# Toward Real-World Computational Nephropathology

Rodrigo T. Calumby [iD],[1] Angelo A. Duarte [iD],[2] Michele F. Angelo,[3] Emanuele Santos [iD],[4] Pinaki Sarder,[5] Washington L.C. dos-Santos [iD],[6] and Luciano R. Oliveira [iD][7]

In pathology, defining consistent criteria for disease diagnosis and prognostication requires ever increasing effort. Nephropathology, in particular, demands greater attention to the correspondence between distinctive patterns of kidney lesions and their associated clinical diagnoses. Some examples include changes in the glomeruar basement membrane and nephrotic syndrome, endocapillary hypercellularity and nephritic syndrome, and extensive glomerular crescents and renal dysfunction.

Although consensus has been achieved for diagnostic criteria and assessing the activity and chronicity of prevalent nephropathies, there remains a lack of substantial agreement in the classification of certain kidney lesions.[1] Some diagnoses require complex procedures to perform lesion scoring, which are labor intensive, time-consuming, and error prone. Automating the steps involved in diagnosis will represent an important achievement in the future of nephropathology. To attain this goal, digital pathology combined with machine intelligence evolves toward automating diagnostic processes in nephropathology. Recent methods exploit the fundamentals of computational analysis to classify, detect, and segment histological structures in digital whole slide images (WSIs) using visual pattern recognition.

The availability of artificial intelligence (AI) systems capable of counting histological structures and scoring disease activity and chronicity in WSI would affect nephropathology practice. Although many machine learning (ML)–based methods have achieved promising results in academic research, evidence needs to be provided regarding applying these methods in the real world.[2] It is due to differences between the settings of experiments conducted under controlled academic conditions versus the wilderness of clinical scenarios.[3] Such differences may become exacerbated when considering the limitations and methodological restraints of controlled studies in computational pathology. These shortcomings have motivated recent proposals detailing general requirements for ML applications in medical imaging.[4] Nevertheless, significant progress must be made regarding improving systems capable of accurate performance in a clinical environment. Despite the importance of incremental advances, most experimental validation continues to be performed using data with limited representativeness, and optimistic validation procedures result in poor generalization assessments and unreliable outcomes. Although these issues restrict progress, improvements in ML methods and computational infrastructure have enabled conducting of more robust experimental research

aimed at performing in a real-world context. Because these works are prone to multiple biases (*e.g.*, observer, recall, and sampling), efforts are needed toward building unbiased histological datasets representative of populations. Recently, National Institutes of Health (NIH) has established several consortia projects for hosting well-curated renal tissue images and omics data to be used by the community. Notable ones are the Human Biomolecular Atlas Program and the Kidney Precision Medicine Project. Other NIH consortiums with data for computational pathology studies are NEPTUNE, CureGN, and GUDMAP. This context prompts reflection on the *status quo* of computational nephropathology (CNP) to shape future research designed to produce intelligent tools suitable for real-world tasks. Although some works[5] broadly overview the current state of the CNP domain, this article addresses the practical experimentation challenges and directions for this field. Figure 1 presents a perspective on managing critical aspects to better conduct academic research in CNP.

## Data

WSI constitutes the core data source for ML in CNP. Although CNP experimentation must consider a plethora of lesion presentations in tissue samples and variations in the slide preparation and digitalization (fixing, embedding, slicing, staining, and scanning), most assessments of CNP methods are performed in narrowly controlled scenarios, involving small datasets that do not reflect real-world data variability and heterogeneity. These limitations hinder the development of accurate and practical computational models. Therefore, future CNP research should consider data from multiple laboratories[6] and seek to represent wide-ranging kidney lesion presentations. To facilitate such data generation, modern auxiliary computing tools can aid in gathering, curation, labeling, augmentation, and performing quality control and quality assessment in larger and varied data collections. In addition, it is necessary to research generalizable ML methods capable of learning from relatively few labeled samples and/or larger unlabeled sets that are easier to obtain.

