KIDNEYS, URETERS, BLADDER, RETROPERITONEUM

Characterization of solid renal neoplasms using MRI-based quantitative radiomics features

Daniela Said[1,2] · Stefanie J. Hectors[1,3,4] · Eric Wilck[3] · Ally Rosen[3,5] · Daniel Stocker[1,6] · Octavia Bane[1] · Alp Tuna Beksaç[7] · Sara Lewis[1,3] · Ketan Badani[7] · Bachir Taouli[1,3]

Abstract

Purpose To assess the diagnostic value of magnetic resonance imaging (MRI)-based radiomics features using machine learning (ML) models in characterizing solid renal neoplasms, in comparison/combination with qualitative radiologic evaluation.

Methods Retrospective analysis of 125 patients (mean age 59 years, 67% males) with solid renal neoplasms that underwent MRI before surgery. Qualitative (signal and enhancement characteristics) and quantitative radiomics analyses (histogram and texture features) were performed on T2-weighted imaging (WI), T1-WI pre- and post-contrast, and DWI. Mann–Whitney U test and receiver-operating characteristic analysis were used in a training set ($n = 88$) to evaluate diagnostic performance of qualitative and radiomics features for differentiation of renal cell carcinomas (RCCs) from benign lesions, and characterization of RCC subtypes (clear cell RCC [ccRCC] and papillary RCC [pRCC]). Random forest ML models were developed for discrimination between tumor types on the training set, and validated on an independent set ($n = 37$).

Results We assessed 104 RCCs (51 ccRCC, 29 pRCC, and 24 other subtypes) and 21 benign lesions in 125 patients. Significant qualitative and quantitative radiomics features (area under the curve [AUC] between 0.62 and 0.90) were included for ML analysis. Models with best diagnostic performance on validation sets showed AUC of 0.73 (confidence interval [CI] 0.5–0.96) for differentiating RCC from benign lesions (using combination of qualitative and radiomics features); AUC of 0.77 (CI 0.62–0.92) for diagnosing ccRCC (using radiomics features), and AUC of 0.74 (CI 0.53–0.95) for diagnosing pRCC (using qualitative features).

Conclusion ML models incorporating MRI-based radiomics features and qualitative radiologic assessment can help characterize renal masses.

Keywords Renal mass · Renal cell carcinoma · Magnetic resonance imaging · Radiomics · Histogram · Texture

Introduction

Renal tumors encompass a large spectrum of neoplasms that vary in clinical behavior, histopathologic features, and genetic expressions. These lesions range from benign (e.g., oncocytoma, angiomyolipoma [AML]) and indolent

✉ Bachir Taouli
 bachir.taouli@mountsinai.org

[1] BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[2] Department of Radiology, Universidad de los Andes, Santiago, Chile

[3] Department of Diagnostic, Molecular and Interventional Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4] Department of Radiology, Weill Cornell Medicine, New York, NY, USA

[5] Department of Radiology, Long Island School of Medicine, NYU-Winthrop Hospital, Mineola, NY, USA

[6] Institute of Interventional and Diagnostic Radiology, University Hospital Zurich and University of Zurich, Zurich, Switzerland

[7] Department of Urology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

tumors (e.g., variants of papillary and chromophobe renal cell carcinomas [RCCs]), to aggressive malignant tumors (e.g., clear cell RCC [ccRCC]) [1–3].

Distinguishing RCC aggressive subtypes from other benign and indolent lesions has been a topic of interest. Malignant and indolent kidney tumors differ in prognosis, biological behavior, and response to available therapies [4]. Percutaneous renal biopsy of renal masses can provide pre-treatment pathologic diagnosis, with a reported accuracy ranging from 70% to 90%. However, biopsy is an invasive procedure with potential complications, and has the possibility of sampling errors and non-diagnostic analysis in up to 20% of cases [5, 6]. Non-invasive methods, such as dynamic contrast-enhanced computed tomography (CT) and contrast-enhanced magnetic resonance imaging (MRI), can provide qualitative assessment for characterizing renal masses. In particular, MRI has been well described in the evaluation of more common subtypes of RCC (ccRCC, papillary [pRCC], and chromophobe), primarily using a combination of T2-weighted imaging (T2-WI) and post-contrast T1-weighted imaging (T1-WI) [7, 8]. Despite this, determining an accurate diagnosis is still challenging, mainly because of known histologic and molecular heterogeneity within different renal neoplasms, subtypes of RCC, and within a single tumor [9, 10].

Radiomics is an emerging field that intends to extract maximal information from standard of care images, to provide information beyond what can be achieved from human imaging interpretation alone [11, 12]. Radiomics features can be divided into semantic (like size, shape, location, and necrosis), and agnostic features (quantitative features defined by an advanced mathematically algorithm) [12, 13]. Statistical features are part of agnostic features, including first-order, second-order, and higher order radiomics features [13, 14]. First-order features (histogram) describe the distribution of pixel intensity values [15]. Second-order features (texture) describe the spatial relationships of voxels. Gray-level co-occurrence matrix (GLCM) texture features introduced by Haralick et al. are the most often analyzed features [16]. Higher order radiomics methods impose filters on the images to extract patterns [13].

Large datasets of features can be extracted from a single image with radiomics. These features can be used in machine learning (ML) algorithms and potentially aid tumor detection, diagnosis, assessment of prognosis, prediction of response to treatment, and monitoring disease status [17].

Machine learning is a subfield of artificial intelligence in which algorithms or classifiers learn patterns from large databases, to generate valuable predictive outputs [18, 19]. In general, ML uses a training set of examples to perform tasks such as feature selection and parameter fitting, and a validation or test set to evaluate the model performance [20]. Random Forest (RF) is a type of ML model known as

a highly effective algorithm for classification. In RF algorithms, multiple decision trees are combined to get strong and robust models, minimizing training errors, and allowing generalization to new datasets [20, 21]. Application of ML models to clinical practice may have beneficial implications, improving healthcare quality and safety [22].

The aim of our study was to assess the diagnostic value of MRI-based radiomics features using ML models in characterizing solid renal neoplasms, in comparison/combination with qualitative evaluation.

Materials and methods

Patient selection

The local institutional review board approved this retrospective single-center study, and HIPAA compliance was maintained throughout the study period. Our center's urological database was queried between August 2015 and June 2019 using the search terms "renal mass," "renal cell carcinoma," "RCC," "nephrectomy," and "partial nephrectomy." This search identified a total of 152 patients with renal masses with preoperative MRI and pathologic correlation. Inclusion criteria were adult (> 18 years) patients with (1) renal mass that underwent total or partial nephrectomy, (2) patients who underwent abdominal MRI before surgery, and (3) patients with solid or mixed lesions (solid and cystic). Exclusion criteria were (1) cystic lesions without any solid component and (2) lesion size less than 1 cm. One hundred twenty-five patients (84 M/41F; mean age 58.8 ± 11.5 years, range 20–83 years) met the inclusion criteria and comprise our study population. Average time between MRI and surgery was 78.1 ± 91 days (range 4–808 days). A flowchart of patient inclusion is shown in Fig. 1.

