The revolving door for AI and pathologists—docendo discimus?

Simone L. Van Es[1], Anant Madabhushi[2,3]

[1]Department of Pathology, School of Medical Sciences, UNSW Medicine, UNSW Sydney, Australia; [2]Department of Biomedical Engineering, Center for Computational Imaging and Personalized Diagnostics, Case Western Reserve University, Cleveland, OH 44106, USA; [3]Louis Stokes Cleveland Veterans Administration Medical Center, Cleveland, OH 44106, USA

Correspondence to: Dr Simone L. Van Es. Department of Pathology, School of Medical Sciences, UNSW Medicine, UNSW Sydney, NSW 2052, Australia. Email: s.vanes@unsw.edu.au.

Response to: Dietz RL, Pantanowitz L. The future of anatomic pathology: deus ex machina? J Med Artif Intell 2019;2:4.

Dietz and Pantanowitz (1) present a well-explained and informative viewpoint on the history, theory, and science behind, as well as, current and potential future uses and challenges of, artificial intelligence (AI) and machine learning (ML), for pathology. They emphasize the importance of development of a "killer suite" of AI applications whose use is evidence-based, that will accelerate acceptance and integration of digital pathology (DP) into diagnostic practice. This is an invited reflection on their editorial content with reference to findings from other groups.

Dietz and Pantanowitz remind us that use of AI in pathology is not new (1). The authors explain that machine learning is a branch of AI, where algorithms are developed from training data to predict outcome for test data. With deep machine learning (DL), a software model of a neural network, with multiple layers, is given data with each successive layer in the network learning from the previous layer. Rather than giving it instructions to perform a task, it is given huge amounts of data to learn the best possible representations to perform the task and to learn how to adapt in the most effective way with increased exposure to data (2). The two main categories of ML, are supervised and unsupervised (2). The unsupervised learning technique identifies hidden patterns or intrinsic structures in the input data and uses these to generate a meaningful output.

DL is being increasingly established over traditional machine learning with potential for more sophisticated performance compared to humans (3,4). The characteristics of tumours and their hosts represent a wealth of data to be mined and the investigative disciplines of medicine are fertile ground for development of sophisticated AI tools. The ability to extract complex information from scanned H&E stained slides, coupled with other laboratory tests, could lead to new diagnostic, and theranostic information (5). Madabhushi *et al.* describe exciting potential for data fusion algorithms combining radiological, histological, and molecular characteristics of a tumour for prognostic and predictive purposes (5,6).

However, just as we are finding that there are areas in tissue pathology where digital pathology is unsuitable, or needs the back-up of traditional glass slides, such as when looking at eosinophils, dysplasia and small microorganisms like helicobacter (7), so too, it is most likely that there will be areas in diagnostic pathology, where AI/ deep machine learning will not be a suitable diagnostic replacement. Dietz and Pantanowitz (1) comment that "AI is a tool and like most tools works best in certain situations". They emphasize that DP and AI are not the "Deus ex machina of anatomical pathology". Vamathevan (2) similarly explains that despite the potentially high value of ML for diagnostic pathology, it unfortunately does not have an all-purpose capacity.

Firstly, although a good ML model can generalize well from training data to test data (2), some AI output can give biased results. Israni *et al.* (8) comment that ""flawed or incomplete data sets that are not inclusive can automate inequality". This includes situations where the original data was of variable quality, with lack of standardisation of scanners and staining of tissue between laboratories, with a paucity of a wide disease range of expert-annotated examples, thus not reflective of the heterogeneity of real-life samples (9,10).

Another challenge is that AI is not able to incorporate contextual knowledge, into the diagnosis. Pathological diagnosis is complex, and for the one pattern under the microscope, there may also be a range of potential diagnoses and it is the clinical information and the ability to interpret that information that needs to be incorporated into the final diagnostic decision. It is this aspect of diagnostic decisions that is hard to quantitate, that "clinical acumen" or "gut feeling" based on years of clinical experience, and depth of knowledge. For machine learning, algorithms cannot make predictions that incorporate human emotion and response to the result. Claridge reminds us that "the best clinicians often make decisions based on their instincts which have developed through experience" when talking about introduction of new decision tools into clinical practice (11).

Schattner describes the persistent rate of diagnostic errors and iatrogenic harm despite "advances in scientific knowledge and technological capabilities" (12). She encourages us to return to 3 clinical paradigms that should underscore all clinical practice: pre-symptomatic diagnosis; skillful history and physical exam in informing decision-making; enhanced attention to patient autonomy and emotional factors. These are features that arguably cannot be incorporated into AI algorithms. AI technology holds great promise to deliver a more sophisticated, efficient and safer health care (3,4), hence the promised optimistic future (7,13). But our clinical acumen needs to be strong and skillful to critically interpret any results. It is our quality assurance and ongoing validation tools that should ensure this, as well as, to continue teaching our medical students the essential thought process and critical interpretation of investigations (14).

