Deep Learning Based on MRI for Differentiation of Low- and High-Grade in Low-Stage Renal Cell Carcinoma

Yijun Zhao, MD,[1] Marcello Chang, MD,[2] Robin Wang, BA,[3] Ianto Lin Xi, BS,[3] Ken Chang, MS,[4] Raymond Y Huang, MD,[5] Martin Vallières, PhD,[6] Peiman Habibollahi, MD,[7] Mandeep S. Dagli, MD,[8] Matthew Palmer, MD,[9] Paul J. Zhang, MD,[9] Alvin C. Silva, MD,[10] Li Yang, MD, PhD,[11] Michael C. Soulen, MD,[8] Zishu Zhang, MD,[1]* Harrison X. Bai, MD,[12]* and S. William Stavropoulos, MD[8]*

Background
Pretreatment determination of renal cell carcinoma aggressiveness may help to guide clinical decision-making.
Purpose: To evaluate the efficacy of residual convolutional neural network using routine MRI in differentiating low-grade (grade I–II) from high-grade (grade III–IV) in stage I and II renal cell carcinoma.
Study Type: Retrospective.
Population: In all, 376 patients with 430 renal cell carcinoma lesions from 2008–2019 in a multicenter cohort were acquired. The 353 Fuhrman-graded renal cell carcinomas were divided into a training, validation, and test set with a 7:2:1 split. The 77 WHO/ISUP graded renal cell carcinomas were used as a separate WHO/ISUP test set.
Field Strength/Sequence: 1.5T and 3.0T/T_2-weighted and T_1 contrast-enhanced sequences.
Assessment: The accuracy, sensitivity, and specificity of the final model were assessed. The receiver operating characteristic (ROC) curve and precision-recall curve were plotted to measure the performance of the binary classifier. A confusion matrix was drawn to show the true positive, true negative, false positive, and false negative of the model.
Statistical Tests: Mann–Whitney *U*-test for continuous data and the chi-square test or Fisher's exact test for categorical data were used to compare the difference of clinicopathologic characteristics between the low- and high-grade groups. The adjusted Wald method was used to calculate the 95% confidence interval (CI) of accuracy, sensitivity, and specificity.
Results: The final deep-learning model achieved a test accuracy of 0.88 (95% CI: 0.73–0.96), sensitivity of 0.89 (95% CI: 0.74–0.96), and specificity of 0.88 (95% CI: 0.73–0.96) in the Fuhrman test set and a test accuracy of 0.83 (95% CI: 0.73–0.90), sensitivity of 0.92 (95% CI: 0.84–0.97), and specificity of 0.78 (95% CI: 0.68–0.86) in the WHO/ISUP test set.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.27153

Received Feb 4, 2020, Accepted for publication Mar 13, 2020.

*Address reprint request to: H.X.B., Providence, RI 02905, USA. E-mail: harrison_bai@brown.edu or S.W., 1Silverstein, 3400 Spruce St., Philadelphia, PA 19104, USA. E-mail: stav@uphs.upenn.edu or Z.Z., No.139 Middle Renmin Road, Changsha, Hunan 410011, China. E-mail: zishuzhang@csu.edu.cn
Yijun Zhao and Marcello Chang share primary authorship

Contract grant sponsor: RSNA fellow research grant; Contract grant number: RF1802; Contract grant sponsor: National Institution of Health/National Cancer Institute; Contract grant number: R03CA249554; Contract grant sponsor: SIR Foundation Radiology Resident Research Grant to H.X.B.; Contract grant sponsor: training grant from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health; Contract grant number: 5T32EB1680; Contract grant sponsor: National Cancer Institute (NCI) of the National Institutes of Health; Contract grant number: F30CA239407 (to K.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This study was partially supported by funds from the Nicolo Foundation.

From the [1]Department of Radiology, The Second Xiangya Hospital, Central South University, Changsha, China; [2]Stanford School of Medicine, Palo Alto, California, USA; [3]Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA; [4]Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Boston, Massachusetts, USA; [5]Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts, USA; [6]Medical Physics Unit, McGill University, Montreal, Québec, Canada; [7]Department of Radiology, Division of Interventional Radiology, UT Southwestern Medical School, Dallas, Texas, USA; [8]Department of Radiology, Division of Interventional Radiology, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania, USA; [9]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [10]Department of Radiology, Mayo Clinical Hospital, Scottsdale, Arizona, USA; [11]Department of Neurology, The Second Xiangya Hospital, Central South University, Changsha, China; and [12]Department of Diagnostic Imaging, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA

Additional supporting information may be found in the online version of this article

Data Conclusion: Deep learning can noninvasively predict the histological grade of stage I and II renal cell carcinoma using conventional MRI in a multiinstitutional dataset with high accuracy.
Level of Evidence: 3
Technical Efficacy Stage: 2