MRI examination

Imaging studies were performed with multichannel MRI systems including 1.5T ($n = 95$; Aera, Espree, Symphony, Amira, Siemens Healthineers; Signa HD, HDxt and Optima, GE medical systems) or 3T ($n = 30$; Skyra and Verio, Siemens Healthineers; Signa HDxt, Discovery 750w, GE Medical Systems) imaging platforms. All patients were imaged in supine position using phased array torso coils. Routine abdominal MRI sequences performed included axial and coronal single-shot fast spin echo T2-WI (HASTE/SSFSE), axial fat-suppressed (FS) fast spin echo T2-WI, diffusion-weighted imaging (DWI, at b values of 50, 400, and 800 s/mm^2) with apparent diffusion coefficient [ADC] maps), and dynamic multiphase T1-WI. For dynamic contrast-enhanced imaging, unenhanced, arterial phase (AP), nephrographic phase (NP), and delayed phase (DP) were obtained using
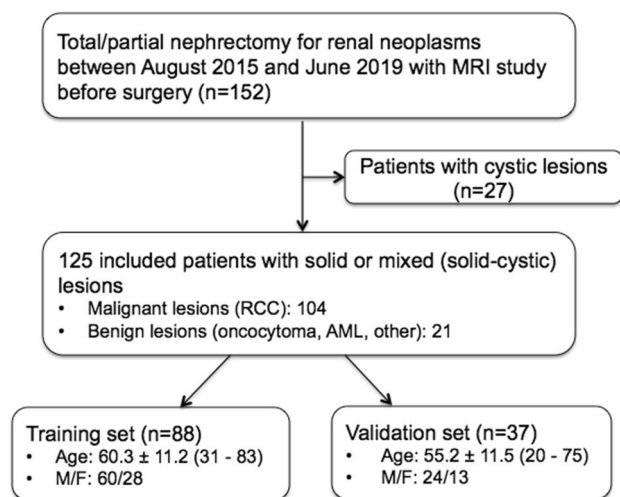
Total/partial nephrectomy for renal neoplasms between August 2015 and June 2019 with MRI study before surgery (n=152)

Patients with cystic lesions (n=27)

125 included patients with solid or mixed (solid-cystic) lesions
• Malignant lesions (RCC): 104
• Benign lesions (oncocytoma, AML, other): 21

Training set (n=88)
• Age: 60.3 ± 11.2 (31 - 83)
• M/F: 60/28

Validation set (n=37)
• Age: 55.2 ± 11.5 (20 - 75)
• M/F: 24/13

Fig. 1 Flowchart of patient population

3D T1-WI fat-suppressed spoiled gradient-recalled echo sequence (VIBE/LAVA) before and after the administration of a gadolinium-based contrast agent (Gadobutrol [Gadavist, Bayer Healthcare]; Gadoterate meglumine [Dotarem, Guerbet]; Gadobenate dimeglumine [Multihance, Bracco Diagnostics]; Gadopentetate dimeglumine [Magnevist, Bayer Healthcare]; Gadoversetamide [Optimark, Guerbet]; Gadoteridol [Prohance, Bracco Diagnostics]; Gadoxetate disodium [Eovist, Bayer Healthcare]; Gadodiamide [Omniscan, GE Healthcare] or unknown). Eighty-four (67.2%) patients had an outside MRI; 4 of them did not have contrast-enhanced images and 33 patients did not have DWI.

Image analysis

Qualitative analysis

The index lesion was defined as the largest pathologically confirmed solid or mixed (solid and cystic) renal mass. Two independent radiologists (with 20 and 14 years of experience in abdominal MRI) identified the index lesion and performed qualitative evaluation, using PACS (Centricity 3.0, General Electric Medical Systems). The observers were aware that the patient had a renal mass; however, they were blinded to the final diagnosis. The evaluation consisted of filling out a form that included tumor size, laterality (right or left), margins (well-defined or ill-defined), composition (solid, cystic or mixed), presence of hemosiderin, tumor fat content, growth pattern (endophytic, < 50% exophytic or > 50% exophytic), collecting system invasion, renal vein invasion, contrast enhancement on AP (hypovascular [lower enhancement compared to renal cortex] or hypervascular [same or higher enhancement compared to renal cortex]), heterogeneity on NP (homogeneous or

heterogeneous nodular enhancement), chemical shift, and T1-WI, T2-WI, diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC) map signal (hypointense, hyperintense, isointense, or heterogeneous) [23]. Presumed benign or malignant appearance of the tumor was also recorded based mainly on signal intensity of the lesion on T2-WI, degree of contrast enhancement on AP, and presence of intra-lesional fat on in- and out-of-phase T1-WI [24].

Quantitative analysis

The same index lesions were evaluated for quantitative analysis. T2-WI, DWI/ADC, and T1-WI pre-and post-contrast (AP, NP, and DP) images were analyzed using OsiriX (Pixmeo SARL, Bernex, Switzerland) software by a third radiologist (with 3 years of experience in body imaging). Regions of interest (ROIs) were placed in the previously defined index lesion including almost the entire area of the solid portion of tumors, avoiding cystic areas and the most peripheral portions to exclude partial volume effects, using OsiriX software. One single-slice ROI was placed in the largest tumor area if the lesion was < 3 cm, and two ROIs were placed on two consecutive slices if the lesion was ≥ 3 cm. Signal intensity (SI) in the lesion ROI was normalized by the mean SI in an ROI placed in the uninvolved renal cortex of the ipsilateral kidney (area 90–100 mm^2), with the purpose to decrease SI variation due to heterogeneity in MR acquisition protocols [25]. Mean enhancement ratios were extracted from the ROIs during AP, NP, and DP.

Radiomics analysis was performed by an MRI physicist (with 4 years of experience), utilizing in-house developed software with MATLAB (MathWorks, Inc. Natick, MA). For each sequence, histogram features (first-order radiomics) included central tendency parameters (mean and median) and heterogeneity parameters (standard deviation (SD), kurtosis, and skewness), with a total of 50 histogram features extracted for each lesion. Fourteen Haralick texture features were calculated from the gray-level co-occurrence matrix (GLCM) for each sequence also, with a total of 140 texture features per lesion. Gray-level intensity data were normalized before the texture extraction in order to standardized the signal intensity range to 0–64. The lists of histogram and texture features are included in Supplemental Fig. 1. Illustration of radiomics workflow is shown in Fig. 2.

