

UC Irvine

UC Irvine Previously Published Works

Title

Artificial Intelligence Assessment of Renal Scarring (AIRS Study)

Permalink

<https://escholarship.org/uc/item/6q6723sp>

Journal

Kidney360, 3(1)

ISSN

2641-7650

Authors

Chantaduly, Chanon
Troutt, Hayden R
Reyes, Karla A Perez
et al.

Publication Date

2022

DOI

10.34067/kid.0003662021

Peer reviewed

Artificial Intelligence Assessment of Renal Scarring (AIRS Study)

Chanon Chantaduly,¹ Hayden R. Troutt,² Karla A. Perez Reyes,² Jonathan E. Zuckerman,³ Peter D. Chang,¹ and Wei Ling Lau ²

Key Points

- In this pilot study, two AI algorithms showed approximately 85% accuracy in predicting kidney fibrosis severity (using kidney biopsies as ground-truth).
- Machine learning algorithms are a promising noninvasive diagnostic tool to quantify kidney fibrosis from CT scans.

Abstract

Background The goal of the Artificial Intelligence in Renal Scarring (AIRS) study is to develop machine learning tools for noninvasive quantification of kidney fibrosis from imaging scans.

Methods We conducted a retrospective analysis of patients who had one or more abdominal computed tomography (CT) scans within 6 months of a kidney biopsy. The final cohort encompassed 152 CT scans from 92 patients, which included images of 300 native kidneys and 76 transplant kidneys. Two different convolutional neural networks (slice-level and voxel-level classifiers) were tested to differentiate severe versus mild/moderate kidney fibrosis ($\geq 50\%$ versus $< 50\%$). Interstitial fibrosis and tubular atrophy scores from kidney biopsy reports were used as ground-truth.

Results The two machine learning models demonstrated similar positive predictive value (0.886 versus 0.935) and accuracy (0.831 versus 0.879).

Conclusions In summary, machine learning algorithms are a promising noninvasive diagnostic tool to quantify kidney fibrosis from CT scans. The clinical utility of these prediction tools, in terms of avoiding renal biopsy and associated bleeding risks in patients with severe fibrosis, remains to be validated in prospective clinical trials.

KIDNEY360 3: 83–90, 2022. doi: <https://doi.org/10.34067/KID.0003662021>

Introduction

Ultrasound-guided percutaneous kidney biopsy remains the standard of care when histologic diagnosis is needed to guide management of proteinuria, microscopic hematuria, transplant rejection, or unexplained kidney dysfunction (1,2). The degree of kidney fibrosis or CKD severity is often unknown at the time of kidney biopsy (3). A kidney biopsy that reveals severe ($> 50\%$) fibrosis may clarify disease diagnosis but is unlikely to change clinical management and places patients at risk of procedure-related bleeding (4). Post-biopsy complications range from transient hematuria to life-threatening hemorrhage, and correlate with risk factors that include elevated blood pressure, advanced age, anemia, low platelet count, reduced kidney function, and hemostasis abnormalities (5). A recent meta-analysis of 118,064 ultrasound-guided kidney biopsies reported incident rates for blood transfusions,

angiographic intervention, and death at 2%, 0.3% and 0.06%, respectively (6).

Assessing CKD severity is critical when considering the risks and benefits of a kidney biopsy, but non-invasive tools to evaluate degree of fibrosis are underdeveloped (7). Studies that explored ultrasound techniques (including shear wave velocity imaging, transient elastography, real-time elastography, Doppler sonography, and ultrasound corticomedullary strain) have noted inconsistent correlation with degree of fibrosis on kidney histology (3,8,9). Furthermore, these methods are strongly dependent on external factors, such as blood pressure, kidney weight, body weight, and the applied transducer force, not to mention high intra- and interobserver variability (3). A recent report found close correlation between ultrasound-determined kidney size with estimated kidney function (eGFR) but did not assess kidney

¹Department of Radiological Sciences and Center for Artificial Intelligence in Diagnostic Medicine, University of California Irvine, Orange, California

²Division of Nephrology, Department of Medicine, University of California Irvine, Orange, California

³Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California

Correspondence: Dr. Wei Ling Lau, Division of Nephrology, Department of Medicine, University of California, Irvine, 101 The City Drive South, City Tower, Suite 400, Orange, CA 92868. Email: wllau@uci.edu

fibrosis (10). Kirpalani *et al.* reported that magnetic resonance imaging elastography was promising in terms of correlating stiffness scores with kidney fibrosis (11), but did not explore machine learning tools as a means to optimize predictive accuracy.

