



Development and External Validation of Radiomics Approach for Nuclear Grading in Clear Cell Renal Cell Carcinoma

Hongyu Zhou, MS^{1,2,4}, Haixia Mao, MD³, Di Dong, PhD^{2,4}, Mengjie Fang, MS^{2,4}, Dongsheng Gu, MS^{2,4}, Xueling Liu, MD³, Min Xu, MD³, Shudong Yang, MD⁵, Jian Zou, PhD⁶, Ruohan Yin, MD⁷, Hairong Zheng, PhD^{1,4}, Jie Tian, PhD^{2,4,8,9}, Changjie Pan, MD⁷, and Xiangming Fang, MD³

¹Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, SZ University Town, Shenzhen, China; ²CAS Key Lab of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China; ³Department of Radiology, Wuxi People's Hospital, Nanjing Medical University, Wuxi, Jiangsu, China; ⁴University of Chinese Academy of Sciences, Beijing, China; ⁵Department of Pathology, Wuxi People's Hospital, Nanjing Medical University, Wuxi, Jiangsu, China; ⁶Center of Clinical Research, Wuxi People's Hospital, Nanjing Medical University, Wuxi, Jiangsu, China; ⁷Department of Radiology, Changzhou No. 2 People's Hospital, Nanjing Medical University, Changzhou, Jiangsu, China; ⁸Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China; ⁹Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi, China

ABSTRACT

Background and Purpose. Nuclear grades of clear cell renal cell carcinoma (ccRCC) are usually confirmed by invasive methods. Radiomics is a quantitative tool that uses non-invasive medical imaging for tumor diagnosis and prognosis. In this study, a radiomics approach was proposed to analyze the association between preoperative computed tomography (CT) images and nuclear grades of ccRCC.

Hongyu Zhou, Haixia Mao, Di Dong, Changjie Pan have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1245/s10434-020-08255-6>) contains supplementary material, which is available to authorized users.

Methods. Our dataset included 320 ccRCC patients from two centers and was divided into a training set ($n = 124$), an internal test set ($n = 123$), and an external test set ($n = 73$). A radiomic feature set was extracted from unenhanced, corticomedullary phase, and nephrographic phase CT images. The maximizing independent classification information criteria function and recursive feature elimination with cross-validation were used to select effective features. Random forests were used to build a final model for predicting nuclear grades, and area under the receiver operating characteristic curve (AUC) was used to evaluate the performance of radiomic features and models.

Results. The radiomic features from the three CT phases could effectively distinguished the four nuclear grades. A combined model, merging radiomic features and clinical characteristics, obtained good predictive performances in the internal test set (AUC 0.77, 0.75, 0.79, and 0.85 for the four grades, respectively), and performance was further confirmed in the external test set, with AUCs of 0.75, 0.68, and 0.73 (no fourth-level data).

Conclusion. The combination of CT radiomic features and clinical characteristics could discriminate the nuclear grades in ccRCC, which may help in assisting treatment decision making.

© Society of Surgical Oncology 2020

First Received: 8 May 2019;
Published Online: 18 May 2020

X. Fang, MD
e-mail: xiangming_fang@njmu.edu.cn

J. Tian, PhD
e-mail: jie.tian@ia.ac.cn

H. Zheng, PhD
e-mail: hr.zheng@siat.ac.cn

Renal cell carcinoma (RCC) is one of the most common malignant tumors of the kidney.¹ Among RCCs, clear cell RCC (ccRCC) is the most common pathological subtype, with a relatively poor prognosis.^{2,3}

In the latest WHO classification of tumors in the urinary system and male genital organs, a pathological nuclear grade system, the International Society of Urological Pathology (WHO/ISUP) grading system, has been established.⁴ This grading system regarded nuclear grade as an important prognostic predictor for ccRCC.^{5–7} Compared with low-grade ccRCCs, high-grade ccRCCs have a high growth rate and a poor prognosis.^{8–10} Therefore, nuclear grades can reflect the aggressiveness of ccRCC¹¹; however, nuclear grade can only be confirmed by pathological tests from biopsy or surgical tissue, which are invasive.¹²

Radiomics is a non-invasive examination approach that has been widely used in diagnosis and treatment outcome evaluations in oncology.^{13–17} Through deep mining the quantitative radiomic features from medical images, radiomics can select key features that reflect clinical or pathological information of the tumors, to assist diagnosis and prognosis.^{18–21} Thus, the radiomics method provides a potential tool for non-invasive nuclear grading.

In this study, we developed a radiomic approach to analyze the association between preoperative CT images and nuclear grades of ccRCC.

MATERIALS

Patients

This retrospective study was performed in accordance with the Declaration of Helsinki and was approved by the ethical committee of two hospitals. The requirement for informed consent was waived.

