OXFORD

Artificial intelligence-based non-invasive tumor segmentation, grade stratification and prognosis prediction for clear-cell renal-cell carcinoma

Siteng Chen[1,§], Dandan Song[2,§], Lei Chen[3,§], Tuanjie Guo[4,§], Beibei Jiang[5], Aie Liu[3], Xianpan Pan[3], Tao Wang[4], Heting Tang[4], Guihua Chen[4], Zhong Xue[3], Xiang Wang[4,*], Ning Zhang[6,*] and Junhua Zheng[1,*]

[1]Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200001, China
[2]Department of Radiology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China
[3]Shanghai United Imaging Intelligence Co., Ltd., Shanghai 201807, China
[4]Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China
[5]Department of Radiology, Erasmus University Medical Center, Rotterdam, P.O. Box 2040, 3000 CA, The Netherlands
[6]Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
*Correspondence: Junhua Zheng, zhengjh0471@sina.com; Ning Zhang, zn12235@rjh.com.cn; Xiang Wang, seanw_hs@163.com
§Siteng Chen, Dandan Song, Lei Chen, and Tuanjie Guo contributed equally to this work.

Abstract

Due to the complicated histopathological characteristics of clear-cell renal-cell carcinoma (ccRCC), non-invasive prognosis before operative treatment is crucial in selecting the appropriate treatment. A total of 126 345 computerized tomography (CT) images from four independent patient cohorts were included for analysis in this study. We propose a V Bottleneck multi-resolution and focus-organ network (VB-MrFo-Net) using a cascade framework for deep learning analysis. The VB-MrFo-Net achieved better performance than VB-Net in tumor segmentation, with a Dice score of 0.87. The nuclear-grade prediction model performed best in the logistic regression classifier, with area under curve values from 0.782 to 0.746. Survival analysis revealed that our prediction model could significantly distinguish patients with high survival risk, with a hazard ratio (HR) of 2.49 [95% confidence interval (CI): 1.13–5.45, $P = 0.023$] in the General cohort. Excellent performance had also been verified in the Cancer Genome Atlas cohort, the Clinical Proteomic Tumor Analysis Consortium cohort, and the Kidney Tumor Segmentation Challenge cohort, with HRs of 2.77 (95%CI: 1.58–4.84, $P = 0.0019$), 3.83 (95%CI: 1.22–11.96, $P = 0.029$), and 2.80 (95%CI: 1.05–7.47, $P = 0.025$), respectively. In conclusion, we propose a novel VB-MrFo-Net for the renal tumor segmentation and automatic diagnosis of ccRCC. The risk stratification model could accurately distinguish patients with high tumor grade and high survival risk based on non-invasive CT images before surgical treatments, which could provide practical advice for deciding treatment options.

Keywords: deep learning, diagnosis, non-invasive, prognosis, prediction, clear-cell renal-cell carcinoma

Introduction

It is estimated that there were 76 080 new cases and 13 780 associated deaths from renal cancer in the USA in 2021.[1] Clinically, clear-cell renal-cell carcinoma (ccRCC) accounts for most of the malignant lesions in the renal system,[2] which generally progresses more aggressively compared with other subtypes. Contrast-enhanced abdominal computed tomography (CT) is recommended for diagnosis of ccRCC,[3] however, it is still challenging to accurately identify ccRCC through visual inspection of CT images, especially distinguishing them from fat poor angiomyolipoma and oncocytoma.[4,5] In addition, the clinical outcome of ccRCC varies from different tumor stages, and patients with tolerable neoplasm suffer from low cure rates even after radical nephrectomy,[6,7] which indicates the importance of accurate tumor characterization and prognosis prediction for treatment planning of ccRCC.

Deep learning (DL), also known as deep neural network, is an artificial intelligence (AI) technique used to mimic the data processing and decision making of the human brain. Unlike other machine learning algorithms that extract image features before learning, DL uses convolutional neural networks (CNNs) to generate features for classification through back-propagation of loss functions based on an optimization algorithm for network learning.[8] Currently, DL has exhibited excellent practicability in medical image processing, especially for unsupervised learning from unstructured image data.[9] Some preliminary studies have confirmed the effectiveness of DL in kidney segmentation and tumor classification based on image data from a single center.[10,11] For example, a multi-scale graph was constructed with nodes representing multi-scale topologies across an organ or tumor regions for volumetric image segmentation.[12] However, there are very few studies focusing on the risk stratification and survival prediction of ccRCC based on CNN from multi-source patient cohorts.

Here, in this study, we developed a CNN-based V-Net algorithm for lesion segmentation of ccRCC from noninvasive CT images. Instead of constructing a complicated graph to learn multi-scale topologies,[12] a multi-resolution strategy is performed by using

Table 1. Clinical characteristics of patients from three independent patient cohorts.

	General cohort (472)	TCGA cohort (184)	CPTAC cohort (41)	KiST cohort (141)	P value	Total (838)
Age (years)					0.099	
≥65	157 (33.3%)	63 (34.2%)	20 (48.8%)	58 (41.1%)		298 (35.6%)
<65	315 (66.7%)	121 (65.8%)	21 (51.2%)	83 (58.9%)		540 (64.4%)
Sex					0.034	
Male	351 (74.4%)	120 (65.2%)	26 (63.4%)	92 (65.2%)		589 (70.3%)
Female	121 (25.6%)	64 (34.8%)	15 (36.6%)	49 (34.8%)		249 (29.7%)
Grade					0.071	
G1	72 (15.3%)	1 (0.6%)	3 (7.3%)	17 (12.1%)		93 (11.1%)
G2	317 (67.2%)	74 (40.2%)	24 (58.5%)	79 (56.0%)		494 (59.0%)
G3	73 (15.4%)	83 (45.1%)	12 (29.3%)	35 (24.8%)		203 (24.2%)
G4	10 (2.1%)	26 (14.1%)	2 (4.9%)	10 (7.1%)		48 (5.7%)
Stage					0.102	
i	433 (91.7%)	95 (51.6%)	19 (46.3%)	96 (68.1%)		643 (76.7%)
ii	29 (6.2%)	17 (9.3%)	5 (12.2%)	5 (3.6%)		56 (6.7%)
iii	10 (2.1%)	46 (25.0%)	13 (31.7%)	25 (17.7%)		94 (11.2%)
iv	0	26 (14.1%)	4 (9.8%)	15 (10.6%)		45 (5.4%)
T stage						
T1	434 (91.9%)	98 (53.3%)	21 (51.2%)	100 (70.9%)	0.090	653 (77.9%)
T2	29 (6.2%)	20 (10.9%)	6 (14.6%)	6 (4.3%)		61 (7.3%)
T3	9 (1.9%)	63 (34.2%)	13 (31.7%)	34 (24.1%)		119 (14.2%)
T4	0	3 (1.6%)	1 (2.5%)	1 (0.7%)		5 (0.6%)
N stage						
N1	3 (0.6%)	4 (2.2%)	2 (4.9%)	5 (3.5%)	0.298	14 (1.7%)
N0	469 (99.4%)	81 (44.0%)	3 (7.3%)	54 (38.3%)		607 (72.4%)
Unknown	0	99 (53.8%)	36 (87.8%)	82 (58.2%)		217 (25.9%)
M stage					0.056	
M1	0	24 (13.0%)	2 (4.9%)	14 (9.9%)		40 (4.8%)
M0	472 (100%)	154 (83.7%)	5 (12.2%)	58 (41.1%)		689 (82.2%)
Unknown	0	6 (3.3%)	34 (82.9%)	69 (49.0%)		109 (13.0%)
Status					0.064	
Dead or with tumor	25 (5.3%)	135 (73.4%)	12 (29.3%)	16 (88.7%)		188 (22.4%)
Alive or without tumor	261 (55.3%)	49 (26.6%)	29 (70.7%)	125 (11.3%)		464 (55.4%)
Unknown	186 (39.4%)	0	0	0		186 (22.2%)

