BJR

https://doi.org/10.1259/bjr.20200002

Received: 20 December 2019 | Revised: 30 March 2020 | Accepted: 23 April 2020

Cite this article as:
Oberai A, Varghese B, Cen S, Angelini T, Hwang D, Gill I, et al. Deep learning based classification of solid lipid-poor contrast enhancing renal masses using contrast enhanced CT. *Br J Radiol* 2020; **93**: 20200002.

FULL PAPER

Deep learning based classification of solid lipid-poor contrast enhancing renal masses using contrast enhanced CT

[1]ASSAD OBERAI, PhD, [2]BINO VARGHESE, [2]STEVEN CEN, PhD, [3]TOMAS ANGELINI, [2]DARRYL HWANG, PhD, [4]INDERBIR GILL, MD, [5]MANJU ARON, MD, [2]CHRISTOPHER LAU, BS and [2,4]VINAY DUDDALWAR, MD, FRCR

[1]Department of Aerospace and Mechanical Engineering, Univ. of Southern California, Los Angeles, CA, USA
[2]Department of Radiology, Univ. of Southern California, Los Angeles, CA, USA
[3]Department of Computer Science, Univ. of Southern California, Los Angeles, CA, USA
[4]Institute of Urology, Univ. of Southern California, Los Angeles, CA, USA
[5]Department of Pathology, Univ. of Southern California, Los Angeles, CA, USA

Address correspondence to: Prof Assad Oberai
E-mail: *aoberai@usc.edu*

Objective: Establish a workflow that utilizes convolutional neural nets (CNN) to classify solid, lipid-poor, contrast enhancing renal masses using multiphase contrast enhanced CT (CECT) images and to assess the performance of the resulting network.
Methods: In this institutional review board approved study of 143 patients with predominantly solid, lipid-poor, contrast enhancing renal lesions (46 benign and 97 malignant), patients with a pre-operative multiphase CECT of the abdomen and pelvis obtained between June 2009 and June 2015 were retrospectively queried. Benign renal masses included oncocytoma and lipid-poor angiomyolipoma and the malignant group included clear cell, papillary, and chromophobe carcinomas.
Region of interests of whole tumor volumes were manually segmented, and CT phase images with the largest cross-section of the segmented tumor in the axial plane were used for assessment. Post-surgical pathological evaluation was used to establish diagnosis.
The segmented images of renal masses were used as input to a CNN. The data were augmented and split into training (83.9%) and validation sets (16.1%) to determine the hyperparameters of the CNN. Thereafter. the performance of the resulting CNN was quantified using eight-fold cross-validation.
Results: The CNN-based classifier demonstrated an overall accuracy of 78% (95% confidence interval: 76–80%), sensitivity of 70% (95% confidence interval: 66–74%), specificity of 81% (79–83%) and an area under the curve of 0.82.
Conclusion: A CNN-based classifier to diagnose solid enhancing malignant renal masses based on multiphase CECT images was developed.
Advances in knowledge: It was established that a CNN-based classifier could be trained to accurately distinguish malignant renal lesions.

INTRODUCTION

Majority of renal tumors are asymptomatic and are identified incidentally in 27–50% of all patients that are imaged.[1,2] While all imaging modalities have their strengths and limitations, the widespread use of CT and contrast enhanced CT (CECT) imaging has led to the increased detection of kidney cancers that might have otherwise remained undetected. CECT images are generated by injecting an intravenous contrast agent into the subject and acquiring images during four distinct time points (pre-contrast, corticomedullary, nephrographic and excretory phases). Conventional diagnosis is based on visual qualification of CECT images.

The vast majority of incidental renal masses imaged with CECT are simple cysts. The cysts are classified using the Bosniak classification.[3] Amongst the non-cystic/solid-lesions there are a proportion of macroscopic fat-containing angiomyolipomas. The remainder are subcategorized into non-enhancing and enhancing during CECT. Of the two groups, the latter, that is predominantly solid renal lesions without macroscopic fat that enhance on CECT, include both benign and malignant tumors and is the cause of most diagnostic errors. Common malignant tumors in this group include clear cell renal cell carcinomas, papillary renal cell

carcinoma, and chromophobe renal cell carcinomas. Benign tumors include oncocytoma and lipid-poor angiomyolipoma.