## Validation and Analysis

Using large volumes of data with robust representativeness, complex ML models can be trained, thanks to modern high-performance computing infrastructure. The association between big data and computational

[1]Advanced Data Analysis and Management Lab, Department of Exact Sciences, University of Feira de Santana, Feira de Santana, Brazil
[2]High Performance Computing Laboratory, Department of Technology, University of Feira de Santana, Feira de Santana, Brazil
[3]Departament of Exact Sciences, University of Feira de Santana, Feira de Santana, Brazil
[4]Center of Sciences, Department of Computing, Federal University of Ceara, Fortaleza, Brazil
[5]Intelligent Critical Care Center, Department of Medicine – Nephrology, University of Florida, Gainesville, Florida
[6]Fundação Oswaldo Cruz, Gonçalo Moniz Institute, Structural and Molecular Pathology Laboratory, Salvador, Brazil
[7]Intelligent Vision Research Lab, Computer Science Department, Institute of Computing, Federal University of Bahia, Salvador, Brazil

**Correspondence:**
Dr. Washington L.C. dos-Santos, Rua Waldemar Falcão 121, Salvador, BA 40296-710, Brazil, or Dr. Luciano R. Oliveira, Av. Milton Santos, S/N, Salvador, BA 40170-110, Brazil. Email: washington.santos@fiocruz.br or lrebouca@ufba.br
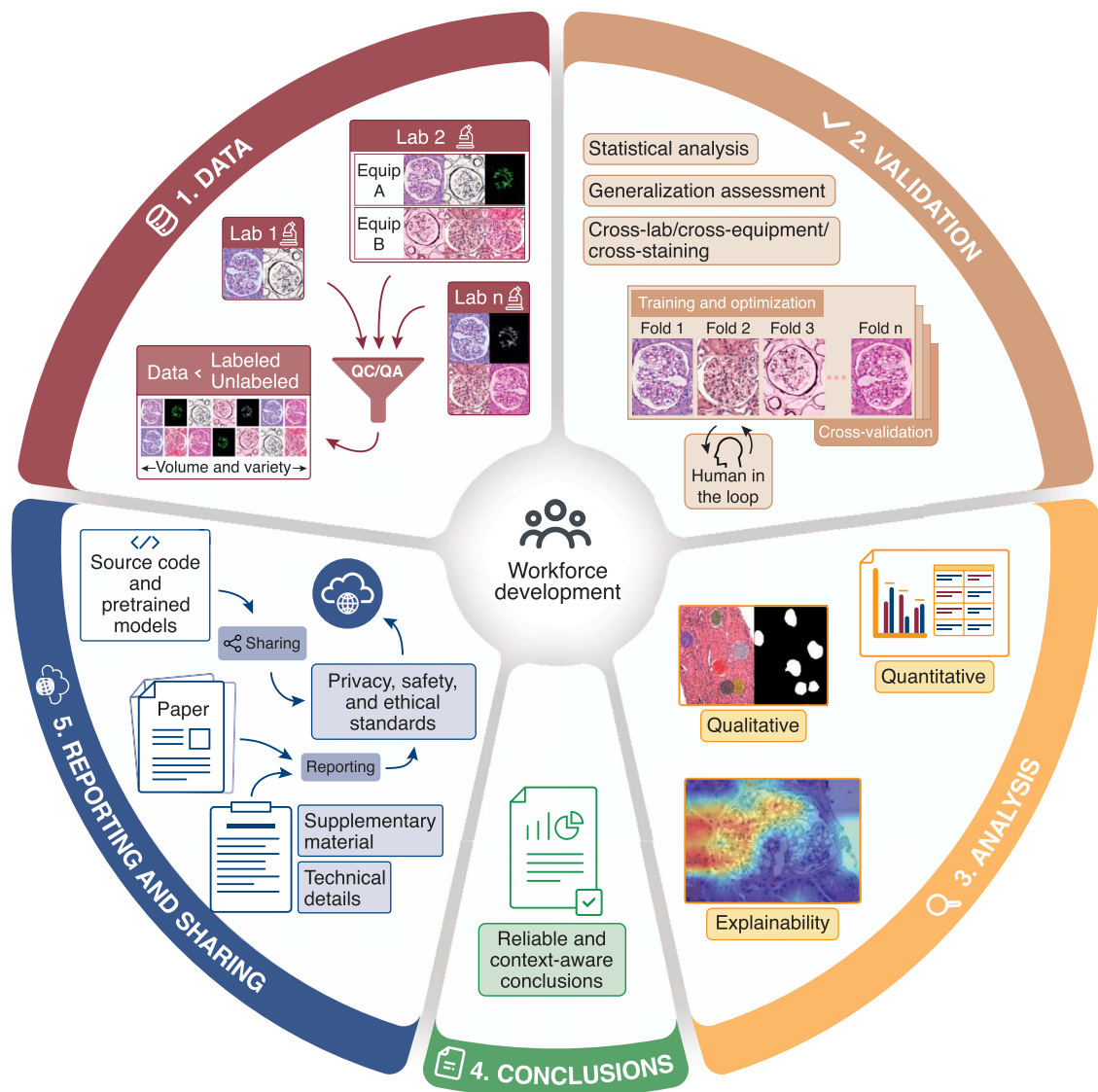
**Figure 1. The perspective of a real-world computational nephropathology (CNP) includes key aspects that must be improved to better conduct academic research.** Specifically, future work must consider the following: (*1*) the gathering and curation of representative labeled and unlabeled data from multiple laboratories (with QC/QA) as well as the training of generalizable ML models even in cases with few data samples; (*2*) the adoption of validation methodologies on the basis of rigorous cross-validation protocols and adequate metrics, enriched with cross-laboratory, cross-equipment, and cross-staining experiments and the exploitation of interactions between nephropathologists and the computational tools (human-in-the-loop); (*3*) the use of complementary qualitative and quantitative analysis, thorough statistical significance assessments, and powerful explainability resources; (*4*) be grounded on robust experimental designs to enable more reliable and context-aware conclusions; and (*5*) adhering to technical and scientific reporting guidelines as well as the sharing of data, source code, and pretrained models following privacy, safety, and ethical standards. Notably, avoiding biases is a transversal concern in CNP studies and must be considered throughout the workflow. All that should be empowered by the next-generation workforce, which must better understand the clinical needs. ML, machine learning; QA, quality assessment; QC, quality control.

