European Radiology (2019) 29:3996–4007
https://doi.org/10.1007/s00330-018-5872-6

COMPUTER APPLICATIONS

Differentiation of clear cell and non-clear cell renal cell carcinomas by all-relevant radiomics features from multiphase CT: a VHL mutation perspective

Zhi-Cheng Li[1] · Guangtao Zhai[2] · Jinheng Zhang[1] · Zhongqiu Wang[3] · Guiqin Liu[4] · Guang-yu Wu[4] · Dong Liang[1] · Hairong Zheng[1]

Received: 1 August 2018 / Revised: 18 October 2018 / Accepted: 31 October 2018 / Published online: 6 December 2018

Abstract

Objectives To develop a radiomics model with all-relevant imaging features from multiphasic computed tomography (CT) for differentiating clear cell renal cell carcinoma (ccRCC) from non-ccRCC and to investigate the possible radiogenomics link between the imaging features and a key ccRCC driver gene—the von Hippel-Lindau (VHL) gene mutation.

Methods In this retrospective two-center study, two radiomics models were built using random forest from a training cohort (170 patients), where one model was built with all-relevant features and the other with minimum redundancy maximum relevance (mRMR) features. A model combining all-relevant features and clinical factors (sex, age) was also built. The radiogenomics association between selected features and VHL mutation was investigated by Wilcoxon rank-sum test. All models were tested on an independent validation cohort (85 patients) with ROC curves analysis.

Results The model with eight all-relevant features from corticomedullary phase CT achieved an AUC of 0.949 and an accuracy of 92.9% in the validation cohort, which significantly outperformed the model with eight mRMR features (seven from nephrographic phase and one from corticomedullary phase) with an AUC of 0.851 and an accuracy of 81.2%. Combining age and sex did not benefit the performance. Five out of eight all-relevant features were significantly associated with VHL mutation, while all eight mRMR features were significantly associated with VHL mutation (false discovery rate-adjusted $p < 0.05$).

Conclusions All-relevant features in corticomedullary phase CT can be used to differentiate ccRCC from non-ccRCC. Most subtype-discriminative imaging features were found to be significantly associated with VHL mutation, which may underlie the molecular basis of the radiomics features.

Key Points

- *All-relevant features in corticomedullary phase CT can be used to differentiate ccRCC from non-ccRCC with high accuracy.*
- *Most RCC-subtype-discriminative CT features were associated with the key RCC-driven gene—the VHL gene mutation.*
- *Radiomics model can be more accurate and interpretable when the imaging features could reflect underlying molecular basis of RCC.*

Zhi-Cheng Li and Guangtao Zhai contributed equally.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5872-6) contains supplementary material, which is available to authorized users.

✉ Guiqin Liu
sdzclgq@126.com

✉ Guang-yu Wu
danielrau@163.com

[1] Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[2] Institute of Image Communication and Network Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

[3] Department of Radiology, The Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China

[4] Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, No. 1630, Dongfang Road, Pudong, Shanghai 200120, China

Keywords Renal cell carcinomas · Diagnostic imaging · Radiomics · von Hippel-Lindau disease

Abbreviations

AUC	Area under the ROC curve
ccRCC	Clear cell renal cell carcinoma
chRCC	Chromophobe renal cell carcinoma
CT	Computed tomography
FDR	False discovery rate
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
GLSZM	Gray level size zone matrix
HIF	Hypoxia-inducible factor
ICC	Intraclass correlation coefficient
MRI	Magnetic resonance imaging
mRMRe	Minimum redundancy maximum relevance ensemble
NGTDM	Neighborhood gray-tone difference matrix
pRCC	Papillary renal cell carcinoma
RCC	Renal cell carcinoma
ROC	Receiver operating characteristic curve
VHL	von Hippel-Lindau

Introduction

Renal cell carcinoma (RCC) accounts for ~2% of all malignancies worldwide [1]. RCC is a heterogeneous malignancy categorized generally into two major groups: clear cell RCC (ccRCC, ~75%) and non-clear cell RCC (non-ccRCC) [2]. Non-ccRCC encompasses distinct histologic and genetic subtypes, including papillary RCC (pRCC, 10–15%), chromophobe RCC (chRCC, ~5%), and many other rare entities [3, 4]. Differentiating ccRCC from non-ccRCC is of essential importance, because they have substantially different prognosis, clinical behavior, genetic expression patterns, and therapeutic pathways. ccRCC has much worse prognosis and accounts for 94% of metastatic RCC [5, 6]. Genetically, ccRCC is characterized by a very high frequency (> 80%) of von Hippel-Lindau (VHL) tumor suppressor gene loss, which constitutes the earliest, truncal, and governing oncogenic driving event for the development of ccRCC [7, 8]. VHL mutation assumes a central place in current understanding of the pathobiology of ccRCC [9]. However, VHL mutations are much less common in non-ccRCC than in ccRCC, and the initiating oncogenic events in non-ccRCC are not driven by VHL [10]. Furthermore, ccRCC and non-ccRCC show different responses to molecularly targeted therapies, especially in advanced and metastatic RCCs [11, 12].

A percutaneous biopsy is a commonly used approach for pretreatment diagnosis of RCC. However, biopsy of RCC has been controversial due to its invasiveness, potential complications, and possible sampling errors [13, 14]. Previous studies have shown that imaging-based approaches allow noninvasive differentiation between ccRCC and other subtypes, based on the contrast enhancing patterns, 2D texture features, or quantitative parameters in either CT or MRI [6, 15–19]. Although promising, these imaging metrics appear to be not pathognomonic for RCC subtypes. Recently, radiomics techniques have emerged as powerful tools to comprehensively characterize microscale information hidden within conventional medical imaging by converting the images into high-throughput quantitative features [20, 21]. Although most radiomics features are difficult to quantify by the naked human eye and are generally not part of the radiologists' lexicon, many evidences [20–23] have demonstrated that radiomics features offer information on cancer phenotype and tumor microenvironment that are distinct and complementary to other pertinent data sources. Radiomics has demonstrated its high accuracy in cancer classification and survival prediction for different cancers [22, 23]. However, few studies have assessed radiomics features for differentiating RCC subtypes [24].