Histopathologic analysis

Histopathologic tissue confirmation for tumor type/subtype of each lesion was extracted from the pathology report of partial (82.4%; $n = 103$) or total nephrectomy (17.6%; $n = 22$).
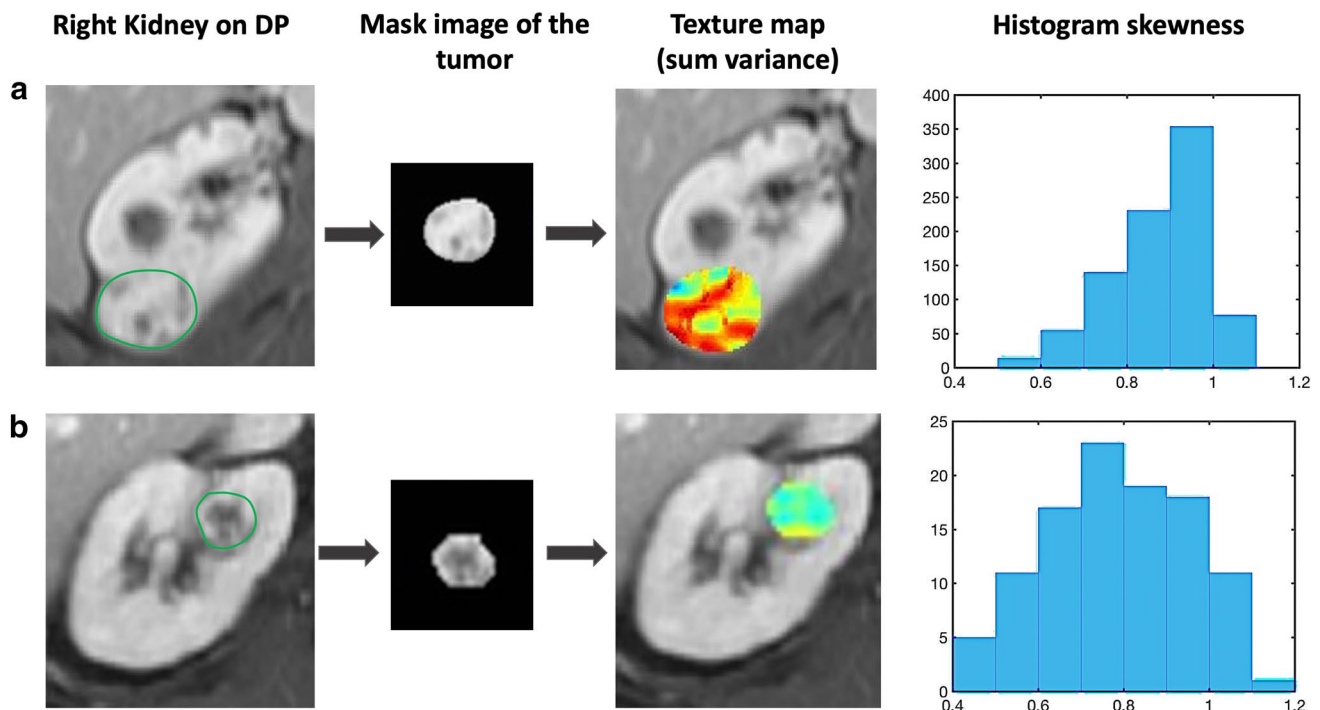
Right Kidney on DP | **Mask image of the tumor** | **Texture map (sum variance)** | **Histogram skewness**

Fig. 2 Illustration of radiomics workflow assessment in two different patients. The figure shows post-contrast T1-weighted images obtained on delayed phase with regions of interest placed on the lesion, mask image of the tumor, texture map (sum variance), and histogram skewness distribution of the lesion. **a** A 32-year-old female patient with right clear cell renal cell carcinoma (ccRCC). **b** A 42-year-old male with right papillary renal cell carcinoma (pRCC) with atypical magnetic resonance imaging (MRI) appearance (hypervascular tumor). Both lesions were characterized as hypervascular malignant tumors, with heterogeneous enhancement. Texture map images showed higher sum variance (texture feature) on ccRCC compared to pRCC. Histogram figures showed skewed distribution to the left on ccRCC, while pRCC showed a more symmetrical histogram distribution. This is a representative example where radiomics features helped differentiating these RCC subtypes

Statistical analysis

Clinical and demographic data were summarized using descriptive statistics expressed as mean ± standard deviation (SD). The study sample was randomly divided in a training set (70%; $n = 88$) and a validation set (30%; $n = 37$) [26]. Chi-square test was used to compare demographics and clinical data of both sets.

We built three classification categories, one to differentiate malignant from benign lesions. The second and third categories were used to differentiate ccRCC and pRCC from all other renal lesions, respectively; given that these two subtypes are the most common [27]. Mann–Whitney U test was used to assess differences in qualitative and quantitative radiomics features among the three classifications, and receiver-operating characteristic (ROC) curves were generated to assess the diagnostic performance of statistically significant features in the training set. ML modeling was performed using qualitative, quantitative radiomics, and combination of features for prediction of tumor diagnosis (RCC, ccRCC, and pRCC) using random forest on MATLAB statistical tools. Only features that showed significance for the

Mann–Whitney U test on the training set were selected as input for the models. Sensitivity and specificity for the models were calculated using Youden index. Missing data were filled using an interpolation imputation method. All statistical tests were conducted in MATLAB (MathWorks, Natick, MA) and SPSS (IBM, Armonk NY). For all tests, a p value below 0.05 was considered statistically significant.

Results

Study population

Our patient population included 104 patients with RCC (83.2%) and 21 with benign lesions (16.8%). RCC subtypes ($n = 104$) included ccRCC ($n = 51$, 49.1%), pRCC ($n = 29$, 27.9%), chromophobe RCC ($n = 12$, 11.5%), and other less common subtypes ($n = 12$, 11.5%). Benign lesions ($n = 21$) consisted of fat poor AML ($n = 8$, 38.1%), oncocytoma ($n = 11$, 52.3%), solitary fibrous tumor ($n = 1$, 4.8%), and mixed epithelial and stromal tumor ($n = 1$, 4.8%). Mean tumor size at pathology was 3.5 cm ± 2.6 (1.0–13 cm).

Demographics and clinical parameters of the patient population are described in Table 1.

Qualitative analysis

Mean tumor size on imaging was 3.5 cm ± 2.7 cm (range 1.0–12.2 cm). Sixty-three tumors were located in the right kidney (50.4%), and sixty-two in the left kidney (49.6%).

