SPECIAL SECTION: RADIOGENOMICS

Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT

Heidi Coy[1] · Kevin Hsieh[2] · Willie Wu[3] · Mahesh B. Nagarajan[1] · Jonathan R. Young[4] · Michael L. Douek[1] · Matthew S. Brown[1] · Fabien Scalzo[5] · Steven S. Raman[6]

Abstract

Purpose Currently, all solid enhancing renal masses without microscopic fat are considered malignant until proven otherwise and there is substantial overlap in the imaging findings of benign and malignant renal masses, particularly between clear cell RCC (ccRCC) and benign oncocytoma (ONC). Radiomics has attracted increased attention for its utility in pre-operative work-up on routine clinical images. Radiomics based approaches have converted medical images into mineable data and identified prognostic imaging signatures that machine learning algorithms can use to construct predictive models by learning the decision boundaries of the underlying data distribution. The TensorFlow™ framework from Google is a state-of-the-art open-source software library that can be used for training deep learning neural networks for performing machine learning tasks. The purpose of this study was to investigate the diagnostic value and feasibility of a deep learning-based renal lesion classifier using open-source Google TensorFlow™ Inception in differentiating ccRCC from ONC on routine four-phase MDCT in patients with pathologically confirmed renal masses.

Methods With institutional review board approval for this 1996 Health Insurance Portability and Accountability Act compliant retrospective study and a waiver of informed consent, we queried our institution's pathology, clinical, and radiology databases for histologically proven cases of ccRCC and ONC obtained between January 2000 and January 2016 scanned with a an intravenous contrast-enhanced four-phase renal mass protocol (unenhanced (UN), corticomedullary (CM), nephrographic (NP), and excretory (EX) phases). To extract features to be used for the machine learning model, the entire renal mass was contoured in the axial plane in each of the four phases, resulting in a 3D volume of interest (VOI) representative of the entire renal mass. We investigated thirteen different approaches to convert the acquired VOI data into a set of images that adequately represented each tumor which was used to train the final layer of the neural network model. Training was performed over 4000 iterations. In each iteration, 90% of the data were designated as training data and the remaining 10% served as validation data and a leave-one-out cross-validation scheme was implemented. Accuracy, sensitivity, specificity, positive (PPV) and negative predictive (NPV) values, and CIs were calculated for the classification of the thirteen processing modes.

Results We analyzed 179 consecutive patients with 179 lesions (128 ccRCC and 51 ONC). The ccRCC cohort had a mean size of 3.8 cm (range 0.8–14.6 cm) and the ONC cohort had a mean lesion size of 3.9 cm (range 1.0–13.1 cm). The highest specificity and PPV (52.9% and 80.3%, respectively) were achieved in the EX phase when we analyzed the single mid-slice of the tumor in the axial, coronal and sagittal plane, and when we increased the number of mid-slices of the tumor to three, with an accuracy of 75.4%, which also increased the sensitivity to 88.3% and the PPV to 79.6%. Using the entire tumor volume also showed that classification performance was best in the EX phase with an accuracy of 74.4%, a sensitivity of 85.8% and a PPV of 80.1%. When the entire tumor volume, plus mid-slices from all phases and all planes presented as tiled images, were submitted to the final layer of the neural network we achieved a PPV of 82.5%.

Conclusions The best classification result was obtained in the EX phase among the thirteen classification methods tested. Our proof of concept study is the first step towards understanding the utility of machine learning in the differentiation of ccRCC from ONC on routine CT images. We hope this could lead to future investigation into the development of a multivariate machine learning model which may augment our ability to accurately predict renal lesion histology on imaging.

Extended author information available on the last page of the article

Introduction

Renal cell carcinoma (RCC), the most common malignancy of the kidney, accounts for approximately 2–3% of all cancers, with a 2% increase in annual incidence [1]. Up to 70% of renal neoplasms are detected incidentally at imaging, and subtyping at the time of initial RCC presentation is clinically relevant, as patients with more indolent disease may undergo active surveillance, while those with more aggressive disease may be triaged to ablation or surgical resection [2–4]. Currently, all solid enhancing renal masses without microscopic fat are considered malignant until proven otherwise, and there are overlapping imaging features between clear cell RCC (ccRCC) and benign oncocytoma (ONC). In addition, multifocal renal ONCs are common, making the diagnosis of ONC with concomitant ccRCC a diagnostic challenge [5]. Since histopathological diagnosis requires a biopsy or surgery, a reliable non-invasive method to characterize incidental renal lesions would have clinical value [3, 6].

In the era of personalized medicine, radiomics may prove useful, converting medical images into datasets with prognostic imaging signatures. This may be useful for machine learning algorithms to construct predictive models by learning the decision boundaries of the underlying data distribution. [7–9]. It has shown promise in the diagnosis of lung, head and neck, breast, and brain tumors [10–16], but has not been validated in RCC, due in part to the genetic heterogeneity of RCC tumors, and the overlap in enhancement features on imaging [2, 17, 18]. As multidetector CT (MDCT) has been advocated for characterization of common renal lesions [2, 3], radiomics may provide a novel approach for determining histology. Previous studies have correlated radiomic features to tumor characteristics including histology, tumor grade, genetic and molecular characteristics, as well as clinical outcomes in patients with renal masses [18–35].

Deep learning may assist radiologists with characterization of incidentally detected renal masses, especially distinguishing ccRCC from its most common benign mimic, ONC utilizing a multiphasic MDCT study routinely used for staging and characterization. Using large MDCT image datasets, radiomics may add a new dimension to personalized medicine by capturing the intra-tumor heterogeneity within renal masses [17]. The TensorFlow™ framework from Google is a state-of-the-art open-source software library that can be used for training deep learning neural networks for performing machine learning tasks. Given the limited size of our dataset of ccRCC and ONC tumors with a pre-operative four-phase MDCT, we made use of transfer learning as implemented through Google's Inception architecture, which allowed us

to exploit a previously trained multi-layer neural network while simply retraining the final layer of the neural network model to develop a deep learning-based renal lesion classifier to discriminate ccRCC from ONC on routine four-phase MDCT images. Multi-layer neural network models typically comprise of several computational layers, each of which functions to abstract the input data in different ways, usually in increasing levels of complexity. This allows them to be very sensitive to complex features in large datasets.

Previous studies have examined the utility of Tensor-Flow™ Inception in various aspects of clinical image interpretation including development of a machine learning-based model for the binary classification of chest x-ray abnormalities [36], development of a breast cancer risk model using a mammographic dataset [37], and classification of striatal signal changes on 123I-ioflupane (DaTscan) SPECT studies in patients with Parkinson's disease [38]. In this study, we investigated the diagnostic value and feasibility of a deep learning-based renal lesion classifier using open-source Google TensorFlow™ Inception in differentiating ccRCC from ONC on routine four-phase MDCT in patients with pathologically confirmed renal masses.

