

Comparative Assessment of Digital Pathology Systems for Primary Diagnosis

Sathyanarayanan Rajaganesan¹, Rajiv Kumar¹, Vidya Rao¹, Trupti Pai¹, Neha Mittal¹, Ayushi Sahay¹, Santosh Menon¹, Sangeeta Desai¹

¹Department of Pathology, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, India

Submitted: 17-Oct-2020

Revised: 09-Dec-2020

Accepted: 14-Jan-2021

Published: 09-Jun-2021

Abstract

Background: Despite increasing interest in whole-slide imaging (WSI) over optical microscopy (OM), limited information on comparative assessment of various digital pathology systems (DPSs) is available. **Materials and Methods:** A comprehensive evaluation was undertaken to investigate the technical performance—assessment and diagnostic accuracy of four DPSs with an objective to establish the noninferiority of WSI over OM and find out the best possible DPS for clinical workflow. **Results:** A total of 2376 digital images, 15,775 image reads (OM - 3171 + WSI - 12,404), and 6100 diagnostic reads (OM - 1245, WSI - 4855) were generated across four DPSs (coded as DPS1, 2, 3, and 4) using a total 240 cases (604 slides). Onsite technical evaluation revealed successful scan rate: DPS3 < DPS2 < DPS4 < DPS1; mean scanning time: DPS4 < DPS1 < DPS2 < DPS3; and average storage space: DPS3 < DPS2 < DPS1 < DPS4. Overall diagnostic accuracy, when compared with the reference standard for OM and WSI, was 95.44% (including 2.48% minor and 2.08% major discordances) and 93.32% (including 4.28% minor and 2.4% major discordances), respectively. The difference between the clinically significant discordances by WSI versus OM was 0.32%. Major discordances were observed mostly using DPS4 and least in DPS1; however, the difference was statistically insignificant. Almost perfect ($\kappa \geq 0.8$)/substantial ($\kappa = 0.6-0.8$) inter/intra-observer agreement between WSI and OM was observed for all specimen types, except cytology. Overall image quality was best for DPS1 followed by DPS4. Mean digital artifact rate was 6.8% (163/2376 digital images) and maximum artifacts were noted in DPS2 ($n = 77$) followed by DPS3 ($n = 36$). Most pathologists preferred viewing software of DPS1 and DPS2. **Conclusion:** WSI was noninferior to OM for all specimen types, except for cytology. Each DPS has its own pros and cons; however, DPS1 closely emulated the real-world clinical environment. This evaluation is intended to provide a roadmap to pathologists for the selection of the appropriate DPSs while adopting WSI.

Keywords: Comparative assessment, digital pathology, digital pathology systems, primary diagnosis, validation, whole-slide imaging

INTRODUCTION

Whole-slide imaging (WSI) allows the digitization of an entire glass slide to produce a digital image, which can be maneuvered and navigated akin to using conventional optical microscopy (OM). Evolving technological advancements, reduced costs, and regulatory approval for WSI systems have paved the way for digital pathology (DP) to move from research and education to routine diagnostic workflow.^[1-3] All these developments justify the feasibility of replacing microscopes with DP systems (DPSs) in the near future. However, this transition from a glass slide to a digital environment presents considerable logistic and organizational challenges.^[2,3] At present, most of the literature related to noninferiority of DP in clinical practice is primarily focused on diagnostic accuracy for primary diagnosis by WSI compared to that by OM.^[4-12]

However, the confounding factors due to different DPSs and specimen types have been inadequately addressed.

DP is a product of complex multistep processes, involving technical (scanner capabilities, software and hardware for viewing and archival of the slides), clinical (specimen type), and organizational (training and pathologist's expertise in DP and institutional information technology [IT] support) factors. As a result, the same tissue sample might appear different

Address for correspondence: Dr. Rajiv Kumar,
Department of Pathology, Tata Memorial Hospital, Homi Bhabha National
Institute, Mumbai, Maharashtra, India.
E-mail: rajiv.kaushal@gmail.com

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Rajaganesan S, Kumar R, Rao V, Pai T, Mittal N, Sahay A, *et al.* Comparative assessment of digital pathology systems for primary diagnosis. *J Pathol Inform* 2021;12:25.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2021/12/1/25/318091>

Access this article online

Quick Response Code:



Website:
www.jpathinformatics.org

DOI:
10.4103/jpi.jpi_94_20

when scanned on different DPSs and assessed using different viewing software. Further, with increasing availability of various DPSs in the market, it is difficult for a pathologist to select an appropriate technology. Thus, understanding all these technical parameters is extremely relevant for adoption of DP, as they might impact digital image quality, workflow, and pathologist's diagnostic capabilities.^[13]

Hence, to facilitate suitable decisions and judicious investments related to the selection of appropriate DP technology, a comprehensive evaluation was undertaken to investigate the comparative technical performance and diagnostic concordance of different DPSs versus conventional OM. We included various specimen types (biopsy, resection specimen, frozen, immunohistochemistry [IHC], and cytology) encountered in routine sign-out for primary diagnosis to test the sustainability of a DP-based diagnostic workflow in a subspecialty setting at a high-volume, tertiary care oncology center.

The aim of this study was to establish the noninferiority of WSI over conventional OM for primary diagnosis and find out the best possible DPS for adopting DP for routine diagnosis.

MATERIALS AND METHODS

This blinded retrospective observational study was performed at a tertiary care oncology center, using archival diagnostic material, following approval from the institutional ethics committee. The vendors of four DPSs consented for the evaluation of their scanners in the study.

Case enrolment

The glass slides of previously reported cases (between June 1, 2018, and September 1, 2018) were retrieved from the archives to represent a standard set of cases encountered in routine practice at our institution. Out of 300 cases initially screened, 240 cases were shortlisted by the enrolment pathologist for evaluation. In each specimen category, i.e., biopsies, resection specimens, cytology (fine-needle aspiration cytology and exfoliative cytology), and frozen sections (intraoperative consultation), 60 cases were included for the evaluation in accordance with CAP recommendations for validation of DP. Each specimen category further included 15 cases from four organ systems – breast, thoracic, gastrointestinal tract, and genitourinary tract (GUT). Additional 10 cases were included as the training set, only to familiarize the participating pathologist with respective DPSs, and were not included in the actual analysis.

Key glass slides were selected from each case by an enrolment pathologist. The glass slides were chosen to adequately represent all the data elements in the standard synoptic format, in addition to primary diagnosis and immunostains, wherever necessary. The glass slides were anonymized and assigned unique study identification codes.

Cases were excluded if they met any of the following criteria: (1) clinical information was not available, (2) selected slides either contained indelible markings or if broken/cracked,

and (3) complex, difficult, and rare cases to reduce the bias, as pathologists often have memory for such cases for a long time. A set of 604 glass slides representing 240 cases formed the study cohort.

Slide scanning

The selected slides were scanned on four WSI scanners after ensuring appropriate calibration and quality control measures. To keep the identity of these respective DPSs confidential, they were anonymized as DPS1, DPS2, DPS3, and DPS4. Although the technical specifications of these DPSs, as well as monitors used were recorded, these details cannot be disclosed. The expert technical personnel provided by the respective vendor scanned all the slides. This was done to ensure that scanning capabilities of DPSs should not get confounded by experience of in-house technical staff handling scanners (operator factor).

Onsite technical evaluation of digital pathology systems capability

Onsite technical scanning capability of each DPS was assessed with respect to the scanning of different types of slides (versatility), successful scanning rate (number of first time right scan and rescans), scanning speed (scan time per slide), and image size for each case. The possible reasons for failed scans were recorded. Cases were subsequently evaluated independently using associated Image Viewer Software (IVS), and display monitors at the workstations were provided by the respective vendors. The images were evaluated using the IVS installed on the provided desktop workstation in three DPSs, while web-based application software was used in one DPS. The monitors used were of minimum display resolution 1920 × 1200 (for screen size of 21 inch at 2 MP). Both scanners and monitors of all the DPS were used in default settings.