On the training set, statistically significant differences for distinguishing between RCC and benign lesions were observed for contrast enhancement on AP ($p = 0.002$; AUC 0.72, $p = 0.001$), and presumed histologic subtype based on enhancement and T2 signal ($p = 0.002$; AUC 0.64, $p = 0.025$).

For differentiating ccRCC from other lesions there were significant differences on contrast enhancement on AP ($p = 0.005$; AUC 0.66, $p = 0.008$); heterogeneity on NP ($p < 0.001$; AUC 0.71, $p < 0.001$); presence of cystic components ($p = 0.007$; AUC 0.64, $p = 0.010$); signal intensity on T2-WI ($p < 0.001$; AUC 0.68, $p = 0.002$); and ADC map signal ($p = 0.019$; AUC 0.65, $p = 0.030$). Presumed benign or malignant MRI appearance showed to be significant in the Mann–Whitney U test ($p = 0.011$), but AUC analysis was not significant, with a value of 0.59 ($p = 0.079$). When comparing pRCC with other lesions, we found significant differences in growth pattern (presence of exophytic component, $p = 0.036$; AUC 0.64, $p = 0.025$); heterogeneity on NP ($p = 0.003$; AUC 0.68, $p = 0.002$); signal intensity on T2-WI ($p < 0.001$; AUC 0.71, $p < 0.001$); contrast enhancement on AP ($p < 0.001$; AUC 0.75, $p < 0.001$); and ADC map signal

Table 1 Demographics and tumor characteristics of our study population

Variable	Training set ($n = 88$)	Validation set ($n = 37$)	Total ($n = 125$)	p***
Age, years (mean, SD)	60.3 (11.2)	55.2 (11.5)	58.8 (11.5)	0.024
Gender, male (%)	60 (68.2%)	24 (64.9%)	84 (67.2%)	0.835
BMI, kg/m^2 (mean, SD)	29.1 (6.2)	28.1 (5.8)	28.8 (6.1)	0.400
Outside MRI, n (%)	62 (70.5%)	22 (59.5%)	84 (67.2%)	0.398
Field strength (T) n (%)				0.324
1.5T	65 (73.9%)	30 (81.1%)	95 (76%)	
3T	23 (26.1%)	7 18.9%)	30 (24%)	
Tumor location				0.561
Right	46 (52.3%)	17 (45.9%)	63 (50.4%)	
Left	42 (47.7%)	20 (54.1%)	62 (49.6%)	
Tumor type, n (%)[a]				0.474
RCC	71 (80.7%)	33 (89.2%)	104 (83.2%)	
AML	5 (5.7%)	3 (8.1%)	8 (6.4%)	
Oncocytoma	10 (11.4%)	1 (2.7%)	11 (8.8%)	
Other benign	2 (2.3%)	0 (0.0%)	2 (1.6%)	
RCC subtype, n (%)[b]				0.366
Clear cell	30 (42.3%)	21 (63.4%)	51 (49.1%)	
Papillary type 1	18 (25.4%)	8 (24.2%)	26 (25%)	
Papillary type 2	3 (4.2%)	0 (0.0%)	3 (2.9%)	
Chromophobe	10 (14.1%)	2 (6.1%)	12 (11.5%)	
Clear cell-papillary	7 (9.9%)	2 (6.1%)	9 (8.7%)	
Other	3 (3.4%)	0 (0.0%)	3 (2.4%)	
Pathological tumor size, cm (mean, SD)	3.5 (2.5)	3.5 (2.8)	3.5 (2.6)	0.990
Radiologic tumor size, cm (mean, SD)	3.4 (2.6)	3.7 (3.1)	3.5 (2.7)	0.590
Surgery type, n (%)				0.120
Partial nephrectomy	76 (86.4%)	27 (72.9%)	103 (82.4%)	
Total nephrectomy	12 (13.6%)	10 (27.1%)	22 (17.6%)	

RCC renal cell carcinoma, *AML* angiomyolipoma, *SD* standard deviation, *T* tesla, *BMI* body mass index, *MRI* magnetic resonance imaging

***Chi-square test or Mann–Whitney U test (where appropriate) for comparison between training and validation sets

[a]Malignant tumors $n = 71$, benign tumors $= 17$ in the training set; malignant tumors $n = 33$ and benign tumors $n = 4$ in the validation set ($p = 0.474$)

[b]Calculated from a total of 71 and 33 RCC in training and validation sets, respectively

$(p = 0.021$; AUC 0.65, $p = 0.02$). Significant qualitative features are summarized in Table 2.

Quantitative analysis

On the training set, five texture features were found to be significant for differentiating benign from malignant tumors, with highest AUC of 0.81 ($p < 0.001$) for ADC homogeneity. Representative cases are shown in Fig. 3.

When comparing ccRCC versus other renal lesions we found significant differences in 22 histogram features, 9 texture features, and mean enhancement ratios. Standard deviation on AP and mean signal intensity in DWI-b800 were the most significant with AUCs of 0.84 ($p < 0.001$) and 0.90 ($p < 0.001$), respectively.

When comparing pRCC versus other tumors, there were significant differences in 19 histogram features, 13 texture features, and enhancement ratios. The highest AUC was 0.85 ($p < 0.001$); obtained in mean and median on AP. Box plots of significant features examples are shown in Fig. 4. Details of all significant features on the training set are included in Supplemental Table 1.

ML random forest models for qualitative features, quantitative radiomics features, and combination of features

Three models were generated for predicting RCC, ccRCC, and pRCC, using qualitative, quantitative, and combination of qualitative and quantitative features.

Only one ML model was significant for differentiating RCC from benign lesions, using combination of qualitative and quantitative radiomics features, with AUC of 0.97 (95% confidence intervals [CI] 0.94–1.00, $p < 0.001$) on the training set (93% sensitivity and 94% specificity), and AUC of 0.73 (95% CI 0.50–0.96, $p = 0.002$) on the validation set (58% sensitivity and 100% specificity).

For diagnosing ccRCC versus all other lesions, radiomics-based ML model had the highest diagnostic performance with AUC of 1.0 (95% CI 1.00–1.00, $p < 0.001$) on the training set (100% sensitivity and 100% specificity), and 0.77 (95% CI 0.62–0.92, $p < 0.001$) on the validation set (95% sensitivity and 56% specificity).

For differentiating pRCC from all other renal tumors, the model with highest AUC was generated using qualitative features, with AUC of 0.91 (95% CI 0.82–0.99, $p < 0.001$) on the training set (91% sensitivity and 81% specificity), and 0.74 (95% CI 0.53–0.95, $p = 0.010$) on the validation set (75% sensitivity and 69% specificity). Complete results of multivariate modeling are summarized in Table 3.

