IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 16, 2019, accepted January 3, 2020, date of publication January 8, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964755

Patch-Based Convolutional Neural Network for Differentiation of Cyst From Solid Renal Mass on Contrast-Enhanced Computed Tomography Images

FATEMEH ZABIHOLLAHY[1], N. SCHIEDA[2], AND E. UKWATTA[3], (Senior Member, IEEE)
[1]Department of System and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada
[2]Department of Radiology, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[3]School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada

Corresponding author: Fatemeh Zabihollahy (fatemehzabihollahy@cmail.carleton.ca)

The work of E. Ukwatta was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery.

ABSTRACT Automated classification of renal masses detected at computed tomography (CT) examinations into benign cyst versus solid mass is clinically valuable. This distinction may be challenging at single-phase contrast-enhanced CE-CT examinations, where cysts may simulate solid masses and where renal masses are most commonly incidentally detected. This may lead to unnecessary and costly follow-up imaging for accurate characterization. In this paper, we describe a patch-based CNN method to differentiate benign cysts from solid renal masses using single-phase CECT images. The predictions of the network for patches extracted from a manually segmented lesion are combined through the majority voting system for final diagnosis. We used a dataset comprised of single-phase CECT images of 315 patients with 77 benign (oncocytomas, and fat poor renal angiomyolipoma) and 238 malignant (renal cell carcinoma including clear cell, papillary, and chromophobe subtypes) tumors. We trained our proposed network using patches extracted and artificially augmented from 40 CECT scans. The presented algorithm was evaluated using 275 unseen CECT test images consisting of 327 renal masses by comparing algorithm-generated labels to those labeled by experts and achieved mean accuracy, precision, and recall of 88.96%, 95.64%, and 91.64%. Our method yielded accuracy of 91.21% $\pm$ 25.88% as mean $\pm$ standard deviation at the patient level. The AUC was reported as 0.804. The results indicate that our algorithm may accurately characterize benign cysts from solid masses with a high degree of accuracy and may be clinically valuable to prevent unnecessary imaging follow-up for characterization in a proportion of patients.

INDEX TERMS Renal mass, benign cyst, malignant, convolutional neural network.

I. INTRODUCTION

Renal cancer is among the 10 most common cancers and renal cell carcinoma (RCC) is the most common type of malignancy of the urogenital tract after prostate cancer [1]. RCCs are most commonly detected incidentally in patients undergoing computed tomography (CT) for other reasons; however, other renal lesions, which are not RCC, are also commonly detected during CT, most commonly benign renal cysts. It is estimated that between 15-40% of patients undergoing CT will have an incidentally discovered renal lesion and among these, the most common are benign cysts [2], [3]. A benign cyst can be diagnosed at CT when it is homogeneous, well-circumscribed and measures water attenuation (<20 Hounsfield Units [HU]). When all of these three criteria are met, a cyst can be diagnosed with a high degree of accuracy [4]. In many instances, a cyst cannot be diagnosed at time of initial CT which may occur due to: internal complexity from protein or hemorrhage or from pseudo-enhancement (which is the artificial increase in attenuation of a cyst due to beam hardening artifact from adjacent iodine in the kidney on single-phase enhanced CT scans) [5]. Indeterminate renal lesions that may be cysts require further imaging with dedicated multi-phase CT or magnetic resonance imaging

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang.

(MRI) or biopsy for definitive diagnosis. The further workup of potentially benign incidental imaging findings is costly and increases patient morbidity.

Automated assessment of RCC is of great interest to the imaging community given the variable inter-observer agreement for diagnosis of renal masses using subjective assessment and relatively low reported accuracy for differentiating benign from malignant solid masses [6]. Automated differentiation of a benign cyst from solid mass is also clinically desirable, as discussed above, and, is an essential step for automated evaluation of solid renal masses for eventual classification into benign neoplasms and RCC. Automated classification of solid renal masses into subtypes using first- and second-order texture features [7]–[10] and machine learning [11]–[15] has been previously studied and is also actively under investigation. However, to best of our knowledge differentiation of benign cysts from solid renal masses has not been studied to date. The purpose of this study is to design and evaluate a patch-based convolutional neural network (CNN) model to differentiate benign renal cyst from solid renal mass (Figure 1). It has been demonstrated that patch-based CNN can outperform image-based CNN for whole slide tissue image classification [16]. In an image-based approach the network is trained and tested using the whole image whereas, in a patch-based approach, small images are extracted around each pixel in the image and used for CNN training and testing. We evaluated the developed algorithm on the largest cohort of renal masses at CECT studied to date for automated diagnosis that was acquired from four CT scanners that assure us of technique robustness and generalizability of results. Moreover, we computed the inter-observer variability of the algorithm. Additionally, we compared the results of the CNN to those of thresholding based on CT number for benign cyst diagnosis.