Case interpretation

The diagnostic evaluation was performed by five pathologists at various stages of their career in the field of diagnostic pathology, with a median of 10 years (range 2–12 years). All readers had some experience using DP, with a median of 1 year (range 0.5–3 years).

In addition, two specialist pathologists also reviewed biopsy specimens of two sites (breast and GUT). Each case was independently assessed five times by every pathologist once using OM and on four DPSs (using respective IVS on compatible display monitor). Each diagnosis by a reading pathologist on a case (whether by WSI or OM) was termed a “read.” Hence, there were 25 “reads” per case, besides the reference (sign-out) diagnosis.

The clinical information was provided to all participating pathologists in electronic formats.

Additional requests for recuts, special stains, and immunostains beyond those included in the study and second opinion were not entertained. Participating pathologists were blinded to the original sign-out diagnosis and also to their own prior impressions. To reduce the recall bias, a minimum washout

period of 2 weeks was observed between successive reporting sessions, and the cases were assigned to the pathologists in a random order.

Diagnostic interpretation included primary diagnosis (top-line diagnosis) along with all data elements (IHCs and special stains) conforming to standard synoptic reporting format was recorded. Diagnostic assessment time for achieving the diagnosis for each case was recorded for both the modalities and was compared for each platform for individual pathologist.

Assessment of diagnostic concordance between whole-slide imaging versus optical microscopy

The original sign-out diagnosis was considered as the reference standard. All discordances between reference diagnoses and OM diagnoses were reviewed independently by subspecialty pathologists, and final reference standards were re-established for analysis. The top-line diagnosis rendered either on WSI or OM by participating pathologists was compared against this reference diagnosis for evaluating the diagnostic concordance [Figure 1]. Diagnostic discrepancies were classified as major and minor discordances (based on the clinical impact) and analyzed separately as per type of specimens based on the criteria's adopted from Li *et al.*^[14]

Assessment of image viewing software, digital image quality, and level of confidence

The IVS used and the image quality for each case on respective DPS were assessed on a scale of 1–3, where 1 represented the worst, 2 average, and 3 best quality. The IVS was evaluated based on an average score of following parameters: overall appearance, ease of navigation, arrangement of cases, panning/zooming, annotation tools, photography quality, and ability to open multiple slides of the same case. Similarly, image quality was assessed based on the brightness, color contrast, color rendition index, i.e., how close it resembled the glass slides, uniformity of the scanning focus, completeness of the sample captured, and any digital artifacts. Digital artifacts including the type of artifacts were recorded for each DPS.

Based on the digital image quality, the level of confidence for reporting for each case and respective DPS was scored on the

scale of 1–3, wherein 1 represented low, 2 average, and 3 high level of confidence. Similarly, level of confidence for diagnosis on OM was also recorded.

Statistical analysis

All statistical analysis was performed by using IBM SPSS Statistics for Windows, Version 25 (IBM Corp., Armonk, New York, USA). The number and percentage of concordances and minor and major discordances by OM and WSI were calculated to determine the accuracy rate. To establish the noninferiority of WSI over OM, the cutoff criteria of <4% as proposed by Bauer *et al.* were adopted.^[5] Inter and intra observer agreement between OM and WSI was estimated through unweighted kappa statistics [Figure 1]. Based on the Landis and Koch guidelines, κ (kappa) values were interpreted as 0–0.2 representing poor agreement, 0.2–0.4 fair, 0.4–0.6 moderate, 0.6–0.8 substantial, and ≥ 0.8 perfect agreement.^[15]

RESULTS

The present study cohort included 240 cases for evaluation and was composed of total 604 slides, i.e., 425 surgical pathology slides (228 - hematoxylin and eosin [H & E], 188 - IHC, and 9 - special stains), 94 cytology slides (61 - Papanicolaou and 33 - May–Grunwald–Giemsa), and 85 frozen section slides (55 - H and E and 30 - toluidine blue slides). A total of 2376 digital images were generated across four DPSs (excluding 40 failed scans). A total of 6100 diagnostic reads (OM - 1245, WSI - 4855) were obtained finally, based on 15,775 image reads (OM - 3171 + WSI - 12,404) by seven evaluating pathologists.

The subsequent results were recorded under three broad categories as follows.

1. Onsite evaluation of DPS technical capability

Scanner capabilities of the four DPSs were compared across the specimen types

- Slide scanning performance:** The first-time successful scanning rate for all specimen types followed the sequence: DPS1 (96.1%) > DPS4 (93.8%) > DPS3 (88.9%) > DPS2 (86.5%). Scanning of cytology slides was particularly challenging across all DPSs, especially with DPS2, which had an overall failure rate of 42.5% even after rescanning, due to which 25 cytology cases corresponding to these failed 40 slides could not be evaluated further on DPS2. Rescanning of the slides after correcting the pre-imaging factors rectified the initial failure rate in all other DPSs [Table 1]. There was no significant difference in the scan rate between H & E and IHC/special stains, as well as between positive and negative IHC slides.
- Scanning time and storage space:** The mean scanning time per slide followed this sequence: DPS4 (126.21 s) < DPS1 (155.7 s) < DPS2 (161.95 s) < DPS3 (183.7 s). Overall digital image output from the DPS3 (208 GB) occupied the least storage

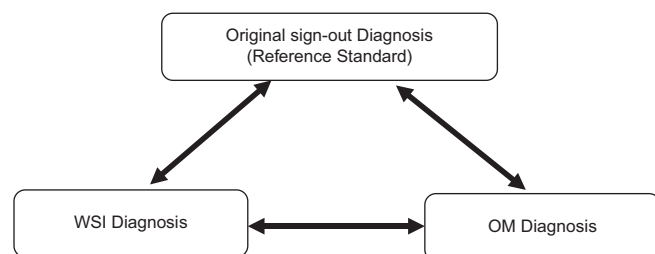


Figure 1: Study design for diagnostic evaluation. (a) Interobserver agreement between reference diagnosis and WSI diagnoses using various digital pathology system (b) Interobserver agreement between reference diagnosis and OM diagnoses of participating pathologist (c) Intraobserver agreement of the OM and WSI diagnoses of individual participating pathologist. OM Optical microscopy, WSI Whole slide imaging

space across all specimen types, followed by DPS2 (405 GB) < DPS1 (868 GB) < DPS4 (934 GB). Among the specimen types, the scanning time for cytology slides was longer, and they occupied more storage space as opposed to the H & E and IHC slides. Further, the mean scanning time and average size of digital images of IHC slides across specimen types (biopsy and resection) were significantly less when compared to corresponding H & E slides (range: 104–143 s vs. 132–158 s; 0.44–0.88 GB vs. 0.675–1.11 GB) [Table 2]

2. Diagnostic assessment: WSI versus OM

The diagnostic assessment of WSI versus OM diagnosis was recorded under following headings.

a. Overall discordance rate and diagnostic accuracy rate as per the specimen type and DPS

A total of 382 discordant reads out of 6100 diagnostic reads were recorded, using both OM (1245 reads) and WSI platforms (4855 reads) by all pathologists, as compared to the reference standard. The overall discordance rate for OM was 4.56% (57/1245) including 2.08% (26/1245) major and 2.48% (31/1245) minor discordances. The overall discordance rate for WSI was 6.68% (325/4855) including 2.4% (117/4855) major and

4.2% (208/4855) minor discordances [Figure 2]. Few of the major discordances observed in this study are illustrated in Figure 3.

The mean diagnostic accuracy rate across the specimen types using different DPSs is summarized in Table 3. The overall diagnostic accuracy for OM and WSI, when compared with the reference standard, was 95.44% and 93.32%, respectively. Considering only major discrepancies, the overall diagnostic accuracy for OM and WSI when compared with the reference standard was 97.92% and 97.6%, respectively. Hence, the difference between the clinically significant discrepancies by WSI and OM diagnosis was 0.32% (95% confidence interval, 0.25–1.0).

The mean difference in the diagnostic accuracy using all DPSs and OM, as compared to the reference standard, was <4% for biopsy, resection specimens, and frozen section diagnosis, thus proving WSI was noninferior to OM for the primary diagnosis of these specimens. However, WSI was inferior to OM for the primary diagnosis in the cytology specimens, as the mean difference in the diagnostic accuracy between WSI (range for various DPSs: 43.59%–87%) and OM (mean: 93.4%) as compared to reference standard was >4%.

