Abdominal Radiology (2021) 46:2052–2063
https://doi.org/10.1007/s00261-020-02832-9

REVIEW

CT-based radiomics for differentiating renal tumours: a systematic review

Abhishta Bhandari[1] · Muhammad Ibrahim[1] · Chinmay Sharma[1] · Rebecca Liong[2] · Sonja Gustafson[2] · Marita Prior[2]

Received: 8 February 2020 / Revised: 6 October 2020 / Accepted: 12 October 2020 / Published online: 2 November 2020

Abstract

Purpose Differentiating renal tumours into grades and tumour subtype from medical imaging is important for patient management; however, there is an element of subjectivity when performed qualitatively. Quantitative analysis such as radiomics may provide a more objective approach. The purpose of this article is to systematically review the literature on computed tomography (CT) radiomics for grading and differentiating renal tumour subtypes. An educational perspective will also be provided.

Methods The Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist was followed. PubMed, Scopus and Web of Science were searched for relevant articles. The quality of each study was assessed using the Radiomic Quality Score (RQS).

Results 13 studies were found. The main outcomes were prediction of pathological grade and differentiating between renal tumour types, measured as area under the curve (AUC) for either the receiver operator curve or precision recall curve. Features extracted to predict pathological grade or tumour subtype included shape, intensity, texture and wavelet (a type of higher order feature). Four studies differentiated between low-grade and high-grade clear cell renal cell cancer (RCC) with good performance (AUC = 0.82–0.978). One other study differentiated low- and high-grade chromophobe with AUC = 0.84. Finally, eight studies used radiomics to differentiate between tumour types such as clear cell RCC, fat-poor angiomyolipoma, papillary RCC, chromophobe RCC and renal oncocytoma with high levels of performance (AUC 0.82–0.96).

Conclusion Renal tumours can be pathologically classified using CT-based radiomics with good performance. The main radiomic feature used for tumour differentiation was texture. Fuhrman was the most common pathologic grading system used in the reviewed studies. Renal tumour grading studies should be extended beyond clear cell RCC and chromophobe RCC. Further research with larger prospective studies, performed in the clinical setting, across multiple institutions would help with clinical translation to the radiologist's workstation.

Keywords Computed tomography · Machine learning · Artificial intelligence · Renal tumours · Radiomics · Grade

Introduction

The rapid development of computational analysis for radiological images represents a major change from conventional visual interpretation. Radiological images are a rich source of quantitative data [1]. Quantitative analysis of images has given rise to the field of radiomics which has resulted in an increasing interest in its potential use in a clinical setting, for example, radiomic application to tumour characterisation (e.g. grading and differentiation) and clinical prediction (e.g. survival) [1–5]. Computational analysis based on data extraction and modelling can detect features within radiological images that are not readily apparent to the human eye. By identifying characteristics such as shape, intensity and texture in heterogeneous volumes of interest (VOI), radiomic-based computational analysis may be a successful tool for radiologists to detect, differentiate and grade tumours and other pathologies.

Radiomic analysis begins with the acquisition of the image. Patients' images are acquired during standard-of-care procedures such as contrast-enhanced computed

✉ Abhishta Bhandari
Abhishta.bhandari@my.jcu.edu.au

[1] Townsville University Hospital, 100 Angus Smith Drive, Douglas, QLD 4814, Australia

[2] Department of Medical Imaging Research Office, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia

tomography (CT). The images are processed to transform them into quantitative data for data mining. This occurs in an ordered sequence of events. First, the image is segmented into a volume of interest [1]. This may be achieved manually whereby trained personnel draw regions of interest (ROI) around tumours on each CT slice. Alternatively, there are various automated techniques to segment tumours—convolutional neural networks being the most well known [6]. This is a machine learning (ML) model biologically inspired by the human visual cortex. Second, features are extracted from each segmented image. Features include the *shape* (geometric parameters) of the tumour, *1st order features* (intensity of voxels) and *2nd order features (*texture of the voxel habitat—i.e. how voxels relate to each other*).* Higher order features (such as fractals and wavelet transformations) are also prevalent within the literature in areas such as glioma radiomic analysis [7].

Redundant features are removed once radiomic data has been extracted. The feature selected data can be used to predict tumour characteristics or clinical sequelae using an algorithm developed through ML. In simple terms, ML is a form of artificial intelligence which involves training a model to recognise features within a dataset. This model is then tested and validated on 'testing' and validation' datasets. Various ML models have been developed such as the *convolutional neural network* [8] whilst others are formed in a tree-like structure such as the *decision tree* [9]. ML algorithms are first applied to a *training* dataset. The trained model is applied to the *testing/validation* dataset for verification. Once verified, the trained model can be applied to the target image for the pathology of interest to be classified. The model's sensitivity and specificity may be determined by, producing a receiver operator or precision recall curve and its associated area under the curve (AUC). Clinical outcomes such as survival can also be predicted with an associated correlation coefficient [10]. The basic steps of a radiomic model for renal tumours are detailed in Fig. 1.

Renal tumours represent a growing interest in radiomic analysis. There are over 400,000 new cases of renal cancer diagnosed globally each year [11]. Predicting tumour grade and subtype prior to a histological diagnosis may guide treatment decisions and aid in prognostication [12–14]. Research in radiomics has been done to *pathologically* grade renal tumours through imaging before formal histological analysis. The radiomic literature reports a number of models able to successfully differentiate between renal tumours such as non-clear cell renal carcinoma (non-clear cell RCC), clear cell renal carcinoma (clear cell RCC) and angiomyolipoma (AML) which have different clinical sequelae [15–17].

