Digital Pathology and Artificial Intelligence: Aspiration or Reality?

Dr Chee Leong **Cheng**

Singapore General Hospital

Disclosure of Relevant Financial Relationships

- Sponsored to attend Philips Digital and Computational Pathology APAC Advisory Board Meeting
- Honorarium from Recordati Rare Disease for speaking engagement
- Honorarium from AstraZeneca Singapore for speaking engagement

Outline

- Can we implement digital pathology for primary diagnosis?
- What are the considerations in implementing digital pathology for primary diagnosis?
- What are the potential for artificial intelligence in digital pathology?
- What are the considerations in implementing artificial intelligence in digital pathology?
- What is the future for diagnostic use of artificial intelligence in digital pathology?

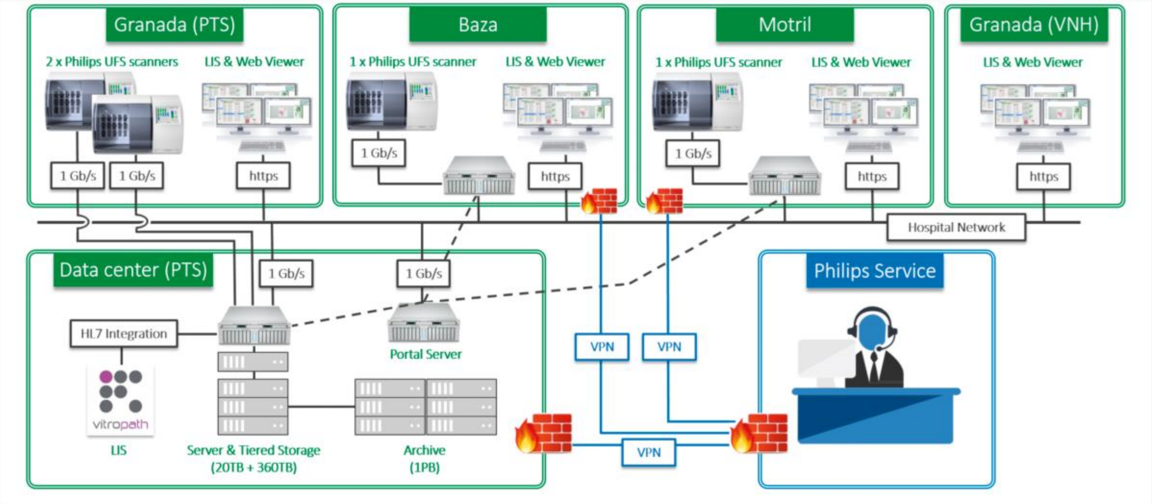
Can we implement digital pathology for primary diagnosis?

Implementations of Digital Pathology for Primary Diagnosis

- Several institutions in US, Europe (e.g. UK, Sweden, Spain, Portugal, France) and Asia (e.g. Japan, Korea) have adopted a fully digital workflow.
- Some published examples:
    - Granada, Spain (Retamero JA, et al. Complete Digital Pathology for Routine Histopathology Diagnosis in a Multicenter Hospital Network. Arch Pathol Lab Med. 2020 Feb;144(2):221-228)
    - Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Portugal (Eloy C, et al. Digital Pathology Workflow Implementation at IPATIMUP. Diagnostics (Basel). 2021 Nov 15;11(11):2111)
    - Caltagirone pathology laboratory at Gravina Hospital, Sicily, Italy (Fraggetta F, et al. A Survival Guide for the Rapid Transition to a Fully Digital Workflow: The "Caltagirone Example". Diagnostics (Basel). 2021 Oct 16;11(10):1916.)
    - Stanford, USA (Rojansky R, et al. Rapid Deployment of Whole Slide Imaging for Primary Diagnosis in Surgical Pathology at Stanford Medicine: Responding to Challenges of the COVID-19 Pandemic. Arch Pathol Lab Med. 2023 Mar 1;147(3):359-367)

Fully digital: Implementation experience

- Granada, Spain (Retamero JA, et al. Complete Digital Pathology for Routine Histopathology Diagnosis in a Multicenter Hospital Network. Arch Pathol Lab Med. 2020 Feb;144(2):221-228)
  - Comprises 4 hospitals, 23 pathologists, fully digital for primary diagnosis in 2016 (6 months implementation)
  - Encompass H&E, immunohistochemistry (IHC) and histochemistry slides.
  - At time of publication, produced 800,000 WSIs from 160,000 specimens; pathologists were able to tolerate a 21% case load increase.



Fully digital: Implementation experience

- IPATIMUP, Portugal (Eloy C, et al. Digital Pathology Workflow Implementation at IPATIMUP. Diagnostics (Basel). 2021 Nov 15;11(11):2111)
  - 25,000 cases per year, 40,000 paraffin blocks, 60,000 slides
  - Workflow design started back in 2016, technology implementation took place in 2019, clinical validation completed in 2020.
  - During clinical validation, adopted a hybrid workflow (glass + digital)
  - At time of publication, 8 out of 14 pathologist using WSIs for primary diagnosis (remaining 6 pathologists are cytopathologists)

Table 1. WSI bright field results of the 8 months operating fully digitally.

	Month								
	July 2020	August 2020	September 2020	October 2020	November 2020	December 2020	January 2021	February 2021	Mean Value
Slides scanned (*n*)	7047	5818	8159	9099	7807	7135	6349	6004	7177
Cases scanned (*n*)	1688	1335	1814	1871	1697	1307	1290	1361	1545
Cases re-scanned (*n*; %) (by technique order)	30; 1.8	23; 1.7	31; 1.7	27; 1.4	5; 0.3	1; 0.1	5; 0.4	4; 0.3	16; 1.0
Cases with good image (%) (by pathologist order)	96.3	97.6	99.0	98.9	98.5	99.1	98.5	98.5	98.3
Cases with glass slides requested (%)	2.1	1.6	1.6	2.1	2.0	2.2	3.3	3.8	2.3

Fully digital: Implementation experience

- Caltagirone pathology laboratory at Gravina Hospital, Sicily, Italy
  (Fraggetta F, et al. A Survival Guide for the Rapid Transition to a Fully Digital Workflow: The "Caltagirone Example". Diagnostics (Basel). 2021 Oct 16;11(10):1916.)
  - Support 7 hospitals. In 2019, 8182 cases per year, 42,245 slides
  - Emphasize on extensive workflow re-design with introduction of extensive tracking, automation and creation of image inventory at various steps.
  - Adoption of LIS driven workflow, accessing case from a worklist and opening virtual slides from virtual slide tray.

Table 1. Automation introduced at every step of the workflow.

Phase	Automation Introduced
Accessioning	Adoption of order entry
	A4 flat scanner to digitize all the paper documents associated with the cases (i.e., endoscopic exams, clinical annotations, etc.)
Grossing	Introduction of a laser block printer at grossing
	Introduction of a camera device to take pictures at the grossing bench Possibility to capture the material in the block
Processing/embedding	Possibility of matching blocks produced at grossing with those sent to processing by using a real-time multi-barcode scanner
Sectioning	Possibility to capture the cut surface of the block for review purposes
	Automated printing of barcodes directly on glass slides rather than on labels
Staining	Automation of requests of histochemical and immunohistochemical stains, which are delivered directly to the stainer
Archiving	Improvement of the archiving of slides and blocks, whose position in the storage trays is random and tracked automatically by barcode scanning

Fully digital: Implementation experience

- Stanford, USA (Rojansky R, et al. Rapid Deployment of Whole Slide Imaging for Primary Diagnosis in Surgical Pathology at Stanford Medicine: Responding to Challenges of the COVID-19 Pandemic. Arch Pathol Lab Med. 2023 Mar 1;147(3):359-367)
  - 2018-2019 yearly volume was 94,000 specimens (daily average 751 blocks, 1588 slides), 52 pathologists.
  - Stepwise subspecialty approach initially since 2018. Implementation accelerated in 2020 during COVID19 pandemic. >95% diagnostic concordance at validation.

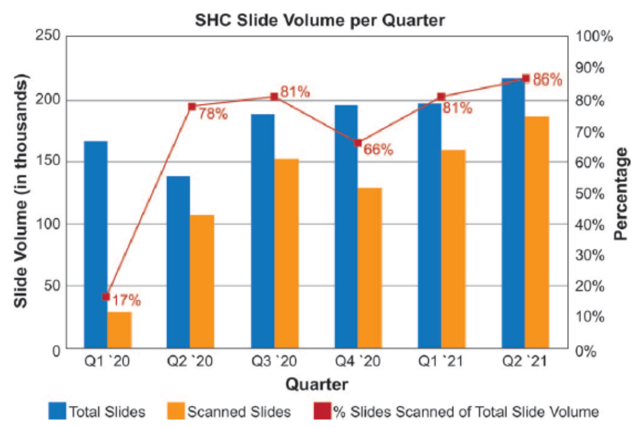


Figure 1. *Stanford Pathology scanning volume by quarter in 2020 and 2021. Total number of slides produced by Stanford Histology Laboratory (blue bars). Total number of slides scanned (yellow bars). Percentage of total slide volume that was scanned (red line). Abbreviations: SHC, Stanford Healthcare; Q1, January through March; Q2, April through June; Q3, July through September; Q4, October through December.*

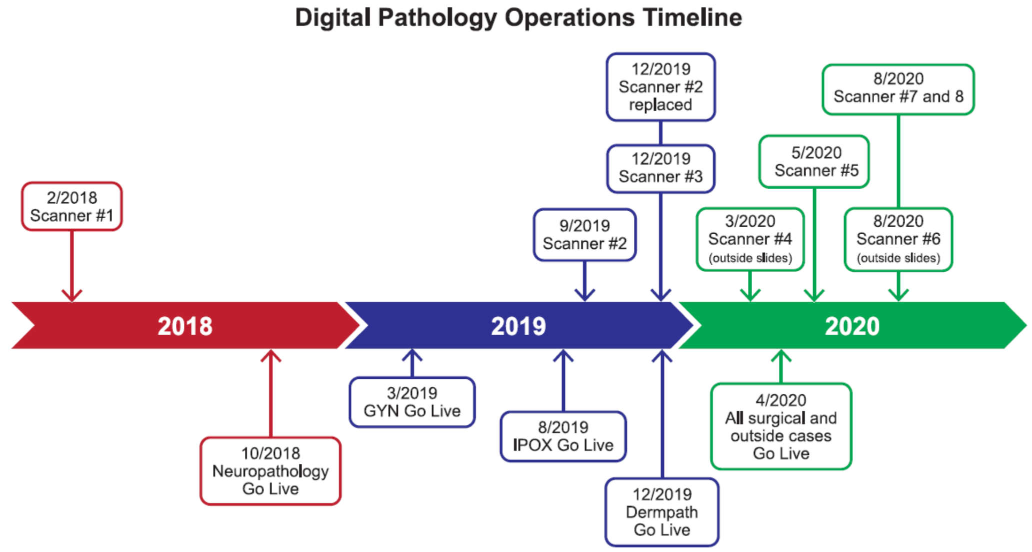


Figure 3. *Timeline of Stanford Pathology whole slide imaging (WSI) implementation beginning February 2018. Blue boxes show points at which scanners were added. Red boxes show the dates of the initial stepwise implementation of WSI for each of 4 subspecialty services followed by broad implementation across all surgical pathology and consult services in April 2020. Abbreviations: GYN, gynecologic pathology; IPOX, immunohistochemistry.*