II. METHOD

A. PATIENT, CT TECHNIQUE, AND MANUAL SEGMENTATION

This retrospective study was approved by the Ottawa Hospital Research Ethics Board. Under the institutional review board (IRB) approval 315 consecutive patients with histologically confirmed solid renal masses and CT without intervening chemo- or radiotherapy were identified. The images used in this study were taken using four CT systems including GE Medical Systems (Lightspeed 64 and Discovery 750 HD), Toshiba (Aquillion 64 and 320), Siemens (Sensation 64), and Philips (Brilliance 64). For each patient, anonymized images in digital imaging and communication in medicine (DICOM) format were exported for axial CECT from our PACS. Images were evaluated by a fellowship-trained genitourinary radiologist with 13 years of experience in renal CT working in conjunction with a research assistant experienced in the analysis and segmentation of CT and magnetic resonance images using ITK-SNAP version 3.2 (Philadelphia, PA). For each patient, the radiologist and research assistant identified each solid renal mass and correlated the location on CT images to the pathology reports. The boundaries of the solid renal masses were then identified and the solid renal mass segmented. Before completing the segmentation, the radiologist reviewed all of the images for other renal cystic or solid masses in the kidneys present on CT images. When additional solid renal masses were identified without histological confirmation, the patient was excluded. When incidental cystic renal masses were encountered that measured $\geq$10 mm with adequate reference standard [17] to confirm benign Bosniak 1 or 2 simple cyst, described above, the cysts were also segmented. Renal cysts measuring $<$10 mm were excluded, as these are considered too small to characterize in clinical practice and of no clinical significance [18].

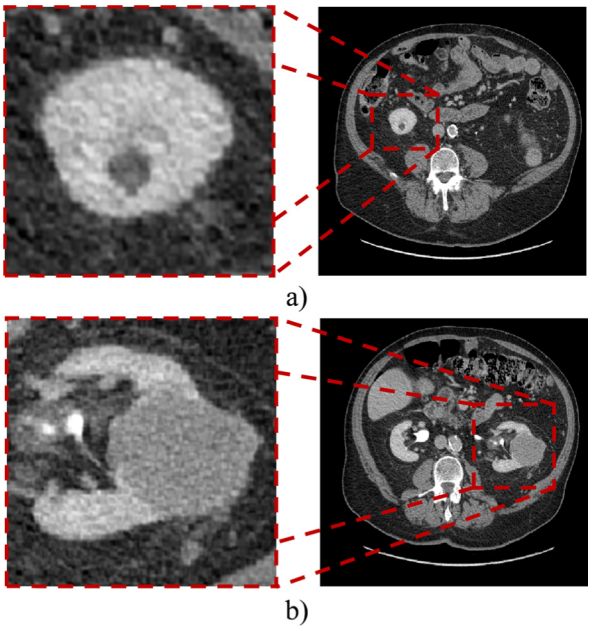

FIGURE 1. An example of a) benign renal cyst b) solid renal mass on CECT images. The 2D CT images are visualized in the standard abdominal window and level set of 400 and 50, respectively.

B. PATCH EXTRACTION AND BALANCING DATASET

For both training and testing phases, renal masses were manually delineated from CECT scans as explained in the previous section. A local image (patch) of size 40×40 pixels around voxels in the renal mass tissue was then derived in the transversal direction with the stride of 5 pixels and the density values were normalized into the range [0,1] across each patch (Figure 2). To label patches as a cyst or solid, the corresponding patch extracted from manual segmentation was used as a reference. As the number of cysts was less than the number of solid renal masses in our dataset, we had fewer patches labeled as cyst than those assigned to the solid renal mass class that caused a class imbalance in the training set. To address this problem, the elastic transformation was applied to the instances from underrepresented class to expand the training dataset. The method described by Simard *et al.* [19] was followed to perform elastic deformation.

TABLE 1. Configuration of the proposed convolutional neural network for classification of extracted patches into cyst and solid renal masses.

Layer	Type	Number of Filters	Fully Connected Unit	Layer	Type	Number of Filters	Fully Connected Unit
1	Convolutional 2D	16	-	17	Batch Normalization	-	-
2	Activation (Relu)	-	-	18	Convolutional 2D	64	-
3	Batch Normalization	-	-	19	Activation (Relu)	-	-
4	Convolutional 2D	16	-	20	Batch Normalization	-	-
5	Activation (Relu)	-	-	21	Maximum Pooling	-	-
6	Batch Normalization	-	-	22	Flatten	-	-
7	Maximum Pooling	-	-	23	Dropout	-	-
8	Convolutional 2D	32	-	24	Fully Connected	-	64
9	Activation (Relu)	-	-	25	Activation (Relu)	-	-
10	Batch Normalization	-	-	26	Dropout	-	-
11	Convolutional 2D	32	-	27	Fully Connected	-	64
12	Activation (Relu)	-	-	28	Activation (Relu)	-	-
13	Batch Normalization	-	-	29	Dropout	-	-
14	Maximum Pooling	-	-	30	Fully Connected	-	2
15	Convolutional 2D	64	-	31	Activation (Softmax)	-	-
16	Activation (Relu)	-	-				