a V Bottleneck-Net (VB-Net) of coarse-resolution and a VB-Net of fine-resolution for kidney and tumor segmentation. We further constructed and verified DL-based models for preoperative Furhman grade classification and prognosis prediction of patients with ccRCC in multi-source patient cohorts. During experiments, ablation study results show the effectiveness of DL segmentation and classification models.

Materials and methods
Patient cohorts and data sources
Four independent patient cohorts were included for analysis from different institutions, including Shanghai General Hospital, Clinical Proteomic Tumor Analysis Consortium (CPTAC),[13,14] the Cancer Genome Atlas (TCGA),[13] and the 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS).[13,15] The inclusion criteria are: (i) pathologically confirmed ccRCC without other malignancies; (ii) complete clinicopathological data; and (iii) access to the original data of corresponding contrast-enhanced CT (corticomedullary phase) images covering the whole malignant tumor. Ethical approval of our study has been obtained from the Research Ethics Committee of Shanghai General Hospital.

Altogether, 472 patients, who underwent partial or radical nephrectomy and were pathologically diagnosed as ccRCC from January 2012 to May 2019 in Shanghai General Hospital, were recruited as the General cohort. In addition, 184 patients from TCGA, 141 patients from KiTS, and 41 patients from CPTAC, who met the inclusion criteria mentioned above, were also included. The clinical characteristics of all 838 patients in this study are shown in Table 1.

CT acquisition
For the General cohort, patients underwent contrast-enhanced abdomen CT scanning after injection of about 70 ml of contrast medium into the antecubital vein at about 3.5 ml/s. Siemens Somatom Force or GE Healthcare were used for CT acquisition, and the slice thickness ranges from 0.6 to 0.625 mm. The helical acquisition mode was adopted with rotation time of 500 ms and pitch of 1.5. Collimation widths were 256 × 0.625 mm or 64 × 0.625 mm for the respective scanners. The voltage, current, and rotation time of the tube were set as 120 kVp, ~150 mAa, and ~0.45 s, respectively. For patients from CPTAC, TCGA, and KiTS, their radiology images were recovered from the Cancer Imaging Archive in DICOM format, which were collected from the standard of care imaging before the pathological diagnosis.

Data pre-processing
For all CT images collected in this study, renal tumors were manually delineated as regions of interest (ROIs) by two senior urologists (>20-years experience). All manual delineations were performed using ITK-SNAP (version 3.8).[16] The images and ROIs
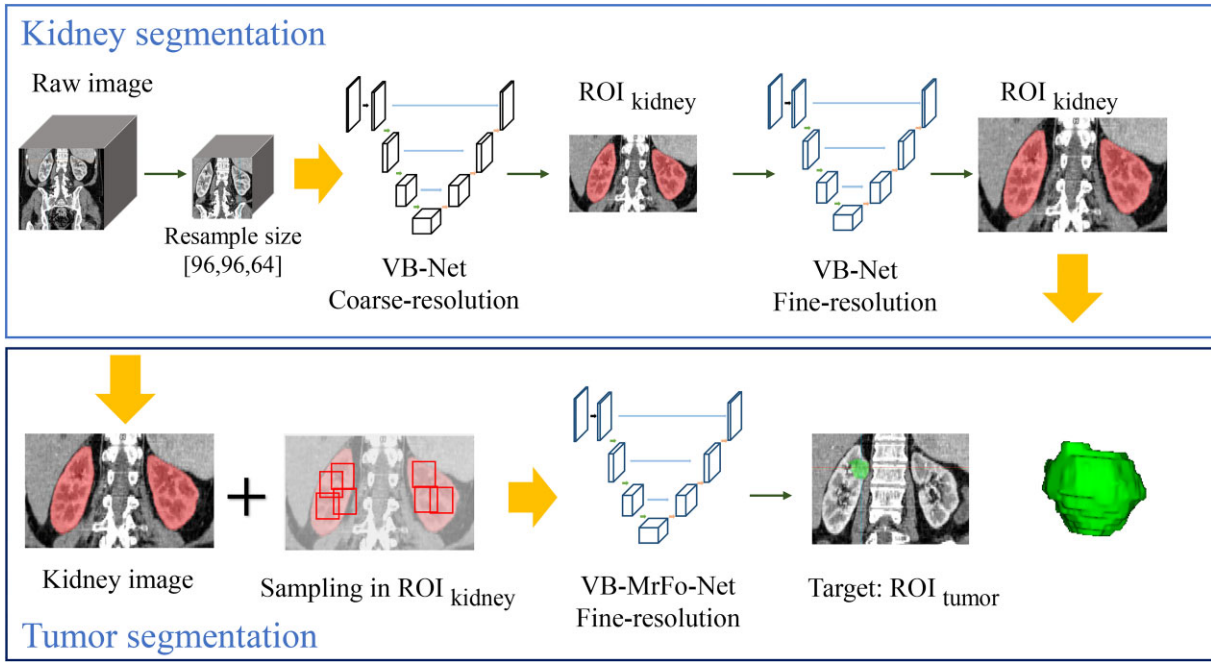
Figure 1. Architecture of VB-MrFo-Net. Top: multi-resolution kidney segmentation, and bottom: focus-organ renal tumor segmentation.

were used to train and evaluate our AI segmentation model. A flowchart of the experiment is shown in Fig. 1.