The inclusion criteria were (1) patients with pathologically confirmed ccRCC; (2) patients who underwent dynamic enhanced kidney CT scans before surgery; and (3) nuclear grades were available from the pathology reports. The exclusion criteria were (1) the CT scan was performed more than 1 week before surgery; and (2) the CT slice thickness was not 5 mm.

Finally, we collected 247 eligible ccRCC patients from center A (Wuxi People's Hospital) and divided them into a training set ($n = 124$) and an internal test set ($n = 123$). Furthermore, an external test set ($n = 73$) was enrolled from center B (Changzhou No. 2 People's Hospital). Details of the three sets are shown in Table 1.

Clinical Characteristics and Nuclear Grade

As shown in Table 1, the preoperative clinical characteristics of patients included sex, age, history of chronic diseases (hypertension and tumor history), chief complaint (gross hematuria or lumbar discomfort), presence of urinary occult blood, renal function, and T staging (determined by the radiologist from a CT scan, according to the American Joint Committee on Cancer [AJCC] guidelines).²²

All pathological sections of patients' surgical specimens were reviewed by two independent histopathologists to estimate the nuclear grades. According to the 2016 WHO/ISUP grading system, the nuclear grades were defined as grades 1, 2, 3, and 4 at $\times 100$ and $\times 400$ magnification pathologic images.⁴ Where necessary, discrepancies were resolved by a third histopathologist.

Computed Tomography Image Acquisition

The CT scans in center A were performed using a Sensation-64 scanner (SOMATOM Definition; Siemens, Munich, Germany), with parameters that included 120 kVp, automatic mAs, a pitch of 0.55, a scan field of view of 360 mm, a pixel matrix size of 512×512 , and a reconstructed slice thickness of 5 mm. The CT scans in center B were performed using a Sensation-64 scanner (SOMATOM Force; Siemens, Germany), with parameters that included 120 kVp, automatic mAs, a pitch of 0.80, a scan field of view of 360 mm, a pixel matrix size of 512×512 , and a reconstructed slice thickness of 5 mm.

During scanning, a non-enhanced CT scan was acquired, followed by post-contrast enhanced scans. Non-ionic iodinated contrast material, Loversol (320 mgI/mL), was intravenously injected with an automatic high-pressure injector, at a flow rate of 3 mL/s, followed by a 30 mL saline chaser. The total dosage of the contrast material was 1.5 mL/kg. Following the initiation of injections, a corticomedullary phase (CMP) scan was performed at 30 s and a nephrographic phase (NP) scan was performed at 70 s. The scan area ranged from the dome of the liver to the symphysis pubis.

Tumor Volume of Interest Segmentation

Manual segmentation of the volume of interest (VOI) of the tumor was conducted using three-dimensional (3D) Slicer software (ITK-SNAP, version 3.6.0; <http://www.itk-snap.org>) on unenhanced, CMP, and NP CT images. The VOIs were segmented by a radiologist with more than 5 years' experience in ccRCC diagnoses. Moreover, 30 patients were randomly selected and segmented by another radiologist to construct a test-retest set and to calculate the

TABLE 1 Clinical information of the ccRCC patients in all sets

Clinical features	Training set				Internal test set				External test set			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Total	23	77	15	9	23	77	14	9	13	46	14	0
Sex												
Male	9	58	11	5	14	51	7	7	7	33	12	0
Female	14	19	4	4	9	26	7	2	6	13	2	0
Age, years	55 ± 11	58 ± 12	56 ± 8	64 ± 10	53 ± 13	60 ± 12	58 ± 16	63 ± 6	56 ± 13	62 ± 12	61 ± 11	0
Hypertension	3	25	6	4	5	23	4	5	4	27	8	0
The chief complaint												
Hematuria	1	4	2	0	2	7	0	1	0	3	3	0
Lumbar discomfort	3	5	1	0	1	4	4	0	3	2	1	0
Other	19	68	12	9	20	66	10	8	10	41	10	0
Serum tumor markers	0	11	2	0	3	9	4	2	–	–	–	–
Renal function	1	15	2	1	4	16	3	3	–	–	–	–
T stage												
1A	19	42	6	0	18	39	3	2	11	30	2	0
1B	2	25	7	5	3	28	5	4	1	16	9	0
2	1	7	0	2	1	4	4	1	0	0	0	0
3	1	2	2	2	1	6	0	2	1	0	3	0
4	0	1	0	0	0	0	2	0	0	0	0	0
Urine occult blood	10	24	4	3	4	27	6	5	0	3	4	0

interclass correlation coefficient (ICC) of radiomic features.²³

Radiomic Feature Extraction

This study used the radiomic features proposed by Aerts et al.²⁴ and Lambin et al.²⁵ We extracted four types of 3D features in the VOI: (1) intensity features, i.e. the first-order statistics features; (2) textural features; (3) shape features, i.e. size, maximum diameter, and maximum section of the tumor; and (4) wavelet features, i.e. intensity and texture features from wavelet decomposition of the original images. For all features, we used the normalization function to obtain the relative value instead of the overall resize. The features were continuous variables and no binning was performed. Note that the variations of a feature between different CT phases were also used as candidate features. In total, we extracted 647 features per phase, 647 variation features per two phases, and 3882 features in total per patient (the process is shown in Fig. 1).