J. MAGN. RESON. IMAGING 2020.

RENAL CELL CARCINOMAS (RCCs) are the most common renal malignancy in adults, affecting ~65,000 new patients each year.[1] The increasing use of cross-section imaging has resulted in RCCs being diagnosed when smaller in size and lower in stage.[2,3] The standard management for RCC is partial or radical nephrectomy.[4] The TNM staging system is used for clinicopathologic staging of RCCs, based on tumor size, tumor extension, lymph node involvement, and metastatic spread.[5] Stages I and II (low stage) RCCs have better prognosis and can mostly be treated with nephron-sparing surgery.[6,7] Less-invasive treatment options, such as percutaneous cryoablation, radiofrequency ablation, and active surveillance, have been investigated in low-stage tumors.[8–10]

Histological grade is an important prognostic indicator for predicting long-term survival in patients with intracapsular (pT1 and pT2) tumor,[11,12] particularly in predicting development of distant metastases after nephrectomy.[13] There are several tumor grading systems for RCC, with Fuhrman grade being one of the most widely used.[14] The association between Fuhrman grade and prognosis is well established.[15] The International Society of Urological Pathologists (ISUP) grading system was proposed in 2012 and endorsed by the World Health Organization (WHO), being redesignated the WHO/ISUP grading system in 2015.[16] WHO/ISUP was widely adopted for RCC grading only very recently and has been proven to be a prognostic indicator of RCC.[17] Biopsy is invasive and subject to poor tumor targeting, sampling error, and false-negative results.[18] Therefore, the ability to assess the aggressiveness of low-stage RCC noninvasively will allow better selection of patients most likely to benefit from ablative therapy or surveillance and can be used to develop appropriate surveillance programs matched to the risk of tumor progression.

Previous studies have demonstrated the value of imaging in differentiating low-grade from high Fuhrman grade RCC.[19–21] Recently, several studies have developed predictive models of Fuhrman grade using computed tomography (CT)-based radiomics on RCC with promising results.[22,23] Deep learning, a type of machine-learning technique that takes raw images as input and applies many layers of transformations to calculate an output signal, has already led to breakthroughs in predicting molecular biomarkers and disease outcomes in other organ systems.[24,25] For example, deep learning has been used for automated segmentation of kidney volume on CT.[26]

The purpose of the current study was to apply a deep learning-based approach based on routine magnetic resonance imaging (MRI) to differentiate low- from high-grade low-stage RCCs in a multicenter cohort.

Materials and Methods

Patient Cohort

Patients with histologically confirmed Fuhrman or WHO/ISUP grading I–IV RCCs from two large academic centers in the United States (HU, Hospital of the University of Pennsylvania and MA, Mayo Clinical Hospital), two hospitals in the People's Republic of China (SX, The Second Xiangya Hospital of Central South University and PH, People's Hospital of Hunan Province) and The Cancer Imaging Archive (TCIA) were retrospectively identified. The study was approved by the Institutional Review Boards (IRBs) of the four hospitals. With the agreement to use TCGA/TCIA data, the IRB approval of our study was waived for TCIA. The informed consent of patients was waived. The inclusion criteria were 1) pathologically confirmed stage I or II RCC with reported histological grade (either Fuhrman grade or WHO/ISUP grade); 2) available preoperative MRI including T_2-weighted (T_2W) and T_1-contrast (T_1C) enhanced sequences; 3) quality of the images was adequate for analysis, without motion or artifacts. If dynamic enhancement was performed, the earliest phase on the T_1C sequence was chosen. Our final cohort consisted 430 RCC lesions (343 lesions from HU, 48 lesions from MA, 14 lesions from TCIA, 10 lesions from SX, and 15 lesions from PH) (Fig. 1). The cohort was partly overlapped with our previous work.[27]

Histopathologic diagnosis was obtained for all 430 tumors after surgical excision; Fuhrman grade was available for 353 tumors according to the 4-tiered Fuhrman classification, while the other 77 tumors were graded by the 4-tiered WHO/ISUP grading system. RCCs were grouped into low-grade (grades I and II) and high-grade (grades III and IV).

Tumor Segmentation

MR images of all patients were loaded into 3D Slicer software (v.4.6), and an abdominal radiologist (Y.Z.) with 5 years of experience reading abdominal MRIs traced the outline of the entire tumor on all contiguous slices where the tumor was visible on the T_2 and T_1C sequences, generating a 3D region of interest.[28] The radiologist was blinded to the histopathological data.