Computed tomography (CT) allows high-resolution imaging of renal tissue with minimal radiation exposure to the patient (7). Machine learning technology in combination with CT imaging is a promising avenue for non-invasive assessment of kidney fibrosis, in lieu of the current histologic standard of care. Deep learning convolutional neural networks (CNN) are a novel form of machine learning, with the capacity to isolate relevant patterns from histology, imaging, or other clinical data useful for disease characterization (12,13). CNN technology for diagnostics has been explored in a variety of settings, ranging from intracranial hemorrhage and cancer to pneumonia and appendicitis (14–17).

The goal of the Artificial Intelligence in Renal Scarring (AIRS) study is to develop CNN tools to quantify kidney fibrosis as a computer-aided diagnostic alternative to renal biopsy. Two CNN algorithms (slice-level and voxel-level classifiers) were trained to analyze CT imaging of native and transplant kidneys to classify degree of fibrosis, on the basis of ground-truth fibrosis scores from kidney biopsies in the same patients. In this pilot analysis, we focused on delineating mild/moderate versus severe (<50% versus $\geq 50\%$) fibrosis as a clinically relevant dichotomy that may be useful for a nephrologist considering kidney biopsy for their patient. We demonstrate that the CNN algorithms are able to differentiate severe from mild/moderate kidney fibrosis with a high degree of accuracy.

Materials and Methods

Patient Database

This was a retrospective analysis of kidney biopsies carried out at the University of California Irvine Medical Center between 2014 and 2019. We identified patients with ≥ 1 abdominal CT scans completed within 6 months of the kidney biopsy. After the initial screen of the medical records system, 42 patients who underwent imaging were excluded for the following reasons: CT scan conducted outside the inclusion period (>6 months from biopsy date, $n=3$); abdominal magnetic resonance imaging but no CT found in the system ($n=6$); patients inadvertently skipped in the annotation process ($n=2$); native kidneys not segmented in a patient who had undergone a transplant ($n=1$); low-quality scan with hardware artifact ($n=1$); and outside imaging studies that were not stored in the long-term imaging archives ($n=29$). The remaining 152 CT scans from 92 patients were downloaded and the axial soft-tissue reconstructed volume was extracted for further analysis. The majority of the CT scans (79%) were conducted without intravenous contrast; 129 scans were conducted on a Philips scanner, 22 on a Siemens scanner, and one on a GE Medical Systems scanner.

A pathologist (J.E.Z.) reviewed a random selection of patients to confirm the degree of fibrosis was accurately documented in kidney biopsy reports. In discordant patients, a second pathologist reviewed the slides to determine ground-truth (see *Acknowledgments*). Ground-truth

degree of kidney fibrosis from biopsy reports was treated as a binary outcome: mild/moderate versus severe (interstitial fibrosis and tubular atrophy <50% versus $\geq 50\%$). Native kidneys in CT scans from patients who were transplanted were automatically scored as severe fibrosis. Demographics and comorbid conditions were compiled in a REDCap database, and CT images were accessed as described below. All research procedures were approved by the University of California Irvine Institutional Review Board.

Annotation

The CT scans were transferred from our hospital's Picture Archiving and Communication System to a secure in-house database. A custom proprietary web-based annotation tool was used to create ground-truth three-dimensional binary masks corresponding to the native right and left kidneys, and any renal transplants, if present. The annotation tool was implemented as a simple brush utility without any thresholding or advanced contouring functionality. Two student researchers (H.R.T., K.A.P.R.) served as the main annotators, and a senior radiologist (P.D.C.) subsequently reviewed each patient and refined region of interest annotations where needed.

Image Preprocessing

Using the annotated kidney volume masks as a template, each individual kidney was cropped and resampled to an isotropic $96 \times 96 \times 96$ voxel volume. This resampling operation was required to ensure all model inputs were of the same matrix size and to accommodate the limitations of graphics processing unit memory. Given the volumes were first cropped to the right and left kidneys, and that the kidneys constitute only a small portion of the original CT volume, the final resampled voxel sizes were similar in resolution to the original data. Each cropped volume was then normalized by clipping all voxel values to a range of -150 to 250 Hounsfield Units and scaled by a factor of $1:50$. Any single exam in this dataset may contain up to three individual cropped kidneys used for algorithm training: (native) left, (native) right, and transplant. For any individual kidney, a total of 96 two-dimensional (2D) images (of size 96×96) were used for training.