Table 2 Qualitative imaging features in the training set (n=88)

Feature	Description	RCC vs. benign			ccRCC vs. other			pRCC vs. other		
		RCC (n=71)	Benign (n=17)	p	ccRCC (n=30)	Other (n=58)	p	pRCC (n=21)	Other (n=67)	p
CE on AP	Hypervascular	27 (38%)	14 (82%)	**0.002**	20 (69%)	21 (37%)	**0.005**	2 (9%)	39 (60%)	**< 0.001**
Composition	Cystic component	23 (32%)	2 (12%)	0.060	14 (47%)	11 (19%)	**0.007**	3 (14%)	22 (32%)	0.103
Heterogeneity on NP	Nodular enhancement	28 (40%)	4 (23.5%)	0.197	19 (66%)	13 (23%)	**< 0.001**	2 (9%)	30 (46%)	**0.003**
T2 signal	Hyperintensity	47 (66%)	11 (65%)	0.912	27 (90%)	31 (53%)	**< 0.001**	7 (33%)	51 (76%)	**< 0.001**
ADC	High ADC (on visual assessment)	16 (31%)	1 (14.3%)	0.378	10 (48%)	7 (18%)	**0.019**	1 (6%)	16 (37%)	**0.021**
Growth pattern	Exophytic	54 (76%)	4 (23%)	0.165	22 (73%)	12 (21%)	0.508	17 (81%)	20 (30%)	**0.036**
Presumed pathology	Malignant	66 (93%)	11 (65%)	**0.002**	30 (100%)	47 (81%)	**0.011**	20 (95%)	57 (85%)	0.225

Significant variables for the three groups are included in the table (significant p values in bold). Description of variables includes positive result based on statistical analysis. p value of Mann–Whitney U test

RCC renal cell carcinoma, *ccRCC* clear cell renal cell carcinoma, *pRCC* papillary renal cell carcinoma, *CE* contrast enhancement, *AP* arterial phase, *NP* nephrographic phase, *ADC* apparent diffusion coefficient
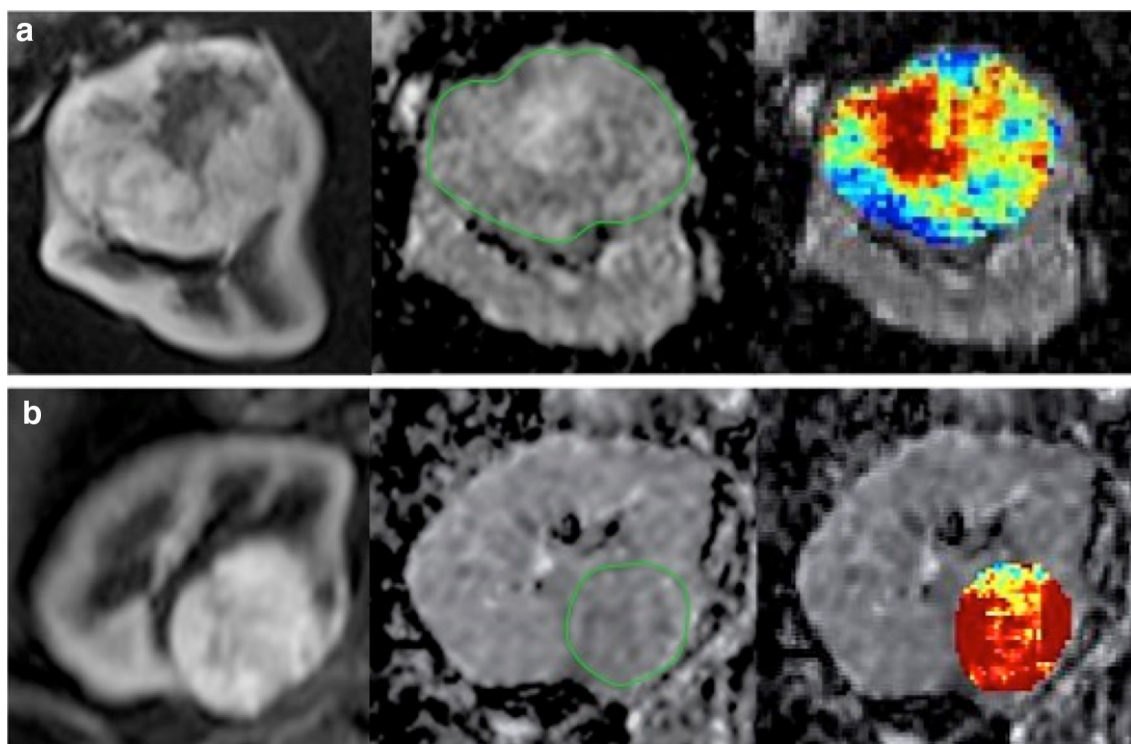
Fig. 3 Representative examples of apparent diffusion coefficient (ADC) homogeneity texture maps helping differentiating oncocytoma from clear cell renal cell carcinoma (ccRCC). T1 post-contrast on arterial phase, ADC map, and texture ADC homogeneity map in **a** a 51-year-old male with left oncocytoma, and **b** a 69-year-old female with right ccRCC; showing higher ADC homogeneity in ccRCC

Discussion

Our study found associations between qualitative and quantitative imaging features with different types of renal neoplasms. Based on informative features, we generated ML models to predict RCC, ccRCC, and pRCC diagnosis. Best diagnostic performances were observed in quantitative radiomics model for differentiating ccRCC, qualitative features model for predicting pRCC, and combination of features model for discriminating malignant lesions. Our study suggests that MRI-based radiomics features may improve characterization of solid renal neoplasms.

There is an increasing amount of evidence supporting the utility of radiomics [28–30].

Medical images are becoming a valuable source of data, and quantitative radiomics features may be used as a non-invasive tool for lesion characterization and classification [11]. There are several studies showing promising results of the utility of radiomics features in differentiating histologic subtypes in renal neoplasms. One study found similar performance to our results using MRI radiomics features in renal tumors evaluation [25]. Several informative features they found were also significant in our study, including skewness on AP and DP, DP variance, and DP sum average. Also, the authors generated models using cross-validation and random forest, showing 79% accuracy for differentiating RCC and ccRCC from oncocytoma, and 78% accuracy for distinguishing pRCC from ccRCC. This study differs from ours because they compared RCC and ccRCC only with oncocytomas as benign lesions, and results were not correlated to qualitative evaluation.