This study will aim to systematically review the literature on CT radiomics to differentiate renal tumours by (1) predicting pathological grade and (2) differentiating between tumour subtypes. CT imaging was identified as the most commonly used modality for incidentally detecting renal tumours. If this imaging can be utilised for initial tumour characterisation [18] then perhaps further imaging may be avoided; this will benefit the patient and reduce the demand on busy clinical departments. Results will be presented in combination with the discussion to facilitate an educational approach.

Methods

A systematic review was performed in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) checklist [19]. Search terms were developed from pilot searches of the literature and

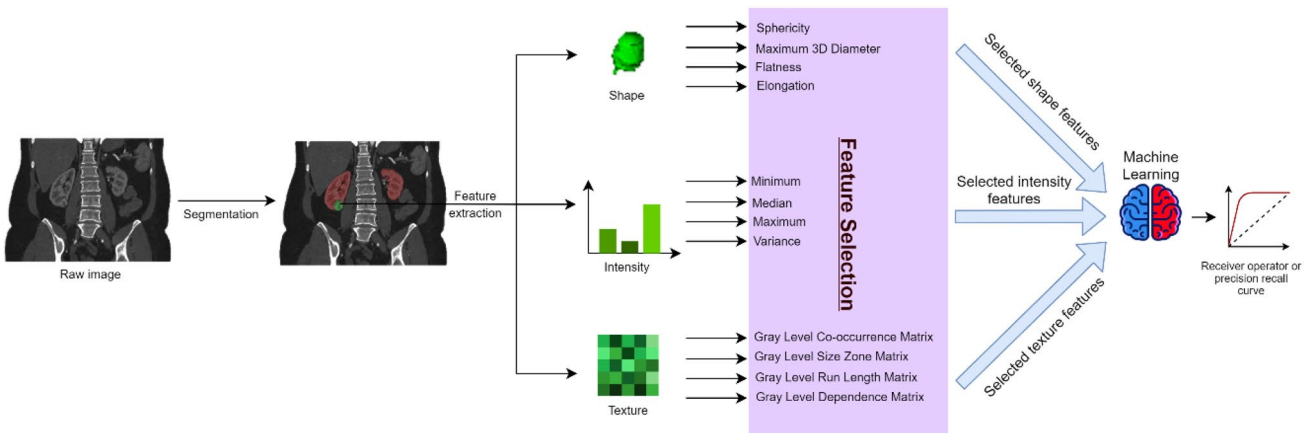

Fig. 1 CT Radiomic Analysis Pipeline for Renal Tumours*:* The CT image is manually or automatically segmented to identify volumes of interest (in this case normal kidney—red; and tumour—green). Data from the volume are extracted for shape, 1st order (intensity), 2nd order (texture), and higher order features (not pictured). Four examples are given in the diagram. These are feature selected. The outputs are then run through a ML model, and a receiver operator curve (ROC) is generated with an area under the curve (AUC)

the PICO (population, intervention, comparison, outcome) framework. Studies of interest involved patients with renal tumours which were either graded or differentiated using CT radiomics with the prediction confirmed by histology. The main outcomes of interest were classification of pathological grade and renal tumour types with sensitivity and specificity of the model measured by the area under the curve (AUC) from the receiver operator or precision recall curve. PubMed, Scopus and Web of Science databases were searched. The search string was: *(renal OR kidney) AND (CT OR "computed tomography") AND radiomic**

Study selection and extraction

Studies were included if they were journal articles that reported: (1) grading or differentiating renal tumour subtypes using radiomic features with ML, (2) adequate information for extraction of pipeline characteristics such as imaging acquisition parameters, segmentation method, features used, ML model and classification of results against histology as the 'gold standard' (such as total/partial nephrectomy or core biopsy), (3) reported an AUC (area under the curve) above 0.8 from a receiver operator or precision recall curve. The associated confidence interval was reported where available. Exclusion criteria were (1) reviews, abstracts, books, opinion articles, (2) non-English articles. Data was extracted by the authors A.B., M.I. and C.S. In addition to AUC, information on segmentation techniques and radiomic features used to grade and differentiate the renal tumours were also extracted. The search was last performed on 09 July,2020.

Data synthesis

The analysis of multiple radiomic pipelines is difficult given their heterogeneity. Features, feature selection method, and ML models differed between the studies; therefore, a meta-analysis was not performed.

Quality assessment

The Radiomic Quality Score (RQS) was applied to assess quality. It is a radiomic-specific scoring system and is based on the Transparent Reporting of a multivariable prediction model for Individual Prognosis OR Diagnosis (TRIPOD) initiative which has a set of recommendations for predictive models [20].

Results and discussion

The literature search found 49 articles from PubMed, 67 from Scopus, and 72 from Web of Science and 21 from the hand search. This gave a total of 209 articles. However, once duplicates were removed 125 articles remained. Titles and abstracts were screened for relevance leaving 24 articles. Full texts were reviewed, 1 article was non-English language, 3 articles did not fit our inclusion criteria of having an AUC > 0.8, and 7 articles did not report an AUC. The results from the thirteen remaining papers are discussed in order of the radiomic pipeline steps, namely image acquisition, segmentation, radiomic features, and machine learning used to grade and differentiate between renal tumours.

Image acquisition

The acquisition parameters of images used in radiomic feature extraction are detailed in Table 1. Slice thickness ranged from 1 to 8 mm, voltage between 120 and 140 kVp, phases included the unenhanced phase (UP), corticomedullary phase (CMP), nephrogenic phase (NP), portal venous phase (PVP), and excretory phase (EP). One study used only unenhanced scans [21]. The amount of contrast injected was between 70 and 150 mL, and the rates were between 3 and 4.5 mL/s. Iodinated contrast used in the studies were Ioversol, Iopamidol, and Iohexol. Three studies [22–24] did not report the type of contrast used.