CNN Approach

Two different custom 2D CNN networks were tested and compared, to differentiate severe from mild/moderate kidney fibrosis. The first algorithm is a standard global slice-by-slice CNN classifier, designed to predict one of three mutually exclusive categories for each 2D image: no-kidney, mild/moderate fibrosis, and severe fibrosis. The second algorithm is a pixel-level CNN classifier implemented through a fully convolutional U-Net architecture to perform simultaneous segmentation and classification. In both models, the final classification ignores slices or voxels without kidneys, and a majority rule is aggregated on the remaining predictions. Both algorithms are implemented using the state-of-the-art squeeze-and-excitation network architecture, the top-performing model of the ImageNet Large Scale Visual Recognition Challenge in 2017 (18). The squeeze-and-excitation network approach enables

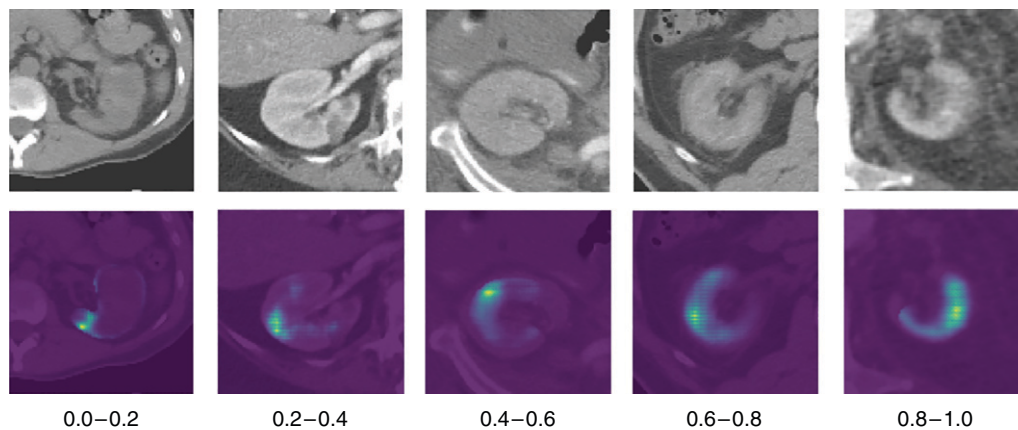


Figure 1. | Prediction heatmaps generated by the deep learning algorithm identifying areas of suspected “severe” fibrosis within the kidney. From left to right, kidneys ranging from normal to severe fibrosis are shown with progressive degrees of estimated fibrotic parenchyma. Final mean softmax normalized predictions for the entire kidney are shown in the bottom row, ranging from 0.0 to 0.2; 0.2–0.4; 0.4–0.6; 0.6–0.8; and 0.8–1.0.

networks to adaptively recalibrate channel-wise feature responses on the basis of different model inputs by learning interdependencies between channels (Figure 1).

Global Slice-by-Slice Classifier

The global CNN classifier is a custom VGG-derived architecture implemented with squeeze excitation modules at each layer. The model input is a single 2D (96×96) slice and the model output is a three-element logit vector representing a three-class prediction. The CNN consists of four convolutional blocks, where each block is defined as a 3×3 convolution, batch normalization, and ReLU repeated three times in total. Subsampling is performed at the end of each block through a convolution with a stride of two. After four convolutional blocks (12 layers), the final feature map is flattened and used as an input into a single fully connected layer.

Voxel-Level Classifier

The voxel-level CNN classifier is a custom U-Net-derived architecture implemented with squeeze excitation modules at each layer (19). The model input is a single 2D (96×96) slice and the model output is a single 2D (96×96) segmentation mask with a three-class prediction at each voxel location. The CNN consists of four contracting convolutional blocks, where each block is defined as a 3×3 convolution, batch normalization, and Leaky ReLU repeated three times total. Subsampling is performed at the end of each block through a convolution with a stride of two. After four convolutional blocks (12 layers), the operations are reversed through an identical network architecture, with the replacement of strided convolutions (subsampling) with convolutional transposes (upsampling).

Neural Network Implementation

A softmax crossentropy loss function is used to optimize both models. Optimization was performed using the Adam technique (20) with exponential decay rates, β_1 and β_2 , set to 0.9 and 0.999, respectively. The learning rate is set to 1×10^{-3} . The batch size is set to eight, with a total of 25,000

training iterations. Xavier normalization is used to initialize the weights before training (21).

Algorithm code was written in Python 3.6, TensorFlow 2.1.0 library, and Keras 1.0.8 library. The network is trained on our in-house graphics processing unit cluster, which contains 48 Nvidia GeForce RTX 2080 Ti and 12 GeForce RTX Titans. On average, the global classifier trained in 30 minutes per experiment, whereas the voxel-based classifier trained in 2–4 hours per experiment.

Evaluation

A five-fold cross-validation strategy was used to evaluate the training process. Although up to three different kidneys from each patient were treated individually during the training process, all volumes from a single patient are stratified into the same crossvalidation group to prevent data leakage. The dataset was split 80:20 and then trained on the 80%, whereas the other 20% was used for validation. The training was repeated five times total with different 80:20 splits until the entire dataset was fully validated.

In addition to experiments on the entire data cohort, additional subcohort analyses were performed to evaluate for potential confounding variables. Stratification was applied on the basis of intravenous contrast status and exclusion of atrophic native kidneys in patients who received a transplant. Furthermore, to evaluate model generalizability, models trained exclusively from Philips and GE scanners were validated exclusively with data from Siemens scanners.

