The need for measurement science in digital pathology

Marina Romanchikova [a,*], Spencer Angus Thomas [a], Alex Dexter [a], Mike Shaw [a], Ignacio Partarrieau [a], Nadia Smith [a], Jenny Venton [a], Michael Adeogun [a], David Brettle [b], Robert James Turpin [c]

[a] National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, United Kingdom
[b] Leeds Teaching Hospitals NHS Trust, St. James's University Hospital, Beckett Street, Leeds, West Yorkshire LS9 7TF, United Kingdom
[c] British Standards Institution, 389 Chiswick High Road, London W4 4AL, United Kingdom

ARTICLE INFO

ABSTRACT

Background: Pathology services experienced a surge in demand during the COVID-19 pandemic. Digitalisation of pathology workflows can help to increase throughput, yet many existing digitalisation solutions use non-standardised workflows captured in proprietary data formats and processed by black-box software, yielding data of varying quality. This study presents the views of a UK-led expert group on the barriers to adoption and the required input of measurement science to improve current practices in digital pathology.

Methods: With an aim to support the UK's efforts in digitalisation of pathology services, this study comprised: (1) a review of existing evidence, (2) an online survey of domain experts, and (3) a workshop with 42 representatives from healthcare, regulatory bodies, pharmaceutical industry, academia, equipment, and software manufacturers. The discussion topics included sample processing, data interoperability, image analysis, equipment calibration, and use of novel imaging modalities.

Findings: The lack of data interoperability within the digital pathology workflows hinders data lookup and navigation, according to 80% of attendees. All participants stressed the importance of integrating imaging and non-imaging data for diagnosis, while 80% saw data integration as a priority challenge. 90% identified the benefits of artificial intelligence and machine learning, but identified the need for training and sound performance metrics.

Methods for calibration and providing traceability were seen as essential to establish harmonised, reproducible sample processing, and image acquisition pipelines. Vendor-neutral data standards were seen as a "must-have" for providing meaningful data for downstream analysis. Users and vendors need good practice guidance on evaluation of uncertainty, fitness-for-purpose, and reproducibility of artificial intelligence/machine learning tools. All of the above needs to be accompanied by an upskilling of the pathology workforce.

Conclusions: Digital pathology requires interoperable data formats, reproducible and comparable laboratory workflows, and trustworthy computer analysis software. Despite high interest in the use of novel imaging techniques and artificial intelligence tools, their adoption is slowed down by the lack of guidance and evaluation tools to assess the suitability of these techniques for specific clinical question. Measurement science expertise in uncertainty estimation, standardisation, reference materials, and calibration can help establishing reproducibility and comparability between laboratory procedures, yielding high quality data and providing higher confidence in diagnosis.

Introduction

Pathology is the backbone of diagnostic medicine and contributes to the majority of clinical pathways. As the pressure on health systems increases worldwide, the need for pathology services follows suit. In the UK, the Royal College of Pathologists reported a 4.5% year-on-year increase in demand for pathology services since 2007.[1b] The COVID-19 pandemic has amplified already mounting pressures on pathologists and increased their willingness to adopt remote ways of working.[2] These pressures can be

alleviated by the digitalisation of pathology services, and some early adopters have shown increased throughput of cases by 21% per year.[3]

Aside from increased throughput, digital pathology (DP) promises many benefits including improved diagnosis,[4] new insights into disease phenotypes and mechanisms,[5,6] validation of diagnosis,[7] and the use of artificial intelligence (AI) to support image quality assurance, prioritisation, review, and diagnosis. The rollout of DP poses a set of challenges including imaging data management, computational complexity, interoperability, image comparability, device and data quality assurance (QA), and

---

Glossary

Terms "calibration", "primary standard", "reference material", "repeatability", "reproducibility", "traceability" and "uncertainty" follow the International Vocabulary of Metrology, 3rd edition.[1]

Term	Definition
AI/ML	Artificial Intelligence/Machine Learning
calibration	Operation that first establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and then uses this information to establish a relation for obtaining a measurement result from an indication
DICOM	Digital Imaging and Communications in Medicine: A standard for capturing and exchanging medical imaging data
DP	Digital Pathology
H&E	Haematoxylin and Eosin stain: The most frequent stain type used in histopathology
interoperability	Ability of two or more systems or components to exchange information and to use the information that has been exchanged[a]
LIS	Laboratory Information System
measurand	quantity intended to be measured
NHS	National Health Service: the main UK healthcare provider
NMI	National Measurement Institute
NPL	National Physical Laboratory
omics	Collective term for biology disciplines that study various molecules including genomics, proteomics, transcriptomics metabolomics and many others
primary standard	Measurement standard established using a primary reference measurement procedure, or created as an artefact, chosen by convention
reference material	Material, sufficiently homogeneous and stable with reference to specified properties, which has been established to be fit for its intended use in measurement or in examination of nominal properties
repeatability	Measurement precision under a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time
reproducibility	Measurement precision under a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects
SOP	Standard Operating Procedure
traceability	Property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty
uncertainty	Non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used
WSI	Whole Slide Imaging

[a] https://www.iso.org/obp/ui/#iso:std:iso:ts:27790:ed-1:v1:en:term:3.39, assessed 14/03/2022.

deployment of artificial intelligence/machine learning (AI/ML) solutions. It is possible that the backlog of patients caused by the pandemic may force many services into premature adoption of DP solutions that are based on unstandardised processes, "black box" software, data of variable quality, and unknown quality.