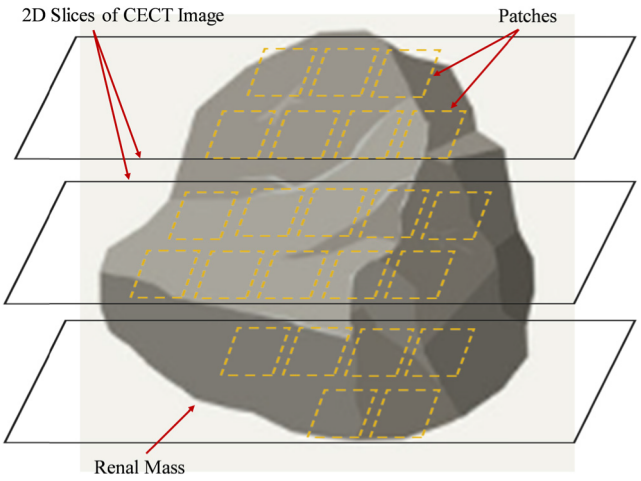

FIGURE 2. A diagram of our proposed method for patch extraction from renal mass tissue.

C. NETWORK ARCHITECTURE

In designing the CNN model, number of layers (including convolutional, maximum pooling, and Dropout), number of filters, number of layers and nodes in the fully connected layer, number of training images, patch size, optimizer, and loss functions were varied. For each set of parameters, a separate network was trained, which was then tested on a subset of unseen images to compare the results and choose the CNN model that yields the highest classification accuracy.

Table 1 displays the configuration of the CNN we used in our experiments to distinguish cyst from solid renal masses. All parameters of the network were adjusted experimentally aimed at raising the network performance. The developed CNN is made up of 6 convolutional (Conv.) layers of size 3×3 with 16, 16, 32, 32, 64, and 64 filters in each layer. Rectified Linear Unit (ReLU) and batch normalization were applied after each Conv. layer. After every two Conv. layers, one maximum pooling (Max. Pool) layer of size 2×2 was employed that calculates the maximum of 4 output nodes. A conventional neural network with two hidden layers of size 64 neurons was included to the end of our proposed network in which three Dropout layers were added before hidden and output layers. We used softmax as an activation function at the last layer of our suggested CNN to map the non-normalized output of a network to a probability distribution over predicted output classes. Categorical cross-entropy and Adadelta were used as loss function and optimizer to train the CNN, respectively.

D. NETWORK TRAINING AND TESTING

Our dataset comprised of 315 CECT images of size $512 \times 512 \times N$ (N: [31,555]). The images were randomly divided into the training ($N = 40$), and testing ($N = 275$) sets. Table 2 shows the number of CECT images used in our experiments broken by the CT systems for training and testing phases, separately. As seen, a combination of data from multiple CT scanners was used for training and testing. This strategy may be limited because it does not account for small differences between CT scanner types; however, using the same technical parameters across systems not only allowed the network to learn the signal characteristics of all scanners during training but also assured us that the results could be more generalizable and reproducible between systems and centers [7]. As we used the patch-based approach for this study, the number of patches derived from 40 images was enough for training the suggested network. Figure 3 shows an overview of our proposed method for renal masses assessment. The network was trained using patches that extracted from training CECT scans and artificially augmented that results in a total of 65287 patches (32837 and 32450 patches

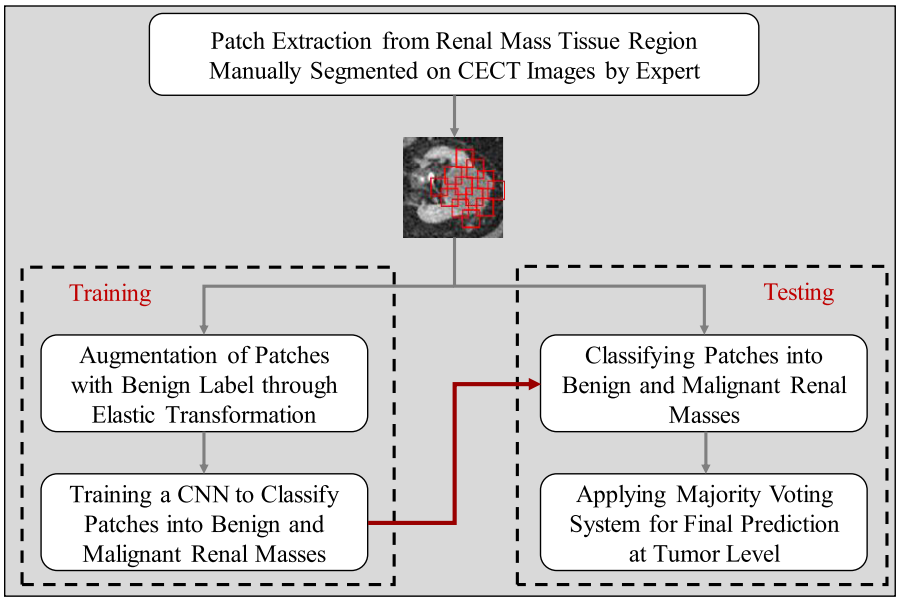

FIGURE 3. A flow diagram of our proposed methodology, where the left and right block diagrams display training and testing procedures.