Materials and methods

Patients and lesions

With institutional review board approval for this 1996 Health Insurance Portability and Accountability Act compliant retrospective study and a waiver of informed consent, we queried our institution's pathology database for histologically proven cases of ccRCC and ONC from January 2000 to January 2016. We then searched our institution's medical records to identify the date of surgery or biopsy, and our image archiving system (Centricity; GE Medical Systems, Milwaukee, WI) to identify those patients who had a preoperative MDCT with a four-phase renal protocol. In those patients who had multiple lesions, in order to avoid potential clustering effects, only one lesion per patient was selected at random and included in the analysis. Figure 1 summarizes the selection of the patient population used in our current study. In three prior studies [29, 30, 35], we reported on the 179 patients included in the current study. The prior studies examined the performance of qualitative and quantitative enhancement features as determined by two expert radiologists and a traditional threshold based CAD algorithm in differentiating clear cell RCC from other RCC subtypes. The objective of this current study was to discriminate ccRCC
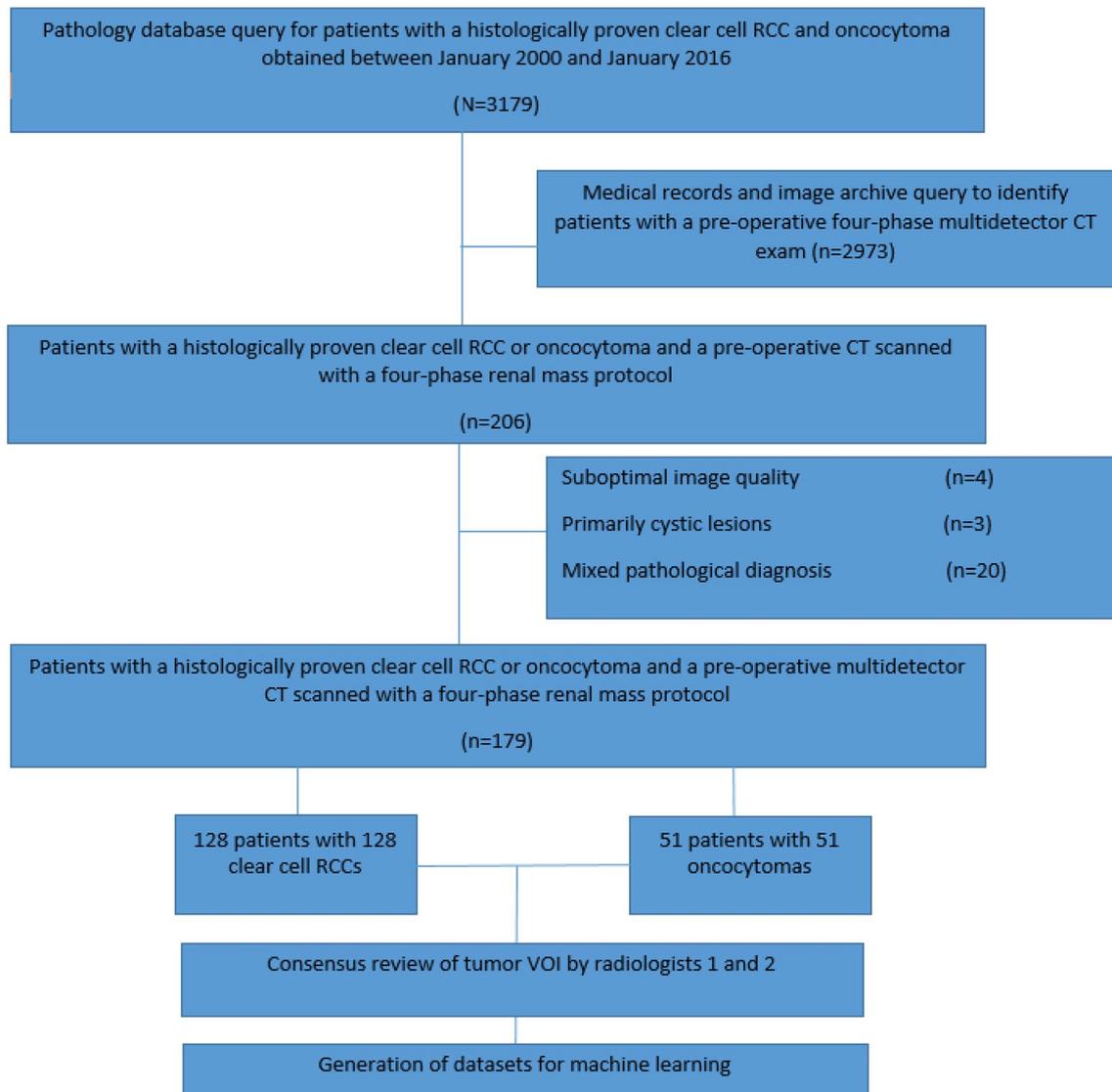
Fig. 1 Patient flow diagram. *CT* computed tomography, *RCC* renal cell carcinoma, *VOI* volume of interest

from ONC with a deep learning-based classifier using open-source Google TensorFlow™ Inception.

Four-phase MDCT acquisition

All patients had a pre-operative MDCT at our institution with an intravenous contrast-enhanced four-phase renal mass protocol. This included an unenhanced (UN) scan and contrast-enhanced acquisitions during the corticomedullary (CM), nephrographic (NP), and excretory (EX) phases acquired from the lung bases to the iliac crests. Pre-operative MDCT was performed on multiple scanners (4-MDCT, 16-MDCT, and 64-MDCT). Acquisition parameters were similar over the multiple generations of scanners: 120 KVp, 200–400 mA based on patient size and 3-mm scan collimation with 3-mm reconstruction intervals. Patients received a power injection of non-ionic IV contrast material (iohexol, Omnipaque 350, GE Healthcare), 35–45 g iodine dosed to weight at a rate of 3 mL/s. Patients weighing less than 45 kg, 45–90 kg, and > 90 kg received 100 mL (35 g iodine), 125 mL (45 g iodine), and 150 mL (54 g iodine) of iohexol, respectively. A bolus tracking system (Care Bolus VB10, Siemens Healthcare or Smart-Prep, GE Healthcare) was used to determine the onset of contrast-enhanced imaging and a region of interest (ROI) was placed in the thoracoabdominal aorta junction, with a trigger set to begin at 150 HU. Images were acquired at 55 s (CM), 120 s (NP), and 8 min (EX) after the threshold of 150 HU was reached in the thoracoabdominal ROI. For the patients imaged with earlier generation helical scanners (4- and 16-MDCT), the

CM and NP phases were performed 15 and 10 s earlier to achieve complete volumetric coverage. Overall, there was very little variation in the timing of image acquisition: all included patients were imaged within 5 s of stated times.