Model

Our neural network model was based on the ResNet50 architecture with ImageNet pretrained weights.[29] Each of our T_1C and T_2 sequences were fed into their own ResNet50 and one of the final layers (with 2024 features) was concatenated. All layers after this were dense layers with the lower part of the model resembling a conventional fully connected neural net. After two more fully connected layers of decreasing size, the resulting dense layer with 24 features

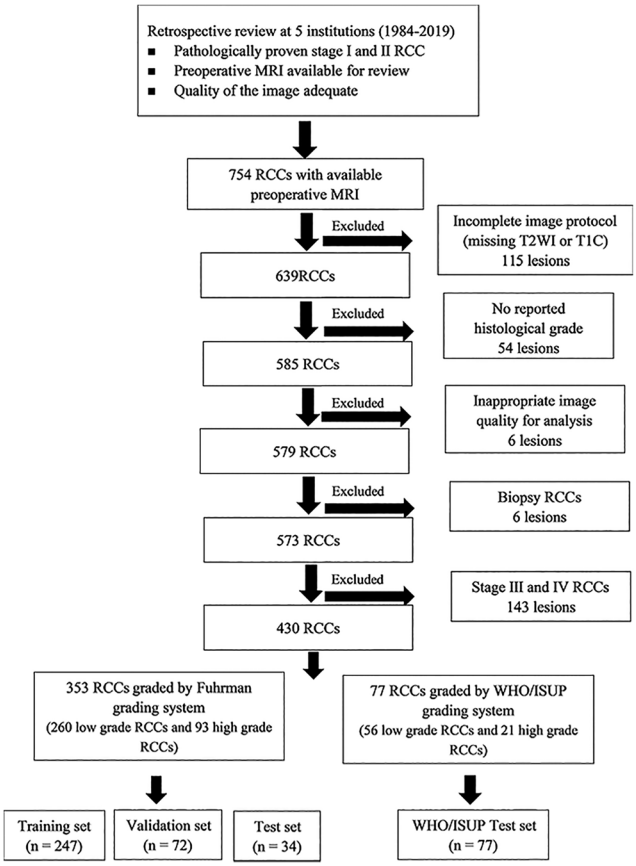

FIGURE 1: Flow diagram of patient inclusion and exclusion.

was concatenated to the clinical variables (an input of 10 features such as age, gender, and tumor size). They were finally fully connected down a dense layer with two nodes for our binary classification. The Dropout layer present in the Dense 24 and Dense 2 layers only stochastically dropped weights in the training stage. Dropout has been shown to reduce overfitting in fully connected neural nets in previous work.[30] The model was optimized with stochastic gradient descent using a learning rate of 0.1, with the learning rate reducing on the plateau of validation loss (reduced by half if validation loss did not decrease for 25 epochs). An illustration of our pipeline is shown in Fig. 2.

Training

Our Fuhrman datasets were partitioned into training, validation, and testing sets at a ratio of 7:2:1. The WHO/ISUP dataset was used as a separate test set. T_1C and T_2 images were registered using SyN registration from the ants Python package. During training, images were augmented dynamically with horizontal flip, vertical flip, shear, and zoom transformations to add variability to the training set. T_1C and T_2 sequences were augmented together, meaning the same transformations were applied to both sequences on a given batch. Image augmentation was performed by the Keras ImageDataGenerator. A batch size of 32 was used for training. The learning rate was reduced on plateau with a patience of 25 and early stopping with a patience of 50 for a maximum of 1000 epochs. A reduced learning rate on plateau and early stopping were implemented using existing Keras callbacks. A stochastic gradient descent optimizer was used with a learning rate of 0.01 and nesterov momentum. For a range of kernel regularizer and activity regularizer values, 10 models were trained and the modeled with the best test area under curve (AUC) receiver operating curve (ROC) was ultimately selected.

Statistical Analysis

Statistical analysis was performed by using SPSS v. 22.0 (SPSS, Chicago, IL) and MiniTab v. 16.0 (Minitab, State College, PA). Comparison between groups was performed by using the Mann–Whitney *U*-test for continuous data and the chi-square test or Fisher's exact test for categorical data. The ROC curve and the Precision-Recall curve were plotted to measure the performance of the binary classifier. Average accuracy, sensitivity, and specificity with the 95% confidence interval (CI) were calculated using the adjusted Wald method.[31] A *P*-value of 0.05 was considered the threshold for significance. A confusion matrix was drawn to show true positive, true negative, false positive, and false negative of the model.

Code

The Keras library[32] with Tensorflow backend[33] was used. Our models were trained on a computer with two NVidia Quadro V100 GPU. The code is publicly available on Github at https://github.com/marcelc21399/Fuhrman_repo.

Results

Patient and Tumor Characteristics

Table 1 shows the clinicopathologic characteristics of our cohort. High-grade RCCs were significantly larger than low-grade RCCs (mean size, 3.7 cm vs. 2.7 cm, $P < 0.001$). The train, validation, and test sets were balanced in terms of age, gender, size, laterality, location, histologic grade, stage, and institution. (Table S1). The MRI parameters are shown in Table S2.

Model Performance

The final model achieved a test accuracy of 0.88 (95% CI: 0.73–0.96), F1 score of 0.81, precision recall AUC of 0.89, sensitivity of 0.89 (95% CI: 0.74–0.96), and specificity of 0.88 (95% CI: 0.73–0.96) in the Fuhrman test set and a test accuracy of 0.83 (95% CI: 0.73–0. 90), F1 score of 0.79, precision recall AUC of 0.89, sensitivity of 0.92 (95% CI: 0.84–0.97), and specificity of 0.78 (95% CI: 0.68–0.86) in the WHO/ISUP test set. The performance of T_1C, T_2, clinical features, and ensemble models among the training, validation, and test cohort is shown in Table 2. Confusion matrices showing true positive, true negative, false positive, and false negative of the model in the training, validation, and test sets are shown in Table S3. The ROC curve and Precision-Recall curve demonstrated high accuracy of the deep-learning models in predicting Fuhrman grade and WHO/ISUP grade in the test sets (Fig. 3).