Measurement science (metrology) concepts such as traceability, calibration, reproducibility, and uncertainty quantification can be used to address challenges in DP implementation and advance both clinical and research pathology to the next level. In the UK, metrological traceability underpins medical radiation dosimetry, where radiation dose measurements on therapy units are traceable to the national standard. The need for metrology has also been realised in quantitative imaging, where accurate estimates of pixel values and associated uncertainties are used as tissue or disease type biomarkers.[8]

This work presents the findings of an online survey and a workshop conducted by the UK's National Physical Laboratory in 2021. It collates the views of DP experts from clinical, pre-clinical, research, industry, and regulatory authorities on the metrology support required to address the key challenges in the area.

Digital pathology workflow

DP workflow stages can be roughly grouped into sample preparation, image acquisition, data analysis, and diagnosis. All of these stages pose challenges some of which are unique to DP, while others are shared with conventional on-microscope histopathology.

During the sample preparation stage a biopsy may undergo dehydration, clearing, fixation, sectioning, embedding, and staining of tissue slices on glass slides, as well as possible transport in between different steps. Variations in sample preparation result in high variability in the resulting images and thus can lead to differences in the diagnosis. Slide quality assurance (QA) is typically performed manually via visual assessment and its outcome is dependant upon the experience of the assessor, although AI/ML tools are being considered for early-stage QA.[9] Unsurprisingly, sample processing presents the largest source of the variability within the DP workflow,[10] with staining or "batch" variability being one of the major obstacles in producing consistent and comparable images.[11]

To digitise a glass slide, a whole slide imaging (WSI) scanner scans the stained tissue slides and produces high resolution images of between 1 and 4 Gigabyte in size. WSI scanners typically include a mechanical stage to feed the slides, a light source, optics, and a digital camera sensor. To date, the scope and frequency of WSI calibration and routine tests vary between vendors and laboratories. The 2018 report points out that, while WSI should undergo ongoing QA, since subjective perception of image quality cannot indicate the image is "fit-for-diagnosis", research in the calibration area "is sparse".[12] The main calibration areas include: (1) dimensionality, (2) illumination, and (3) colour, with the latter being the most widely recognised challenge. Colour calibration comprises: (a) internal colour calibration which involves removing the variabilities arising from the scanning process itself, and (b) external colour calibration that focusses on the standardisation of the display, accounting for the monitor's effect on perceived colour and the viewing environment. Existing literature indicates that colour management in DP is challenging due to the lack of standards,[13] while the digitisation of the slide introduces further lack of colour control and compounds the issue.

In additional to conventional stains such as haematoxylin-eosin (H&E) and periodic-acid-Shiff (PAS) that focus on tissue morphology, pathologists can use a range of other techniques. These may include "omics" to provide genetic, proteomic, or metabolomic information, using blood, urine, liquid biopsies,[14] or novel imaging techniques such as mass spectrometry imaging, imaging mass cytometry, immunohistochemistry, Raman microscopy, and many others. Combining multiple techniques offers advantages over histopathology, since a combination of measurements makes it is possible to differentiate diseases where single measurands do not.[15] Such multidimensional data can improve accuracy of diagnosis, including tumour grading[16] and measuring intra-tumour heterogeneity.[17]

The major barriers to wide clinical uptake of these techniques can be broken down into 3 categories: (i) the data collection and sample

preparation differ between modalities and can be more difficult to compare;[18,19] (ii) the data themselves are much larger and more complex meaning data analysis is more challenging;[20] and, (iii) importantly, these methods are not yet fully understood within a clinical setting and are not yet approved by regulating bodies for clinical use.[21]

There is a growing body of research on AI/ML applications for various stages of the DP workflow, including diagnosis, pre-diagnostic slide QA,[22,23] and colour normalisation.[24] A PubMed search on "AI digital pathology" shows a 4-fold growth from 43 publications in 2016 to 190 in 2021. At the same time, the uptake of AI/ML tools in clinical routine is relatively slow. The roadblocks to adoption include large image size, image artefacts, colour variations, regulatory approval, as well as lack of access to large well-annotated datasets, lack of protocols for training and validation of algorithms, and challenges regarding interpretation/explainability of results.[4,25,26] Attempts to increase the amount of data for algorithm development are being undertaken in several national and European projects that aim to create a large-scale AI/ML WSI data processing cloud for algorithm developers and clinical users.[c,d]

Data exchange connects all stages of the DP workflow. WSI data-related challenges can be subdivided into 2 categories: (1) lack of vendor-neutral data format and (2) lack of metadata standards to capture the knowledge about how the image was obtained. The lack of widespread adoption of vendor-neutral WSI formats makes data access, processing and exchange challenging and highly dependent on laboratory, equipment, and software setup. Slow adoption of vendor-neutral formats such as DICOM or OME-TIFF has negative impact on sharing of WSI data within and outside of a hospital.[27]

