NEW ARTICLE

This article was posted on the Archives Web site as a New Article. New Articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the Archives. New Articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2023-0214-OA

The print version of this manuscript will replace the New Article version at the above DOI once it is available.
Digital Pathology in the Detection of Infectious Microorganisms

An Evaluation of Its Strengths and Weaknesses Across a Panel of Immunohistochemical and Histochemical Stains Routinely Used in Diagnostic Surgical Pathology

Mehrvash Haghighi, MD; Clare Bryce, MD; John D. Paulsen, MD; Shafinaz Hussein, MD; Brandon Veremis, DDS; Christian Salib, MD; Roshanak Allaly, MD; Mega Lahori, MBBS; Yansheng Hao, MD, PhD; Yuanxin Liang, MD; Arnold Szporn, MD; William Westra, MD

Context.—The diagnosis of some infectious diseases requires their identification in tissue specimens. As institutions adopt digital pathology for primary diagnosis, the limits of microorganism detection from digital images must be delineated.

Objective.—To assess the reliability of microorganism detection from digitized images of histochemical and immunohistochemical stains commonly used in pathology.

Design.—Original glass slides from 620 surgical pathology cases evaluated for the presence of infectious microorganisms were digitized. Immunohistochemical stains included those for herpes simplex virus (n = 100), Giemsa for cytomegalovirus (n = 100), Helicobacter pylori (n = 100), and spirochetes (n = 80). Histochemical stains included mucicarmine for Cryptococcus spp (n = 20), Grocott methenamine silver for fungi (n = 100), Giemsa for H pylori (n = 100), and Ziehl-Neelsen for acid-fast bacilli (n = 20). The original diagnosis based on the glass slides was regarded as the reference standard. Six pathologists reviewed the digital images.

Results.—Digital review was generally associated with high (ie, ≥90%) specificity and positive predictive value owing to a low percentage of false positive reads, whereas a high percentage of false negatives contributed to low sensitivity and negative predictive value for many stains. Fleiss κ showed substantial interobserver agreement in the interpretation of Grocott methenamine silver and immunostains for herpes simplex virus, H pylori, and cytomegalovirus; moderate agreement for spirochete, Ziehl-Neelsen, and mucicarmine; and poor agreement for Giemsa.

Conclusions.—Digital immunohistochemistry generally outperforms histochemical stains for microorganism detection. Digital interpretation of Ziehl-Neelsen and mucicarmine stains is associated with low scores for interrater reliability, accuracy, sensitivity, and negative predictive value such that it should not substitute for conventional review of glass slides.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2023-0214-OA)

As many aspects of medical practice have become digitized in recent years, surgical pathology is also experiencing a digital transformation whereby traditional glass slides are scanned and then viewed on a computer monitor as digital images. At a time of increasing adoption of digital pathology, the gap between technical promise and quality of diagnostic performance is a matter of concern. To address this concern, the College of American Pathologists (CAP) in collaboration with the American Society for Clinical Pathology and the Association for Pathology Informatics recently updated their guideline for validating whole slide imaging for primary diagnosis. This guideline has been useful for the validation of routine hematoxylin-eosin and immunohistochemistry (IHC) slides, but other important aspects of diagnostic surgical pathology such as the interpretation of microorganism slides must also be addressed. Although histochemical studies are routinely used to enhance the visualization of infectious organisms, assuming that the glass slide experience can be generalized to the digital pathology platform without appropriate validation could contribute to significant diagnostic error and negatively impact patient care.

Currently there are technical aspects of digital pathology that may adversely affect the performance of microorganism detection. Due to set parameters for whole slide image...
scanners, for example, small and isolated organisms may easily elude digital detection when objective magnification is limited and point of focus is restricted by single-plane scanning. The impact of these and other factors on reliable detection have yet to be defined. The incorporation of digital pathology into daily diagnostic practices must first be informed by an analysis of its ability to detect microorganisms across various histochemical studies.

