Current and future applications of artificial intelligence in pathology: a clinical perspective

Emad A Rakha,1 Michael Toss,1 Sho Shiino,1 Paul Gamble,2 Ronnachai Jaroenstip1, Craig H Mermel,2 Po-Hsuan Cameron Chen2

ABSTRACT

During the last decade, a dramatic rise in the development and application of artificial intelligence (AI) tools for use in pathology services has occurred. This trend is often expected to continue and reshape the field of pathology in the coming years. The deployment of computational pathology and applications of AI tools can be considered as a paradigm shift that will change pathology services, making them more efficient and capable of meeting the needs of this era of precision medicine. Despite the success of AI models, the translational process from discovery to clinical applications has been slow. The gap between self-contained research and clinical environment may be too wide and has been largely neglected. In this review, we cover the current and prospective applications of AI in pathology. We examine its applications in diagnosis and prognosis, and we offer insights for considerations that could improve clinical applicability of these tools. Then, we discuss its potential to improve workflow efficiency, and its benefits in pathologist education. Finally, we review the factors that could influence adoption in clinical practices and the associated regulatory processes.

INTRODUCTION

The use of information technology (IT) tools for automated tissue processing, staining and generation of synoptic reports, have been performed in pathology laboratories for a long time.1,2 However, the introduction of digitalised whole slide image (WSI) technology has attracted the greatest attention of pathologists/clinicians and the use of digital pathology (DP) is now becoming commonplace.3,5 Currently, in the UK, there are several pathology laboratories that are either fully or partially digitalised, and various governmental funds are provided to increase the adoption of this technology. The introduction of DP for primary diagnosis in pathology has been a paradigm shift in the clinical practice. DP allows remote diagnostic reporting with flexible working hours, ensures service delivery even in challenging situations such as COVID-19 pandemic, and facilitates second and expert opinion reporting. More importantly, the adoption of DP enables the implementation of artificial intelligence (AI)-based algorithms in routine practice.1,4,9 Concrete applications include workflow-related algorithms such as automated triaging and quality control processes, assisted read of WSI and integration of multimodal data to support patients management. Of the recent advances in computational pathology (online supplementary appendix), the development of image-based diagnostic, prognostic and predictive algorithms have led to increasing interest in the applications of AI-based tools.

The most commonly used image-based AI algorithms in DP are convolutional neural networks.10,11 These applications use digitalized pathological WSI as input and learn associations between specific features and labels of interest such as diagnoses from pathologists, underlying molecular characteristics and outcome measures such as patients’ survival or response to adjuvant/neoadjuvant therapy.12 AI has also been shown to improve concordance between pathologists in many areas, such as the assessment of certain subjective features like cytonuclear pleomorphism/degree of atypia, granularity and counting mitotic figures, scoring of tumour infiltrating lymphocytes and assessment of proliferation (Ki67) index.4,15-18 Meanwhile, these algorithms have the potential to go beyond the visual assessment of the main histopathological features to capture subtle patterns which are currently beyond human recognition such as tumour microenvironment. AI can combine the information obtained from such myriad of features to make the overall diagnoses, and then link these features to other patients-related data to provide potentially valuable information on disease behaviour, patients’ outcome or prediction of response to therapy.19,20 Thus, AI technology can be used to support the overall reporting system, and measure morphological features more objectively as well as to speed up reporting time. AI-aided reporting of certain features or lesions will also enable pathologists to focus on challenging cases and meet the increasing workload demands. Implementation of such technology in the workflow of pathology service is not to replace the human resources including pathologists, and laboratory technicians, but to provide support for them and improve diagnostic and performance efficiency with better allocation of resources, increased cost-effectiveness of the service and more consistent pathology reviews.21

Despite these potential applications and benefits of the AI, its use in routine pathology services remains to be encouraged, various scenarios need to be considered and the whole process related to its implementation and the end-to-end use in routine practice needs to be contemplated. The focus of this review is the clinical application of AI in pathology and progress towards the ultimate goal of using this technology to improve pathology service and patient care.