Diagnostic assessment according to the DPS used is illustrated in Figure 2. Major and minor discordances across the four DPSs

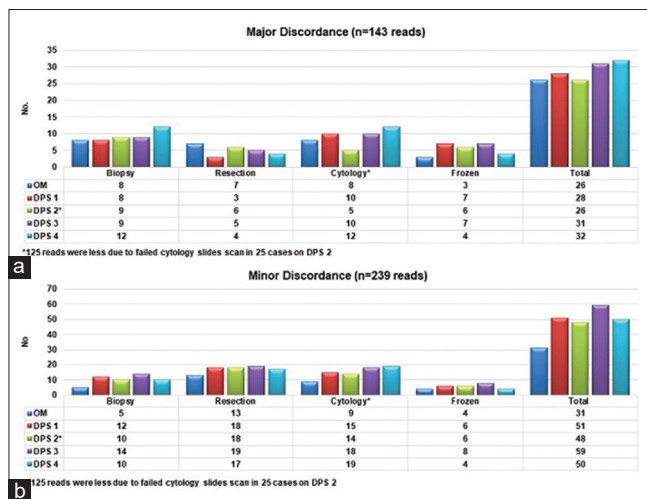


Figure 2: Bar diagram showing major discordances (a) and minor discordances (b) across various digital pathology systems and specimen types

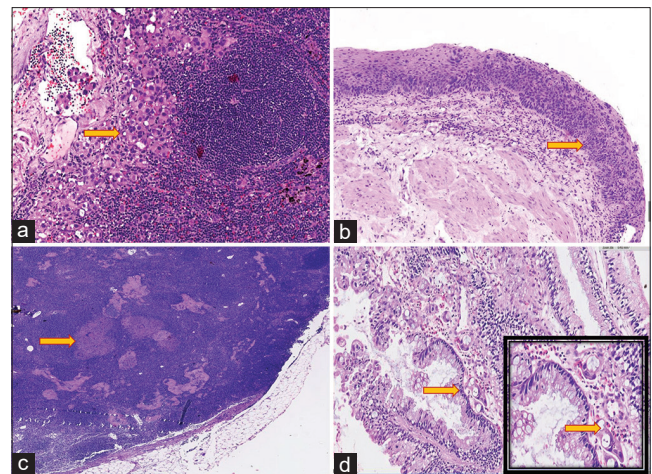


Figure 3: Photomicrograph of few discordant cases. (a) Scanty focus of metastatic carcinoma in lymph node, (b) focal dysplasia at esophageal cut margin, (c) scanty focus of granuloma in lymph node, (d) foci of intra mucosal adenocarcinoma in serrated adenoma which were missed on whole slide imaging diagnosis

Table 1: Scanning performance of digital pathology system

Specimen type	DPSs (%)							
	DPS1		DPS2		DPS3		DPS4	
	1 st scan	Re-scan	1 st scan	Re-scan	1 st scan	Re-scan	1 st scan	Re-scan
Biopsy	189/195 (96.9)	6/6 (100)	185/195 (95)	10/10 (100)	180/195 (92.3)	15/15 (100)	187/195 (96)	8/8 (100)
Resection	220/230 (96)	10/10 (100)	217/230 (94)	13/13 (100)	212/230 (92)	18/18 (100)	218/230 (95)	12/12 (100)
Cytology	89/94 (95)	5/5 (100)	39/94 (41)	15/55 (27.3)	64/94 (68)	30/30 (100)	79/94 (84)	15/15 (100)
Frozen	83/85 (98)	2/2 (100)	82/85 (96)	3/3 (100)	81/85 (95)	4/4 (100)	83/85 (98)	2/2 (100)

DPSs: Digital pathology system

Table 2: Scanning time and average storage space utilization by various digital pathology system

Specimen type	DPSs, mean (range)							
	DPS1		DPS2		DPS3		DPS4	
	Scanning time (s)	Storage (GB)	Scanning time (s)	Storage (GB)	Scanning time (s)	Storage (GB)	Scanning time (s)	Storage (GB)
Biopsy (H & E) (n=62)	139.24 (54-625)	1 (0.14-2.4)	138 (31-470)	0.32 (0.05-1.5)	145 (82-302)	0.14 (0.03-0.3)	109.5 (50-316)	1.25 (0.54-3.9)
Biopsy (IHC) (n=128)	98.08 (24-342)	0.5 (0.41-2.59)	116 (26-387)	0.12 (0.01-0.5)	120 (65-240)	0.12 (0.02-0.8)	85.75 (18-281)	1 (0.45-2.2)
Resection (H and E) (n=166)	155.25 (41-620)	1.68 (0.12-5.67)	168.25 (22-360)	0.97 (0.19-2.20)	182 (67-346)	0.24 (0.2-0.69)	126.5 (42-320)	1.73 (0.37-5.6)
Resection (IHC) (n=60)	120.75 (24-342)	1.22 (0.09-2.84)	164 (26-320)	0.62 (0.16-2.5)	170 (110-278)	0.22 (0.02-1.9)	121.5 (24-275)	1.46 (0.46-2.8)
Cytology (n=94)	272.97 (91-620)	2.6 (0.82-5.51)	61.5 (34-102)*	0.16 (0.05-0.34)*	245 (85-400)	0.78 (0.17-2.5)	195 (35-374)	2.31 (0.41-4.1)
Frozen (n=85)	148.08 (25-379)	1.59 (0.11-3.28)	223.5 (22-380)	0.85 (0.1-2.34)	240 (28-420)	0.2 (0.17-0.5)	119 (18-270)	1.34 (0.10-3.1)

*40 cytology slides completely failed to scan on DPS2. DPSs: Digital pathology system, IHC: Immunohistochemistry, H & E: Hematoxylin and eosin

were as follows: DPS1 (79/1245 [6.34%] including 28 major discordance [2.24%] and 51 minor discordance [4.09%]), DPS2 * (74/1120 [6.60%] including 26 major discordance [2.32%] and 48 minor discordance [4.28%] *125 reads were less due to failed cytology slides scan in 25 cases), DPS3 (90/1245 [7.23%] including 31 major discordance [2.48%] and 52 minor discordance [4.17%]), and DPS4 (82/1245 [6.58%] including 32 major discordance [2.57%] and 50 minor discordance [4.016%]). This difference was not statistically significant.

b. Interobserver agreement for WSI and OM for primary diagnosis

Almost perfect agreement ($\kappa > 0.8$) was noted for resection specimens and frozen section specimens, and there was substantial agreement ($\kappa = 0.6-0.8$) for biopsy specimens across all the DPSs. This underscores that all DPSs had reproduced similar results to one another and with OM, across the above-mentioned specimen subtypes. In cytology specimen, there was an almost perfect agreement ($\kappa > 0.8$) for OM and substantial agreement ($\kappa = 0.6-0.8$) for the various digital DPSs, highlighting that all DPSs could not reproduce similar results as with OM when using the same slides. Although DPS2 had high failure rates, the interobserver agreement for successfully scanned digital images was similar to that with other DPSs [Figure 4] and Supplementary Table 1.

c. Intraobserver agreement for WSI and OM for primary diagnosis

All five pathologists had an almost perfect agreement ($\kappa > 0.8$) between WSI and OM for resection and frozen subtypes. For biopsy specimens, almost perfect agreement ($\kappa > 0.8$) between WSI and OM on different DPSs was observed for three pathologists and substantial agreement ($\kappa = 0.6-0.8$) for other two pathologists. For cytology specimens, two pathologists had an almost perfect intraobserver agreement ($\kappa = 0.8$) between WSI and OM, other two had a substantial intraobserver agreement ($\kappa = 0.6-0.8$) between WSI and OM, and one had a moderate intraobserver agreement ($\kappa = 0.4-0.6$) between WSI and OM [Figure 5 and Supplementary Table 2].