Additionally, there is the "black-box" approach of DL methods that Dietz and Pantanowitz (1) refer to, and that other groups emphasize (2), as the lack of transparency in the rationale behind DL decisions for classification tasks. The lack of interpretability in how DL arrives at its output, also makes it hard to troubleshoot difficulties. Vamathevan et al. (2) point out that for histological diagnosis of complex cancers such as melanoma, where diagnostic stakes are very high (arguably one of the most difficult cancer diagnoses to make and one that is associated with high rates of litigation) (15,16), this 'black-box' may become the choke for regulatory agencies because a suitable explanation as to how the result was derived with the DL process, evades us. Vamathevan et al. also alludes to the human factor in this acceptance of AI (2); are we able to trust a result derived in such a way, enough to rely on it into the diagnostic workflow?

Dietz and Pantanowitz (1) suggest that the enormous benefit from DL may not be realized in anatomical pathology for some time. But the question is, what do we actually want AI to do for investigative medicine. Is it cost-saving, time-saving, to provide validation and quality, or perhaps better accuracy and thus safer practice? And will cost of validation and integration be prohibitive?

Dietz and Pantanowitz (1) anticipate that 'regulatory bodies will in the near future approve deep learning techniques that arrive at a diagnosis through a "black box."' However, they wisely advise that we must proceed cautiously during "this dawn of AI in pathology". Every test we employ in medicine, pathology or otherwise is imperfect with false positives and negatives. Predictive and prognostic tools delivered through AI would be no different, underscoring our constant obligation as practitioners to use cautious skepticism rather than blind acceptance when considering any result.

Dietz and Pantanowitz are right to be cautious (1) because we have no full sense of how the machine is learning, often seeming to target alternative foci in an image compared to the human eye (5). However, looking at this from a different perspective, DL/AI is opening our eyes to different ways to assess tumour tissue and its microenvironment. Lee et al. (17) showed that machine learning focused on benign tissue surrounding prostate tumour which turned out to have prognostic value. Similar findings were described by Beck et al. (18), where stromal characteristics surrounding breast carcinoma tumour cells rather than the cells themselves carried a stronger association with survival. The machines have alerted us to important prognostic features within the tissue that historically have not seemed important when diagnostically analysing cancer tissue under the microscope. Perhaps this suggests that once the machines receive their high volume, high quality standardized training data, that our true pathological future will be the machines training us.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Mingyu Chen (Department of General Surgery, Sir Run-Run Shaw Hospital, Zhejiang University, Hangzhou, China).

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2019.05.02). Dr Van Es is a member of the Australian NPAAC (National Pathology Accreditation Advisory Council) Digital Pathology Drafting committee; She is also involved in developing digital pathology resources for the Royal College of Pathologists of Australasia (RCPA). Dr. Madabhushi is an equity holder in Elucid Bioimaging and in Inspirata Inc. He is also a scientific advisory consultant for Inspirata Inc. In addition he has served as a scientific advisory board member for Inspirata Inc, Astrazeneca and Merck. He also has sponsored research agreements with Philips and Inspirata Inc. His technology has been licensed to Elucid Bioimaging and Inspirata Inc. He is also involved in a NIH U24 grant with PathCore Inc, and 3 different R01 grants with Inspirata Inc.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Dietz RL, Pantanowitz L. The future of anatomic pathology: deus ex machina? J Med Artif Intell 2019;2:4.
2. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019. [Epub ahead of print].
3. Djuric U, Zadeh G, Aldape K, et al. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. NPJ Precis Oncol 2017;1:22.
4. Olsen TG, Jackson BH, Feeser TA, et al. Diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology. J Pathol Inform 2018; 9:32.
5. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. Med Image Anal 2016;33:170-5.
6. Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology. Annu Rev Biomed Eng 2016;18:387-412.
7. Van Es SL. Digital Pathology: semper ad meliora. Pathology 2019;51:1-10.
8. Israni ST, Verghese A. Humanizing artificial intelligence. JAMA 2019;321:29-30.
9. Janowczyk A, Zuo R, Gilmore H, et al. HistoQC: An open-source quality control tool for digital pathology slides. JCO Clin Cancer Inform 2019;3:1-7.
10. Leo P, Elliott R, Shih NNC, et al. Stable and discriminating features are predictive of cancer presence and Gleason grade in radical prostatectomy specimens: a multi-site study. Sci Rep 2018;8:14918.
11. Claridge LC. Should we replace the art of clinical acumen with the science of clinical decision rules? BMJ 2010;341:c5204.
12. Schattner A. Clinical paradigms revisited. Med J Aust 2006;185:273-5.
13. Glassy EF. Digital pathology: quo vadis? Pathology 2018;50:375-6.
14. Van Es SL, Grassi T, Velan GM, et al. Inspiring medical students to love pathology. Hum Pathol 2015;46:1408.
15. Marghoob AA, Changchien L, DeFazio J, et al. The most

common challenges in melanoma diagnosis and how to avoid them. Australas J Dermatol 2009;50:1-13; quiz 14-5.

16. Abikhair MR, Mahar PD, Cachia AR, et al. Liability in the context of misdiagnosis of melanoma in Australia. Med J Aust 2014;200:119-21.

17. Lee G, Veltri RW, Zhu G, et al. Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: preliminary findings. Eur Urol Focus 2017;3:457-66.

18. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011;3:108ra113.