Model Building and Evaluation

First, we used the test–retest data ($n = 30$), segmented by different radiologists, to evaluate the reproducibility of features between different segmentations. We preserved features with an ICC value < 0.75 . Second, we used the maximizing independent classification information (MICI) criteria function²⁶ to select the features associated with nuclear grades. On this basis, we used the recursive feature elimination with cross-validation (RFECV) method and the random forest (RF) classifier to find the best radiomic features for nuclear grading.²⁷ Finally, the selected features were combined with clinical characteristics to generate a final combined model. We also built a clinical model using only clinical characteristics, and a radiomic model using only radiomic features, for comparison.

The ability of radiomic features and models to categorize nuclear grades was evaluated by the area under the curve (AUC) of the receiver operator characteristic curve (ROC).

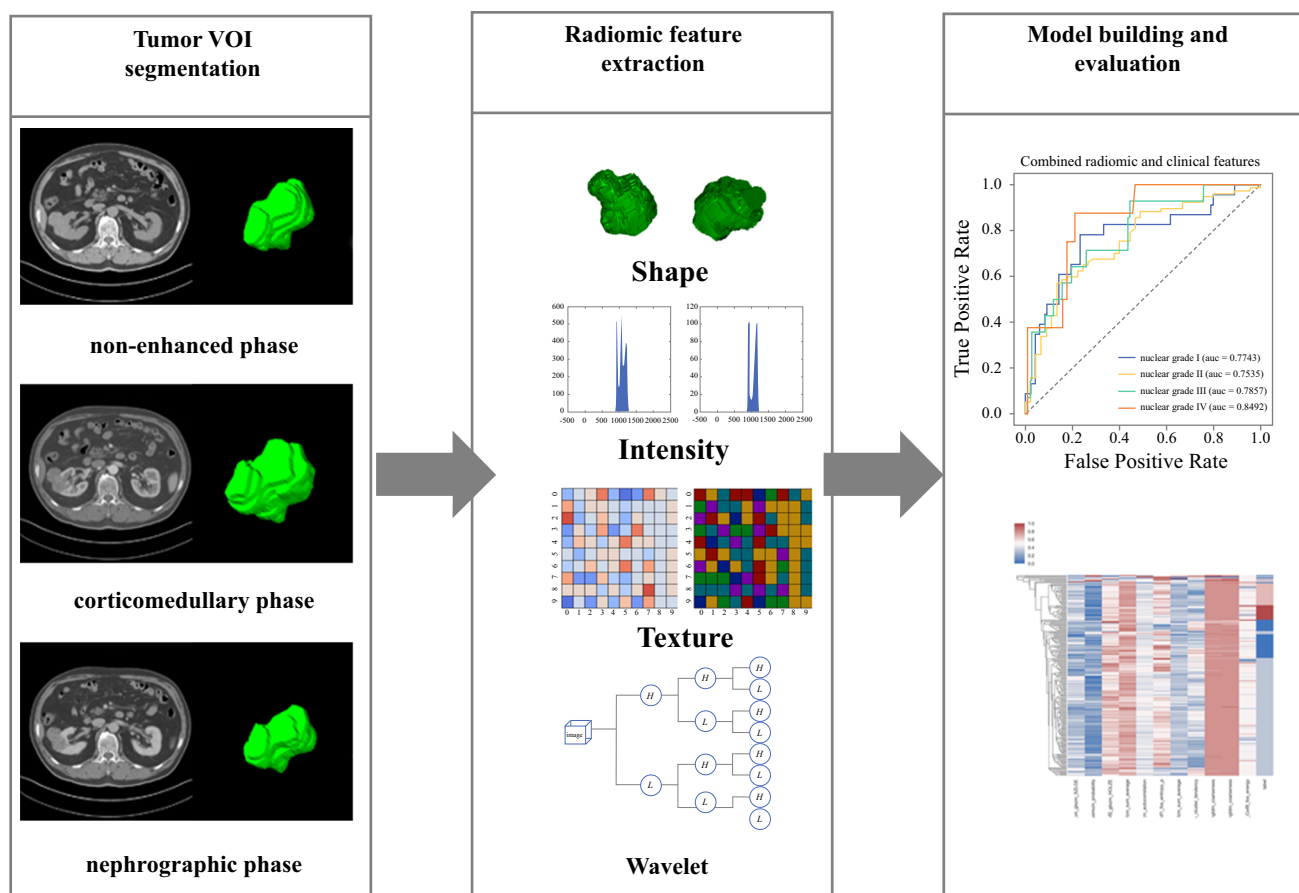


FIG. 1 Radiomics workflow diagram. VOI volume of interest, AUC area under the curve