TABLE 2. The number and variety of CT systems used in our experiments for training and testing phases.

CT Scan System	GE Medical Systems	Toshiba	Siemens	Philips	Total
Train	18	17	1	4	40
Test	159	93	3	20	275
Total	177	110	4	24	315

labeled as cyst and solid renal mass respectively) for training. We implemented the algorithm in Python using Keras library built on top of Tensorflow. The network was trained for 250 epochs on Intel Core i7, 2.8 GHz on a graphics processing unit (GPU)-accelerated computing platform, which took approximately 110 minutes. In the testing phase, the extracted patches around each voxel from manually segmented renal mass tissue were passed to the trained CNN to be labeled as benign cyst or solid renal mass. For each renal mass, if the number of predicted patches with the label of benign cyst was more than that of solid, the renal mass was categorized as cyst and vice versa. For cases with an equal number of patches labeled with benign cyst and solid renal masses, the patch-extraction was repeated with the different stride value (3 or 4 pixels) to assure of existing unequal number of cyst and solid labels as a prerequisite for applying majority voting rule.

E. EVALUATION METRICS

We used accuracy, precision, and recall to assess the performance of our algorithm for renal mass classification (i.e., benign cyst vs. solid renal mass).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{1}$$

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

and

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. These metrics were computed per tumor basis. Therefore, if there were multiple tumors in one kidney, each tumor was considered separately in computing the metrics. We also computed the accuracy of renal mass classification on a per-patient basis. To this end for patients with multiple tumors, the mean accuracy was computed for each patient and the average accuracy across the whole test dataset was reported. Moreover, the receiver operating characteristic (ROC) curve, which graphs the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings was plotted. The area under the ROC curve (AUC) was reported as a measure of our proposed method for renal mass assessment performance across all possible classification thresholds. We measured 95% confidence interval and p-value for AUC. To this end, the AUC computation was bootstrapped 10,000 times by sampling predicted and true labels for each lesion in the test dataset with replacement. The lower and upper bounds at the 95% level of the histogram of the bootstrapped AUC values considered as the confidence intervals. Given 0.5 as the null hypothesis value of AUC, the p-value was a matter of counting the proportion of time AUC was less than or equal to 0.5.

We used the coefficient-of-variation (CV) [20] in accuracy of the algorithm in classifying renal masses into benign cyst and solid renal masses on a per-patient basis to evaluate the variability among users.

To diagnose a simple cyst on CT, the lesion should be completely homogeneous and measure $<$ 20 HU in density [4]. A cyst measuring $\geq$ 20 HU is technically indeterminate on CT and would require further characterization typically with a follow-up CT or MRI performed with and without IV contrast to assess for the presence of enhancement. Follow-up imaging results in unnecessary health care expenditure and increases patient morbidity. Since many cysts will spuriously measure above 20 HU at single-phase CECT (due to pseudo-enhancement) higher attenuation thresholds have been proposed to improve accuracy including 30 HU [17] and more recently 40 HU [21]. A limitation of using higher attenuation thresholds to diagnose simple cysts is the potential false positive results and misclassification of solid masses which show low levels of enhancement and measure <40 HU. We calculated the number of cysts and solid masses with average HU $\geq$20 HU, $\geq$30 HU and $\geq$40 HU to determine what proportion of cysts and masses would be accurately classified using only attenuation measurements. Since attenuation measurements are performed in clinical practice by manual placement of a circular region of interest (ROI) in the center of a lesion encompassing approximately 2/3 of its' surface area and avoiding the edges to prevent inclusion of adjacent structures, we trimmed the boundary of the cysts and tumors for three iterations using the erosion operation in python to sample the center of tumor for density computation. The HU was then estimated at the tumor level. The ROC for HU was also plotted, where the diagnosis of a benign cyst was true positive when HU $<$ 20, and false negative when HU $\geq$ 20. The diagnosis of solid mass was considered as false positive for HU $<$ 20, and true negative otherwise.