The heterogeneity of WSI formats causes poor integration of DP within the hospital IT infrastructure. A notable consequence is the inadequate data exchange between WSI scanners and a laboratory information system (LIS) that holds sample processing and patient metadata vital for image interpretation. Although proprietary WSI/LIS solutions exist, they do not solve the issue of inter-institution or inter-system data exchange that is the cornerstone of DP workflow. Efforts to create working specifications for WSI data exchange are being undertaken by the Integrating the Health-care Enterprise (IHE) initiative in collaboration with DICOM working group 26.[28]

Poor interoperability of WSI data is compounded by the lack of consensus on minimum metadata on sample handling, scanner settings, and image processing that limits the diagnostic value of WSI data, its re-use and sharing. Failure to supply appropriate metadata and lack of reporting standards have been recognised as key contributors to the reproducibility crisis in digital medicine,[29] and open repositories such as BioArchive attempt to tackle the issue by specifying their own metadata requirements.[30]

The role of standards

The uptake of DP is greatly limited by the lack of systems interoperability and agreed good practice for data exchange.[27] Added to this is the complexity surrounding AI/ML solutions, and the need to ensure they are safe, secure, and perform in accordance with their intended use.

The adoption of standards – agreed organisational and technical best practices – can help to address the barrier to adoption of DP. Standards are a foundation for innovation, providing a key mechanism by which to diffuse knowledge of what good looks like. They provide a route to consensus between collaborating businesses and expert organisations, and enable policymakers to educate the market. Standards are voluntary agreements of good practice, commonly taking the form of guidance, codes of practice, and specifications. They can inform both the technical and procedural aspects of data exchange, along with the conformity to regulations.

Standards play an important role in DP, from improving data management and interoperability, to ensuring the quality and competence of medical laboratories, and ensuring the safety and performance of medical diagnostic equipment and software. On a country level, standards sit within the National Quality Infrastructure, alongside testing

and measurement to ensure validity and consistency, and certification and accreditation that ensures standards are applied correctly and with competency by the personnel who carry out testing and inspection.

Role for metrology

Metrology is the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science. Metrology methods help to obtain a reproducible measurement result with known uncertainty that is traceable to a reference and the International System of Units (SI).[31] Metrology principles that could be applied to DP workflow include traceability, calibration, uncertainty analysis, reproducibility, and comparison (Fig. 1).

A well-established example of metrology applied in medicine is the traceability chain for UK radiation dosimetry, within which all hospitals delivering radiation therapy own a measurement device (secondary standard) that is regularly calibrated against the national primary standard. In the DP context, a similar traceability chain could be established for WSI image acquisition. Each WSI scanner would be regularly characterised using test objects with known physical properties to measure inter- and intra-device variability. These test objects could be traced to the national standard to assess variations between test objects. Having accurate and up-to-date information about WSI scanner performance, e.g., scanner calibration data, would aid image normalisation and colour re-scaling to match user preferences. Most importantly, normalised WSI datasets could be used to define reproducible quantitative imaging biomarkers with known uncertainties, and could be used to train AI/ML algorithms.

Efforts to apply metrology methods to obtain digital quantitative biomarkers are underway in medical imaging,[32–34] where they are led by the Quantitative Imaging Biomarkers Alliance.[e]

National Measurement Institutes such as NPL are well-placed to develop reference methods and standards for DP and biosciences. The following sections describe the landscape study carried out by NPL in 2021 to gain the views of DP stakeholders on what metrology interventions are most needed by the community.

Methods

Participants for the study were selected through an established network of contacts, through word of mouth, and through identification of stakeholders via internet searches. We aimed to recruit at least four representatives from the professional backgrounds summarised in Fig. 2, while ensuring there were at least 2 organisations in each category. The resulting cohort of participating organisations is presented in Table 1.

The online survey conducted in October–December 2020 contained 31 questions on participants background (n = 6), imaging data acquisition and storage (n = 10), QA and regulations (n = 5), AI and ML applications (n = 4), future directions in the area (n = 4), and further comments (n = 2). The survey was accompanied by a set of interviews with participants from healthcare (n = 3), pre-clinical (n = 2), software vendors (n = 1), and WSI device manufacturers (n = 3).

The survey and interview findings were used to set out 5 discussion groups for the workshop: (1) sample preparation and processing, (2) equipment calibration, (3) image processing and analysis, (4) omics and novel imaging techniques, and (5) data integration.

The workshop attendees were allocated to the discussion groups so that: (1) each group included at least 1 participant from healthcare, pre-clinical, device/software vendor, and regulatory background and (2) participants in each group represented different organisations.