Segmentations of renal tumours

Segmentation methods are detailed in Table 2. All thirteen studies utilised manual components for segmentation to achieve their first step in the radiomics pipeline (See Fig. 1). To avoid partial volume effect, segmentations were done 1–3 mm inside the tumour margin [2, 22, 25, 27–30] in some studies and skipping slices in the superior and inferior poles [2, 25, 27]. Others tended to incorporate the whole tumour slice during segmentation [16, 21, 24] without segmenting inside the tumour margin. Two studies used a single-slice axial ROI and texture radiomics which classified renal subtype [24] and grade [26] (AUC = 0.90 and with AUC = 0.91, respectively).

Manual segmentation is sub-optimal due to operator subjectivity and its time-consuming nature. Accurate and reliable automated tissue segmentation software should be developed and clinically validated as it will provide greater efficiency in the radiomic pathway [31]. As a response to the lack of automated segmentation, the Medical Imaging Computing and Computer Assisted Intervention (MICCAI) society developed the KiTS19 (Kidney Tumour Segmentation) Grand Challenge where scientists compete using algorithms to automate the segmentation of kidney tumours. The KiTS19 challenge is associated with a single institution-derived database which is freely available on GitHub (https://github.com/neheller/kits19). It contains arterial phase abdominal CT scans of kidney tumour patients. The manual segmentations were done by medical students

Table 1 CT acquisition parameters for images used in radiomic feature extraction to differentiate renal tumours

Author and year	kVp, mA	Slice thickness (mm)	Preprocessing	Phase/s	Time of phase post-IV contrast injection (s)	Amount of IV contrast	Rate of injection (mL/s)	Type of contrast
Grading								
Sun 2019 [3]	120 kVp, tube current automatic adjustment technology	1.25	$1 \times 1 \times 1$ cm^3 voxel sizes	CMP NP EP	25–30 60–70 120–180	100 mL	4.5	Iohexol
Shu 2019 [2]	120 kV 250–400 mA	5	Unavailable	CMP NP	25 70	1.0 mL/kg	3.5	Iopromide
Lin 2019 [25]	Unavailable	1 or 3	0.625×0.625 mm^2 pixel size	UP CMP NP	30 90	70-100 mL	3	Iopamidol, Iohexol
Bektas 2019 [22]	120 kVp 100–500 mA	1–2	0.3×0.3 mm^2 pixel size ± 3 sigma technique for signal intensity	PVP	60	2 mL/kg, max = 150 mL	Unavailable	Unavailable
Schieda 2018 [21]	120 kVp 100–500 mA	Axial: 2.5–5 Coronal and sagittal: 2.5–3	Unavailable	N/A—unenhanced CT	N/A—unenhanced CT	N/A—unenhanced CT	N/A—unenhanced CT	N/A—unenhanced CT
Subtype differentiation								
Yang 2020 [24]	120–140 kVp mA unavailable	Unavailable	Unavailable	UP CMP NP EP	Unavailable	Unavailable	Unavailable	Unavailable
Sun 2020 [26]	120 kVp 180 mA	1.5	Unavailable	CMP NP	10 s after 100HU threshold in supraceliac abdominal aorta 40 s after CMP	1.5 mL/kg (90–120 mL)	3	Iopromide
Ma 2020 [27]	120 kVp 200 mA	5	Unavailable	CMP NP	15 30 100HU threshold in coeliac artery	90–100 mL	3	Iopromide
Erdim 2020 [28]	120 kVp 100–500 mA	Between 1 and 3	1×1 mm^2 pixel size ± 3 sigma technique for signal intensity grey-level discretisation (64 discrete grey levels)	UP CMP	Unavailable	1-2 mL/kg	Unavailable	Non-ionic contrast

Table 1 (continued)

Author and year	kVp, mA	Slice thickness (mm)	Preprocessing	Phase/s	Time of phase post-IV contrast injection (s)	Amount of IV contrast	Rate of injection (mL/s)	Type of contrast
Cui 2019 [29]	Unavailable	1 or 3	Resolution of 512 × 512	CMP	30	70–100 mL	3	Iopamidol
			0625 × 0.625 mm^2 pixel size	NP	90			
Li 2019 [23]	120 kVp mA unavailable	5–8	Unavailable	CMP NP	30–200	Unavailable	3	Unavailable
Feng 2018 [30]	120 kVp 200 mA	5	Unavailable	CMP NP	30 90	90	3	Iohexol
Yu 2017 [16]	120 kVp 200–650 mA	1.25	Unavailable	PVP	70	100	3	Ioversol, Iopamidol

kVp kilovoltage peak, *mA* milliamps, *UP* unenhanced phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *EP* excretory phase, *PVP* portal venous phase

Table 2 Segmentation methods to grade and differentiate subtypes in renal tumours

Author and year	Segmentation method
Grading	
Sun 2019 [3]	Manually selecting the largest axial ROI
Shu 2019 [2]	Manual delineation of contiguous slices "slightly within" the borders of the tumour and excluding first and last slices
Li 2019 [23]	Manual segmentation by contiguous slices
Bektas 2019 [22]	Polygonal ROIs 1 mm inside the lesion contour line
Schieda 2018 [21]	Axial ROI in middle of lesion incorporating the outer margin Axial ROI in between middle and superior pole incorporating the outer margin Axial ROI in between middle and inferior pole incorporating the outer margin
Differentiation of subtype	
Yang 2020 [24]	Largest single-slice axial ROI on CMP first, then UP, NP and EP
Sun 2020 [26]	Semi-automated: manual ROI with perilesional normal tissue then a dichotomous classification algorithm to separate from perilesional normal tissue and then manually outlining the contour of the tumour to give a whole tumour segmentation
Ma 2020 [30]	Manual segmentation by contiguous slices 2–3 mm inside the visible tumour margin for the whole tumour excluding first and last slices
Erdim 2020 [28]	Manual segmentation by contiguous slices 1 mm inside the visible tumour margin
Cui 2019 [29]	Manual segmentation on contiguous slices at 3 mm inside the visible tumour margin
Lin 2019 [25]	Manual segmentation by 10 mm from superior/inferior pole and at 15 mm intervals, 3 mm inside tumour margin
Feng 2018 [30]	Manual segmentation by contiguous slices 2–3 mm from the tumour margin
Yu 2017 [16]	Manual segmentation in 1.25 mm increments on 10 contiguous axial slices in the mid-portion of the tumour