CT image analysis

Tumor volume of interest

To extract features to be used for the deep learning model, the entire renal mass was contoured in the axial plane in each of the four phases using an in-house developed FDA 510 K cleared software. Each slice of the segmentation was manually edited to ensure the contour was limited to the target lesion and was slightly within the lesion borders to exclude volume averaging from adjacent perirenal fat or normal renal parenchyma. This resulted in a 3D volume of interest (VOI) representative of the entire renal mass (Fig. 2). All VOIs were reviewed by two board certified genitourinary fellowship trained abdominal radiologists with 20 (blinded for review) and 11 (blinded for review) years of clinical experience, both of whom were blinded to all patient information, including the histopathologic diagnosis. The two radiologists reviewed and confirmed the 3D VOI independently to ensure only tumor was included in the VOI, but did not record a diagnosis. All segmentation tumor VOI discrepancies were resolved by consensus. Histopathologic assignment after nephrectomy or biopsy served as the reference standard for all tumors.

Deep learning analysis

Google TensorFlow™ Inception

Currently, TensorFlow™ Inception only supports 2D images as input. Given that the CT imaging data were acquired as a volumetric set of DICOM images, we investigated thirteen different approaches to convert the acquired volumetric data into a set of 2D Joint Photographic Experts Group (JPEG)

images that adequately represented each tumor. A detailed summary of the individual processing modes is available in Table 1. The DICOM data to JPEG conversion was performed using in-house software developed on Python 3.7 using the Python Imaging Library (PIL) module with default JPEG compression settings [39].

Processing modes

First, for each tumor, we obtained the largest tumor diameter in the axial, coronal and sagittal planes and converted it to a JPEG image as it has been shown that encompassing all three planar dimensions on post-contrast CT can improve discrimination between ONC and ccRCC [18]. All three planar dimensions were submitted as a grayscale JPEG image in each of the three post-contrast phases to the final layer of TensorFlow™ Inception. This resulted in three datasets, one for each post-contrast phase (CM, NP, and EX) referred to in shorthand as **ACM1** (Fig. 3), **ANM1** and **AEM1**, respectively.

Second, we used the largest tumor diameter, as well as the image before it and after it, in the axial, coronal, and sagittal planes. This allowed the tumor to be characterized using a greater volume, while maintaining the constraint of TensorFlow™ Inception, which requires use of a single 2D image to represent the tumor. Additionally, In order to allow for a richer characterization of homogeneity and heterogeneity within the tumor [17] we encoded the three slices as the red, green, and blue (RGB) channels of a JPEG image. When the R, G, and B values are similar, the result is a neutral color (similar to the grayscale images), but when the R, G, and B values are different, the colors stray from grayscale with stronger red, blue, or green components. Thus, the heterogeneity or homogeneity in the central location of the tumor can be better captured through our RGB approach; rather than using a single slice alone. Given that the differences in heterogeneity of ccRCC and ONC is an important property when a radiologist is determining its classification on routine clinical images [21], we hoped the level of detail captured with
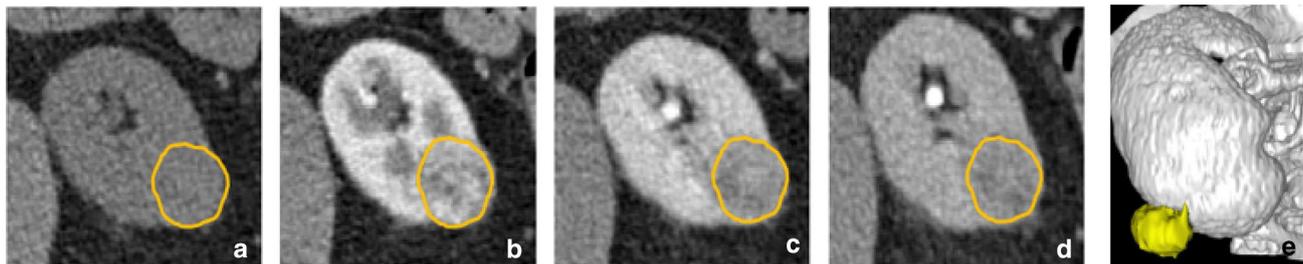

Fig. 2 An example of how a VOI for each tumor was obtained. Figure shows a clear cell RCC at axial multidetector CT in a 67-year-old man that was segmented in the unenhanced (**a**), corticomedullary (**b**), nephrographic (**c**) and excretory phase (**d**) resulting in a 3D VOI representative of the entire lesion in each phase (**e**). *RCC* renal cell carcinoma, *CT* computed tomography, *VOI* volume of interest

Table 1 Summary of the notation for each image processing mode submitted to the final layer of the neural network used to classify clear cell RCC and oncocytoma on MDCT using the Google TensorFlow™ Inception framework

Processing mode	Imaging plane	Imaging phase on MDCT	Slice	Grayscale or RGB JPEG	Single or triplet
ACM1	Axial, coronal and sagittal	Corticomedullary	Mid-slice of tumor	Grayscale	Single
ANM1	Axial, coronal and sagittal	Nephrographic	Mid-slice of tumor	Grayscale	Single
AEM1	Axial, coronal and sagittal	Excretory	Mid-slice of tumor	Grayscale	Single
ACM3	Axial, coronal and sagittal	Corticomedullary	Middle 3 slices	RGB	Triplet
ANM3	Axial, coronal and sagittal	Nephrographic	Middle 3 slices	RGB	Triplet
AEM3	Axial, coronal and sagittal	Excretory	Middle 3 slices	RGB	Triplet
ZC	Axial	Corticomedullary	Total tumor volume	RGB	Triplet
ZN	Axial	Nephrographic	Total tumor volume	RGB	Triplet
ZE	Axial	Excretory	Total tumor volume	RGB	Triplet
ATM3	Axial, coronal and sagittal	Unenhanced, corticomedullary, nephrographic and excretory	Mid-slice of tumor	RGB	Triplet

A all imaging planes (Axial, Sagittal, Coronal), *C* corticomedullary phase, *E* excretory phase, *JPEG* joint photographic experts group, *M* largest tumor diameter located in the middle of the 3D tumor contour, *MDCT* multidetector computed tomography, *N* nephrographic phase, *RGB* red, green, blue, *T* tiled phases, *Z* entire tumor volume, *1* single. *3* triplet

Corticomedullary Phase
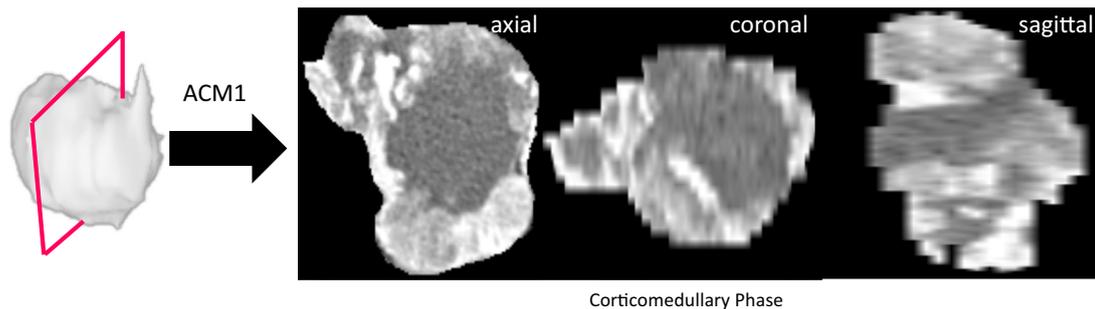
Fig. 3 Example of the ACM1 processing mode

our RGB approach could improve classifier performance. We created three such datasets, one for each of the three post-contrast phases (CM, NP, and EX); referred to in shorthand as ACM3, ANM3, and AEM3 (Fig. 4), respectively.