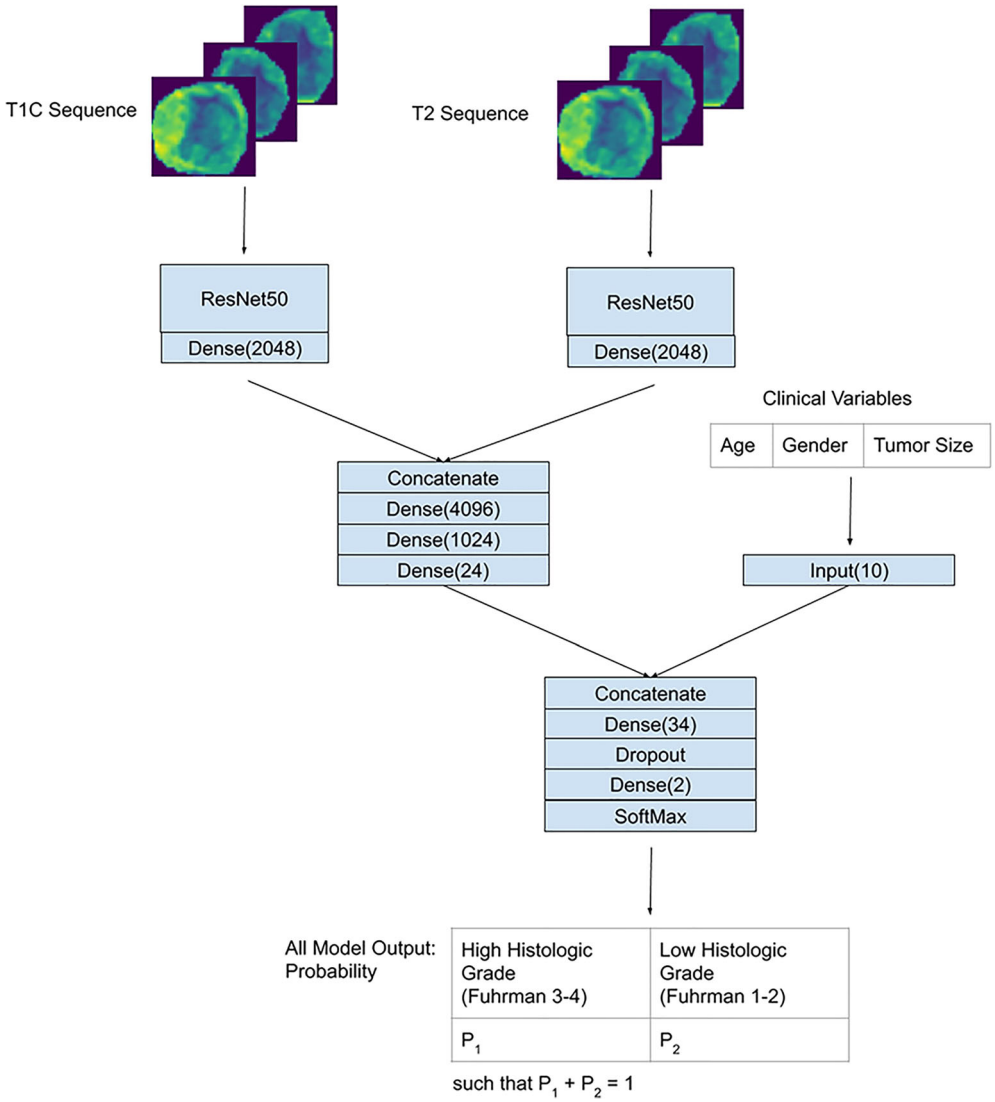

FIGURE 2: An illustration of the analysis pipeline.

Discussion

Tumor stage and grade are considered to be the most important prognostic factors in RCC.[34] Clinical staging of RCC using preoperative imaging studies have demonstrated high accuracy.[35] Pretreatment determination of RCC histological grade noninvasively is challenging but important for clinical decision-making. In this study we developed a deep-learning model based on routine MRI to predict the histological grade in stages I and II RCC in a large multicenter cohort with high accuracy.