Statistics

Our study goal was to distinguish between mild/moderate fibrosis from severe fibrosis on kidney CT imaging. Majority rule was used in both approaches (global slice-by-slice and voxel-based) to determine the severity of fibrosis. To evaluate the performance between the two approaches, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated and compared. Area under the receiver operating curve (AUC) was also calculated by varying the softmax

score threshold for the fibrosis classification and the resulting curve drawn for both approaches to compare. Descriptive statistics are reported as mean±SD.

Results

Both models were trained and validated on a total of 152 CT scans from 92 patients, which contained 148 left kidneys, 152 right kidneys, and 73 transplant kidneys. Patient characteristics are summarized in Table 1; 55 patients had minimal or mild (<25%) fibrosis, 14 patients were classified as moderate (25%–50%) fibrosis, and 23 patients had severe (>50%) fibrosis. Average eGFR was 37±26 versus 25±18 versus 16±13 ml/min per 1.73 m², respectively, in the mild versus moderate versus severe fibrosis groups. Patients with mild/moderate fibrosis were grouped for analysis against patients with severe fibrosis on kidney CT imaging; average kidney volume was 217 cm³ (range 48–668) versus 76 cm³ (range 3–347) in the mild/moderate versus severe (<50% versus ≥50%) fibrosis groups. There was one patient biopsy that yielded discordant fibrosis scores (comparing medical chart report with current pathologist review), and ground-truth was determined *via* blinded review by a second pathologist (see *Acknowledgments*).

On crossfold validation, the slice-by-slice model yielded an overall AUC of 0.917, whereas the voxel-based approach yielded an overall AUC of 0.922 (Figure 2). The model accuracy, sensitivity, specificity, PPV, and NPV at various thresholds evaluated across both of the CNN algorithms are summarized in Table 2. At a prediction threshold of 0.5, the slice-by-slice approach gave us an accuracy of 0.831, sensitivity of 0.817, specificity of 0.852, PPV of 0.886, and NPV of 0.767. At the same prediction threshold of 0.5, the voxel-based approach yielded similar results with an accuracy of 0.879, sensitivity of 0.853, specificity of 0.916, PPV of 0.935, and NPV of 0.816. The voxel-based approach when compared with ground-truth pathology reports yield 13 false positives (mild/moderate fibrosis classified as severe), 32 false negatives (severe fibrosis not detected; 15% miss rate), 187 true positives, and 141 true negatives.

Results from stratified subcohort analyses are shown in Table 3. Overall performance across both models was similar after exclusion of postcontrast exams and/or the exclusion of atrophic native kidneys in patients who have undergone a transplant. A minor decrease in performance (0.86 accuracy) was noted specifically for the voxel-based CNN strategy after removing postcontrast exams; all other permutations demonstrated an accuracy of ≥0.90. Additionally, overall model generalizability was preserved after stratification of training set (Philips, GE) and validation set (Siemens) exclusively by manufacturer.

A comparison of our prediction tools and other CNN studies that evaluated kidney disease parameters is summarized in Table 4.

Discussion

We initiated the AIRS study to develop machine learning tools to quantify kidney fibrosis as a noninvasive diagnostic alternative to renal biopsy. In a cohort of 92 patients, we analyzed 300 native kidney images and 73 transplant

Table 1. Patient characteristics in the Artificial Intelligence in Renal Scarring study

Variable	Patients, n (%)
Sex	
Male	45 (49)
Female	47 (51)
Ethnicity	
Non-Hispanic White	16 (17)
Hispanic White	31 (33)
Black	4 (4)
Asian	24 (26)
Other/mixed race	17 (19)
Native versus transplant kidney biopsy	
Native	45 (49)
Transplant	47 (51)
Kidney function, eGFR ml/min per 1.73 m²	
Degree of fibrosis on biopsy	
Mild (n=55)	
eGFR <15	14 (25)
15–29	12 (22)
30–44	9 (16)
45–60	6 (11)
>60	14 (25)
Moderate (n=14)	
eGFR <15	4 (29)
15–29	7 (50)
30–44	1 (7)
45–60	1 (7)
>60	1 (7)
Severe (n=23)	
eGFR <15	15 (65)
15–29	6 (26)
30–44	0
45–60	2 (9)
>60	0
Etiology of kidney disease diagnosed on biopsy^a	
Interstitial fibrosis and tubular atrophy	
Acute tubular necrosis	
Lupus nephritis	
Membranous nephropathy	
Glomerulosclerosis	
IgA nephropathy	
Diabetic nephropathy	
Focal segmental glomerulosclerosis	
Transplant rejection	

Mean age (±SD) was 50±20 years.

