The role of IT in pathology: A discussion of the potential impacts and associated challenges of machine learning in clinical practice and research

Word count: 979

Artificial intelligence (AI) has the potential to bring unimaginable benefits to human society, not least in the field of medicine. There are those who are not so optimistic about the influence of this technology, however, which some believe may cause unemployment and social inequality, amongst various other problems[1]. The application of machine learning (ML) algorithms to virtual slides and molecular datasets will allow the development of a deeper understanding of disease processes, but has also led to speculation that histopathologists may, in the not-so-distant future, be replaced by such algorithms[2]. ML will certainly change day-to-day practice in pathology, though it is likely that complete replacement of the histopathologist will not occur without AI which equals or surpasses human intelligence.

Whole-slide imaging permits the use of ML techniques for the automated interpretation of tissue samples[3]. The ability of pigeons to identify breast cancer clearly shows that the accurate analysis of tissue does not require any innately human qualities[4]. ML has already been successfully used for the diagnosis, staging, and grading of several cancers with a degree of confidence approaching that of expert pathologists[5–7]. Significant technical hurdles remain, however, before this approach can be reliably used to improve routine clinical practice. Performance consistently higher than the most experienced human pathologist would likely require 'training' of algorithms on an extremely large amount of data[8]. A sufficient volume of data, albeit in analogue form, does exist - millions of slides are held in storage around the world - but utilisation of this resource is currently complicated by factors such as variation in staining protocols, which make it difficult for ML algorithms to develop generalisable conclusions[9,10], as well as costs associated with the scanning process[11]. Additionally, each high-resolution virtual slide contains approximately ten gigabytes of data[12], meaning that this approach is computationally intensive, possibly to a degree that is impractical for many current research centres[13].

Despite this, it is conceivable that solutions to these problems will be found in the near future. 'Supervised' ML, where a computer programme is instructed to examine images for pre-defined features[14], such the number of observable mitotic figures, will aid pathologists by reducing the amount of time spent on tasks where tissue features must be quantified, leaving them free to spend more time on the interpretation of difficult cases[13]. The role of pathologists in medicine is likely to be more significantly influenced by 'unsupervised', or 'deep' ML. This form of ML, employed by Google's DeepMind project, which famously defeated the world *Go* champion in 2015[15], differs from supervised ML in that the programmes used are 'domain-agnostic', and are not instructed to look for specific features[8]. Modelled on neural networks, unsupervised algorithms are capable of identifying solutions to problems that humans are incapable of envisioning[14]. There is a very real possibility that this approach could allow algorithms to exceed the capabilities of any human histopathologist in the interpretation of slides[16]. Whilst this may unsettle some members of the profession, effective, automated slide interpretation techniques could dramatically widen the availability of pathology services in less developed countries, in addition to reducing the incidence of diagnostic errors worldwide[17].

This technology may lead to some downsizing of the histopathology workforce, or at least reduce the amount of time spent conducting clinical work. Nevertheless, it seems highly unlikely that pathologists are imminently about to be replaced by machines, for several reasons. Firstly, macroscopic examination of tissue, with dissection of resection or autopsy tissue and sampling of material for tissue blocks, requires clinical knowledge and technical skill; such a task is likely to be much more challenging for a machine to perform compared to the examination of standardised images. Additionally, the role of the pathologist is not limited to examination of tissue in isolation. Pathologists must integrate different types of macroscopic, microscopic, and molecular information from a range of different sources, along with, crucially, clinical variables, which are difficult to convert into a standardised form for easy interpretation by an algorithm[13].

Furthermore, production of programmes that can function as well as human pathologists in the interpretation of slides will not affect demand for research-active pathologists. In the past decade, the amount of experimental data produced in the field of biology has skyrocketed, largely as a result of advances in high-throughput omics technologies[18]. The progression of biomedicine is now much more significantly limited by the ability of scientists to process and interpret this data, rather than generate it[19]. ML algorithms and their analytical capabilities will be undoubtedly be invaluable in solving the ‘data deluge’ problem[20], but will require appropriate guidance from clinician-scientists trained in pathology.