There was no statistical significant correlation between the discrepancy rates and the clinical experience of pathologists in this study.

d. Diagnostic assessment time; WSI versus OM

The mean time taken for diagnosis for various specimen types by OM and WSI platforms by the individual pathologist was recorded as illustrated in Figure 6a and b and Supplementary Tables 3 and 4. The overall mean assessment time per case using OM was 47.60 s and for WSI using DPS1, DPS2, DPS3, and DPS4 was 71.65, 56.55, 55.29, and 55.38 s, respectively, with an average of 59.72 s. Hence, the diagnostic assessment time required for OM was less as opposed to WSI across all specimen types by all pathologists. The overall difference between the mean reading times for OM diagnosis and WSI

Table 3: Mean Diagnostic accuracy as per specimen subtypes: OM/ WSI Vs. Reference Standard

Specimen type	Mean diagnostic accuracy [% (range)]				
	Reference versus OM	Reference versus DPS 1	Reference versus DPS 2*	Reference versus DPS 3	Reference versus DPS 4
Biopsy	94.3 (90-98.3)	92.7 (86.7-96.6)	94.0 (90.0-98.3)	91.0 (88.0-95.4)	91.7 (88.3-96.7)
Resection specimen	92.64 (86.6-98.3)	91.9 (85.0-98.3)	90.34 (86.7-96.7)	90.34 (86.7-96.7)	91.9 (85.0-98.3)
Frozen section	97.68 (95-100)	94.34 (91.7-96.7)	95.6 (95.00-98.3)	94.28 (88.1-98.3)	97.66 (96.70-98.3)
Cytology	93.34 (90-96.7)	85.02 (76.7-91.70)	43.59 (40-46.70)	84.68 (81.7-86.7)	83.3 (80.00-93.3)

*125 reads were less due to failed cytology slides scan in 25 cases on DPS 2. DPSs: Digital pathology system, OM: Optical microscopy

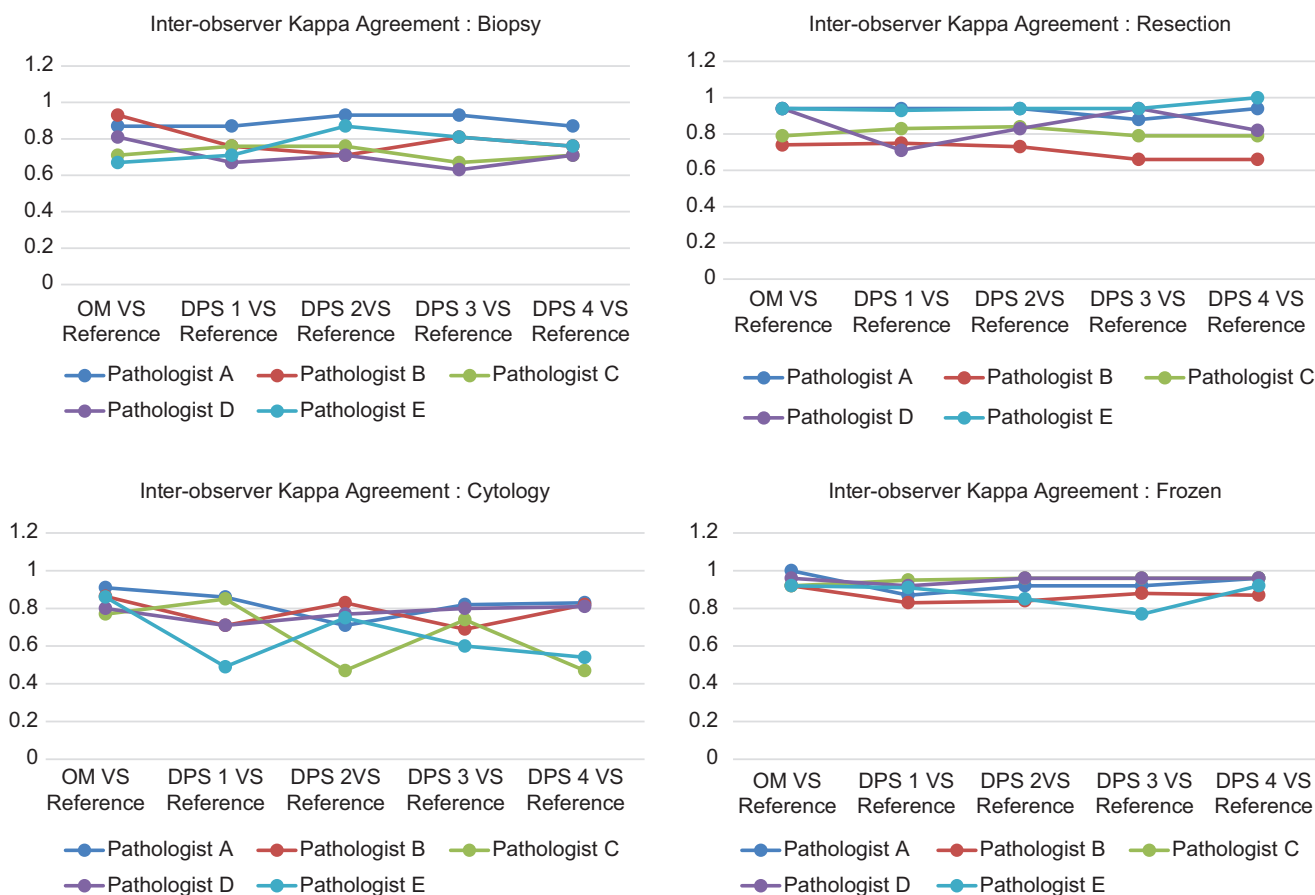


Figure 4: Line diagram showing inter observer agreement between optical microscopy and whole slide imaging for various digital-pathology-systems as compared to reference standard for different specimen types

diagnosis was 12 s per slide and maximum difference observed in DPS1, except for cytology cases. Pathologist E took least time for the assessment using both OM and WSI among five pathologists. Total diagnostic time spent for cytology and frozen section as compared to biopsy and resection specimens was less, for both OM and WSI, because fewer slides per case were included for these specimen types in this study.

3. Assessment of digital image quality, artifacts, IVS, and level of confidence

The overall image quality was best in DPS1, followed by DPS4 [Figure 7a]. No statistically significant correlation between the number of discrepancies and image quality of a particular DPS could be established.

Mean digital image artifacts rate was 6.8% (163/2376 digital images) across all the DPSs (excluding failed scans) [Figure 7b]. The maximum number of digital artifacts were noted in DPS2 (*n* = 77) followed by DPS3 (*n* = 36). Common artifacts were out of focus images (either focal or diffuse) observed in H & E slides on DPS4 and DPS3 and stitching errors in cytology/H & E slides on DPS2, resulting in a very limited area in digital format available for assessment [Figure 8]. None of the digital slides, except the one which were completely out of focus (*n* = 4), were deferred during evaluation.

Most of the pathologists (4 out of 5) preferred viewing software of DPS1 and DPS2 based on the survey of the parameters

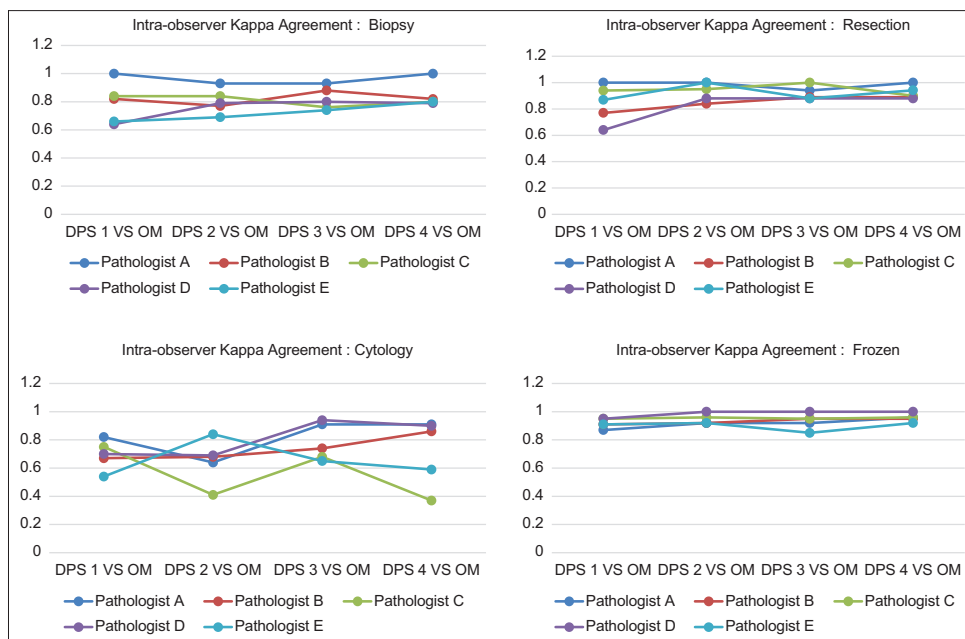


Figure 5: Line diagram showing intraobserver agreement between optical microscopy and whole-slide imaging for various digital pathology systems of individual pathologist for different specimen types

