-order, texture, filter, and wavelet features and the embeddings extracted from ResNet in PyRadiomics. The components for which the explanation ratio was greater than 0.8 were retained. The principal component numbers for the shape, first-order, texture, filter, wavelet, and ResNet features were 2, 3, 3, 6, 10, and 75, respectively. We concatenated these principal components to obtain the radiological feature representation. For the radiological unimodal model that calculates the RPRS, these features were input as covariates into the Cox proportional hazards model for prognostic modeling.

CPRS construction

With respect to the clinical information, all the features were encoded as multiclass, ordinal, or continuous variables. Afterward, for the unimodal

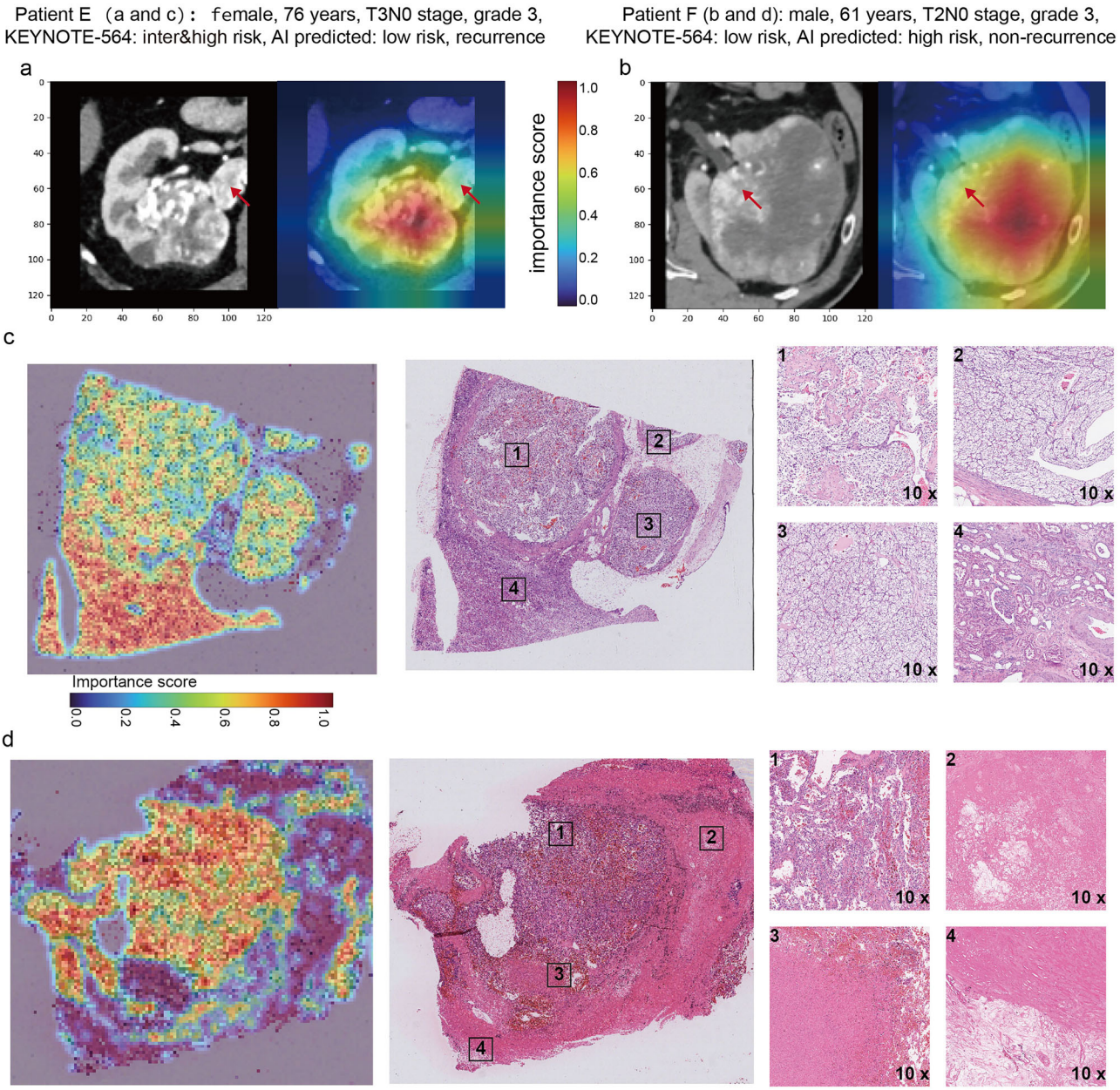


Fig. 6 | Analysis of representative cases of misclassification. Grad-CAM analysis of images obtained with radiological and histopathological modalities. The clinical information of representative patients is included above the figure: gender, age, TNM stage, tumor grade, risk group defined by our model, and clinical outcome (recurrence status). On the basis of the importance scores, the deep red areas suggest greater contributions to the prediction of tumor recurrence, and the deep blue areas indicate lower contributions. **a** Radiological modality: High-attention regions (deep red) localize primarily in the tumor mass, while areas with renal vein tumor thrombus (yellow–green) receive insufficient attention, indicating that the model failed to focus on capturing the tumor thrombus area. **b** Radiological modality: Central necrotic areas (deep red) dominate attention maps, with peripheral enhancing tumor regions (yellow–green) being overlooked, revealing an inadequate model balance between necrosis and peripheral enhancing tumor regions. **c** Histopathological modality: Our model focused more on normal renal tissue (Region 4, deep red) than on tumor tissue (Regions 1-3, yellow–green). **d** Histopathological modality: Our model primarily identified the tumor tissue region (Region 1, deep red) but missed the necrotic tissue areas (Regions 2–4, deep blue).

clinical model that calculates the CPRS, we used the Cox proportional hazards model to analyze the clinical information for the multivariate survival analysis.

MPRS construction

To construct a multimodal prediction model for MPRS calculation, the histopathological, radiological, and clinical embeddings were concatenated as 365-dimensional features, comprising 256-dimensional histopathological features, 99-dimensional radiological features, and 10-dimensional clinical features. The 365-dimensional features were processed by means of the deepSurv network[39] to output the MPRS. The deep survival network that we employed extends the traditional Cox model, which uses linear modeling for the covariates. Specifically, the hazard of patient X can be represented as follows:

$$\lambda(t|X) = \lambda_0(t) \cdot e^{h(x)}$$

where $\lambda_0(t)$ is the baseline hazard. In the traditional Cox model, $h(x) = \beta^T x$, whereas for deepSurv, h represents learnable multilayer perceptron layers.

Additionally, three classical machine learning models adapted for survival analysis were employed for comparison: RSF, GBS, and SSVM.

Statistical analysis

For baseline characteristics, age and disease-free survival (DFS) are presented as medians (quartiles) because of their nonnormal distributions as continuous variables, with group differences assessed by means of the Kruskal–Wallis H test. Categorical variables are reported as frequencies and percentages, and group differences were evaluated by means of Pearson's chi-square test or Fisher's exact test, as appropriate.

Model performance was evaluated by means of the C-index and time-dependent receiver operating characteristic (ROC) curves and the area under the curve (AUC) at 3 years and 5 years, and the numbers of patients at risk for time-dependent ROC analysis were provided (Table S4). Calibration curves were generated in R (using the rms and survival packages) to assess the agreement between the predicted and observed survival probabilities in the three cohorts. Cox proportional hazards models predicting 3-year and 5-year survival based on the MPRS risk score were fitted. Calibration was performed using 1000 bootstrap resamples, stratifying patients into quintiles (five risk groups). In each risk group, observed event probabilities were estimated via the Kaplan–Meier method to account for right-censoring and compared against model-predicted survival probabilities after bootstrap bias correction.

To transform the continuous MPRS output into clinically actionable high-risk and low-risk groups, we identified an optimal cutoff point using time-dependent ROC analysis (implemented via the timeROC package). This analysis evaluated the performance of the model for predicting DFS at both the 3-year endpoint and the 5-year endpoint. Candidate cutoffs were determined at each time point by maximizing the Youden index (sensitivity + specificity - 1). Owing to substantially higher rates of censoring at the 5-year endpoint than at the 3-year endpoint, which introduces greater uncertainty in true event status assignment for cutoff estimation, we prioritized the candidate cutoff derived from the 3-year analysis for final clinical stratification. Patients with MPRS values ≥ the 3-year Youden-derived cutoff (-1.955) were classified as being at high risk, while those below this value were classified as being at low risk.

Survival analysis was performed with the Kaplan–Meier method, with significant differences evaluated by means of the log-rank test. All the statistical analyses were conducted with R software (version 4.22). A two-sided p value < 0.05 was considered to indicate statistical significance.