III. RESULTS

Our proposed method for differentiating benign cyst from solid renal masses yielded mean accuracy, precision, and recall of 88.96% (291/327), 95.64%, and 91.64% on 275 CECT test images including 327 renal masses (39 benign cysts and 288 solid renal masses) on a per-renal mass basis. The confidence interval for accuracy, precision, and recall at a 95% level was (0.859 - 0.917), (0.935 - 0.975), and (0.888 - 0.942) respectively.Our methodology reported accuracy of 91.21% $\pm$ 25.88% as mean $\pm$ standard deviation at the patient level. The average running time required to classify a manually segmented renal masses from a typical CECT test image on Intel Core i7, 2.8 GHz using a GPU-accelerated computing platform was 14.46 $\pm$ 12.30 sec. Figure 4 shows the confusion matrix for the developed method where the number of misclassified cases are shown for each type of renal masses. The average of HU for misclassified cystic renal masses using our developed method was 51.

The ROC curve was graphed for our proposed methodology for diagnosis of benign cyst versus solid renal masses as shown in Figure 5. The AUC was estimated as 0.804. Figure 6 displays the histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. The confidence intervals at the 95% level was (0.739 - 0.866) with p-value $<$ 0.0001.

Figure 7 shows examples of cases in which our developed model failed to classify tumors correctly. In order to investigate whether there is a correlation between tumor size and misclassification, the volume of the renal masses was measured. Although the majority of small renal masses were correctly classified, the size of approximately 80% of the misclassified cases was less than one-fifth of the average

N = 327	Predicted Cyst	Predicted Solid
Cyst	27	12
Solid	24	264

FIGURE 4. Confusion matrix shows the performance of our proposed method for classification of cyst and solid renal masses on CECT images.

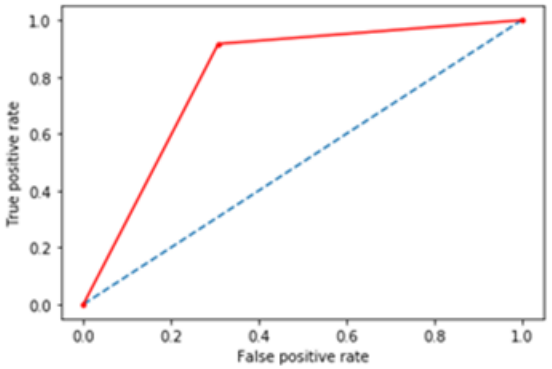

FIGURE 5. ROC curve displays the performance of our designed patch-based CNN model for diagnosis of cyst versus solid renal masses, where true positive rate (sensitivity) against false positive rate (1 - specificity) has been plotted.

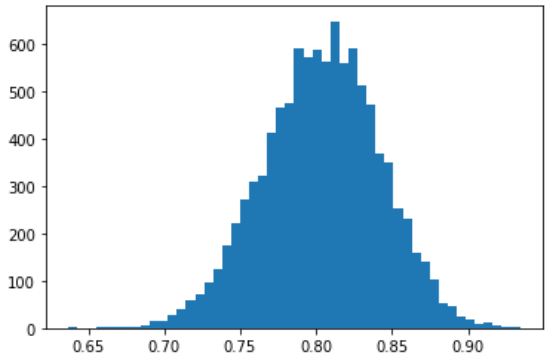

FIGURE 6. Histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation.

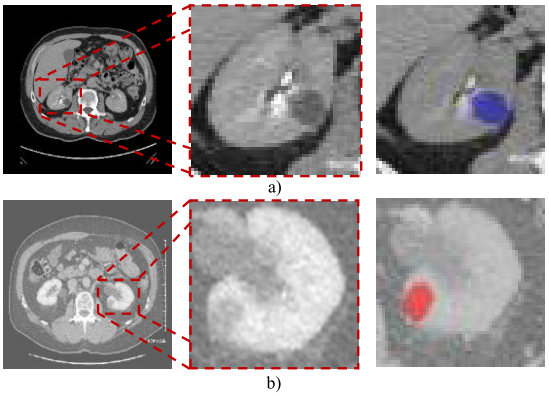

FIGURE 7. Example results of misclassified (a) benign cyst and (b) solid renal mass using our algorithm. The original CECT images are shown in the first column. The third column shows the location of tumor where blue and red show benign cyst and solid renal mass, respectively. The CECT images are visualized in the standard abdominal window and level set of 400 and 50.

TABLE 3. Accuracy of classification of benign cysts from solid masses as evaluated using independent segmentation of two observers.

Observer (O)	Accuracy (%)	Precision (%)	Recall (%)
O1	96.97	100.0	96.55
O2	93.94	96.55	96.55

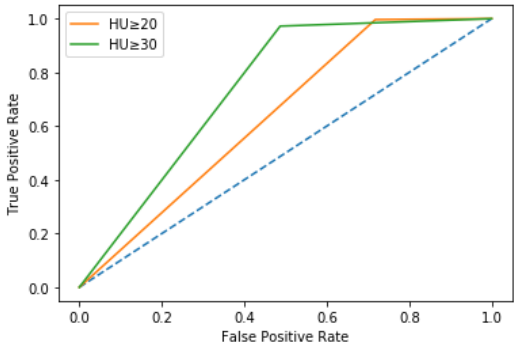

FIGURE 8. ROC curve displays the performance of conventional assessment for diagnosis of cyst versus solid renal masses, where true positive rate (sensitivity) against false positive rate (1 - specificity) has been plotted.

