Forging the tools for a computer-aided workflow in transplant pathology

Lancet Digit Health 2021
Published Online
November 15, 2021
https://doi.org/10.1016/S2589-7500(21)00254-5
See Online/Articles
https://doi.org/10.1016/S2589-7500(21)00211-9

To date, the application of machine learning to digital pathology has mainly focused on cancer diagnosis, with immune-mediated diseases following in tow. However, a new study by Jesper Kers, Roman Bülow, and colleagues[1] adds to a small but growing body of work attempting to enhance kidney pathology through the application of deep learning to whole slide image data.

Biopsies of kidney transplants are central to patient management. Specialised histopathologists apply the Banff classification for allograft pathology[2] to assign Banff lesion scores to biopsy features, and integrate these with ancillary studies (serology, immunohistochemistry, or electron microscopy) to reach a diagnosis or several diagnoses that inform patient management.

Previous digital pathology applications to immune-mediated kidney diseases have focused on applying traditional[3,4] and deep learning[5–9] image segmentation techniques to facilitate feature extraction. Segmented features (eg, glomeruli or tubules) can be comprehensively described by shape parameters and then compared with pathologist evaluation, or related to clinical outcome. Augmented ResNet,[8] Inception v3,[7] and U-net[9] architectures have shown promise for these tasks. Semantic segmentation studies are limited by a reliance on labelled ground truth data, a scarce commodity that requires substantial pathologist time. The study by Kers, Bülow, and colleagues[1] obviates this requirement by focusing instead on a classification task: assigning each biopsy a label on the basis of a simplified version of Banff classification categories,[2] extracted from biopsy reports.

Serial convolutional neural networks performed the two-step process of first classifying cases into normal or disease classes, then further classifying disease into rejection or other diseases. The resulting models were validated using an internal testing set (normal area under the receiver operating characteristic curve [AUROC] 0·87 [CI 0·85–0·88]; disease 0·87 [0·86–0·88]; other diseases 0·75 [0·72–0·77]; rejection 0·75 [0·73–0·76]) and an external dataset (normal AUROC 0·83 [0·80–0·85]; disease 0·83 [0·73–0·91]; other diseases 0·61 [0·50–0·74]; rejection 0·61 [0·51–0·70]). Visualisation techniques (Occlusion Sensitivity and gradient-weighted Class Activation Mapping) were applied to identify tiles of the whole slide image that were considered highly predictive of the label. This study is, to our knowledge, the first to attempt automated classification of rejection in transplant biopsies, using an impressive dataset of 5844 whole slide images from 1948 patients across three centres. The dataset includes not only haematoxylin and eosin-stained sections, but also sections with special stains that are routinely used in renal pathology (periodic acid-Schiff and Jones silver stain).

The authors propose two potential clinical applications for the model. First, visualisation techniques that could direct the pathologist's eye to highly predictive tiles. Second, the model could flag cases requiring urgent clinical action (ie, cases with rejection) for prioritised reporting. The second application is somewhat undermined by the clinical reality that rate of organ function deterioration most often dictates urgency of patient management, in which case a finding of a normal biopsy is just as important.

Visualisation techniques might hold more promise as potential computer-aided detection (CAD) tools, which have, for example, been shown in clinical trials to improve the accuracy of breast cancer detection.[10] It is encouraging that the highly predictive regions for rejection contained important features for a diagnosis of rejection, including interstitial inflammation, peritubular capillaritis, and tubulitis. However, other important features of rejection (glomerulitis, glomerular capillary wall double contours, and vasculitis) were not represented, risking forcing learned biases onto the observer and potentially hiding rare but important features. The predictive tiles that are shown in the paper rarely contained the glomeruli and arteries that the pathologist would need to identify lesions and were often sampled from the medulla.[1] These observations go counter to standard pathology, where lesion identification and pathological diagnosis are reliant on adequate cortical sampling, quantified by counting glomeruli and arteries. The authors rightly highlight the class imbalance caused by more prevalent cortical and medullary tubulointerstitial representation in the

training dataset. Uncovering the underlying decision-making causality behind potential CAD tools would be useful, as so-called explainable artificial intelligence is important for future clinical adoption. Finally, as also highlighted by the authors, detecting rejection is only one of many important tasks performed by the transplant pathologist: detection and reporting of recurrent disease, infections, and drug toxicity also have substantial implications for patient management.

The Article by Kers, Bülow, and colleagues provides a valuable contribution to the rather sparse field of digital pathology in renal medicine, with a justifiable focus on improving pathology workflow. Future models (appendix p 1) could be trained with rejection subtype labels, perhaps also incorporating clinical data (eg, creatinine) to improve workflow decisions. To be truly transformative for patient care, a machine learning solution might take a more modular approach by applying serial classification and segmentation models, each dedicated to reproducing the complex clinical tasks performed by the renal pathologist.

See Online for appendix

We declare no competing interests. We thank Dr Bernhard Kainz (Department of Computer Science, Imperial College London, London, UK) for kindly reading and commenting on the manuscript.

*Callum Arthurs, Candice Roufosse
c.arthurs@imperial.ac.uk

Centre for Inflammatory Diseases, Department of Immunology and Inflammation, Imperial College London, London, UK (CA, CR); Department of Cellular Pathology, North West London Pathology, London, UK (CA, CR)

1 Kers J, Bülow RD, Klinkhammer BM, et al. Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study. *Lancet Digit Health* 2021; published online Nov 15. https://doi.org/10.1016/S2589-7500(21)00211-9.

2 Roufosse C, Simmonds N, Clahsen-van Groningen M, et al. A 2018 reference guide to the Banff classification of renal allograft pathology. *Transplantation* 2018; **102:** 1795–814.

3 Brazdziute E, Laurinavicius A. Digital pathology evaluation of complement C4d component deposition in the kidney allograft biopsies is a useful tool to improve reproducibility of the scoring. *Diagn Pathol* 2011; **6** (suppl 1)**:** S5.

4 Farris AB, Ellis CL, Rogers TE, Lawson D, Cohen C, Rosen S. Renal medullary and cortical correlates in fibrosis, epithelial mass, microvascularity, and microanatomy using whole slide image analysis morphometry. *PLoS One* 2016; **11:** e0161019.

5 Bukowy JD, Dayton A, Cloutier D, et al. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *J Am Soc Nephrol* 2018; **29:** 2081–88.

6 Hermsen M, de Bel T, den Boer M, et al. Deep learning-based histopathologic assessment of kidney tissue. *J Am Soc Nephrol* 2019; **30:** 1968–79.

7 Kannan S, Morgan LA, Liang B, et al. Segmentation of glomeruli within trichrome images using deep learning. *Kidney Int Rep* 2019; **4:** 955–62.

8 Ginley B, Lutnick B, Jen K-Y, et al. Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol* 2019; **30:** 1953–67.

9 Bouteldja N, Klinkhammer BM, Bülow RD, et al. Deep learning-based segmentation and quantification in experimental kidney histopathology. *J Am Soc Nephrol* 2021; **32:** 52–68.

10 Hadjiiski L, Sahiner B, Chan HP. Advances in computer-aided diagnosis for breast cancer. *Curr Opin Obstet Gynecol* 2006; **18:** 64–70.