Another study reported that several MRI texture features had excellent diagnostic performance in differentiating ccRCC from non-ccRCC (AUC > 0.8), including T2-WI entropy, DWI standard deviation at b-500 and b-1000, ADC mean, and skewness on T1-WI and on AP [28]. From these features, we agree on ADC mean as an informative feature that showed high AUC value (0.80) when comparing ccRCC to other tumors.

Recently, a study proposed radiologic-radiomics ML models for differentiation of benign and malignant solid renal masses [31]. They extracted CT-based radiomics features in different renal tumors and compared it also to clinical radiologic evaluation. In concordance with our study, they found that models incorporating radiologic assessment and radiomics features may help in the differentiation of different renal solid tumors, with sensitivity up to 90%, and specificity up to 91.7%.

Our results are relevant to clinical practice because of the heterogeneous nature of renal tumors, it is necessary

Fig. 4 Box plot distributions of significant radiomics features. Renal cell carcinomas (RCCs) vs. other; RCCs showed higher texture values (homogeneity and maximal correlation) on diffusion-weighted image (DWI [low *b* value]) and apparent diffusion coefficient (ADC). Clear cell RCC (ccRCC) vs. other; ccRCC showed higher standard deviation on arterial phase (AP), and lower median values on DWI (high *b* value). Papillary RCC (pRCC) vs. other; pRCC showed lower mean and median values on AP and nephrographic phase (NP), respectively
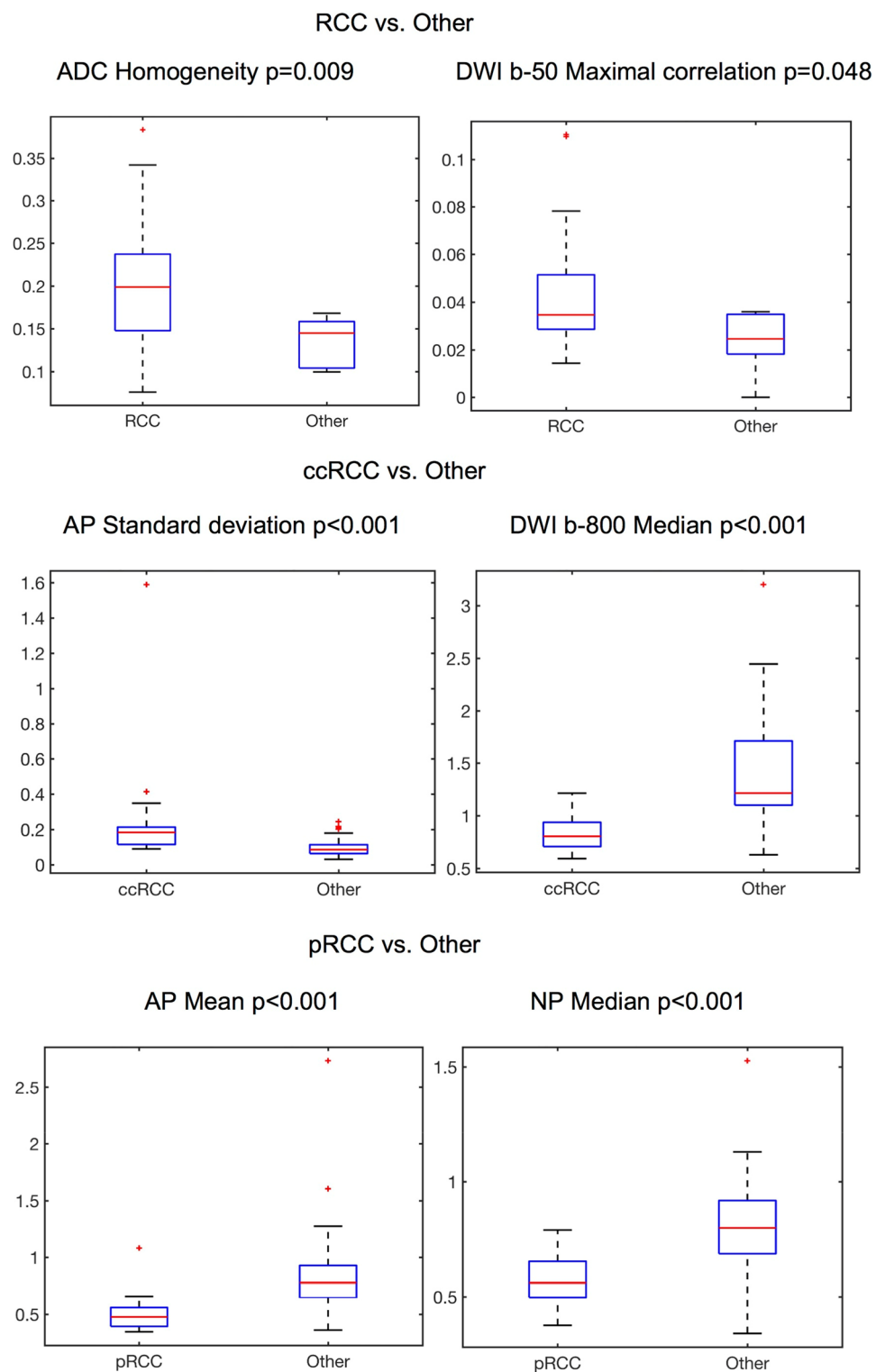
to better characterize this disease [10]. Radiomics analysis can add complementary information for tumor characterization, which may not be perceptible to human eye [12]. These features can be used to generate ML models and aid radiological evaluation of renal tumors non-invasively. Also, in the era of personalized medicine, radiomics in combination with histopathologic, genetic, and metabolic datasets may help to improve patient management, and might be used as a biomarker that would help in tumor characterization, treatment selection, and prognosis [32, 33].

