Title: Artificial intelligence in digital pathology: *A roadmap to routine use in clinical practice.*

Short Running Title: Artificial intelligence, a roadmap to clinical use.

Authors: Richard Colling[1], Helen Pitman[2], Karin Oien[3], Nasir Rajpoot[4], Philip Macklin[5], CM-Path AI in Histopathology Working Group[†], David Snead[6*], Tony Sackville[7*], Clare Verrill8[*]

[1]Nuffield Division of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK OX3 9DU

[2]National Cancer Research Institute, Angel Building, 407 St John Street London UK EC1V 4AD

[3]Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Glasgow UK G61 1QH

[4]Department of Computer Science, University of Warwick, Coventry CV4 7AL

[5]Nuffield Department of Medicine, University of Oxford, NDM Research Building, Old Road Campus, Oxford UK OX3 7FZ

[6]PathLAKE (Director) and Histopathology, University Hospitals Coventry and Warwickshire NHS Trust, University Hospital, Clifford Bridge Road, Coventry UK CV2 2DX

[7]British In Vitro Diagnostics Association, 299 Oxford Street, London UK W1C 2DZ

[8]PathLAKE (Principle Investigator), Nuffield Department of Surgical Sciences and Oxford NIHR Biomedical Research Centre, University of Oxford, John Radcliffe Hospital, Oxford, UK OX3 9DU

*Joint senior authors contributed equally

[†]please see end of manuscript for full list of members

Correspondence to: R Colling richard.colling@ndcls.ox.ac.uk

Conflict of Interest: None

Word Count: 4391

Abstract

The use of artificial intelligence will likely transform clinical practice over the next decade and the early impact of this will likely be the integration of image analysis and machine learning into routine histopathology. In the UK and around the world, a digital revolution is transforming the reporting practice of diagnostic histopathology and this has sparked a proliferation of image analysis software tools. While this is an exciting development that could discover novel predictive clinical information and potentially address international pathology work-force shortages, there is a clear need for a robust and evidence-based framework in which to develop these new tools in a collaborative manner that meets regulatory approval. With these issues in mind, the NCRI Cellular Molecular Pathology (CM-Path) initiative and the British in vitro Diagnostics Association (BIVDA) has set out a roadmap to help academia, industry and clinicians develop new software tools to the point of approved clinical use.

Key Words: digital, pathology, image, analysis, artificial, intelligence, evidence-based

Introduction

The integration of artificial intelligence (AI) will be one of the biggest transformations for medicine in the next decade and histopathology is right at the centre of this revolution. The value, both for medical practice and creating business and wealth from AI has been recognised across the world and in particular by the UK Government who published an Industrial Life Sciences Strategy in August 2017 [[1]. Histopathology was highlighted in the report as *''being ripe for innovation"* and *''where modern tools should allow digital images to replace the manual approach based on microscopy"* in addition to *''the opportunity to create AI-based algorithms that could provide grading of tumours and prognostic insights that are not currently available through conventional methodology".*

Much of the workflow of histopathology departments remains largely unchanged for decades, although some processes can be automated e.g. immunohistochemistry and more recently routine molecular testing has been incorporated for some disease types. The adoption of digital pathology (DP) technologies to replace microscopy has been slow and adoption of the use of image analysis/AI tools to augment the workflow or solve capacity issues is limited. Algorithms have the potential to either perform routine tasks which are currently undertaken by pathologists, or provide new insights into disease, which are not possible by a human observer [2].

Innovate UK recently awarded £50M to create five new centres of excellence for DP and imaging using AI medical advances [3]. The centres will aim to realise the benefits of AI in pathology by speeding up diagnosis, improving outcomes, providing better value for money and allowing clinicians to spend time on other tasks. The vision is a

healthcare service which transforms the NHS into an ecosystem of enterprise and innovation that allows technology to flourish and evolve. Two of the five centres focus entirely on DP AI, with a third centre focussing on imaging and DP. These new DP centres are known as PathLAKE, a DP consortium led by University Hospitals Coventry and Warwickshire NHS Trust and also including Oxford, Belfast and Nottingham, the Leeds-led Northern Pathology Imaging Co-operative (NPIC) and the pan-Scottish iCAIRD (Industrial Centre for AI Research in Digital Diagnostics). Each centre was awarded funding in partnership with industry, who will make significant in-kind investments.