IV intravenous, *ROI* region of interest, *UP* unenhanced phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *EP* excretory phase

under the guidance of a urological surgeon. A total of 210 lesions were used for the training set, and 90 of the lesions were used for the testing set. Clinical information associated with the images was also available and includes the type of surgery, risk factors for renal cancer, comprehensive clinical outcomes and histological characteristics [32]. Segmentation consistency between automated and manual for the challenge is described using the Dice-Sorensen coefficient (DSC) which lies between 0 and 1. A perfect consistency equalled 1, whilst 0 equalled no consistency.

The first and second place winners of the MICCAI KiTS19 challenge both used the U-Net segmentation method

[33]. This segmentation structure is based on a U-shaped computational architecture which "contracts" the image data as it propagates downwards and "expands" the data as it goes upward forming a U. The KiTS19 challenge included segmentations performed on the kidney, the tumour and the composite of both. The winning team's U-Net segmentation method had DSC values for the tumour: 0.85, kidney: 0.97 and kidney plus tumour composite: 0.90 [17]. Development of automated segmentation methods would help improve efficiency. The studies to date have incorporated manual steps in segmentation.

Radiomic features

Texture-type features were the most frequently used radiomic feature in the studies from this systematic review [3, 16, 21–30]. Table 3 reports texture features used in differentiating renal tumours from this search. Other radiomic features utilised include shape, intensity and wavelet (higher order) features. We provide definitions for each texture feature for understanding. Texture features used were grey-level co-occurrence matrix (GLCM) and grey-level run length matrix (GLRLM), followed by grey-level difference matrix (GLDM), grey-level size zone matrix (GLSZM), autoregressive model, neighbouring grey-tone difference matrix (NGTDM) and gradient. Feature extraction should also occur on common image quality parameters for generalised use across multi-institutional contexts.

Application of machine learning

Support vector machine (SVM) models were the most common method of differentiating renal tumours using CT radiomics [2, 3, 16, 22–24, 29, 30]. SVM models involve separating data values into a binary classification system (i.e. voxels belonging to either healthy tissue or tumour) using a *hyperplane* constructed by the ML process. The hyperplane is constructed with the widest margins between the two datasets. Figure 2 demonstrates a linear SVM model. It illustrates how the hyperplane is represented by a linear function that separates the blue values from the red values. The values that lie on the margins of the hyperplane are the support vectors—from which the name is derived. Other ML models used were logistic regression [21, 27], decision tree [25] and random forest [28].

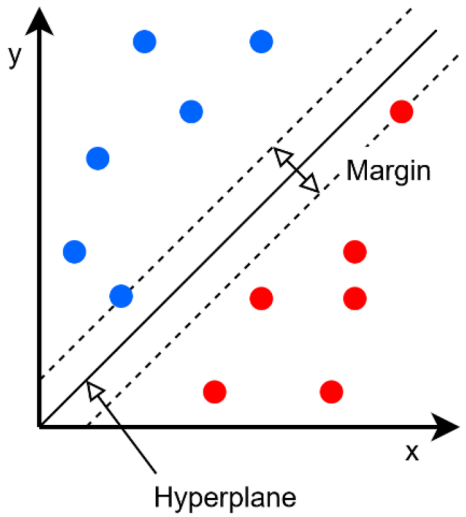

Fig. 2 Support Vector Machine: Schematic representation of a linear SVM. The red dots represent voxels belonging to tumour, and the blue dots are voxels from healthy tissue. These are separated by a hyperplane. Machine learning is used to construct the hyperplane, with the goal being maximal separation between tumour and healthy tissue voxels. Data points on the margins of separation are the support vectors

Table 3 Radiomic texture features used in differentiating renal tumours

Radiomic texture features	Description [19]	References
Gradient	Distortion of texture moving away from a particular point	[16]
GLCM	How two voxels in an image are dependent on each other within that image	[2, 3, 16, 21–23, 25–29]
GLDM	Assesses grey-level dependencies within an image. A grey-level dependency is defined as the number of voxels in connection within a certain distance	[24, 28]
GLRLM	Assesses grey-level runs. These are the length in number of voxels, of consecutive voxels that have an equal grey-level value	[2, 3, 16, 21–27]
GLSZM	Assesses grey-level zones within an image. A grey-level zone is defined as the number of connected voxels that share the same grey-level intensity	[2, 23–25, 28]
NGTM	Assesses the difference between grey value of a particular voxel and average grey values of its neighbours	[24, 28]
Autoregressive model	A random process that gives an output variable that is linearly dependent on its previous values	[22, 27]

GLCM grey-level co-occurrence matrix, *GLRLM* grey-level run length matrix, *GLDM* grey-level difference matrix, *GLSZM* grey-level size zone matrix, *NGTDM* neighbouring grey-tone difference matrix

Grading and differentiating

Table 4 below shows articles found from the systematic review that graded tumours as high-grade and low-grade tumours using radiomic features. To be included in the table, the study had to directly compare two grade categories (high or low) and report an AUC. Three grading systems were used—the Fuhrman (I-IV), International Society of Urological Pathology (ISUP) grade (I-IV) or Paner system. High-grade tumours were defined as being Fuhrman III–IV, ISUP III-IV, Paner system 2 or 3 out of 3 [34]. Low-grade tumours were Fuhrman I–II, ISUP I–II, Paner system: 1 out of 3 [34]. Only studies with AUCs above 0.8 were included.