RESEARCH AND CLINICAL APPLICATIONS OF AI IN PATHOLOGY

The application of AI in pathology generally fall into the following main categories (figure 1):

© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.
Diagnostic applications

AI approaches have been used in a variety of tasks such as object recognition, detection and segmentation. Current evidence indicates that computer vision algorithms may be capable of extracting several features from WSI to make a diagnostic prediction.22–27 For example, various diagnostic AI tools have been developed to make histological diagnosis equivalent to the average pathologist and provide information that can be challenging for pathologists to identify.23 25 27 Accurate assessment of quantitative features such as immunohistochemical biomarker assessment, counting cells and evaluating the degree of various tissue features such as spatial arrangement of cells, density of structures, pattern of distribution and architecture of the tissue are other important advantages of AI application.28 29 Meanwhile, AI tools can also help in the standardisation of the histological scoring criteria of morphological features that represent a continuum biological processes such as the Gleason score for prostatic carcinoma and breast cancer grading.30 31

Applications of AI in pathology diagnosis may also include rare and complex cases that are often challenging to general pathologists who rarely come across them. AI has the power of searching for images that are similar to query images from large histopathological databases, which can be defined as content-based image retrieval (CBIR). CBIR is extremely useful for the pathological diagnosis process by improving the probability of reaching a correct diagnosis for difficult cases by quickly retrieving similar cases from accessible pathological image databases containing such rare and complex cases.22 32 Here similarity reflects associated histopathological features rather than simple image similarity.

However, several factors need to be considered during the design, development and validation of the diagnostic AI tools to improve their adoption into clinical practice. Based on the aim of the diagnostic algorithms, these include the following:

A. Independent reporting algorithms: These include AI algorithms that can provide diagnosis without input from the pathologists. Examples include screening algorithms that can identify normal tissue (ie, colonic, gastric biopsies, prophylactic and reduction breast specimens and so on) and provide automated reports on such cases. Although this appears as the easiest task, caveats should be considered such as auxiliary tests in discrepant cases. The developers of these algorithms should consider the wide varieties of normal tissue during their developments, to avoid missing rare microlesions or focal lesions that may be missed during the screening process but have significant clinical implications. The majority of normal pathology samples are not exactly normal but part of a spectrum of morphology that lack significant pathological abnormalities and the appearances is often variable based on age, location and even the biopsy techniques.

B. Diagnosis-aided tools: These include algorithms that assess one of several features on the slides such as tumour grade, type and extent. Pathologists typically assess multiple features and combine all these features to reach a diagnosis. Developers of such single features diagnostic algorithms should pay a lot of attention to the way these algorithms will be integrated in the diagnostic workflow and the usability of such tools by the pathologists. The added value of such AI algorithms will be based on several variables, including the features to be assessed, the time the algorithms results will be provided and the usability of the integrated algorithms. For instance, the potential advantage of an AI algorithm to score breast cancer grading will not be primarily workflow improvements—experienced pathologists are capable of reading slides and reporting very efficiently—rather, the main added value will be in terms of objectivity and prognostic superiority when compared with an average interobserver variation of breast cancer grading. Although recent publications have demonstrated the potential of AI algorithms to perform equally or even superior to clinicians in various clinical tasks, it is important to consider the context of the intended clinical use case. For example, for lymph node metastasis detection, the AI can have suboptimal, but complementary operating points as pathologists, but superior performance when serving as a pathologist assistive tool.34

C. Automated quantification of specific features: Automated quantification of immunohistochemistry (IHC) staining intensity, receptors and other markers scoring in the breast and other tissue is attracting a lot of interest, and more recently, efforts have been made to predict breast cancer biomarkers directly from the H&E slides.33 However, these algorithms need to consider some items for clinical application such as added value in terms of cost saving and the level of impact of their added value of objectivity and accuracy of the AI-aided scoring. Scoring receptors and IHC prognostic and predictive biomarkers in breast cancer, for instance, takes around 1 to 2 min of pathologists’ time per marker routinely. Pitfalls of scoring of receptors are related to many preanalytical and analytical variables, and not only to scoring the positive cells in a given sample. Receptor assessment for therapeutic use should be limited to invasive tumour cells. In the breast, for instance, distinction between in situ and invasive carcinoma can be very challenging and IHC for myoepithelial markers might be needed to make such a distinction. Similar to pathologists, internal and external controls should be considered, and discrepancy between morphology and receptor status may be a clue to identify either false staining pattern or misdiagnosis of the morphological features on H&E. Some markers could be expressed in normal and malignant cells, such as oestrogen and progesterone receptors in the breast. Ki67 index is usually expressed in the inflammatory cells. Thus, the automated tools that used to estimate the proportion of positive cells within a WSI, must consider not only how to count, but, more importantly, what and where to count.