All CT images were first resampled using tri-linear interpolation to the same spatial resolution: $0.7 \times 0.7 \times 2.5$ (mm³). Then, the intensities of each volume were truncated to the Hounsfield unite (HU) range of [−200, 500] to remove the irrelevant details and linearly normalized into the range of [−1,1]. Considering the GPU memory, the input 3D patch size was set to $96 \times 96 \times 64$ (voxel). Due to the limited amount of training images, we applied image augmentation on the training dataset including shifting, rotation, and scaling.

Cascade multi-resolution and focus-organ VB-nets

In our previous work,[17] we used VB-Net to replace the conventional convolutional layers in V-Net[18] based on the bottleneck structure. In this study, we further improve VB-Net and develop a multi-resolution and focus-organ framework (VB-MrFo-Net) using a cascade framework (Fig. 1). The bottleneck structure consists of three convolutional layers with a small model size, which makes it easier to deploy to either cloud or mobile applications than the V-Net structure. As 3D medical images are often large in size, passing the whole 3D image volume into a network will consume large GPU memory. Therefore, we use a multi-resolution strategy, and the coarse-resolution is used to roughly localize the boundary of the whole kidney at lower resolution $2.5 \times 2.5 \times 2.5$ (mm³), while the kidney boundary is accurately delineated at high resolution $0.7 \times 0.7 \times 0.7$ (mm³). After renal segmentation, we utilize a focus-organ strategy to segment the tumor within the kidney region. Since a tumor may grow across the edge of the kidney, the mask of kidney is enlarged by dilating using morphological operation, and the filter size is $7 \times 7 \times 7$ (mm³). For renal tumor segmentation, we utilized a focus-organ strategy to segment the tumor within the segmented kidney region. Since the tumor may grow across the edge of the kidney, we further enlarged the bound-

ary by dilating the mask of kidney with a filter size of $7 \times 7 \times 7$ (mm³). The model size of VB-Net is similar to VB-MrFo-Net, 8.8 MB. As the sampling region within the segmented kidney is much smaller than the whole CT image, it is easier to train VB-MrFo-Net than VB-Net in steps of optimization and convergence. The parameters of VB-MrFo-Net were, step learning rate schedule was initialized as 1e-4, and the Adam optimizer was set as momentum = 0.9, decay = 1e-4, betas = (0.9, 0.999). Data augmentation methods include rotating, scaling, flipping, shifting, and adding noise. The raining procedure includes 2501 epochs; global sampling was set for the coarse model, and mask sampling for the fine model.

In experiments, we adopted a generalized Dice loss function to only focus on the target (foreground voxels) in the whole image. The loss function was defined based on the Dice coefficient (range of [0, 1]) as follows,

$$\text{Dice} = \frac{2\sum_i^N p_i * g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2},$$

where N is the number of voxels, p_i is the prediction result of the i-th voxel belonging to the target region, g_i denotes whether the i-th voxel belongs to the ground truth (delineated target) or not (1 means yes, 0 means no).

Prediction model

After tumor segmentation, radiomics analysis was further performed using the uAI Research Portal (United Imaging Intelligence, China), which is a clinical research platform implemented by Python programming language (version 3.7.3). The pre-processing of feature extraction includes CT image resampling, intensity normalization, and feature normalization. Specifically, the CT images were resampled to the same spatial resolution, $0.7 \times 0.7 \times 2.5$ (mm³), and the intensity values of each image were linearly normalized into the range [−1,1]. The extracted features from ROI were normalized by the Z-score method before

(A)

(B)
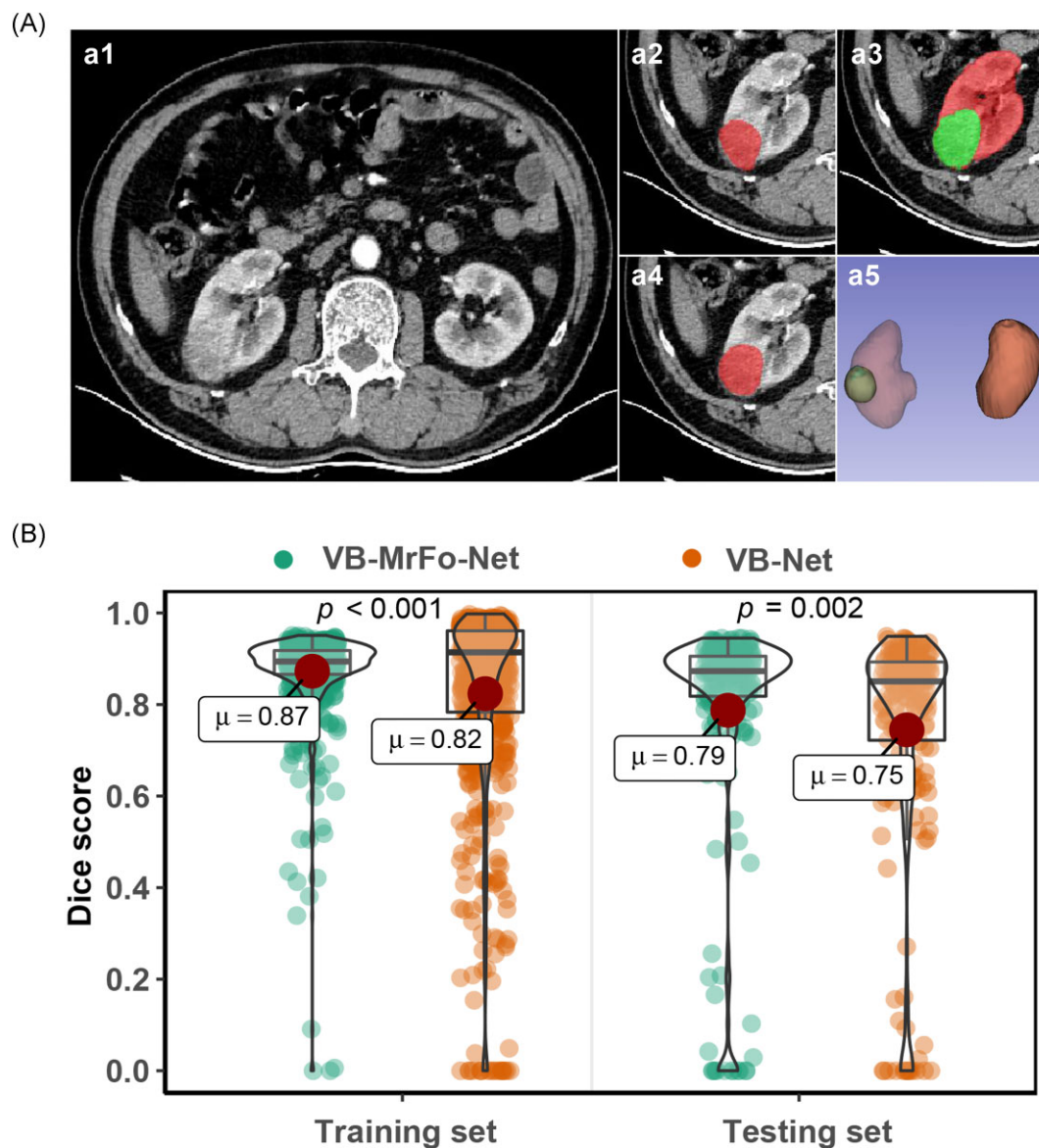
Figure 2. Segmentation of ccRCC by AI models in CT images (60 year old male). (**A**) A representative segmentation from a 60-year male patient, including original CT image (a1); manual delineation (a2), segmented ROIs by VB-Net (red indicates kidney and green indicates tumor) (a3), segmented ROIs by VB-MrFo-Net (a4), and 3D visualization for segmentation of VB-MrFo-Net (a5). (**B**) Comparison of Dice between VB-MrFo-Net and VB-Net in the training and the testing datasets.