Impact of whole slide imaging on reading time

- Baidoshvili A, et al. A whole-slide imaging based workflow reduces the reading time of pathologists. Pathol Int. 2023 Mar;73(3):127-134.
  - Compare reading time between conventional optical microscope and digital whole slide images (WSI) workflow at Laboratory for Pathology Eastern Netherlands (LabPON, Hengelo, The Netherlands), where WSIs has been used for primary diagnosis for a number of years.
  - Two arms: 4 pathologists in conventional, 5 pathologists in digital, with random distribution of consecutive cases across both arms; duration of study was 2 months
  - For conventional workflow, time capture is through scanning of barcodes in glass slide; for digital workflow was based on image management system log files.
  - Cases categorised by tissue type show comparable or lower reading time for digital compared to conventional arm, with an average time advantage of 12.3% for digital workflow, in the context of a fully integrated digital set up with pathologists familiar with digital reading.
  - Proposed that past studies demonstrating increased reading time for digital workflow is likely due to unfamiliarity with digital slides, and non-ideal digital integration

SGH: Expanding our capacity towards a fully digital workflow

2013/2014
- Routine Single Slides
- Subset of Workload

Evolve and Expand

2022/2023

2023/2024

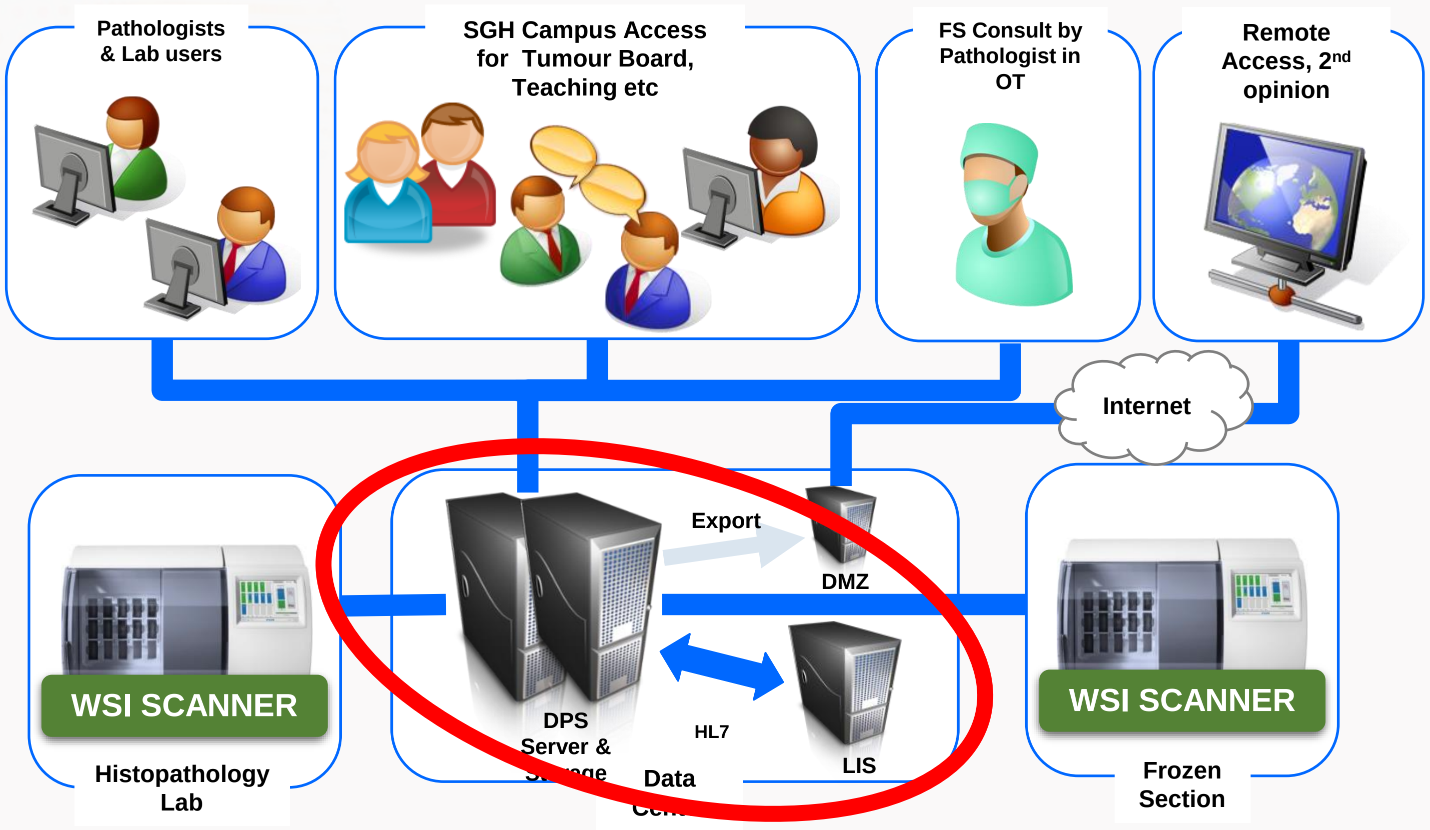
- Expansion of slide scanning capacity to cover majority of workload
- Evolution of slide scanning platform
- Double slide / Wholemount scanning
- Cytology scanning (education)
- Adoption of purpose built platforms (e.g. education/research)

Can we implement digital pathology for primary diagnosis?

- **Yes, but……**

What are the considerations in implementing digital pathology for primary diagnosis?

Integrated digital set up



Reference: Cheng CL, Azhar R, Sng SH, et. J Clin Pathol. 2016 Sep;69(9):784-92..

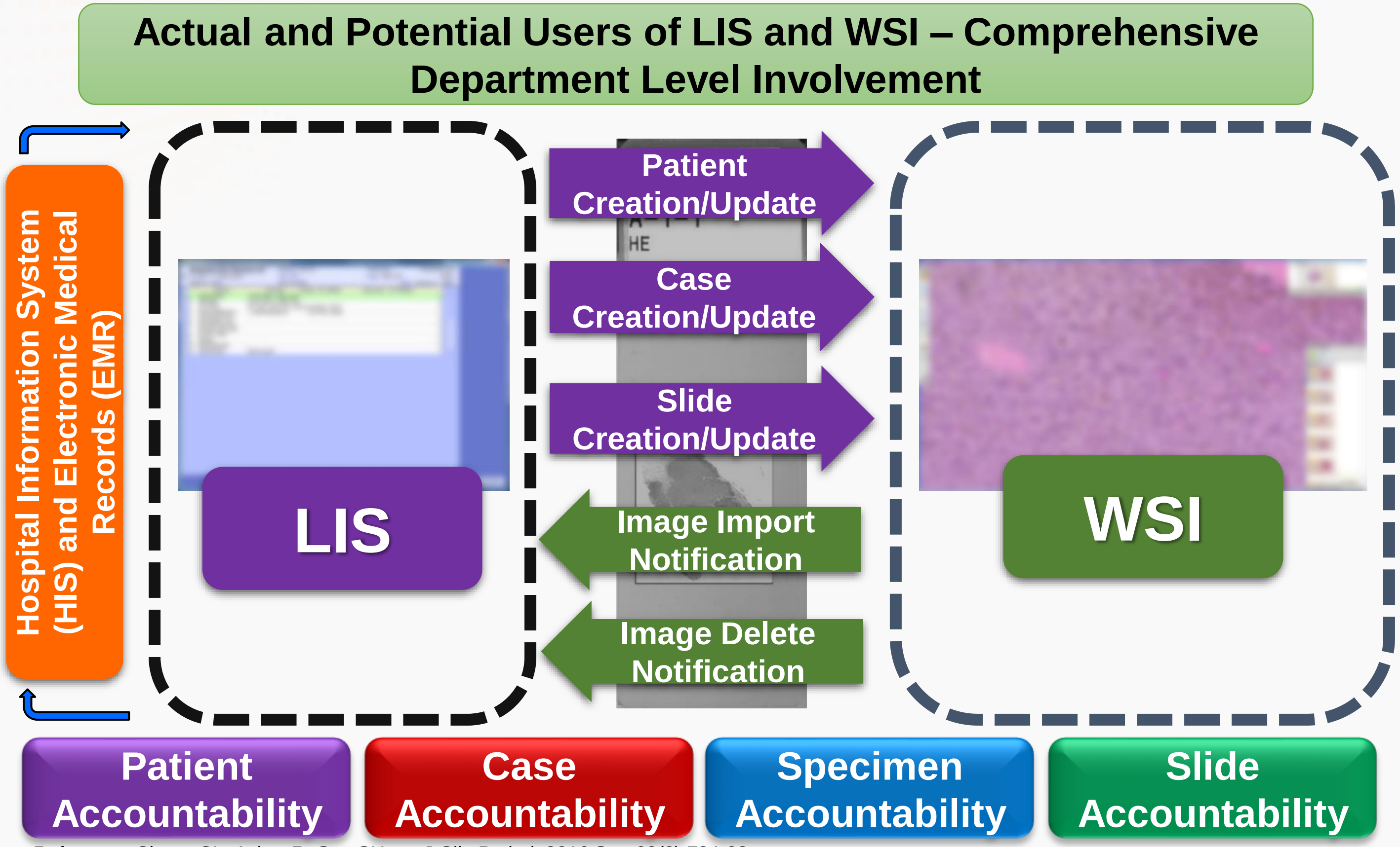
Integration with laboratory information system (LIS)

- Avoid double entry of patient and case data into separate systems
- Updates to patient and case metadata in LIS can be transmitted and consumed by DPS through automated means
- Provide sustainable information management and readily scalable slide scanning.
- Provides the potential for synergy of functionality and business rules between DPS and LIS, and to overcome system limitations by novel means

Challenges with LIS Integration

- LIS have many users and intertwined system flows that may influence metadata updates to DPS.
- Require a comprehensive Implementation approach, including consideration of exceptions to ensure information integrity.
- Roll back of “bad” design decisions, particularly beyond “go-live” may not be easy or even possible, and workarounds may be cumbersome.
- Deep engagement with LIS vendor to go through this digital pathology journey together is necessary for creative solutioning.

Integrated workflow with LIS



Reference: Cheng CL, Azhar R, Sng SH, et. J Clin Pathol. 2016 Sep;69(9):784-92..

