Nephrol Dial Transplant (2022) 37: 438–440
doi: 10.1093/ndt/gfaa181
Advance Access publication 30 September 2020

Artificial intelligence: is there a potential role in nephropathology?

Meyke Hermsen [1], Bart Smeets [1], Luuk Hilbrands [2] and Jeroen van der Laak [1,3]

[1]Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands, [2]Department of Nephrology, Radboud University Medical Center, Nijmegen, The Netherlands and [3]Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

Correspondence to: Jeroen van der Laak; E-mail: Jeroen.vanderlaak@radboudumc.nl

RENAL HISTOPATHOLOGY

Examination of a kidney biopsy is frequently required to diagnose the type and stage of a kidney disease or to determine the cause of kidney transplant dysfunction. Using various (immuno)histochemical stainings, biopsy slides are visually assessed by the pathologist for the recognition of distinctive patterns leading to a diagnosis. In addition, grading systems are used to express the severity of pathological changes, e.g. the extent of inflammation in (un)scarred renal parenchyma [1]. Although pathologists are very well trained in this type of pattern recognition and quantification, the resulting scores remain semi-quantitative, are not always reproducible and have limited predictive value in clinical practice. Moreover, the scoring of tissue slides in large research settings can be a tedious task. As a result, there is a need for tools that facilitate objective, quantitative scoring in renal pathology, possibly leading to the discovery of hallmarks that can (better) predict the course of the renal disease or evaluate the response to treatment. Artificial intelligence (AI) has the potential to yield such tools [2, 3].

AI IN HISTOPATHOLOGY

The field of AI, originating in the 1950s, witnessed several so-called 'AI winters' (periods of strongly reduced interest) before it reached its current wide applicability. Today's AI is mostly based on machine learning (ML), which can be described as the fully automatic discovery of patterns in large datasets by computers. Two branches of ML are distinguished: supervised and unsupervised ML [4]. Unsupervised ML is used in cases where no outcome or ground truth annotations are available, aiming to detect inherent structure in the data. It was, for instance, shown that, using electronic health record data, unsupervised ML could identify subgroups of patients with an increased risk of developing conditions like diabetes and schizophrenia [5].

More widely used is supervised ML, where a computer learns to predict the correct label of a sample by being trained on a large number of training samples with accompanying 'ground truth' labels [6]. The most important innovation in ML in the last decade is the possibility to train multilayered ('deep') neural networks ['deep learning' (DL)] [4]. A specific subtype of deep neural networks, convolutional neural networks (CNNs), is especially powerful for the analysis of (medical) images [4]. Combined with the possibilities of scanning entire histopathological slides at high resolution [yielding whole-slide images (WSIs)], the use of CNNs creates entirely new opportunities for analysing tissue sections.

Development of CNNs typically requires a large, annotated data set. In histopathology, we distinguish between pixel-level ('strong') labels (for instance, by manually delineating glomeruli in renal biopsy WSIs) and 'weak' image-level labels (e.g. by labelling an entire renal transplant biopsy WSI positive for rejection). In AI development, the set of available WSIs is typically subdivided in (non-overlapping) training, validation and test sets. The CNN is trained using the training set, while the training process is monitored using the CNN performance on the validation set. After achieving the optimal CNN, the test set is used to yield unbiased performance data (Figure 1).

CURRENT STATUS

In general, the use of DL in histopathology is still in a research stage. Studies have shown that for certain well-described tasks the DL algorithms can perform as well as trained humans. Also, for some applications, the performance of a pathologist can be improved (either in efficiency or accuracy) by DL [7]. The extent of DL research for renal pathology is still relatively limited, mostly relying on supervised techniques using strongly labelled data sets. CNNs have been developed for the detection, classification and segmentation of renal structures (e.g. glomeruli, tubuli and interstitium), tubular atrophy and glomerular lesions [2, 8]. Comparing the output of CNNs to existing grading systems showed encouraging preliminary results. Although less frequent, CNNs have been trained using weakly labelled data sets as well, for instance, for predicting renal dysfunction, proteinuria and renal survival [9].