power can leverage quantitative evaluations on the basis of rigorous cross-validation protocols, adequate metrics, and thorough statistical analysis to yield reliable generalization assessments.[7] The entire validation process can also be enriched with cross-laboratory, cross-equipment, and cross-staining studies. In addition, visual error analysis is expected to improve discussions on CNP experimentation by providing qualitative information. Along with clear descriptions of experimental limitations, meticulous qualitative and quantitative analysis will allow for deeper assessments of proposed methods and a clearer understanding of their capabilities,

shortcomings, and potential to evolve. Moreover, interactions between nephropathologists and computational tools can guide the evolution of ML models through human-in-the-loop supervision (annotation and feedback) and model training while at the same time calling attention to improving explainability—a key step to achieving clinical translatability.

## Reporting and Sharing

The evolution toward real-world–oriented CNP encompasses standardized practices to ensure proper peer review,

reproducibility, replicability, and repeatability. Adopting technical and scientific reporting guidelines can bring strategic advances to the field. Many ongoing initiatives are developing reporting guidelines for AI in health care, such as STARD-AI[8] and TRIPOD-AI,[9] which focus on the preclinical and offline validation of predictive models for diagnosis and prognosis. With respect to sharing, although gaps remain in the standardization of digital nephrology image data,[10] advances in CNP methodology can be powerfully leveraged if researchers make data, source code, and pretrained models available. Using transfer learning through pretrained deep-learning models allows for knowledge sharing and accelerates system development while avoiding sensitive data reveal. Sharing these elements must follow privacy, safety, and ethical standards to protect patient anonymity, guarantee controlled access, and adhere to fair research principles.

## Reliable Conclusions

Considering the effects of medical imaging analysis on the practice of nephropathology, it is important for preclinical studies to be grounded by robust experimental design. Otherwise, it might bias conclusions because of the lack of representative supporting data, proper validation methodology, and complementary quantitative and qualitative analysis.[3] In other words, real-world applications demand substantial datasets properly representing visual phenomena. It is also necessary for a well-defined experimental methodology that allows for comprehensive performance assessments, strict statistical validations that confirm the performance of the methods, and clear reporting of all research performed. These elements enable studies to draw more reliable and clear context-aware conclusions.