A small number of approved image analysis tools exist, e.g. oestrogen receptor status, but their use is not widespread. The barriers to uptake are multifactorial, but uncertainty around the accreditation is a significant contributor. In the UK, for example, laboratories are strongly encouraged to be assessed by the UK Accreditation Service (UKAS) to establish competence in applied-for activities, assessed against ISO 15189 (2012) [4]. AI tools should be no different. Although quantification tools may assist pathologists and reduce the subjectivity of human observers, the notion that AI will replace the need for pathologists to make even simple interpretative judgements is one that the pathology community struggles with [5]. It is likely that outputs generated by such tools will increase the complexity of the information that needs to be assimilated into integrated diagnostic reports as part of a modern precision medicine driven approach with pathology forming part of the “big data’’ set [6].

The first major step in adopting DP is the introduction of digital whole slide imaging (WSI) into routine practice. This is now well evidence-based and will provide the

infrastructure and initial datasets for building AI tools [7-9]. With departments now beginning to make the digital transition [10], and in the context of current and near future predicted shortages of pathology staff [11,12], the opportunity for computer-aided diagnosis (CAD) will almost certainly become the real focus of DP research over the next 10 to 15 years.

With this in mind, in June 2018 the NCRI Cellular Molecular Pathology Initiative (CM-Path) [13] joined forces with the British In Vitro Diagnostics Association (BIVDA) [14] and organised a workshop with academic, clinical, regulatory and industry leaders to look at the use of AI in a clinical histopathology environment. The aim was to understand the path from tool concept, through development to full roll-out in a routine histopathology workflow, understanding the roadmap and the challenges at each stage. The objective was to understand why such tools have had limited uptake thus far, in order to understand the barriers before a larger number of products hit the market. Understanding the process involved in clinical adoption from concept through to clinical practice will enable more confidence in understanding of the steps necessary to support appropriate adoption. The different groups present, reflected the differing expertise needed to achieve this, with pathologists often holding the clinical expertise and cohorts with industry the market expertise. The group was completed by regulators and accreditors. Here, we report the output from the workshop, present our road map (Figure 1) for developing new tools and outline the components needed in AI tool development (Table 1) for clinical use.

Potential Applications

The potential applications of AI in DP are wide ranging, but the focus of interest now is largely based around digital image analysis (DIA). Established image analysis involves a combination of manual or computer aided image processing techniques (such as colour correction, filtering and other basic manipulation methods) and user-driven feature classification and extraction (e.g. edge detection, pixel intensity thresholding, mathematical transformations) based on pre-defined parameters. Newer methodologies, often termed artificial intelligence (AI), are based on machine-learning algorithms, whereby an automated computer program runs the image analysis and uses various statistical methods to model the output data to progressively fit ('learn') to some defined outcome of interest. For example, this could be the likelihood that a specific diagnosis is present in the image, or the likelihood that the tumour in an image will respond to chemotherapy. An AI program can be 'trained' with example images (supervised learning) or the software can be allowed to discover key features that fit the outcome for itself (unsupervised learning). In either case, AI tools can be user directed (run on demand by pathologists or laboratory staff) or can be completely automated and the extent of interaction with an AI tool by the pathologist can vary from the user deciding to run a program and evaluating the quality of the output, to simply reporting the output from an automated analysis that has run in the background. Practical applications may include immunohistochemistry (IHC) biomarker detection and scoring (for example, Her-2 and Ki67 tools are already available with many other markers in development), disease quantification, morphometrics, tumour detection and cancer grading, and rare event screening (e.g. highlighting samples where tumour or micrometastases are detected and need pathologist review, and those which are negative and may not need review) [15-19].