Third, an RGB JPEG image from three mid-slices of the axial, sagittal, and coronal plane was created from UN, CM, NP, and EX phases. We incorporated the UN phase in this processing mode to determine if this would improve the classification accuracy, as it has shown some utility in the differentiation of ccRCC from ONC on MDCT [34]. This resulted in 12 RGB images (4 phases x 3 cutting planes) used to represent the corresponding tumor in the dataset; referred to in shorthand as ATM3 (Fig. 5).

Fourth, as prior work has shown that whole lesion parameters provide a slight improvement in the discrimination of ccRCC from ONC on multiphasic CT [19], we incorporated the entire tumor volume acquired in the axial plane to determine if this would improve the classification accuracy of ccRCC and ONC. These were also encoded in groups of three adjacent slices as the red, green, and blue channels

of a RGB JPEG image. We created three such datasets, one for each of the three post-contrast phases (CM, NP, and EX); referred to in shorthand as ZC (Fig. 6), ZN, and ZE, respectively.

Finally, based on the performance of the individual processing modes, we combined the modes with the best performance to see if this would improve the classification accuracy.

Supervised learning

For the machine learning task, we used the Tensorflow API from Google™ [40]. Given the limited size of the dataset, building an entire multi-layer neural network from scratch was not feasible. Instead, we pursued a simple transfer learning approach that allows us to exploit a previously trained neural network model and retrain the existing weights for new classes; the image data from the current study were used to train the final layer of the neural network model.
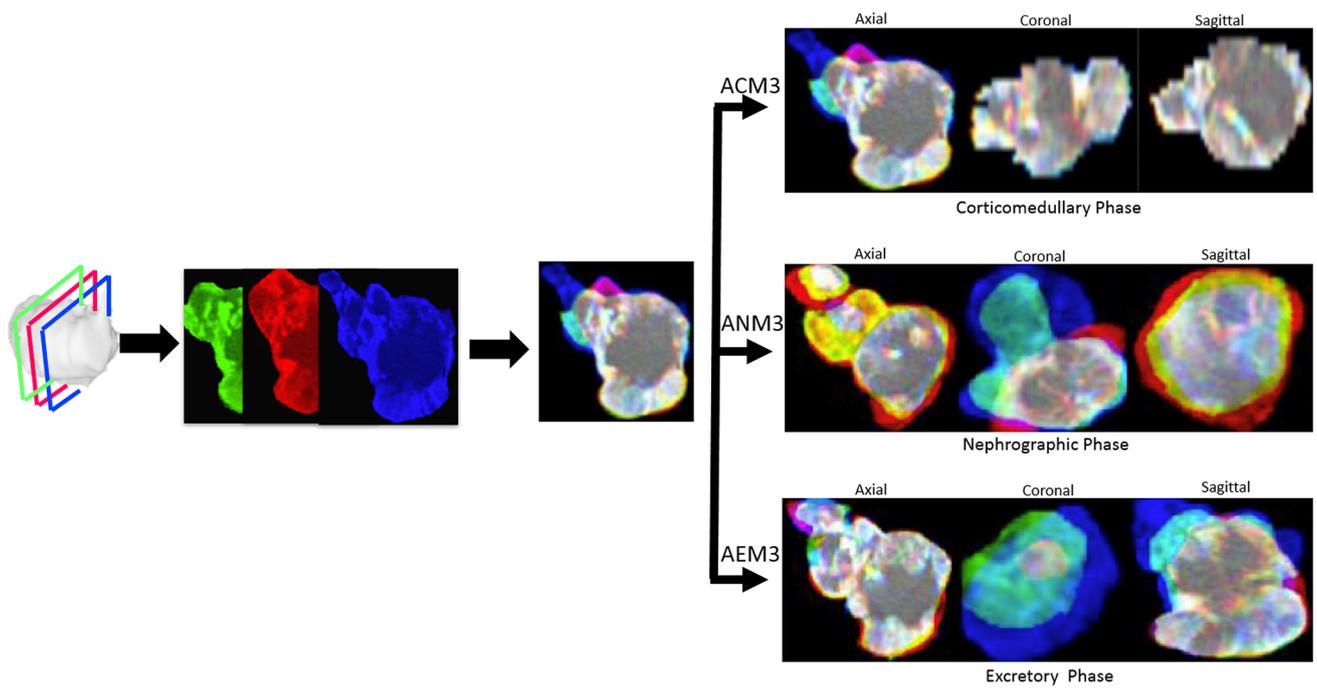
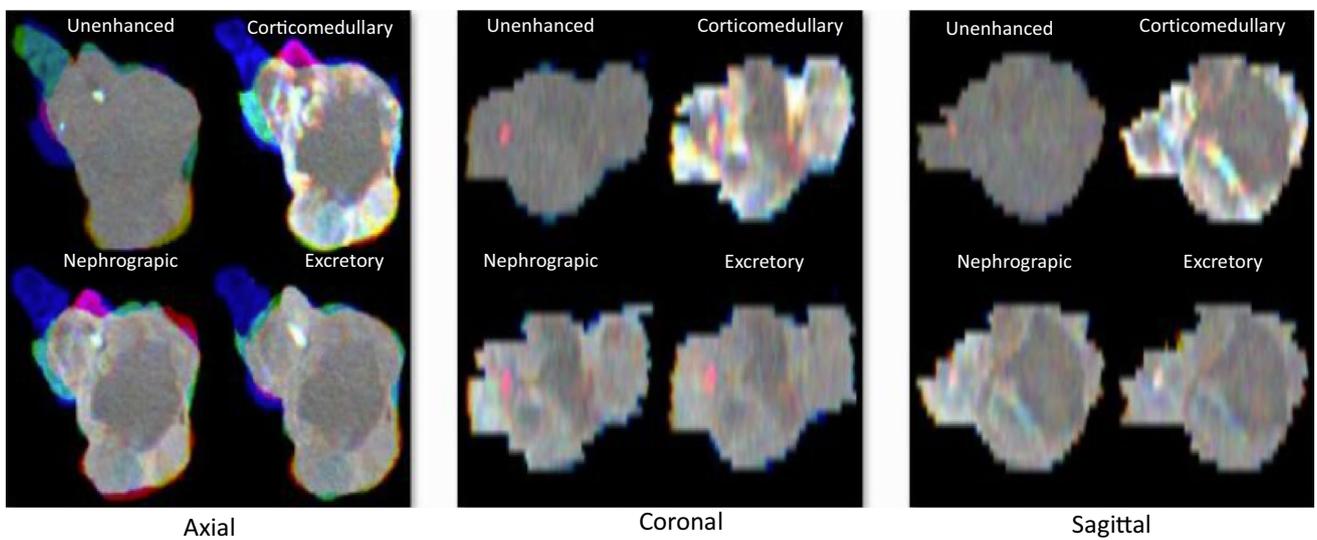**Fig. 4** Example of the processing modes using a RGB JPEG triplet image
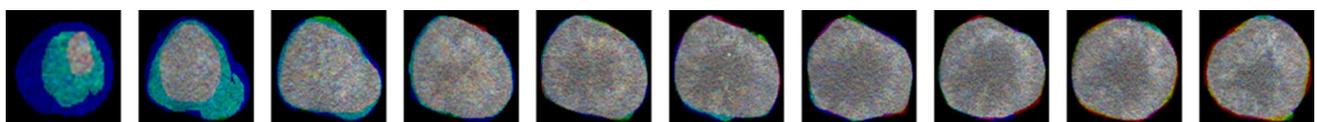