Validation of Digital Pathology/WSI

- Digital pathology/WSI is a modality which needs to be validated like a “test” in the diagnostic laboratory context, but with a user/operator element
  - Validation that the technology meets the minimum requirement to be implemented across the laboratory
  - Validation of the user/operatory of the technology (i.e. the pathologists)

Validation of Digital Pathology/WSI

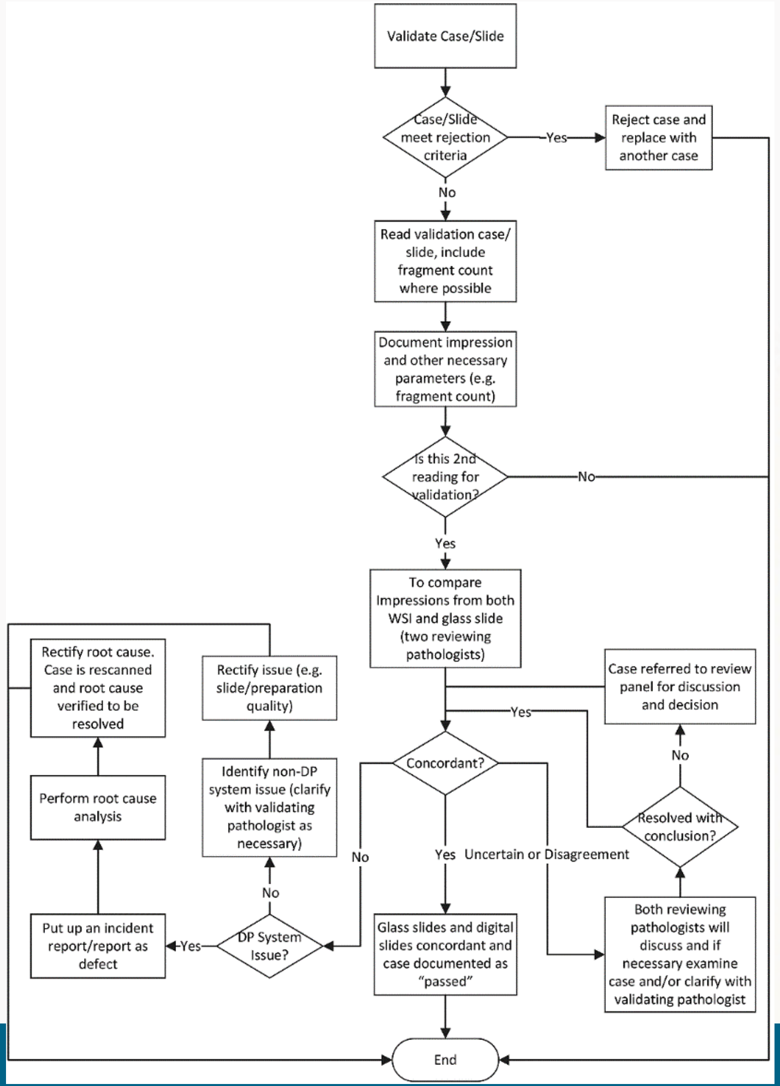
- College of American Pathologists (Andrew J Evans, et al; Validating Whole Slide Imaging Systems for Diagnostic Purposes in Pathology: Guideline Update From the College of American Pathologists in Collaboration With the American Society for Clinical Pathology and the Association for Pathology Informatics. Arch Pathol Lab Med 1 April 2022; 146 (4): 440–450)
  - 3 strong recommendations, 9 good practice statements

Recommendation	Strength of Recommendation
1. The validation process should include a sample set of at least 60 cases for one application, or use case (eg, hematoxylin-eosin–stained sections of fixed tissue, frozen sections, hematology), that reflect the spectrum and complexity of specimen types and diagnoses likely to be encountered during routine practice. The validation should include another 20 cases to cover additional applications such as immunohistochemistry or other special stains if these applications are relevant to an intended use and were not included in the 60 cases mentioned above	Strong
2. The validation study should establish diagnostic concordance between digital and glass slides for the same observer (ie, intraobserver variability). If concordance is less than 95%, laboratories should investigate and attempt to remedy the cause	Strong
3. A washout period of at least 2 weeks should occur between viewing digital and glass slides	Strong

Validation of Digital Pathology/WSI

Validation of DPS in SGH

- Followed published CAP recommendations.
- The number of cases involved, pathologists involved, validation record templates, validation action plan (Figure on the right) are approved by the steering committee (Chaired by HOD) prior to execution.
- Completed key validations prior to clinical go-live of the system.



Validation Template used in SGH

CASE NO.: **BIOPSY NO.:** **DATE:**

SINGAPORE GENERAL HOSPITAL – DEPARTMENT OF PATHOLOGY
PHILIPS DIGITAL PATHOLOGY SOLUTION VALIDATION NOV/DEC 2013

PATHOLOGIST: **DATE SLIDES SEEN:**

MATERIAL TYPE: FROZEN SECTION

READ NUMBER: FIRST READ (GLASS SLIDES)

For Use By DP Technician

NUMBER OF FRAGMENTS ON GLASS SLIDES:

NUMBER OF FRAGMENTS: (PLEASE STATE IF NOT POSSIBLE TO COUNT)	

IMPRESSION:

Is the reading completed within 7min (1 slide) / 14 min (2 slides) / 21 min (3 slides)?
(please tick as appropriate) ☐ **Yes** ☐ **No**
If "No", please state approximately how long was required: __________

PATHOLOGIST SIGNATURE:

CASE NO.: **BIOPSY NO.:** **DATE FIRST READ:**

SINGAPORE GENERAL HOSPITAL – DEPARTMENT OF PATHOLOGY
PHILIPS DIGITAL PATHOLOGY SOLUTION VALIDATION NOV/DEC 2013

PATHOLOGIST: **DATE CURRENT READ:**

MATERIAL TYPE: FROZEN SECTION

READ NUMBER: SECOND READ (WHOLE SLIDE IMAGE)

NUMBER OF FRAGMENTS: (PLEASE STATE IF NOT POSSIBLE TO COUNT OR SECTION APPEARS OBVIOUSLY INCOMPLETE)	

IMPRESSION:

(Please continue next page if space is insufficient)

Is the reading completed within 7min (1 slide) / 14 min (2 slides) / 21 min (3 slides)?
(please tick as appropriate) ☐ **Yes** ☐ **No**
If "No", please state approximately how long was required: __________

Please rate the quality of the whole slide image (circle selection): [Poor] 1 2 3 4 5 [Excellent]
Comments on image quality:

PATHOLOGIST SIGNATURE:

For Use By DP Technician

NUMBER OF FRAGMENTS COMPARED TO GLASS SLIDES:

For Use By Review Personnel Only

INTRAOBSERVER COMPARISON: *(Please tick as appropriate)*
☐ No discrepancy ☐ Minor discrepancy ☐ Major discrepancy ☐ Uncertain if Major Discrepancy

CONCLUSION BETWEEN TWO REVIEWERS: *(Please tick as appropriate)*
☐ Concordant ☐ Not concordant ☐ Refer to review panel

Comments:

Post-Deployment Validation - SGH

- Paired delivery of glass slides and WSI (i.e. slides scanned prior to delivery of glass slides) on randomly selected cases reflecting the lab’s case load.
- Pathologist compare WSI with glass slides at the same time and answer the following questions
    1. *Initial diagnostic impression on WSI*
    2. *Does the impression of the digital slides match that of the glass slides?* (Choose from “Yes”, “No”, “Somewhat” and provide reasons for the latter two)
    3. *Please rate the quality of the whole slide images* (choose from “1” [worst] and “5” [excellent])
- Duration of 6 months, which is also an exercise to allow pathologists across the department to gain experience and confidence with WSI, as well as technical refinement.

Post-Deployment Validation - SGH

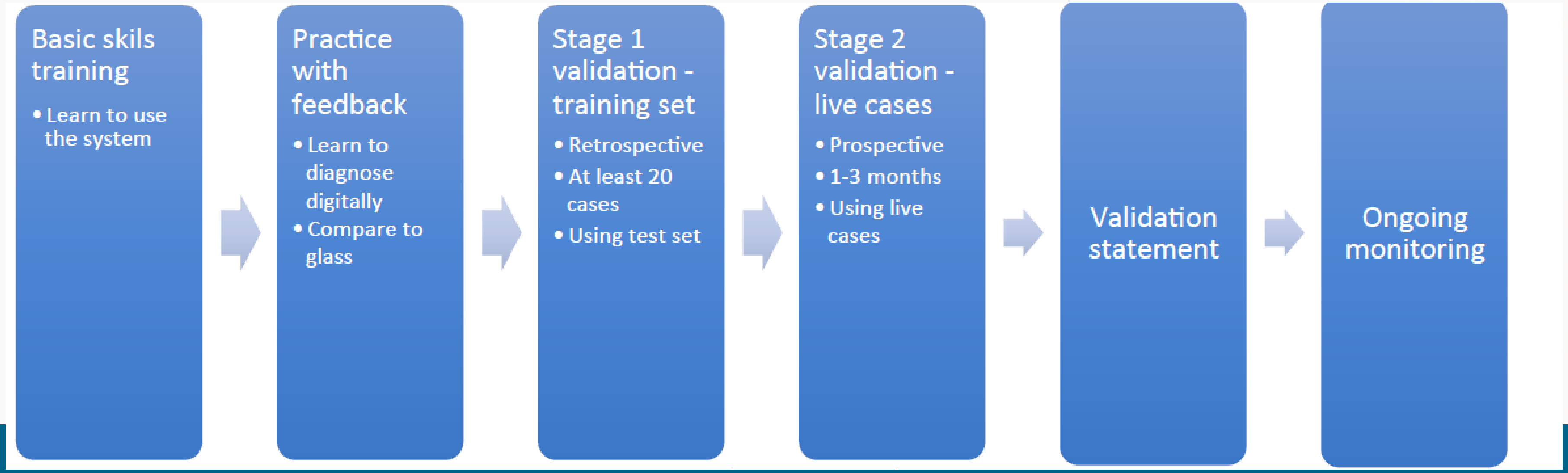
Type	DOES THE IMPRESSION ON THE DIGITAL SLIDES MATCH THAT OF GLASS SLIDES?	No. of Responses	% Response (by type)
FS	Somewhat	4	2.76%
FS	Yes	141	97.24%
	Total FS	145	
Routine	No	8	0.40%
Routine	Somewhat	71	3.57%
Routine	Yes	1912	96.03%
	Total Routine	1991	
	Grand Total	2053	

Paired glass slides and digital slides for routine cases at point of reporting

Good feedback received in general – about 4% will still require glass slide review

Validating DP for Each User/Operator

- The Royal College of Pathologists (United Kingdom). Best practice recommendations for implementing digital pathology (last reviewed Jan 2023, source: https://www.rcpath.org/static/f465d1b3-797b-4297-b7fedc00b4d77e51/Best-practice-recommendations-for-implementing-digital-pathology.pdf)
  - Emphasised on "*pathologist-led, self-validation*" process, with "*cautious approach, including ready recourse to conventional microscopes when needed*", such that the pathologist become "*confident and knowledgeable users of digital pathology*"



Other Considerations

- Equipments
  - Displays
  - Pointing/Navigation devices
  - Network connections (including VPNs in remote access)
- Quality Control and Assurance
- Environmental factors, including ambient lighting and screen fatigue

Examples of issues that remains challenging

- Identification of Helicobacter pylori organisms primarily on H&E slides
- Assessment of immunohistochemistry (IHC) intensity for membrane markers

Identification of *Helicobacter pylori*

- Currently, most diagnosis of *Helicobacter pylori* is based on H&E stained slides, with immunohistochemistry (IHC)/histochemistry used in a subset based on situation.
- Chen W, et al. Comparing Accuracy of Helicobacter pylori Identification Using Traditional Hematoxylin and Eosin-Stained Glass Slides With Digital Whole Slide Imaging. Lab Invest. 2024 Jan;104(1):100262.
  - Involved 30 gastric biopsies (20 H.pylori positive, 10 H.pylori negative, IHC ground truth) and 13 users (7 GI subspecialty pathologist, 3 community general pathologists, 3 pathology trainees)
  - Compare pathologist identification of H.pylori using glass slides vs whole slide images: Diagnostic accuracy of glass vs digital H&E being 81% vs 72% ($p = 0.0142$)
  - If IHC is used (14 cases with most interobserver disagreement), diagnostic accuracy is comparable: 96% glass vs 99% digital.
  - Recommend to review glass slides and/or special/IHC stains, particularly where there is discordance between nature of inflammation and *H.pylori* status.