Several studies have reported significant survival differences among the different tumor grades.[11,13,15] A significant association between Fuhrman grade and other pathological variables, such as primary tumor pathological stage, lymph node involvement, metastasis, and venous invasion have also reported in the literature.[15,36] Similarly, certain imaging features have been known to correlate with Fuhrman grade. Recently, Ding et al proposed a predictive model with a radiomics approach based on CT images consisting of texture and nontexture features to identify high Fuhrman grade RCC with an AUC of 0.88 in a training cohort and 0.77 in a validation cohort.[23] However, radiomics has the limitation of relying on the computation of a selection of manually formulated or "handcrafted" features, which may not capture the full range of information contained in the images and can be low in reproducibility. Deep learning takes raw images as input, and applies many layers of image filtering transformations to calculate an output signal. As data are used to train and update the image filtering transformations' weights, the machine "learns" to increase the accuracy of its predictions. The many degrees of freedom allow the algorithm to learn complex patterns with a high level of abstraction, many of which the human eye cannot see. This may explain the high accuracy of our algorithm.

TABLE 1. Patient Demographics, Clinical Features, and Tumor Characteristics Between Low and High Histological Grade Groups

	Low-grade (grade I and II) N = 311	High-grade (grade III and IV) N = 119	*P* value
Age, median, range (years)	61.0 (27–85)	61.0 (28–81)	0.779
Gender			0.022*
Male	205 (65.9%)	92 (77.3%)	
Female	106 (34.1%)	27 (22.7%)	
Race			0.495
White	203 (65.3%)	81 (68.1%)	
Black	65 (20.9%)	26 (21.8%)	
Asian	32 (10.3%)	7 (5.8%)	
Unknown	11 (3.5%)	5 (4.1%)	
Subtype			0.001*
Clear cell	217 (69.8%)	70(58.8%)	
Papillary	58 (18.6%)	42 (35.3%)	
Chromophobe	8 (2.6%)	1 (0.8%)	
Clear cell papillary	21 (6.8%)	4 (3.4%)	
Multilocular cystic	6 (1.9%)	0 (0%)	
Unclassified	1 (0.3%)	2 (1.7%)	
Laterality			0.531
Left	149 (47.9%)	53 (44.54)	
Right	162 (52.1%)	66 (55.5%)	
Location			0.402
Upper	107 (34.4%)	36 (30.2%)	
Interpole	115 (37.0%)	51 (42.9%)	
Lower	89 (28.6%)	32 (26.9%)	
Tumor size, median, range (cm)	2.7 (0.9–17.2)	3.7 (1.0–15.0)	<0.001*
T stage			<0.001*
T1a	242 (77.8%)	66 (55.5%)	
T1b	60 (19.3%)	41 (34.5%)	
T2a	4 (1.3%)	9 (7.5%)	
T2b	5 (1.6%)	3 (2.5%)	

*Statistically significant.

In this study the residual convolutional neural network model combining conventional MR sequences (T_2W, T_1C) and 10 clinical variables (such as age, gender, and tumor size) achieved high accuracy in differentiating low- from high-grade RCCs. Our model was based on the ResNet50 architecture, which has been shown to decrease overfitting and address vanishing gradients of deep neural nets.[29] The augmentation technique was used in our training set, which

TABLE 2. Deep-Leaning Model Performance for Overall Cohort in Training, Validation, and Test Sets

Modality	F1 Score	ROC AUC	PR AUC	Acc (95% CI)	TPR (95% CI)	TNR (95% CI)	PPV	NPV	FDR
Training									
Clinical	0.53	0.70	0.53	0.66 (0.60–0.71)	0.58 (0.51–0.64)	0.70 (0.64–0.76)	0.49	0.77	0.50
T_2WI	0.72	0.87	0.76	0.77 (0.72–0.82)	0.87 (0.82–0.90)	0.73(0.67–0.78)	0.61	0.91	0.39
T_1C	0.64	0.81	0.66	0.71 (0.65–0.76)	0.77 (0.71–0.82)	0.68 (0.62–0.73)	0.54	0.85	0.46
Ensemble	0.84	0.99	0.98	0.90 (0.86–0.93)	1.00 (0.98–1.00)	0.85 (0.82–0.91)	0.73	1.00	0.27
Validation									
Clinical	0.61	0.79	0.62	0.71 (0.60–0.81)	0.67 (0.55–0.76)	0.74 (0.63–0.83)	0.56	0.82	0.44
T_2WI	0.63	0.78	0.63	0.71 (0.59–0.80)	0.76 (0.65–0.84	0.68 (0.57–0.78)	0.54	0.85	0.46
T_1C	0.66	0.79	0.62	0.71 (0.60–0.81)	0.82 (0.71–0.89)	0.67 (0.55–0.76)	0.55	0.88	0.45
Ensemble	0.84	0.99	0.94	0.90 (0.81–0.95)	1.00 (0.94–1.00)	0.87(0.77–0.93)	0.73	1.00	0.27
Fuhrman test									
Clinical	0.53	0.75	0.54	0.66 (0.49–0.80)	0.56 (0.40–0.71)	0.71 (0.54–0.84)	0.50	0.76	0.50
T_2WI	0.67	0.77	0.55	0.72 (0.55–0.85)	0.81 (0.65–0.91)	0.68(0.51–0.81)	0.57	0.88	0.43
T_1C	0.65	0.73	0.59	0.70 (0.53–0.83)	0.81 (0.65–0.91)	0.65 (0.48–0.78)	0.54	0.87	0.46
Ensemble	0.80	0.91	0.77	0.88 (0.73–0.96)	0.89 (0.74–0.96)	0.88 (0.73–0.96)	0.73	0.96	0.27
WHO/ISUP test									
Clinical	0.56	0.73	0.65	0.73 (0.62–0.81)	0.45 (0.35–0.57)	0.89 (0.80–0.95)	0.71	0.73	0.25
T_2WI	0.52	0.70	0.65	0.63 (0.51–0.72)	0.55 (0.43–0.65)	0.67(0.56–0.77)	0.50	0.71	0.50
T_1C	0.54	0.69	0.61	0.61 (0.50–0.71)	0.61 (0.49–0.71)	0.62 (0.51–0.72)	0.49	0.72	0.51
Ensemble	0.79	0.94	0.89	0.83 (0.73–0.90)	0.92 (0.84–0.97)	0.78 (0.68–0.86)	0.69	0.95	0.26