Concept Development

The first step for DP is the transition from traditional microscopy to digital slides. The first stage of creating any new AI application (often called ‘app’ or ‘tool’) however is concept development: identifying the clinical need and defining the potential solution. Currently, ideas for new tools come from a variety of interested parties including industry (biotechnology companies, drug company companion diagnostics), academia (academic pathologists, computer scientists, engineers), practicing histopathologists and clinical staff (e.g. oncologists) – many of whom are working on similar projects and often repeating work being done elsewhere (see Table 1). This is the first major challenge – definition of the clinical need and who should be making those decisions and setting priorities around algorithm building. Industry and academia often have different perspectives on what tools should be developed as different measures of success are applied – typically a successful commercial product in industry versus grant funding and academic publications. Although most companies solicit specialist advice to guide the direction of suitable potential candidate applications for development, companies are often pulled in other directions by existing technology preferences and platforms, access to technical expertise and resources, and IP in the form of patents, technology, know-how, market positioning etc. They are likely to prefer to use proprietary technologies at the early stages of development as this is seen as the most protectable route to a return on their investment. This may result in a disconnect between what is launched commercially and what is actually required by the end users of the products in the delivery of the clinical services they provide. In the UK, the newly formed network of national AI centres of excellence is expected to be pivotal in them bringing the diverse groups of health and academic institutions, entrepreneurs and commerce together.

Ethics and Funding

AI tool development must consider the need for Research and Ethics Council (REC) approval, which is generally required in the research and trial stages. Developers have to comply with the ethics of using patient data for research development, commercial gain and return for the NHS. Mindful of the value of patient data for research and the challenges of obtaining consent for its use the NHS is establishing the National NHS Opt Out Scheme to provide individual patients with some control over what purposes their data is used for. Individual institutions may have in addition local procedures for allowing opt out of the use of their data for research and it is important that all of these factors are understood and followed in the design stage of AI tool development. There are many parallels to be drawn with the therapeutics pipeline; whilst successful products will pass through the entire pathway, most likely supported by sequential funding rounds from differing sources, many products are bound to fail at some point. Measurable outcomes of success are important in enabling rational decisions over which products should be supported, and this is relevant to each stage of the pathway, up to and including justification of the tool for review and being recommended for use in clinical guidelines, e.g. by the National Institute for Health and Care Excellence (NICE) in the UK. This typically requires evidence of financial or resource savings, improvements in quality, patient impact, and is thus often difficult to prove, particularly when the solution involves significant transformation, workflow redesign and financial investment.

Development

Once an idea has been conceived and collaboration established, the cycle of tool development is a helpful model to understand the process of creating the software. This includes defining pre-processing steps (defining the output needed, designing the algorithm to obtain this), the analysis stage (pilot or larger follow-up sample), data analytics (collection, organisation, storage and processing of raw data, statistical analysis of comparison data). This will inevitably require several cycles of trial and error to get the tool working well and refining the methodology; this process could be thought of being akin to the pre-trial early drug development. There is often a pilot stage trial to ascertain if the tool is likely to be of clinical use and there may be some overlap with early development and later validation steps.

Validation and Regulation

The introduction of any new test requires an evidence-based approach to validation and this forms a key component of regulation. The new in vitro device regulation (IVDR) requirements set out very specific and detailed guidance on validation and we summarised our recommendations for a number of key components of validation in Table 1. In laboratory medicine there is usually a distinction between a technical or analytical validation (the test measures exactly what it is supposed to measure, evaluated usually on a deliberately selected population of cases) and a clinical evaluation (the test performs well in routine clinical practice, evaluated ideally on an unselected and unbiased population of patients) [20], This part of the process could be thought of as similar to Phase I (analytical validation) and Phase II/III (clinical

validation) drug development. Measures of laboratory and clinical validation should be established for any new (index) test against a current gold standard (reference) test.

In image analysis, an analytical (Phase I) validation is often achieved by comparing a tool with so called ‘ground truth’, for example comparing an AI tool count for Ki67 positive cells on several idealised images compared with a very detailed cell count made manually acting as a gold standard. Comparison of any DP technology or technique will need to be compared with the performance of Human Pathologists with their inherent irreproducibility and day to day performance variation. Defining ground truth in this situation is inherently difficult and requires careful study design and an acceptance of the weaknesses of the current gold standard reference method. The end result must produce a final dataset which can be used to demonstrate (for regulatory approval and accreditation) the validity of the app. A clinical (Phase II/III) validation involves higher level trials in large patient unselected and blinded datasets. An example could be comparing the performance of a Ki67 tool with pathologists in assigning a grade to all neuroendocrine tumours that come through a department over a set period of time.

It is likely that for many AI tools it will be difficult to obtain ground truth and there may not be any comparable (gold standard) test currently in use by pathologists. In this scenario, the validation will primarily be a clinical one and hinge on robust and reproducible validations in large patient cohorts with detailed outcome data. One of the most pressing issues is the relative lack of such required cohorts for validation. In those that exist with mature data, logistical challenges of getting slides scanned are often prohibitive. Those who have access to such cohorts are often unwilling to share.