Two rounds of group discussions were used to identify the challenges in the area and to outline whether metrology support is required. Group discussions were led by moderators with expertise in the subject matter and captured by designated notetakers. The intermittent findings of both
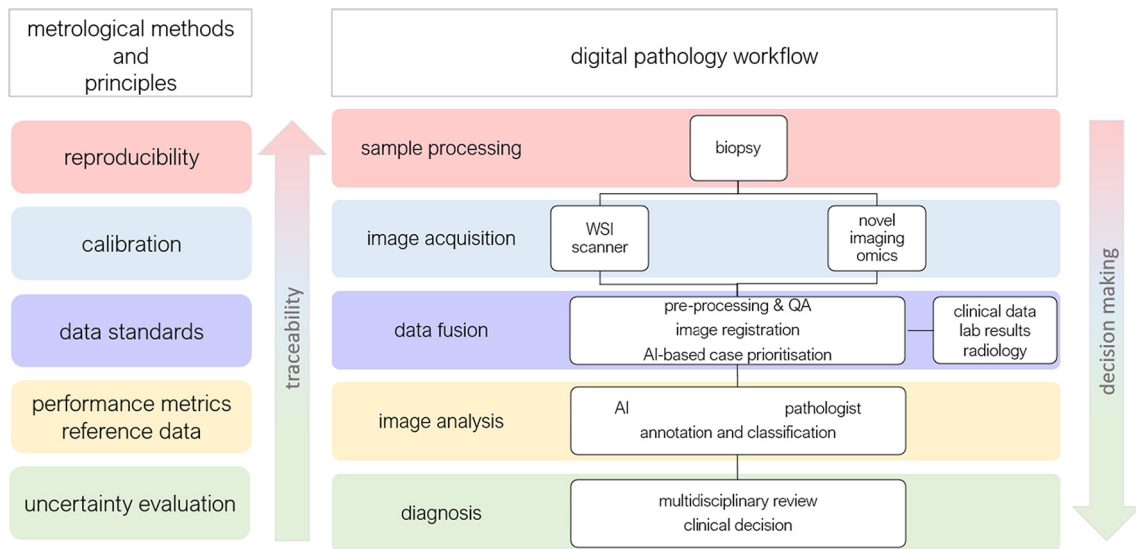

Fig. 1. Critical metrology principles applied to a digital pathology workflow. The same principles apply for non-WSI modalities, e.g., RAMAN, MSI, omics, radiology etc.

discussion sessions were shared with all workshop attendees during the joint session to ensure adequate coverage. Following the workshop, the notes from each group discussion were shared with the group members for review and collated to a single report.

Results

Table 1 lists the organisations that took part in the workshop and survey. A comprehensive summary of the findings is presented in the NPL report "Metrology for Digital Pathology".[35]

Online survey findings

Twenty-four respondents with backgrounds in the health sector (n = 5), academia (n = 3), regulatory bodies (n = 2), drug development (n = 7), vetinary medicine (n = 2), WSI device manufacturing (n = 2), and software development (n = 3) completed the survey. Sixteen participants worked in pathology laboratories. The laboratories varied in the number of slides processed per year: 100–1000 (n = 2, 13%), 1000–5000 (n = 5, 32%), and over 5000 (n = 9, 56%). All laboratories exchanged WSI data with other centres, while most also exchanged glass slides (n = 12, 75%) and tissue samples (n = 11, 69%).
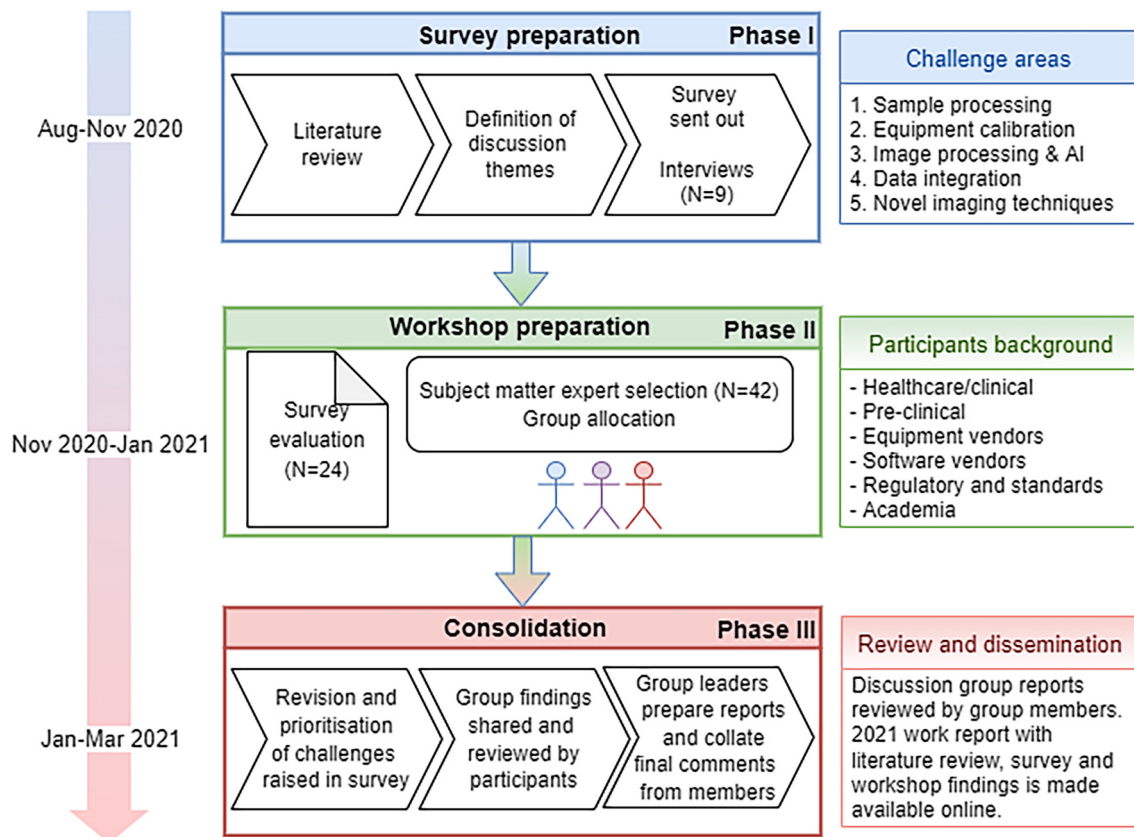

Fig. 2. Information collection process used in this study.