Fig. 5 Example of the ATM3 processing mode

Fig. 6 Example of the ZC processing mode

Neural network training

In this study, we used the Inception v3 architecture model [41] which was previously trained on images with ImageNet [42]. The top layer of the Inception v3 model outputs a 2048-D vector (also known as bottleneck in TensorFlow terminology) for each input image. We then trained a Softmax layer to serve as the final layer of the model; this layer would assign fuzz probabilities for any specific tumor as belonging to either ONC or ccRCC class of tumors. Since the classification task involved 2 classes, this training step corresponded to learning 2 + 2048*2 bias and weight parameters of the neural network model.

Training was performed over 4000 iterations. In each iteration, 90% of the data were designated as training data and the remaining 10% served as validation data to assist in optimization of model parameters while preventing any over-fitting and maximize generalizability of the model. Test data were separated prior to the training; therefore, no data were designated as test data during the neural network training with the Inception model.

Classification

A leave-one-out cross-validation scheme was implemented where data from 178 subjects were used for training to build a classifier model which was then evaluated on the 179th subject. Since the tumor in each subject is represented by multiple images, the classifier was first used to generate fuzzy probabilities for each image as belonging ONC or ccRCC classes. These fuzzy probabilities were thresholded at 50% to determine the classified label for each image. An unweighted voting scheme using the labels of these individual images determined the final tumor classification for that subject (ccRCC = positive, ONC = negative). This process was repeated 179 times in order to get a distribution of classifier performance metrics. The final classification task was implemented using in-house software developed with Python 3.7.

Statistical analysis

Statistical analysis was performed using statistical software (MedCalc, version 18, MedCalc Software). Accuracy, sensitivity, specificity, positive (PPV) and negative predictive (NPV) values, and CIs were calculated. The CIs for accuracy, sensitivity, and specificity are exact Clopper–Pearson CIs. The CIs for the PPV and NPV are calculated using the log method. All CIs are expressed at 95%.

Results

Patient and lesion characteristics

We analyzed 179 consecutive patients with 179 lesions (128 ccRCC and 51 ONC). Table 2 describes the clinicopathologic characteristics of the entire cohort. The mean patient age was 64 years (range 22–91). The majority of the patients with ccRCC had a T1 (78%) lesion with a mean size of 3.8 cm (range 0.8–14.6 cm). The mean lesion size for patients with an ONC was 3.9 cm (range 1.0–13.1 cm) with no significant differences in lesion size between the two cohorts. For all patients, the mean time from pre-operative CT to surgery or biopsy was 68 days (range, 0–865 days). Histopathology was diagnosed in 58 (32%) lesions from 20-gauge percutaneous core biopsy, 77 (43%) lesions from partial nephrectomy, 26 (15%) lesions from total nephrectomy, and 18 (10%) lesions from radical nephrectomy (Table 2).

Diagnostic performance metrics

When we compared the single mid-slice of the tumor in the axial, coronal, and sagittal plane submitted as a single image to the final layer of TensorFlow™ in the CM, NP and EX phases (processing modes ACM1, ANM1, and AEM1, respectively), we found a similar accuracy, sensitivity, and NPV among the three methods. The highest specificity and PPV (52.9% and 80.3%, respectively) were achieved in the EX phase (processing mode AEM1) (Table 3).

When we increased the number of mid-slices of the tumor to three, and encoded them as a RGB JPG image submitted to the final layer of TensorFlow™ in the CM, NP and EX phases (processing modes ACM3, ANM3, and AEM3, respectively), we found the highest classification accuracy was also in the EX phase, with an accuracy of 75.4%. Using the AEM3 mode in the EX phase also increased the sensitivity to 88.3% and the PPV to 79.6%. The specificity ranged from 39.2% to 45.1% using triplet images, and the highest NPV achieved was 59.5% in the EX phase with the AEM3 mode. (Table 3).

Using the entire tumor volume (processing modes ZC, ZN, ZE) in the axial plane also showed that classification performance was best in the EX phase with an accuracy of 74.4%, a sensitivity of 85.8% and a PPV of 80.1%. However, the specificity (44.9%) and NPV (55.0%) did not show any improvement with this method in this phase. Using this whole volume approach showed limited classification improvement in the CM and NP phases (Table 3).

The highest PPV (82.5%) was obtained using three mid-slices of the axial, sagittal and coronal plane in the UN,

Table 2 Patient and lesion characteristics

Characteristic	All lesions ($n=179$)	Clear cell RCC ($n=128$)	Oncocytoma ($n=51$)
Gender			
Male	115 (64)	80 (63)	35 (69)
Female	64 (36)	48 (37)	16 (31)
Mean age (years)*	66 (22–91)	62 (22–91)	69 (38–87)
Method of specimen acquisition			
Biopsy	58 (32)	33 (26)	25 (49)
Partial nephrectomy	77 (43)	56 (44)	21 (41)
Total nephrectomy	26 (15)	22 (17)	4 (8)
Radical nephrectomy	18 (10)	17 (13)	1 (2)
Mean time From pre-operative MDCT to biopsy or surgery (days)*	68 (0–865)	75 (0–756)	61 (0–865)
Laterality of tumor			
Left	94 (53)	68 (53)	26 (51)
Right	85 (47)	60 (47)	25 (49)
Mean lesion size (cm)*	3.9 (0.8–14.6)	3.8 (0.8–14.6)	3.9 (1.0–13.1)
Pathologic tumor stage			
T1a	–	73 (57)	–
T1b	–	27 (21)	–
T2	–	6 (5)	–
T3	–	22 (17)	–
T4	–	0 (0)	–