feature selection. The radiomics features were computed from ROIs based on PyRadiomics.[19–21] A total of 2600 features were computed from the tumor ROI in each CT image, including 25 imaging filters and 104 original features grouped as: 18 first-order statistics, 14 shape, 21 texture including Gray-Level Co-occurrence Matrix (glcm), 16 Gray-Level Run-Length Matrix (glrlm), 16 Gray-Level Size-Zone Matrix (glszm), 14 Gray-Level Dependence Matrix (gldm), and 5 Neighboring Gray-Tone Difference Matrix (ngtdm). The least absolute shrinkage and selection operator regression (LASSO) was further applied to obtain an optimal feature subset. The pre-processing includes normalization of Z-score, and the alpha value for LASSO was set to 0.0152. Finally, we built three machine-learning classifiers, i.e. logistic regression

(LR), random forest (RF), and support vector machine (SVM), based on the radiomics features selected. Five-fold cross-validation was applied for validation.

Statistical analysis

R 3.6.2 was used for data analyses and visualization in this study. Continuous variants were analyzed by the Mann–Whitney U test. Quantitative variants were compared through the Chi-square test or the Fisher exact method. Survival analysis was performed using a Kaplan–Meier (KM) curve with hazard ratio (HR) and 95% confidence interval (CI) for comparing different overall survival (OS) or disease-free survival (DFS) outcomes. We also carried out receiver operating characteristic curve (ROC) analysis and calculated area
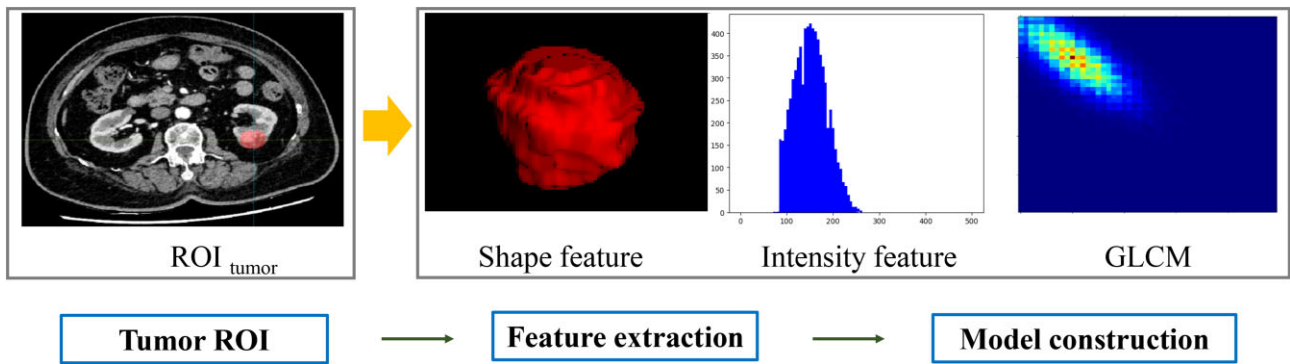
Figure 3. Flowchart for tumor grade prediction model.

Table 2. The performance of tumor grade predicting models constructed by LR, SVM, and RF by a 5-fold cross-validation method.

Performance		Manual labeled ROIs			AI segmented ROIs		
		LR	SVM	RF	LR	SVM	RF
Training	Sensitivity	0.721	0.732	0.817	0.716	0.789	0.742
	Specificity	0.699	0.737	0.757	0.606	0.495	0.629
	Accuracy	0.705	0.736	0.775	0.639	0.583	0.662
	AUC	0.782	0.799	0.866	0.730	0.727	0.756
	(95% CI)	(0.737–0.827)	(0.756–0.841)	(0.833–0.900)	(0.680–0.781)	(0.676–0.778)	(0.710–0.803)
Testing	Sensitivity	0.726	0.71	0.661	0.726	0.823	0.613
	Specificity	0.659	0.63	0.754	0.616	0.464	0.616
	Accuracy	0.68	0.655	0.725	0.65	0.575	0.615
	AUC	0.746	0.717	0.729	0.721	0.719	0.703
	(95% CI)	(0.667–0.825)	(0.632–0.801)	(0.645–0.813)	(0.643–0.799)	(0.641–0.798)	(0.621–0.786)

under curve (AUC) to evaluate the accuracy of the tumor grade prediction. P value < 0.05 was set as the significance level.

Results

Patient cohorts

The clinical characteristics of all 838 patients recruited are shown in Table 1. The subjects were randomly divided into a training set (638 patients) and a testing set (200 patients) (Fig. S1, see online supplementary material).

Tumor segmentation based on VB-MrFo-Net

Kidney was automatically segmented by VB-Net. Dice score was used for evaluation of the segmentation accuracy. Based on the KiTS19 dataset, the mean Dice score for the kidney segmentation model was 0.974, ranging from 0.921 to 0.990.[17] Based on the VB-MrFo-Net, we trained a tumor segmentation model for ccRCC using the training set. A paradigm of tumor segmentation results is shown in Fig. 2A, the segmented ROI by VB-MrFo-Net (Fig. 2A(a4)) was close to ground truth (Fig. 2A(a2)), while VB-Net was prone to over-segmentation (Fig. 2A(a3)). Compared with manual annotation (ground truth) of the tumor region, the proposed VB-MrFo-Net achieved better performance than VB-Net (Fig. 2B), with Dice score of 0.87 vs 0.82 ($P < 0.001$) and 0.79 vs 0.75 ($P = 0.002$) in the training set and testing test, respectively. Similarly, the Dice score of VB-MrFo-Net was higher than that of VB-Net based on the public dataset and the General cohort, respectively (Fig. S2, see online supplementary material).