Studies that graded tumours with an AUC > 0.8 dealt with clear cell RCCs apart from one study on chromophobe RCCs [21]. The most accurate predictor was using the multilayer perceptron (MLP) ML model on texture features to discriminate between high- (III–IV) and low-grade (I–II) clear cell RCCs using the ISUP system (AUC = 0.978) [2]. The most accurate discriminator between high and low-grade renal tumours using the Fuhrman grading system was achieved by decision tree ML model and texture features (AUC = 0.87) [25]. Other Fuhrman grading pipelines include SVM with texture features, wavelet features and PVP CT (AUC = 0.869) [22] and SVM with shape, intensity and texture features achieving an AUC = 0.822 [25]. Schieda et al. 2018 [21] proposed another type of grading system by Paner [34], and an AUC = 0.84 was achieved using logistic regression and texture features for chromophobe RCC. CT contrast phases may influence the radiomic analysis for differentiating high-grade and low-grade clear cell RCC. Lin et al. 2019 [25] showed that ML with radiomics achieved an AUC = 0.87 based on three-phase CT (precontrast, nephrogenic and corticomedullary) and was superior to solely just using UP, CMP, NP individually on CT.

The ISUP grading system has been used in the top radiomic pipeline [26]. The ISUP grading system has been proposed to account for deficiencies in the Fuhrman grading system. The Fuhrman system is more complex and involves three parameters: nuclear size, shape and nucleolar prominence, without clarity on how to weigh conflicting information between them, leading to interpretation errors and poor to moderate inter-observer reproducibility. The ISUP grading system is based on the assumption that nucleolar grade alone is sufficient for grading clear cell and papillary RCC, which was shown to result in higher inter-observer consistency [35]. Given the ease of use, potential less inter-observer variability and higher classification in renal radiomics it may be pertinent that more studies use the ISUP grading system.

The main findings of studies that address differentiation of renal tumour types are summarised in Table 5. Only those studies that used radiomic features and reported an AUC > 0.8 were included in the table.

Differentiation between malignant and benign renal tumours is important for clinical decision-making regarding invasive procedures [4]. Benign (oncocytoma and

Table 4 Grading renal tumours using CT radiomics

Author and year	Grading system	Number of lesions (*n*)	Radiomic features and selection method	ML model and classification results	Histopathology specimen type
Sun 2019 [3]	ISUP (I–II) versus ISUP (III–IV)	Clear cell RCCs = 227	Intensity Texture Wavelet LASSO	SVM classified between grades with an AUC = 0.91 (95%CI: 0.65–0.99) from CMP and NP	Nephrectomy
Shu 2019 [2]	ISUP (I–II) versus ISUP (III–IV)	Clear cell RCCs = 260	Shape Intensity Texture LASSO	MLP classified between grades with an AUC = 0.978 (95%CI: 0.957–0.995) from CMP and NP	Nephrectomy
Lin 2019 [25]	Fuhrman (I–II) versus Fuhrman (III–IV)	Clear cell RCCs = 232	Shape Intensity Texture Unclear feature selection method	Decision tree ML model classified between grades with an AUC = 0.87 from UP, CMP and NP	Nephrectomy
Bektas 2019 [22]	Fuhrman (I–II) versus Fuhrman (III–IV)	Clear cell RCCs = 54	Texture Wavelet Wrapper-based feature selection algorithm	SVM classified between grades with an AUC = 0.86 from portal venous phase	Nephrectomy
Schieda 2018 [21]	High-grade (2 or 3 out of 3) and low-grade (1 out of 3) on system proposed by Paner [34]	Chromophobe RCCs = 37	Texture Unclear feature selection method	Logistic regression identified high-grade tumours with a maximal AUC = 0.84 from UP	Nephrectomy

ISUP International Society of Urologic Pathologists, *LASSO* least absolute shrinkage and selection operator, *UP* unenhanced phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *AUC* area under the curve, *MLP* multi-layer perceptron

Table 5 Differentiating renal tumour subtypes using CT radiomics

Author and year	Subtypes	Number of lesions (n)	Radiomic features and selection method	ML model and classification results	Histopathology specimen type
Yang 2020 [24]	RCC (< 4 cm) Fat-poor AML (< 4 cm)	118 45	Texture *t*-score	SVM classified between (< 4 cm) fat-poor AML and RCC with an AUC = 0.90 from UP, CMP, NP, EP	Nephrectomy
Sun 2020 [26]	Comparison 1: Clear cell RCC versus primary RCC + chromophobe RCC	Clear cell RCC = 190 Papillary RCC = 26 Chromophobe RCC = 38	Shape Intensity Texture *with* radiologist assessment	Comparison 1 using SVM: AUC = 0.93 (95% CI 0.89–0.95)	Total nephrectomy
	Comparison 2: Clear cell RCC versus fat-poor AML + oncocytomas	Fat-poor AML = 26 Oncocytoma = 10	RFE-SVM	Comparison 2 using SVM: AUC = 0.94 (95% CI 0.89–0.96)	
	Comparison 3: Primary RCC + chromophobe RCC versus fat-poor AML + oncocytoma			Comparison 3 using SVM: AUC = 0.87 (95% CI 0.79–0.93) All radiomic features extracted from CMP	
Ma 2020 [27]	Fat-poor AML versus clear cell RCC	Fat-poor AML = 22 Clear cell RCC = 62	Intensity Texture LASSO	Logistic regression with radiologist-measured features with radiomics applied to UP, CMP and NP images using SVM produced an AUC = 0.964 (95% CI = 0.824 – 1.000)	Partial or total nephrectomy
Erdim 2020 [28]	Benign versus malignant tumours	Malignant Clear cell RCC = 25 Papillary RCC = 23 Chromophobe RCC = 15 Benign Oncocytoma = 10 Fat-poor AML = 11	Texture Nested approach using a wrapper algorithm with a greedy stepwise searching method	Random forest ML model produced an AUC = 0.915 from UP and CMP	Core biopsy or nephrectomy
Cui 2019 [29]	Fat-poor AML (< 4 cm) RCC (< 4 cm)	Fat-poor AML from RCC Fat-poor AML from clear cell RCC Fat-poor AML from non-clear cell RCC	Texture *with* radiologist input Boruta feature selection SVM-RFECV	Fat-poor AML can be differentiated from RCC with an AUC = 0.96 from UP, CMP, NP using SVM in combination with radiologist input Fat-poor AML can be differentiated from clear cell RCC with an AUC = 0.97 from UP, CMP, NP using SVM in combination with radiologist input Fat-poor AML can be differentiated from non-clear cell RCC with an AUC = 0.89 from UP, CMP, NP using SVM in combination with radiologist input	Radical or partial nephrectomy
Li 2019 [23]	Chromophobe RCC and renal oncocytoma	Chromophobe RCC = 44 Renal oncocytoma = 17	Intensity Texture LASSO	The highest classifier, SVM had an AUC = 0.964 (SE: ± 0.054) from CMP and NP	Nephrectomy
Feng 2018 [30]	RCC (≤ 4 cm) Fat-poor AML (≤ 4 cm)	41 17	Texture RFE + SMOTE	SVM differentiated between (≤ 4 cm) RCC and fat-poor AML with an AUC = 0.955 from UP, CMP, NP	Total or partial nephrectomy