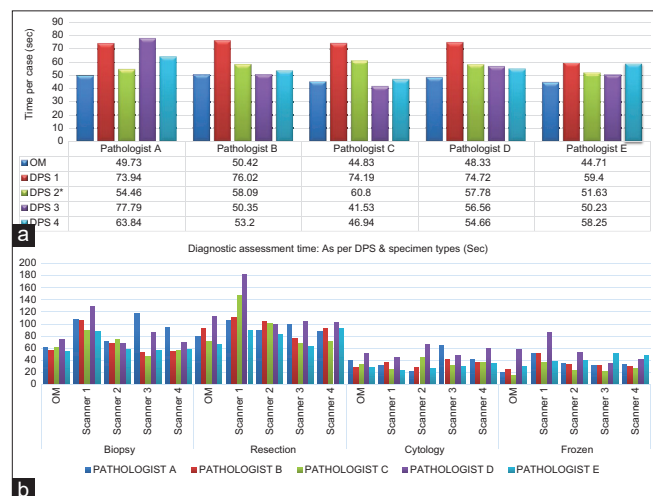


Figure 6: Bar diagram showing mean diagnostic assessment time across various digital pathology systems according to pathologists (a) and as per specimen types (b)

for the assessment of IVS by the participating pathologist. The DPS1 and DPS2 were almost consistent in reproducing the original color of the glass slides. Images appeared more basophilic in DPS4, whereas they were more eosinophilic in DPS3 [Figure 9].

Based on the mean score of the participating pathologist, level of confidence for digital reporting was highest for DPS1 followed by DPS2 for biopsy, resection, and frozen section cases. The overall level of confidence for cytology evaluation was average, irrespective of the DPS type [Figure 10]. The low level of confidence was recorded, especially for the evaluation of negative lymph nodes in the digital images of frozen section slides and exfoliative cytology.

DISCUSSION

Current technological advancements in DP have revolutionized the practice of pathology. Results of some recent studies demonstrating high concordance rates between the OM and WSI diagnoses have justified the feasibility of replacing microscopes with digital platforms in the future.^[4-12] However, this transition from a glass slide to digital is not straightforward. Hence, robust validation of WSI platforms is required before adoption into clinical practice, as recommended by the recently published guidelines.^[16-18]

With significant improvement in technology and increasing availability of a large number of DPSs, it becomes very challenging for the pathologist to select an appropriate platform, which can be easily incorporated in the routine laboratory workflow. Further, there is a paucity of published literature on the comparative technical evaluation of various DPSs.

This is a unique comprehensive comparison study incorporating a spectrum of DPSs, specimen types, and cases posing different levels of complexity and encompassing each and every component of WSI, including technical factors as well as diagnostic performance using different DPSs. This is likely to emulate the real-world clinical scenario in keeping with the CAP recommendations for validation and adopting DP. This study was unique since each case was interpreted five times by each individual pathologist (once using OM and four times on different WSI platforms). Hence, despite low number of cases enrolled in the present study, i.e., 604 glass slides from 240 cases, the final evaluation was based on a significantly large number of observations, i.e., 15,575 image reads (OM - 3171 + WSI - 12,404) and 6100 diagnostic

Table 4: Comparison amongst major published WSI validation studies in literature

	Campbell et al., 2012^[4]	Bauer et al., 2013^[5]	Snead et al., 2016^[7]	Tabata et al., 2017^[19]	Mills et al., 2018^[8]	Mukhopadhyay et al., 2018^[9]	Hanna et al., 2018^[10]	Borowsky et al., 2020^[11]	Current study
Number cases	212	607	3017	1070	510	1992	204	2045	240
Number of diagnostic Reads	424	1025	3017	1070	3060	15925	1075	15031	6100
Number of pathologists involved	2	2	17	9	3	16	8	19	7
Training period	Not specified	Yes	Yes	Yes	Not specified	Yes	Not specified	Not specified	Yes
Scanner model	iScan	Aperio scope XT	Omnyx VL4	Philips intellisite, AperioAT2, nanozoomer, FINO	Philips intellisite	Philips intellisite	Aperio AT2	Aperio AT2	4 different DPSs*
Scanning magnification	20×	20×	40×	Both 20× and 40×	40×	40×	40×	20×	Both 20× and 40×.
Wash-out period	Not specified	1 year	3 week	2 weeks	6 week	4 week	13 weeks	31 days	2 weeks
Consecutive cases selected	Yes	Yes	Yes	Not specified	Not specified	No	Not specified	Not specified	No
Randomization of cases	Not specified	Not specified	Not specified	Not specified	Yes	Yes	Yes	Yes	Yes
Original diagnosis blinded	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Both WSI and OM evaluation done by same pathologist	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Diagnostic concordance rate (%)	97.70	99.10	97.70	99.1	95.60	99	99.30	96.36	97.60
Evaluation criteria	Diagnostic assessment	Diagnostic assessment	Diagnostic assessment	Diagnostic assessment	Diagnostic assessment	Diagnostic assessment	Diagnostic and technical assessment	Diagnostic and technical assessment	Diagnostic and technical assessment

*Identity of the DPS was kept confidential as per agreement with vendors. DPSs: Digital pathology system, OM: Optical microscopy, WSI: Whole-slide imaging

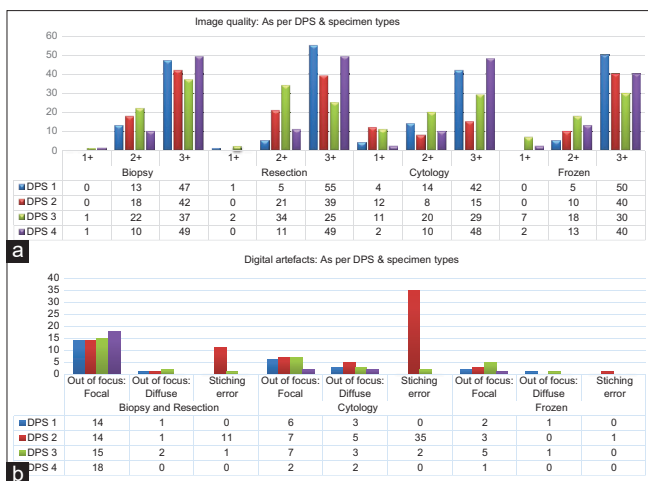


Figure 7: Bar diagram showing level of image quality (a) and types of digital artifacts (b) across various digital-pathology-systems and specimen types

reads (OM - 1245 + WSI - 4855). Thus, the present study represents the third largest series worldwide, in terms of number of diagnostic reads comparing WSI versus OM with enough statistical power for analysis, and is first of its kind wherein comparison amongst 4 DPS were performed, as summarized in Table 4.^[4,5,7-11,19]

While deploying DPS, it is recommended to check the scanning capabilities based on realistic clinical tissue specimens to ensure appropriate implementation of DPS. These relevant technical aspects were highlighted by this onsite technical evaluation of the various DPSs, as discussed in ensuing paragraphs.