Most radiomics studies focus on high-throughput data mining to extract features useful for accurate predictions [22–24]. An important limitation of existing studies is that the biological meaning of selected imaging features has been seldom assessed [25]. Beyond accuracy, it is more meaningful to build an interpretable radiomics model by selecting features with biological relevance, as in the way of genomic analysis to identify genes relevant to cancer [26]. To achieve both accuracy and biological relevance, a major challenge is to discover imaging features that (1) can achieve an accurate classification and (2) may decipher mechanisms underlying the classification model. In machine learning community, this is referred to as the all-relevant feature selection problem [27, 28], which has received much interest in selection of cancer-related genes [26, 29]. To our knowledge, such an accurate and interpretable radiomics model built with all-relevant features associated with RCC subtypes has not yet been described, and the biologic underpinning of such imaging features has not been assessed.

In this hypothesis-generating study, we first investigated the performance of all-relevant radiomics features from multiphasic CT in differentiating ccRCC from non-ccRCC. We further explored the underlying molecular basis of the identified all-relevant features by assessing their possible radiogenomics link with the key ccRCC driver gene—the VHL gene mutation. We also tried to reveal the possible reason and manner of the radiogenomics association between ccRCC-discriminative imaging features and ccRCC-relevant genes.

Material and methods

Patients

This two-center retrospective study was approved by both the Renji Hospital (RH) Ethics Committee and The Affiliated Hospital of Nanjing University of Chinese Medicine (AHNUCM) Ethics Committee. Informed consent was obtained from all patients. From the pathology and radiology databases in the above two institutions, we performed a retrospective search to derive consecutive patients with (1) newly diagnosed histologically confirmed ccRCC, pRCC, or chRCC and (2) pretreatment multiphase (including non-enhanced, corticomedullary, and nephrographic phases) CT images from January 2015 to October 2017. The exclusion criteria were previous biopsy or surgery for renal tumor. We finally identified 255 patients with 255 tumors, where 188 were ccRCC and 67 were non-ccRCC (36 pRCC and 31 chRCC). The 170 patients from one center, RH, were used as the training cohort, while the 85 patients from the other center, AHNUCM, were used as an independent validation cohort. VHL mutation status was available for 119 patients, which was used as a radiogenomics cohort.

CT imaging

All CT examinations were performed with helical scanners using triphasic renal protocol on CT systems (VCT LightSpeed, Discovery CT, or Optima 670, GE Healthcare). All patients were administered 400–500 mL of water orally 20 min before the examination. After unenhanced CT scans, corticomedullary and nephrographic phase images were then obtained from the diaphragm through the kidneys beginning 30 s and 65–75 s after injection of non-ionic contrast medium (Iopamiro, Bracco Imaging, or Iopromide, Bayer Schering Pharma) at a dose of 2 mL/kg. For the RH patients, the imaging parameters of each phase were 120 kVp, 110–380-mA tube current, and a 5.0-mm slice thickness. For the AHNUCM patients, the parameters were 120 kVp, 200–400 mA, and a 5.0-mm thickness.

VHL mutation status testing

Tumor tissue from each patient in the radiogenomics cohort was selected on the basis of hematoxylin and eosin-stained sections to ensure a minimum of 75% tumor cells in the samples. Tumor and normal tissue was scraped away from the paraffin blocks using a scalpel. Mutation analyses of the entire coding regions of VHL genes were performed by using polymerase chain reaction amplification and bidirectional Sanger sequencing method, as previously described in [30].

Image preprocessing and tumor segmentation

Image preprocessing was performed to standardize both geometry and intensity properties of the CT images. The aim was to extract robust features and ensure model reproducibility. First, all voxels were isotropically resampled into $1 \times 1 \times 1$ mm^3 using linear interpolation. Then, rigid registration was performed on triphasic CT images for each patient using corticomedullary phase image as a template with the mutual information similarity metric. Three-dimensional (3D) tumor contours were manually delineated slice-by-slice using the ITK-SNAP software by a radiologist (G.W. with 8 years' experience in urologic imaging). To find robust features against the intra- and inter-observer delineation variations, the delineation was repeated on 40 patients (with 20 randomly chosen ccRCC, 10 pRCC, and 10 chRCC) by the same radiologist (G.W.) for intra-observer assessment and by another radiologist (G.L. with 5 years' experience in urologic imaging) for inter-observer assessment, yielding a test-retest data set and a multiple delineation data set respectively. To normalize the intensity range, voxels in the segmented tumor with intensities out of the range $\mu \pm 3\sigma$ were excluded in subsequent analysis (μ and σ were mean and standard deviation for the tumor area). This intensity normalization approach has been successfully used to reduce interscanner variation in CT texture analysis [18].

Radiomics feature extraction

For each segmented 3D tumor, we extracted quantitative texture features from each phase using an in-house developed program. The texture features described the high-order spatial distributions of the intensity within the 3D tumor. Fifty-two texture features were extracted from each phase using several different methods, including gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM), as summarized in Table 1. The detailed calculation of the texture features can be found in [20, 22]. Finally, for each tumor, 156 features were extracted from the triphasic CT.

Feature selection

First, features with low repeatability were excluded from subsequent analysis. Intra- and inter-observer repeatability for each feature was measured by the intraclass correlation coefficient (ICC) calculated on the test-retest data set and multiple delineation data set respectively. Features with ICC of less than 0.85 were discarded.

The remaining features were further analyzed for all-relevant selection. According to Kohavi's definition [31], the relevance of features can be defined in terms of their impact on the performance of an ideal Bayes classifier. Random forest

Table 1 A summary of the radiomics features extracted. Homogeneity and informational measure of correlation have two calculation methods, respectively, which can be found in [22]

Types	Feature names
GLCM features	Contrast, correlation, autocorrelation, energy, variance, dissimilarity, entropy, sum average, sum entropy, sum variance, difference variance, difference entropy, cluster prominence, cluster shade, maximum probability, homogeneity 1/2, informational measure of correlation 1/2, inverse difference moment normalized, inverse difference normalized
GLRLM features	Short run emphasis, long run emphasis, gray-level non-uniformity, run length non-uniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, run length variance, short run low gray-level emphasis, short run high gray-level emphasis, gray-level variance, long run low gray-level emphasis, long run high gray-level emphasis
GLSZM features	Small zone emphasis, large zone emphasis, gray-level non-uniformity, zone size non-uniformity, zone percentage, low gray-level zone emphasis, high gray-level zone emphasis, zone-size variance, small zone low gray-level emphasis, small zone high gray-level emphasis, gray-level variance, large zone low gray-level emphasis, large zone high gray-level emphasis
NGTDM features	Coarseness, contrast, busyness, complexity, strength

algorithm natively produces an importance measure for a feature and requires little parameter tuning, thusly popular in gene selection tasks [32]. Here, we used a random forest-based wrapper algorithm, Boruta, to select all-relevant features [33]. Boruta evaluated the feature relevance by comparing the importance of original features with that achieved by artificially added random features. A random forest algorithm was performed iteratively to measure the feature importance, where irrelevant features were discarded progressively. To reach statistical significance, the algorithm repeatedly calculated all possible feature combinations, generating an all-relevant subset of features.