^aSome patients had >1 kidney disease etiology.

kidneys from 152 abdominal CT scans. We found the global slice-by-slice and voxel-based CNN models were similar in differentiating severe from mild/moderate kidney fibrosis, with AUC 0.917 versus 0.922, and PPV 0.886 versus 0.935, when compared with the ground-truth from kidney biopsy reports. Both CNN models performed consistently when tested against different CT scanners and when atrophic native kidneys in patients who have undergone a transplant were excluded; the global slice-by-slice model had slightly better prediction accuracy when only noncontrast CT scans were tested.

Diagnostic kidney biopsy remains the standard of care when histologic diagnosis is needed to guide management

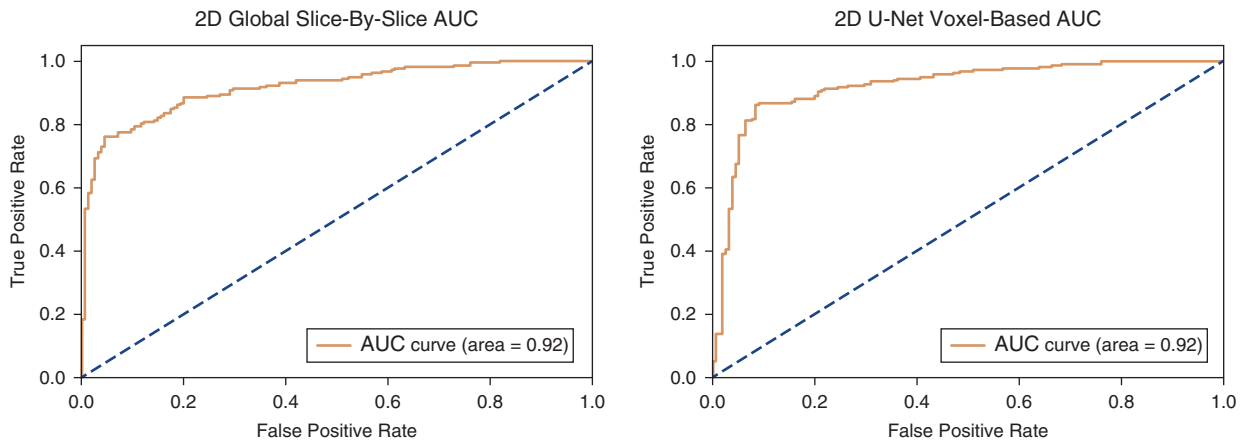


Figure 2. | Area under the receiver operating curve (AUC) was calculated by varying the softmax score threshold for kidney fibrosis classification. On crossfold validation, the slice-by-slice model yielded an overall AUC of 0.917, whereas the voxel-based approach yielded an overall AUC of 0.922.

Table 2. Summary of inference thresholds for predicting severe patients of kidney fibrosis in the two models

Two-Dimensional Global Slice-by-Slice Threshold						Two-Dimensional U-Net Voxel-Based Threshold					
Threshold	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Threshold	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
>0.1	0.753	0.940	0.490	0.722	0.854	>0.1	0.818	0.936	0.652	0.791	0.878
>0.2	0.810	0.913	0.665	0.793	0.844	>0.2	0.847	0.917	0.748	0.837	0.866
>0.3	0.826	0.894	0.729	0.823	0.831	>0.3	0.847	0.881	0.800	0.861	0.827
>0.4	0.839	0.867	0.800	0.859	0.810	>0.4	0.871	0.867	0.877	0.909	0.824
>0.5	0.831	0.817	0.852	0.886	0.767	>0.5	0.879	0.853	0.916	0.935	0.816
>0.6	0.831	0.794	0.884	0.906	0.753	>0.6	0.869	0.835	0.916	0.933	0.798
>0.7	0.831	0.775	0.910	0.923	0.742	>0.7	0.861	0.807	0.935	0.946	0.775
>0.8	0.839	0.757	0.955	0.959	0.736	>0.8	0.839	0.766	0.942	0.949	0.741
>0.9	0.802	0.679	0.974	0.974	0.683	>0.9	0.815	0.720	0.948	0.952	0.707

Table 3. Sub-cohort analyses to test performance of the two convolutional neural networks models by type of computed tomography scanner, noncontrast versus contrast scans, and exclusion of native kidneys from patients who have undergone a transplant

Model	Training set	Validation set	Correctly predicted normal or severe fibrosis
2D global slice-by-slice model	Training set: Philips and GE scanner	Validation set: Siemens scanner	Correctly predicted normal or severe fibrosis
	All data (n=320)	All data (n=53)	49/53 (92%)
	Exclude contrast CT scans (n=286)	Exclude contrast CT scans (n=43)	40/43 (93%)
Exclude contrast CT scans, exclude native kidneys in transplant patients (n=159)	Exclude contrast CT scans, exclude native kidneys in transplant patients (n=29)	26/29 (90%)	
2D U-Net voxel-based model	Training set: Philips and GE scanner	Validation set: Siemens scanner	Correctly predicted normal or severe fibrosis
	All data (n=320)	All data (n=53)	49/53 (92%)
	Exclude contrast CT scans (n=286)	Exclude contrast CT scans (n=43)	37/43 (86%)
Exclude contrast CT scans, exclude native kidneys in transplant patients (n=159)	Exclude contrast CT scans, exclude native kidneys in transplant patients (n=29)	26/29 (90%)	

2D, two-dimensional; CT, computed tomography.