The successful scanning rates and rescan rates for the surgical pathology specimen as observed in this study were within the same range as reported by previous studies.^[9,10] All the DPSs were compatible with the existing glass slides used in our institute for various specimen types, except for DPS2, wherein many cytology slides failed to scan ($n = 40/95$). Thickness of slides used for cytology preparation seemed to contribute to the higher failure rate, despite repeated scanning attempts, and thus constitutes an important determining factor. Further, challenges in scanning of the cytology slides are well documented and need additional refinement in the technology.

Scanning time and storage requirement based on the data generated per digital slide have implications on the turnaround time (TAT) and investment for digital archival. As a variety of specimen types of different tissue areas are encountered during routine reporting, the industry standard 15 mm × 15 mm area for the assessment of these parameters is less informative for practical purposes. However, very limited information based on the realistic clinical setting is available on this aspect. Hanna *et al.* based on 204 cases documented the median WSI file size of 1.54 GB, scan time/slide of 6 min 24 s, and scan area of 32.1 mm × 18.52 mm using 40 × equivalent resolution (0.25 μm/pixel).^[10] Snead *et al.* recorded a digital archive space of 2.22 TB (mean size per

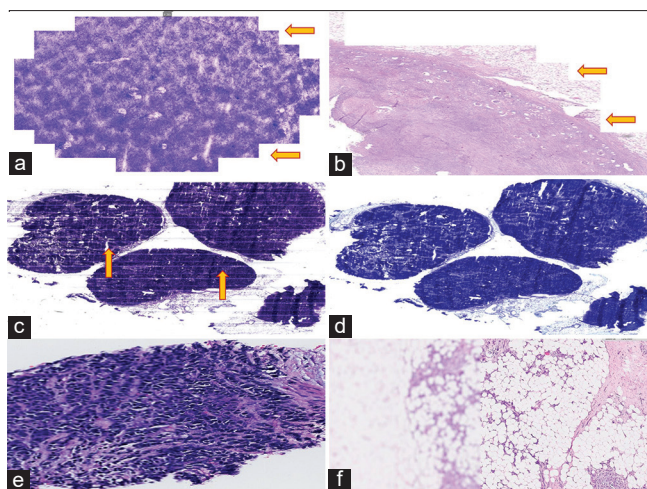


Figure 8: Photomicrograph showing various digital artifacts: Stitching artifact (Arrow) in cytology (a) and resection specimens (b); line artifacts (Arrow) on digital pathology system 3 (c, - toluidine blue; ×5) and good quality image on digital pathology system 2 (d, toluidine blue; ×5) for the same glass slide; focal blurring in biopsy case (e); focal out of focus area in resection specimen (f)

case of 189 MB) in a cohort of 3017 cases (including 2666 biopsies, 340 resection, and 11 frozen cases).^[7] The current study has provided more realistic information on average scan time and image size for each specimen type using 4 different scanners [Table 2]. Interestingly, the mean scanning time and digital storage space requirements for IHC slides were found significantly less by approximately 50% as compared to the corresponding H & E slides, across all DPSs. None of the prior studies had documented this finding. Further, cytology and frozen section slides consumed more time to scan and storage as opposed to the H & E and IHC slides in all the scanners. The lossy compression mode used in DPS3 resulted in the smaller image file sizes as compared to other systems. Choosing appropriate file compression method would facilitate the data handling in routine DP practice.

A meta-analysis review of the prior concordance studies demonstrated an increase in overall diagnostic concordance over the years.^[20,21] This may be due to a combination of improvement in technology as well as increased familiarity and confidence of the pathologists using DPSs over time. Of all prior studies, the diagnostic intraobserver concordance reported ranged from 63% to 100%, with the mean diagnostic concordance rate of 92.4%.^[4-12] Overall, the diagnostic accuracy rate of WSI in the current study bearing equal clinical impact as compared to OM was 97.6%. Hence, our results were similar to those reported by other authors (ranges from 95.6% to 99.1%) as summarized in Table 4.

In a recent review, Williams *et al.* evaluated discordance rate based on data from 23 DP validation papers including 8069 glass digital comparisons and 335 discordances were recorded amounting to the rate of 4% including predominantly minor discordances.^[20] Other studies have reported a discordance rate of 3.0% for 607 consecutive daily clinical cases and 0.89% for

3017 cases.^[5,7] The overall discordance rate in our study for WSI was 6.68% (including 2.40% - major and 4.28% - minor) as opposed to 4.56% (including 2.08% - major and 2.48% - minor) for OM, when compared to reference standard. The possible reason for higher disparity in our cohort as compared to reported in the literature may be due to evaluation by multiple pathologists with varying degree of experience, inclusion of cytology specimens, and documentation of both minor as well as major discrepancies. However, it is difficult to determine whether the discordances recorded were due to a technical problem attributable to DPS or random error by individual pathologists. We did not find any platform-specific problem attributing to the statistical differences in diagnostic discrepancy among the various DPSs used in this study. However, a minimum number of clinically relevant major discordances were noted in DPS 1. Mukhopadhyay *et al.* and Tabata *et al.* did not find any correlation between the DPS used and discrepancy rates. Besides these studies, none have addressed this issue.^[9,19]

A wide variety of cases included in this series allowed us to evaluate a diagnostic assessment of WSI as per specimen

types. WSI was consistently proved to be noninferior to OM in biopsy, resection, and frozen specimens, as the mean difference between WSI and OM diagnosis as compared to the reference standard was <4%. The inter- and intra-observer agreement between WSI and OM for primary diagnosis of cytology specimens in this study was not substantial as reported by others.^[22-24] There were not only more diagnostic discrepancies by WSI but also a high number of technical challenges in cytology specimens as compared to other specimens. These technical issues (e.g., the image layering, i.e., multiplane z-stacks, to provide depth-of-focus) should be addressed properly to improve the evaluation of cytology specimens by WSI in the future.

The diagnostic assessment time for reporting using digital images is likely to impact the efficiency of WSI over the standard OM. Diagnostic assessment time using WSI has been reported longer than traditional microscopy in three out of four prior validation studies.^[25-28] Hanna *et al.* observed a median of 19 s (26%) increase per slide and 2 min 57 s increase per case when signing out digitally.^[10] We also observed that overall time required for OM was less as opposed to WSI across all scanners and specimen types. Further, as study progressed, pathologists spent relatively less time for WSI evaluation. More time spent on WSI might be related to unfamiliarity of the participating pathologist, learning curve with WSI, subspecialty reporting, screen loading time, and maneuverability involved in digital reporting. The time spent on digital reporting will definitely reduce in the future with more practice, repetitive use, and training on WSI.

In addition to the diagnostic accuracy, we have documented the impact of the digital image quality, artifacts, and user-friendliness of IVS on DP. These aspects were not elaborated in any of the previous validation studies, as the evaluation was primarily done using a single DPS. Color variation and presence of digital artifacts influence the quality of the digital images. The final color of digital images depends on illumination, magnification, image capture, compression, and storage.^[29] Because each scanner company uses different image processing algorithms, variation in color contrast and intensity can be observed. Gray *et al.* demonstrated a mean difference of 7% in H & E color ratio, while scanning the

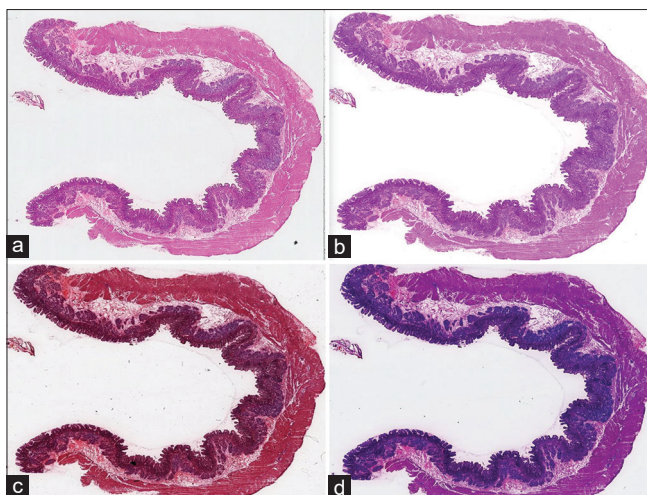


Figure 9: Photomicrograph of resection margin of intestine, depicting wide variation in the color of the image by various digital pathology systems (×2): digital pathology system 1 (a), digital pathology system 2 (b), digital pathology system 3 (c), and digital pathology system 4 (d)