Table 3 Diagnostic performance of machine learning models based on qualitative and quantitative radiomics features

Category	Set	AUC _p_ value Sensitivity Specificity	Qualitative	Quantitative	Combination
RCC vs. benign	Training	AUC (95% CI)	0.76 (0.65–0.87)	0.96 (0.92–1.00)	0.97 (0.94–1.00)
		_p_	**< 0.001**	**< 0.001**	**< 0.001**
		Sensitivity	80.3%	90.1%	93.0%
		Specificity	64.7%	94.1%	94.1%
	Validation	AUC (95% CI)	0.5 (0.2–0.81)	0.63 (0.36–0.9)	0.73 (0.5–0.96)
		_p_	0.5	0.17	**0.02**
		Sensitivity	18.1%	57.6%	57.6%
		Specificity	100%	100%	100%
ccRCC vs. other	Training	AUC (95% CI)	0.85 (0.75–0.94)	1.0 (1.0–1.00)	1.0 (1.0–1.00)
		_p_	**< 0.001**	**< 0.001**	**< 0.001**
		Sensitivity	76.7%	100%	100%
		Specificity	77.6%	100%	100%
	Validation	AUC (95% CI)	0.72 (0.56–0.89)	0.77 (0.62–0.92)	0.74 (0.58–0.90)
		_p_	**0.004**	**< 0.001**	**0.002**
		Sensitivity	61.9%	95.2%	95.2%
		Specificity	75.0%	56.3%	56.3%
pRCC vs. other	Training	AUC (95% CI)	0.91 (0.82–0.99)	1.0 (0.98–1.00)	1.0 (0.98–1.00)
		_p_	**< 0.001**	**< 0.001**	**< 0.001**
		Sensitivity	90.5%	95.2%	100%
		Specificity	80.6%	100%	95.5%
	Validation	AUC (95% CI)	0.74 (0.53–0.95)	0.73 (0.51–0.94)	0.72 (0.51–0.94)
		_p_	**0.01**	**0.02**	**0.02**
		Sensitivity	75.0%	62.5%	75.0%
		Specificity	69.0%	82.8%	72.4%

Significant _p_ values in bold

RCC renal cell carcinoma, _ccRCC_ clear cell carcinoma, _pRCC_ papillary RCC, _AUC_ area under the curve, _CI_ confidence interval

Our findings have several future applications such as facilitating workflow in busy clinical settings. Further studies are needed to incorporate these ML models in deep learning and AI algorithms including tasks like lesion detection/segmentation, characterization, and prediction of diagnosis and treatment response [34, 35]. Future studies should also evaluate the use of radiomics and ML models to assess treatment response on minimally invasive treatment modalities, such as ablative therapies, which have been used as an alternative treatment to surgery for T1a stage RCC, especially for medically fragile patients [36].

Our study has several limitations. It was conducted in a single center and the sample size was relatively small, with only 16.8% of benign lesions. This difference in numbers may affect the lack of significant features when comparing benign versus malignant tumors. Selection bias may exist because we retrospectively analyzed only patients with renal tumors that underwent surgery with prior abdominal MRI. Variability in MRI protocols/scanners may have affected the robustness and reproducibility of radiomics features [11,

37]. Results can also be affected by the fact that segmentation of the lesion was performed by one observer, and we did not measure inter-observer variability [11, 38]. Our results can potentially be improved in the future, with a larger and multicentric cohort of patients, standardization in MRI protocols, and development of new ML techniques.

In conclusion, this study showed that ML models incorporating MRI-based radiomics features and qualitative radiologic assessment can help characterize renal masses.

Compliance with ethical standards

References

1. Leibovich BC, Lohse CM, Crispen PL, Boorjian SA, Thompson RH, Blute ML, Cheville JC (2010) Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. J Urol 183 (4):1309-1315. https://doi.org/10.1016/j.juro.2009.12.035

2. Reuter VE, Presti JC, Jr. (2000) Contemporary approach to the classification of renal epithelial tumors. Semin Oncol 27 (2):124-137

3. Collins S, McKiernan J, Landman J (2006) Update on the epidemiology and biology of renal cortical neoplasms. J Endourol 20 (12):975-985. https://doi.org/10.1089/end.2006.20.975

4. Hsieh JJ, Le V, Cao D, Cheng EH, Creighton CJ (2018) Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. J Pathol 244 (5):525-537. https://doi.org/10.1002/path.5022

5. Lebret T, Poulain JE, Molinie V, Herve JM, Denoux Y, Guth A, Scherrer A, Botto H (2007) Percutaneous core biopsy for renal masses: indications, accuracy and results. J Urol 178 (4 Pt 1):1184-1188; discussion 1188. https://doi.org/10.1016/j.juro.2007.05.155

6. Silverman SG, Gan YU, Mortele KJ, Tuncali K, Cibas ES (2006) Renal masses in the adult patient: the role of percutaneous biopsy. Radiology 240 (1):6-22. https://doi.org/10.1148/radiol.2401050061

7. Sun MR, Ngo L, Genega EM, Atkins MB, Finn ME, Rofsky NM, Pedrosa I (2009) Renal cell carcinoma: dynamic contrast-enhanced MR imaging for differentiation of tumor subtypes–correlation with pathologic findings. Radiology 250 (3):793-802. https://doi.org/10.1148/radiol.2503080995

8. Young JR, Coy H, Kim HJ, Douek M, Lo P, Pantuck AJ, Raman SS (2017) Performance of Relative Enhancement on Multiphasic MRI for the Differentiation of Clear Cell Renal Cell Carcinoma (RCC) From Papillary and Chromophobe RCC Subtypes and Oncocytoma. AJR Am J Roentgenol 208 (4):812-819. https://doi.org/10.2214/ajr.16.17152

9. Durinck S, Stawiski EW, Pavia-Jimenez A, Modrusan Z, Kapur P, Jaiswal BS, Zhang N, Toffessi-Tcheuyap V, Nguyen TT, Pahuja KB, Chen YJ, Saleem S, Chaudhuri S, Heldens S, Jackson M, Pena-Llopis S, Guillory J, Toy K, Ha C, Harris CJ, Holloman E, Hill HM, Stinson J, Rivers CS, Janakiraman V, Wang W, Kinch LN, Grishin NV, Haverty PM, Chow B, Gehring JS, Reeder J, Pau G, Wu TD, Margulis V, Lotan Y, Sagalowsky A, Pedrosa I, de Sauvage FJ, Brugarolas J, Seshagiri S (2015) Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. Nat Genet 47 (1):13-21. https://doi.org/10.1038/ng.3146

10. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366 (10):883-892. https://doi.org/10.1056/nejmoa1113205

11. de Leon AD, Kapur P, Pedrosa I (2019) Radiomics in Kidney Cancer: MR Imaging. Magn Reson Imaging Clin N Am 27 (1):1-13. https://doi.org/10.1016/j.mric.2018.08.005

12. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. Radiology 278 (2):563-577. https://doi.org/10.1148/radiol.2015151169

13. Lewis S, Hectors S, Taouli B (2020) Radiomics of hepatocellular carcinoma. Abdom Radiol (NY). https://doi.org/10.1007/s00261-019-02378-5

14. Yang F, Ford JC, Dogan N, Padgett KR, Breto AL, Abramowitz MC, Dal Pra A, Pollack A, Stoyanova R (2018) Magnetic resonance imaging (MRI)-based radiomics for prostate cancer radiotherapy. Transl Androl Urol 7 (3):445-458. https://doi.org/10.21037/tau.2018.06.05