For comparison, another widely used feature selection method, minimum redundancy maximum relevance (mRMR) ensemble (mRMRe), was also investigated [34]. The mRMR is designed for finding a subset of both relevant and complementary features, where the relevance is characterized in terms of feature mutual information [35]. mRMRe is a bootstrap ensemble of mRMR for building more robust models.

Classification

Based on the selected Boruta features and mRMRe features, two random forest models were built respectively for classification of ccRCC and non-ccRCC. The third random forest classifier was built by combining both Boruta features and clinical variables (sex and age). The tree number of all random forest classifiers was set to 350. The Gini index was used as importance measure [36].

Statistical analysis

All statistical analysis was done with R software, version 3.5.1. p value of less than 0.05 was considered significant. To achieve model reproducibility, the data set in our study, including extracted features, radiomics models, statistical analysis, and their R implementation, was deposited into the public database (https://github.com/Z-CLi/Model-Evaluation.git).

The differences in sex, age, and RCC subtypes between the training and validation cohorts were assessed by using χ^2 test or Wilcoxon rank-sum test, where appropriate. All three classification models were trained on the training cohort and tested on the independent validation cohort. The classification performance was assessed in terms of area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and F1 score. In all tests, ccRCC was regarded as a positive result. The AUCs were statistically compared between different classifiers using the DeLong method [37]. All indices were calculated for both training and validation cohorts. To further investigate the potential two-subtype classification power of the three models, their performance of differentiating ccRCC from pRCC in the validation cohort (chRCC were removed) and differentiating ccRCC from chRCC in the validation cohort (pRCC were removed) was also tested.

Based on the radiogenomics cohort, Wilcoxon rank-sum test was performed to assess the association between the selected imaging features and the VHL mutation. Both feature subsets selected by Boruta and mRMRe were analyzed. Multiple hypothesis correction was performed by means of false discovery rate (FDR) adjustment using the Benjamini-Hochberg method [38].

Results

Patients

No significant difference was found in patient and tumor characteristics between the training and validation

Table 2 Patient and tumor characteristics of the study population. The difference in sex and subtypes between training and validation cohort was assess by using χ^2 test, while the difference in age was assessed by using Wilcoxon rank-sum test

Characteristic	Training cohort	Validation cohort
No. of patients	170 (66.67%)	85 (33.33%)
No. of tumors	170	85
Sex		
Male	113 (66.47%)	56 (65.88%)
Female	57 (33.53%)	29 (34.12%)
Age (years)		
Mean (range)	58.49 (21–84)	58.89 (33–81)
≤ 60	83 (48.82%)	41 (48.24%)
> 60	87 (51.18%)	44 (51.76%)
Subtype		
ccRCC	125 (73.53%)	63 (74.12%)
Non-ccRCC	45 (26.47%)	22 (25.88%)
pRCC	25	11
chRCC	20	11
VHL available	119	
ccRCC	89 (74.79%)	
Non-ccRCC	30 (25.21%)	
Mutated	76 (63.87%)	
Unmutated	43 (36.13%)	

cohorts ($p = 0.86$ to 0.93). Demographic and tumor characteristics of all patients are summarized in Table 2.

Feature selection

After intra- and inter-observer repeatability test, 133 features remained. Finally, eight all-relevant features from the corticomedullary phase image were selected by the Boruta algorithm. Meanwhile, the mRMRe method selected eight features comprising seven from nephrographic phase and one from corticomedullary phase. The feature selection results are summarized in Table 3.

Model validation

The ROC curves of the Boruta and mRMRe models for differentiating ccRCC from non-ccRCC in both training and validation cohorts are shown in Fig. 1. The performance is summarized in Table 4. The importance of each selected feature is shown in Table 3. The model built with Boruta features achieved an AUC of 0.949 (95% confidence interval [CI] 0.889, 0.993) and an accuracy of 92.9% in the validation cohort, which outperformed the mRMRe model with an AUC of 0.851 (95% CI 0.742, 0.936) and an accuracy of 81.2%. The DeLong analysis found significant difference between their AUCs ($p = 0.023$). Combining Boruta features with age and sex obtained an AUC of 0.951 (95% CI 0.882, 0.996) and an accuracy of 92.9% in the validation cohort. No significant difference was found between the AUCs of the combined model and the Boruta model ($p = 0.853$).

Table 3 A summary of the feature selection results for Boruta (a) and mRMRe (b), where cp, np, Impt, and pFDR are short for corticomedullary phase, nephrographic phase, importance measure, and false discovery rate (FDR)-adjusted p value, respectively. Wilcoxon rank-sum test was performed to assess the association between the selected imaging features and the VHL mutation

No.	Selected Boruta features	Phase	Impt	p	pFDR
(a) The selected Boruta features					
f_1	GLCM.Informational Measure of Correlation 1	cp	18.25	0.008	0.022
f_2	GLCM.Informational Measure of Correlation 2	cp	16.10	0.009	0.023
f_3	GLCM.Correlation	cp	7.08	0.006	0.022
f_4	GLSZM.Small Zone Emphasis	cp	6.71	0.008	0.023
f_5	NGTDM.Contrast	cp	5.93	0.028	0.064
f_6	GLCM.Cluster Prominence	cp	5.92	0.048	0.099
f_7	GLCM.Sum Entropy	cp	4.21	0.253	0.350
f_8	NGTDM.Complexity	cp	1.57	0.007	0.022
No.	Selected mRMRe features	Phase	Impt	p	pFDR
(b) The selected mRMRe features					
l_1	GLCM.Inverse Difference Normalized	cp	14.77	0.007	0.022
l_2	NGTDM.Contrast	np	14.68	0.007	0.022
l_3	NGTDM.Complexity	np	10.35	0.010	0.026
l_4	GLSZM.Zone Percentage	np	6.62	0.004	0.021
l_5	GLCM.Homogeneity 1	np	6.18	0.006	0.022
l_6	GLCM.Dissimilarity	np	4.80	0.004	0.021
l_7	GLCM.Inverse Difference Moment Normalized	np	4.14	0.004	0.021
l_8	GLCM.Difference Entropy	np	4.05	0.004	0.021