Statistical Analyses

All statistical analyses were performed using Python version 3.6, while the RFECV feature selection, RC classifier, and AUC calculation used the Python Scikit-learn package.^{28,29}

RESULTS

Based on the test-retest data set, we obtained 2668 robust radiomic features with an ICC > 0.75 (Fig. 2). A total of 46 features determined by the MICI method were selected as candidates for classification, which included the maximum, minimum, median, and range values of the first three-order wavelet decompositions in the three phases of CT images, and the variations of texture features and entropy between the three phases.

Based on both the RFECV method and the RF model assessment, we obtained the most effective 11 features to build a radiomic model. Selected features included (1) texture features of CMP: gray-level co-occurrence matrix (GLCM) after wavelet decomposition, and gray level size zone matrix (GLSZM); (2) texture features of the NP:

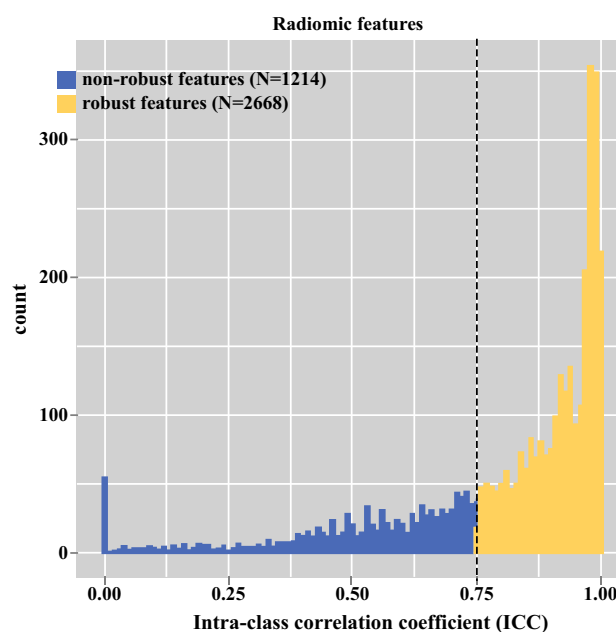


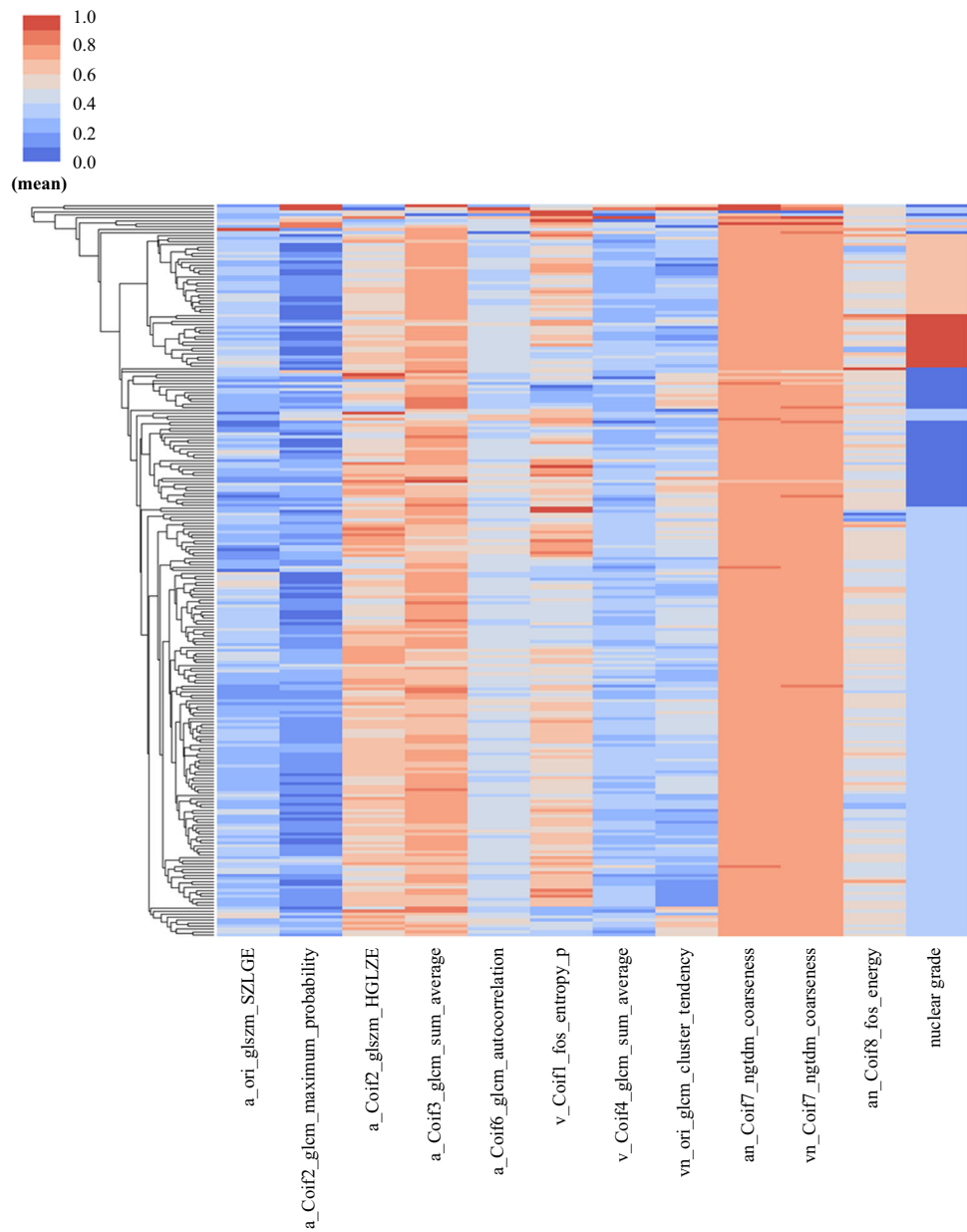
FIG. 2 Stable radiomic feature selection using test-retest set