Kidney cancer is among the 10 most common cancers.[4] In 2019, the American Cancer Society estimated 73,820 new cases of kidney cancer and 14,770 deaths from this disease. The 5 year survival rate reduces from 93% in low-risk groups to 69% in high risk groups of patients with localized kidney cancer. However, following the spread of cancer, viz. metastatic disease, these rates plummet to 12%.

Qualitative (visual) image analysis of CECT images is hindered by the inherent intra- and intertumor heterogeneity in renal masses, interobserver differences and errors, and the daunting task of comparing across different phases to glean differences in contrast. These challenges can be addressed by supplementing the visual subjective assessment with quantitative assessment through radiomics, *i.e.* the high-throughput extraction of quantifiable metrics from standard of care images. Modern machine learning algorithms, such as, convolutional neural nets (CNN)-based classifiers have the potential to take this one step further, in that they learn and then extract the most influential features for clinical decision-making.[5] In this study, we explore the use of these algorithms in accurately classifying renal tumors.

METHODS AND MATERIALS

Patient data

This IRB-approved study included 143 patients with predominantly solid, lipid-poor, contrast enhancing renal lesions who underwent partial or radical nephrectomy and had pre-operative multiphase CECT (Table 1). Patients who underwent pre-operative multiphase CECT of the abdomen and pelvis from June 2009 to June 2015 were retrospectively identified. Pathologic evaluation was performed by experienced genitourinary pathologists post resection. Uncommon tumors such as sarcomas and juxtaglomerular tumors were excluded.

CECT imaging

All scans were obtained using a previously reported imaging protocol[6,7] on the same scanner (Brilliance 64, Philips Healthcare) (Figure 1). The CT scans were obtained during patient breath-holding with the following parameters: 120 kVp, variable tube current, slice thickness of 0.5 mm with reconstruction interval of 2 mm. An unenhanced CT scan of the of the abdomen was obtained first, followed by three contrast-enhanced scans in the corticomedullary (30 s), nephrographic (90 s), and excretory (5–7 min) phases. Approximately 100–150 ml of non-ionic water-soluble i.v. iodinated contrast medium (iopamidol, Isovue 350, Bracco Imaging) dosed to weight was administered with a power injector at a rate of 5 ml s^{-1}.

Tumor segmentation and phase co-registration

Using Synapse 3D software (FujiFilm, Stamford, CT), an experienced radiologist manually segmented the renal tumors as three-dimensional (3D) regions of interest.[6,7] The nephrographic phase was used as the reference template for subsequent co-registering in other phases.[6,7] Two-dimensional images of all tumors capturing the largest tumor diameters in each phase in the axial projection were selected and used as inputs to the CNN model (Figures 1 and 2).

Neural network architecture

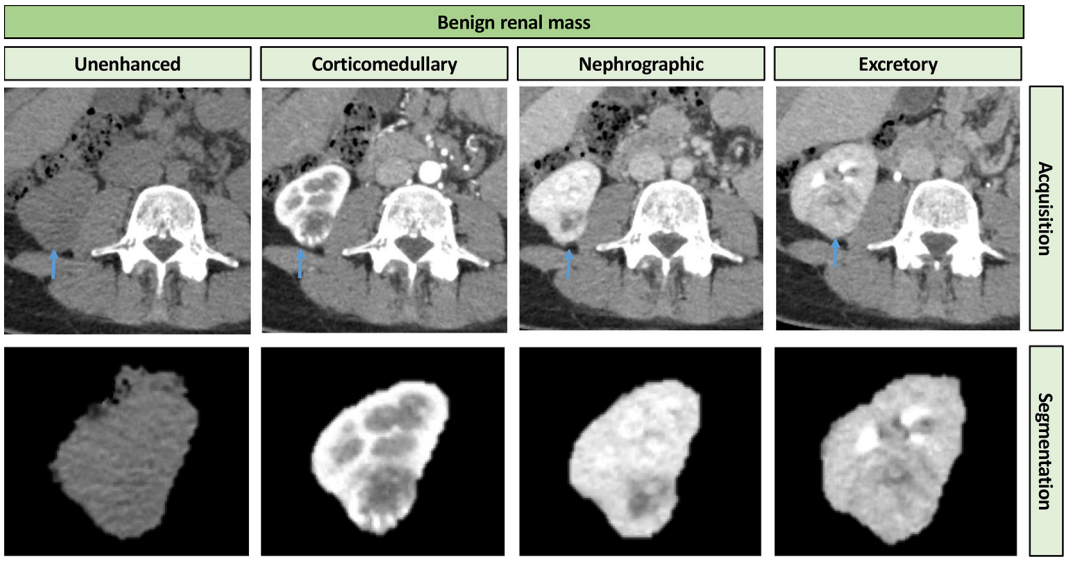
For the sake of simplicity, and in order to accommodate the relatively small amount of data, a simple CNN architecture was selected and trained end-to-end. The input to the CNN was the sequence of four segmented CECT images corresponding to the axial projection. The network comprised of three convolution layers, each with ReLU activation and a stride of 2. The first layer