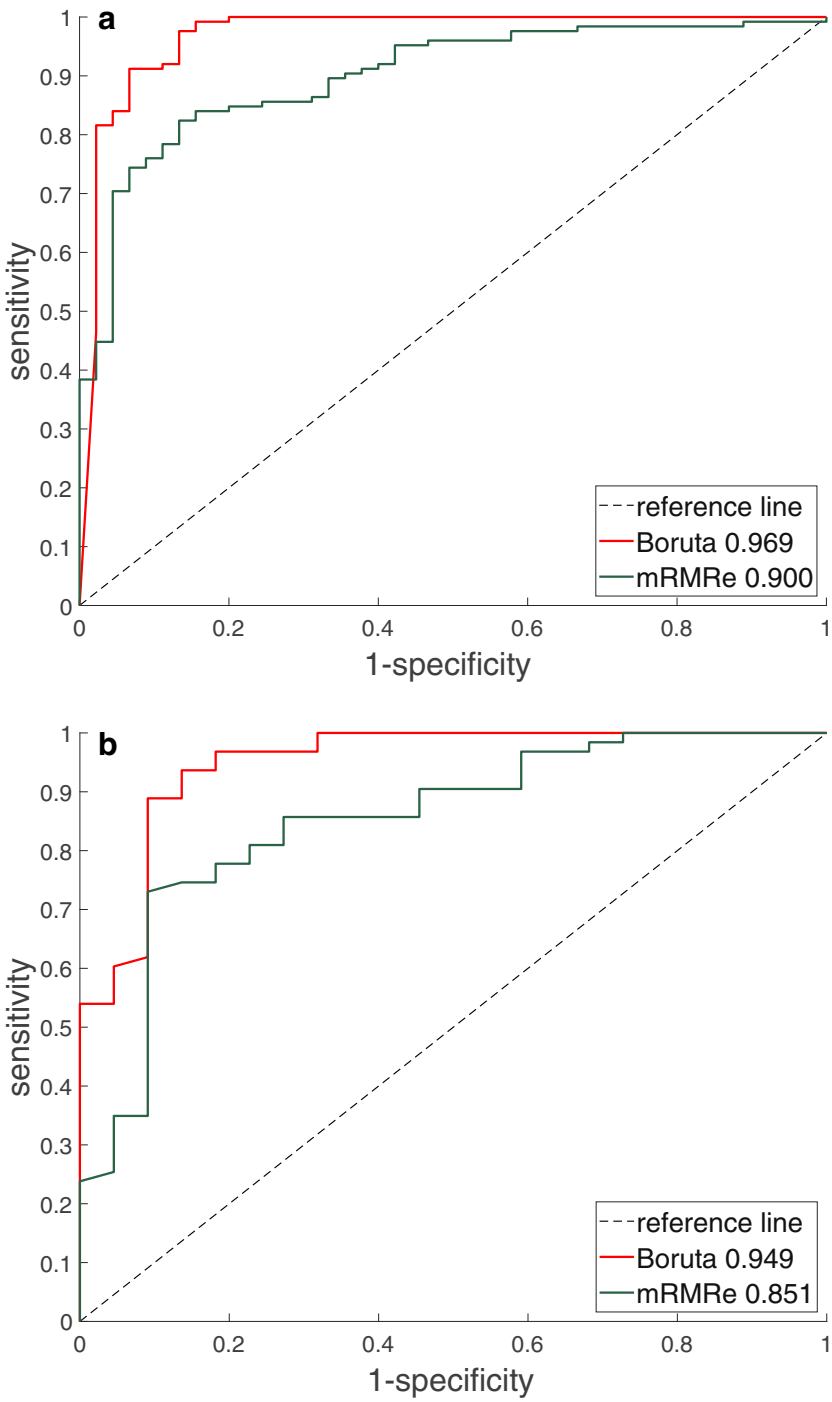

Fig. 1 ROC curves of both Boruta and mRMRe radiomics models for differentiating ccRCC from non-ccRCC in both training (**a**) and validation cohort (**b**)

The ROC curves for differentiating ccRCC from pRCC and differentiating ccRCC from chRCC are shown in Fig. 2. The two-subtype classification performance is summarized in Table 5. Both accuracy and AUC of the Boruta model were higher than those of the mRMRe model. However, no significant difference was found between their AUCs ($p = 0.056$ and 0.154 respectively). To further reveal the relevance of the Boruta features with subtypes, the feature maps for a ccRCC patient, a pRCC patient, and a chRCC patient are shown in Fig. 3.

VHL association

All eight mRMRe features were significantly associated with the VHL mutation (FDR-adjusted $p < 0.05$). Before FDR

Table 4 A performance summary of the radiomics models and the combined model in differentiating ccRCC from non-ccRCC. ACC, AUC, SEN, SPE, PPT, NPV, and F1 are short for accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, positive prediction value, negative prediction value, and F1 score, respectively

Model	Training cohort							Independent validation cohort						
	ACC	AUC	SEN	SPE	PPV	NPV	F1	ACC	AUC	SEN	SPE	PPV	NPV	F1
Boruta	0.947	0.969	0.976	0.867	0.953	0.929	0.964	0.929	0.949	0.968	0.818	0.938	0.900	0.953
mRMRe	0.853	0.900	0.952	0.578	0.862	0.813	0.905	0.812	0.851	0.952	0.409	0.822	0.750	0.882
Combined	0.947	0.966	0.976	0.867	0.953	0.929	0.964	0.929	0.951	0.968	0.818	0.938	0.900	0.953