entropy and GLCM after wavelet decomposition; (3) variation of texture features between the CMP and the non-enhanced phase: the energy and neighborhood gray tone difference matrix (NGTDM); and (4) variation of texture features between the NP and the non-enhanced phase: the GLCM and NGTDM. The clustered map of the features in the radiomic model is shown in Fig. 3.

The ROC curves of the radiomic model, clinical model, and the combined model are shown in Fig. 4. The AUC values of the four-class classifications of the radiomic model reached 0.71, 0.73, 0.81, and 0.80, respectively, in the internal test set. The clinical model (including sex, age, T stage, hypertension, chief complaint, and urine occult blood) showed high performance in predicting the grade 4

with an AUC of 0.88 in the internal test set, indicating the clinical factors had a high sensitivity in discriminating grade 4. Both the radiomic and clinical models worked worse in the external test set. Combined with radiomic features and clinical characteristics, the final combined model had an AUC of 0.77, 0.75, 0.79, and 0.85 in the internal test set. The AUC of the combined model on the external test set reached 0.75, 0.68, and 0.73 (there was no fourth-level grade in the external test set). We found that the addition of clinical characteristics to the radiomic model could help improve performance. Compared with the radiomic and clinical models, the combined model performed better and was more robust.

FIG. 3 Radiomic feature cluster diagram. ‘a’ and ‘v’ represent the features extracted in the corticomedullary phase and nephrographic phase CT, respectively; ‘an’ and ‘vn’ represent the variation between the corticomedullary phase, nephrographic phase, and non-enhanced phase feature; and ‘Coif1’ means that the feature is decomposed by the first-order wavelet. *CT* computed tomography