Table 1

Workshop participants background and their allocation to discussion groups. Asterisk (*) denotes participants who took part in a phone interview before the workshop.

Organisation name	Organisation type	Workshop discussion groups	Survey
Ashford and St Peter's Hospitals NHS Foundation Trust	Healthcare provider	Equipment Calibration	Y
AstraZeneca	Industry (pharmaceutical)	Image Analysis, Novel Imaging, Sample Processing	Y
British Standards Institution	Standards bodies	Data Integration	N*
Cambridge University Hospitals	Healthcare provider	Sample Processing	Y
Charles River Laboratories	Industry (pharmaceutical)	Image Analysis	Y
DesAcc	Industry (software)	Data Integration	Y*
DICOM	Standards bodies	Data Integration	N
FFEI Ltd	Industry (equipment)	Equipment Calibration	N
GE Healthcare	Industry (software)	Image Analysis	N
GlaxoSmithKline	Industry (pharmaceutical)	Data Integration, Equipment Calibration, Sample Processing	Y*
IHE	Standards bodies	Data Integration	N
Leeds Teaching Hospitals NHS Trust	Healthcare provider	Equipment Calibration	N*
Leica Biosystems	Industry (equipment)	Image Analysis	Y
NHS Digital	Healthcare provider	Novel imaging	Y
Nikon Corporation	Industry (equipment)	Equipment Calibration	Y
Norfolk and Norwich University Hospitals NHS Foundation Trust	Healthcare provider	Data Integration	Y
Paige.AI	Industry (software)	Data Integration, Image Analysis	Y
Royal Brompton and Harefield Hospitals	Healthcare provider	Sample Processing	Y
Royal Surrey County Hospital	Healthcare provider	Equipment Calibration	N*
Royal Veterinary College	Academia	Sample Processing	Y*
Smith + Nephew	Industry (equipment)	Sample Processing	Y
Surrey Heartlands	Healthcare provider	Novel Imaging	Y
University College London	Academia	Image Analysis	Y
University of Cambridge	Academia	Image Analysis, Equipment Calibration	N*
University of Dundee	Academia	Data Integration	Y
University of Oslo	Academia	Image Analysis	N
University of Surrey	Academia	Image Analysis	Y*
Zeiss	Industry (equipment)	Equipment Calibration	N*

Imaging data acquisition, storage, and processing

Most respondents used multiple imaging modalities, with WSI being the most common (n = 20), followed by immunohistochemistry (n = 14) and then immunofluorescence multiplex assays (n = 11). WSI data was predominantly stored in a vendor-specific format (n = 16), while purpose-built open source formats DICOM for WSI (n = 7) and OME-TIFF (n = 7) as well as generic image formats (n = 10) were also in use. Two respondents used in-house developed formats.

Regarding ease of access to WSI data, most respondents reported it was easy to access and locate the images (n = 23), while most (n = 22) also noted that they needed to combine WSI data with other data sources such as electronic patient records (n = 12), study protocols (n = 12), laboratory results (n = 10), radiology (n = 10), omics (n = 9), and other data (n = 5) for diagnosis. Access to the non-WSI data sources has posed issues for some respondents (n = 16), with key problems being segregated data storage (n = 5), lack of interoperability between different systems hindering transfer of metadata and annotations (n = 11), absence of necessary metadata (n = 7), image registration (n = 2), and large file sizes (n = 3).

Quality assurance and regulations

75% of respondents (n = 18) came from workplaces where good laboratory practice applied. Of these, 76% (n = 13) also applied good clinical practice, while some laboratories also implemented current good manufacturing practice, relevant ISO standards, and software-specific in-house practices. Quality assurance protocols were regularly applied to tissue management equipment, image analysis tools, and imaging equipment, whereby calibration tools for assessment of uniformity of illumination within the field, magnification, and colour calibration were of equal interest.

AI and ML applications

From 79% of responders who were able to use AI/ML tools within their workflow (n = 19), only 16% (n = 3) showed no interest in AI applications (Fig. 3). Cell type annotation and disease classification were most frequent use cases (n = 8 and n = 7, respectively), and AI-assisted image quality assurance was the least frequent case (n = 2). At the same time, cell type annotation was the task where users showed
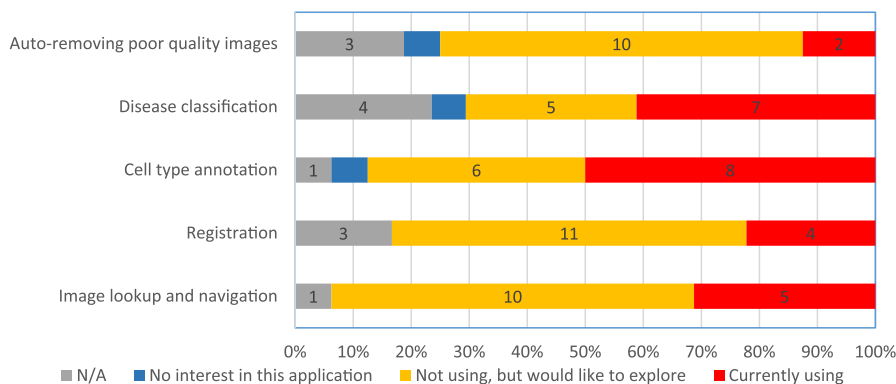