Table 5 (continued)

Author and year	Subtypes	Number of lesions (n)	Radiomic features and selection method	ML model and classification results	Histopathology specimen type
Yu 2017 [16]	Papillary RCC versus other	Clear cell RCC = 46	Intensity Texture	Linear SVM differentiated: papillary RCC from other tumours with an AUC = 0.92; chromophobe RCC from other tumours with an AUC = 0.82; clear cell RCC with an AUC = 0.91 from portal venous phase	Nephrectomy
		Papillary RCC = 41	Laws' features		
	Chromophobe RCC versus other	Chromophobe RCC = 22	Unclear feature selection method		
		Renal oncocytoma = 10			
	Clear cell RCC versus other				

SVM support vector machines, *RFE* recursive feature elimination, *RFECV* recursive feature elimination with cross-validation, *LASSO* least absolute shrinkage and selection operator, *SMOTE* synthetic minority over-sampling technique, *UP* unenhanced phase, *CMP* corticomedullary phase, *NP* nephrogenic phase, *EP* excretory phase, *AUC* area under the curve, *SE* standard error

fat-poor AML) versus malignant (clear cell RCC, papillary RCC, chromophobe RCC) differentiation can be achieved with an AUC = 0.915 [28]. Difficulty arises in differentiating fat-poor AML from RCCs as the RCC may be incorrectly suspected due to the lack of macroscopic fat. Four studies examined fat-poor AML differentiation [24, 27, 29, 30] and three in lesions smaller than 4 cm [4, 29, 30]. The highest classifying pipeline used a combination of radiomics and human interpretation [29]. Other studies found that papillary, chromophobe and clear cell RCC can be differentiated from other RCC with an AUC = 0.92, 0.81 and 0.91, respectively [16]. In addition, chromophobe RCC can be differentiated from renal oncocytoma with an AUC = 0.964 [27].

Recent research has been performed comparing qualitative interpretation with radiomic ML pipelines. Sun et al. [26] compared radiologists' interpretation of the tumour subtype based on clinical experience with radiomic analysis using an SVM ML model on textural features of CT scans. The study aimed to differentiate among clear cell RCCs, primary RCCs, chromophobe RCCs, clear cell RCCs and fat-poor AML. The promising finding is that combining radiologist and radiomic classification can improve performance. For example, for differentiation between fat-poor AML and clear cell RCC/renal oncocytoma, radiomics showed an accuracy of 61.9%, whilst radiologist accuracy ranged from 73.5 to 95.1% [26]. Combining both the radiomic model and radiologist interpretation, the study demonstrated an accuracy of 85.8%. Cui et al. 2019 [29] found that fat-poor AML can be differentiated from all other RCC with an AUC of 0.96 (accuracy = 87%) on unenhanced phase, corticomedullary and nephrogenic phases in lesions < 4 cm. In combination with radiologist input this increased to an AUC = 0.96 (accuracy = 93%). This radiologist input involved scoring the measurement of the attenuation of the tumour as compared to the cortex (hypo-, iso-, hyper attenuation, amount of exophytic growth (< 50% and ≥ 50%) and homogeneity of the tumour (marked: ≥ 50% heterogeneous and mild: < 50% heterogeneous) and completely homogenous tumour).

Quality assessment

Figure 3 demonstrates the Radiomic Quality Score [36] graphically. The RQS demonstrated a median score of 12/36 (range: 5/36 -15/36). This was 33.3% (range: 13.9–41.7%) of the maximum score. Main deficiencies relevant for integration into clinical practice were that there were no cost-effectiveness analyses in all studies, no application of the models in a clinical setting and no prospective studies, and validation was mainly only done in single institutions—hence generalisability to other contexts and scanners may be an issue. Further large-scale studies are warranted in

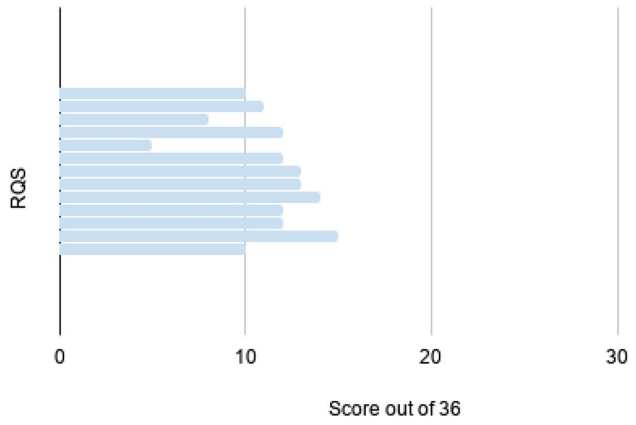

Fig. 3 Radiomic Quality Score (RQS) of Included Studies

multi-institutional settings for further substantiation of applicability in clinical settings.