Table 1. Description of study population

	Overall	Benign	Malignant
Sample size (N)	143	46	97
Age (Mean [Std.dev])	61.73 [12.81]	61.8 [14.5]	61.7 [12.1]
Gender	41F, 102M	17F, 29M	24F, 73M
Subclassification			
Benign (*N* = 46)	**Oncocytoma**	26	
	Size (mean ± sd)	3.4 ± 1.7 cm	
	Lipid-poor AML	20	
	Size (mean ± sd)	4.4 ± 3.3 cm	
Malignant (*N* = 97)	**ccRCC**	70	
	Size (mean ± sd)	3.68 ± 1.6 cm	
	pRCC	17	
	Size (mean ± sd)	4.4 ± 3.0 cm	
	chRCC	10	
	Size (mean ± sd)	3.7 ± 2.1 cm	

AML, Angiomylipoma; F, female; M, male; Std.dev, standard deviation; ccRCC, clear cell renal cell carcinoma; chRCC, chromophobe renal cell carcinoma; pRCC, papillary renal cell carcinoma.

Figure 1. Multiphase CT images demonstrate a heterogenous enhancing mass at the lower pole of the right kidney in a 43-year-old female which was later found to be an oncocytoma at surgery (top). Manually segmented regions of interest (renal mass) were used as inputs to the CNN model for tumor characterization (bottom). CNN, convolutional neural nets.



comprised of 64 channels of five-by-five filters, the second layer comprised of 128 channels of five-by-five filters, and the third layer comprised of 256 channels of three-by-three filters. The output of the third layer was flattened and transformed into a vector of size 64 through a fully connected layer, followed by a 2-class soft-max classifier. A dropout of 50% was implemented in the fully connected layer.

Training details

Each CECT image sequence was transformed through reflection and rotation into seven additional sequences for data augmentation (Figure 3). The data were split into a training (83.9%, 120 subjects) and another set to determine the hyperparameters of the network (16.1%, 23 subjects), while ensuring that the image sequences in each set belonged to distinct subjects. The hyperparameters included the number of layers, the size of the convolutions, the learning rate (0.3), batch size (20), and the number of epochs (200). A mean-square error loss and the stochastic gradient descent (SGD) optimizer were used. No regularization was used, and it was observed that the SGD minimizer produced stable results even in this limit. This remarkable stability of the SGD minimizer has been recently proven in.[8] Class imbalance was addressed by using a class weight of 2:1 (benign: malignant) in the loss function.

Figure 2. Multiphase CT images in an asymptomatic 67-year-old male with a heterogenous enhancing right renal mass identified incidentally that proved to be clear cell renal cell carcinoma (top). Manually segmented regions of interest (renal masses) were used as inputs to the CNN model for tumor characterization. CNN, convolutional neural nets.