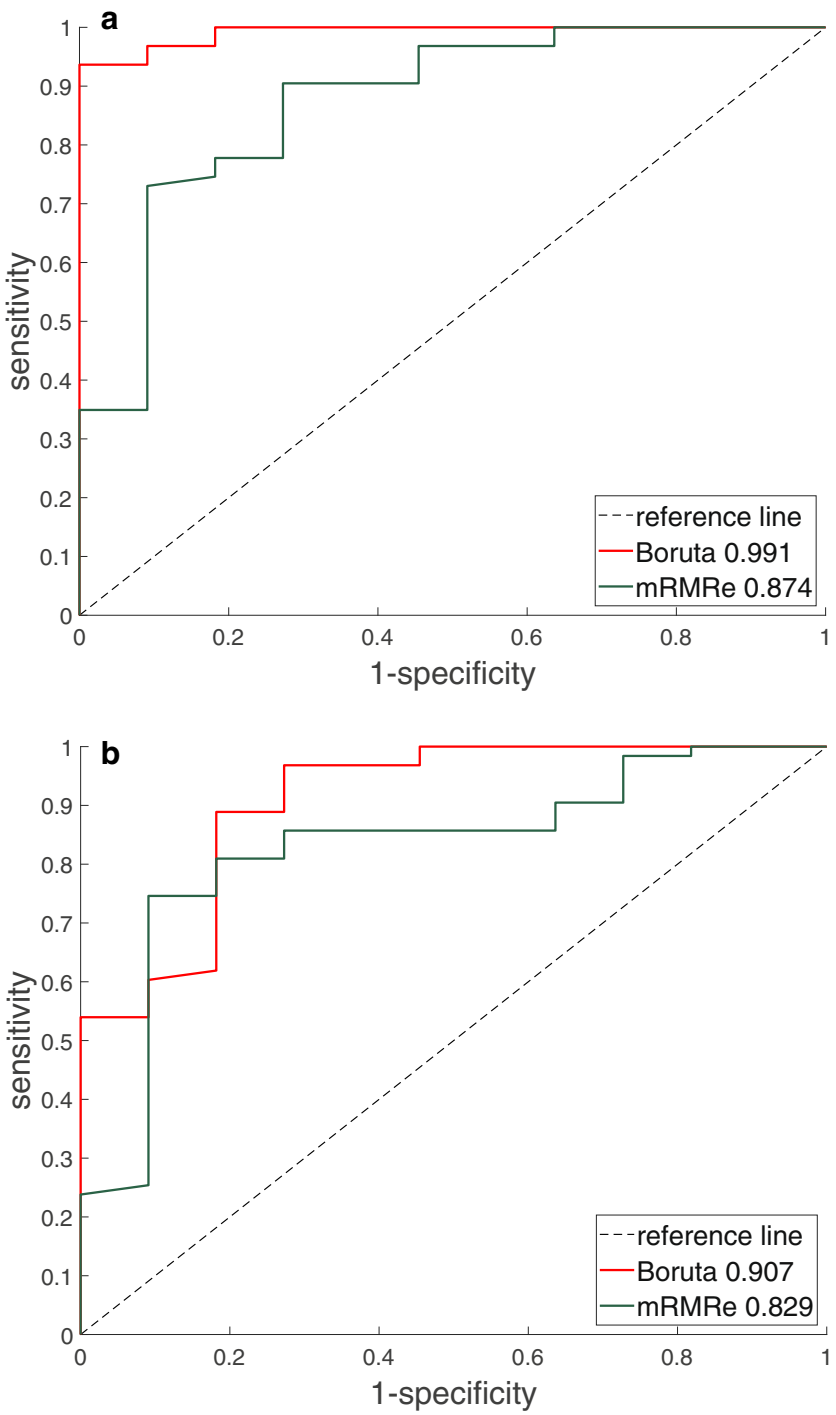

Fig. 2 ROC curves of both Boruta and mRMRe radiomics models for differentiating ccRCC from pRCC (**a**) and from chRCC (**b**)