Identification of *Helicobacter pylori*

- Employ IHC/special stains on all cases?
  - Increased cost
  - Artefactual staining may cause false positive intepretation
- IHC/special stains on selected cases?
  - Subjectivity for assessment of degree of inflammation considered “significant”
- Possibility of artificial intelligence assisted identification of *H.pylori* and associated inflammation?

IHC membrane marker intensity assessment

- Magnification “rule” exists for several predictive markers:
  - For HER2 IHC assessment in breast cancer (Tan PH et al. Practical approach to scoring HER2 immunohistochemistry in breast cancer in the wake of updated guidelines. Histopathology. 2024 Mar;84(4):715-718)
    - Weak : Clearly observed at 40x objective, recognised at 20x and up to 10x
    - Moderate : Readily observed at 20x and easily visualised at 10x and 4x objective
    - Strong 3+ : Unequivocally strong and readily noted at low magnification
  - For PD-L1 22C3 CPS assessment in urothelial cancer, the vendor interpretation manual stated “*Evaluation of membrane staining should be performed at no higher than 20× magnification. Slide reviewer should not perform the CPS calculation at 40× magnification*” (Reference: https://www.agilent.com/cs/library/usermanuals/public/29276_22C3_pharmdx_uc_interpretation_manual_us.pdf)
  - The colour intensity and mean width of DAB-precipitates at various scores is proposed to be related to the detection resolution of the microscope objectives. (Scheel AH, et al. Physical basis of the 'magnification rule' for standardized Immunohistochemical scoring of HER2 in breast and gastric cancer. Diagn Pathol. 2018 Mar 12;13(1):19)

IHC membrane marker intensity assessment

Magnification	40x	20x	10x	5x
Consensus Intensity-Score	'1+'	'2+'		'3+'
Num. Aperture	0.65 – 0.75	0.40 – 0.50	0.25 – 0.30	0.12 – 0.15
Resolution [µm]	0.40 – 0.46	0.60 – 0.75	1.0 – 1.20	2.0 – 2.50
DAB-Precipitate width ± SD [µm]	0.64 ± 0.1	1.0 ± 0.23		2.14 ± 0.4

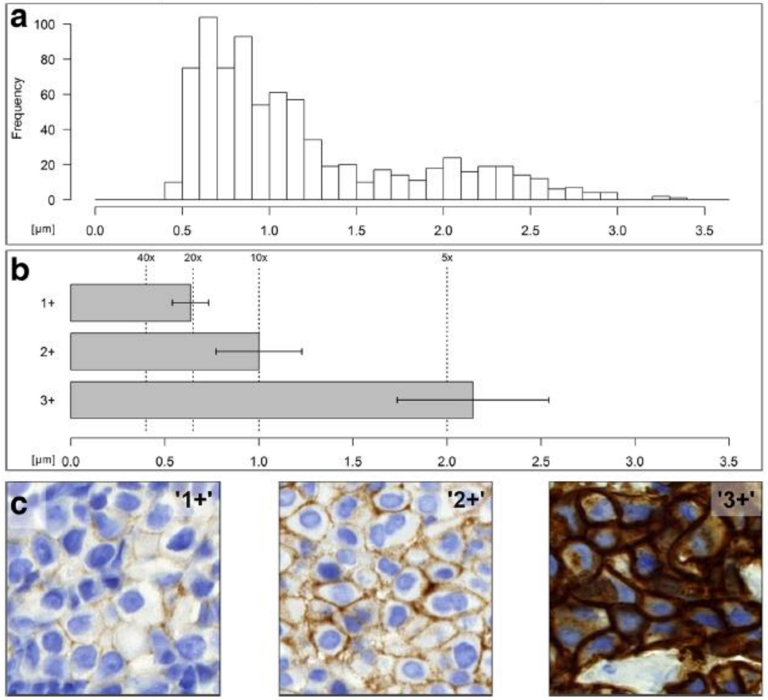


Fig. 1 Her2-IHC scoring categories reflect DAB-precipitate widths. Table: Microscope objectives have a fixed resolution that depends on the numerical aperture (Range: Values of common objectives). DAB-precipitates in HER2-IHC differ in width according to the intensity score. **a** histogram: Summary of 1200 DAB-precipitate-width measurements in µm. **b** bar chart: Mean DAB-width (bars) ±SD (antennae); resolution of standard microscope objectives (dashed lines). **c** images: Representative HER2-IHC stainings of invasive ductal breast carcinomas according to intensity score

Reference: Scheel AH, Penault-Llorca F, Hanna W, Baretton G, Middel P, Burchhardt J, Hofmann M, Jasani B, Rüschoff J. Physical basis of the 'magnification rule' for standardized Immunohistochemical scoring of HER2 in breast and gastric cancer. Diagn Pathol. 2018 Mar 12;13(1):19. doi: 10.1186/s13000-018-0696-x. PMID: 29530054; PMCID: PMC5848460.

IHC membrane marker intensity assessment

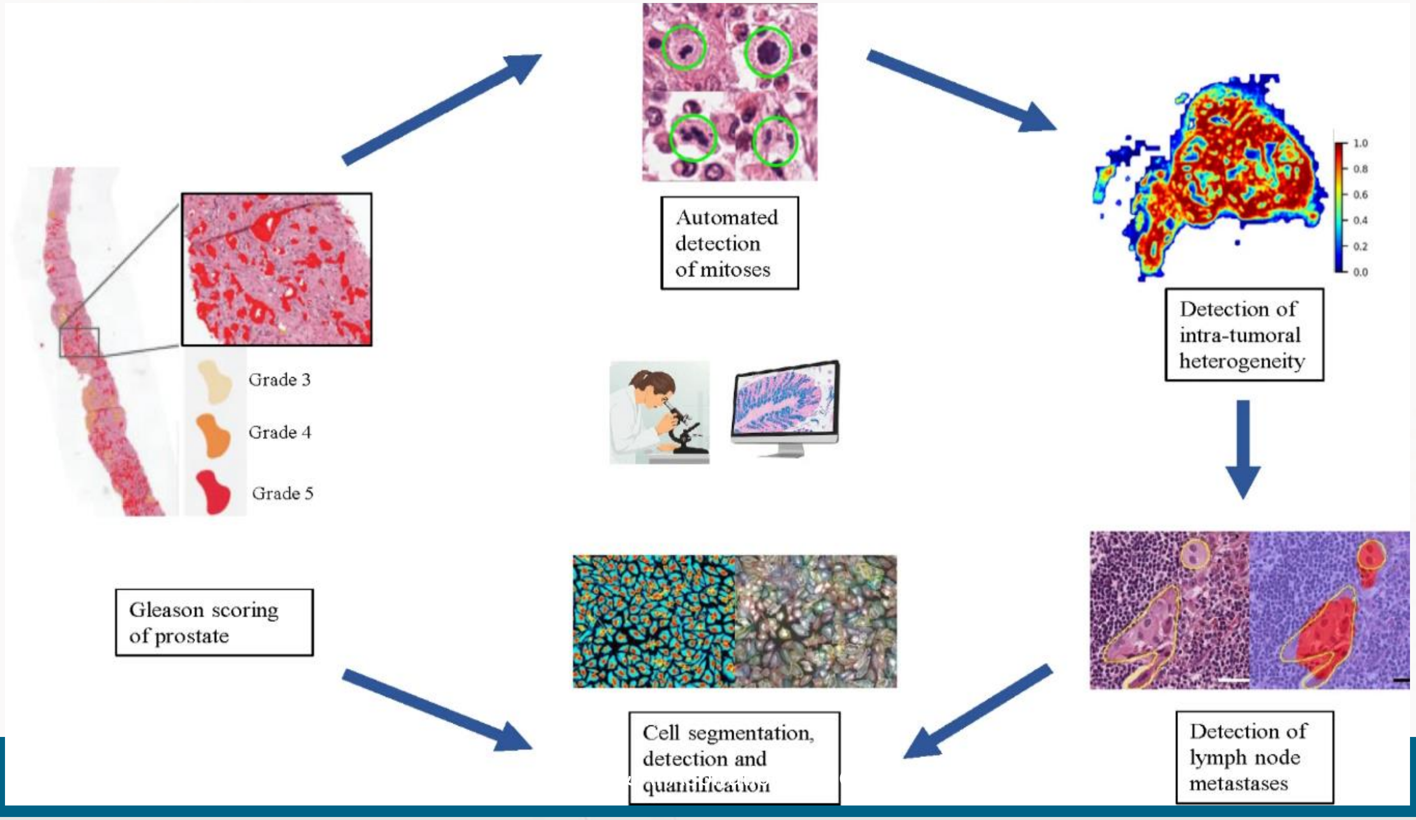
- How should we extend these magnification rules used for physical optical microscopes to whole slide images (WSI)?
  - Can the objective equivalents used on WSI viewers be applied?
  - What are the guidelines around membrane intensity interpretation/scoring for WSIs?
- Compare with controls of 1+, 2+, 3+ (e.g. HER2)
- Are biomarker automated image analysis more reliable since there is a morphometric component to this assessment?

What are the potential for artificial intelligence in digital pathology?

Current Application of Artificial Intelligence in Diagnostic Digital Pathology

- Applications remains assistive to help the pathologist with parameter assessment and diagnosis

Figure reference and source: Shafi S, Parwani AV. Artificial intelligence in diagnostic pathology. Diagn Pathol. 2023 Oct 3;18(1):109. doi: 10.1186/s13000-023-01375-z. PMID: 37784122; PMCID: PMC10546747.



Example: Prostate Cancer

- Increasing demand and workload for prostate cancer biopsy arising from prostate cancer screening
- Increasing number of biopsies per case, from 12 core transrectal biopsy to transperineal biopsy that combine both systematic and MRI targeted cores that readily generates 30 to 40 cores and beyond (at least 3 times increase in processing workload per case)
- Evolving refinement of assessment
  - Past: presence or absence of adenocarcinoma with overall Gleason score per case (4 to 6 data points/parameters each case)
  - Current: adenocarcinoma measurement of length and extent of involvement , Gleason score and grade grouping per core (i.e. at least 4 to 6 data points/parameters per core x 30 to 40 cores for each case)
- Role for artificial intelligence to assist with this increase in demand

Example: Prostate Cancer

- In 2021, FDA authorizes prostate cancer detection tool utilizing artificial intelligence on whole slide image (WSI)
- At least 5 commercially available products for prostate cancer detection on WSI
- Raciti P, et al. Clinical Validation of Artificial Intelligence-Augmented Pathology Diagnosis Demonstrates Significant Gains in Diagnostic Accuracy in Prostate Cancer Detection. Arch Pathol Lab Med. 2023 Oct 1;147(10):1178-1185. doi: 10.5858/arpa.2022-0066-OA. PMID: 36538386.
  - Evaluated a commercially available clinical grade AI solution using 610 prostate core needle biopsy WSIs from 218 institutions
  - 18 pathologists were involved in the study, including two genitourinary subspecialists, with or without assistance of the AI solution
  - Sensitivity of AI assisted reads increased by 8% (from 88.7% to 96.6%, $p < 0.001$)) (reduce errors of detection by 70%)
  - Specificity of AI assisted reads increased by 0.7% (from 97.3% to 98%, $p=0.02$)
  - Accuracy gains of 3.5% while accuracy loss is 0.5% with AI assisted reads. Accuracy gains all show concordant AI results.