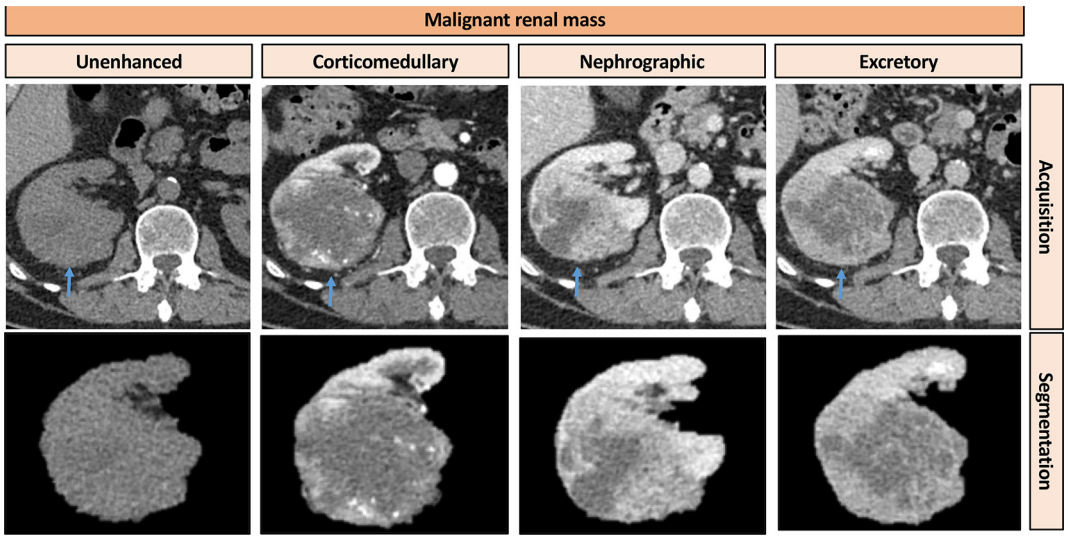
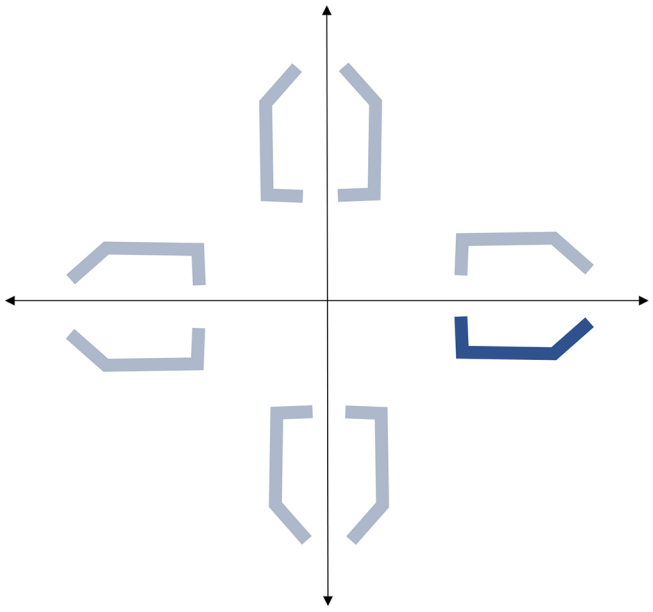


Figure 3. Data augmentation through rotation and reflection. A feature in the original image is represented in dark color. Seven copies obtained through rotation and reflection are shown in lighter color.



Statistical and data analysis

Eightfold cross-validation was performed to test the performance of the network. In every cross-validation run CECT image sequences from 128 distinct subjects were used to train the network; sequences from the remaining 15 subjects were used to test it. There was no overlap between testing sets. Data from the 23 subjects, which were used to determine the hyperparameters of the network, were excluded from the testing sets.

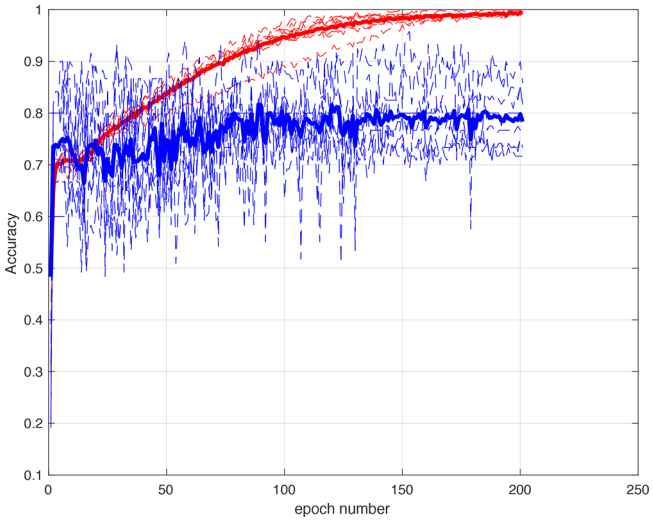
The output of the soft-max classifier is the predicted probability of belonging to the malignant class. When this probability is greater than a threshold, the network returns with a malignant diagnosis. Using the predicted probability from the cross-validation runs, the receiver operating characteristic (ROC) curve was generated by varying this threshold, and the area under the curve (AUC) was reported. Youden's index was used to further decide the diagnostic threshold in classifying malignant from benign lesions. The accuracy, sensitivity, and specificity of the network with binomial 95% confidence interval (CI) were reported based on the threshold.