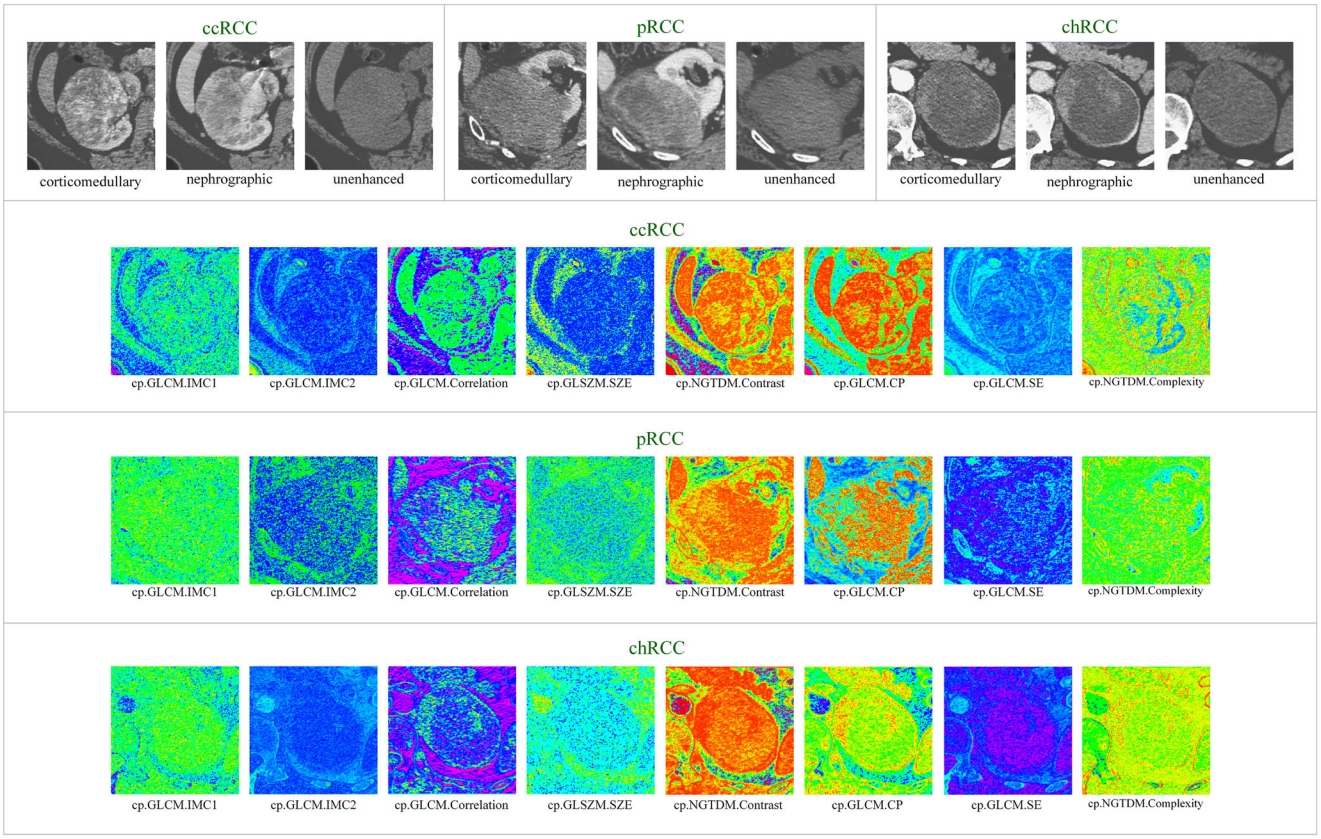

Fig. 3 Radiomics feature maps of the 8 selected all-relevant Boruta features from corticomedullary phase for a ccRCC patient (top), a pRCC patient (middle), and a chRCC patient (bottom), respectively. The prefix cp in feature names was short for corticomedullary phase. The feature names can be found in Table 3. The corticomedullary phase image for ccRCC demonstrated an avidly enhancing tumor. The feature maps illustrated how each all-relevant feature radiologically quantified the regional variations. Specifically, GLCM.IMC, GLCM.Correlation, GLCM.CP, GLCM.SE, NGTDM.Contrast, and NGTDM.Complexity measured the informational correlation, linear dependencies, intensity proximity, disorder randomness, intensity variation, and visual complexity of the region of interest, respectively; GLSZM.SZE gauged the small flat zone with certain gray level. GLCM.Dissimilarity, GLCM.Homogeneity, GLCM.IDN, GLCM.IDMN, and GLCM.DE measured the difference, inverse difference, normalized inverse difference, normalized second moment of inverse difference, and difference disorder between gray levels of neighboring voxels, respectively. GLSZM.ZP was the ratio of the number of small flat zones to the total number of all possible small flat zones

correction, seven out of eight Boruta features were significantly associated with the VHL mutations ($p < 0.05$). After controlling for FDR, five Boruta features were of sufficient strength to retain significance (FDR-adjusted $p < 0.05$). The p value and FDR-adjusted p value for each feature are shown in Table 3. A radiogenomics map was created in Fig. 4 showing each Boruta and mRMRe feature per mutation.

Discussion

Although multiphasic CT has demonstrated useful in differentiation of RCC subtypes [6, 15], few studies have been done on assessing the radiomics features for differentiating RCC subtypes [24], and no studies have investigated the underlying molecular basis of the subtype-discriminative imaging features. Previous studies have evaluated the enhancement patterns [6], iodine quantification [15], and 2D textures [16–18] in CT images, and apparent diffusion coefficients [19] and other quantitative parameters [39] in MRI. One recent work in [24] investigated 43 radiomics features from 2D venous phase CT slices. In our study, for a more comprehensive characterization of the tumor phenotypes and microenvironment, we extracted 156 features derived from 3D tumor regions in triphasic CT. To enhance the reproducibility and usefulness of radiomics models, a guideline was proposed in [20] (http://www.radiomics.world) to assess all key components in radiomics studies. Our study was designed in strict accordance with the guidance in [20] and considered all controllable aspects that may affect the reproducibility of the results, including the CT image normalization, intra- and inter-observer reproducibility, external validation, and data/code availability. For example, we normalized both geometry and intensity properties of the CT images. These efforts may offer penitential to improve the classification performance. For another example, as suggested in [20], to validate a radiomics

Table 5 A performance summary of the radiomics models and the combined model in differentiating ccRCC from pRCC and from chRCC, respectively. ACC, AUC, SEN, SPE, PPT, NPV, and F1 are short for accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, positive prediction value, negative prediction value, and F1 score, respectively

Model	Differentiating ccRCC from pRCC							Differentiating ccRCC from chRCC						
	ACC	AUC	SEN	SPE	PPV	NPV	F1	ACC	AUC	SEN	SPE	PPV	NPV	F1
Boruta	0.959	0.991	0.968	0.909	0.984	0.833	0.976	0.932	0.907	0.968	0.727	0.953	0.800	0.961
mRMRe	0.892	0.874	0.952	0.545	0.923	0.667	0.938	0.851	0.829	0.952	0.273	0.882	0.500	0.916
Combined	0.959	0.994	0.968	0.909	0.984	0.833	0.976	0.932	0.908	0.968	0.727	0.953	0.800	0.961

model internal and/or preferably external validation must be performed. In our two-center study, the model was trained on a training cohort from one center and further tested on an external validation cohort from the other center.

Beyond an accurate classification, learning an interpretable model with features biologically relevant to the target could be more meaningful in understanding the mechanism of a radiomics model. This study further explored the underlying molecular basis of the identified imaging features by assessing the possible radiogenomics association with the VHL mutation. The basic hypothesis of radiogenomics is that expression of specific genes or driver mutations could impact the imaging features [21, 22]. Large-scale genomics studies have shown the VHL loss occurs in > 80% of ccRCC cases [8]. The VHL loss leads to unregulated accumulation of oncogenic hypoxia-inducible factor (HIF) proteins [9, 40], which in turn results in uncontrolled activation of HIF-targeted genes that regulate angiogenesis, glycolytic metabolism, and apoptosis [2]. Therefore, most ccRCC tumors are rich in glycogens and lipids, and are highly vascular [2, 10], which may underlie the regional microenvironmental changes visible at contrast-enhanced CT images. Accordingly, it is a reasonable hypothesis that radiomics imaging features that identify ccRCC tumor could be in certain way associated with VHL gene mutation.

To validate this hypothesis, one major challenge was from high-throughput radiomics data to discover imaging features that are subtype-discriminative and biologically relevant. This can be accomplished by means of the all-relevant feature selection, which is designed for selecting target-relevant feature subset with meaningful insight into the investigated target itself [27, 28]. The random forest-based all-relevant feature selection algorithm, Boruta, has been proven to be the most robust in gene selection [26]. Recently, Boruta has been