15. Just N (2014) Improving tumour heterogeneity MRI assessment with histograms. Br J Cancer 111 (12):2205-2213. https://doi.org/10.1038/bjc.2014.512

16. Haralick R, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Transactions on systems, man, and cybernetics 11 (6):610-621

17. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J (2019) The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics 9 (5):1303-1322. https://doi.org/10.7150/thno.30309

18. Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. Science 349 (6245):255-260. https://doi.org/10.1126/science.aaa8415

19. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, Sethi RK (2019) Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev. https://doi.org/10.1007/s10143-019-01163-8

20. Deo RC (2015) Machine Learning in Medicine. Circulation 132 (20):1920-1930. https://doi.org/10.1161/circulationaha.115.001593

21. Kruppa J, Liu Y, Biau G, Kohler M, Konig IR, Malley JD, Ziegler A (2014) Probability estimation with machine learning methods for dichotomous and multicategory outcome: theory. Biom J 56 (4):534-563. https://doi.org/10.1002/bimj.201300068

22. Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2018) Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 19 (6):1236-1246. https://doi.org/10.1093/bib/bbx044

23. Shinagare AB, Vikram R, Jaffe C, Akin O, Kirby J, Huang E, Freymann J, Sainani NI, Sadow CA, Bathala TK, Rubin DL, Oto A, Heller MT, Surabhi VR, Katabathina V, Silverman SG (2015) Radiogenomics of clear cell renal cell carcinoma: preliminary findings of The Cancer Genome Atlas-Renal Cell Carcinoma (TCGA-RCC) Imaging Research Group. Abdom Imaging 40 (6):1684-1692. https://doi.org/10.1007/s00261-015-0386-z

24. Kay FU, Canvasser NE, Xi Y, Pinho DF, Costa DN, Diaz de Leon A, Khatri G, Leyendecker JR, Yokoo T, Lay AH, Kavoussi N, Koseoglu E, Cadeddu JA, Pedrosa I (2018) Diagnostic Performance and Interreader Agreement of a Standardized MR Imaging Approach in the Prediction of Small Renal Mass Histology. Radiology 287 (2):543-553. https://doi.org/10.1148/radiol.2018171557

25. Hoang UN, Mojdeh Mirmomen S, Meirelles O, Yao J, Merino M, Metwalli A, Marston Linehan W, Malayeri AA (2018) Assessment of multiphasic contrast-enhanced MR textures in differentiating small renal mass subtypes. Abdom Radiol (NY). https://doi.org/10.1007/s00261-018-1625-x

26. Mehta P, Wang CH, Day AGR, Richardson C, Bukov M, Fisher CK, Schwab DJ (2019) A high-bias, low-variance introduction to

Machine Learning for physicists. Phys Rep 810:1-124. https://doi.org/10.1016/j.physrep.2019.03.001

27. Hirsch MS, Signoretti S, Dal Cin P (2015) Adult Renal Cell Carcinoma: A Review of Established Entities from Morphology to Molecular Genetics. Surg Pathol Clin 8 (4):587-621. https://doi.org/10.1016/j.path.2015.09.003

28. Goyal A, Razik A, Kandasamy D, Seth A, Das P, Ganeshan B, Sharma R (2019) Role of MR texture analysis in histological subtyping and grading of renal cell carcinoma: a preliminary study. Abdom Radiol (NY). https://doi.org/10.1007/s00261-019-02122-z

29. Yu H, Scalera J, Khalid M, Touret AS, Bloch N, Li B, Qureshi MM, Soto JA, Anderson SW (2017) Texture analysis as a radiomic marker for differentiating renal tumors. Abdom Radiol (NY) 42 (10):2470-2478. https://doi.org/10.1007/s00261-017-1144-1

30. Bonekamp D, Kohl S, Wiesenfarth M, Schelb P, Radtke JP, Gotz M, Kickingereder P, Yaqubi K, Hitthaler B, Gahlert N, Kuder TA, Deister F, Freitag M, Hohenfellner M, Hadaschik BA, Schlemmer HP, Maier-Hein KH (2018) Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. Radiology 289 (1):128-137. https://doi.org/10.1148/radiol.2018173064

31. Sun XY, Feng QX, Xu X, Zhang J, Zhu FP, Yang YH, Zhang YD (2020) Radiologic-Radiomic Machine Learning Models for Differentiation of Benign and Malignant Solid Renal Masses: Comparison With Expert-Level Radiologists. AJR Am J Roentgenol 214 (1):W44-W54. https://doi.org/10.2214/ajr.19.21617

32. Gupta K, Miller JD, Li JZ, Russell MW, Charbonneau C (2008) Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. Cancer Treat Rev 34 (3):193-205. https://doi.org/10.1016/j.ctrv.2007.12.001

33. Shinagare AB, Krajewski KM, Braschi-Amirfarzan M, Ramaiya NH (2017) Advanced Renal Cell Carcinoma: Role of the Radiologist in the Era of Precision Medicine. Radiology 284 (2):333-351. https://doi.org/10.1148/radiol.2017160343

34. Uthoff J, Stephens MJ, Newell JD, Jr., Hoffman EA, Larson J, Koehn N, De Stefano FA, Lusk CM, Wenzlaff AS, Watza D, Neslund-Dudas C, Carr LL, Lynch DA, Schwartz AG, Sieren JC (2019) Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. Med Phys 46 (7):3207-3216. https://doi.org/10.1002/mp.13592

35. Abajian A, Murali N, Savic LJ, Laage-Gaupp FM, Nezami N, Duncan JS, Schlachter T, Lin M, Geschwind JF, Chapiro J (2018) Predicting Treatment Response to Intra-arterial Therapies for Hepatocellular Carcinoma with the Use of Supervised Machine Learning-An Artificial Intelligence Concept. J Vasc Interv Radiol 29 (6):850-857 e851. https://doi.org/10.1016/j.jvir.2018.01.769

36. Yu J, Zhang X, Liu H, Zhang R, Yu X, Cheng Z, Han Z, Liu F, Hao G, Mu MJ, Liang P (2020) Percutaneous Microwave Ablation versus Laparoscopic Partial Nephrectomy for cT1a Renal Cell Carcinoma: A Propensity-matched Cohort Study of 1955 Patients. Radiology 294 (3):698-706. https://doi.org/10.1148/radiol.2020190919

37. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I (2019) Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology 291 (1):53-59. https://doi.org/10.1148/radiol.2019182023

38. Rios Velazquez E, Aerts HJ, Gu Y, Goldgof DB, De Ruysscher D, Dekker A, Korn R, Gillies RJ, Lambin P (2012) A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. Radiother Oncol 105 (2):167-173. https://doi.org/10.1016/j.radonc.2012.09.023