Prediction of tumor grade based on manual annotations and AI segmentation

The construction of the grade prediction model is shown in Fig. 3. Firstly, we trained the prediction model based on manual annotations. A total of 22 features were selected by LASSO and then applied to build the prediction model. A prediction model based on AI segmentations was also developed and verified in the same way. The performance of the prediction model is shown in Table 2, including accuracy, specificity, sensitivity, and AUC. It was found that the best predicting model was the LR classifier. For manually labeled ROIs, the AUC values were 0.782 and 0.746 in the training and testing set, respectively. For AI segmented ROIs, LR achieved AUC values of 0.730 and 0.721 in the training and testing set, respectively.

Automatic survival stratification model for patients with ccRCC

Since the tumor grade has been reported to be significantly associated with clinical prognosis of patients with ccRCC, we next explored whether our AI-based prediction model could be used for automatic survival stratification for patients with ccRCC before surgical treatment. The LR classifier model was used for risk stratification due to its superior performance for tumor grading. By performing KM curve survival analysis, our prediction model could significantly distinguish patients with high survival risk, with HR of 2.49 (95%CI: 1.13–5.45, $P = 0.023$) in the General cohort (Fig. 4B). Excellent survival risk stratification performance had also been verified in the TCGA cohort, the CPTAC cohort, and the KiTS cohort, with HRs of 2.77 (95%CI: 1.58–4.84, $P = 0.0019$,

(A)

(B)

(C)

(D)
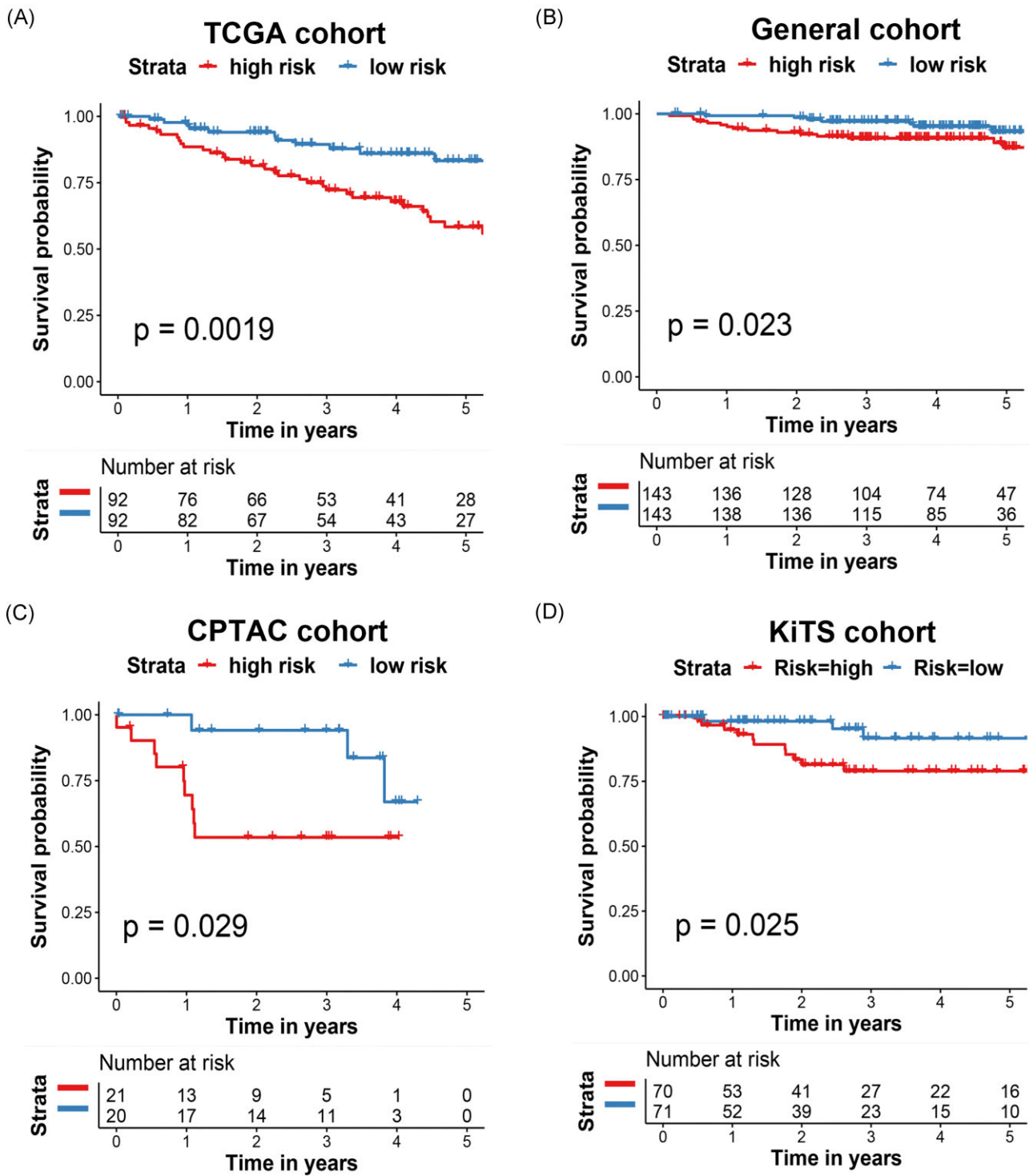
Figure 4. Automatic survival risk stratification model for patients with ccRCC. (**A**, **D**) Kaplan–Meier survival analysis of overall survival stratified by automatic survival risk for ccRCC in the TCGA cohort and the KiTS cohort, respectively; (**B**, **C**) Kaplan–Meier survival analysis of disease-free survival stratified by automatic survival risk for ccRCC in the General cohort and the CPTAC cohort, respectively.

Fig. 4A), 3.83 (95%CI: 1.22–11.96, P = 0.029, Fig. 4C), 2.80 (95%CI: 1.05–7.47, P = 0.062, Fig. 4D), respectively.