RESULTS

Training details

The network described in the Methods section was implemented in Keras using Tensorflow.[9] All training and prediction runs were performed on a six core 2.9 GHz Intel Core i9 MacBook Pro machine. The average time for the network to train on sequence of images from 128 subjects was 16 min. The variation of prediction accuracy and test accuracy for each of the eight cross-validation runs is shown in Figure 4. With increasing epochs, the average training accuracy approached 1; while the average testing accuracy, corresponding to a threshold of 0.5 for the likelihood of malignancy, approached 0.79.

Figure 4. Training and testing accuracy (red and blue, respectively) as a function of epoch number for the cross-validation runs. Thin, dashed curves indicate individual runs and thick, solid curves indicate the average curve.



Statistical and data analysis

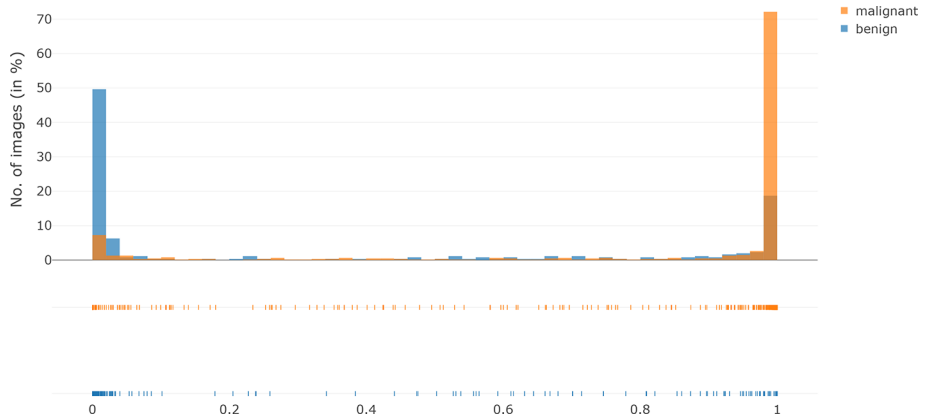
Each cross-validation run resulted in the probability of a malignant diagnosis for 120 images sequences (15 subjects times 8 augmented sequences). Across all eight cross-validation runs this resulted in 960 predictions of the probability of malignancy. A histogram of these data are shown in Figure 5. Here, the probability of malignancy for subjects with surgery-proven malignant diagnosis is shown in orange, and the predicted probability of malignancy for subjects with surgery-proven benign diagnosis is shown in blue. We note that the predicted probability for the surgery-proven malignant masses is clustered around one, while that for the surgery-proven benign masses is clustered around zero. This points to the efficacy of the network in successfully segregating these populations.

For the ROC curve constructed based on the 8-fold cross-validation predicted probability (Figure 6), an AUC of 0.82 (95% CI: 0.8–0.84) was determined. Youden's index suggested that a predicted probability threshold >0.27, should be used to classify a lesion as malignant. Based on this threshold, an overall accuracy of 78% (95% CI: 76–80%), sensitivity of 70% (95% CI: 66–74%), and specificity of 81% (95% CI: 79–83%) was determined.

DISCUSSION

We present a simple CNN-based framework to classify lipid-poor enhancing renal masses into "benign" and "malignant" categories using multiphase CECT images. When trained on 128 subjects this network yielded an accuracy of 78%, sensitivity of 70%, specificity of 81%, and an AUC of 0.82. While influenced by tumor size,[10] the reported sensitivity and specificity for predicting RCC based on CT imaging is between 60–79% and 44–100%, respectively.[11,12] Of the various renal masses, only an angiomyolipoma (AML) can be detected using CT alone, based on the visualization of fat. Techniques such as Tc99m sestamibi SPECT/CT imaging have also been shown to characterize indeterminate renal masses with a sensitivity and specificity of 87.5

