Editorial
Computational Pathology

UNTIL recent years, histopathologists have analysed tissue sections and diagnosed diseases including cancer primarily by using a microscope. The introduction of high resolution and high throughput digital scanners has enabled digitizing entire glass slides to generate high-resolution whole-slide images (WSI), de facto giving rise to the field of *Digital Pathology*. Since then, an increasing number of pathology laboratories have transitioned to a digital pipeline, which offers advantages such as remote diagnosis (e.g., for a second opinion), reduction of some routine work in the lab and partly reduction of physical storage. However, perhaps the most revolutionizing aspect of digital pathology is that it enables *image analysis* in pathology using *machine learning*. This field has come to be known as *Computational Pathology*.

In parallel with the introduction of digital pathology, the computer science community has witnessed the rise of *deep learning* (DL), a subfield of machine learning and in turn *artificial intelligence* (AI), inspired by the way the brain works, which can learn and make predictions solely based on raw data, such as images. Motivated by advances in the computer vision community, where using large-scale datasets such as ImageNet have allowed computers to achieve superhuman performance at image classification, deep learning has quickly become a part and parcel of computational pathology. This has fostered the development of computer models trained using large sets of *manually annotated* digital pathology images to address tasks such as tissue detection, segmentation and classification, often reaching human-level performance. The application focus of these models has been very broad, ranging from computer aided diagnosis to precision medicine.

Computer-aided diagnosis (CAD) systems, such as algorithms for tumor detection or tumor grading, aim at assisting pathologists in routine diagnostics. As a result, AI-based systems have shown to reach and even outperform pathologists in diagnostic tasks such as detection of metastasis in lymph nodes [1] grading prostate cancer biopsies [2], [3] and to achieve a level of effectiveness to be considered as clinically applicable [4], [5]. However, most of these systems have been solely validated in research settings, and deployment of such systems in clinical practice results in a completely new set of research questions and implications. When applied to routine diagnostics, these systems will have to face variations arising at every step of the processing pipeline, which the human visual system adapts to quite easily, based on expert knowledge and experience, but might represent a cause of failure of computer models.

In order to properly address these potential issues, mechanisms of *quality control* (QC) should be implemented, as well as tools to measure how *certain* a model is about the prediction it makes, and how the test environment differs, in terms of data distribution, from the one used to train the model. Examples of such a *domain shift* between the training and test set are: differences in tissue staining, scan resolution, tissue preparation, etc. In the specific case of stain variation, several methods of *stain normalization* and stain augmentation have been proposed in the literature, initially based on traditional image analysis methods and recently on deep learning.

Another strong focus of computational pathology is on precision medicine, which addresses relevant questions about the clinical outcome of a patient in terms of treatment response, disease recurrence, and survival. Recent work in this field has used AI to analyze H&E-stained slides and predict genetic mutations [6], [7] and survival [8], [9]. At the same time, there is a growing interest in the field of immune oncology to describe and understand the tumor microenvironment (TME). In an era when new cancer treatments such as immunotherapy are increasingly being used, understanding the differences between patients and identifying potential responders at an early stage is key to personalize treatment and deliver better healthcare. Histopathology and immunohistochemistry offers a unique opportunity to understand those differences, where multiple tissue (serial) sections can be stained with different *immunohistochemical markers* to target a broad spectrum of different cell phenotypes. In particular, there has been a growing interest in the interplay between the immune system and cancer, with the prominent role of *tumor infiltrating lymphocytes* (TIL) and their role in the TME.

Finally, research is ongoing to tackle challenges such as the need for *large sets of manual annotations* to build effective AI system, where making annotations does not scale with the increasing amount of available data, and how to efficiently train DL models end-to-end using entire whole-slide images, which intrinsically poses a problem due to large size of WSI.

This special issue provides a snapshot of recent advances in research and applications of Computational Pathology. The thirteen papers included in this issue present novel research of high interest in the computational pathology community. This special issue arose primarily from, but is not limited to, extended versions of selected contributions to the 2nd MICCAI Workshop on Computational Pathology (COMPAY19) held in Shenzhen, China, in October 2019. In the next sections, we present the main highlights of each paper, grouping them into four main topics.

Digital Object Identifier 10.1109/JBHI.2021.3052029

I. Quality Control of Data and Prediction Models

The papers by Wright *et al.* [10] Cicalese *et al.* [11] and Stacke *et al.* [12] focus on the importance of quality control of pathology image data and on assessing the reliability and uncertainty of the models when deployed in routine diagnostics.