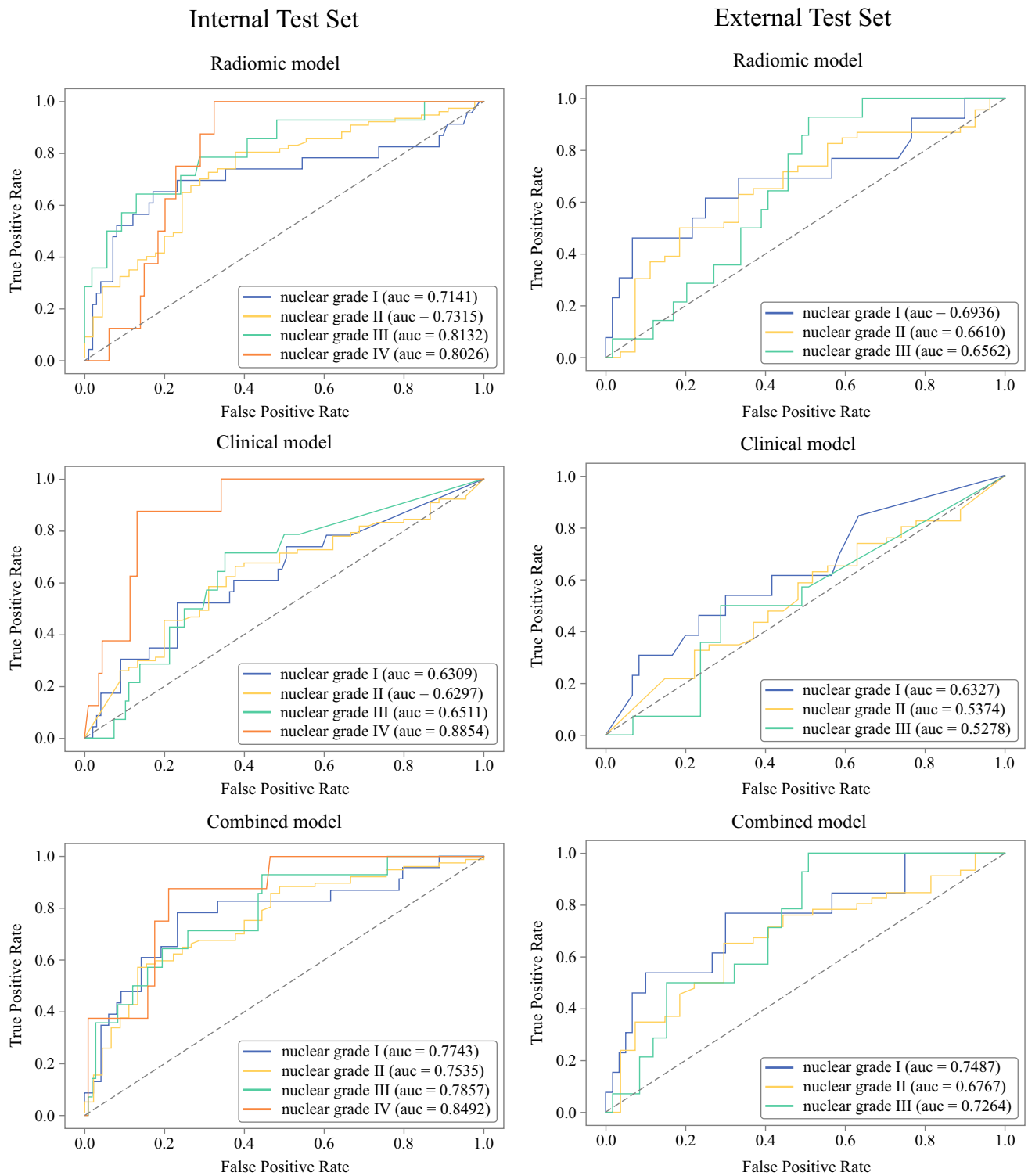


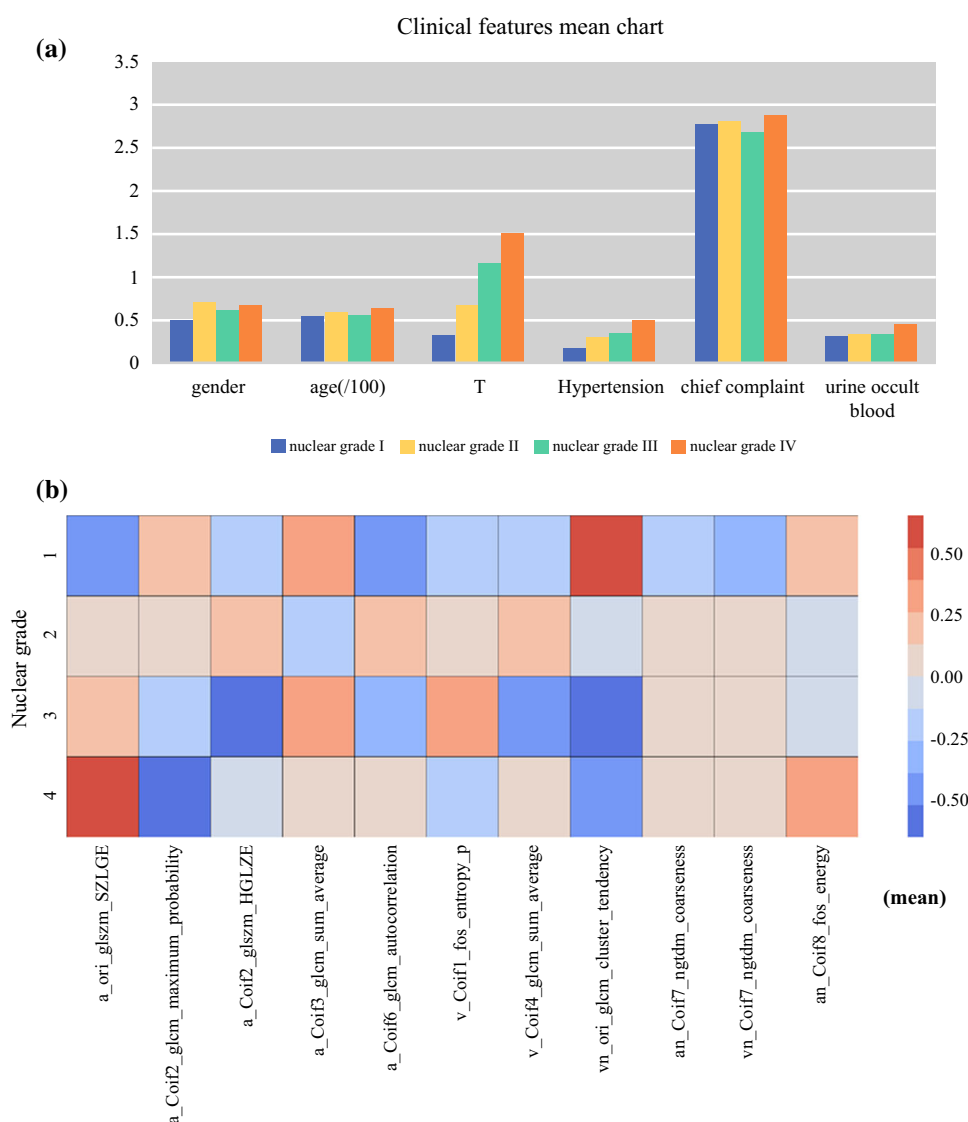
FIG. 4 ROC curves of different models on the internal validation set and external validation set. *ROC* receiver operating characteristic, *AUC* area under the curve