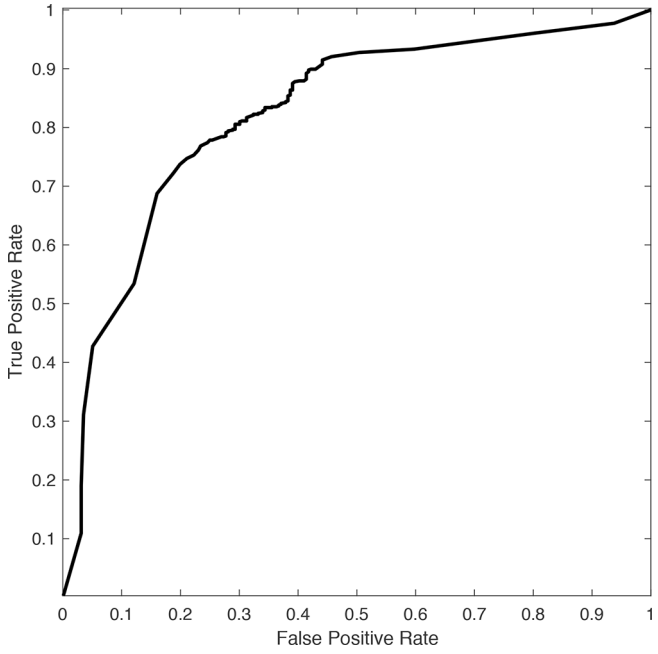
Figure 5. Histogram of the predicted likelihood of malignancy for the test cohort for true malignant (orange) and true benign (blue) sequences. In the lines below, each orange/blue tick mark indicates a true malignant/benign sequence.



and 95.2%, respectively.[13] However, Tc99m sestamibi is not available commonly and differentiates between benign and low grade malignant from more aggressive renal carcinomas. Our findings of an AUC of 0.82, with an overall accuracy of 78% (95% CI: 76–80%), sensitivity of 70% (95% CI: 66–74%), and specificity of 81% (95% CI: 79–83%) using a Youden's index of greater than 0.27 is better than visual assessment and comparable to Tc99m sestamibi SPECT/CT based assessment of renal masses. Our approach also has the advantage of using clinically performed imaging with no additional imaging.

To our knowledge, only three other studies have focused on characterizing lipid-poor renal masses using multiphase CECT and CNNs and none of these solve the precise classification problem we are solving. Further, in contrast to the present study, they all rely on transfer learning to handle the problem of paucity of data. In Coy et al,[14] Google's Inception model was selectively re-trained on CECT data using the so-called transfer learning approach[15] in order to distinguish between ccRCC and Oncocytoma. Using different types of input data streams ranging from single plane multiphase images in the simplest case, to multiphase images from multiple planes plus segmented 3D volume images in the most complex case, accuracies of 60.9–76% were reported. In Lee et al[16] also, transfer learning was used in conjunction with a complex input stream that comprised of around 70 handcrafted features and 4000 machine-learned features to achieve an accuracy of 76%. In contrast to these, we developed a relatively simple network that was trained end-to-end without transfer learning and achieved an accuracy of 79%. By virtue of its simplicity, our network is faster to train and use, and easier to analyze. Further, in contrast to previous studies, where the input stream of two-dimensional CECT images was supplemented with other complex data to achieve the best-case accuracy of 76%, our input stream utilized only CECT image slices, and was therefore easier to implement. Finally, we note that in Han et al,[17] the authors used a weighted sum of multiphase CECT images as input to a CNN-based classifier to solve the related problem of sub classifying different types of renal cell carcinomas (clear cell, papillary and chromophobe) and reported a classification accuracy of 85%. Similar to the studies described above, and in contrast to our approach, the authors in that study also used transfer learning to retrain an existing network (GoogLeNet[18]).

Figure 6. ROC curve for the eightfold cross-validation cohort generated using the likelihood of malignance predicted by the CNN-based classifier. CNN, convolutional neural nets; ROC, receiver operating characteristic.



A comparison of the performance of the CNN classifier with typical results reported from visual assessment which report AUC ~0.65[19] reveals that the CNN is more accurate. A comparison with similar studies performed with techniques that extract and quantify explicit features within images and report an AUC ~0.85[7] reveals that the CNN-based classifier performs similarly, while circumventing the tedious step of requiring the user to extract these image features.

It is worth noting that the CNNs developed in Pan Yang, Lee et al and Han et al and,[15–17] as well as the CNN described in this manuscript utilize images from about the same number of