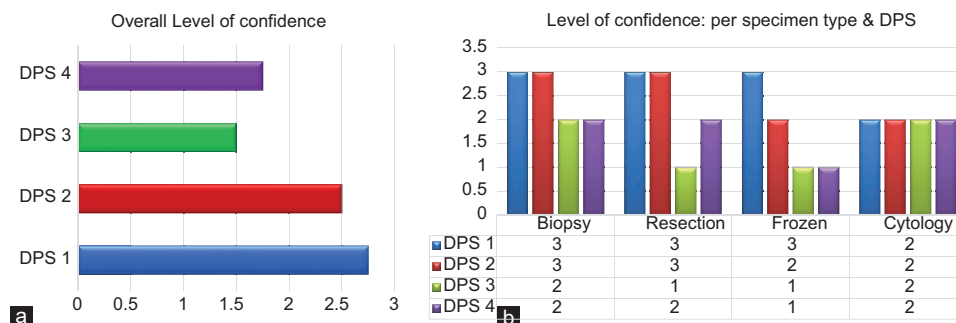


Figure 10: Bar diagram showing overall level of confidence across various digital pathology systems (a) and according to specimen types (b)

same slide into different scanners on the same day.^[30] We noticed that the DPS1 and DPS2 were almost consistent in reproducing the original color of the glass slides. Image quality in the DPS3 and DPS4 was largely attributable to poor color fidelity. Establishment of global color standardization by using universal calibration of WSI images and pseudostaining, i.e., digital superposition of color to WSI as opposed to actual staining, is recommended as possible solution to overcome the color-related issue for WSI.^[31]

Digital artifacts can mask the diagnostic material and cause error in reporting. In the current study, mean digital artifacts rate was 6.8% (163/2376 digital images) across all the scanners. Jukić *et al.* have reported digital artifacts in 8.6% (77/900) of the digital images.^[32] Common artifacts encountered in our study were out-of-focus (either focal or diffuse) areas ($n = 113$) and stitching errors ($n = 50$). Maximum digital artifacts were observed in DPS2 ($n = 77$), followed by DPS3 ($n = 36$). Digital artifacts were frequent in the cytology images across all DPSs which could possibly be attributed to (1) variability of specimen thickness due to nonuniform material distribution, (2) low cellularity, (3) material outside the coverslips, and (4) the markings toward the edges of slide, especially due to use of diamond pencil for labelling.

Although the image quality plays a very crucial role in the diagnostic assessment and confidence for digital reporting, no statistical correlation between the discordance rate and image quality of a particular scanner could be established based on this study. Suboptimal image quality can compromise the diagnostic efficiency and increase the reluctance to adopt the DP. Hence, proper calibration and daily quality control checks are recommended to improve the image quality. In the future, the scanner systems should be enabled to raise a flag/alarm if any scanned image has a significant digital artifact and the slides should be subjected to auto-rescan to improve the TAT and diagnostic efficiency.

Since pathologists are accustomed and trained on OM for rendering diagnosis, an efficient and user-friendly IVS is mandatory in order to shift to digital reporting in routine practice. In the current study, most pathologists preferred IVS of DPS1 and DPS2, as the pattern of case arrangement, panning, and navigation of slides resembled routine OM. IVS of DPS3 and DPS4 had frequent issues due to slow response and auto-closures, especially when multiple images were opened simultaneously. Vendor neutral IVS, which can allow access to the digital images scanned on different DPSs, might be a better option to overcome these issues.

Finally, level of confidence for digital reporting was also evaluated, which was highest for DPS1 followed by DPS2 for biopsy, resection, and frozen section cases. Overall level of confidence for cytology evaluation was average, irrespective of the scanner type, wherein high numbers of discordances were recorded. Improvised scanning of cytology slides with resultant enhanced image quality coupled with training and repeated use of WSI is likely to improve the level of

diagnostic confidence for digital reporting in routine clinical practice.

Thus, understanding and evaluation of all these technical parameters of the DPS, in addition to the diagnostic assessment, are pivotal for the successful implementation of DP in routine surgical pathology workflow. Further, in addition to the mandatory initial validation of DPS, daily quality control and monitoring of both preimaging (e.g., slide preparations, technical personnel) and imaging factors (e.g., scanner calibration) are mandatory for successful adoption of DP for primary diagnosis and artificial intelligence algorithms.

There were a few limitations of this study. Cases from only four subspecialties were included in this study. Additional subspecialty-specific studies with larger numbers of cases to unravel the anatomic site-specific interpretation issues for WSI are required for complete validation. We intentionally did not include difficult cases in our study cohort as pathologists tend to have a long memory for such cases. We could not record information about the optimization of network connectivity in this study as the images were directly viewed on the computer system associated with the scanner and not on the hospital network. Hence, network-related issues between scanners were not addressed in this study. Finally, although the wash-off period of 2 weeks was observed between the two readings, the recall bias among the participating pathologist could not be completely ruled out as they observed each case 5 times.

Based on our results, it can be concluded that irrespective of the DPS used, WSI could be deemed as noninferior to OM for all specimen types except for cytology specimens. The DPS used did not significantly influence the diagnostic capability of the pathologists, and we were able to record high levels of intraobserver equivalence. Technical refinements for scanning cytology specimens, autonavigation, and training of pathologists can substantially address the existing issues. Each scanner had its own pros and cons for various parameters assessed. Based on both the technical and diagnostic performance, DPS1 closely emulated the real-world clinical environment, when compared with OM.

DP remains as a dynamic complex algorithm of technical factors, case-related parameters, and pathologist experience and training levels. Training and institutional validation are indispensable after making decisions on technical parameters customized to one's needs. The study highlights that irrespective of the DPS used, pathologists adapt and autotune to technical influencing factors in unquantifiable measures. This evaluation will provide a roadmap to pathologists for selection of the appropriate DPSs, which can best suit their environment in routine clinical practice, while adopting the WSI technology.