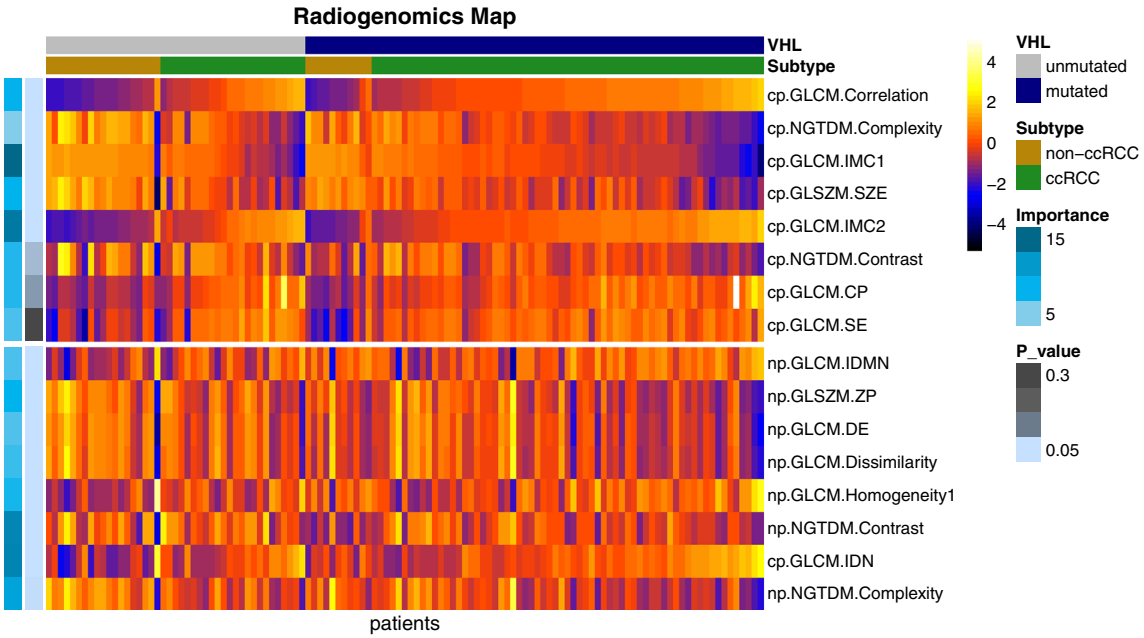

Fig. 4 Radiogenomics map of both Boruta (top eight) and mRMRe features per mutation in the radiogenomics cohort. The prefixes cp and np in feature names were short for corticomedullary phase and nephrographic phase respectively. Each row represented a feature and each column represented a patient. The difference of each feature between ccRCC and non-ccRCC can be observed, indicating the subtype-discriminative power of the features. All features except cp.GLCM.SE, cp.GLCM.CP, and cp.NGTDM.Contrast were significantly associated with VHL mutations

successfully used as a feature selection tool in both radiomics [41, 42] and radiogenomics studies [43]. Another popular feature selection method in genomic analysis, mRMRe, was also employed for comparison. Essentially, mRMR is not designed for solving all-relevant problem but for selecting minimally redundant features that maximize the feature relevance measured by mutual information [35].

Our results showed that the Boruta features outperformed the mRMRe features with a significantly better AUC and higher accuracy (Table 4 and Fig. 1) in differentiating ccRCC from non-ccRCC. The Boruta model also performed better in distinguishing ccRCC from chRCC and from pRCC respectively (Table 5 and Fig. 2). Compared with Boruta features, the mRMRe features had a comparable sensitivity but a rather low specificity and a low negative predictive value (Tables 4 and 5). Interestingly, the Boruta and mRMRe features were totally different (considering CT phase), as shown in Table 3. This implies that all-relevant features do not equal features with maximum relevance; maximizing the mutual information relevance may not always maximize the classification accuracy. Our results highlighted the value of the 3D texture features from corticomedullary phase CT, which were more discriminative than those from nephrographic phase in differentiating ccRCC from non-ccRCC. One possible reason may be that features from corticomedullary phase characterize more tumorous vascular structures, which are rich in ccRCC (Fig. 3) driven by the VHL mutation [2, 10]. The feature maps in Fig. 3 give an example of how all-relevant imaging features radiologically quantified the regional and microenvironmental variations within the tumor for each subtype. For example, in the case of ccRCC with rich vascular structures in corticomedullary phase CT images, voxels of vascular structures within the tumor are highly correlated, therefore, in Fig. 3, the GLCM.Correlation feature had higher values and exhibited visible difference between ccRCC and non-ccRCC (more detailed description of the features can be found in the Supplementary Material).

The radiogenomics results showed that all eight mRMRe features were significantly associated with VHL mutation. However, after FDR correction, three all-relevant Boruta features (f_5, f_6, and f_7 in Table 3) were not significantly associated with VHL mutation. It indicated that all-relevant features do not mean all-VHL-relevant features, and all-VHL-relevant imaging features may be less optimal in differentiating ccRCC from non-ccRCC. Consistent evidences have also shown that VHL loss alone is insufficient to induce ccRCC [2]. Recent large-scale sequencing studies have discovered additional novel driver genes involved in the ccRCC pathogenesis, including PBRM1 (29–41%), SETD2 (8–12%), BAP1 (6–10%), KDM5C (4–7%), and MTOR (5–6%) [9, 10]. A radiogenomics study in [44] revealed significant association between CT imaging features and mutations of VHL, BAP1, and KDM5C in ccRCC. Another study in [45] also found significant association between imaging features and mutations of BAP1 and MUC4 in ccRCC. Therefore, we hypothesize that the three non-VHL-relevant features—f_5, f_6 and f_7—might be associated with other ccRCC-relevant gene mutation, which requires future investigation.

Moreover, our study shows how the performance of radiomics could depend not only on data mining but also on biology: although elaborate radiomics model achieves a good prediction of clinical outcome, the model can be more accurate and interpretable when the imaging features reflect underlying molecular basis of cancer. There were several limitations to our study. First, although this study was based on two-center cohorts, larger prospective cohorts from more centers should be involved to improve the machined learning-based model. Second, more unknown radiogenomics links should be explored. For example, the potential link between radiomics features and gene mutations needed for ccRCC to develop, such as PBRM1, SETD2, BAP1, KDM5C, and MTOR, may provide us with more insights. Third, previous studies have shown the value of multiparametric MRI in ccRCC differentiation [19, 39]. Incorporating features derived from these MR modalities may potentially improve the model performance.

Additionally, all radiomics analysis in this study was performed based on an in-house developed program. There also exist several publicly available open-source tools for the extraction of radiomics features from medical images, such as pyradiomics (available at https://github.com/Radiomics), an R package named Radiomics, and Imaging Biomarker Explorer (IBEX) [46]. There are currently no FDA-cleared or FDA-approved radiomics devices. In 2017, FDA issued a de novo classification of such devices as class II [47], and FDA described the generic form of these devices as "Radiological computer-assisted diagnostic (CADx) software for lesions suspicious for cancer."

In conclusion, our study suggests that all-relevant radiomics features from corticomedullary phase CT can differentiate ccRCC from non-ccRCC with high accuracy. Our results have revealed the significant radiogenomics association between the subtype-discriminative CT features and the key ccRCC driver gene—the VHL gene mutation—which may underlie the molecular basis of the discriminative imaging features and warrants further investigation.