subjects (100–200) for training. This number is much smaller than the number of images used to train CNNs for tasks related to natural images, which is typically around 10^5–10^6. In order to address this paucity of data, the authors in Pan Yang, Lee et al and Han et al [15–17] have relied on transfer learning. In this approach, the starting point is a pre-trained CNN whose weights have been determined for an image classification task that uses a large number of images. Thereafter, the weights of the outer layers of this network are re-evaluated when it is adapted to the task for which a small set of images are available. In our work, we employ a different approach. First, we employ data augmentation to increase the number of images in our training through image rotation and reflection. Second, and perhaps more importantly, we control the complexity of our network by ensuring that it is not very deep and therefore has relatively few parameters that are determined during training. For example, when compared with the Google Inception three network which has around 24 million parameters, the CNN in this manuscript has only 538 thousand parameters. This ensures that relatively small amount of data available to train this network are adequate. A test to determine whether the amount of data used to train an ML algorithm are adequate involves examining its generalizability. A network that suffers from overfitting will perform very well on the training set, but will fail on the test set. In other words, it will not generalize well to data that it has not seen. In our case, the accuracy of the net for the training set is around 0.99 and this drops to 0.79 for the cross-validation test set. This indicates that while the algorithm does generalize well enough for it to be useful, it will likely benefit from more data.

This was a retrospective study. Based on a previously published study from our group and analysis of existing knowledge, the subcohort most likely to lead to misclassification was the category with solid enhancing masses. Two additional classes of lesions were excluded. These included a lipid predominant lesion is visually consistent diagnosis of angiomyolipoma and cystic lesions which are visually classified using the Bosniak classification. We therefore selectively chose the included cohort of more difficult subset of patients. We have also previously reported on the radiomic evaluation of this subset.[7] We understand that subjective segmentation can affect radiomics results. To alleviate this issue, we conducted an interobserver agreement on tumor segmentation. For a subset of 15 renal masses, tumor boundaries were independently drawn by three radiologists (with varying experience of interpreting oncologic CECT images), who were blinded to the pathological results or to each other's segmentations. The inter-rater agreement among radiologists regarding manual tumor segmentations revealed an interclass correlation coefficient of 0.97 ± 0.07, indicating excellent reliability of the segmentation results.

The present study can be improved and extended in several ways. The inclusion of data from more patients would improve the robustness of the network, and facilitate the training of a deeper, more accurate, network. The inclusion of 3D image sequences might also improve the performance of the network. The current workflow involves the subjective step of image segmentation. An easier option would be to utilize deep learning algorithms to automate this task, or to work with unsegmented CT images and train the network to determine the most critical regions. An additional improvement would be to extend the network to multiple scanners and protocols, and all renal mass types. We used only the largest cross-sectional image for analysis. While this corresponds to the most common plane used for visual qualification in conventional clinical practice, this may affect classification and analysis in heterogenous tumors. In the future we will be addressing this issue withmultiple planes as well as a more comprehensive volumetric evaluation of the renal mass.

As with many other artificial intelligent applications, CNNs possess the capability to sift through copious volumes of data, here images, to identify disease patterns that can provide information regarding diagnosis, grading and prognosis. Once validated using multisite data, techniques such as ours can be used to form clinical decision support systems. These can be deployed within the clinical setting to improve clinical confidence and optimize treatment decisions.[20] While current applications of CNNs in medical imaging research are promising, better research and reporting using multi-institutional, large sample size data are needed to reliably evaluate the power of deep learning in healthcare settings.[21]

In summary, we have demonstrated the applicability of a deep-learning-based method in classifying malignant renal masses using multiphase CECT images. Using a modest amount of training data from 128 subjects, the network produced accuracy metrics that were better than visual quantification and compared well with the most sophisticated image processing approaches. It is likely that the accuracy and the robustness of this network, or others derived from it, will improve as more data are made available.

CONFLICT OF INTEREST

There was no conflict of interest in the data presented in this paper. Disclosures for the authors are, Dr. Vinay Duddalwar. Consultant: Radmetrics, Intuitive systems; Advisory board: DeepTek Grant recipient: Samsung healthcare. Prof. Assad Oberai. Grant recipient: Simmetrix Inc.; Grant recipient: IBM Corp.

FUNDING

This project has received funding from the Whittier Foundation. The project described was supported in part by award number P30CA014089 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health