[10], the authors explore the effect of variations that naturally arise at every step of the digital pathology pipeline in image analysis. They used quantification of a well-known prognostic biomarker such as the tumor-stroma ratio (TSR) in colorectal cancer as a use case, assessed by both a previously published algorithm and pathologists on a set of more than 2000 cases from a clinical trial. Using TSR as a benchmark, they showed that disagreement between algorithm and pathologists' score could be attributed to artifacts (e.g., presence of coverslip, air bubbles, tissue folds) and to morphological characteristics of the tissue itself (e.g., poorly differentiated tumor, presence of necrosis) or due to the preparation procedure (e.g., weak staining). They also showed that training the same algorithm with only "good" cases (i.e., excluding cases with the aforementioned problems) reduces the performance of the model, due to lack of heterogeneity in the training set, which hampers the generalization.

[11], the authors tackle the problem of uncertainty estimation in classification of kidney biopsies for diagnosis of lupus nephritis. They adopt a Bayesian approach to estimate uncertainty at the level of glomeruli classification, and then take it to a framework for kidney classification at whole-slide image level. At this step, they only considered highly certain glomeruli and reject classification results with high uncertainty, based on a proposed selection heuristic. They showed that this technique is effective to improve classification accuracy at kidney-level classification, when compared to a CNN-based approach that does not take uncertainty into account. Next to its use as a way to improve classification accuracy, the same approach to estimate uncertainty is of high relevance when AI algorithms are deployed into clinical routine.

[12], the authors propose a "representation shift" metric, to quantify the magnitude of model-specific domain shift (e.g., differences in slide preparation, staining protocol, scanner properties, etc.) between the data that was used to train the model and the data that the model has to make predictions on when deployed in the clinic. They show that the proposed metrics correlate with drop in performance when a model is applied to a test set with a domain shift with respect to the training set, outperforming existing approaches to measure domain shift. The presented method can be seen as a tool complementary to uncertainty estimation, to measure the (un)reliability of a model and its predictions, which is relevant for understanding if a deployed prediction model is performing as expected. Methods for stain normalization are a set of algorithms often present in the toolkit of computational pathology developers, which can be used to mitigate the domain shift caused by intrinsic difference in staining across different pathology laboratories, or even within the same laboratory when staining happens at different time points. In this special issue, the paper by Zheng *et al.* [13] proposed a novel approach to stain normalization using the recent capsule network model and dynamic routing algorithm, which was shown to outperform existing methods of stain normalization and stain augmentation on most of the considered benchmarks.

II. Computational Pathology for Computer-Aided Diagnosis

The papers by Vu *et al.* [14] Lei *et al.* [15] and Zhang *et al.* [16] focus on computer-aided diagnosis.

[14], the authors present an approach to assist tumor grading based on unsupervised learning. A model based on Conditional Generative Adversarial Networks is trained to learn characterization of benign tissue, which they call "BenignGAN", by asking the model to reconstruct the tissue itself from the map of its image edges. The idea behind this method is that the trained model is not capable of reconstructing malignant tissue when the tumor grade, i.e., the difference with respect to healthy epithelial cells, increases. Using colorectal cancer as a use case, they show that the distance between benign and malignant epithelial regions increases as cell differentiation decreases, from well-differentiated (low-grade) tumor to poorly-differentiated (high-grade) tumor. This work goes in the direction of using AI to potentially reformulate the currently used grading scheme, moving from fixed categories (e.g., Gleason grading) currently used in the clinic, to interpreting grade as a continuous transition from benign to malignant, potentially carrying additional information with potential predictive and prognostic value.

In the context of tumor grading, automated mitotic count, one of the components of breast cancer grading, is addressed [15] adopting recent advances in computer vision and DL such as the Region Proposal Network and an Attention Model.

[16], the results of the ACDC challenge on segmentation of lung cancer in both surgical resections and diagnostic biopsies are presented. They show that some of the presented methods have performance within the inter-observer variability. These methods are potentially applicable to detection of lung tumor in routine diagnostic, for example by scanning preoperative biopsies, as well as in lung cancer research, by accurately segmenting the tumor and separating it from other components of the TME.

III. Computational Pathology for Precision Medicine

The papers by Krijgsman *et al.* [17] Solorzano *et al.* [18] and Lahiani et al [19] focus on computational pathology methods to investigate protein co-expression, which is plays a central role in research and development of digital biomarkers that can offer new insights relevant for the clinical decision making process.