As shown in Fig. 5a, the mean values of the clinical characteristics for each grade in the internal test set showed that nuclear grade 4 had the worst conditions (T stage, hypertension, urine occult blood); hence, the clinical model

was effective in discriminating grade 4 in our study. Figure 5b shows the mean distribution of the final radiomic features used in the radiomic and combined models.

FIG. 5 Analysis of the effective features. **a** Distribution of the mean values of different clinical features. **b** Distribution of the 11 radiomic features.

Note: sex (0: female; 1: male), age (divided by 100), T stage (0: T1A; 1: T1B; 2: T2; 3: T3; 4: T4), hypertension (0: without hypertension; 1: with hypertension), chief complaint (1: hematuria; 2: lumbar discomfort; 3: other), urine occult blood (0: without urine occult blood; 1: with urine occult blood)



DISCUSSION

In this multicenter study, we investigated the relationship between nuclear grades of ccRCC, with radiomic features in non-enhanced phase, CMP, and NP CT images, and found that the radiomic features in all CT phases were associated with nuclear grades. The combination of radiomic features and clinical characteristics could well discriminate nuclear grade, and it performed better than the radiomic and clinical models.

We found that the texture features and perplexity of the tumor VOI had significant variation in different nuclear grades. The WHO/ISUP grading system defined tumor grades 1–3 based on nucleolar prominence, and defined grade 4 based on the presence of pronounced nuclear pleomorphism, tumor giant cells, and/or rhabdoid and/or sarcomatoid differentiation. The proportion of nucleoli,

nuclear pleomorphism, tumor giant cells, and rhabdoid or sarcomatoid differentiation under microscope reflected cell density and tumor heterogeneity, which could be associated with texture features and perplexity of the tumor in the CT images.

Ding et al. performed a preliminary study on nuclear grading.³⁰ Compared with their work, we used the latest urological pathology (WHO/ISUP) grading system standard (4-level grading) instead of the Fuhrman grading system (2-level grading). In terms of method, our study used CMP and NP CT images, which were not included in the work of Ding et al. Moreover, the wavelet transform and variation of feature between different phases were used in our study to generate more features.

Our study had some limitations. First, although this study was performed on two centers, the sample size was relatively small. Further validation of our model on large-

scale data sets is necessary. Second, we used a CT slice thickness of 5 mm and did not consider other thicknesses. Third, this was a retrospective research study and further prospective validation is needed.

CONCLUSIONS

This study proposed a non-invasive method based on preoperative CT images and radiomics methods for nuclear grading in ccRCC, which could assist clinicians in diagnosis and treatment decision making.

ACKNOWLEDGEMENT The authors would like to acknowledge the instrumental and technical support of the Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

AUTHOR CONTRIBUTIONS Research idea and study design: XF, HZ, JT. Data acquisition: HM, XL, MX, SY, JZ, RY. Data analysis/interpretation: HZ, DD, MF, DG, HZ, XF, JT. Supervision or mentorship: DD, HZ, XF, JT, CP, and XF.