Data in parenthesis are the percentage unless otherwise noted

RCC renal cell carcinoma, *MDCT* multidetector computed tomography

*Data in parentheses are the range

Table 3 Performance of a deep learning classifier of clear cell RCC and oncocytoma using the Google TensorFlow™ Inception framework

Processing mode	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Single mid-slice					
ACM1	64.8% (57.3–71.8%)	77.3% (69.1–84.3%)	33.3% (20.8–47.9%)	74.4% (70.1–78.3%)	37.0% (26.2–49.2%)
ANM1	60.9% (53.3–68.1%)	71.9% (63.3–79.5%)	33.3% (20.8–47.9%)	73.0% (68.4–77.2%)	32.1% (22.7–43.2%)
AEM1	69.8% (62.5–76.5%)	76.6% (68.3–83.6%)	52.9% (38.5–67.1%)	80.3% (75.0–84.7%)	47.4% (37.5–57.5%)
Three mid-slice					
ACM3	69.3% (62.0–75.9%)	78.9% (70.8–85.6%)	45.1% (31.3–59.7%)	78.3% (73.5–82.5%)	46.0% (35.2–57.2%)
ANM3	69.8% (62.5–76.5%)	82.0% (74.3–88.3%)	39.2% (25.8–53.9%)	77.2% (72.8–81.1%)	46.5% (34.5–59.0%)
AEM3	75.4% (68.4–81.5%)	88.3% (81.4–93.3%)	43.1% (29.4–57.8%)	79.6% (75.3–83.3%)	59.5% (45.3–71.2%)
Whole tumor volume					
ZC	69.5% (62.1–76.3%)	79.4% (71.3–86.1%)	43.8% (29.5–58.8%)	78.7% (74.0–82.8%)	44.7% (33.6–56.4%)
ZN	71.6% (64.3–78.1%)	84.3% (76.7–90.1%)	38.8% (25.2–53.8%)	78.1% (73.8–81.7%)	48.7% (35.8–61.9%)
ZE	74.4% (67.2–80.7%)	85.8% (78.5–91.4%)	44.9% (30.7–59.8%)	80.1% (75.6–84.0%)	55.0% (41.9–67.5%)
Tiled (all axes and all phases)					
ATM3	72.6% (65.5–79.0%)	82.8% (75.1–88.9%)	47.1% (32.9–61.5%)	79.7% (75.0–83.7%)	52.2% (40.3–63.8%)
Combined modes					
ATM3 + ZC	69.8% (62.5–76.5%)	73.4% (64.9–80.9%)	60.8% (46.1–74.2%)	82.5% (76.7–87.0%)	47.7% (38.8–56.7%)
ATM3 + ZN	76.0% (69.0 -82.0%)	85.2% (77.8–90.8%)	52.9% (38.5–67.1%)	82.0% (77.1– 86.0%)	58.7% (46.6–69.9%)
AEM3 + ZE	74.3% (67.3–80.5%)	82.0% (74.3–88.3%)	54.9% (40.3–68.9%)	82.0% (76.9–86.2%)	54.9% (43.8–65.5%)

RCC renal cell carcinoma, *PPV* positive predictive value, *NPV* negative predictive value

CM, NP, and EX phases combined with the full volume the CM phase and submitted to TensorFlow's final layer of the neural network as a set of 12 RGB tiled images to represent the tumor (processing mode ATM3 + ZC. Using this approach, the accuracy, sensitivity, specificity, and NPV remained similar to the other twelve methods (Table 3).

Discussion

Although MDCT is most commonly used for detection, visual characterization, and clinical staging of renal masses, there remains a need for a non-invasive risk stratification, particularly in challenging diagnostic cases that involve discrimination between the variably aggressive ccRCC, and benign ONC [43]. A multivariate diagnostic approach including a radiomics machine learning-based analysis and tumor biopsy, could improve confidence in the visual characterization to enable more appropriate triage. This model has shown promise in renal tumor classification and grading [16, 43, 44]. Whether it can assist the accurate differentiation of ccRCC from ONC on routine CT images has not yet been fully elucidated.

In this pilot study, we demonstrated that TensorFlow™ Inception shows promise as a machine learning classifier to triage malignant ccRCC lesions from ONC, as it achieved a high PPV of 82.5%, when the entire tumor volume, plus mid-slices from all phases and all planes presented as tiled images, were submitted to the final layer of the neural network. An advantage of using TensorFlow™ Inception is the ability to preprocess images to elucidate as much information as possible in order to capture the inherent radiomic and biologic heterogeneity of ccRCC [17, 21]. The study cohort has previously been reported on three prior studies [29, 30, 35] which investigated the ability of two experienced radiologists to distinguish ccRCC from ONC based on qualitative and quantitative enhancement features. Our current study had comparable, and in some instances higher, results with regards to improving the accuracy, sensitivity and PPV in distinguishing ccRCC from ONC. However, in the prior work, two experienced radiologists had a higher NPV and specificity indicating they were able to correctly classify an ONC more frequently than our current deep learning classifier. These results may due to our sample size and if we included more subjects to power the deep learning study, we could improve upon the NPV and specificity. It may also be due, in part, to the experience of the radiologists practicing at our institution which is a large tertiary referral center. Further investigation is needed with a larger sample size for the deep learning classifier in comparison to radiologists of varying level of experience to determine how to definitively improve upon classifying ONC.

Previous studies have investigated several quantitative methods on imaging, including enhancement thresholds and texture features in individual post-contrast phases in an attempt to improve diagnostic accuracy in distinguishing ccRCC from ONC [19–35]. When we investigated each individual post-contrast phase, we found the best classification result was obtained in the EX phase with processing mode AEM3, where the three mid-slices of the tumor in the axial, coronal, and sagittal planes were submitted to the final layer of the neural network, with an accuracy of 75.4% and a sensitivity of 88.3%. This was similar to prior work [30, 34, 35, 43] that has shown improved differentiation of ccRCC from ONC using enhancement and texture features quantified in the EX phase on MDCT, likely due to the differences in the cellularity and stroma between these two tumor types [5]. However our study also demonstrated further investigation is needed to improve ONC classification, as the NPV ranged from 32.1 to 59.5% and the specificity ranged from 33.3 to 60.8% among the thirteen processing modes analyzed.