Fig. 3. Use of AI/ML applications in the digital pathology pipeline. There were zero "No interest" responses for "Registration" and "Image lookup and navigation" categories, and 1 response for each of the other categories.
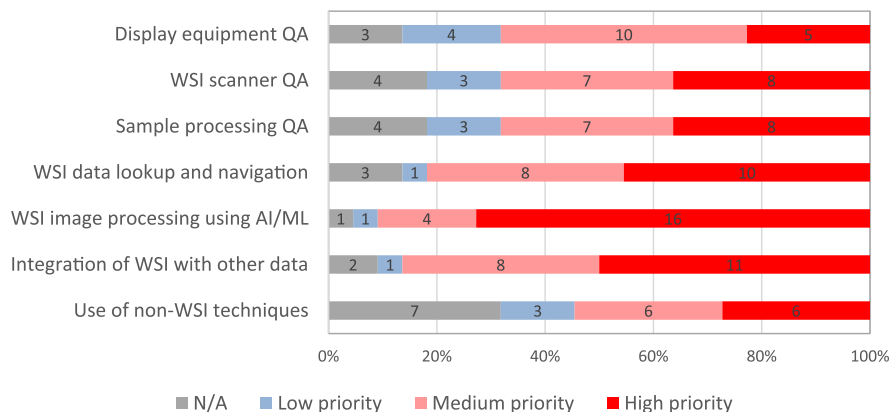
Fig. 4. Prioritised challenges in digital pathology (n = 22).

least confidence (only 39% stated they are confident in the results), while automated quality assurance was the use case where most users (61%) declared high level of confidence.

Performance and reliability of AI tools were accessed by comparison to human operator(s) performance (n = 6), or other software tools (n = 1), while some respondents used multi-centre studies (n = 2).

Use of novel imaging techniques and future work directions

Novel imaging modalities including targeted and untargeted metabolic and proteomic imaging, as well as genomic imaging, were used by relatively few laboratories (5%–16%, mean: 11%). Thereby high numbers of participants were interested in using all or some of these techniques in the future (48%–68%, mean: 58%). Consequently, nearly all participants were interested in software tools to combine data from multiple modalities (Fig. 4). This finding is consistent with the lack of data integration stated in section "Imaging data acquisition, storage, and processing".

Workshop findings

Table 2 presents the selected findings of 2 discussion sessions on: (1) sample preparation and processing, (2) equipment calibration, (3) image processing and analysis, (4) omics and novel imaging techniques, and (5) data integration.

Discussion

This study describes the views of stakeholders on the metrology support required to address pressing challenges in digital pathology. The desired interventions and rationales behind them follow the order of the DP pipeline stages.

Sample handling

It has been recognised that sample handling and preparation present the largest source of variability. It is therefore crucially important to develop vendor-agnostic metrology methods to improve intra- and inter-laboratory reproducibility, e.g., calibration procedures in combination with standardised digital QA and QC. Working with end users, metrologists can develop fit-for-purpose reference materials, quality metrics and standardisation routines for H&E, immunohistochemical, in situ hybridisation, and other tissue staining protocols. The introduction of such metrics and computerised QA and QCs would benefit the community by improving system performance, maintaining reproducibility within and between laboratories and by accelerating the implementation of new techniques.

WSI calibration

The need for sharing and comparative analysis of digital images across sites and the advance of AI/ML systems place new demands on the consistency of image datasets. Such datasets can only be obtained if WSI instruments are subjected to regular and reliable calibration to improve image comparability and reproducibility. National Measurement Institutes (NMIs) should work together with vendors, AI developers, and users to develop calibration procedures and physical artefacts such as calibration slides to assess the imaging device performance including dimensionality, illumination, and colour. Some efforts to provide reliable calibration objects to assess colour reproduction have been undertaken, but need to be developed further to meet the area's needs. In the longer term, creating physical calibration objects that are: (a) validated through round-robin multi-NMI trials and (b) traceable to primary standards and the SI will underpin quantitative imaging in pathology.

Novel imaging

Beyond WSI histopathology, there is a growing interest in novel imaging modalities such as super-resolution and light sheet microscopy, Raman spectroscopy, mass spectrometry imaging, and many others. Understanding the uses, advantages and limitations of these techniques in combination with histopathology requires validation and standardisation efforts from end users and metrologists. Good practice recommendations can then be developed for specific tasks such as diagnosis of a particular disease or stage, quantitative pathology, and integration of conventional H&E imaging with other experimental techniques.

Image analysis

AI/ML tools are increasingly used for image quality assurance, image analysis, annotation, review prioritisation, and disease classification. It is essential that such systems are based on explainable and trustworthy mechanisms that can account for uncertainties and provide visualisation tools to aid the interpretation of their results. The development of trustworthy AI systems requires large volumes of well annotated WSI data with consistent metadata that includes WSI device calibration and sample handling information. Lastly, AI systems should include clear guidelines of their potential usage scenarios or be "intelligent" enough to reject input data when it is not suitable.