Pathologists assessment with an optical microscope is often considered to represent the ground truth, and this is a controversial assumption. Interobserver variability and subjectivity mean that the observations and annotations of one pathologist should not necessarily be considered ground truth. This is especially true when one is building tools where the ground truth is subjective e.g. Gleason grading of prostate cancer [21]. Validation and testing by multiple pathologists and in multiple laboratories are usually required.

Bringing AI algorithms into diagnostic practice creates interesting new challenges around the legal implications of a pathologist signing out a report using AI. The pathologist would be required to be confident in the output of the algorithm in order to integrate it into the main report and any algorithms used would need to have been through appropriate validation and verification. The need for pathologists to build trust in new digital systems which may be seen as opaque or “black box” technologies could put a natural but important brake on the speed of adoption of AI in digital pathology. This could act as a focus for closer collaboration between the industry and end users to deliver robust applications that pathologists are happy to rely on when preparing and signing out their reports. The fact that AI researchers are now beginning to focus on (a) providing confidence estimates with their predictions/results and (b) localising pathology-related features should help with allaying concerns about interpretability and building trust. Besides, there is also need for regulatory processes to learn from the experience of medical imaging community in evaluating the performance of algorithms for various challenge contests [22]. The future educational needs of the pathology community will change, bringing a need for at least a basic working knowledge of how such algorithms function with some pathologists taking on a more advanced ‘’computational pathologist’’ role. Similar to many other diagnostic platforms

(e.g. molecular diagnostics assays), we suggest that any new AI tool would fall under the European Medical Devices Regulation 2002 [23] and are probably best regarded as in vitro diagnostic devices (IVD). In the UK currently, the competent authority for medical device regulation is the Medicines and Healthcare Products Regulatory Agency (MHRA) and, like elsewhere in the European Economic Area, devices must be approved via the conformité Europeéne – in vitro diagnostic device (CE-IVD) legislative process (IVD Directive 98/79/EC). For most devices (including WSI imaging systems), this has until recently been via the self-certification route. However, there is currently a transition phase to the new In-Vitro Diagnostic Medical Devices Regulation (2017/746) (IVDR) 11a. Under the new regulations devices are given a risk classification (Class A-D), with WSI imaging systems deemed Class C. The IVDR sets out a new pathway for certification that will be carried out by approved Notified Bodies [24-26]. It is likely that the regulatory changes will continue to apply in UK, after its withdrawal from the European Union (EU). The impact of these new regulatory changes on development of AI tools is uncertain at this stage, but we recommend that all AI tools should undergo CE-IVD marking. This will require additional clinical evidence, rigour and assessment by Notified Bodies in addition to existing requirements for conformity, including situations where machine learning technology is used, and where self-learning systems result in modification to algorithms and data analysis workflows that are different from what was originally submitted to gain the accreditation in the first place.

In the US, medical devices are classified based on likely patient risk (Class I-III). Class II & III devices (~60% of devices) are required to undergo Premarket Approval (PMA) unless there is a specific exemption such as the Humanitarian Device Exemption or approval under the Premarket Notification [(510(k)] route for devices which are similar

to existing PMA approved devices [7,27]. Previously, the FDA classified WSI imaging systems as Class III however in 2017 the FDA classified the Philips IntelliSite Pathology Solution (and concurrently by default classified all generic WSI systems) as a Class II device (although with special controls) and granted permission for the system to be marketed via the [510(k)] route [28]. The route to marketing approval in the US may change however. The FDA is piloting a new streamlined approval route specifically for digital health products, known as the Software Precertification (Pre-Cert) Pilot Program. This route would presumably include diagnostic image analysis software and AI-based technologies [29].

An additional consideration is the use of in-house lab developed methods and tools (often called Lab Developed Tests) which in Europe are currently governed and controlled under 'Health Institution Exemption' to the IVD Directive 11d. These will be subject to the new in vitro diagnostic medical device regulation (2017/746) and the new medical device regulation (2017/745), in particular, the provisions of Article 5(5) of both IVDR and medical devices regulations (MDR). Application of the exemption are currently the subject of a consultation exercise by MHRA 11b. Health Institutions making or modifying and using a medical device or IVD can be exempt from some of the provisions of the regulations provided products meet the relevant General Safety and Performance Requirements. Health institutions will need to have an appropriate quality management system in place, a justification for applying the exemption and technical documentation in place. Some of this information will need to be publicly available.