To our knowledge, our study is the first to investigate the use of TensorFlow™ Inception to discriminate ccRCC from ONC on multiphasic CT. A limited number of studies have attempted machine learning applied to texture feature analysis to differentiate ONC from other tumor types. Yu et al. [26] and Raman et al. [45] found that the use of machine learning further improved upon the area under the curve (AUC) achieved with individual texture features. Both of these studies used the AUC as an indicator of classifier performance, but considering their limited data set and class imbalance, AUC is not the best metric to determine the diagnostic utility of a machine learning classifier [43, 46].

Our study had some limitations. First, we only reported on the utility of these methods in discriminating ccRCC from ONC. However, distinguishing between these two most commonly encountered lesions is likely most clinical impactful, as these are the most difficult to differentiate in routine clinical practice, and 15–20% of patients who undergo surgical intervention end up having a benign ONC [6, 44]. Second, one shortcoming of TensorFlow™ Inception is that although it is a pre-trained model, the ImageNet database which it was previously trained on is not restricted to medical images. It is possible that having a TensorFlow™ based neural network trained with only medical images could improve the classification performance. However, this involves building a significantly larger database of medical images, an activity that is currently underway at our institution. Third, TensorFlow™ Inception currently supports analysis of single images and not volumetric data. By analyzing the largest tumor diameter in each phase in the axial, coronal and sagittal planes, as well as adjacent slices, submitting the entire tumor volume as individual images, and investigating a tiled approach, we were able to characterize the tumor using a greater volume while maintaining the constraint of using 2D images to represent the

tumor. Fourth, approximately a third of the patients included in our study had their histological diagnosis obtained from percutaneous core biopsy. While biopsy has become a standard of care at our institution to diagnose renal tumors, and the specimens are reviewed by a genitourinary subspecialty pathologist, this may be a limitation due to the heterogeneity of the tumor type. Fifth, although the classes are unbalanced, they are comparable to those in previous investigations [18–34] and incidence of disease in the population [1, 47]. Our dataset was further limited because we only chose those patients with a dedicated four-phase renal mass protocol, however this reduced the variability inherent in different scanning protocols and CT scanners. Additionally, we used 90% of the data as a designated training dataset, and the remaining 10% served as the validation dataset to prevent any over-fitting and maximize the generalizability of the model.

As previous work has shown, evaluating individual parameters in isolation is not always useful in the differentiation of ccRCC from ONC [3, 21, 34, 45]. Our proof of concept study is the first step towards understanding the utility of machine learning. The ability to submit DICOM and/or 3D images to TensorFlow™ for implementing convolutional neural networks will be an exciting next step for machine learning in radiology as this is an open-source platform. Further investigation in a large prospective cohort is needed to further validate this measure and improve diagnostic accuracy in differentiating ccRCC from ONC on routine CT images. We hope this could lead to future investigation into the development of a multivariate machine learning model which may augment our ability to accurately predict common lesion histology on imaging. This may also entail correlating machine learning results with genomic data, and incorporation of these results into a nomogram to predict disease specific survival. Given that the genotype of a tumor is likely to influence its phenotype, which can manifest as differences in appearance on imaging, we hope these phenotypes can be further elucidated with machine learning to assist the clinician in triaging patients with a renal mass to the appropriate therapeutic pathway.

Compliance with ethical standards

Conflicts of interest All authors have no conflicts of interest.

Informed consent All data was acquired with IRB approval, followed HIPAA guidelines, and with a waiver of informed consent.

References

1. Global Burden of Disease Cancer Collaboration. JAMA Oncol 2015; 1: 505-527
2. Sasaguri K, Takahashi N. CT and MR imaging for solid renal mass characterization. European Journal of Radiology 2018; 99: 40-54
3. Silverman SG, Israel GM, Herts BR, Richie JP: Management of the incidental renal mass. Radiology 2008; 249: 16-31
4. Liu N, Huang D, Cheng X, et al. Percutaneous radiofrequency ablation for renal cell carcinoma vs. partial nephrectomy: Comparison of long-term oncologic outcomes in both clear cell and non-clear cell of the most common subtype. Urol Oncol 2017; 35(8): 530.e6. https://doi.org/10.1016/j.urolonc.2017.03.014.
5. Ishigami K, Jones AR, Dahmoush L, Leite LV, Pakalniskis MG, Barloon TJ. Imaging spectrum of renal oncocytomas: a pictorial review with pathologic correlation. Insights into Imaging 2015;6(1):53-64
6. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. Solid renal tumors: an analysis of pathological features related to tumor size. J Urol 2003; 170 (6 Pt 1): 2217-20
7. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. J. Mach. Learn. 2003; 3(3): 1157–1182
8. Aerts HJ, Velazquez E, Leijenaar R, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014; 5 (4006). https://doi.org/10.1038/ncomms5006.
9. Gillies RJ, Kinahan P, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016; 278: 563-577
10. Song J, Liu Z, Zhong W, et al. Non-small cell lung cancer: quantitative phenotypic analysis of CT images as a potential marker of prognosis. Sci. Rep 2016; 6: 38282
11. Kickingereder P, Burth S, Wick A, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. Radiology 2016; 280(3): 880–889
12. Vallières M, Freeman CR, Skamene SR , El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys. Med. Biol. 2015; 60(14), 5471–96.
13. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014; 5: 4006. https://doi.org/10.1038/ncomms5006
14. Hatt M, Tixier F, Pierce L, Kinahan PE, Le rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 2017; 44(1): 151-165.
15. Yip, S. S. F. & Aerts, H. J. W. L. Applications and limitations of radiomics. Phys. Med. Biol. 61(13), R150–R166 (2016)
16. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning method for Quantitative Radiomic Biomarkers. Sci Rep 2015; 5:13087. https://doi.org/10.1038/srep13087
17. Gerlinger, Rowan, Horswell S, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. New England Journal of Medicine 2012; 366(10) 883-92.
18. Yap F, Hwang D, Cen S, et al. Quantitative Contour Analysis as an Image-Based Discriminator between Benign and Malignant Renal Tumors. Urology 2018; 114:121-127
19. Chen F, Gulati M, Hwang D, et al. Voxel-based whole lesion enhancement parameters: a study of its clinical value in differentiating clear cell renal cell carcinoma from renal oncocytoma. Abdominal Radiology 2017; 42: 552-560
20. Jamshidi N, Jonasch E, Zapala M, et al. The Radiogenomic Risk Score: Construction of a Prognostic Quantitative, Noninvasive Image-based Molecular Assay for Renal Cell Carcinoma. Radiology 2015; 277:114-123
21. Lubner M, Stabo N, Abel EJ, Munoz del Rio A, Pickhardt P. CT Textural Analysis of Large Primary Renal Cell Carcinomas:

Pretreatment Tumor Heterogeneity Correlates With Histologic Findings and Clinical Outcomes. AJR 2016; 207: 96-105.