Future directions

Translation of research into the clinical sphere remains challenging. Limitations of existing research include a lack of prospective data, studies carried out in single centres, the need to adapt models to a clinical setting and a lack of cost analysis. There is also a need for automation of steps such as segmentation. As a stepping stone, it may be beneficial to compare or combine computerised techniques with qualitative interpretation. Only two studies in this review compared human observers with radiomics, and found that together, this improved classification performance [26, 37]. Other gaps in the renal tumour CT radiomic literature include the fact that only one higher order feature was examined (wavelet), and grading studies were limited mainly to one tumour type (clear cell RCC). Other areas of interest include the use of CT radiomics to predict clinical outcomes in patient with renal tumours. In lung tumours, for example, clinical outcomes such as disease survival [38] and overall survival [39] have been reported indicating a role for prognostication using radiomics. There have also been studies in non-small cell lung cancer to predict the absence of distant metastasis [40] and response to neoadjuvant chemotherapy [41].

Conclusion

CT radiomics shows promise in grading and differentiating renal tumours. Studies have been performed to differentiate between clear cell RCC, fat-poor angiomyolipoma, papillary RCC, chromophobe RCC, renal oncocytoma (RO) with AUC ranging from 0.82 to 0.96 in eight studies. The renal tumour grading studies focused on clear cell RCC (AUC = 0.82–0.978) and chromophobe RCC (AUC = 0.84) with further work needing to be done with other renal tumour types. Challenges remain to translate radiomic pipelines for use at the radiologist's workstation.

References

1. 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
2. Shu J, Wen D, Xi Y, Xia Y, Cai Z, Xu W, Meng X, Liu B, Yin H (2019) Clear cell renal cell carcinoma: Machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade. European journal of radiology 121:108738. https://doi.org/10.1016/j.ejrad.2019.108738
3. Sun X, Liu L, Xu K, Li W, Huo Z, Liu H, Shen T, Pan F, Jiang Y, Zhang M (2019) Prediction of ISUP grading of clear cell renal cell carcinoma using support vector machine model based on CT images. Medicine 98 (14):e15022. https://doi.org/10.1097/md.0000000000015022
4. Yang G, Gong A, Nie P, Yan L, Miao W, Zhao Y, Wu J, Cui J, Jia Y, Wang Z (2019) Contrast-Enhanced CT Texture Analysis for Distinguishing Fat-Poor Renal Angiomyolipoma From Chromophobe Renal Cell Carcinoma. Molecular imaging 18:1536012119883161. https://doi.org/10.1177/1536012119883161
5. Zhang GM, Shi B, Xue HD, Ganeshan B, Sun H, Jin ZY (2019) Can quantitative CT texture analysis be used to differentiate subtypes of renal cell carcinoma? Clinical radiology 74 (4):287-294. https://doi.org/10.1016/j.crad.2018.11.009
6. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E (2019) Convolutional neural networks for radiologic images: a radiologist's guide. Radiology 290 (3):590-606
7. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D (2012) Radiomics: the process and the challenges. Magnetic resonance imaging 30 (9):1234-1248
8. Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, Zhai G (2017) A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Scientific reports 7 (1):1-8
9. Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova NJBSP, Control (2019) Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. 52:456-462
10. Che D, Liu Q, Rasheed K, Tao X (2011) Decision tree and ensemble learning algorithms with their applications in bioinformatics. In: Software tools and algorithms for biological systems. Springer, pp 191–199
11. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 68 (6):394–424
12. Deng Y, Soule E, Samuel A, Shah S, Cui E, Asare-Sawiri M, Sundaram C, Lall C, Sandrasegaran K (2019) CT texture analysis in the differentiation of major renal cell carcinoma subtypes and correlation with Fuhrman grade. European radiology 29 (12):6922-6929. https://doi.org/10.1007/s00330-019-06260-2
13. He X, Zhang H, Zhang T, Han F, Song B (2019) Predictive models composed by radiomic features extracted from multi-detector computed tomography images for predicting low- and high- grade clear cell renal cell carcinoma: A STARD-compliant article. Medicine 98 (2):e13957. https://doi.org/10.1097/md.0000000000013957