Example: Prostate Cancer

- Eloy C, et al. Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch. 2023 Mar;482(3):595-604. doi: 10.1007/s00428-023-03518-5. PMID: 36809483; PMCID: PMC10033575.
  - Evaluated a commercially available clinical grade AI solution using 105 prostate core needle biopsy WSIs (matched confirmation with immunohistochemistry available)
  - Evaluated on 4 pathologists with or without assistance of the AI solution, only on H&E slides
  - Comparative performance with or without assistance of AI
  - However, with AI assistance, 20% fewer immunohistochemistry request (i.e. lower cost), 40% fewer second opinions and 30% less borderline results (i.e. increase diagnostic confidence), 20% less time for reading and reporting cases (i.e. improved turnaround time)

Example: Prostate Cancer

- Lami K, et al. Validation of prostate and breast cancer detection artificial intelligence algorithms for accurate histopathological diagnosis and grading: a retrospective study with a Japanese cohort. Pathology. 2024 Apr 19:S0031-3025(24)00101-6. doi: 10.1016/j.pathol.2024.02.009. Epub ahead of print. PMID: 38719771.Evaluated a commercially available clinical grade AI solution using 100 consecutive prostate core needle biopsy cases (697 H&E WSIs)
  - Had pretest technical validation done on 32 prostate core biopsy of various conditions (separate from 100 test biopsies), all of which were correctly classified. Preset algorithm settings were maintained.
  - Algorithm produced cancer probability score and indicate Gleason score above 3+3, with alerts generated if cancer was missed or gleason score were higher than original 3+3.
  - 5 cases originally diagnosed as benign (among 36 benign cases in total) has high cancer probability, and one of the case was revised to malignant diagnosis.
  - 4 cancer cases originally with Gleason 3+3 were noted to have higher Gleason score by algorithm and the Gleason score was revised in all these cases.
  - One Gleason 3+3 cancer case has a low cancer probability. (false negative)
  - Overall AUC was 0.969 in distringuishing between benign and ASAP/cancer cases.

Example: Breast Cancer

- Successful breast cancer screening programme with improvements in radiology imaging technology also increase the demand for pathology evaluation of breast core biopsies.
- Increasingly sensitive radiology imaging techniques with detection of distinct lesions in the breast, requiring pathology evaluation of multiple distinct lesions where present.
- Improvements in therapy (including neo-adjuvant setting) demand refinement of pathology evaluation beyond just presence or absence of neoplasia, including extent, grading and biomarker analysis used in patient selection.

Example: Breast Cancer

- Development of clinically deployable AI algorithm in breast biopsy have slightly lagged behind prostate biopsy due to greater tissue structural complexity and greater variety of lesions in breast.
- Sandbank J, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. NPJ Breast Cancer. 2022 Dec 6;8(1):129. doi: 10.1038/s41523-022-00496-w. PMID: 36473870; PMCID: PMC9723672.
    - Validated a currently commercially available AI solution on 841 WSIs (436 cases) across two institutions
        - Invasive carcinoma detection: 95.5% sensitivity, 93% specificity, AUC 0.99
        - In situ carcinoma detection: 93.2% sensitivity, 93.8% specificity, AUC 0.98
    - Deployed in clinical setting on 5,954 cases (12,031 WSIs) as second read, with detection of several misdiagnosed cases (4 described in the publication)

Example: Breast Cancer

- Lami K, et al. Validation of prostate and breast cancer detection artificial intelligence algorithms for accurate histopathological diagnosis and grading: a retrospective study with a Japanese cohort. Pathology. 2024 Apr 19:S0031-3025(24)00101-6. doi: 10.1016/j.pathol.2024.02.009. Epub ahead of print. PMID: 38719771.Evaluated a commercially available clinical grade AI solution using 100 breast core biopsy cases (258 H&E WSIs)
  - Had pretest technical validation done on 54 breast core biopsy of various conditions (separate from 100 test biopsies). Preset algorithm settings were maintained.
  - Algorithm produced invasive cancer and DCIS probability score and heatmap of a variety of significant histologic findings, with alerts generated if DCIS or invasive cancer was missed.
  - Among the 45 invasive cancer cases, all but one had high invasive cancer score. (the one outlier has medium invasive cancer score; none with low invasive cancer score).
  - 17 DCIS only cases had high DCIS score by algorithm. Two DCIS cases has high invasive cancer score as well, one of which was deemed uncertain for invasion and one case was maintained as DCIS.
  - Two intraductal papilloma case has high DCIS score by algorithm (false positive)
  - For distinguishing between invasive cancer from other diagnoses, AUC was 0.997; for distinguishing between DCIS from benign cases, AUC was 0.996

Example: Breast Cancer

- Challa B, et al. Artificial Intelligence-Aided Diagnosis of Breast Cancer Lymph Node Metastasis on Histologic Slides in a Digital Workflow. Mod Pathol. 2023 Aug;36(8):100216. doi: 10.1016/j.modpat.2023.100216. Epub 2023 May 12. PMID: 37178923.
    - Evaluated a commercially available algorithm for detection of metastasis in a total of 336 sentinel lymph nodes (SLNs) from 168 consecutive breast carcinoma resection specimens in clinical setting.
    - Detected all SLN metastasis in validation cohort of 234 SLNs (46 positive), showing 100% sensitivity, 41.5% specificity, 29.5% PPV, and 100% NPV (useful for screening out negative cases)
    - Consensus cohort of 102 SLN with matched immunohistochemistry (IHC) examined by 3 pathologists, with concordance between AI aided (H&E only) and IHC aided review of 99%, with less time consumed on AI aided review.
        - One false negative consensus conclusion on AI aided review was highlighted by AI solution as possible metastasis with moderate confidence, and positive focus is smaller than that seen on the level/section used for IHC.

Quantitative Image Analysis (QIA): HER2 IHC

- QIA on HER2 IHC is acknowledged as one of the modalities that can be used for diagnostic assessment
- Several 510(k) FDA-cleared and CE-IVD cleared QIA algorithms for HER2 IHC exist in the market
- QIA is reported to be an objective and reproducible tool for HER2 IHC scoring and may reduce equivocal score (may reduce need for ISH testing)

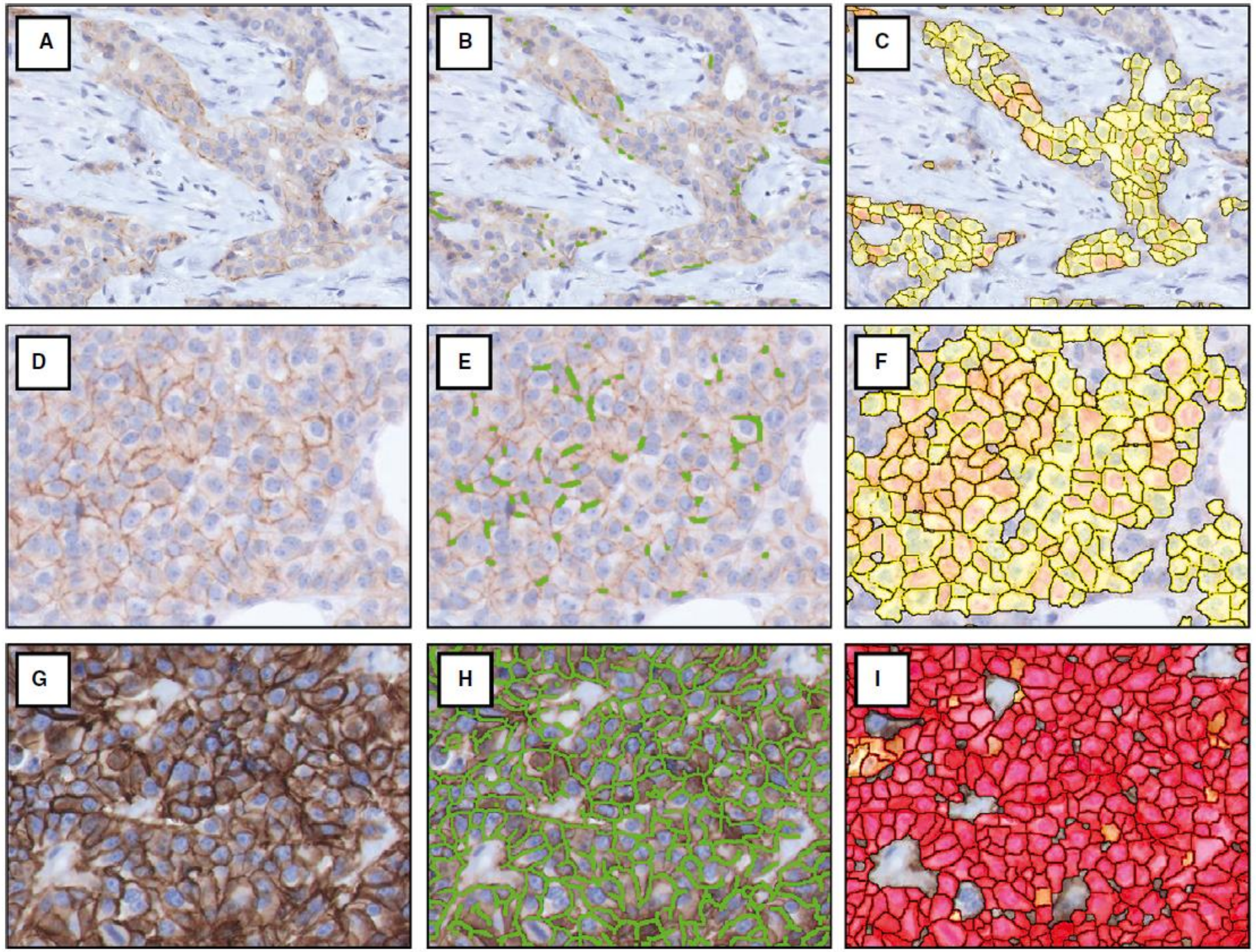


Figure 1. Digital image analysis of HER2 immunohistochemistry by two DIA platforms. HER2 score 1+ (A–C), score 2+ (D–F) and score 3+ (G–I). Images without DIA mark-up (left column), with DIA in platform A (middle column) and with DIA in platform B (right column). HER2, human epidermal growth factor 2; DIA, digital image analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure Source Reference: Koopman T, et al. What is the added value of digital image analysis of HER2 immunohistochemistry in breast cancer in clinical practice? A study with multiple platforms. Histopathology. 2019 May;74(6):917-924. doi: 10.1111/his.13812. Epub 2019 Apr 1. PMID: 30585668; PMCID: PMC6850320.