**MATERIALS AND METHODS**

Four IHC stains and 4 histochemical stains that are commonly used in diagnostic surgical pathology to detect microorganisms in tissue sections were evaluated by digital pathology (Table 1). The pathology laboratory information system was queried for those stains ordered from 2010 to 2021. A total of 678 cases were collected. Cases were selected to include an equal distribution of negative and positive slides. All original glass slides were inspected to ensure the presence of organisms and the quality of slides; and slides that were faded, scratched, or broken were excluded from further analysis. Slides were anonymized and the reviewers were blinded to the original diagnosis. The slides were scanned with a Philips ultra-fast scanner (UFS, Philips Healthcare, Amsterdam, the Netherlands) at 3–40 magnification with no z-stacking. The scanned images were inspected for quality control. Four slides were excluded from the study due to persistent blurred images after multiple rounds of rescanning.

The slides were sorted in random order and grouped based on the type of study as follows: herpes simplex virus IHC (Cell Marque, Rockford, IL; 362A-78-ASR) (n = 100), cytomegalovirus IHC (Millipore, Billerica, MA; CMV/MAB810) (n = 100), *Helicobacter pylori* IHC (Dako, Agilent technologies, Carpinteria, CA; H pylon/ Helicobacter pylori (B047101-1) (n = 100), spirochete (BioCare medical, Walnut Creek, CA; Treponema pallidum/ACA 135 A, B,C) IHC (n = 79), Grocott methenamine silver (GMS) for fungi (n = 100), and Ziehl-Neelsen (ZN) for acid-fast bacilli (n = 20) (Table 1). In this study, the *Treponema pallidum* (Spirochete) IHC was used to detect *Brachyspira* spp (the *Treponema pallidum* is known to cross-react with *Brachyspira* spp). Six board-certified pathologists (2 hematopathologists, 1 oral pathologist, 1 cytopathologist, 1 gastrointestinal pathologist, and 1 neuropathologist) reviewed the digital images. The participants’ experience ranged from 2 to 10 years with limited digital pathology experience. The pathologists reviewed the digital images and documented the result as positive, negative, or equivocal. The original interpretation rendered by the glass slide review was considered the reference standard.

Statistical analyses were performed using IBM SPSS V.28 (Chicago, Illinois; 2021). We measured the diagnostic value by assessing the pooled diagnostic sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. The equivocal values were excluded from pooled observed data for performance assessment. Fleiss \( \kappa \) was computed to assess the level of agreement between 6 observers, \( P < .01 \) was considered significant. The strength of the interobserver agreement was interpreted as follows: <0.2, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, excellent. The original diagnosis was compared to the digital image results to determine concordance rates for each observer. Equivocal results were considered discrepant. The color contrast between the stained organism and the background tissue was measured by using the online color-pair contrast testing website.6 The eyeprover tool of the application was positioned above the intended organism or stain to convert the color to an RGB value, and the contrast ratio was displayed as x:1.

**RESULTS**

Of the 678 cases identified, there were a total of 620 slides that were satisfactory for digital scanning. The scanning system failed to detect the tissue on 1 spirochete IHC stain and 3 mucicarmine stains despite multiple attempts, and these were removed from the study set.

**Statistical Performance**

The diagnoses based on review of the digital images were compared to the original diagnoses based on evaluation of the glass slides. For each detection assay, the ability to discern the presence or absence of microorganisms was statistically measured in terms of accuracy, sensitivity, and specificity, PPV and NPV were calculated. The total number of false positive and false negative interpretations in each category is illustrated in Figure 1. These results are detailed in Table 2.

![Figure 1. Percentage of false negative and false positive values in each category.](image-url)
Table 2. Statistical Evaluation of Individual Microorganism Detection Assays

<table>
<thead>
<tr>
<th>Microorganism Detection Assays</th>
<th>Immunochemical Stains</th>
<th>Histochemical Stains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Parameter</td>
<td>Accuracy, %</td>
<td>94.77</td>
</tr>
<tr>
<td></td>
<td>Sensitivity, %</td>
<td>95.43</td>
</tr>
<tr>
<td></td>
<td>Specificity, %</td>
<td>93.44</td>
</tr>
<tr>
<td></td>
<td>PPV, %</td>
<td>96.73</td>
</tr>
<tr>
<td></td>
<td>NPV, %</td>
<td>90.96</td>
</tr>
<tr>
<td>Abbreviations: CI, confidence interval; CMV, cytomegalovirus; GMS, Grocott methenamine silver; HP, Helicobacter pylori; HSV, herpes simplex virus; NPV, negative predictive value; PPV, positive predictive value; ZN, Ziehl-Neelsen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Herpes Simplex Virus IHC.—Herpes simplex virus IHC diagnostic values surpassed all other assays in all indicators except for specificity and PPV. Across the statistical categories, it showed the highest accuracy (94.8%), highest sensitivity (95.4%), and highest NPV (91.0%). All herpes simplex virus IHC diagnostic indicators were above 90% and distributed within a narrow range (91.0%–96.7%).