22. Yin Q, Hung SC, Wang L, et al. Associations between Tumor Vascularity, Vascular Endothelial Growth Factor Expression and PET/MRI Radiomic Signatures in Primary Clear Cell Renal Cell Carcinoma: Proof of Concept Study. Sci Rep 2017; 7: 43356

23. Shinagare A, Krajewski K, Braschi-Amirfarzan M, Ramaiya N. Advanced Renal Cell Carcinoma: Role of the Radiologist in the Era of Precision Medicine. Radiology 2017; 284(2): 333-351

24. Karlo C.A., Di Paolo P.L., Chaim J., et al. Radiogenomics of clear cell renal cell carcinoma: Associations between CT imaging features and mutations. Radiology. 2014; 270:464–471.

25. Shinagare A.B., Vikram R., Jaffe C., et al. Radiogenomics of clear cell renal cell carcinoma: Preliminary findings of the cancer genome Atlas-Renal Cell Carcinoma (TCGA-RCC) imaging research group. Abdom. Imaging. 2015;40:1684–1692

26. Yu H, Scalera J, Khalid M, et al. Texture analysis as a radiomic marker for differentiating renal tumors. Abdom Radiol 2017; 42 (10): 2470-2478

27. Lee HS, Hong H Jung DC, Park s, Kim J. Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification. Med Phys. 2017 Jul;44(7):3604-3614.

28. Jagga Z, Gupta D. Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms. BMC Proc 2014; 8 (Suppl 6): S2

29. Young JR, Margolis D, Sauk S, Pantuck AJ, Sayre J, Raman SS: Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT. Radiology 2013; 267(2):444-453

30. Lee-Felker S, Felker E, Tan N, et al: Qualitative and Quantitative MDCT Features for Differentiating Clear Cell Renal Cell Carcinoma From Other Solid Renal Cortical Masses. AJR Am J Roentgenol 2014; 203(5):W516-W524

31. Zhang J, Lefkowitz RA, Ishill NM, et al. Solid renal cortical tumors: differentiation with CT. Radiology 2007; 244(2):494-504

32. Pierorazio PM, Hyams ES, Tsai S, et al. Multiphasic attenuation patterns of small renal masses (≤4cm) on preoperative computed tomography: utility for distinguishing subtypes of renal cell carcinoma, angiomyolipoma, and oncocytoma. Urology 2013; 81(6):1265-1271

33. Bird V, Kanagarajah P, Morillo G, et al. Differentiation of oncocytoma and renal cell carcinoma in small renal masses (<4cm): the role of 4-phase computerized tomography. Worl J Urol 2011; 29:787-792

34. Pano B, Macias N, Salvador R, et al. Usefulness of MDCT to Differentiate Between Renal Cell Carcinoma and Oncocytoma: Development of a Predictive Model. AJR 2016; 764-774.

35. Coy H, Young JR, Douek M, et al. Quantitative computer-aided diagnostic algorithm for automated detection of peak lesion attenuation in differentiating clear cell from papillary and chromophobe renal cell carcinoma, oncocytoma, and fat-poor amgiomyolipoma on multiphasic multidetector computed tomography. Abdom Radiol 2017; 42(7): 1919-1928

36. Yates EJ, Yates LC, Harvey H. Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. Clinical Radiology 2018; 73(9): 827-831

37. Ha R, Chang P, Karcich J, et al. Convolutional Neural Network Based Breast Cancer Risk Stratification Using a Mammographic Dataset. Academic Radiology 2018 July 31. pii: S1076-6332(18)30334-9. doi: 10.1016/j.acra.2018.06.020. [Epub ahead of print]

38. Zhang YC, Kagen AC. Machine Learning Interface for Medical Image Analysis. J Digit Imaging 2017; 30(5): 615-621

39. Python Imaging Library (PIL). Available at http://www.pythonware.com/products/pil/. Accessed on 30 Jan 2018.

40. Tensorflow. Available at https://www.tensorflow.org. Accessed on 30 Jan 2018.

41. Szegedy C, Vanhoucke V, Loeffe S, Shlens J, and Wojna Z. "Rethinking the Inception Architecture for Computer Vision," arXiv: 1512.00567v3, 2015. arxiv.org/abs/1512.00567.

42. Russakovsky O, Deng J, Su H, et al. "Imagenet large scale visual recognition challenge," arXiv: 1409.0575v3, 2014. arxiv.org/abs/1409.0575.

43. Varghese B, Chen F, Hwang D, et al. Differentiation of Predominantly Solid Enhancing Lipid-Poor Renal Cell Masses by Use of Contrast-Enhanced CT: Evaluating the Role of Texture in Tumor Subtyping. AJR 2018; 211: 1-9

44. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012; 30(9): 1234-48

45. Raman SP, Chen Y, Schroeder JL, Huang P, Fishman EK. CT texture analysis of renal masses. Acad Radiol 2014; 21:1587–1596

46. Lobo JM, Jimenez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 2008; 17: 145-151

47. American Cancer Society (2018) Cancer facts & figures 2016. Atlanta: American Cancer Society

Affiliations

Heidi Coy[1] · Kevin Hsieh[2] · Willie Wu[3] · Mahesh B. Nagarajan[1] · Jonathan R. Young[4] · Michael L. Douek[1] · Matthew S. Brown[1] · Fabien Scalzo[5] · Steven S. Raman[6]

✉ Heidi Coy
hcoy@mednet.ucla.edu

Kevin Hsieh
kevin.hsieh@ucla.edu

Willie Wu
williewu@ucla.edu

Mahesh B. Nagarajan
mnagarajan@mednet.ucla.edu

Jonathan R. Young
jryoung@ucla.edu

Michael L. Douek
mdouek@mednet.ucla.edu

Matthew S. Brown
mbrown@mednet.ucla.edu

Fabien Scalzo
fab@cs.ucla.edu

Steven S. Raman
sraman@mednet.ucla.edu

1   Department of Radiological Sciences, David Geffen School
    of Medicine at UCLA, Ronald Reagan-UCLA Medical
    Center, 924 Westwood Boulevard, Suite 615, Los Angeles,
    CA 90024, USA

2   Department of Computer Science, School of Engineering
    at UCLA, NRB 635 Charles E. Young Dr. South Suite 116,
    BOX 951769, Los Angeles, CA 90095-1769, USA

3   Department of Bioengineering, School of Engineering
    at UCLA, NRB 635 Charles E. Young Dr. South Suite 116,
    BOX 951769, Los Angeles, CA 90095-1769, USA

4   Department of Radiology, City of Hope National Medical
    Center, 1500 East Duarte Road, Duarte, CA 91010, USA

5   Department of Neurology, David Geffen School of Medicine
    at UCLA, NRB 635 Charles E. Young Dr. South Suite 116,
    BOX 951769, Los Angeles, CA 90095-1769, USA

6   Department of Radiological Sciences, David Geffen School
    of Medicine at UCLA, Ronald Reagan-UCLA Medical
    Center, 757 Westwood Plaza, RRUMC 1621H, Box 957437,
    Los Angeles, CA 90095, USA