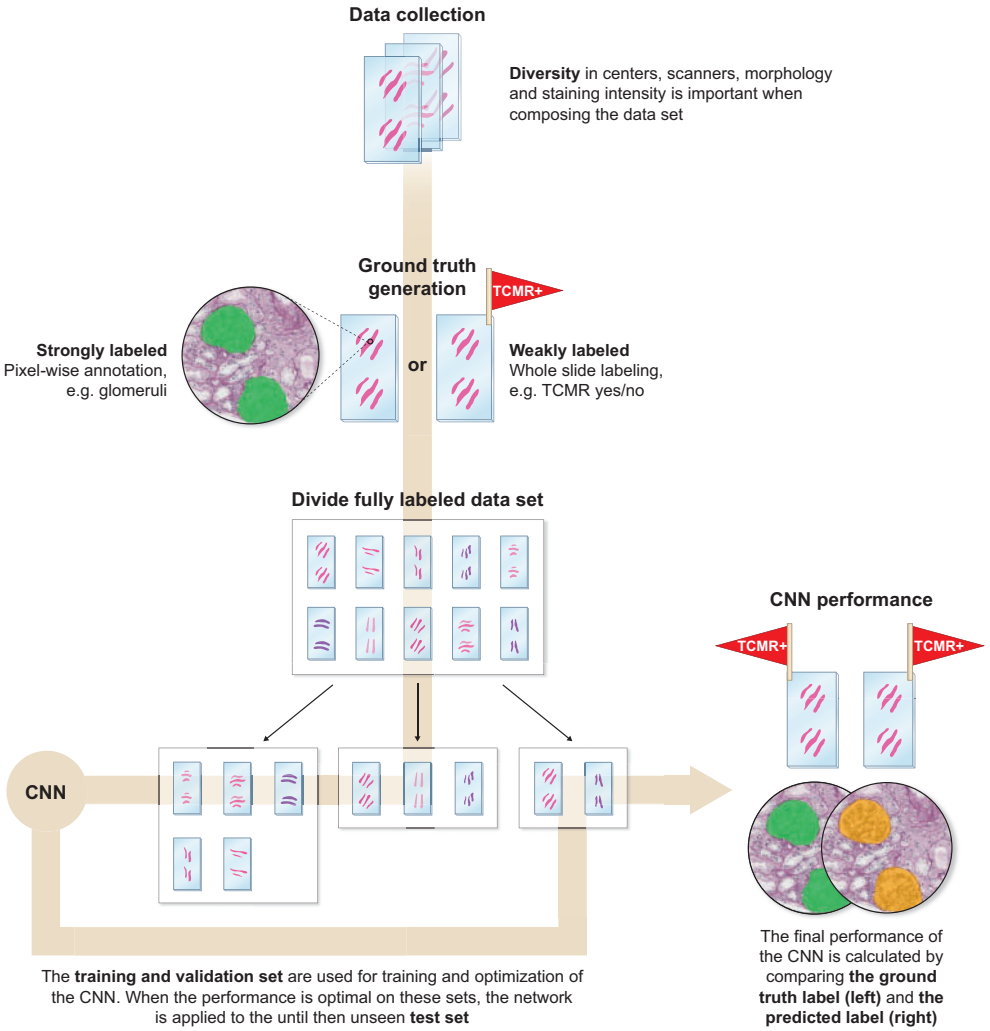

FIGURE 1: Overview of the CNN development pathway. Data collection: a large amount of data is required, preferably covering (colour and morphologic) variations that also occur in daily practice. Ground truth generation: generate labels accompanying the data set elements, either providing strong (e.g. annotating individual glomeruli and assigning the 'glomerulus' class to all pixels within these annotations) or weak labels (labelling the entire image with a single class, e.g. TCMR positive or negative). Subdivide the fully labelled data set: the set is divided in a representative training, validation and test set. CNN: the neural network is trained using the training set, while regularly checking its performance on the validation set to prevent the network from memorizing the training set ('overtraining'). CNN performance: once the CNN achieves its optimal performance on the validation set, unbiased performance metrics are determined using the (so-far unused) test set.