Cox regression analysis confirmed that our automatic survival risk stratification (AI-SRS) could act as an independent prognostic factor for patients with ccRCC (Table 3). Further decision analysis revealed that our noninvasive AI-SRS bore comparison with the current tumor grading system in OS (Fig. 5A) and DFS (Fig. 5B)

prediction, while the latter required pathological images and professional pathologists.

Biological associations for the AI-SRS model

We also explored the potential biological associations behind the models. As shown in Fig. 6A, correlation analyses were performed between 20 selected radiologic features and OS status or

Table 3. Cox: regression analyses of AI-SRS in four independent cohorts.

General cohort	HR (95% CI)	P	CPTAC cohort	HR (95% CI)	P
Age	1.05 (1.01–1.09)	0.014		0.99 (0.94–1.05)	0.846
Sex	3.99 (0.94–16.97)	0.061		1.08 (0.32–3.58)	0.904
Stage	2.76 (1.70–4.48)	<0.0001		5.48 (1.89–15.88)	0.0017
Grade	6.31 (3.6–11.05)	<0.0001		1.34 (0.59–3.01)	0.484
AI-SRS	127.4 (19.7–824.0)	<0.0001		35.69 (2.13–598.4)	0.0129
TCGA cohort	HR (95% CI)	P	KiTS cohort	HR (95% CI)	P
Age	1.04 (1.01–1.06)	0.0028		1.02 (0.98–1.07)	0.276
Sex	0.71 (0.40–1.25)	0.234		1.40 (0.48–4.08)	0.54
Stage	1.93 (1.51–2.48)	<0.0001		2.36 (1.54–3.61)	<0.0001
Grade	2.24 (1.49–3.36)	<0.0001		3.55 (1.88–6.68)	<0.0001
AI-SRS	7.65 (2.48–23.63)	<0.0001		13.76 (2.11–89.64)	0.0061
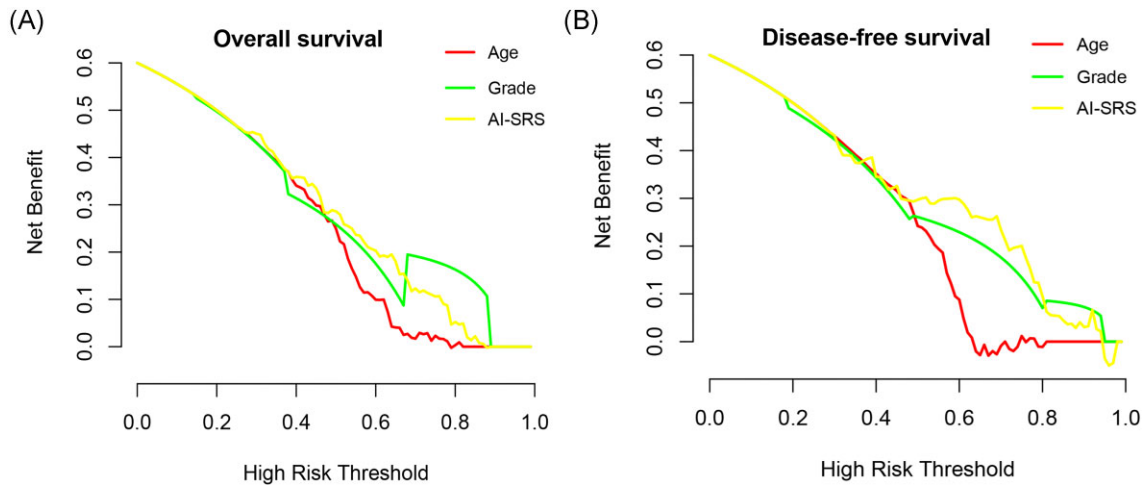
Figure 5. Decision curve analyses for the automatic survival risk stratification model. (**A**) Decision curve analyses for overall survival; (**B**) decision curve analyses for disease-free survival.

DFS status, respectively. When the threshold value of correlation coefficients was set to 0.25, feature 4, the 3D Max-diameter (3DMD, Fig. 6B) of lesion in CT imaging, was found to be the most important parameter for the AI-SRS model. Further visual assessment revealed that 3DMD was associated with higher tumor grade (Fig. 6C). More interestingly, 3DMD was up-regulated in patients with tumor lymphatic metastasis (Fig. 6D) or tumor distant metastasis (Fig. 6E), which could act as a noninvasive and excellent marker for predicting the status of tumor lymphatic metastasis and tumor distant metastasis before operative treatments, with AUCs of 82.4 and 87.3% (Fig. 6F), respectively.

Discussion

Traditionally, image reading is done by visual inspections of the shape, size, edge characteristics, and gray-scale values of object of interest in CT images. For renal carcinoma with high heterogeneity, large amounts of important information are missing from manual observations of CT images. Currently, radiomics based on machine learning in CT images has been gradually used in prediction of prognosis and treatment response.[22,23] However, these manually formulated characters from CT images might be nonredundant due to the incomplete information unscrambled from original image data.[24]

The application of AI technology in image processing has largely affected traditional medical imaging diagnosis and tumor risk stratification.[25,26] Benefiting from the increasing availability of large datasets of digital images, radiologists are now cooperating more with biomedical engineering scientists and exploring the potential of a new branch in the field of radiology. Using CNN and recurrent neural network, noninvasive tracking of the tumor phenotype for predicting pathologic response may contribute to the progress of personalized therapy.[27] The Google TensorFlow had also been used to distinguish ccRCC from oncocytoma on multiphasic CT images.[28]