ROC AUC: area under receiver operating characteristic curve; PR AUC: area under Precision-Recall curve; TPR: true positive rate or sensitivity; TNR true negative rate or specificity; Acc: accuracy; PPV: positive predictive value; NPV: negative predictive value; FDR: false discovery rate; N/A: not applicable; 95% CI: 95% confidence interval.

allows a further increase in the size of the cohort and prevents overfitting.[29]

In our study we included 353 Fuhrman grading RCCs and 77 WHO/ISUP grading RCCs. The model was trained and validated on Fuhrman grading RCCs, and achieved a comparative high accuracy in the WHO/ISUP test set, supporting the generalizability of our approach on the WHO/ISUP grading system. RCC consists of different subtypes that are unique, with their own genetic, biological, and histological characteristics.[37] Most radiologic studies only focused on the clear-cell subtype. In the study conducted by Cornelis et al, no significant association between Fuhrman grade and radiologic feature was found for papillary or chromophobe RCCs.[19] This may be related to the fact that the Fuhrman grade system is better applied to clear-cell RCCs than to other subtypes.[38–40]

Limitations

First, selection bias may exist because we only included all RCC tumors with pathological-proven histological grades. Second, only conventional MRI was used. Inclusion of advanced MR technologies such as DWI, BOLD-MRI, and SWI may further increase model accuracy. Third, manual segmentations were performed in this study by a radiologist with 5 years of experience reading abdominal MRI. Automatic segmentation based on deep learning has been successfully performed in other organs[41,42] and kidney volume on CT with comparable accuracy to expert segmentation.[26] However, deep-learning segmentation of renal tumor on MRI is very challenging due to the small size of these tumors and other organs in the abdomen that have similar intensity to the tumors. We have identified no study in the literature reporting such successful segmentation. Last, a large amount