[17], the authors investigate the interplay between tumor and cytotoxic T-cells in the tumor microenvironment in the context of ER-positive breast cancer. For each case, two serial sections were obtained, one stained with H&E and one stained with CD8. The H&E slide was analyzed to segment the main morphological components of the tissue, such as the tumor and the stroma. Post-processing the output allowed to define the region of the central tumor, as well as the invasive margin. After image registration, the H&E and the CD8 slide were aligned, and lymphocytes detected in the CD8 slide were linked to morphological regions derived from the H&E slide. As a result, correlation was shown between some statistics of CD8 density distribution and clinical outcome. These results are in line with ongoing research on the predictive and prognostic value of tumor infiltrating lymphocytes. The connection between serial sections stained with H&E and CD8 via image registration may open the door to further research into a cost-effective and efficient

characterization of the tumor microenvironment, where multiple markers can be investigated without the need for multiplexing technology.

[18], the authors present a pipeline to analyze protein co-expression in tissue microarray (TMA) slides. The pipeline was entirely built using open-source software, and applied to 142 gastric cancer cases stained with both E-cadherin and CD44v6. The method allowed to a) align multiple slides, b) evaluate alignment quality, c) find the tumor region via classification on the CD44v6 marker, and d) quantify protein expression. A "co-expression color map" was proposed, to easily identify and visualize regions with different levels of co-expression.

[19], the authors address the problem of synthesizing digitally stained slides from using Cycle Generative Adversarial Networks (CycleGAN). They propose a novel Perceptual Embedding Consistency loss term to deal with the tiling effect obtained applying CycleGAN to a whole-slide image in a sliding window fashion, necessary to avoid memory problems due to the size of WSI. They showcase the proposed method synthesizing FAP-CK slides from H&E, and showed good performance when comparing the artificial slide with the actual serial section stained with the same marker, as well as good agreement of pathologists when assessing the real and the virtual slide. When validated against clinical molecular or clinical targets, this approach could be applicable to the analysis of protein co-expression in settings with a limited set of markers of limited number of slides (e.g., sections from tumor biopsies).

IV. Novel AI-Based Technology for Computational Pathology

The last set of papers by Mormont *et al.* [20] Stenman *et al.* [21] and Wieslander *et al.* [22] present novel AI-based technology to support development of future computational pathology systems for decision support.

[20], the authors present a valid alternative to commonly used DL models pre-trained on ImageNet. Since a single large-scale alternative to ImageNet does not exist in the field of digital pathology, the authors assembled publicly available datasets into a pool of 22 classification tasks with almost 900000 images, which they used to design a framework based on multi-task learning to pre-train a DL model. They showed that such a pre-trained model could be used for fine-tuning and for feature extraction in downstream tasks, reporting performance at least comparable or significantly better than an ImageNet pre-trained model. This multi-task pre-trained model has the potential to become largely adopted by the computational pathology community as the base model to develop new DL systems, replacing ImageNet pre-trained networks. This adoption can be facilitated by the publicly available code and pre-trained models, which have been recently released [23].

Similar to [17] and [18] the work [21] relies on a combination of multiple markers, in this case to generate a reference standard to build DL models. The authors used slides from papillary thyroid carcinoma cases stained with H&E, which were digitized, de-stained and re-stained with an anti CD45 marker. As done [24] after image registration, positive regions in IHC were extracted, post-processed, and linked to morphological regions in H&E, and finally used to train CNN for tissue segmentation in H&E. Next to showing good segmentation performance, the relevance of this work is in using IHC to generate a reference standard to train downstream DL models, with limited or no need for manual annotations, which eases the generation of high-quality annotations when large-scale data is used.

Finally, [22] the authors proposed a hierarchical approach to the analysis of WSI, to tackle the problem of memory limitation due to the size of digital pathology images. Focusing on COPD and idiopathic pulmonary fibrosis (IPF) and on quantifying region-specific drug response in lung tissue, they first decompose the slide into multiple regions of interest, detected automatically using CNNs at low-resolution, followed by segmentation of multiple morphological components at a medium level of resolution. One of the novel aspects of their approach was the use of statistical learning theory of Conformal Prediction (CP), which provides confidence measure for a prediction, applied here to pixel-wise classification in semantic segmentation at high resolution. Finally, they aggregate findings at WSI level in the detected regions to make a prediction about drug response at WSI level using CP. This work is complementary to recent developments in the field of end-to-end learning using entire WSI as input to the model, and training it to predict targets such as presence of cancer, genetic mutations, treatment response, or survival. These techniques can enable the use of very large-scale training sets where only image-level labels are available, can target end-points that are highly relevant to the patient (e.g., survival), and potentially allow discovery and understanding of novel underlying mechanisms in the tumor microenvironment.