FUNDING This work was supported by the National Key R&D Program of China (2017YFC1308700, 2017YFA0205200, 2017YFC1309100, 2016YFC0103803, 2016YFC0103001), National Natural Science Foundation of China (91959130, 81971776, 81771924, 812716298, 81271629, 81227901), Beijing Natural Science Foundation (L182061), Wuxi Medical Innovation Team Program (CXTD002), Bureau of International Cooperation of Chinese Academy of Sciences (173211KYSB20160053), and the Youth Innovation Promotion Association CAS (2017175).

DISCLOSURE Hongyu Zhou, Haixia Mao, Di Dong, Mengjie Fang, Dongsheng Gu, Xueling Liu, Min Xu, Shudong Yang, Jian Zou, Ruohan Yin, Hairong Zheng, Jie Tian, Changjie Pan, and Xiangming Fang have declared no conflict of interest.

REFERENCES

- Bhatt JR, Finelli A. Landmarks in the diagnosis and treatment of renal cell carcinoma. *Nat Rev Urol*. 2014;11:517–25.
- Störkel S, Eble JN, Adlakh MD, et al. Classification of renal cell carcinoma. *Cancer*. 1997;80:987.
- Rabjerg M. Identification and validation of novel prognostic markers in Renal Cell Carcinoma. *Dan Med J*. 2017;64:B5339.
- Moch H, Cubilla AL, Humphrey PA, et al. The 2016 WHO classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *Histopathology*. 2016;46:93–105.
- Patard JJ, Leray E, Rioux-Leclercq N, et al. Prognostic value of histologic subtypes in renal cell carcinoma: a multicenter experience. *J Urol*. 2006;175:2763–71.
- Ljungberg B, Bensalah K, Canfield S, et al. EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015;67:913–24.
- Zhu YH, Wang X, Zhang J, et al. Low enhancement on multiphase contrast-enhanced CT images: an independent predictor of the presence of high tumor grade of clear cell renal cell carcinoma. *AJR Am J Roentgenol*. 2014;203:295–300.
- Coy H, Douek M, Young J, et al. Differentiation of low grade from high grade clear cell renal cell carcinoma neoplasms using a CAD algorithm on four-phase CT. *J Clin Oncol*. 2016;34(15 Suppl):4564.
- Leibovich BC, Blute ML, Cheville JC, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer*. 2003;97:1663–71.
- Erdoğan F, et al. Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol*. 1982;58:655–63.
- Bretheau D, Lechevallier E, De FM, et al. Prognostic value of nuclear grade of renal cell carcinoma. *Cancer*. 1995;76:2543.
- Sekar RR, Patil D, Pearl J, et al. The relationship between pre-operative c-reactive protein and Fuhrman nuclear grade in stage T1 renal cell carcinoma. *J Urol*. 2016;195:e1033.
- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
- Zhou H, Dong D, Chen B, et al. Diagnosis of distant metastasis of lung cancer: based on clinical and radiomic features. *Transl Oncol*. 2018;11:31–6.
- Song J, Shi J, Dong D, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res*. 2018;24(15):3583–92.
- Dong D, Tang L, Li Z-Y, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol*. 2019;30:431–8.
- Dong D, Zhang F, Zhong L-Z, et al. Development and validation of a novel MR imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: a randomized controlled trial substudy (NCT01245959). *BMC Med*. 2019;17(1):190.
- Peng H, Dong D, Fang M, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res*. 2019;25(14):4271–9.
- Zhu X, Dong D, Chen Z, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol*. 2018;28(7):2772–8.
- Yang L, Dong D, Fang M, et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? *Eur Radiol*. 2018;28(5):2058–67.
- Wang S, Zhou M, Liu Z, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. *Med Image Anal*. 2017;40:172–83.
- Edge SB, Byrd DR, Compton CC, et al. American Joint Committee on Cancer (AJCC) cancer staging manual; 2010.
- Rios VE, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res*. 2017;77:3922.
- Aerts HWJL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
- Lambin P, Rth L, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749.
- Wang J, Wei JM, Yang Z, Wang SQ. Feature selection by maximizing independent classification information. *IEEE Trans Knowl Data Eng*. 2017;29:828–41.
- Granitto PM, Furlanello C, Biasioli F, Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom Intell Lab Syst*. 2006;83:83–90.
- Buitinck L, Louppe G, Blondel M, et al. API design for machine learning software: experiences from the scikit-learn project. *Eprint Arxiv*; 2013.

29. Pedregosa F, Gramfort A, Michel V, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2016;12:2825–30.
30. Ding J, Xing Z, Jiang Z, et al. CT-based radiomic model predicts high grade of clear cell renal cell carcinoma. *Eur J Radiol.* 2018;103:51–6.

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