Data integration

The intelligent use of DP data by clinicians and AI systems requires that the images are supplied with appropriate metadata and can be reliably

Table 2

Prioritised challenges and metrology support required in Digital Pathology grouped by domains. Priorities denote how urgently the issue should be addressed, with "high" meaning "as soon as possible" and "medium" meaning "within 1–3 years from now."

Challenge	Priority	Metrology support required	Domain
Quality framework is not well developed or taken up by the community. Where quality frameworks are deployed, their implementation varies and is poorly documented	High	Develop standardised operation, quality assurance and quality control procedures for manual and computerised processes. Define framework for traceble capture of information.	Sample preparation
Lack of comparability of in-house sample processing methods with those used by Contract Research Organisation laboratories	Medium	Support the development of reference materials for comparison between staining protocols and sites to monitor and track performance. Prioritise standards and metrics to track stain intensity and to quantify variability.	
Differences in methods and frequency of instrument calibration across laboratories lead to significant variations in results and impact the downstream analysis	High	Develop good practice guides and robust calibration protocols to reduce interlaboratory variability, subjectivity, and bias.	Equipment calibration
Lack of certified methods, instruments, and calibration artefacts to assess WSI device performance	Medium	NMIs should develop vendor-agnostic SI-traceable physical calibration artefacts to quantify variations in colour, focus, luminance, depth of field, and image quality.	
Issues in transfer of pathologists' annotations and WSI metadata limit the availability of datasets for training of AI/ML systems and junior pathologists	High	Introduce data standards for images, annotations and linking of both. Use vendor-neutral format to that can be transferred between systems.	Image processing and analysis
Lack of transparency and explainability in how the system arrived at the result. No mechanisms for calculating and communitating the uncertainty of the results	Medium	Provide mechanisms for standardised and trustworthy outputs from data analysis and training of pathologists in how to interpret the results. Develop DP-specific metrics for evaluation of AI-based system performance.	
Lack of knowledge and training in use of novel imaging techniques	Medium	Support the validation and standardisation of novel imaging techniques to ensure confidence in their capabilities and to understand their uses, advantages, and limitations in DP	Omics and novel imaging techniques
Slow uptake of experimental techniques in clinical practice	Medium	Design a training program on good practices and the use of novel imaging techniques to aid their adoption in DP and to disseminate the interdisciplinary knowledge.	
Heterogeneity of image formats and metadata contents hinders the data exchange between devices, software systems, and institutions	High	Encourage use of a vendor-neutral format for within DP pipeline. Agree on a set of minimum metadata to promote traceability, account for variations, and enable meaningful analysis.	Data integration
Segregation of data between WSI and laboratory/clinical systems hinders data collection, costs time, and increases potential for errors	High	Define tests for DP interoperability, including benchmarking tests to ensure that solutions work at scale, and guides on practical implementations to make the standards work for end users	

linked to other data sources such as radiology or laboratory findings. Unlike radiology, where the Digital Imaging and Communications in Medicine standard (DICOM) enables data exchange with the hospital information system, DP systems tend to use proprietary storage and have limited interoperability. Using a vendor-agnostic image format such as DICOM for WSI will solve this issue and provide a common "language" to exchange data with other clinical systems. However, DICOM alone does not guarantee data quality. It is perfectly possible to have a well-shaped DICOM WSI file that has insufficient information about the image context and is clinically meaningless. Therefore, the community needs robust metadata standards to capture sample handling, imaging device setting, and image pre-processing steps. These metadata should be captured using consistent clinical terminologies, ontologies, and units of measurement. DICOM WSI offers suitable mechanisms to encode metadata in widely used clinical nomenclatures such as SNOMED CT, LOINC, and UCUM. NMIs can develop minimum metadata guidelines in collaboration with DICOM working group 26. Agreeing a vendor-neutral format for annotations and defining the minimum metadata to be stored with the image will make the DP imaging pathway ready for AI/ML deployment, increasing the throughput, reliability, and adoption of DP.

Summary

Across all challenge areas, there was a call for metrological guidance and shaping an appropriate training program to upskill the existing staff and to educate the new generation of pathologists on the role and value of metrology in creating reliable and trustworthy digital pathology solutions.

Web resources

Accessed between November 2021 and March 2022.

- Metrology for Digital Pathology: Digital pathology cross-theme project report http://eprintspublications.npl.co.uk/9279/1/AS102.pdf
- European Metrology Network for traceability in laboratory medicine: https://www.euramet.org/european-metrology-networks/laboratory-medicine/
- "Testing times to come? An evaluation of pathology capacity across the UK", Cancer Research UK, 2016 https://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruk.pdf