The development of clinical AI tools by individual institutions will need to conform to any new regulations, even if only intended for use within their own institutions. However, the benefits and opportunities afforded by DP based systems, on which AI

tools depend and run, largely arise from the ability to use them in collaborative professional networks over wide areas and between institutions. In pathology, the professional norm of collaborating on cases and seeking second opinions will increasingly require AI tools to be used in a standardised way between institutions, and will require either exemptions to the legislation, or conformance to it that is consistent with the emerging DP enabled infrastructure.

The variability in performance of in-house developed tests is cited as one of the main reasons for limiting their use to intra-institution application, and to the move to requiring their accreditation and conformance to the new legislation. Tools labelled purely for research projects with no medical purpose can be considered for Research Use Only (RUO) and exempt from the IVD Directive 11c (devices for performance evaluation are subject to the regulation set out above) [30,31].

Regulatory advice can be sought from authorities. In the US, this would be the Food and Drug Administration (FDA), in the UK this would be the MHRA. The latter recommend initial informal enquiries to regulators MHRA can be made via email (Innovationoffice@mhra.gov.uk or Devices.Regulatory@mhra.gov.uk). The MHRA publishes a variety of guidance documents [25,26], including on medical devices, and offers a scientific advice service in the context of medicines development. In addition, the Innovation Office provides a free single point of access to expert regulatory information, advice and guidance that helps organisations of all backgrounds and sizes develop innovative technologies.

Implementation

Implementation involves two main areas of focus: test introduction and accreditation. To introduce a new test there needs to a be a clinical need, review of the market, review of the literature evidence and writing a business case to fund it via healthcare budgets. In the case of in-house developed tests, much of this work should have been done but when buying in a new CE-IVD marked test, this can be a big undertaking. Once a test has been commissioned for use, adhering to accreditation requirements for any new tool providing data used in clinical reporting would be encouraged (for both in-house and regulatory approved tests). In the UK, this process would be provided by the United Kingdom Accreditation Service (UKAS), meeting the requirements of ISO 15189:2012 [32]. All diagnostic laboratory staff will be familiar with the usual processes of this (see Figure 1) that include Standard Operating Procedure (SOP) documentation, test verification (checking a previously validated test is working correctly in your lab by running on a set of known cases), documentation, audit cycle, calibration records, non-conformity handling, error and adverse event reporting, staff training and participating in External Quality Control (EQA) via a scheme such as the UK National EQA Scheme (NEQAS). Any in-house modifications to the tool (adjusting user preferences, algorithm tweaks, change of computer equipment and screens, change of slide scanners etc.) require each step of the accreditation process to be updated and may need to meet the requirements of the IVDR health institution exemption. An immediately obvious issue is the need for EQA scheme, which currently do not exist, to be up and running – however plans to start such a scheme are underway.

It is beyond the scope of this paper to outline all the working issues of digital pathology and this is well covered by others, [15,33] but clearly a major step in the implementation of any AI tool in histopathology is the digitization of pathology

departments to begin with and until this happens it is unlikely that AI tools will be widely adopted. Although this transition will take some time, AI tools could be adopted in limited circumstances in the meantime, with individual cases scanned where needed. The challenges of course will include issues around financing scanners and software and long-term data storage is a problem. The RCPath recommends storage of images for at least two laboratory inspection cycles [33] and this requires many terabytes of data – often the biggest cost of digitisation a department will face.

A further major challenge for AI tool development and implementation is platform variety, integration and interoperability. In echoes of the early days of immunohistochemistry and molecular diagnostics, is the emergence of multiple parallel and competing platforms and methodologies, often based on proprietary technologies and vendor specific workflows. The health service sector conversely requires measurable reliability and interoperability, to enable for example running an AI tool from one vendor on another vendor's platform, and on samples processed in separate laboratories. All of these requirements need to be clearly understood and addressed in the regulatory process to deliver a useable and standardised routine workflow in the laboratory framework. An essential issue is data compatibility and a standard, universal file format (that maintains functionality for legacy data) for digital WSI has yet to be practically implemented. Although many manufacturers claim that their systems are open to other vendors' file formats, progress is slow and in practice there remain many difficulties. Many are now working towards a pathology version of the DICOM (Digital Imaging and Communications in Medicine) format and once agreed this will need to cope with the adaptations and advancements delivered by technological progression.