Data availability

After signing a data transfer agreement and obtaining ethical approval. Upon reasonable request to the corresponding author (Wei Zhai, jacky_zw2002@hotmail.com), the data will be made available to interested research partners.

Code availability

The source code is available online (https://github.com/jackyzw2002/RenalxAI-prognosis/tree/master).

Received: 5 March 2025; Accepted: 24 September 2025;
Published online: 17 November 2025

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **74**, 229–263 (2024).
2. Moch, H. et al. The 2022 world health organization classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *Eur. Urol.* **82**, 458–468 (2022).
3. Motzer, R. J. et al. NCCN guidelines® insights: kidney cancer, version 2.2024: featured updates to the NCCN guidelines. *J. Natl. Compr. Cancer Netw.* **22**, 4–16 (2024).
4. Haas D. N. B. Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): a double-blind, placebo-controlled, randomised, phase 3 trial.387. 2016;
5. Choueiri, T. K. et al. Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. *N. Engl. J. Med.* **385**, 683–694 (2021).
6. Choueiri, T. K. et al. Overall survival with adjuvant pembrolizumab in renal-cell carcinoma. *N. Engl. J. Med.* **390**, 1359–1371 (2024).
7. Leibovich, B. C. et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer* **97**, 1663–1671 (2003).
8. Zisman, A. et al. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. *JCO* **20**, 4559–4566 (2002).
9. Rini, B. et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: development and validation studies. *Lancet Oncol.* **16**, 676–685 (2015).
10. Wei, J. H. et al. Predictive value of single-nucleotide polymorphism signature for recurrence in localised renal cell carcinoma: a retrospective analysis and multicentre validation study. *Lancet Oncol.* **20**, 591–600 (2019).
11. Cao, K. et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat. Med* **29**, 3033–3043 (2023).
12. Sirinukunwattana, K. et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* **70**, 544–554 (2021).
13. Lee, Y. et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nat Biomed Eng*. Published online August 18, https://doi.org/10.1038/s41551-022-00923-0.2022.
14. Gao, P. et al. Interpretable multi-modal artificial intelligence model for predicting gastric cancer response to neoadjuvant chemotherapy. *Cell Rep. Med.* **5**, 101848 (2024).
15. Vanguri, R. S. et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat. Cancer* **3**, 1151–1164 (2022).
16. Boehm, K. M. et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **3**, 723–733 (2022).
17. Zhu, P., et al. Tumor contour irregularity on preoperative CT predicts prognosis in renal cell carcinoma: a multi-institutional study. *eClinicalMedicine* **75**, 102775 (2024).
18. Pallwein-Prettner, L. et al. Assessment and characterisation of common renal masses with CT and MRI. Published online 2011.
19. Lebacle, C. et al. Epidemiology, biology and treatment of sarcomatoid RCC: current state of the art. *World J. Urol.* **37**, 115–123 (2019).
20. Mouallem, N. E., Smith, S. C. & Paul, A. K. Sarcomatoid renal cell carcinoma: biology and treatment advances. *Urol. Oncol.: Semin. Original Investig.* **36**, 265–271 (2018).
21. Kirkali, Z. & Van Poppel, H. A critical analysis of surgery for kidney cancer with vena cava invasion. *Eur. Urol.* **52**, 658–662 (2007).
22. Khene, Z. E. et al. Application of Machine Learning Models to Predict Recurrence After Surgical Resection of Nonmetastatic Renal Cell Carcinoma. *European Urology Oncology*. Published online August 2022: S2588931122001377. https://doi.org/10.1016/j.euo.2022.07.007.
23. Nie, P. A CT-based deep learning radiomics nomogram outperforms the existing prognostic models for outcome prediction in clear cell renal cell carcinoma: a multicenter study. *European Radiology*. Published online 2023.
24. Chen, S. et al. Machine learning-based pathomics signature could act as a novel prognostic marker for patients with clear cell renal cell carcinoma. *Br. J. Cancer* **126**, 771–777 (2022).
25. Gui, C. P. et al. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. *Lancet Digit. Health* **5**, e515–e524 (2023).