Bui MM, et al. Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer: Guideline From the College of American Pathologists. Arch Pathol Lab Med. 2019 Oct;143(10):1180-1195. doi: 10.5858/arpa.2018-0378-CP. Epub 2019 Jan 15. PMID: 30645156; PMCID: PMC6629520.

QIA HER2 IHC : Influence of Staining Quality

- Palm et al (2023) utilised a closed antibody clone, WSI scanning and image analysis solution for evaluation of HER2 expression in invasive carcinoma.
  - In the preliminary cohort of 475 IHC slides, where manual assessment comprise **181 score 0 (38.1%)**, 156 score 1+ (32.8%), **87 score 2+ (18.3%)**, 51 score 3+ (10.7%), AI IHC algorithm found **22 score 0 (4.6%)**, 137 score 1+ (28.8%), **254 score 2+ (53.5%)**, 62 score 3 (13.1%); overall concordance only 38.3%.
  - 184 of 475 (38.7%) of discrepancies may have significant impact on the additional testing or HER2 status.
  - Deviations in staining protocol from vendor’s recommended protocol (including incubation time with specific reagents and antibody, counterstaining time) were identified to be a key factor for the discrepancy and were tuned and restained for the subsequent study cohort (only 4.5% of study cohort show discordance with significant diagnostic impact)

Palm C, etc. Determining HER2 Status by Artificial Intelligence: An Investigation of Primary, Metastatic, and HER2 Low Breast Tumors. Diagnostics (Basel). 2023 Jan 3;13(1):168. doi: 10.3390/diagnostics13010168. PMID: 36611460; PMCID: PMC9818571.

QIA HER2 IHC : Algorithm and Calibration

- Koopman et al (2019) compared two commercially available QIA for evaluation of HER2 expression in 152 primary invasive breast carcinomas from consecutive resection specimens scanned into WSI
  - QIA Platform A was not further calibrated, QIA platform B was calibrated (calibration done using 20 cases that are not part of study cohort, calibration optimised colour classification, cell classification, membrane detection, membrane completeness)
  - HER2 QIA applied to 3 annotated representative areas of 1.5mm^2; highest HER2 score among annotated area is concluded as HER2 score for the case
  - Comparison between QIA and manual scoring, between both QIA, and between QIA with "standard diagnostics" results (interplatform agreement using linear weighted kappa statistics; comparison with "standard diagnostics" using sensitivity, specificity, positive predictive value [PPV] and negative predictive value [NPV])
    - Standard diagnostics refer to detection of HER2 positive: IHC 3+ and ISH positive 2+ cases
  - Manual scoring: 114 score 0 and 1+ (75%), **26 score 2+ (17.1%)**, 12 score 3+ (7.9%)
  - Platform A: 139 score 0 and 1+ (91.4%), **2 score 2+ (1.3%)**, 11 score 3+ (7.2%)
  - Platform B: 114 score 0 and 1+ (75%), **24 score 2+ (15.8%)**, 14 score 3+ (9.2%)
  - Manual scoring vs QIA: "Moderate" agreement for platform A (kappa = 0.6, 82.9%) and "almost perfect" agreement for platform B (kappa = 0.85, 91.1%)
  - Platform A vs Platform B: "Moderate" interplatform agreement (kappa = 0.58, 62.2%)
  - Standard diagnostics vs QIA: Platform A 81.3% sensitivity, 100% specificity, 100% PPV, 97.8% NPV and three false-negative cases; Platform B 100% sensitivity, specificity, PPV and NPV.
  - QIA comparable to manual scoring with potential to improve scoring reproducibility, but algorithm with less 2+ cases may have false negative cases. (effect of calibration?)

Koopman T, et al. What is the added value of digital image analysis of HER2 immunohistochemistry in breast cancer in clinical practice? A study with multiple platforms. Histopathology. 2019 May;74(6):917-924.

QIA HER2 IHC: "HER2 Low" Assessment

- Wu et al (2023) examined impact of custom built AI-assisted interpretation of HER2 expression (focusing on score 0 vs 1+) with an augmenting reality module attached to a microscope
  - Manual consensus scoring assessment by 2 or 3 experienced pathologists used as gold standard
  - Involved 15 pathologists of varying experiences, two round ring study of HER2 expression in 246 invasive ductal carcinoma, NOS. (120 score 0, 126 score 1+) with 2 week wash out period between manual and AI assisted read
  - For AI assisted read, selected 5 representative typical visual fields under 40 x objective (after examination of whole slide at low power)
  - Accuracy of assessment improved with AI assisted read compared to manual scoring (0.93 vs 0.8), with greatest improvement of accuracy seen among junior pathologists.
  - Consistency of AI assessment measured by intraclass correlation coefficient (ICC) improved with AI assisted read compared to manual scoring (0.812 vs 0.542)
  - Limitations:
    - AI assisted scoring is based on selected field of views (FOVs) which can be subjective
    - AI model cannot automatically differentiate between in situ and invasive carcinomas (need to select correct FOVs or manually annotate regions of interest to limit assessment to invasive carcinoma)

Wu S, et al. The Role of Artificial Intelligence in Accurate Interpretation of HER2 Immunohistochemical Scores 0 and 1+ in Breast Cancer. Mod Pathol. 2023 Mar;36(3):100054.

QIA HER2 IHC : “HER2 Low” Assessment

- Jung et al (2022), presented an abstract on validation study of AI assisted analysis of HER2 expression across all scores.
  - Validation set of 209 WSIs of HER2 IHC assessed by 3 pathologists, with AI assisted reading done where AI model is discrepant from pathologist assessment without AI.
  - Main aim is to study consistency and interobserver variability of assessment, as measured by Fleiss kappa value
  - Fleiss kappa value improved from 0.512 without AI assistance, to 0.762 with AI assistance across all scores. (Improvement from 49.3% concordant to 74.6% concordant)
  - Specific to score 1+ assessment, Fleiss kappa value improved from 0.242 without AI assistance, to 0.687 with AI assistance. (Improvement from 25.7% concordant to 68.9% concordant)

Jung,et al. [Artificial intelligence-powered human epidermal growth factor receptor 2 (HER2) analyzer in breast cancer as an assistance tool for pathologists to reduce interobserver variation.]
Journal of Clinical Oncology 2022 40:16_suppl, e12543

Example: Gastric Biopsy

- Identification of *Helicobacter pylori (HP)* and intestinal metaplasia in gastric biopsy is one of the most common tasks for pathologists.
- As much as 20% of the histopathology laboratory workload may be related to gastric biopsy. (personal experience)
- *HP* may produce inflammatory response that can be a clue to possible presence of *HP*, although presence of *HP* must be confirmed on routine histopathology, or, in some cases, via special stains (Giemsa, Warthin Starry) or immunohistochemistry.
  - Special stains and immunohistochemistry may be associated with added cost.

Example: Gastric Biopsy

- Lin et al. Two-tiered deep-learning-based model for histologic diagnosis of Helicobacter gastritis. Histopathology. 2023 Nov;83(5):771-781. doi: 10.1111/his.15018. Epub 2023 Jul 31. PMID: 37522271. Developed a two-stage model to identify *HP* related gastritis and localise *HP* organisms
  - Utilise weakly supervised enhanced streaming convolutional neural network (ESCNN) methodology at slide level to identify HP related inflammatory response on whole slide images (885 WSIs training of which HP-positive in 439 and 190 test dataset, all at 40x objective magnification)
    - 190 test dataset was confirmed for HP positivity vs HP negativity via IHC
  - For auxillary model for HP localisation, use 9 WSI with 824 annotated HP loci as training, and 3 HP positive and 2 HP negative WSIs as test dataset.
    - Used traditional machine learning methods rather than deep learning to minimise overfitting.
  - Sensitivity and specificity for HP related gastritis is 93.3% and 90.1% respectively
    - Compared to pathologist sensitivity and specificity of 93.3% and 84.2% respectively
  - Exact localisation of HP has an average precision of 0.5796
    - Note that we do not have to identify all HP foci for diagnosis in clinical practice

What are the considerations in implementing artificial intelligence in digital pathology?

Machine Learning in Whole Slide Images

- High resolution, large dimension and large file size of whole slide images (WSIs) remains challenging
- Typically managed as smaller patches with less pixels for machine learning and inference

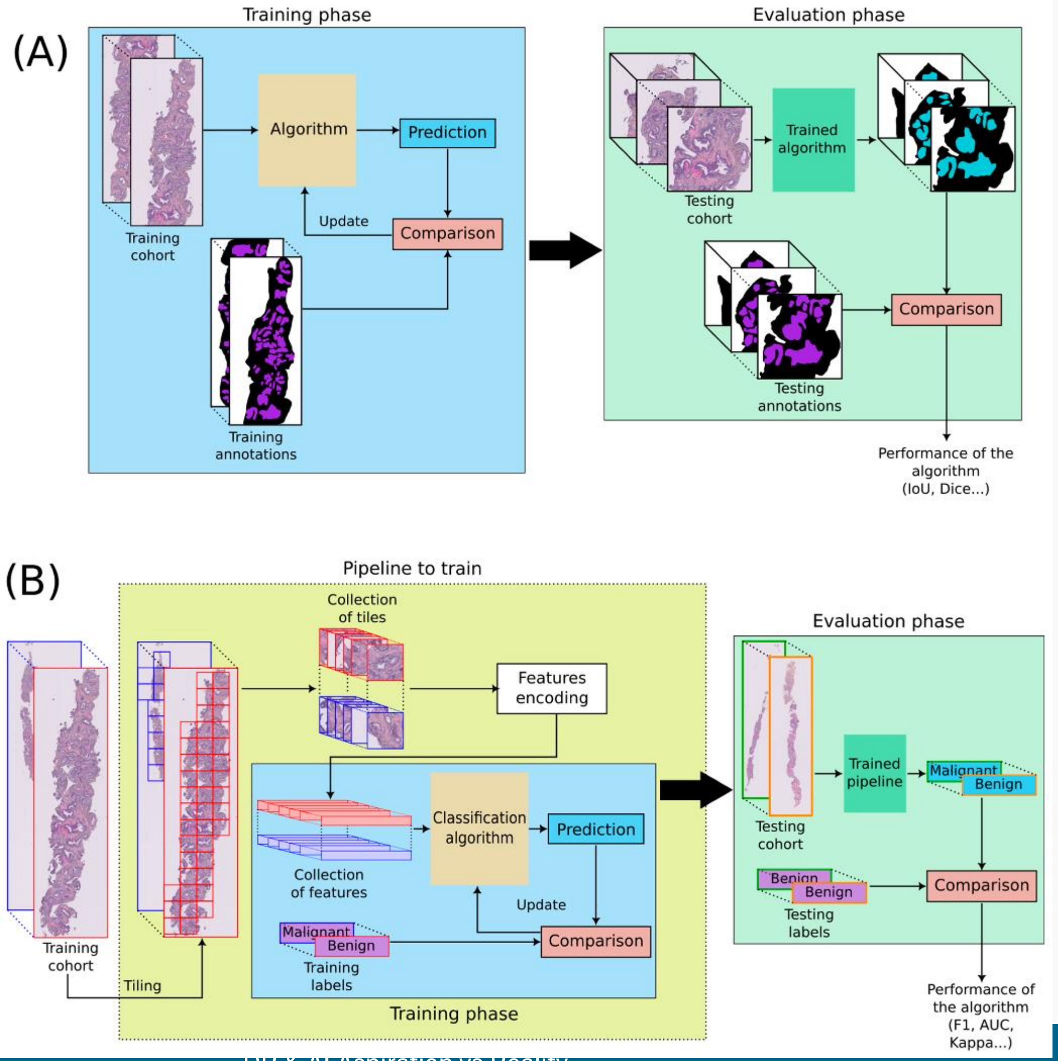


Figure Reference and Source: Rabilloud N, Allaume P, Acosta O, De Crevoisier R, Bourgade R, Loussouarn D, Rioux-Leclercq N, Khene ZE, Mathieu R, Bensalah K, Pecot T, Kammerer-Jacquet SF. Deep Learning Methodologies Applied to Digital Pathology in Prostate Cancer: A Systematic Review. Diagnostics (Basel). 2023 Aug 14;13(16):2676. doi: 10.3390/diagnostics13162676. PMID: 37627935; PMCID: PMC10453406.