Impact on work force

The introduction of new technology and tests into clinical practice has an impact on the laboratory workflow and the staff (laboratory and pathologist) training. As discussed earlier, compliance with UKAS accreditation will require laboratories to amend their scope of practice, and assess any tool prior to implementation, measuring the observed performance against what is expected (verification). The Innovate UK initiative to build a network of UK AI centres will provide an important network of well-resourced laboratories which will be able to offer leadership and exemplar practices for this sector over the coming years.

Less obvious but no less important is the effect of AI on pathologists and technicians using the technology in practice. There is an opportunity for pathologists in particular to come to rely too heavily on AI support leading to a degradation of diagnostic ability. Individual departments will need to understand how the implementation of such tools affects pathologists daily practice in order to understand these risks and provide support and assessment to protect and monitor their competence to guard against any atrophy of diagnostic skills. The UK Royal College of Pathologists (RCPath) have produced guidance on DP in clinical practice [17] but this does not cover the used of CADs. Additional work is required to address this emerging gap, which also needs to be factored into pathologists' training.

Conclusions

Much of what is discussed here is a distillation of the experiences of those who have come from varied background and have been involved in isolated parts of the road map. By coming together at the workshop in June, as a group we were able to consolidate these ideas and formulate our road map for developing AI software applications for use in histopathology practice. We feel strongly that a UK-wide strategy should be urgently developed for AI and DP. This technology really offers a chance to transform histopathology practice in the face of the extremely challenging problems the profession is facing. With proper slide image management software, integrated reporting systems, improved scanning speeds and high-quality images, DP systems will provide time and cost saving benefits over the traditional microscope approach and improve the age-old problem of inter-observer variation. Real and significant barriers to this are the introduction of tools without the proper regulatory-driven, evidence-based validation, the resistance of developers (academic and industry) not to collaborate and the need for commercial integration and open-source data formats.

Ethics: ethical approval was not applicable for this work.

Acknowledgements: CV is part funded by the Oxford NIHR Biomedical Research Centre. CV, DS and NR are part of the PathLAKE digital pathology consortium. These new Centres are supported by a £50m investment from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI).

Statement of Author Contributors:

This work was produced by the NCRI CM-Path initiative's Technology and Informatics (TI) workstream's (Digital Pathology sub-stream) in collaboration with the British in vitro Diagnostics Association. The article crystallises the consensus of academics, industry leaders and clinicians (the CM-Path AI in Histopathology Working Group) following a workshop in 2018. R Colling is a Clinical Lecturer in Histopathology at Oxford University and member of the TI workstream who led the collating of consensus opinion at the workshop and led the writing of the article, including reviewing the relevant literature. H Pitman is the CM-Path TI programme manager and assisted with the overall editing of the article and collating consensus opinion. K Oien is a Reader at Glasgow University with an interest in oncological predictive algorithm development and directly helped develop several technical aspects of the article. N Rajpoot is Professor in Computational Pathology at the University of Warwick and helped develop computational aspects of the article. P Macklin is a DPhil (PhD) student at Oxford University and assisted with editing the manuscript. The original roadmap project was conceived by C Verrill, D Snead and T Sackville and all three equally contributed to the editing and direction of the article. D Snead is Professor in Histopathology at University Hospitals Coventry and Warwickshire NHS Trust and the PathLAKE lead. T Sackville is an independent IVD industry consultant and chairs the digital pathology working group at BIVDA. C Verrill is Associate Professor in Cellular Pathology at the University of Oxford and is the CM-Path TI workstream lead. C Verrill, T Sackville and D Snead are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