Table 4. Summary of studies that utilized convolutional neural networks to quantify parameters of kidney disease

Study	Parameter of Interest	Modality	Number of Subjects	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Area Under the Curve
2D global slice-by-slice model in current report	Kidney fibrosis	CT	92	0.817	0.852	0.886	0.767	0.831	0.917
2D U-Net voxel model in current report	Kidney fibrosis	CT	92	0.853	0.916	0.935	0.816	0.879	0.922
Abdeltawab (2019) (7)	Early transplant kidney dysfunction	MRI	56	0.933	0.923	NR	NR	0.929	0.93
Kuo (2019) (25)	eGFR	US	1297	0.607	0.921	NR	NR	0.856	0.904
Sabanayagam (2020) (26)	eGFR	RP	6485	0.83	0.83	0.54	0.96	NR	0.911
Chen (2020) (27)	Kidney tumors	CT	100	0.77	0.93	NR	NR	0.9714	NR

2D, two-dimensional; CT, computed tomography; MRI, magnetic resonance imaging; NR, not reported; US, ultrasound; RP, two-field retinal photography.

of proteinuria, transplant rejection or unexplained kidney dysfunction (1,2). However, the decision to pursue a native kidney biopsy may be complex, especially if the patient is on anticoagulation therapy or has anemia or thrombocytopenia, which may increase the risk for major complications, including blood transfusions, angiographic intervention, and death (overall incident rates 2%, 0.3% and 0.06%, respectively) (6). Noninvasive tools to evaluate the degree of kidney fibrosis are lacking (7), and to date have not utilized machine learning tools to minimize subjectivity and bias. These tools can prove useful when assessed in the context of other markers of chronic kidney impairment (e.g., elevated parathyroid hormone and phosphate). In particular, prediction of severe (>50%) fibrosis may justify avoidance of kidney biopsy because the patient would not be a candidate for immune-modulating therapy (i.e., the biopsy would not change clinical management). This information is valuable, for example, when patients present with advanced kidney failure and unknown medical history. Additional passes with the biopsy needle to optimize sampling of glomeruli for histology from a fibrotic kidney may lead to an increased risk of bleeding (22). If severe degree of fibrosis is predicted on CNN analysis of a CT scan, the medical team may decide to limit biopsy sampling or avoid kidney biopsy altogether, after shared decision making with the patient. Of note, a diagnostic native kidney biopsy may still be indicated in a seemingly end-stage kidney per nephrologist discretion to guide the evaluation of transplant candidacy or to determine the etiology of a systemic disease.

We noted that a prediction threshold of 0.5 performed the best with respect to future applicability in clinical decision making, whereby the goal is to identify patients with advanced kidney fibrosis, and thus avoid invasive biopsy and its potential bleeding complications discussed above. With the voxel-based CNN model, a threshold of 0.5 would identify 85% of individuals with advanced kidney fibrosis with a PPV of 94% (Table 2). Raising the threshold to 0.9 would marginally improve PPV to 95% but would drastically decrease sensitivity from 85% to 72% (i.e., would miss 28% of individuals with severe fibrosis).

Currently, the primary limitation of the proposed algorithm is the requirement for a manual region of interest to be defined for each kidney before deep learning prediction. It should be noted, however, that although three-dimensional kidney contours were generated as part of this study, the prediction for each new patient (i.e., separate from training) required only a coarse bounding-cube around the kidney(s) to be defined. Algorithms for deep learning-based whole kidney segmentation are being developed, which would facilitate a fully automated end-to-end process (23,24). Another limitation is that heterogeneity in the degree of fibrosis across a single kidney may be a source of error in algorithm ground-truth and validation. In this study, we decided to prioritize the trade-off of including more potentially noisy data (entire kidneys were segmented) rather than utilizing a small amount of clean data; empirically, this strategy is common in big data deep learning tasks, and our observation of an algorithm prediction AUC of 0.91+ suggests the degree of noise introduced by this strategy is modest. Future refinement of these CNN prediction algorithms will be explored *via* incorporation of

clinical laboratory values relevant to CKD, such as creatinine, parathyroid hormone, and hemoglobin. Pending larger datasets, the algorithm output can be further optimized to include more expansive quantitative binning (degree of fibrosis <25%, 25%–50%, 50%–75%, and >75%). Finally, given the current algorithm is trained and validated on data from a single institution, further work is needed to evaluate generalizability beyond the studied patient cohort.