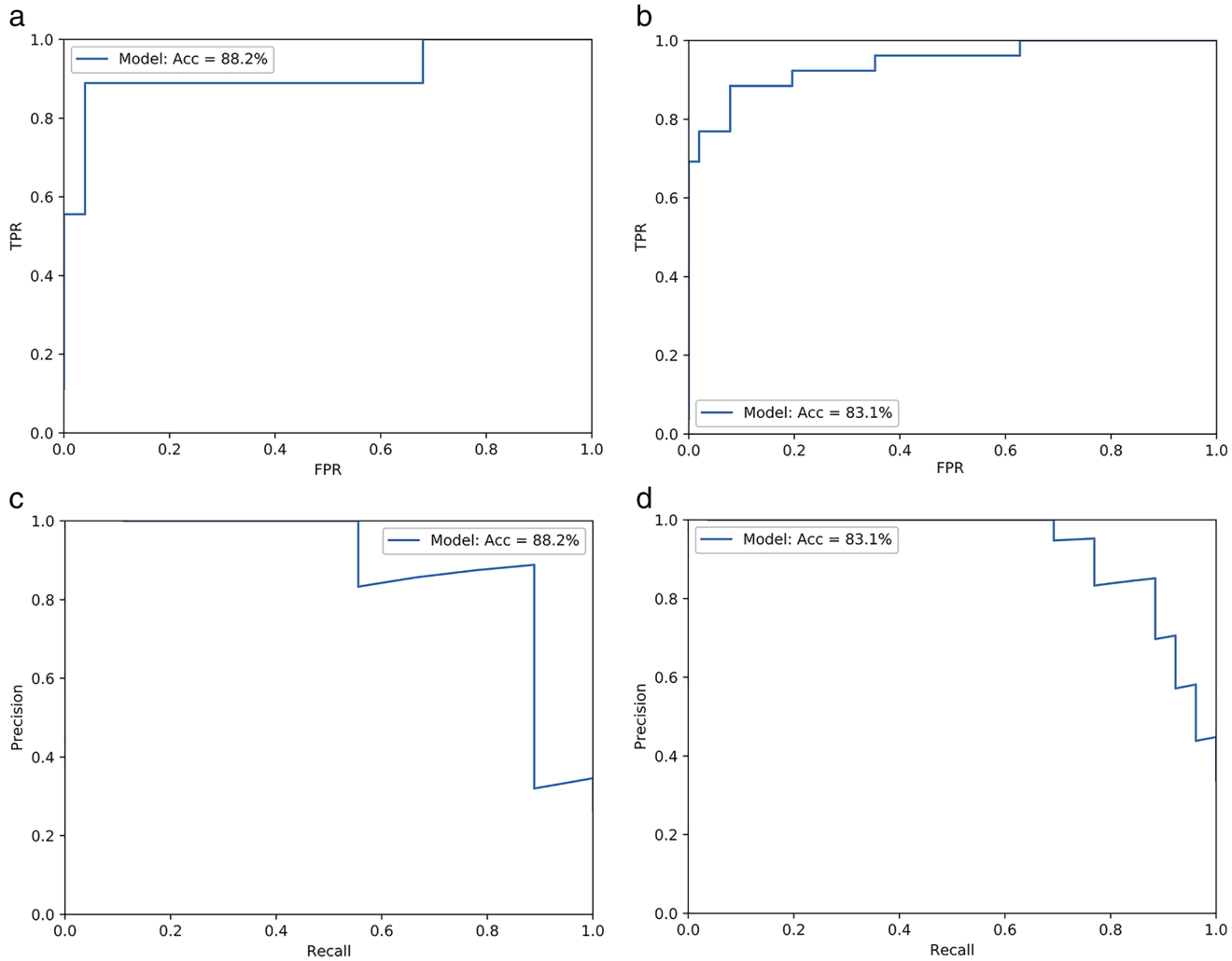

FIGURE 3: ROC curve (a,b) and Precision-Recall curve (c,d) for the final model on the Fuhrman test set and the WHO/ISUP test set.

data is essential for training of deep neural networks. Although data augmentation partially overcame this challenge, algorithm development would benefit from a larger patient cohort, especially considering the heterogeneity of different image acquisition parameters at different institutions.

Conclusion

In this study a deep-learning-based model using conventional MRI was developed to noninvasively differentiate low from high histological grade in stages I and II RCC with high accuracy. If further validated, the model can help guide management in a clinical setting, and better triage patients for active surveillance vs. treatment with ablation or resection.

Acknowledgments

We thank Sukhdeep Khurana, Aidan McGirr, An Xie, and Jianbin Liu for help in data collection.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013;63(1):11-30.
2. Cooperberg MR, Mallin K, Ritchey J, Villalta JD, Carroll PR, Kane CJ. Decreasing size at diagnosis of stage 1 renal cell carcinoma: Analysis from the national cancer data base, 1993 to 2004. J Urol 2008;179(6): 2131-2135.
3. Jung EJ, Lee HJ, Kwak C, Ku JH, Moon KC. Young age is independent prognostic factor for cancer-specific survival of low-stage clear cell renal cell carcinoma. Urology 2009;73(1):137-141.
4. Jonasch EGJ, Rathmell WK. Renal cell carcinoma. Isr Med Assoc J 2014;349(4):249-259.
5. Ficarra V, Galfano A, Mancini M, Martignoni G, Artibani W. TNM staging system for renal-cell carcinoma: Current status and future perspectives. Lancet Oncol 2007;8(6):554-558.
6. Cooperberg MR, Mallin K, Kane CJ, Carroll PR. Treatment trends for stage I renal cell carcinoma. J Urol 2011;186(2):394-399.
7. Andre B, Sebastien C, David C, Georges-Pascal H, Gill IS. Minimally invasive nephron-sparing surgery. Urol Clin North Am 2008;30(3): 551-579.
8. Boorjian SA, Blute ML. Intermediate comparison of partial nephrectomy and radiofrequency ablation for clinical T1a renal tumors. Br J Urol Int 2007;26(1):102.