## Concluding Remarks

Parallel advancements in photonics have enabled the identification of molecular markers in digital histology images, vertebrate codex using the gene breaking protein trap library, and deep genomics information using spatial transcriptomics. Another milestone on the future roadmap of using AI as an assistive tool for clinical diagnosis will be the capability of inferring subcellular diversity abundance using specific molecular motifs on the basis of digital histology image data alone.

Integrating the workforce from the engineering and clinical domains is of critical importance so that these groups can exchange ideas more effectively and not work in isolation. This process will empower the next-generation workforce to understand clinical needs better while also using high-performance computational tools to address specific requirements.

Finally, advances in ML-based CNP prompt a debate on the plausibility of digitally assisted diagnosis in clinical practice. The reliability of AI systems remains a concern,[11] and to gain trust, we suggest deep diving into the dimensions discussed here regarding representative real-world data curation, rigorous validation protocols, critical and enriched analysis, meticulous reporting, and broader sharing to ultimately ground experimental designs and support reliable conclusions.

## Author Contributions

**Conceptualization:** Michele F. Angelo, Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos, Pinaki Sarder.
**Formal analysis:** Luciano R. Oliveira.
**Funding acquisition:** Washington L.C. dos-Santos, Pinaki Sarder.
**Investigation:** Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos, Pinaki Sarder.
**Methodology:** Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos, Pinaki Sarder.
**Project administration:** Michele F. Angelo, Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Pinaki Sarder.
**Resources:** Michele F. Angelo, Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos.
**Supervision:** Pinaki Sarder.
**Validation:** Washington L.C. dos-Santos, Luciano R. Oliveira.
**Visualization:** Michele F. Angelo, Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos.
**Writing – original draft:** Rodrigo T. Calumby, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Pinaki Sarder.
**Writing – review & editing:** Michele F. Angelo, Washington L.C. dos-Santos, Angelo A. Duarte, Luciano R. Oliveira, Emanuele Santos, Pinaki Sarder.

## References

1. Trimarchi H, Barratt J, Cattran DC, et al. Oxford classification of IgA nephropathy 2016: an update from the IgA nephropathy classification working group. *Kidney Int.* 2017;91(5):1014–1021. doi:10.1016/j.kint.2017.02.003

2. Huo Y, Deng R, Liu Q, Fogo AB, Yang H. AI applications in renal pathology. *Kidney Int.* 2021;99(6):1309–1320. doi:10.1016/j.kint.2021.01.015

3. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology.* 2018;286(3):800–809. doi:10.1148/radiol.2017171920

4. de Hond AAH, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med.* 2022;5(1):2. doi:10.1038/s41746-021-00549-7

5. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, Balis U. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol.* 2020;16(11):669–685. doi:10.1038/s41581-020-0321-6

6. Fontinele J, Souza L, Chagas P, et al. Boundary-aware glomerulus segmentation: toward one-to-many stain generalization. *Comput Med Imaging Graph.* 2022;100:102104. doi:10.1016/j.compmedimag.2022.102104

7. Mongan J, Moy L, Kahn- CE Jr. Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. *Radiol Artif Intell* 2020;2(2):e200029. doi:10.1148/ryai.2020200029

8. Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open.* 2021;11(6):e047709. doi:10.1136/bmjopen-2020-047709

9. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open.* 2021;11(7):e048008. doi:10.1136/bmjopen-2020-048008

10. Gurcan MN, Tomaszewski J, Overton JA, Doyle S, Ruttenberg A, Smith B. Developing the Quantitative Histopathology Image Ontology (QHIO): a case study using the hot spot detection problem. *J Biomed Inform.* 2017;66:129–135. doi:10.1016/j.jbi.2016.12.006

11. Kwade Z. Clinical evaluation of software. *RF Q.* 2022;2(1):23–34. https://www.raps.org/news-and-articles/news-articles/2022/3/clinical-evaluation-of-software.