In a proportion of cases, pathologists make their diagnosis not only based on H&E findings but also with the aid of auxiliary tests most commonly IHC and the choice of IHC markers is
based on several variables, including judgement of the pathologist. Using a panel of markers, interpretation of the intensity, pattern and frequency of staining, location and subcellular localisation of expression, and negative expression of other markers are often used in combination together with H&E findings and other metadata to make a diagnosis. An AI algorithm making a diagnosis based on H&E only is less likely to work for such diagnostically challenging cases. AI tools that can predict IHC marker expression based on H&E findings look very promising but the diagnostic value of such information when it is used in such a context remains to be proven. In addition, if these tools are used to replace the existing companion diagnostics, the current data on the benefit of targeted therapy is based on expression of the relevant protein or specific genomic alterations as assessed using conventional methods. Application of these tools to the clinical practice will need sufficient evidence that the response to the specific targeted therapy as predicted by the AI tools is not inferior to that predicted by the conventional methods. Using AI tools to predict IHC marker expression from H&E findings for diagnostic purposes should also consider the fact that in routine practice, we use IHC in morphologically borderline cases rather than in cases with distinct features. Therefore, enrichment of these cases with borderline morphology in the training and validation of these algorithms will be needed if they are to be of useful clinical application.

Prognostic and predictive applications
One of the most important potential applications of AI in pathology is to predict patient prognosis and response to specific therapy based on the morphological features. While the prognostic value of some well-established pathological features, such as tumour grade and subtype, has been demonstrated, directly linking pathology images with a myriad of features including those of the tumour, surrounding microenvironment and genetic profiles with survival outcomes and treatment response for adjuvant/neoadjuvant therapy remains largely unexplored.

In fact, pathologists use a limited number of morphological findings on the tissue sections to determine tumour histological type and grade. This is not only based on the correlation with outcome, but also on their ability to identify these features using their visual identification capacity and skills and to achieve an acceptable level of concordance. Furthermore, integration of multiple factors, such as microenvironment patterns and various morphological patterns into a single prognostic score or index is quite challenging for human capacity. For pathologists to identify histological subtypes of invasive breast carcinoma for instance, it is important that they identify these morphologies as malignant lesions, of invasive nature and of breast origin regardless of their difference in behaviour. Association with outcome and response to therapy are typically evaluated following their histological identification using the defined set of features. However, image-based AI tools can provide a novel classification system that is based on clinical relevance according to outcome and response to therapy, and correlate a set of features out of a myriad of features related to tumour cells morphology, architecture, stroma and microenvironment, and presence of lymphovascular invasion, as well as correlate all these features to defined end points of clinical relevance such as probability of recurrence or metastasis and degree of response to specific therapy. Finally, all histological features combine into one classification system to reflect tumour nature and behaviour without the need to specify the different morphological features identified separately by pathologists. Those can be the important ways forward for patients’ management and precision medicine.

Integration with patient genomic and genetic profiles
A lot of attention is being paid to the ability of AI tools in pathology to relate morphological phenotypes to patient tumours genetic/genomic profiles, which is essential for understanding the biological mechanisms underpinning cancer development and for selecting targeted therapy. Identifying the association between morphological features and tumour genetic profiles or predicting the underlying molecular alteration based on the morphological features appears to be a successful and easy approach. However, integration of the large-scale genomic data such as next-generation sequencing (NGS) data remains a challenge and sufficient information on the value of the NGS massive data items, alone and when combined with other features, must be obtained before such algorithms can be used in patient diagnosis. The other challenges include the extremely high-dimensional data of both imaging and genomic data, and that the interaction between imaging and genomic features remains to be defined. Although integrating the imaging and molecular features can provide a comprehensive view of individual tumours, including how to develop, train and validate such models that are capable of tackling such sophisticated multidimensional data remains a challenge.

Pathology workflow efficiency
Improving the efficiency of pathology services remains a vital healthcare goal and an important step for timely and appropriate patient management. With increasing digitisation of cellular pathology service, the application of AI becomes possible and the potential of improving the efficiency of the service by using various AI tools becomes more important than ever before. AI applications could have the benefit for pathology services as follows: human error reduction in specimen handling and processing, faster turnaround times, workload management, quality control and quality assurance measures, automatic requests of relevant tests in certain cases and automatic reporting. To achieve these, pathology workflow AI tools will need to be integrated into existing laboratory information management systems (LIMS), and the DP image management system to enable integrated access to WSIs and the associated metadata for the relevant algorithms to work in a fully automated process.