To our knowledge, the AIRS study is the first to utilize CNN models to predict the degree of kidney fibrosis from imaging scans. The results are promising and provide a basis for testing these CNN tools in prospective trials, to validate their utility in clinical decision making when it is unclear whether a patient's kidney disease is acute versus chronic.

Disclosures

J. Zuckerman reports having consultancy agreements with Leica Biosystems; reports receiving honoraria from ApotheCom; and reports being a scientific advisor or member of Pathologyoutlines.com. P. Chang reports having consultancy agreements with, and receiving research funding from, Canon Medical; and reports having an ownership interest in Avicenna.ai. W.L. Lau reports having consultancy agreements with Ardelyx and Fresenius; reports receiving research funding from the American Heart Association, Hub Therapeutics, and the National Institutes of Health; and reports receiving honoraria from Roche Australian Nephrology Preceptorship. All remaining authors have nothing to disclose.

Funding

This project was funded by a University of California, Irvine Department of Medicine Chair Research Award 2019–2021.

Acknowledgments

The authors thank Dr. Anthony Sisk (Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at University of California Los Angeles) for blinded review of discordant biopsy patients to determine ground truth.

Author Contributions

W.L. Lau conceptualized the study; C. Chantaduly, K. Perez Reyes, H. Troutt, and J. Zuckerman were responsible for the data curation; P. Chang, C. Chantaduly, and J. Zuckerman were responsible for the formal analysis; W.L. Lau was responsible for the funding acquisition; C. Chantaduly and K. Perez Reyes were responsible for the investigation; P. Chang, C. Chantaduly, and J. Zuckerman were responsible for the methodology; W.L. Lau and H. Troutt were responsible for the project administration; P. Chang was responsible for the resources and software; P. Chang and W.L. Lau provided supervision; C. Chantaduly and H. Troutt wrote the original draft; P. Chang, C. Chantaduly, W.L. Lau, K. Perez Reyes, and H. Troutt reviewed and edited the manuscript.

References

- Luciano RL, Moeckel GW: Update on the native kidney biopsy: Core curriculum 2019. *Am J Kidney Dis* 73: 404–415, 2019 <https://doi.org/10.1053/j.ajkd.2018.10.011>
- Hogan JJ, Mocanu M, Berns JS: The native kidney biopsy: Update and evidence for best practice. *Clin J Am Soc Nephrol* 11: 354–362, 2016 <https://doi.org/10.2215/CJN.05750515>