14. Meng F, Li X, Zhou G, Wang Y (2017) Fuhrman grade classification of clear-cell renal cell carcinoma using computed tomography image analysis. Journal of Medical Imaging and Health Informatics 7 (7):1671-1676. https://doi.org/10.1166/jmihi.2017.2184
15. Sung CK, Kim SH, Woo S, Moon MH, Kim SY, Kim SH, Cho JY (2016) Angiomyolipoma with minimal fat: differentiation of morphological and enhancement features from renal cell carcinoma at CT imaging. Acta radiologica (Stockholm, Sweden : 1987) 57 (9):1114–1122. https://doi.org/10.1177/0284185115618547
16. 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. Abdominal radiology (New York) 42 (10):2470-2478. https://doi.org/10.1007/s00261-017-1144-1
17. Badri AV, Waingankar N, Edwards K, Kutikov A, Parsons RB, Chen DY, Smaldone MC, Viterbo R, Greenberg RE, Uzzo RG (2019) Non-contrast imaging characteristics of papillary renal cell carcinoma: implications for diagnosis and subtyping. Canadian Journal of Urology 26 (5):9916-9921
18. Guðmundsson E, Hellborg H, Lundstam S, Erikson S, Ljungberg B (2011) Metastatic Potential in Renal Cell Carcinomas ≤7 cm: Swedish Kidney Cancer Quality Register Data. European Urology 60 (5):975-982. https://doi.org/10.1007/s00261-017-1144-1
19. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the P-DTAG (2018) Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA 319 (4):388-396. https://doi.org/10.1001/jama.2017.19163
20. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Annals of internal medicine 162 (10):735-736. https://doi.org/10.7326/l15-5093-2
21. Schieda N, Lim RS, Krishna S, McInnes MDF, Flood TA, Thornhill RE (2018) Diagnostic Accuracy of Unenhanced CT Analysis to Differentiate Low-Grade From High-Grade Chromophobe Renal Cell Carcinoma. AJR American journal of roentgenology 210 (5):1079-1087. https://doi.org/10.2214/ajr.17.18874
22. Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, Erdim C, Kilickesmez O (2019) Clear Cell Renal Cell Carcinoma: Machine Learning-Based Quantitative Computed Tomography Texture Analysis for Prediction of Fuhrman Nuclear Grade. European radiology 29 (3):1153-1163. https://doi.org/10.1007/s00330-018-5698-2
23. Li Y, Huang X, Xia Y, Long L (2019) Value of radiomics in differential diagnosis of chromophobe renal cell carcinoma and renal oncocytoma. Abdominal radiology (New York). https://doi.org/10.1007/s00261-019-02269-9
24. Yang R, Wu J, Sun L, Lai S, Xu Y, Liu X, Ma Y, Zhen X (2020) Radiomics of small renal masses on multiphasic CT: accuracy of machine learning–based classification models for the differentiation of renal cell carcinoma and angiomyolipoma without visible fat. European radiology 30 (2):1254-1263. https://doi.org/10.1007/s00330-019-06384-5doi:
25. Lin F, Cui EM, Lei Y, Luo LP (2019) CT-based machine learning model to predict the Fuhrman nuclear grade of clear cell renal cell carcinoma. Abdominal radiology (New York) 44 (7):2528-2534. https://doi.org/10.1007/s00261-019-01992-7
26. 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 American journal of roentgenology 214 (1):W44-w54. https://doi.org/10.2214/ajr.19.21617
27. Ma Y, Cao F, Xu X, Ma W (2020) Can whole-tumor radiomics-based CT analysis better differentiate fat-poor angiomyolipoma from clear cell renal cell caricinoma: compared with conventional CT analysis? Abdominal radiology (New York) 45 (8):2500-2507. https://doi.org/10.1007/s00261-020-02414-9
28. Erdim C, Yardimci AH, Bektas CT, Kocak B, Koca SB, Demir H, Kilickesmez O (2020) Prediction of Benign and Malignant Solid Renal Masses: Machine Learning-Based CT Texture Analysis. Academic radiology. https://doi.org/10.1016/j.acra.2019.12.015
29. Cui EM, Lin F, Li Q, Li RG, Chen XM, Liu ZS, Long WS (2019) Differentiation of renal angiomyolipoma without visible fat from renal cell carcinoma by machine learning based on whole-tumor computed tomography texture features. Acta radiologica (Stockholm, Sweden : 1987) 60 (11):1543–1552. https://doi.org/10.1177/0284185119830282
30. Feng ZC, Rong PF, Cao P, Zhou QY, Zhu WW, Yan ZM, Liu QY, Wang W (2018) Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. European radiology 28 (4):1625-1633. https://doi.org/10.1007/s00330-017-5118-z
31. Sharma N, Aggarwal LM (2010) Automated medical image segmentation techniques. J Med Phys 35 (1):3-14. https://doi.org/10.4103/0971-6203.58777
32. Heller N, Sathianathen N, Kalapara A, Walczak E, Moore K, Kaluzniak H, Rosenberg J, Blake P, Rengel Z, Oestreich M (2019) The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:190400445
33. Ronneberger O, Fischer P, Brox T U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, 2015. Springer, pp 234–241
34. Paner GP, Amin MB, Alvarado-Cabrero I, Young AN, Stricker HJ, Moch H, Lyles RH (2010) A novel tumor grading scheme for chromophobe renal cell carcinoma: prognostic utility and comparison with Fuhrman nuclear grade. The American journal of surgical pathology 34 (9):1233-1240. https://doi.org/10.1097/PAS.0b013e3181e96f2a
35. Samaratunga H, Gianduzzo T, Delahunt B (2014) The ISUP system of staging, grading and classification of renal cell neoplasia. Journal of kidney cancer and VHL 1 (3):26
36. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology 14 (12):749-762. https://doi.org/10.1038/nrclinonc.2017.141
37. Cui EM, Lin F, Li Q, Li RG, Chen XM, Liu ZS, Long WS (2019) Differentiation of renal angiomyolipoma without visible fat from renal cell carcinoma by machine learning based on whole-tumor computed tomography texture features. Acta Radiologica 60 (11):1543-1552. https://doi.org/10.1177/0284185119830282
38. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J, Liang C (2016) Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non—small cell lung cancer. Radiology 281 (3):947-957
39. Song J, Liu Z, Zhong W, Huang Y, Ma Z, Dong D, Liang C, Tian J (2016) Non-small cell lung cancer: quantitative phenotypic analysis of CT images as a potential marker of prognosis. Scientific reports 6:38282
40. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, Court LE (2014) Prognostic value and reproducibility of pretreatment

CT texture features in stage III non-small cell lung cancer. International Journal of Radiation Oncology* Biology* Physics 90 (4):834-842

41. Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, Aerts HJ (2017) Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. Journal of Thoracic Oncology 12 (3):467-476