

innovations and this functionality should be developed. The report of Potluri and colleagues¹ reveals the depth of patient trust in their care providers to provide them with best evidence-based advice regarding their transplant options. It is our responsibility to ensure we maximize the information obtained from their experience to better inform decision making for future patients.

DISCLOSURES

Dr. Gill and Dr. Formica have no disclosures. Dr. Levitsky is a speaker for Gilead.

REFERENCES

1. Reese P, Potluri V, Goldberg D, Mohan S, Bloom R, Sawinski D, et al.: National trends in utilization and one-year outcomes with transplantation of HCV-Viremic kidneys. *J Am Soc Nephrol*, 30: 1939–1951, 2019
2. Durand CM, Bowring MG, Brown DM, Chattergoon MA, Massacesi G, Bair N, et al.: Direct-acting antiviral prophylaxis in kidney transplantation from Hepatitis C Virus-infected donors to noninfected recipients: An open-label nonrandomized trial. *Ann Intern Med* 168: 533–540, 2018
3. Goldberg DS, Abt PL, Blumberg EA, Van Deerlin VM, Levine M, Reddy KR, et al.: Trial of transplantation of HCV-infected kidneys into uninfected recipients. *N Engl J Med* 376: 2394–2395, 2017
4. Levitsky J, Formica RN, Bloom RD, Charlton M, Curry M, Friedewald J, et al.: The American Society of Transplantation consensus Conference on the use of Hepatitis C viremic donors in solid organ transplantation. *Am J Transplant* 17: 2790–2802, 2017

See related article, “National Trends in Utilization and 1-Year Outcomes with Transplantation of HCV-Viremic Kidneys,” on pages 1939–1951.

Machine Learning Comes to Nephrology

Kevin V. Lemley

Division of Nephrology, Children’s Hospital Los Angeles, Los Angeles, California; and Department of Pediatrics, Keck School of Medicine of the University of Southern California, Los Angeles, California

JASN 30: 1780–1781, 2019.

doi: <https://doi.org/10.1681/ASN.2019070664>

Nephrology has been affected for decades by advances in technology: from Western blotting and reporter gene assays to multiphoton imaging and next-generation sequencing.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Kevin V. Lemley, Division of Nephrology, Department of Pediatrics, Children’s Hospital Los Angeles, MS#40, 4650 Sunset Boulevard, Los Angeles, CA 90027. Email: klemley@chla.usc.edu

Copyright © 2019 by the American Society of Nephrology

Initially exotic techniques eventually find their way into nephrology’s DNA.

Among the latest exotic technologies affecting nephrology is machine learning (ML). Just one manifestation of artificial intelligence,¹ ML is a collection of computationally intensive statistical learning techniques. It is the result of the challenge of big and high-dimensional data and the development of the hardware (graphical processing units, mass digital memory storage) and software needed to understand it. The rapid expansion of whole slide imaging by digital slide scanners has provided fertile ground for ML in renal pathology. Rather than trying to build in hardcoded models on the basis of prior knowledge and rules to predict outcomes, ML allows a program to learn from experience alone, improving its performance iteratively on a training set by comparing its predictions to authoritatively labeled cases (ground truth) and adjusting a very large number of weighting parameters in the model so as to minimize a loss function, which represents the distance between prediction and truth. This characterizes supervised learning. This process of parameter adjustment is iterated many dozens of times to “train” the model. Each complete passage of training data through the model (an epoch) may require many millions of calculations. Optimization (fine-tuning) of the hyperparameters (such as learning rate, depth of the model, and number of epochs), which determine the overall architecture of the model, is performed on a separate validation set, typically 20% of all the training data. Finally, the model is tested on annotated data that was held out from the training and validation sets. If this is not feasible because of the small number of cases, other options include k-fold crossvalidation and using bagged (bootstrap aggregated) data.

Uses of ML in nephrology include predicting AKI or time to allograft loss from clinical features, recognizing specific histologic features in a biopsy, choosing an optimal dialysis prescription, or mining text in the electronic health record to find specific cases. Unsupervised learning methods, less often applied in biomedicine, recognize clusters of unlabeled individuals by their proximity in some multidimensional feature space. These methods are typically used to reduce data dimensionality, cluster data or detect patterns, and find outliers.