TABLE 4. The performance of different techniques for benign cyst diagnosis on CECT test images.

Method	Correctly Classified Benign Cysts	Incorrectly Classified Benign Cysts
CNN	27	12
HU≤20	14	25
HU≤30	22	17

tumor volume, which suggests that the proposed algorithm is relatively less accurate in identifying smaller renal lesions.

A subset of our test dataset was used to evaluate the inter-observer variability of our algorithm. We randomly selected 30 images from patients with different renal mass subtypes including oncocytomas, fpAml, clear cell RCC, papillary RCC, and chromophobe RCC (six from each subtype). Two observers participated in this study. In addition to the initial segmentations performed (described earlier) by observer 1, a second observer (observer 2) who was a non-radiologist shown only where the location of renal masses was in each patient, and independently segmented the tumors on all slices in which they appeared. The inter-observer CV was computed as 8.8%. As seen, the algorithm yielded a very small inter-observer CV. In addition, the results for the classification of benign cysts from solid masses performed equally well using segmentations from both observers (Table 3). This can be expected because in our method, small patches are extracted from the region of interest and the majority voting system is applied for final prediction. Therefore, minor discrepancies in manual segmentation at the boundary of the renal masses do not substantially affect the accuracy of classification at the patient level.

Our analysis of the average attenuation of benign cysts in the test dataset revealed that 71.79% of cysts measured $\geq$20 HU, 48.72% $\geq$30 HU and 38.46% $\geq$40 HU. Figure 8 shows the ROC curve for the diagnosis of a simple cyst from solid renal masses using attenuation with the different threshold values of 20 HU and 30 HU. AUC was 0.639 and 0.743 for HU $\geq$ 20 and 30, respectively.

The results of our study revealed that automated diagnosis of benign cyst is more challenging than that of solid tumors. We examined the performance of CNN for cyst cases and compared the results with the traditional method, where benign cysts are differentiated from solid tumors based on the HU threshold. Table 4 shows the results of different methods for cyst detection, where the CNN approach is more accurate for cyst identification than thresholding based on CT numbers. It is remarkable that we measured the CT numbers in HU in 3D while in practice the slice-wise measurement is performed. Therefore, it is likely to have a greater number of incorrectly diagnosed benign cyst using HU thresholding-based technique.

IV. DISCUSSION

In this study, we described a CNN-based method to differentiate benign renal cysts from solid renal masses. Our proposed method was evaluated on a relatively large test dataset compared to previous studies have done on automated evaluation of renal masses and achieved highly accurate results. Using the patch-based CNN model for classification, our proposed algorithm benefits from the majority voting system for the final prediction that improves the performance of the designed model for renal mass evaluation. Our results may be useful for groups actively investigating automated diagnosis

of renal masses since an initial step after kidney segmentation is to differentiate benign cysts (which are ubiquitous in the population) from any solid renal mass. Our results may also be clinically useful, since a proportion of benign cysts may appear complex and indeterminate in CT, requiring further costly work-up in many instances and our results indicate that the proposed model can be highly accurate to diagnose renal cysts from solid masses at single-phase enhanced CT exams. In training our CNN model, we had substantially fewer images with cysts. We employed elastic transformation to artificially augment data from training samples of patches with cyst label that was the key concept in designing our method. Elastic transformation not only enhanced available data for training instances but also reduced the generalization error as this operation allows the network to learn invariance to such deformation which is a common variation in CT images due to the dynamic motion of abdominal tissues. The low inter-observer CV of our algorithm suggests high reproducibility. This is particularly useful in multi-center clinical trials and when using large registries of tumors, where segmentation could be performed by multiple observers.

We evaluated the average density of benign cysts in our test dataset which demonstrated that the majority of cysts measured greater or equal than 20 HU which, when applied strictly, would require that these lesions be further characterized with additional dedicated imaging with renal mass protocol CT or MRI. The increased attenuation of benign cysts in our study is likely due to the fact that only single-phase enhanced CT images were evaluated where pseudo-enhancement spuriously increases attenuation values. Comparing the ROC curves for HU and our proposed methodology shows the potential clinical value of our method which may help to characterize a subset of benign cysts which may otherwise be considered indeterminate. In the recently published Bosniak version 2019 guideline, an upper threshold of 30 HU has been proposed to diagnose benign cysts from solid masses and to partially mitigate the effects of pseudo-enhancement [17]; however, this proposed threshold requires further validation. When applied in our cohort, 48.72% % of cysts measured >30 HU, whereas 2.78% of solid masses measured <30 HU.