CM-Path AI in Histopathology Working Group

Velicia Bachtiar

Richard Booth

Alyson Bryant

Joshua Bull

Jonathan Bury

Fiona Carragher

Richard Colling

Graeme Collins

Clare Craig

Maria Freitas da Silva

Daniel Gosling

Jaco Jacobs

Lena Kajland-Wilén

Johanna Karling

Darragh Lawler

Stephen Lee

Philip Macklin

Keith Miller

Guy Mozolowski

Richard Nicholson

Daniel O'Connor

Mikkel Rahbek

Nasir Rajpoot

Alan Sumner

Dirk Vossen

Kieron White

Charlotte Wing

Corrina Wright

References

1. GOV.UK. Life sciences: industrial strategy Crown Copyright. [Accessed February 5 2019]: Available from: https://www.gov.uk/government/publications/life-sciences-industrial-strategy
2. Bychkov D, Linder N, Turkki R*, et al.* Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018; **8**: 3395.
3. GOV.UK. Artificial Intelligence to help save lives at five new technology centres [Accessed February 14 2019]: Available from: https://www.gov.uk/government/news/artificial-intelligence-to-help-save-lives-at-five-new-technology-centres
4. UKAS. Medical Laboratory accreditation (ISO 15189) [Accessed February 5 2019]: Available from: https://www.ukas.com/services/accreditation-services/medical-laboratory-accreditation-iso-15189/
5. van Laak J RN, Vossen D. The Promise Of Computational Pathology: Part 1 [Accessed February 5 2019]: Available from: https://thepathologist.com/inside-the-lab/the-promise-of-computational-pathology-part-1
6. Verrill C. Floppy Disks to Diagnostics thepathologist.com. [Accessed February 14 2019]: Available from: https://thepathologist.com/inside-the-lab/floppy-disks-to-diagnostics
7. Abels E, Pantanowitz L. Current State of the Regulatory Trajectory for Whole Slide Imaging Devices in the USA. *J Pathol Inform* 2017; **8**: 23.
8. Lee JJ, Jedrych J, Pantanowitz L*, et al.* Validation of Digital Pathology for Primary Histopathological Diagnosis of Routine, Inflammatory Dermatopathology Cases. *Am J Dermatopathol* 2018; **40**: 17-23.
9. Snead DR, Tsang YW, Meskiri A*, et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016; **68**: 1063-1072.
10. Williams BJ, Lee J, Oien KA*, et al.* Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. *J Clin Pathol* 2018; **71**: 463-466.
11. Medical Research Council. Molecular Pathology Review [Accessed January 21st 2016,]: Available from: http://www.mrc.ac.uk/documents/pdf/mrc-molecular-pathology-review
12. CRUK. UK's pathology services at tipping point [Accessed June 27 2018]: Available from: http://www.cancerresearchuk.org/about-us/cancer-news/press-release/2016-11-23-uks-pathology-services-at-tipping-point
13. CM-Path N. What is CM-Path? [Accessed February 14 2019]: Available from: https://cmpath.ncri.org.uk/about/
14. BIVDA. [Accessed February 14 2019]: Available from: https://www.bivda.org.uk/About-BIVDA
15. Jon G, Darren T. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology* 2017; **70**: 134-145.
16. Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. *Comp Struct Biotechol J* 2018; **16**: 34-42.
17. RCPath. Diagnostic digital pathology strategy [Accessed April 18 2019]: Available from: https://www.rcpath.org/asset/2248BB71-B773-4693-945BFFDA593F2F2F/
18. Awan R, Sirinukunwattana K, Epstein D*, et al.* Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. *Sci Rep* 2017; **7**: 16852.
19. Sirinukunwattana K, Snead D, Epstein D*, et al.* Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. *Sci Rep* 2018; **8**: 13692.
20. Mattocks CJ, Morris MA, Matthijs G*, et al.* A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur J Hum Genet* 2010; **18**: 1276-1288.
21. Robinson M, James J, Thomas G*, et al.* Quality assurance guidance for scoring and reporting for pathologists and laboratories undertaking clinical trial work. *J Pathol Clin Res* 2018.