Quality control of the slides currently is very important though it is a time-consuming step in the diagnostic pathways. AI tools can identify the spectrum of error on the scanned slides in terms of staining quality, tissue representation on the slide, processing errors, quality of staining of the external, and on slide control samples and even the quality of tissue fixation. For workflow management, AI can help in the triage of cases into high or low priority, sort and classify cases to ensure that urgent cases are always reviewed first, and check and slide identification matching. Other applications can include the use of diagnostic AI tools to reduce the need for double reporting of diagnostically subjective cases. The potential business advantage of DP is going to come mainly from efficiencies and AI is likely to be the main source of improving efficiency of the pathology services.

AI in pathology training and education
AI tools can help further training of the next generation of pathologists, provide automated annotations and other interaction functions to create a dynamic teaching environment for trainees. Additionally, diagnostic AI tools can be used to support
Review

primary reporting by pathology trainees and biomedical scientists. Such educational models will be complementary to the conventional educational processes provided by the specialists, at least in the early transitional period, and will add invaluable resources for knowledge. Integration of AI tools in the reporting workflow can provide trainees with additional information such as lists of differential diagnosis and potential auxiliary tests that can be requested, the level of difficulties and subjectivity of the diagnosis of the lesion and the relevant educational resources which potentially improve their training.

APPLYING AI IN PATHOLOGY CLINICAL PRACTICE

The intended clinical use case of the AI tools needs to be considered early in the process of development and validation. Therefore, it is critical for the development of clinically applicable and relevant AI applications to adopt a multidisciplinary approach with input from not only AI researchers and pathologists but also other stakeholders such as oncologists and surgeons. The negative impact of impractically high expectations of the technology on the clinical applicability and usage of the technology, and the balance between the performance and the technology with the real-life benefits should be considered. Despite the race by research groups and industry to produce and license their AI products, it should be advised that sustainability and clinical utility of such products be considered. Healthcare systems funding such AI-based projects should consider the clinical utility of the tools and not just their analytical performance. The clinical utility measures the overall benefits from the algorithms, including the costs, risks and added value compared with the existing practice. Algorithms are evaluated by comparing the algorithm’s predictions with a reference standard on a held-out set. Therefore, a holistic approach is required with consideration of the whole context such as the intended use case of the algorithms, the evaluated algorithm performance, the existing and future platforms and environment in which AI tools will be integrated and used, and the potential clinical utility.

As pathologists use certain evidence-based clinical and molecular data of known clinical values to make a diagnosis, it is expected that image-based AI tools would use the same well-defined clinical and genomic data to reach the same level of confidence in making a diagnosis as pathologists do. Randomised control studies are usually the best method to achieve such a level of confidence, but they are challenging to carry out in the diagnostic pathology field. Retrospective cohorts with high-quality data gathering could be used with caution to achieve reliable results. In summary, understanding the clinical relevance of the input data for AI algorithms is critical for its success. Care should be paid as the clinical impact of certain features when analysed alone may be different when integrated with other data, and the validity of each data item when used alone and in combination should be confirmed before its integration into any AI tool.

REGULATORY ISSUES AND AI APPLICATION IN ROUTINE PRACTICE

The Food and Drug Administration (FDA), the European Medicines Agency (EMA) and other regulatory bodies have created a regulatory framework for AI-based computer aided medical software and devices. The regulatory framework provides suggested methods for evaluating the performance of image-based AI algorithms to ensure the rigour and consistency in the reported metrics. One of the aims for these image-based AI tools is to provide digital assistance for pathologists to improve the diagnostic accuracy and efficiency. Other usages may include the prediction of outcome or predict a patient’s response to therapy. Therefore, the level of evidence required for approval should consider the appropriate context. Regardless of whether an AI diagnostic tool is approved or not, it must be validated in the laboratory before being placed into clinical use. However, the details of the validation study, such as the number, the spectrum and the complexity of specimens, the limits of acceptability, the clinical setting and relevance to the intended clinical use remain to be defined.

CHALLENGES AND OPPORTUNITIES

Despite the provisional success of AI image models in pathology and other healthcare disciplines, their implementation in routine practice has been slower than expected. One of the main reasons for this phenomenon is arguably the ‘black box’ nature of these models: a lack of understanding of why the algorithm makes such predictions. Machine learning models can be thought of as high-dimensional function approximations. Training a neural network in practice means fitting the model parameters to best associate a set of inputs with a corresponding set of outputs. Unless explicitly designed for interpretability, or probed after training with a variety of techniques, the model produced by this training process will not necessarily conform to human understanding. To increase the acceptability of AI applications in routine healthcare, attempts are currently being made by algorithm developers to incorporate inputs from the end users of the AI algorithms. These methods are expected to improve explainability and interpretability of AI applications, improve the perception of AI by its end users and provide a way for improvement based on clear factors and thereby increase their acceptability.