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Köhler R. The international vocabulary of metrology, 3rd edition: basic and general concepts and associated terms. Why? How?. Transverse Disciplines in Metrology [Internet]. John Wiley & Sons, Ltd; 2009. p. 233–238. [cited 2022 Mar 14]. Available from: https://doi.org/10.1002/9780470611371.ch21.
2. Browning L, Fryer E, Roskell D, et al. Role of digital pathology in diagnostic histopathology in the response to COVID-19: results from a survey of experience in a UK tertiary referral hospital. J Clin Pathol 2021 Feb 1;74(2):129–132.
3. Retamero JA, Aneiros-Fernandez J, Del Moral RG. Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network. Arch Pathol Lab Med 2020 Feb;144(2):221–228.
4. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019 Nov;16(11):703–715.
5. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, Balis UGJ. Digital pathology and computational image analysis in nephropathology. Nat Rev Nephrol 2020 Nov;16(11):669–685.
6. Burlutskiy N, Waring P, Hipp J. The revival of the H&E with artificial intelligence. J Clin Anat Pathol 2020;5(2):108–114. https://doi.org/10.47275/2332-4864-108http://www.clinpathology.com/the-revival-of-the-he-with-artificial-intelligence/. Accessed November 18, 2022.
7. Snead DRJ, Tsang Y-W, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology 2016 Jun;68(7):1063–1072.
8. Cashmore MT, McCann AJ, Wastling SJ, McGrath C, Thornton J, Hall MG. Clinical quantitative MRI and the need for metrology. Br J Radiol 2021 Apr 1;94(1120):20201215.
9. Niazi MKK, Parwani AV, Gurcan M. Digital pathology and artificial intelligence. Lancet Oncol 2019 May;20(5):e253–e261.
10. Marée R. The need for careful data collection for pattern recognition in digital pathology. J Pathol Inform 2017;8(1):19.
11. Aeffner F, Zarella D M, Buchbinder N, et al. Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. J Pathol Inform 2019;10(9). https://doi.org/10.4103/jpi.jpi_82_18https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6437786/. Accessed November 18, 2022.
12. Cross S, Furness P, Igali L, Snead D, Treanor D. Best practice recommendations for implementing digital pathologyThe Royal College of Pathologists. . 2018.1(1):1–38. v1: https://www.rcpath.org/uploads/assets/f465d1b3-797b-4297-b7fedc00b4d77e51/Best-practice-recommendations-for-implementing-digital-pathology.pdf, 1 2018. Accessed November 18, 2022.
13. Clarke EL, Treanor D. Colour in digital pathology: a review. Histopathology 2017 Jan;70(2):153–163.
14. Palmirotta R, Lovero D, Cafforio P, et al. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. Therapeutic Advances in Medical Oncology 2018;10. https://doi.org/10.1177/1758835918794630. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6116068/.
15. Zhou Y, Liu C-H, Wu B, et al. Optical biopsy identification and grading of gliomas using label-free visible resonance Raman spectroscopy. J Biomed Opt 2019 Sep;24(9):1-12.
16. Morais CLM, Lilo T, Ashton KM, et al. Determination of meningioma brain tumour grades using Raman microspectroscopy imaging. Analyst 2019;144(23):7024–7031.
17. Abdelmoula WM, Balluff B, Englert S, et al. Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data. Proc Natl Acad Sci 2016 Oct 25;113(43):12244.
18. Goodwin RJA. Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences. J Proteomics 2012 Aug;75(16):4893–4911.
19. Everall NJ. Confocal Raman microscopy: common errors and artefacts. The Analyst 2010;135(10):2512.
20. Shakya R, Nguyen TH, Waterhouse N, Khanna R. Immune contexture analysis in immuno-oncology: applications and challenges of multiplex fluorescent immunohistochemistry Clin Transl Immunol [Internet] 2020 Jan. [cited 2021 Mar 31];9(10). Available from: http://dx.doi.org/10.1002/cti2.1183.
21. Finlayson D, Rinaldi C, Baker MJ. Is infrared spectroscopy ready for the clinic? Anal Chem 2019 Oct 1;91(19):12117–12128.
22. Wang M, Kothari S, Phan J. Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. J Pathol Inform 2013;4(1):22.
23. Senaras C, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. PLOS ONE 2018 Oct 25;13(10), e0205387.
24. Lafarge MW, Pluim JPW, Eppenhof KAJ, Moeskops P, Veta M. Domain-adversarial neural networks to address the appearance variability of histopathology images. In: *Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, et al. eds.* Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing; 2017. p. 83–91.
25. Serag A, Ion-Margineanu A, Qureshi H, et al. Translational AI and deep learning in diagnostic pathology. Front Med 2019 Oct 1;6:185.
26. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J 2018;16:34–42.
27. Clunie DA. DICOM format and protocol standardization—a core requirement for digital pathology success. Toxicol Pathol 2021 Jun 1;49(4):738–749.
28. Dash N, Jones N, Merrick R, et al. Integrating the health-care enterprise pathology and laboratory medicine guideline for digital pathology interoperability. J Pathol Inform 2021;12(1):16.
29. Stupple A, Singerman D, Celi LA. The reproducibility crisis in the age of digital medicine. Npj Digit Med 2019 Jan 29;2(1):1–3.
30. Sarkans U, Chiu W, Collinson L, et al. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. Nat Methods 2021 Dec;18(12):1418–1422.
31. Czichos H, Saito T, Smith LE. *Springer Handbook of Metrology and Testing.* Springer Science & Business Media. 2011.
32. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. Radiology 2015 Dec;277(3):813–825.
33. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015 Feb;24(1):27–67.
34. European Society of Radiology (ESR). ESR statement on the validation of imaging biomarkers. Insights Imaging 2020 Jun 4;11(1):76.
35. Adeogun M, Bunch J, Dexter A, et al. *Metrology for Digital Pathology. Digital pathology cross-theme project report [Internet]* 2021. [cited 2022 Mar 3]. Available from: http://dx.doi.org/10.47120/npl.AS102.