26. 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. *International Journal of Surgery*. Published online March 4, 2024. https://doi.org/10.1097/JS9.0000000000001222.
27. Vorontsov, E. et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat. Med.* **30**, 2924–2935 (2024).
28. Xiang, J. et al. A vision-language foundation model for precision oncology. *Nature* **638**, 769–778 (2025).
29. Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med.* **30**, 850–862 (2024).
30. Martínez-Salamanca, J. I. et al. Prognostic impact of the 2009 UICC/AJCC TNM staging system for renal cell carcinoma with venous extension. *Eur. Urol.* **59**, 120–127 (2011).
31. Amin, M. B. et al. The Eighth Edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA CANCER J CLIN*. 67.2017;
32. Correa, A. F. et al. Predicting renal cancer recurrence: defining limitations of existing prognostic models with prospective trial-based validation. *JCO* **37**, 2062–2071 (2019).
33. Khene, Z. E., Bex, A. & Bensalah, K. Adjuvant therapy after surgical resection of nonmetastatic renal cell carcinoma: one size does not fit all. *Eur. Urol.* **81**, 432–433 (2022).
34. Naito, H. et al. Postoperative recurrence factors in patients with pT3N0M0 clear cell renal cell carcinoma, called M0 intermediate-high-risk group of the KEYNOTE 564 trial. Preprint at https://doi.org/10.21203/rs.3.rs-2905509/v1 (2023).
35. Liu, Z. et al. Re: Overall Survival with Adjuvant Pembrolizumab in Renal-cell Carcinoma. *Eur. Urol.* **86**, 482–483 (2024).
36. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, g7594 (2015).
37. Altman, D. G., McShane, L. M., Sauerbrei, W. & Taube, S. E. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med.* **10**, 51 (2012).
38. Otsu, N. Threshold Selection Method From Gray-level Histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
39. Katzman, J. L. et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18**, 24 (2018).
40. Selvaraju, R. R. et al. In *Proceedings of the IEEE international conference on computer vision.* 618–626.
41. Ilse, M., Tomczak, J. & Welling, M. In *International conference on machine learning.* 2127–2136 (PMLR).

Acknowledgements

This study was supported by National Natural Science Foundation of China (82173214, 82473386); National Key Research and Development Program of China (2022YFC2505305); National Major Science and Technology Infrastructure for Translational Medicine (Shanghai) 2024 "Open Research Program" Key Projects (TMSK-2024-119); Shanghai Municipal Education Commission's Artificial Intelligence-Powered Research Paradigm Reform and Discipline Advancement Program (BJ1-9000-25-8010); Shanghai 2023 "Science and Technology Innovation Action Plan" Key Project on Computational Biology (23JS1400803); Basic-Clinical Collaborative Innovation Project from Shanghai Immune Therapy Institute; Shanghai Science and Technology Commission Medical Innovation Clinical Research Special Project (23Y21900400); Excellent Project of Shanghai Municipal Health Commission (20244Z0004); Shanghai Key Program of Computational Biology (23JS1400800); Innovative Research Team of High-level Local Universities in Shanghai (SHSMUZLCX20212601); Simulation RCT Study of Shanghai Shenkang Three Year Action Plan (SHDC2024CRI042); Shanghai Jiao Tong University Medical Engineering Interdisciplinary Research Fund - Key Project (YG2024ZD07); Renji Hospital Clinical Research Incubation Fund (RJPY-DZX-004). We acknowledge the TCGA database for providing the data. Workflow of the study (Figure 1) was created with Figdraw.

Author contributions

X.Z. (Xinyi Zang), H.L., K.W., C.L., J.Z. (Junhua Zheng), Z.Y., and W.Z. were responsible for the design and conception of the study. Y.X. and Z.Y. constructed models. X.Z. (Xinyi Zang), Y.X. and H.X. finished writing the original draft. M.S., N.H., A.H., L.X., J.Z. (Jin Zhu), Z.X., and J.W. (Jianning Wang) accessed and verified the data. T.C., Z.L., X.P., and X.Z. (Xiangyu Zi) performed a literature review. Z.W., J.X., D.C., and Y.Y. were consultants with expertise in pathology. J.W. (Jieying Wang) was a consultant with expertise in statistics. W.K. and Y.C. reviewed and refined all automatically segmented masks. X.W., J.H., J.Z. (Jin Zhang), Y.H., D.K.W. LEUNG, and J.Y.C. TEOH were responsible for reviewing and editing papers. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-02034-x.

Correspondence and requests for materials should be addressed to Keliang Wang, Chaozhao Liang, Junhua Zheng, Zhangsheng Yu or Wei Zhai.

Reprints and permissions information is available at http://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