Other issues that need to be considered include the usability and added value of the proposed AI tools, and the potential disruption to the current workflow. Specifically, with respect to added value, it is important to note that, besides their capacity to combine complex image-based data with genomic, radiological and other clinical data, the number of diagnostic components measured by these algorithms is important to promote their use in routine clinical practice. For example, a primary invasive tumour detection tool that can also identify lymph node metastasis would significantly help reduce the time spent by pathologists. However, this is only possible if the algorithms assess all the features assessed by the pathologists in the lymph nodes. The algorithms must be able to diagnose other malignant and benign entities—metastatic carcinoma mimics—that may be incidental in the lymph nodes. The same argument applies to lesions in the primary tissue.

Another instance of potential added value of AI algorithms— or the absence of it—is in the grading of breast cancer and assessment of receptor expression status. Since grading and receptor status assessment are relatively stable and routine work is undertaken by the pathologists, the value added by the development of an algorithm solely for tumour grading or receptor assessment may not justify the cost and disruption of pathology workflow associated with the adoption of AI algorithms. Conversely, some prostate cancer detection tools which assist in reducing turnaround time, IHC requirements and reporting time, and provide added measurements such as Gleason score and tumour burden assessment will increase the usability of the tool. Therefore, AI tools can only add value to clinical practice if they are developed to function in the areas of most need such as in screening normal cases or in identifying cases that require double reporting, or in aspects in which multiple components
need to be assessed at once, rather than in the measurement of one or a few components. Such applications will help to provide the practicality and the best use of this technology.

The clinical relevance of the integration of the AI tools in the pathology workflow and patient pathways, as well as their cost-benefit ratio remain to be considered. Most pathology departments use different versions of LIMS, with or without integration or incorporation into the image viewing software. In fact, many departments, which currently use DP have independent imaging viewing software and LIMS and so need a middleware to combine both their functionalities. But the associated costs need to be considered in view of IT infrastructure and a large amount of data storage costs. Because WSIs takes hundreds of gigabytes of space, long-term storage cost is massive and comprise a burden on most hospitals. At the same time, current rules mandate the long-term storage of glass slides which doubles the cost of storage of the diagnostic materials. Worldwide, hospitals are considering the cost of improving the IT infrastructure to meet the requirement of DP and the added costs of long-term massive storage, and changing the LIMS increases the challenge. Moreover, for the multimodal use case, AI systems will need to access radiology images, WSIs images, genomic and in situ hybridisation data, this leads to additional complication in integrating these independent data sources.

CONCLUSION
There has been a sharp rise in the development and application of AI tools, including image-based algorithms for the use in pathology service, and it is expected to dominate the field of pathology in the coming years. The deployment of computational pathology and application of pathology-related AI tools can be considered a paradigm shift that will change the way pathology services are managed and make them not only more efficient but also capable of meeting the needs of this era of precision medicine. Development of pathology-based AI tools needs input from a multidisciplinary team in which pathologists and users should have great input to improve the adoption of these technology-driven applications. A combination of AI and pathologists can yield results that are more accurate, consistent, timely and useful beyond a human’s ability. AI can provide analytical tools to streamline the complex, multitstep pathology case life cycle in pathology laboratories, from accession to archiving. This can provide not only workflow automation but also analytics dashboard and data repository that can improve efficiency by self-learning from previous experience and help understand laboratory productivity, quality and efficiency, in addition to helping to allocate future resources to areas in more need. AI can also improve streamlining the whole process by aligning the laboratory technical components of the case pathway with the pathologists reporting components. Improving the efficiency of pathology service workflow, trainee and junior pathologists reporting, timely reporting by pathologists, cost-efficient diagnostic, prognostic/predictive algorithms and production of multidimensional output of pathology reports, and combining with image and genomic/genetic data are some of the expected benefits of AI technology application in routine practice. AI applications will also lead to an advanced diagnostic, enabling researchers and clinical teams to share knowledge and use computational algorithms to assess and contribute valuable insights that can ultimately lead to a more informed and detailed pathology diagnosis. This integration will help advance the future of precision oncology and can result in personalised care plans.

Correction notice This article has been corrected since it was published Online First. Author names and affiliations have been corrected.

Handling editor Runjan Chetty.

Acknowledgements We would like to thank Dr Henry Ebili for his kind contribution to this review.

Contributors All authors reviewed the paper.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer-reviewed.

ORCID iDs
Emad A Rakha http://orcid.org/0000-0002-5009-5525
Michael Toss http://orcid.org/0000-0002-9077-3984

REFERENCES