The manual segmentation of kidney and renal tumor is time-consuming and inconsistent between different doctors. Therefore, a VB-Net[17] was proposed to segment kidney and tumor in contrast-enhanced CT, which was modified from V-Net[18] and was initially used to segment the prostate by training an end-to-end fully convolutional network on magnetic resonance imaging. Here, we created a novel VB-MrFo-Net, which consists of a multi-resolution part and a focus-organ part. The Dice value of the VB-MrFo-Net model achieved 0.787 in automatic segmentation ccRCC in test contrast-enhanced CT imaging, which is higher than traditional VB-Net (0.745). In several cases the regions of tumors did not segment well compared to the ground truth in this study. There were four segmentation failure cases where the tumors, which were large and severe, grew outside the kidney. Three segmentation failure cases occurred because the tumor size was small and the boundary was not very well-defined. In addition, one failure case was because only one kidney (surgical excision)
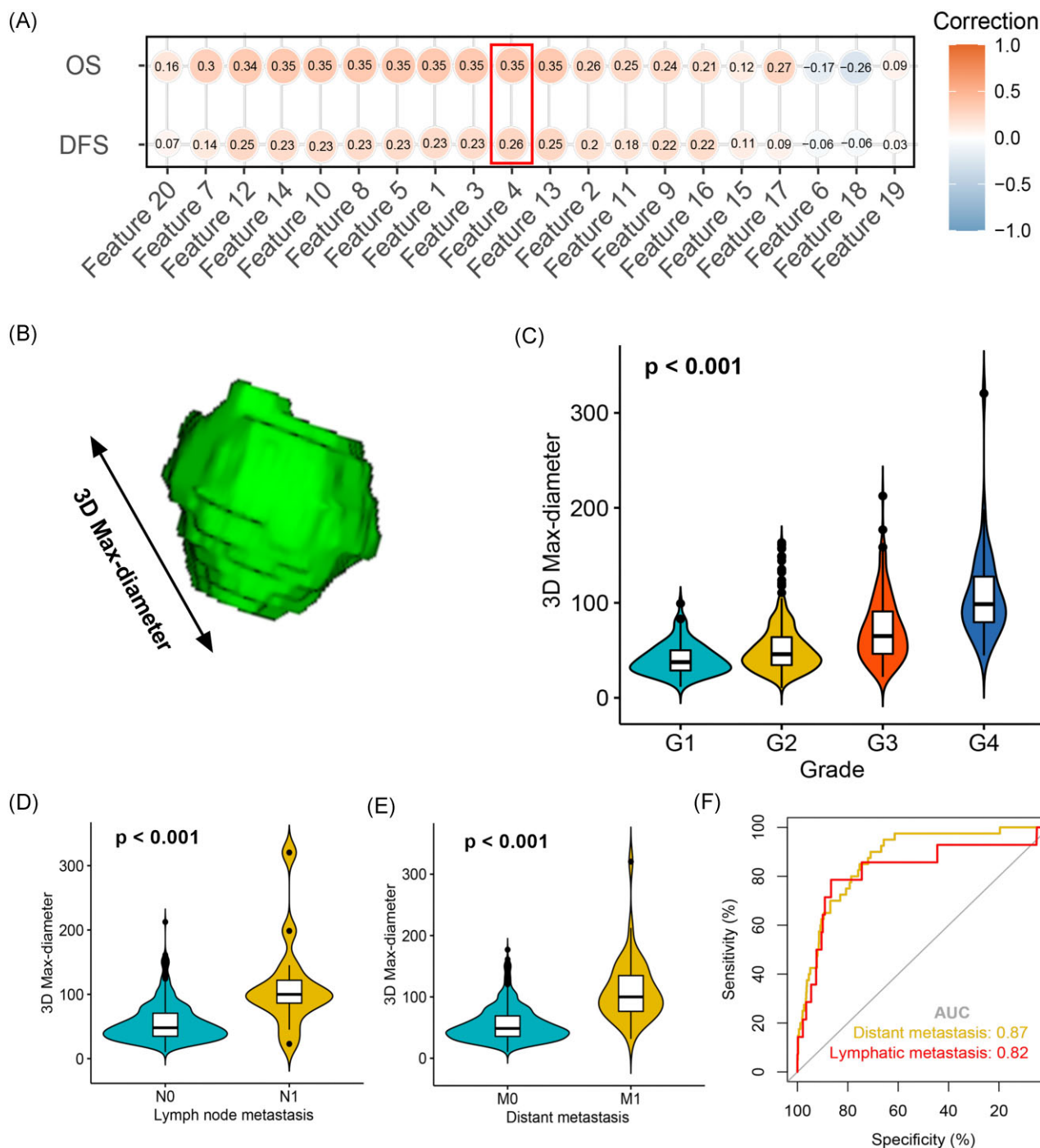
Figure 6. Biological associations for the AI-SRS model. (**A**) Correlation analyses of selected radiologic features and OS status or DFS status; (**B**) 3D visualization for segmentation of 3D Max-diameter; (**C**) comparison of 3D Max-diameter among patients with different tumor grades; (**D**, **E**) comparison of 3D Max-diameter among patients with tumor lymphatic metastasis and tumor distant metastasis, respectively. (**F**) Receiver operating characteristic curves for the prediction of lymphatic metastasis status and tumor distant metastasis status.

existed in the CT image, the other image failed to segment due to noise.

Semantic segmentation of kidney and renal tumor is crucial for quantitatively characterizing lesions. Based on a DL algorithm, a 3D U-Net was trained for automated detection of renal cyst and exhibited excellent practicability in tumor detection.[29] A CNN framework was also proposed to classify solid, lipid-poor, and contrast-enhancing renal masses in multi-phase contrast-enhanced CT images.[30] The overall accuracy of the CNN-based

classifier is 0.78, and the results illustrated what could be trained for the accurate classification of malignant renal lesions. However, most of these studies are based on small sample sizes from a single center, without effective external validation.

In this study, by applying multi-institutional patient cohorts, we developed and verified an automatic risk stratification model based on CNN. Our radiomics model could accurately distinguish patients with high tumor grade, who might suffer from poor clinical outcomes. This is of great significance since there is still a

requirement for effective methods of decision making for radical or partial nephrectomy for patients with ccRCC. Our study could provide importance noninvasive and practical advice before operative treatment of ccRCC.

We also note some limitations of this study. First, part of the images recruited for analysis were acquired from public databases, which might be associated with potential bias due to the inevitably unequal information available. Furthermore, this is a retrospective study, which might need further prospective validation in randomized controlled trials. Additionally, considering different CT scanning protocols among different patient cohorts, the homogeneity of CT images might be debased due to potential heterogeneity, which might affect the analysis.

Conclusions

In conclusion, we proposed a novel VB-MrFo-Net for renal tumor segmentation and analysis, and applied radiomics and machine learning algorithms for automatic diagnosis of ccRCC. The risk stratification model could accurately distinguish patients with high tumor grade and high survival risk based on noninvasive CT images before surgical treatments, which could provide practical advice for deciding treatment options. However, prospective validation in randomized controlled trials is still necessary for further application of our AI-based models in clinical practice.

Supplementary data

Supplementary data is available at *PCMEDI* online.

Acknowledgments

Author contributions

S.T.C.: data curation, methodology, validation, writing—original draft. D.D.S.: visualization, validation, writing—original draft. L.C.: visualization, software, writing—original draft. T.J.G.: data curation, validation, writing—original draft. B.B.J.: data curation. A.E.L.: methodology. X.P.P.: data curation. T.W.: data curation. H.T.T.: methodology. G.H.C.: methodology. Z.X.: software. X.W.: conceptualization, writing—review and editing. N.Z.: conceptualization, writing—review and editing. J.H.Z.: supervision, writing—review and editing. All authors were involved in manuscript editing and manuscript review.