Francesco Ciompi
Diagnostic Image Analysis Group
Department of Pathology
Radboud University Medical Center
6525 EZ Nijmegen, The Netherlands
francesco.ciompi@radboudumc.nl

Mitko Veta
Eindhoven University of Technology
5600 MB Eindhoven, The Netherlands
m.veta@tue.nl

Jeroen van der Laak
Diagnostic Image Analysis Group
Department of Pathology
Radboud University Medical Center
6525 EZ Nijmegen, The Netherlands
jeroen.vanderlaak@radboudumc.nl

Nasir Rajpoot
University of Warwick
CV4 7AL Coventry, UK.
n.m.rajpoot@warwick.ac.uk

References

[1] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *J. Amer. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, 2017.

[2] W. Bulten *et al.*, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study," *Lancet Oncol.*, vol. 21, no. 2, pp. 233–241, 2020.

[3] P. Ström, *et al.*, "Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study," *Lancet Oncol.* vol. 21, pp. 222–232, 2020.

[4] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Med.*, vol. 25, pp. 1301–1309, 2019.

[5] Z. Song *et al.*, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Commun.*, vol. 11, 2020, Art. no. 4294.

[6] N. Coudray *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, pp. 1559–1567, 2018.

[7] J. Nikolas Kather *et al.*, "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature Med.*, vol. 25, pp. 1054–1056, 2019.

[8] P. Courtiol *et al.*, "Deep learning-based classification of mesothelioma improves prediction of patient outcome," *Nature Med.*, vol. 25, pp. 1519–1525, 2019.

[9] E. Wulczyn *et al.*, "Deep learning-based survival prediction for multiple cancer types using histopathology images," *PLoS One*, 2020.

[10] A. I. Wright, C. M. Dunn, M. Hale, G. A. Hutchins, D. Treanor, "The effect of quality control on accuracy of digital pathology image analysis," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3046094.

[11] P. A. Cicalese, A. Mobiny, Z. Shahmoradi, X. Yi, C. Mohan and H. Van Nguyen, "Kidney level Lupus Nephritis classification using uncertainty guided bayesian convolutional neural networks," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3039162.

[12] K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "Measuring domain shift for deep learning in histopathology," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3032060.

[13] Y. Zheng *et al.*, "Stain standardization capsule for application-driven histopathological image normalization," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2983206.

[14] Q. D. Vu, K. Kim, J. T. Kwak, "Unsupervised tumor characterization via conditional generative adversarial networks," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2983206.

[15] H. Lei, S. Liu, A. Elazab, X. Gong, and B. Lei, "Attention-guided multi-branch convolutional neural network for mitosis detection from histopathological images," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3027566.

[16] Z. Li *et al.*, "Deep learning methods for lung cancer segmentation in whole-slide histopathology images – the ACDC@LungHP challenge 2019," to be published, doi: 10.1109/JBHI.2020.3039741

[17] D. Krijgsman *et al.*, "Quantitative whole slide assessment of tumor infiltrating CD8-positive lymphocytes in ER positive breast cancer in relation to clinical outcome," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3003475.

[18] L. Solorzano *et al.*, "Towards automatic protein co-expression quantification in immunohistochemical TMA slides," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.3008821.

[19] A. Lahiani, I. Klaman, N. Navab, S. Albarqouni, and E. Klaiman, "Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2975151.

[20] R. Mormont, P. Geurts, and R. Maree, "Multi-task pre-training of deep neural networks for digital pathology," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2992878.

[21] S. Stenman *et al.*, "Antibody supervised training of a deep learning based algorithm for leukocyte segmentation in papillary thyroid carcinoma," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2994970.

[22] H. Wieslander *et al.*, "Deep learning with conformal prediction for hierarchical analysis of large-scale whole-slide tissue images," *J. Biomed. Health Inf.*, to be published, doi: 10.1109/JBHI.2020.2996300.

[23] https://github.com/waliens/multitask-dipath

[24] W. Bulten *et al.*, "Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard," *Nature Sci. Rep.*, vol. 9, no. 1, 2019.