Funding This study has received funding from the National Natural Science Foundation of China (no. 61571432) and Shenzhen Basic Research Program (JCYJ20170413162354654).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Hairong Zheng.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Zhi-Cheng Li) has significant statistical expertise.

Informed consent Written informed consent was obtained from all patients.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- multicenter study

References

1. Ferlay J, Soerjomataram I, Ervik M et al (2013) Cancer incidence and mortality worldwide: IARC cancer base No. 11. Available via: http://globocan.iarc.fr. Accessed 22 July 2018
2. Hsieh JJ, Purdue MP, Signoretti S et al (2017) Renal cell carcinoma. Nat Rev Dis Primers 3:17009
3. Durinck S, Stawiski EW, Pavía-Jiménez A et al (2015) Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. Nat Genet 47:13–21
4. Bellmunt J, Dutcher J (2013) Targeted therapies and the treatment of non-clear cell renal cell carcinoma. Ann Oncol 24:1730–1740
5. Cheville JC, Lohse CM, Zincke H, Weaver AL, Blute ML (2003) Comparisons of outcome and prognostic features among histologic subtypes of renal cell carcinoma. Am J Surg Pathol 27:612–624
6. Young JR, Margolis D, Sauk S, Pantuck AJ, Sayre J, Raman SS (2013) Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT. Radiology 267:444–453
7. Cancer Genome Atlas Research Network (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499:43–49
8. Hakimi AA, Pham CG, Hsieh JJ (2013) A clear picture of renal cell carcinoma. Nat Genet 45:849–850
9. Gossage L, Eisen T, Maher ER (2015) VHL, the story of a tumour suppressor gene. Nat Rev Cancer 15:55–64
10. 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 244:525–537
11. Armstrong AJ, Halabi S, Eisen T et al (2016) Everolimus versus sunitinib for patients with metastatic non-clear-cell renal cell carcinoma (ASPEN): a multicentre, open-label, randomised phase 2 trial. Lancet Oncol 17:378–388
12. Fernández-Pello S, Hofmann F, Tahbaz R et al (2017) A systematic review and meta-analysis comparing the effectiveness and adverse effects of different systemic treatments for non-clear cell renal cell carcinoma. Eur Urol 71:426–436
13. Leveridge MJ, Finelli A, Kachura JR et al (2011) Outcomes of small renal mass needle core biopsy, nondiagnostic percutaneous biopsy, and the role of repeat biopsy. Eur Urol 60:578–584
14. Pandharipande PV, Gervais DA, Hartman RI et al (2010) Renal mass biopsy to guide treatment decisions for small incidental renal tumors: a cost-effectiveness analysis. Radiology 256:836–846
15. Mileto A, Marin D, Alfarocordoba M et al (2014) Iodine quantification to distinguish clear cell from papillary renal cell carcinoma at dual-energy multidetector CT: a multireader diagnostic performance study. Radiology 273:813–820
16. Raman SP, Chen Y, Schroeder JL, Huang P, Fishman EK (2014) CT texture analysis of renal masses: pilot study using random forest classification for prediction of pathology. Acad Radiol 21:1587–1596
17. Hodgdon T, McInnes MD, Schieda N, Flood TA, Lamb L, Thornhill RE (2017) Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images? Radiology 276:787–796
18. Feng Z, Rong P, Cao P et al (2018) Machine learning-based quantitative texture analysis of CT images of small renal masses: differentiation of angiomyolipoma without visible fat from renal cell carcinoma. Eur Radiol 28:1625–1633
19. Hötker AM, Mazaheri Y, Wibmer A et al (2017) Differentiation of clear cell renal cell carcinoma from other renal cortical tumors by use of a quantitative multiparametric MRI approach. AJR Am J Roentgenol 208:W85–W91
20. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749
21. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
22. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
23. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 34:2157–2164
24. Yu HS, Scalera J, Khalid M et al (2017) Texture analysis as a radiomic marker for differentiating renal tumors. Abdom Radiol (NY) (10):2470–2478
25. Kiessling F (2018) The changing face of cancer diagnosis: from computational image analysis to systems biology. Eur Radiol 28: 3160–3164
26. Kursa MB (2014) Robustness of random forest-based gene selection methods. BMC Bioinformatics 15:8
27. Nilsson R, Peña JM, Björkegren J, Tegnér J (2007) Consistent feature selection for pattern recognition in polynomial time. J Mach Learn Res 8:589–612
28. Rudnicki WR, Wrzesień M, Paja W (2015) All relevant feature selection methods and applications. In: Stańczyk U, Jain LC (Eds.) Feature selection for data and pattern recognition. Springer, Berlin, pp 11–28
29. Guo P, Luo Y, Mai G et al (2014) Gene expression profile based classification models of psoriasis. Genomics 103:48–55
30. Hakimi AA, Chen YB, Wren J et al (2013) Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol 63:848–854
31. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324
32. Díaz-Uriarte R, Alvarez de Andrés S (2006) Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7(3)
33. Kursa MB, Rudnicki WR (2010) Feature selection with the Boruta package. J Stat Softw 36:1–13
34. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B (2015) mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics 29:2365–2368
35. Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27:1226–1238
36. Louppe G, Wehenkel L, Sutera A, Geurts P (2013) Understanding variable importances in forests of randomized trees. Adv Neural Inf Process Syst:431–439
37. DeLong ER, DeLong DM, Clarke-Pearson DL (1998) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics:837–845

38. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300
39. Gaing B, Sigmund EE, Huang WC et al (2015) Subtype differentiation of renal tumors using voxel-based histogram analysis of intravoxel incoherent motion parameters. Invest Radiol 50:144–152
40. Semenza GL (2013) HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. J Clin Invest 123: 3664–3671
41. Baessler B, Mannil M, Oebel S, Maintz D, Alkadhi H, Manka R (2017) Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on nonenhanced cine MR images. Radiology 286:103–112
42. Li ZC, Bai H, Sun Q et al (2018) Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. Eur Radiol 28:3640–3650
43. Yamamoto S, Korn RL, Oklu R et al (2014) ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. Radiology 272:568–576
44. Karlo CA, Di Paolo PL, Chaim J et al (2014) Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. Radiology 270:464–471
45. Shinagare AB, Vikram R, Jaffe C et al (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 40:1684–1692
46. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE (2015) IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys 42:1341–1353
47. Food and Drug Administration (2017) Device classification under Section 513(f)(2)(de novo). Available via http://www.accessdata.fda.gov/cdrh_docs/pdf17/DEN170022.pdf. Accessed 18 Oct 2018