22. Maier-Hein L, Eisenmann M, Reinke A*, et al.* Why rankings of biomedical image analysis competitions should be interpreted with care. *Nat Commun* 2018; **9**: 5217.
23. Copyright C. The Medical Devices Regulations 2002 [Accessed June 27 2018]: Available from: http://www.legislation.gov.uk/uksi/2002/618/contents/made
24. García-Rojo M DD, Muriel-Cueto P, Atienza-Cuevas L, Domínguez-Gómez M, Bueno G. New European union regulations related to whole slide image scanners and image analysis software. *J Pathol Inform* 2019; **10**.
25. GOV.UK. Medicines and Healthcare products Regulatory Agency [Accessed February 14 2019]: Available from: https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency
26. GOV.UK. Medical devices: EU regulations for MDR and IVDR [Accessed February 5 2019]: Available from: https://www.gov.uk/guidance/medical-devices-eu-regulations-for-mdr-and-ivdr
27. HHS. Consumers (Medical Devices) [Accessed February 14 2019]: Available from: https://www.fda.gov/medicaldevices/resourcesforyou/consumers/default.htm
28. HHS. FDA Approval [Accessed February 14 2019]: Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf16/den160056.pdf
29. FDA. Digital Health Software Precertification (Pre-Cert) Program [Accessed February 14 2019]: Available from: https://www.fda.gov/medicaldevices/digitalhealth/digitalhealthprecertprogram/default.htm
30. Enzmann H, Meyer R, Broich K. The new EU regulation on in vitro diagnostics: potential issues at the interface of medicines and companion diagnostics. *Biomark Med* 2016; **10**: 1261-1268.
31. Favaloro EJ, Plebani M, Lippi G. Regulation of in vitro diagnostics (IVDs) for use in clinical diagnostic laboratories: towards the light or dark in clinical laboratory testing? *Clin Chem Lab Med* 2011; **49**: 1965-1973.
32. UKAS. Our Role [Accessed June 27 2018]: Available from: https://www.ukas.com/about/our-role/
33. RCPath. Best pratice recommendations for implementing digital pathology [Accessed April 18 2019]: Available from: https://www.rcpath.org/uploads/assets/uploaded/d6b14330-a8b9-4f5e-bbe443f0d56de24a.pdf

Table 1. The various tasks that we recommend need to be completed when developing and using an AI tool in clinical practice. Regulatory approval in the UK is managed by the Medicines and Healthcare Products Regulatory Authority (MHRA), in Europe this is done via conformité Europeéne - in vitro diagnostic device (CE marking) licensing, and in the US regulation is handled by the Food and Drug Administration (FDA). There are new UK regulatory requirements required for IVDR approval – for a more detailed description of these, please refer to MRHA publications [25,26], In the UK, accreditation is regulated by the UK Accreditation Service (UKAS) and management guidelines are compiled by the National Institute for Healthcare Excellence (NICE). *PPV=positive predictive value, NPV=negative predictive value.*

Development (Design stage)	**Analytical Performance (Phase I)**	**Clinical Performance (Phase II/III)**	**Clinical Practice (Post-marketing)**
Identifying clinical need	Determining testing protocol and specimen handling	Diagnostic accuracy (sensitivity and specificity, PPV, NPV, likelihood ratios, expected values in normal and affected populations)	Obtaining regulatory approval
Literature review and status quo	Establishing markers of test performance (analytical sensitivity, specificity, trueness (bias), precision (repeatability and reproducibility), accuracy (resulting from trueness and precision), limits of detection and quantitation, measuring range, linearity, cut-offs)	Diagnostic reproducibility	National management guideline approval
Research the market for existing solutions (also required for health institute exemption)		Comparisons with gold standards	Compliance with accreditation
Scientific rationale for new test methodology (sound basic science/mechanistic approach), establishing scientific validity		Prognostic studies (survival analyses, Kaplan-Meier plots, odd ratios)	On-going audit cycle of performance and review of clinical experience of new devices
Collaborative approach and multidisciplinary input		Assessing the significance of potential clinical benefits / losses	On-going EQA or equivalent independent measure of performance
Obtaining funding and skills to support work		Practicalities of using in clinical setting	Business case for on-going funding
Ethics approval		Health economics assessments	
Prototype production			
Pilot trial and error, design refinement			

Figure 1. The digital pathology AI development ‘road map’. This diagram describes the recommended steps in the development of AI and other digital pathology tools for use in laboratories. The order of events is given as a guide only and in some circumstances flexibility will be needed. In the UK, accreditation is regulated by the UK Accreditation Service (UKAS) and management guidelines are compiled by the National Institute for Health and Care Excellence (NICE). Regulators in the UK are the Medicines and Healthcare Products Regulatory Agency (MHRA), in Europe this is via conformité Europeéne - in vitro diagnostic device (CE marking) licensing, and in the US, regulation is handled by the Food and Drug Administration (FDA). *PPV=positive predictive value, NPV=negative predictive value, EQA=external quality control.*