This study has several limitations. One limitation of our technique is that it requires expert manual identification of renal masses on CECT images. In practice, this is not a major limitation since renal lesion detection is rarely a clinical problem and renal lesion segmentation is by far less time consuming than whole kidney segmentation. Future studies are required to develop a fully automated approach including renal mass detection and segmentation. In this research, we used patch-based CNN that inherits all limitations of patch-based approach including redundant computation due to the overlap exists between adjacent patches, being slow, and incapability of capturing both local and global features simultaneously [22]. Although the sample size in this study was quite large and images were acquired from several CT scanners, images were taken from a single institution retrospective cohort that may create the potential for population bias. Evaluation of our technique in other patient populations is required. The number of reported cysts measuring ≥ 20 HU and ≥ 30 HU at enhanced CT is higher than expected and this most likely relates to the automated HU calculation derived from the whole cyst rather than the single slice HU measurement using ROI placement in clinical practice. Though we attempted to simulate manual ROI placement in practice by trimming the volume of HU measurement in our study, it is likely that the whole cyst analysis resulted in an increased mean HU value per cyst compared to what would be expected by using single slice ROI measurements. Nevertheless, the practice of ROI placement in clinical practice is not without its' own limitations subject to observer differences in the size of ROI and placement of ROI which has been reported to vary reported HU measurements and accuracy of classification [23].

V. CONCLUSION

We described a deep learning-based model that differentiates renal lesions on CECT as benign cysts compared to solid masses. The results indicate that our algorithm is highly accurate in characterizing benign cysts from solid masses and maybe clinically valuable to prevent unnecessary imaging follow-up for characterization in a proportion of patients.

ACKNOWLEDGMENT

F. Zabihollahy acknowledges the Ontario Graduate Scholarship (OGS).

REFERENCES

[1] A. S. cancer. (Jan. 2018). *Key Statistics About Kidney Cancer*. Accessed: Sep. 6, 2019. [Online]. Available: https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html

[2] H. J. Meyer, A. Pfeil, D. Schramm, A. G. Bach, and A. Surov, "Renal incidental findings on computed tomography: Frequency and distribution in a large non selected cohort," *Medicine*, vol. 96, no. 26, 2017, Art. no. e7039.

[3] S. D. O'Connor, P. J. Pickhardt, D. H. Kim, M. R. Oliva, and S. G. Silverman, "Incidental finding of renal masses at unenhanced CT: Prevalence and analysis of features for guiding management," *Amer. J. Roentgenol.*, vol. 197, no. 1, pp. 139–145, Jul. 2011.

[4] S. D. O'Connor, S. G. Silverman, I. K. Ip, C. K. Maehara, and R. Khorasani, "Simple cyst–appearing renal masses at unenhanced CT: Can they be presumed to be benign?" *Radiology*, vol. 269, no. 3, pp. 793–800, Dec. 2013.

[5] S. Krishna, C. Murray, M. Mcinnes, R. Chatelain, M. Siddaiah, O. Al-Dandan, S. Narayanasamy, and N. Schieda, "CT imaging of solid renal masses: Pitfalls and solutions," *Clinical Radiol.*, vol. 72, no. 9, pp. 708–721, Sep. 2017.

[6] S. K. Kang, W. C. Huang, P. V. Pandharipande, and H. Chandarana, "Solid renal masses: What the numbers tell us," *Amer. J. Roentgenol.*, vol. 202, no. 6, pp. 1196–1206, Jun. 2014.

[7] T. Hodgdon, M. D. F. Mcinnes, N. Schieda, T. A. Flood, L. Lamb, and R. E. Thornhill, "Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images?" *Radiology*, vol. 276, no. 3, pp. 787–796, Sep. 2015.

[8] N. Schieda, R. E. Thornhill, M. Al-Subhi, M. D. F. Mcinnes, W. M. Shabana, C. B. Van Der Pol, and T. A. Flood, "Diagnosis of sarcomatoid renal cell carcinoma with CT: Evaluation by qualitative imaging features and texture analysis," *Amer. J. Roentgenol.*, vol. 204, no. 5, pp. 1013–1023, May 2015.