Cytomegalovirus IHC.—All diagnostic performance indicators were higher than 83% and distributed in a relatively narrow range (83.8%–99.0%). It demonstrated an accuracy of 91.5%, a sensitivity of 86.3%, a specificity of 98.8%, a PPV of 99.0%, and an NPV of 83.8%.

Helicobacter pylori IHC.—The diagnostic performance indicators were comparable to cytomegalovirus IHC and distributed in a relatively narrow range from 83% to 100%. There were no false positive interpretations, so specificity and PPV were 100%. The lowest value was for NPV (83.3%).

Spirochete IHC.—The diagnostic values of spirochete IHC were distributed in a relatively narrow range (78.7%–98.7%). It demonstrated an accuracy of 88.5%, a sensitivity of 78.7%, a specificity of 98.7%, a PPV of 98.4%, and an NPV of 81.7%.

GMS for Fungi.—The performance of histochemical stains were consistently below the IHC studies with the notable exception of the GMS stain for fungi. All indicators were above 88% and distributed in a narrow range (88.7%–96.4%). It demonstrated an accuracy of 92.8%, a sensitivity of 90.7%, a specificity of 95.6%, a PPV of 96.4%, and an NPV of 88.4%.

Giems Stain for H pylori.—Giems staining underscores the superiority of IHC compared to histochemical stains, at least when it comes to the detection of H pylori. Giems staining underperformed IHC across all indicators. It demonstrated an accuracy of 80.1%, a sensitivity of 74.2%, a specificity of 90.8%, a PPV of 92.7%, and an NPV of 69.2%. These diagnostic indicators were distributed throughout a wide range (69.2%–90.1%).

Mucicarmine Stain for Cryptococcus spp.—The performance of mucicarmine staining for the detection of Cryptococcus spp was poor. Despite a specificity value and PPV of 100%, the remaining diagnostic values were low. The accuracy was 64.2%, the sensitivity was 62.3%, and the NPV was only 12.3%. These diagnostic indicators were distributed throughout a wide range (12.3%–100.0%).

ZN Stain for the Detection of Acid-Fast Bacilli.—As with the mucicarmine stain, review of the scanned images of the ZN stains had significant problems in detecting its targeted microorganism. Of all the detection assays, it was associated with the lowest sensitivity rate (52.7%) and second lowest accuracy rate (67.0%) and NPV (52.1%). The specificity rate was 92.7% and the PPV was 93.0%.

For 175 of 3696 interpretations (0.04%) the pathologist’s read was recorded as equivocal. In the statistical calculations, these equivocal readings were considered discrepant. The Giems category had the highest percentage (6.8%) with 41 equivocal out of 600 occurrences and the H pylori IHC had the lowest (1.6%) with 10 equivocal out of 600 occurrences (Figure 2A). The total number of equivocal values reported by individual pathologists was considered as an additional performance indicator (Figure 2B). The number of equivocal values ranged from a low of 2 (observer 4) to a high of 57 out of 616 interpretations (observer 1).
Interobserver Agreement

Fleiss $\kappa$ analysis was used to evaluate the reliability of agreement between the 6 study pathologists when reviewing the digital images (Table 3). Across all stains evaluated, the Fleiss $\kappa$ coefficient was statistically significant ($P < .001$). There was a substantially good interrater agreement (0.61–0.80) for the interpretation of herpes simplex virus IHC, cytomegalovirus IHC, $H$ pylori IHC, and the GMS stain. The agreement level for spirochete IHC, mucicarmine staining for $Cryptococcus$ spp, and ZN staining for acid-fast bacilli was moderate (0.41–0.60). Interoobserver reliability was poor for the Giemsa category (0.21–0.40). The diagnostic concordance between glass and digital images in each category for different observers is illustrated in Figure 3.