The current DL techniques in renal histopathology mainly focus on single slide, bright field microscopy. This makes them very suitable for applications in renal transplant pathology, e.g. for quality assessment through glomerular counting and for the scoring of (peri)tubular or interstitial inflammation. The recent installation of the Banff Digital Pathology Working Group underlines the interest of the field in AI and will foster the application of DL in transplant pathology [10].

CHALLENGES

One of the largest limitations for AI development is the lack of publicly available large data sets with accompanying clinical metadata. Even though abundant numbers of biopsies exist, they reside in archives all over the world, are mostly not digitized and meta-data (if readily available) are not standardized. To further complicate matters, development of robust algorithms (insensitive to inter-laboratory variations) requires multicentre data sets. Moreover, in case of supervised ML, the requirement for pathologists' involvement to annotate cases may limit progress considerably.

Many clinicians and medical researchers do not fully understand how CNNs works, resulting in an inability to recognize the opportunities and pitfalls. While understandable, this might result in underestimation as well as overestimation of the potential of AI. AI training programmes specifically tailored for people working in medicine are therefore advised.

While current applications of AI are largely focused on evaluation of kidney transplant pathology, its use in the diagnosis of native kidney diseases lags behind. For the precise diagnosis of glomerular diseases, immunofluorescence (IF) and electron microscopy (EM) are usually required in addition to bright field microscopy. Use of AI here requires CNNs that are specifically trained for the analysis of IF and EM images, potentially even

combining multimodal data. An additional challenge in this field is posed by the rarity of most glomerular diseases, hampering the collection of large data sets that are generally needed for training a CNN. A joined effort of renal pathologists, nephrologists and computer scientists from different centres is required to address and solve this matter.

CONCLUSION

Currently developed DL algorithms were shown to yield objective and reproducible data from kidney WSIs. Deployed in a clinical setting, this may result in increased accuracy and reproducibility of diagnostics. The availability of accurate quantitative descriptors can help tailor treatment for individual patients. In research settings, the potential of AI to analyse large numbers of WSIs in a highly standardized manner will aid the development of novel biomarkers. For both clinical use and research, DL may increase efficiency by taking away certain tedious tasks from pathologists. To fully exploit these opportunities, establishment of large, well-curated multicentre data sets is required, as well as training and involvement of clinicians.

FUNDING

M.H. and J.v.d.L were supported by ERACoSysMed's SysMIFTA project, as part of the European Union's Horizon 2020 Framework Programme (grant number 9003035004), during the conduct of the study.

CONFLICT OF INTEREST STATEMENT

J.v.d.L. reports personal fees from Philips (Eindhoven, the Netherlands) and ContextVision (Linköping, Sweden) and grants from Sectra (Linköping, Sweden) and Philips (Eindhoven, the Netherlands), outside the submitted work.

REFERENCES

1. 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: 1795–1814
2. Hermsen M, de Bel T, den Boer M *et al.* Deep learning-based histopathological assessment of kidney tissue. *J Am Soc Nephrol* 2019; 30: 1968–1979
3. Santo BA, Rosenberg AZ, Sarder P. Artificial intelligence driven next-generation renal histomorphometry. *Curr Opin Nephrol Hypertens* 2020; 29: 265–272
4. Litjens G, Kooi T, Ehteshami Bejnordi B *et al.* A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60–88
5. Miotto R, Li L, Kidd BA *et al.* Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016; 6: 26094
6. Abels E, Pantanowitz L, Aeffner F *et al.* Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol* 2019; 249: 286–294
7. Steiner DF, MacDonald R, Liu Y *et al.* Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol* 2018; 42: 1636–1646
8. Ginley B, Lutnick B, Jen KY *et al.* Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol* 2019; 30: 1953–1967
9. Kolachalama VB, Singh P, Lin CQ *et al.* Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep* 2018; 3: 464–475
10. Farris AB, Moghe I, Wu S *et al.* Banff Digital Pathology Working Group: going digital in transplant pathology. *Am J Transplant* 2020; Doi: 10.1111/ajt.15850

Received: 19.5.2020; Editorial decision: 28.5.2020