3. Berchtold L, Friedli I, Vallée J-P, Moll S, Martin P-Y, de Seigneux S: Diagnosis and assessment of renal fibrosis: The state of the art. *Swiss Med Wkly* 147: w14442, 2017
 4. Fiorentino M, Bolignano D, Tesar V, Pisano A, Van Biesen W, D'Arrigo G, Tripepi G, Gesualdo L; ERA-EDTA Immunonephrology Working Group: Renal biopsy in 2015: From epidemiology to evidence-based indications. *Am J Nephrol* 43: 1–19, 2016 <https://doi.org/10.1159/000444026>
 5. Brachemi S, Bollée G: Renal biopsy practice: What is the gold standard? *World J Nephrol* 3: 287–294, 2014 <https://doi.org/10.5527/wjn.v3.i4.287>
 6. Poggio ED, McClelland RL, Blank KN, Hansen S, Bansal S, Bomback AS, Canetta PA, Khairallah P, Kiryluk K, Lecker SH, McMahon GM, Palevsky PM, Parikh S, Rosas SE, Tuttle K, Vazquez MA, Vijayan A, Rovin BH; Kidney Precision Medicine Project: Systematic review and meta-analysis of native kidney biopsy complications. *Clin J Am Soc Nephrol* 15: 1595–1602, 2020 <https://doi.org/10.2215/CJN.04710420>
 7. Abdeltawab H, Shehata M, Shalaby A, Khalifa F, Mahmoud A, El-Ghar MA, Dwyer AC, Ghazal M, Hajjdiab H, Keynton R, El-Baz A: A novel CNN-based CAD system for early assessment of transplanted kidney dysfunction. *Sci Rep* 9: 5948, 2019 <https://doi.org/10.1038/s41598-019-42431-3>
 8. Wang Z, Yang H, Suo C, Wei J, Tan R, Gu M: Application of ultrasound elastography for chronic allograft dysfunction in kidney transplantation. *J Ultrasound Med* 36: 1759–1769, 2017 <https://doi.org/10.1002/jum.14221>
 9. Preuss S, Rother C, Renders L, Wagenpfeil S, Büttner-Herold M, Slotta-Huspenina J, Holtzmann C, Kuechle C, Heemann U, Stock KF: Sonography of the renal allograft: Correlation between doppler sonographic resistance index (RI) and histopathology. *Clin Hemorheol Microcirc* 70: 413–422, 2018 <https://doi.org/10.3233/CH-189306>
 10. Kuo C-C, Chang C-M, Liu K-T, Lin W-K, Chiang H-Y, Chung C-W, Ho M-R, Sun P-R, Yang R-L, Chen K-T: Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *Digit Med* 2: 1–9, 2019
 11. Kirpalani A, Hashim E, Leung G, Kim JK, Krizova A, Jothy S, Deeb M, Jiang NN, Glick L, Mnatzakanian G, Yuen DA: Magnetic resonance elastography to assess fibrosis in kidney allografts. *Clin J Am Soc Nephrol* 12: 1671–1679, 2017 <https://doi.org/10.2215/CJN.01830217>
 12. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, Francis JM, Salant DJ, Chitalia VC: Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep* 3: 464–475, 2018 <https://doi.org/10.1016/j.ekir.2017.11.002>
 13. Sumathipala Y, Lay N, Turkbey B, Smith C, Choyke PL, Summers RM: Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks. *J Med Imaging (Bellingham)* 5: 044507, 2018 <https://doi.org/10.1117/1.JMI.5.4.044507>
 14. Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, Chen J, Abcede H, Shafie M, Sugrue L, Filippi CG, Su M-Y, Yu W, Hess C, Chow D: Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. *AJNR Am J Neuroradiol* 39: 1609–1616, 2018 <https://doi.org/10.3174/ajnr.A5742>
 15. Rajaraman S, Candemir S, Kim I, Thoma G, Antani S: Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Appl Sci (Basel)* 8: 8, 2018 <https://doi.org/10.3390/app8101715>
 16. Yoo S, Gujrathi I, Haider MA, Khalvati F: Prostate cancer detection using deep convolutional neural networks. *Sci Rep* 9: 19518, 2019 <https://doi.org/10.1038/s41598-019-55972-4>
 17. Park JJ, Kim KA, Nam Y, Choi MH, Choi SY, Rhie J: Convolutional-neural-network-based diagnosis of appendicitis via CT scans in patients with acute abdominal pain presenting in the emergency department. *Sci Rep* 10: 9556, 2020 <https://doi.org/10.1038/s41598-020-66674-7>
 18. Hu J, Shen L, Sun G: Squeeze-and-excitation networks. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, pp 2011–2023
 19. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional networks for biomedical image segmentation. In: *MICCAI 2015*, edited by Navab N, Hornegger J, Wells WM, Frangi AF, *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp 234–241 https://doi.org/10.1007/978-3-319-24574-4_28
 20. Kingma DP, Ba J: Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs]* [Internet] 2017 Available at: <http://arxiv.org/abs/1412.6980>. Accessed March 2, 2021
 21. Glorot X, Bengio Y: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp 249–256, 2010
 22. Nguyen L, Souccar S, Zuckerman JE, Chen JLT, Katrivesis J, Abi-Jaoudeh N, Reddy U, Baraghoush A, Morrison DE, Li X, Wang B, Lau WL: Kidney biopsy: challenges with periprocedural management [published online ahead of print October 2021]. *J Nephropathol* 10.34172/jnp.2021.xx
 23. Türk F, Lüy M, Barışçı N: Kidney and renal tumor segmentation using a hybrid V-Net-based model. *Mathematics* 8: 1772, 2020 <https://doi.org/10.3390/math8101772>
 24. Sharma K, Rupprecht C, Caroli A, Aparicio MC, Remuzzi A, Baust M, Navab N: Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. *Sci Rep* 7: 2049, 2017 <https://doi.org/10.1038/s41598-017-01779-0>
 25. Kuo C-C, Chang C-M, Liu K-T, Lin W-K, Chiang H-Y, Chung C-W, Ho M-R, Sun P-R, Yang R-L, Chen K-T: Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *NPJ Digit Med* 2: 29, 2019 <https://doi.org/10.1038/s41746-019-0104-2>
 26. Sabanayagam C, Xu D, Ting DSW, Nusinovic S, Banu R, Hamzah H, Lim C, Tham Y-C, Cheung CY, Tai ES, Wang YX, Jonas JB, Cheng C-Y, Lee ML, Hsu W, Wong TY: A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health* 2: e295–e302, 2020 [https://doi.org/10.1016/S2589-7500\(20\)30063-7](https://doi.org/10.1016/S2589-7500(20)30063-7)
 27. Chen G, Ding C, Li Y, Hu X, Li X, Ren L, Ding X, Tian P, Xue W: Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform. *IEEE Access* 8: 100497–100508, 2020 <https://doi.org/10.1109/ACCESS.2020.2995310>
- Received:** June 2, 2021 **Accepted:** November 11, 2021
 C.C. and H.R.T. contributed equally to this manuscript.
 See related editorial, “Kidney Fibrosis Assessment by CT Using Machine Learning,” on pages 1–2.