It is important that the profession quickly adopts the use of ML in both clinical and research settings, which will require changes in working and educational practices. Provision of adequate data science training alongside current clinical training in order to permit proper utilisation of ML may prove difficult, but will quickly yield significant benefits[2]. Additionally, it is likely that increases in efficiency gained through the use of ML in routine clinical practice will lead to many pathologists spending increasing amounts of time conducting research, rather than clinical work. Trainees must, therefore, continue to be encouraged to undertake research degrees, with a stronger focus on applied data science.

The suggestion that developments in ML will render clinical histopathologists obsolete in the near future[13] is based on a flawed understanding of their role. Histopathologists have an obligation to employ ML in clinical practice if this will improve patient care, but should not fear unemployment as a result of these techniques. It is certainly true that the development of artificial 'general' intelligence, able to perform any intellectual task as well as or better than a human[1], might bring an end to the era of human pathologists. No profession, however, is entirely safe from this eventuality, the existential significance of which should perhaps cause greater concern than its potential impacts on future employment prospects.

References

1. Bostrom N. Superintelligence: Paths, Dangers, Strategies. 1st ed. Oxford University Press; 2014.

2. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. Journal of the American Medical Association. 2016;316(22):2353–4.

3. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific Reports. 2016;6(26286).

4. Levenson RM, Krupinski EA, Navarro VM, Wasserman EA. Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. PLoS ONE. 2015;10(11):1–21.

5. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis Detection in Breast Cancer Histology Images using Deep Neural Networks. Medical Image Computing and Computer-Assisted Intervention. 2013;411–8.

6. Su H, Xing F, Kong X, Xie Y, Zhang S, Yang L. Robust Cell Detection and Segmentation in Histopathological Images Using Sparse Reconstruction and Stacked Denoising Autoencoders. Medical Image Computing and Computer-Assisted Intervention. 2015;9351:383–90.

7. Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nature Communications. 2016;7(12474).

8. Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging. 2014;1(3).

9. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics. 2016;7(1):29.

10. Vandenberghe ME, Scott MLJ, Scorer PW, Söderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. Scientific Reports. 2017;7(45938).

11. Boyce BF. Whole slide imaging: uses and limitations for surgical pathology and teaching. Biotechnic & Histochemistry. 2015;1–10.

12. Chlipala E, Elin J, Eichhorn O, Krishnamurti M, Long RE, Sabata B, et al. Archival and Retrieval in Digital Pathology Systems. Digital Pathology Association. 2016;

13. Granter SR, Beck AH, Papke DJ. AlphaGo, Deep Learning, and the Future of the Human Microscopist. Archives of Pathology & Laboratory Medicine. 2017;141(5):619–21.

14. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nature Reviews: Genetics. 2015 May 7;16(6):321–32.

15. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529(7587):484–9.

16. Granter SR, Beck AH, Papke DJ. Straw men, deep learning, and the future of the human microscopist: Response to “artificial intelligence and the pathologist: Future frenemies?” Archives of Pathology and Laboratory Medicine. 2017;141(5):624.

17. Sharma G, Carter A. Artificial Intelligence and the Pathologist: Future Frenemies? Archives of Pathology & Laboratory Medicine. 2017 May;141(5):622–3.

18. Gehlenborg N, O’Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. Nature Methods. 2010;7(3s):S56–68.

19. Bellazzi R, Diomidous M, Sarkar IN, Takabayashi K, Ziegler A, McCray AT. Data Analysis and Data Mining: Current Issues in Biomedical Informatics. Methods of Information in Medicine. 2011;50(6):536–44.

20. Berger B, Peng J, Singh M. Computational solutions for omics data. Nature Reviews: Genetics. 2013 Apr 18;14(5):333–46.