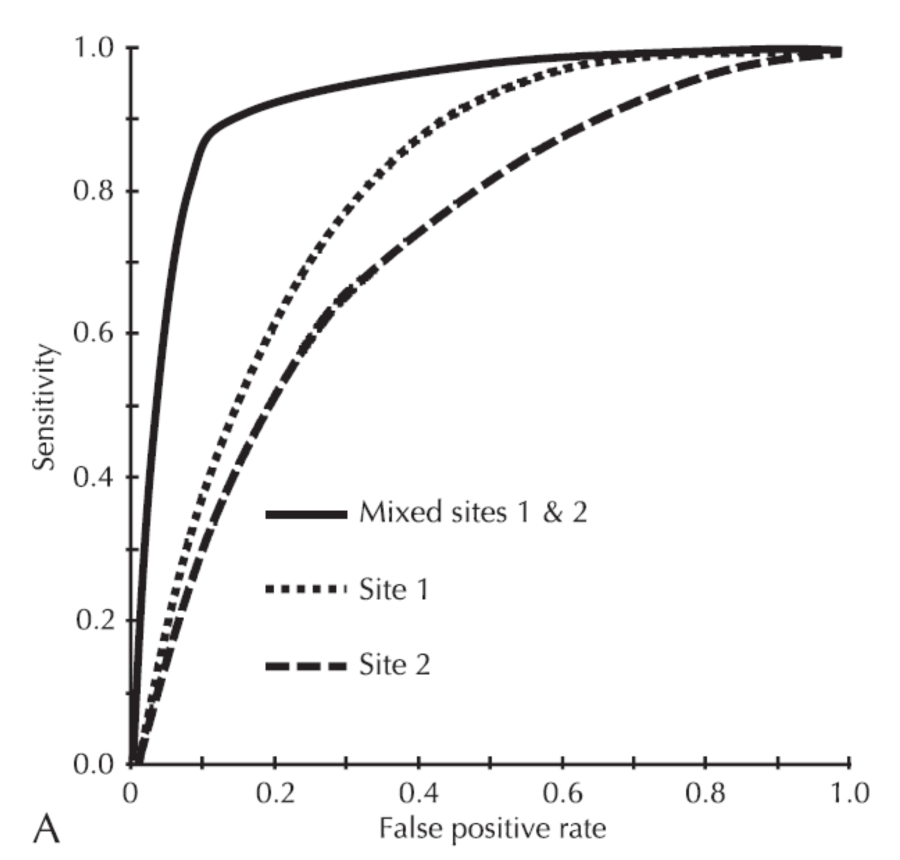
AI Generalisability in DP

- Generalisation of AI algorithm requires measuring performance of similar dataset outside the development site.
- Poor generalisability means algorithm cannot perform well outside data from development site.
  - Many of the AI algorithm are developed using data from one site only.
  - Model development should encompass data from a variety of sources.
- Aggregating data from multiple sources may partially mitigate the problem but performance of algorithm may not be optimal for site specific data compared to training on own site data.
- Other strategies to mitigate include process standardization and image normalization.
- Local verification and tuning (including e.g. transfer learning) of AI algorithm developed elsewhere may be important steps in AI adoption.

Reference: Harrison JH, et al. Introduction to Artificial Intelligence and Machine Learning for Pathology. Arch Pathol Lab Med. 2021 Oct 1;145(10):1228-1254.

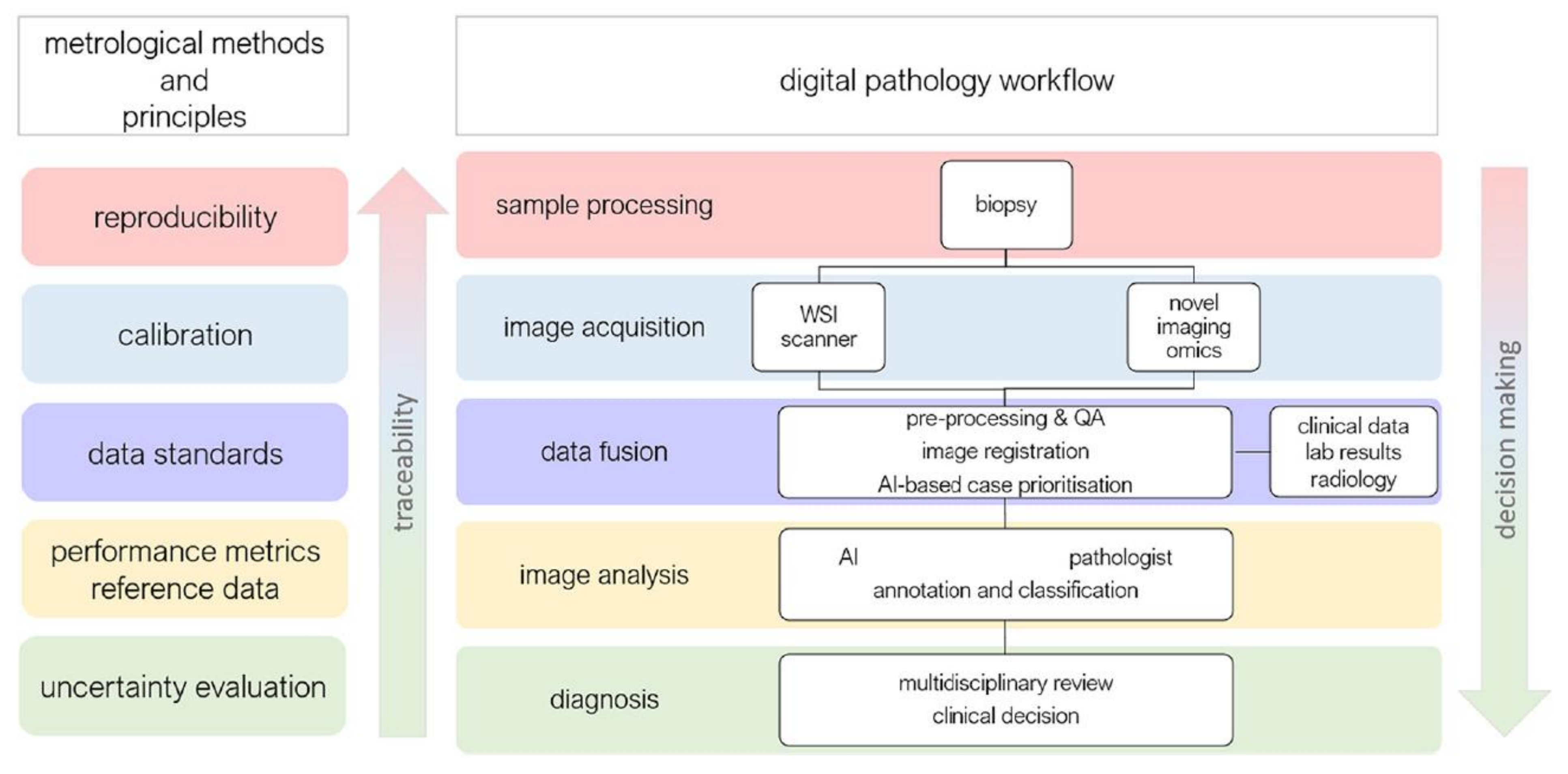
AI Generalisability in DP

- Training from mixed set of data from multiple site may perform well on mixed data, but sub-optimally on site specific data



Reference: Harrison JH, et al. Introduction to Artificial Intelligence and Machine Learning for Pathology. Arch Pathol Lab Med. 2021 Oct 1;145(10):1228-1254.

Considerations of Areas for Standardization for Better AI Application in DP



Reference: Romanchikova M, et al. The need for measurement science in digital pathology. J Pathol Inform. 2022;13:100157. doi: 10.1016/j.jpi.2022.100157. Epub 2022 Nov 10. PMID: 36405869; PMCID: PMC9646441.

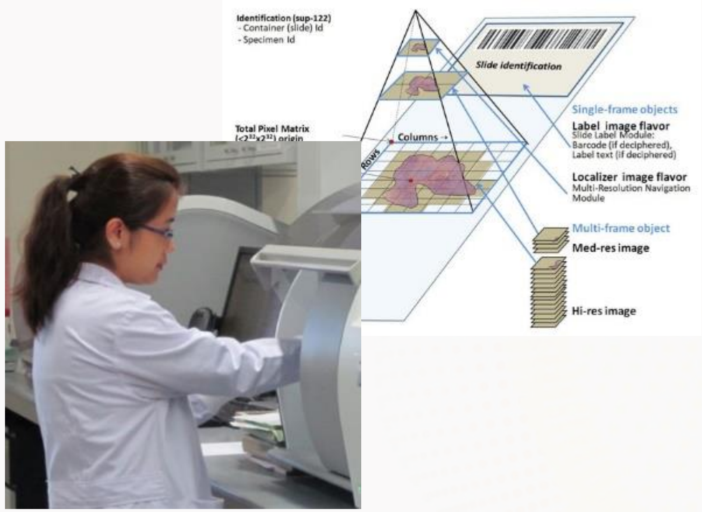
Considerations for Standardization: AI in DP

Specimen/Slide Processing/Preparation (Input Quality)

- Tissue size and placement
- Tissue section thickness
- Stain protocol (including tissue fixation requirement)
- Slide quality assessment (QC/QA)

Image Acquisition and Metadata Linkage

- Image resolution and file encoding/format
- Metadata field definition and terminology standards (bringing meaning to image)
- Interoperability standards (e.g. DICOM), including between scanning equipments, image management and AI solution
- Day-to-day image quality control (QC/QA)