Acknowledgment

We acknowledge all the digital pathology system vendors who agreed to participate in this comparative assessment and Mrs. Rashmi Sarang for her assistance in data management in electronic format.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Cornish TC, Swapp RE, Kaplan KJ. Whole-slide imaging: Routine pathological diagnosis. *Adv Anat Pathol* 2012;19:152-9.
- Griffin J, Treanor D. Digital pathology in clinical use: Where are we now and what is holding us back? *Histopathology* 2017;70:134-45.
- Evans AJ, Salama ME, Henricks WH, Pantanowitz L. Implementation of whole slide imaging for clinical purposes: Issues to consider from the perspective of early adopters. *Arch Pathol Lab Med* 2017;141:944-59.
- Campbell WS, Lele SM, West WW, Lazenby AJ, Smith LM, Hinrichs SH. Concordance between whole-slide imaging and light microscopy for routine surgical pathology. *Hum Pathol* 2012;43:1739-44.
- Bauer TW, Schoenfeld L, Slaw RJ, Yerian L, Sun Z, Henricks WH. Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med* 2013;137:518-24.
- Thrall MJ, Wimmer JL, Schwartz MR. Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2015;139:656-64.
- Snead DR, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68:1063-72.
- Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, *et al.* Diagnostic efficiency in digital pathology: A comparison of optical versus digital assessment in 510 surgical pathology cases. *Am J Surg Pathol* 2018;42:53-9.
- Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaiya S, Cacciabeve NG, *et al.* Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (Pivotal Study). *Am J Surg Pathol* 2018;42:39-52.
- Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, *et al.* Whole slide imaging equivalency and efficiency study: Experience at a large academic center. *Mod Pathol* 2019;32:916-28.
- Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, *et al.* Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology: A multicenter, double-blinded, randomized study of 2045 cases. *Arch Pathol Lab Med* 2020;144:1245-53.
- Rao V, Subramanian P, Sali AP, Menon S, Desai SB. Validation of whole slide imaging for primary surgical pathology diagnosis of prostate biopsies. *Ind J Pathol Microbiol* 2021;64:78-83. [Doi: 10.4103/IJPM.IJPM_855_19].
- US Food and Drug Administration Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices. Available from: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM435355.pdf>. [Last accessed on 2021 Feb 11].
- Li X, Liu J, Xu H, Gong E, McNutt MA, Li F, *et al.* A feasibility study of virtual slides in surgical pathology in China. *Hum Pathol* 2007;38:1842-8.
- Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363-74.
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, *et al.* Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137:1710-22.
- García-Rojo M. International clinical guidelines for the adoption of digital pathology: A review of technical aspects. *Pathobiology* 2016;83:99-109.
- Royal College of Pathologists (RCP). Best Practice Recommendations for Implementing Digital Pathology. London, UK: RCP; 2018.
- Tabata K, Mori I, Sasaki T, Itoh T, Shiraiishi T, Yoshimi N, *et al.* Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. *Pathol Int* 2017;67:547-54.
- Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. *Arch Pathol Lab Med* 2017;141:1712-8.
- Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: A systematic review. *Arch Pathol Lab Med* 2017;141:151-61.
- Wilbur DC. Digital cytology: Current state of the art and prospects for the future. *Acta Cytol* 2011;55:227-38.
- Evered A, Dudding N. Accuracy and perceptions of virtual microscopy compared with glass slide microscopy in cervical cytology. *Cytotechnology* 2011;22:82-7.
- Donnelly AD, Mukherjee MS, Lyden ER, Bridge JA, Lele SM, Wright N, *et al.* Optimal z-axis scanning parameters for gynecologic cytology specimens. *J Pathol Inform* 2013;4:38.
- Gui D, Cortina G, Naini B, Hart S, Gerney G, Dawson D, *et al.* Diagnosis of dysplasia in upper gastro-intestinal tract biopsies through digital microscopy. *J Pathol Inform* 2012;3:27.
- Jen KY, Olson JL, Brodsky S, Zhou XJ, Nadasdy T, Laszik ZG, *et al.* Reliability of whole slide images as a diagnostic modality for renal allograft biopsies. *Hum Pathol* 2013;44:888-94.
- Vodovnik A. Diagnostic time in digital pathology: A comparative study on 400 cases. *J Pathol Inform* 2016;7:4-8.
- Velez N, Jukic D, Ho J. Evaluation of 2 whole-slide imaging applications in dermatopathology. *Hum Pathol* 2008;39:1341-9.
- Clarke EL, Treanor D. Colour in digital pathology: A review. *Histopathology* 2017;70:153-63.
- Gray A, Wright A, Jackson P, Hale M, Treanor D. Quantification of histochemical stains using whole slide imaging: Development of a method and demonstration of its usefulness in laboratory quality control. *J Clin Pathol* 2015;68:192-9.
- Kather JN, Weis CA, Marx A, Schuster AK, Schad LR, Zöllner FG. New colors for histology: Optimized bivariate color maps increase perceptual contrast in histological images. *PLoS One* 2015;10:e0145572.
- Jukić DM, Drogowski LM, Martina J, Parwani AV. Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images. *Arch Pathol Lab Med* 2011;135:372-8.

Supplementary Table 1: Interobserver agreement (Kappa) as per specimen subtypes: Optical microscopy/whole-slide imaging versus standard reference

Specimen type	Mean Kappa (range)				
	Reference versus OM	Reference versus DPS1	Reference versus DPS2*	Reference versus DPS3	Reference versus DPS4
Biopsy	0.80 (0.67-0.93)	0.76 (0.67-0.87)	0.80 (0.71-0.93)	0.77 (0.63-0.93)	0.76 (0.71-0.87)
Resection specimen	0.83 (0.55-0.95)	0.83 (0.63-0.93)	0.86 (0.73-0.94)	0.87 (0.66-0.94)	0.846 (0.66-1)
Frozen section	0.95 (0.92-1)	0.90 (0.63-0.93)	0.91 (0.86-0.96)	0.90 (0.78-0.96)	0.94 (0.88-0.96)
Cytology	0.84 (0.77-0.91)	0.73 (0.63-0.93)	0.71 (0.49-0.83)	0.61 (0.49-0.83)	0.70 (0.48-0.83)

DPSs: Digital pathology system, OM: Optical microscopy

Supplementary Table 2: Intraobserver agreement (Kappa) as per specimen subtypes: Optical microscopy versus whole-slide imaging

	Biopsy	Resection	Frozen section	Cytology
Pathologist A	0.97 (0.94-1)	0.986 (0.94-1)	0.92 (0.81-0.96)	0.83 (0.64-0.92)
Pathologist B	0.83 (0.78-0.88)	0.9854 (0.77-.896)	0.94 (0.91-0.96)	0.74 (0.68-.87)
Pathologist C	0.82 (0.76-0.85)	0.95 (0.90-1)	0.96 (0.96-0.96)	0.56 (0.38-0.75)
Pathologist D	0.76 (0.65-0.80)	0.824 (0.64-0.89)	0.99 (0.96-1)	0.81 (0.7-0.95)
Pathologist E	0.72 (0.66-0.80)	0.922 (0.86-1)	0.90 (0.85-0.93)	0.66 (0.55-0.84)

Supplementary Table 3: Overall diagnostic assessment time across various digital pathology system; optical microscopy versus whole-slide imaging (sec per case)

	OM			DPS1			DPS2*			DPS3			DPS4		
	Average	Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum
Pathologist A	49.73	10.75	156.25	73.95	13.75	191.75	54.46	11.5	121.25	77.79	13.25	203.75	63.84	11.75	142
Pathologist B	50.42	8	132	76.02	13	205	58.09	13.5	164.75	50.35	12.25	118.75	53.20	13	155.5
Pathologist C	44.83	8.25	119	74.19	16.25	146.5	60.80	16	146.75	41.53	11	117.5	46.94	15	106.5
Pathologist D	48.33	9	135.75	74.72	14.33	181.08	57.78	13.67	144.25	56.56	12.17	146.67	54.66	13.25	134.67
Pathologist E	44.71	8.75	113.5	59.41	13.75	142	51.63	11	118.5	50.23	12	123.75	58.25	10.75	156.25

*125 reads were less due to failed cytology slides scan in 25 cases on DPS 2. DPSs: Digital pathology system, OM: Over optical microscopy

Supplementary Table 4: Mean diagnostic assessment time across various digital pathology system according to pathologists and as per specimen types (sec per case)

	Pathologist A	Pathologist B	Pathologist C	Pathologist D	Pathologist E
Biopsy					
OM	60.52	55.5	60.45	74.63	55.07
DPS1	107.68	105.17	89.8	129.4	87.23
DPS2	71.65	68.27	75.15	68.35	57.9
DPS3	116.72	53.03	46.64	85.4	56.6
DPS4	94.17	55.05	55.67	69.42	58.18
Resection					
OM	79.39	91.87	70.65	112.1	66.12
DPS1	105.02	111.49	146.29	181.14	89.37
DPS2	88.67	104.2	100.37	99.62	82.6
DPS3	98.75	75.57	66.9	103.47	63.18
DPS4	88.19	93.25	70.3	102.63	92.63
Cytology					
OM	39.72	28.8	32.82	50.67	28.45
DPS1	31.59	35.89	24.82	44.12	22.55
DPS2	22.23	27.34	43.87	65.32	25.97
DPS3	64.56	41.9	30.89	47.6	29.21
DPS4	40.73	35.61	36.13	59.6	35.12
Frozen					
OM	19.3	25.5	15.4	57.9	29.2
DPS1	51.49	51.51	35.86	86.39	38.48
DPS2	35.3	32.53	23.82	52.12	40.03
DPS3	31.14	30.9	21.68	34.31	51.93
DPS4	32.27	28.9	25.67	40.57	47.05

*125 reads were less due to failed cytology slides scan in 25 cases on DPS2. DPSs: Digital pathology system, OM: Optical microscopy