Data availability

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

Conflict of Interest

Lei Chen, Aie Liu, Xianpan Pan, and Zhong Xue are employees of Shanghai United Imaging Intelligence Co., Ltd.

References

1. Siegel RL, Miller KD, Fuchs HE, *et al.* Cancer statistics, 2021. *CA Cancer J Clin* 2021;**71**:7–33. doi:10.3322/caac.21654.

2. Humphrey PA, Moch H, Cubilla AL, *et al.* The 2016 WHO classification of tumours of the urinary system and male genital organs-part B: Prostate and bladder tumours. *Eur Urol* 2016;**70**:106–19. doi:10.1016/j.eururo.2016.02.028.

3. Ljungberg B, Bensalah K, Canfield S, *et al.* EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol* 2015;**67**:913–24. doi:10.1016/j.eururo.2015.01.005.

4. Choudhary S, Rajesh A, Mayer NJ, *et al.* Renal oncocytoma: CT features cannot reliably distinguish oncocytoma from other renal neoplasms. *Clin Radiol* 2009;**64**:517–22. doi:10.1016/j.crad.2008.12.011.

5. Hindman N, Ngo L, Genega EM, *et al.* Angiomyolipoma with minimal fat: Can it be differentiated from clear cell renal cell carcinoma by using standard MR techniques? *Radiology* 2012;**265**:468–77. doi:10.1148/radiol.12112087.

6. Liu L, Qi L, Li Y, *et al.* Retroperitoneoscopic partial nephrectomy for moderately complex ventral hilar tumors: Surgical technique and trifecta outcomes from a single institution in China. *J Laparoendosc Adv Surg Tech A* 2017;**27**:812–7. doi:10.1089/lap.2016.0194.

7. Yi X, Xiao Q, Zeng F, *et al.* Computed tomography radiomics for predicting pathological grade of renal cell carcinoma. *Front Oncol* 2020;**10**:570396. doi:10.3389/fonc.2020.570396.

8. Gulshan V, Peng L, Coram M, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**:2402–10. doi:10.1001/jama.2016.17216.

9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–44. doi:10.1038/nature14539.

10. Han S, Hwang SI, Lee HJ. The classification of renal cancer in 3-phase CT images using a deep learning method. *J Digit Imaging* 2019;**32**:638–43. doi:10.1007/s10278-019-00230-2.

11. Tanaka T, Huang Y, Marukawa Y, *et al.* Differentiation of small (≤ 4 cm) renal masses on multiphase contrast-enhanced CT by Deep learning. *AJR Am J Roentgenol* 2020;**214**:605–12. doi:10.2214/AJR.19.22074.

12. Xuan P, Bi H, Cui H, *et al.* Graph based multi-scale neighboring topology deep learning for kidney and tumor segmentation. *Phys Med Biol* 2022;**67**:225018. doi:10.1088/1361-6560/ac9e3f.

13. Clark K, Vendt B, Smith K, *et al.* The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* 2013;**26**:1045–57. doi:10.1007/s10278-013-9622-7.

14. Clark DJ, Dhanasekaran SM, Petralia F, *et al.* Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* 2019;**179**:964–83. doi:10.1016/j.cell.2019.10.007.

15. Heller N, Isensee F, Maier-Hein KH, *et al.* The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *Med Image Anal* 2021;**67**:101821. doi:10.1016/j.media.2020.101821.

16. Yushkevich PA, Piven J, Hazlett HC, *et al.* User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 2006;**31**:1116–28. doi:10.1016/j.neuroimage.2006.01.015.

17. Gao Y, Ma Y, Mu G, *et al.* Automatic MR kidney segmentation for autosomal dominant polycystic kidney disease. In: *Computer-Aided Diagnosis*, 2019. doi:10.1117/12.2512372.

18. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation.

In: *2016 Fourth International Conference on 3D Vision (3DV)*, 2016. doi:10.1109/3DV.2016.79.

19. Wang W, Gu D, Wei J, *et al*. A radiomics-based biomarker for cytokeratin 19 status of hepatocellular carcinoma with gadoxetic acid-enhanced MRI. *Eur Radiol* 2020;**30**:3004–14. doi:10.1007/s00330-019-06585-y.

20. Ji GW, Zhu FP, Xu Q, *et al*. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* 2019;**50**:156–65. doi:10.1016/j.ebiom.2019.10.057.

21. Kim J, Choi SJ, Lee SH, *et al*. Predicting survival using pretreatment CT for patients with hepatocellular carcinoma treated with transarterial chemoembolization: Comparison of models using radiomics. *AJR Am J Roentgenol* 2018;**211**:1026–34. doi:10.2214/AJR.18.19507.

22. Sun R, Limkin EJ, Vakalopoulou M, *et al*. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;**19**: 1180–91. doi:10.1016/S1470-2045(18)30413-3.

23. Zheng X, Yao Z, Huang Y, *et al*. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun* 2020;**11**:1236. doi:10.1038/s41467-020-15027-z.

24. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, *et al*. Radiomics of CT features may Be nonreproducible and redundant: Influence of CT acquisition parameters. *Radiology* 2018;**288**:407– 15. doi:10.1148/radiol.2018172361.

25. Marostica E, Barber R, Denize T, *et al*. Development of a histopathology informatics pipeline for classification and prediction of clinical outcomes in subtypes of renal cell carcinoma. *Clin Cancer Res* 2021;**27**:2868–78. doi:10.1158/1078-0432.CCR-20-4119.

26. Lu MY, Williamson DFK, Chen TY, *et al*. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;**5**:555–70. doi:10.1038/s41551-020-00682-w.

27. Xu Y, Hosny A, Zeleznik R, *et al*. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019;**25**:3266–75. doi:10.1158/1078-0432.CCR-18-2495.

28. Coy H, Hsieh K, Wu W, *et al*. Deep learning and radiomics: The utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT. *Abdom Radiol (NY)* 2019;**44**:2009–20. doi:10.1007/s00261-019-01929-0.

29. Lin Z, Cui Y, Liu J, *et al*. Automated segmentation of kidney and renal mass and automated detection of renal mass in CT urography using 3D U-net-based deep convolutional neural network. *Eur Radiol* 2021;**31**:5021–31. doi:10.1007/s00330-020-07608-9.

30. Oberai A, Varghese B, Cen S, *et al*. Deep learning based classification of solid lipid-poor contrast enhancing renal masses using contrast enhanced CT. *Br J Radiol* 2020;**93**: 20200002. doi:10.1259/bjr.20200002.