One of the most often used ML architectures is a convolutional neural network (CNN), a type of artificial neural network that underlies “deep learning.” This powerful tool was used in two papers published in this issue of *JASN*.^{2,3} The workings of a CNN can be reasonably well explained by a familiar biologic system. In analogy to the mammalian visual system, a multilayered system of interconnecting neurons converts the primitive events of retinal photoreception (corresponding to pixels of an image) to the activation of the proverbial “grandmother” cell in the visual cortex, the final integrating neuron that fires only when the retina is exposed to grandma’s image, as a form of recognition or classification. The “hidden layers” between the visible input and output layers consist of layers of neurons wherein activation of a

downstream neuron is controlled by a small number of immediately upstream neurons. A typical CNN architecture can involve a million neurons (whose local interactions are regulated by millions of weighting parameters) sitting in dozens of alternating convolutional and pooling layers, followed by one or more “fully connected” layers, the neurons of which integrate the outputs from all of the neurons in the preceding layer to complete the classification task. The stacking of hidden layers adds to the complexity of image features that can be recognized, giving the “depth” to deep learning. The nodes connecting and integrating neuron outputs represent various types of nonlinear activation functions that give the system the ability to represent highly complex, nonlinear relationships among the input data.

One of the basic tasks of CNNs in biomedicine is segmentation of the pixels of an image into defined components, such as a glomerulus⁴ on a histologic slide, an angiomyolipomata or other mass on a computed tomography image, *etc.* In addition to recognizing histologic “primitives,” CNNs can be trained to predict abstract outcomes such as 5-year renal survival.⁵

ML has inherent strengths and weaknesses. CNNs are a classic example of a black box: they produce an output (the number of glomeruli in a tissue section) from an input (digital image of the tissue section) without explicitly indicating how they got this value. This could be called the challenge of intelligibility. Techniques exist to mine the CNN for the pivotal features used for object recognition, such as saliency maps.⁶ CNN image classification may be inordinately sensitive to changes in the image (orientation, staining) that present no challenge to a human observer. Models must be tested to assure that they are robust against such seemingly trivial image distortions. Conversely, in contrast to human evaluators, CNNs may be sensitive to “subvisual” features not even perceptible to humans,⁷ giving CNNs the prospect of doing something more than just imitating a pathologist.

Unlike a classic CNN such as AlexNet, which is trained to recognize a thousand different types of objects (dogs, cats, planes) present in an image by training on over a million annotated examples, typical annotated training sets in biomedicine may consist of only dozens^{2,3} or hundreds of images, although training sets of >100,000 annotated images have been used.⁸ The effort involved in careful annotation of nephrology image training sets represents a significant limitation. Such “sparseness” of training sets can be compensated for through a number of strategies, such as data augmentation and transfer learning. ML models are generally quite susceptible to overfitting, endangering their generalizability to new cases. A number of techniques, such as regularization, exist to mitigate this risk. Generally, limiting the number of epochs of training to that which minimizes the total error on validation sets is required. ML models on human data (to the degree that they are “unintelligible”) also carry the risk of inadvertently baking racial or socioeconomic biases inherent in the training data into their models. It is therefore important to test models on diverse populations.

The current emphasis on reproducibility in research⁹ applies to studies using ML, where it may meet some of its greatest challenges. It is likely that both the computer code of the models and raw training, validation, and test data will have to be made available to reviewers and others.

How do we assess the performance of ML approaches to prediction or recognition? Gold standards may be hard to achieve. For example, recognizing glomeruli would seem to be a task for which a pathologist’s eye should be perfect for annotation, but significant interobserver variability in such seemingly easy recognition tasks exists. Even so, any CNN classification or prediction models should be tested against clinician or pathologist performance, both for accuracy and time of execution.¹⁰ The acceptable error rate will depend on the specific task for which the ML is being used. Ultimately, clinical utility must be demonstrated and US Food and Drug Administration approval will likely be required. There are indications that the best solution may be augmented intelligence, with clinician and ML working together.

DISCLOSURES

None.

REFERENCES

- Torres R, Olson E: AI: What have you done for us lately? *J Am Soc Nephrol* 29: 2031–2032, 2018
- Ginley B, Lutnick B, Jen K-Y, Fogo A, Jain S, Rosenberg A, et al.: Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol* 30: 1953–1967, 2019
- Hermesen M, de Bel T, den Boer M, Steenberg EJ, Kers J, Florquin S, et al.: Deep learning-based histopathological assessment of renal tissue. *J Am Soc Nephrol* 30: 1968–1979, 2019
- Bukowy JD, Dayton A, Cloutier D, Manis AD, Staruschenko A, Lombard JH, et al.: Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *J Am Soc Nephrol* 29: 2081–2088, 2018
- Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, et al.: Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep* 3: 464–475, 2018
- Simonyan K, Vedaldi A, Zisserman A: Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv* 1312.6034v2, 2014
- Madabhushi A, Lee G: Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal* 33: 170–175, 2016
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542: 115–118, 2017
- Briggs JP: The quality of reporting of kidney research: A challenge to JASN. *J Am Soc Nephrol* 30: 1–2, 2019
- De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 24: 1342–1350, 2018

See related articles, “Computational Segmentation and Classification of Diabetic Glomerulosclerosis,” and “Deep Learning–Based Histopathologic Assessment of Kidney Tissue,” on pages 1953–1967 and 1968–1979, respectively.