[9] N. Schieda, R. S. Lim, S. Krishna, M. D. F. Mcinnes, T. A. Flood, and R. E. Thornhill, "Diagnostic accuracy of unenhanced CT analysis to differentiate low-grade from high-grade chromophobe renal cell carcinoma," *Amer. J. Roentgenol.*, vol. 210, no. 5, pp. 1079–1087, May 2018.
[10] K. Sasaguri, N. Takahashi, D. Gomez-Cardona, S. Leng, G. D. Schmit, R. E. Carter, B. C. Leibovich, and A. Kawashima, "Small (< 4 cm) renal mass: Differentiation of oncocytoma from renal cell carcinoma on biphasic contrast-enhanced CT," *Amer. J. Roentgenol.*, vol. 205, no. 5, pp. 999–1007, 2015.
[11] L. Zhou, Z. Zhang, Y.-C. Chen, Z.-Y. Zhao, X.-D. Yin, and H.-B. Jiang, "A deep learning-based radiomics model for differentiating benign and malignant renal tumors," *Transl. Oncol.*, vol. 12, no. 2, pp. 292–300, Feb. 2019.
[12] S. P. Raman, Y. Chen, J. L. Schroeder, P. Huang, and E. K. Fishman, "CT texture analysis of renal masses: Pilot study using random forest classification for prediction of pathology," *Academic Radiol.*, vol. 21, pp. 1587–1596, Dec. 2014.
[13] H. Lee, H. Hong, J. Kim, and D. C. Jung, "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.*, vol. 45, no. 4, pp. 1550–1561, Apr. 2018.
[14] H. S. Lee, H. Hong, D. C. Jung, S. Park, and J. Kim, "Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification," *Med. Phys.*, vol. 44, no. 7, pp. 3604–3614, Jul. 2017.
[15] H. Coy, K. Hsieh, W. Wu, M. B. Nagarajan, J. R. Young, M. L. Douek, M. S. Brown, F. Scalzo, and S. S. Raman, "Deep learning and radiomics: The utility of Google TensorFlow Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT," *Abdominal Radiol.*, vol. 44, no. 6, pp. 2009–2020, Jun. 2019.
[16] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016.
[17] S. G. Silverman, I. Pedrosa, J. H. Ellis, N. M. Hindman, N. Schieda, A. D. Smith, E. M. Remer, A. B. Shinagare, N. E. Curci, S. S. Raman, S. A. Wells, S. D. Kaffenberger, Z. J. Wang, H. Chandarana, and M. S. Davenport, "Bosniak classification of cystic renal masses, version 2019: An update proposal and needs assessment," *Radiology*, vol. 292, no. 2, pp. 475–488, Aug. 2019.
[18] B. R. Herts, S. G. Silverman, N. M. Hindman, R. G. Uzzo, R. P. Hartman, G. M. Israel, D. A. Baumgarten, L. L. Berland, and P. V. Pandharipande, "Management of the incidental renal mass on CT: A white paper of the ACR incidental findings committee," *J. Amer. College Radiol.*, vol. 15, no. 2, pp. 264–273, Feb. 2018.
[19] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, 2003.
[20] M. Bland. (Oct. 16, 2006). *How Should I Calculate a Within-Subject Coefficient of Variation?* Accessed: Dec. 12, 2019. [Online]. Available: https://www-users.york.ac.uk/~mb55/meas/cv.htm
[21] E. M. Hu, J. H. Ellis, S. G. Silverman, R. H. Cohan, E. M. Caoili, and M. S. Davenport, "Expanding the definition of a benign renal cyst on contrast-enhanced CT: Can incidental homogeneous renal masses measuring 21–39 HU be safely ignored?" *Academic Radiol.*, vol. 25, pp. 209–212, Dec. 2018.
[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
[23] A. T. Taner, N. Schieda, and E. S. Siegelman, "Pitfalls in adrenal imaging," *Seminars Roentgenol.*, vol. 50, no. 4, pp. 260–272, Oct. 2015.

FATEMEH ZABIHOLLAHY received the B.A.Sc. and M.A.Sc. degrees in biomedical engineering from Shahid Beheshti University, Iran, in 2001, and Carleton University, Canada, in 2016, respectively. She is currently pursuing the Ph.D. degree with Carleton University. She worked in a medical devices industry as an R&D Engineer for more than ten years. Her research interest includes application of deep learning techniques for medical image analysis.

N. SCHIEDA is currently an Abdominal Radiologist with The Ottawa Hospital and an Associate Professor in radiology with The University of Ottawa. He is also the Director of Abdominal and Pelvic MRI and Prostate Imaging, The Ottawa Hospital, and is actively involved in clinical research with his research focusing mainly on the CT and MR imaging of genitourinary malignancies, especially kidney and prostate cancer and studying various body MRI applications. He regularly presents his work at international conferences and has authored over 100 peer-reviewed publications on these topics. He is an active member of the Radiological Society of North America, European Society of Radiology, and Society of Abdominal Radiology. He is also a panel member of the American College of Radiology Appropriateness Criteria for urological imaging and a member of the Society of Abdominal Radiology disease focused panel on renal cell carcinoma. He was recently appointed as the Chair of the American College of Radiology genitourinary continuing professional improvement program. He is an Associate Scientific Editor in MRI for *European Radiology*.

E. UKWATTA (Senior Member, IEEE) received the master's and Ph.D. degrees from Western University, Canada, in 2009 and 2013, respectively. He was a multicenter Postdoctoral Fellow with Johns Hopkins University and University of Toronto. He is currently an Assistant Professor with the School of Engineering, University of Guelph, Canada, and an Adjunct Professor in systems and computer engineering with Carleton University, Canada. He has been an author/coauthor of more than 75 journal articles and conference proceedings. His research interests include medical image segmentation and registration, deep learning for computer-aided diagnosis, and computational modeling. He is also a Professor Engineer in Canada.