9. Guazzoni G, Cestari A, Buffi N, et al. Oncologic results of laparoscopic renal cryoablation for clinical T1a tumors: 8 years of experience in a single institution. Urology 2010;76(3):624-629.
10. Heuera R, Guazzoni G, Kirkali Z, Marberger M, Richie JP, JJMCHdl R. A critical analysis of the actual role of minimally invasive surgery and active surveillance for kidney cancer. Eur Urol 2010;57(2):223-232.
11. Medeiros LJ, Gelb AB, Weiss LM. Renal cell carcinoma. Prognostic significance of morphologic parameters in 121 cases. Cancer 1998;61(8): 1639-1651.
12. Andrea M, Leon L, Riccardo M, Cesare S. Prognostic value of nuclear grading in patients with intracapsular (pT1-pT2) renal cell carcinoma. Long-term analysis in 213 patients. Cancer 2010;94(10):2590-2595.
13. Bretheau D, Lechevallier E, De FM, Sault MC, Rampal M, Coulange C. Prognostic value of nuclear grade of renal cell carcinoma. Cancer 1996; 76(12):2543-2549.
14. Erdoğan F, Demirel A, Polat O. Prognostic significance of morphologic parameters in renal cell carcinoma. Int J Clin Pract 2004;58(4):333-336.
15. Novara G, Martignoni G, Artibani W, Ficarra V. Grading systems in renal cell carcinoma. J Urol 2007;177(2):430-436.
16. Moch HAM, Delahunt B. Tumours of the kidney. In: HP MH, Ulbright T, Reuter VE, editors. *Pathology and genetics: Tumors of the urinary system and male genital organs*. Zurich, Switzerland: IARC Press; 2015. p 7-10.
17. Dagher J, Delahunt B, Rioux-Leclercq N, et al. Clear cell renal cell carcinoma: Validation of WHO/ISUP grading. Histopathology 2017;71(6): 918–925.
18. Shannon BA, Cohen RJ, De BH, Davies RJ. The value of preoperative needle core biopsy for diagnosing benign lesions among small, incidentally detected renal masses. J Urol 2008;180(4):1257-1261.
19. Cornelis F, Tricaud E, Lasserre AS, et al. Multiparametric magnetic resonance imaging for the differentiation of low and high-grade clear cell renal carcinoma. Eur Radiol 2015;25(1):24-31.
20. Vargas HA, Delaney HG, Delappe EM, et al. Multiphasic contrast-enhanced MRI: Single-slice versus volumetric quantification of tumor enhancement for the assessment of renal clear-cell carcinoma Fuhrman grade. J Magn Reson Imaging 2013;37(5):1160-1167.
21. Zhu YH, Wang X, Zhang J, Chen YH, Kong W, Huang YR. Low enhancement on multiphase contrast-enhanced CT images: An independent predictor of the presence of high tumor grade of clear cell renal cell carcinoma. Am J Roentgenol 2014;203(3):295-300.
22. Hannu H, Darryl H, Steven C, et al. CT prediction of the Fuhrman grade of clear cell renal cell carcinoma (RCC): Towards the development of computer-assisted diagnostic method. Abdom Imaging 2015;40(8): 3168-3174.
23. 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;103:51-56.
24. Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for determination of IDH status in low- and high-grade gliomas from MR imaging. Clin Cancer Res 2018;24(5):1073-1081.
25. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 2017; 19(6):862-870.
26. Sharma K, Rupprecht C, Caroli A, et al. Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. Sci Rep 2017;7(1):2049.
27. Xi IL, Zhao Y, Wang R, et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. Clin Cancer Res 2020.
28. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D SLICER as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012;30(9):1323-1341.
29. Huang G, Sun Y, Liu Z, Sedra D, Weinberger KQ. Deep networks with stochastic depth. In: European Conference on Computer Vision. Cham, Switzerland; 2016. p 646–661.
30. Srivastava GH N, Krizhevsky A. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929-1958.
31. AlanAgresti CB. Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 1998;52(2):119-126.
32. Chollet F. Keras. *GitHub repository* 2015, San Francisco: GitHub. https://github.com/fchollet/keras.
33. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. *TensorFlow: a system for large-scale machine learning*. arXiv; 2016 https://arxiv.org/pdf/1605.08695.pdf.
34. Lam JS, Shvarts O, Leppert JT, Figlin RA, Belldegrun AS. Renal cell carcinoma 2005: New frontiers in staging, prognostication and targeted molecular therapy. J Urol 2005;173(6):1853-1862.
35. Yadlapalli SB, Shi D, Vaishampayan U. Renal cell carcinoma: clinical presentation, staging, and prognostic factors. In: Lara PN, Jonasch E, editors. *Kidney Cancer: Principles and Practice*. Cham: Springer International Publishing; 2015. p. 105-121.
36. Ficarra V, Righetti R, Martignoni G, et al. Prognostic value of renal cell carcinoma nuclear grading: Multivariate analysis of 333 cases. Urol Int 2001;67(2):130-134.
37. Srigley JR, Delahunt B, Eble JN, et al. The International Society of Urological Pathology (ISUP) Vancouver classification of renal neoplasia. Am J Surg Pathol 2013;37(10):1469-1489.
38. Meskawi M, Sun M, Ismail S, et al. Fuhrman grade has no added value in prediction of mortality after partial or radical nephrectomy for chromophobe renal cell carcinoma patients. Mod Pathol 2013;26(8):1144-1149.
39. Sika-Paotonu D, Bethwaite PB, Mccredie MR, William JT, Delahunt B. Nucleolar grade but not Fuhrman grade is applicable to papillary renal cell carcinoma. Am J Surg Pathol 2006;30(9):1091-1096.
40. Delahunt B, Sika-Paotonu DP, Mccredie M, Martignoni G, Eble J, Jordan T. Fuhrman grading is not appropriate for chromophobe renal cell carcinoma. Am J Surg Pathol 2007;31(6):957-960.
41. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 2016;35(5):1240-1251.
42. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med Image Anal 2016;30:108-119.