REFERENCES

1. Loo RK, Lieberman SF, Slezak JM, Landa HM, Mariani AJ, Nicolaisen G, et al. Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. *Mayo Clin Proc* 2013; **88**: 129–38. doi: https://doi.org/10.1016/j.mayocp.2012.10.004
2. Ljungberg B, Hanbury DC, Kuczyk MA, Merseburger AS, Mulders PFA, Patard J-J, et al. Renal cell carcinoma guideline. *Eur Urol* 2007; **51**: 1502–10. doi: https://doi.org/10.1016/j.eururo.2007.03.035
3. Prasad SR, Dalrymple NC, Surabhi VR. Cross-Sectional imaging evaluation of renal masses. *Radiol Clin North Am* 2008; **46**: 95–111. doi: https://doi.org/10.1016/j.rcl.2008.01.008
4. Survival rates for kidney cancer. American cancer Society website..
5. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *in: Advances in neural information processing systems* 2012;pp. : 1097–105.
6. Chen F, Gulati M, Hwang D, Cen S, Yap F, Ugwueze C, et al. Voxel-Based whole-lesion enhancement parameters: a study of its clinical value in differentiating clear cell renal cell carcinoma from renal oncocytoma. *Abdom Radiol* 2017; **42**: 552–60. doi: https://doi.org/10.1007/s00261-016-0891-8
7. Varghese BA, Chen F, Hwang DH, Cen SY, Desai B, Gill IS, 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 Am J Roentgenol* 2018; **211**: W288–96. doi: https://doi.org/10.2214/AJR.18.19551
8. SS D, Lee JD, Li H, Wang L, Zhai X. Gradient descent finds global minima of deep neural networks. 2018; **03804**. arXiv preprint arXiv:1811.
9. Keras. Google Website.. 2017Last modified: October 25. Available from: https://www.tensorflow.org/guide/keras [Last accessed: March 24, 2019].
10. Thompson RH, Kurta JM, Kaag M, Tickoo SK, Kundu S, Katz D, et al. Tumor size is associated with malignant potential in renal cell carcinoma cases. *J Urol* 2009; **181**: 2033–6. doi: https://doi.org/10.1016/j.juro.2009.01.027
11. Kim JH, Sun HY, Hwang J, Hong SS, Cho YJ, Doo SW, et al. Diagnostic accuracy of contrast-enhanced computed tomography and contrast-enhanced magnetic resonance imaging of small renal masses in real practice: sensitivity and specificity according to subjective radiologic interpretation. *World J Surg Oncol* 2016; **14**: 260. doi: https://doi.org/10.1186/s12957-016-1017-z
12. Heidenreich A, Ravery V, .European Society of Oncological Urology Preoperative imaging in renal cell cancer. *World J Urol* 2004; **22**: 307–15. doi: https://doi.org/10.1007/s00345-004-0411-2
13. Gorin MA, Rowe SP, Baras AS, Solnes LB, Ball MW, Pierorazio PM, et al. Prospective Evaluation of (99m)Tc-sestamibi SPECT/CT for the Diagnosis of Renal Oncocytomas and Hybrid Oncocytic/Chromophobe Tumors. *Eur Urol* 2016; **69**: 413–6. doi: https://doi.org/10.1016/j.eururo.2015.08.056
14. Coy H, Hsieh K, Wu W, Nagarajan MB, Young JR, Douek ML, et al. Deep learning and radiomics: the utility of Google TensorFlow™ inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT. *Abdominal Radiology* 2019; **44**: 2009–20. doi: https://doi.org/10.1007/s00261-019-01929-0
15. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010; **22**: 1345–59. doi: https://doi.org/10.1109/TKDE.2009.191
16. Lee H, Hong H, Kim J, Jung DC. Deep feature classification of angiomyolipoma without visible fat and renal cell carcinoma in abdominal contrast-enhanced CT images with texture image patches and hand-crafted feature concatenation. *Med Phys* 2018; **45**: 1550–61. doi: https://doi.org/10.1002/mp.12828
17. Han S, Hwang SI, Lee HJ. The classification of renal cancer in 3-Phase CT images using a deep learning method. *J Digit Imaging* 2019; **32**: 638–43. doi: https://doi.org/10.1007/s10278-019-00230-2
18. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov Det al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015. pp. 1–9.
19. Shin T, Duddalwar VA, Ukimura O, Matsugasumi T, Chen F, Ahmadi N, et al. Does computed tomography still have limitations to distinguish benign from malignant renal tumors for radiologists? *Urol Int* 2017; **99**: 229–36. doi: https://doi.org/10.1159/000460303
20. Kunapuli G, Varghese BA, Ganapathy P, Desai B, Cen S, Aron M, et al. A decision-support tool for renal mass classification. *J Digit Imaging* 2018; **31**: 929–39. doi: https://doi.org/10.1007/s10278-018-0100-0
21. Choudhury A. Ai in healthcare: improving human interface for patient safety. better standards needed to make artificial intelligence user-friendly for clinicians. *Better Standards Needed to Make Artificial Intelligence User-Friendly for Clinicians* 2020;In, : 33–7. February 1, 2020.