Considerations for Standardization: AI in DP

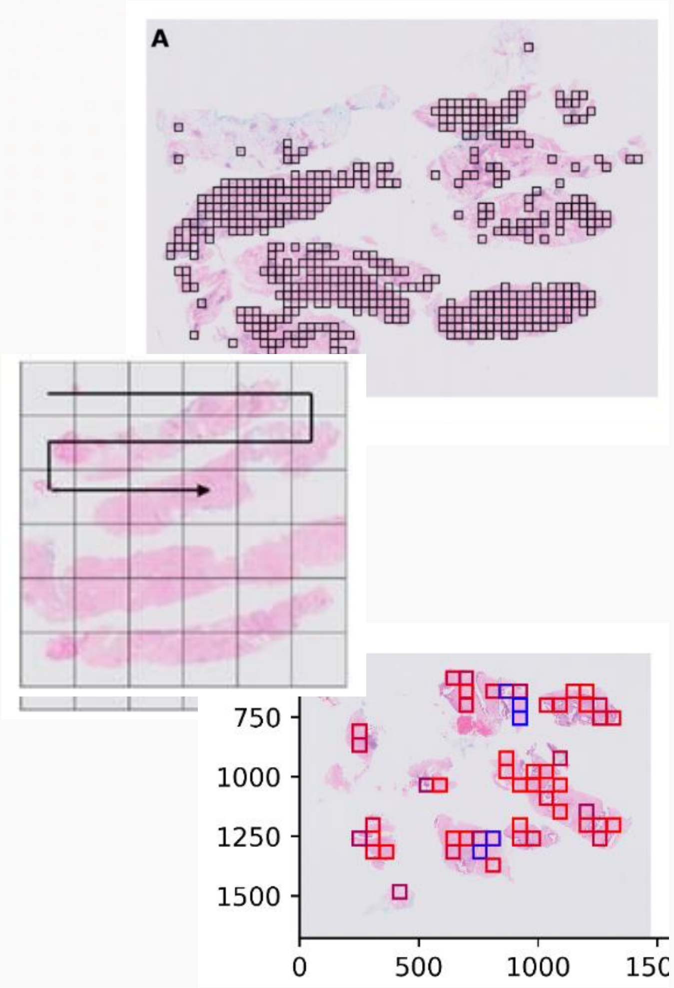


Image Analysis/AI Algorithm Development and Deployment

- Annotation and its linkage with image, including transferability and reusability (interoperability), as well as terminology standards against features
- Optimise quality of inputs prior to analysis, e.g. filtering of slide preparation and image artefacts (e.g. folded tissue, entrapped air bubbles in slides, out of focus areas)
- Colour normalisation and data augmentation
- Solution optimisation for site specific adoption, including e.g. transfer learning
- Validation protocol against clinical samples
- Output format (e.g. heatmaps), explainability, information and terminology standards

What is the future for diagnostic use of artificial intelligence in digital pathology?

What should we expect in the foreseeable future?

- Berbís MA, et al. Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. EBioMedicine. 2023 Feb;88:104427.
  - Collected perspectives from 24 subject matter experts: Role of AI in pathology by 2030
  - Consensus opinion were attained in 141 survey items (out of 180 items, ~78%)

What should we expect in the foreseeable future?

Examples of high consensus statements

By 2030, the probability of these AI tools being routinely used in pathology labs is ...

AI application	Item #	Mode (%)	Mean (SD)	Median (IQR)	Likelihood
Identification of micrometastases	78	7 (50.0)	6.17 (1.09)	6.5 (6.0–7.0)	Certain
Detection of lymph node metastases	79	7 (54.2)	6.33 (0.87	7.0 (6.0–7.0)	Certain
Quantification of IHC or IF stains, such as Ki-67, ER, PgR, PD-L1	85	7 (70.8)	6.67 (0.56)	7.0 (6.0–7.0)	Certain
Quantification of number of mitoses in H&E-stained images	86	7 (50.0)	6.33 (0.76)	6.5 (6.0–7.0)	Certain
Counting lymphocytes	87	7 (50.0)	6.42 (0.65)	6.5 (6.0–7.0)	Certain
Automated ordering of IHC for specific applications/assisting with selection of immunohistochemical stains needed	61	6 (45.8)	5.46 (0.93)	6.0 (5.0–6.0)	Very likely
Automated QA/QC of IHC positive and negative controls	62	6 (54.2)	5.75 (0.90)	6.0 (5.0–6.0)	Very likely
Proposing specific IHC or other molecular methods to solve a specific diagnostic problem	68	6 (41.7)	5.17 (1.34)	5.5 (5.0–6.0)	Very likely
Prioritization of cases (such as cases with neoplasia and infectious organisms in immunosuppressed patients)	69	6 (45.8)	5.50 (1.10)	6.0 (5.0–6.0)	Very likely
Quality control of whole-slide images (scanning process), and detection of poor-quality slides (tissue folds, poor staining)	73	6 (66.7)	6.13 (0.68)	6.0 (6.0–6.5)	Very likely

Berbís MA, et al. Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. EBioMedicine. 2023 Feb;88:104427.

What should we expect in the foreseeable future?

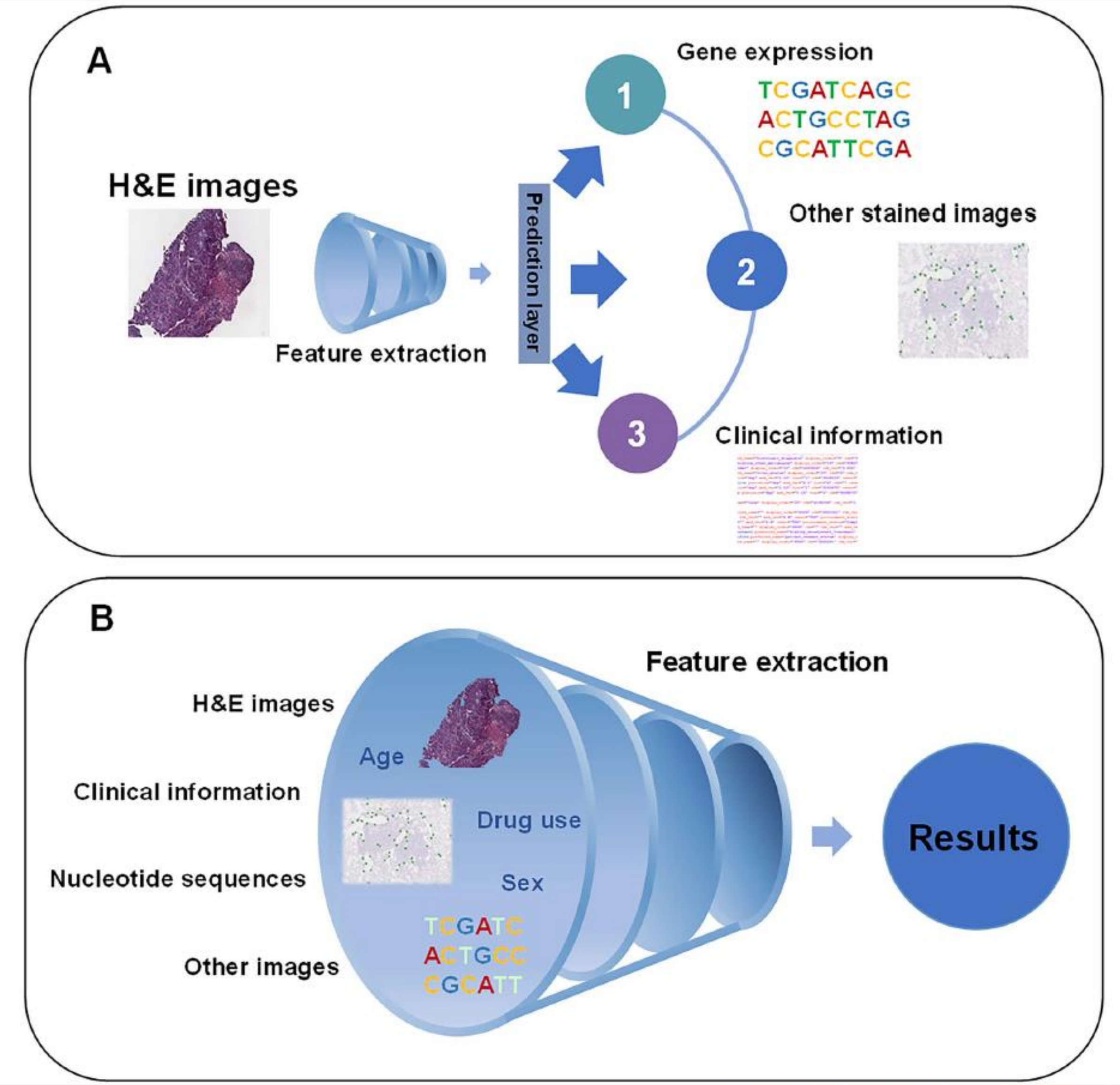
Examples of high consensus statements

By 2030, the probability of these tasks being fully delegated to AI in pathology labs is ...

Task	Item #	Mode (%)	Mean (SD)	Median (IQR)	Likelihood
Verification of positive and negative controls for IHC	124	6 (58.3)	5.71 (0.91)	6.0 (5.0–6.0)	Very likely
Prioritization of cases	125	6 (50.0)	5.54 (1.47)	6.0 (5.0–6.0)	Very likely
Triage of cases to appropriate pathologists	126	6 (45.8)	5.46 (1.25)	6.0 (5.0–6.0)	Very likely
Contextual data lookup on patients from the EHR relevant to the pathology case being reviewed	127	6 (50.0)	5.25 (1.15)	6.0 (5.0–6.0)	Very likely
Slide QC (e.g., detection of tissue folds and tears, stain quality evaluation, etc.)	128	6 (58.3)	5.88 (1.03)	6.0 (6.0–6.0)	Very likely
Screening for microorganisms, such as AFB and *H. pylori*	129	6 (58.3)	5.96 (0.75)	6.0 (6.0–6.0)	Very likely

Berbís MA, et al. Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. EBioMedicine. 2023 Feb;88:104427.

Multimodal AI



Qiao Y, at al. Multi-modality artificial intelligence in digital pathology. Brief Bioinform. 2022 Nov 19;23(6):bbac367. doi: 10.1093/bib/bbac367. PMID: 36124675; PMCID: PMC9677480.

Multimodal AI

Targeted Oncology

NEWS ▾ CONFERENCES ▾ MEDIA ▾ PUBLICATIONS ▾ CME/CE RESOURCES ▾ SUBSCRIBE

Choose Specialty ▾

Spotlight - Biomarker-Driven Lung Cancer | GIST | HER2-Positive Breast Cancer | Chronic Lymphocytic Leukemia | Small Cell Lung Ca

Advertisement

MJH life sciences knowledge that matters

Improving patient lives is not what we do, it is who we are.

NCCN Recommends First AI Prognostic Tool in Prostate Cancer

March 5, 2024

Sabrina Serani

News Article

The National Comprehensive Cancer Network has included ArteraAI in its clinical practice guidelines as a predictive test for localized prostate cancer.

Conclusion

- Implementing digital pathology for primary diagnosis is possible but careful attention to upstream and downstream workflows and clinical validation, at both the lab and individual pathologist level, is key to success
- Digital pathology is an evolving field and some issues remains to be addressed.
- The use of artificial intelligence on whole slide images is likely an inevitable progression, although issues with implementation and generalisability requires further study.
- Multimodal artificial intelligence model with inclusion of whole slide images is progressively being developed and its clinical potential recognised.

Thank You

Dr Chee Leong **Cheng**