Color Contrast of Digitalized Stains

“Color contrast” refers to how bright and dark colors appear against each other on screens as perceived by the human eye. The difference is indicated with a ratio ranging from 1:1 (lowest possible) to 21:1 (highest possible, as in a black text on a white background). Table 4 indicates the contrast ratio between the organism and background as measured from the digital images. The GMS stain for detection of fungi was associated with the highest contrast ratio (12.52). The mucicarmine stain for the detection of $Cryptococcus$ spp was associated with the lowest contrast ratio (1.08). The ZN stain for detection of acid-fast bacilli also had a very low contrast (2.7) which made the visual detection of pink organism in a background of pale blue very difficult. The brown labeling of the IHC probes provided good contrast ratio when staining was robust as was present with IHC stains for cytomegalovirus (6.7) and herpes simplex virus (7.9). However, the contrast ratio significantly decreases when IHC was weak as in the detection of $H$ pylori (2.5) and spirochetes (1.25).

DISCUSSION

Digital pathology is transforming the practice of diagnostic surgical pathology. The review of digitalized whole slide images is fast, reliable, efficient, and easily learned even by pathologists steeped in the glass slide tradition. But much caution is needed before generalizing the experience with routinely stained slides to other aspects of surgical pathology such as the interpretation of microorganism assays, where unforeseen limitations could adversely impact patient care. The CAP guideline has provided important validation requirements for primary diagnostic purposes, but various diagnostic adjuncts such as microorganism stains still require vigorous validation to identify pitfalls prior to widespread implementation.10–16 A few studies have used a digital platform to correlate detection of select microbes (eg $H$ pylori and mycobacteria18) as a function of resolution and z-stack scanning, but our study is the first assessment of digital

![Figure 2](image)

Figure 2. A, The percentage of equivocal values in each category. B, Number of equivocal values read by each observer.

Table 3. Fleiss $\kappa$ Strength of Agreement Across 6 Pathologists

<table>
<thead>
<tr>
<th>Immunohistochemical Stains</th>
<th>Histochemical Stains</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV</td>
<td>CMV</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.69</td>
</tr>
<tr>
<td>Strength of agreement</td>
<td>Substantial</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviations: CMV, cytomegalovirus; GMS, Grocott methenamine silver; HP, Helicobacter pylori; HSV, herpes simplex virus; ZN, Ziehl-Neelsen.
pathology performance across a spectrum of histochemical stains and IHC stains widely used to detect microorganisms.

Our findings demonstrated a few important trends when applying digital pathology to the detection of microorganisms. First, digital review is generally associated with high (ie 90% or higher) specificity and PPV owing to a low percentage of false positive reads. On the other hand, a high percentage of false negatives often contributed to unacceptably low sensitivity and NPV. As a result, sensitivity and NPV was outperformed by specificity and PPV across all categories. For diagnostic surgical pathology practices utilizing a digital platform, it may be prudent to review the glass slides for all cases interpreted as negative from the digital images, at least until technologic advances can diminish the false negative rate.

A second notable trend was that IHC generally outperforms histochemical stains when it comes to detecting microorganisms from scanned slides. This trend was underscored with *H pylori* detection. Giemsa staining was associated with a low diagnostic confidence level in detecting *H pylori*, and it had the highest percentage of equivocal and false positive values. In contrast, IHC staining for *H pylori* had the lowest number of equivocal values among all categories, no false positives, and high interobserver agreement. Like the Giemsa stain, the other histochemical stains (ie, mucicarmine for the detection of *Cryptococcus* spp and ZN for the detection of acid-fast bacilli) should not be used for screening purposes due to unacceptably low accuracy, sensitivity, and NPV rates coupled with a low degree of interobserver agreement. The notable exception was GMS for the detection of fungi.

There are a number of technical, pathological, and professional factors that may contribute to errors when looking for microbes on digital images, particularly when occurring in combination. Not all of these are unique to the scanned images, but may be exacerbated using the digital platform. We were able to quantitate color contrast from the digital images, and this helped confirm that the contrast between the stained organism and its background may play a major role in microbe detection. There was a strong inverse relationship between high color contrast and false negative results. The size and shape of the microorganism may, understandably, may also affect the ability to perceive their presence. For example, larger objects capture attention in a visual search more than smaller items. Herpes viral inclusions and some fungal species (eg, *Cryptococcus* spp) are relatively large in size and associated with lower false negative reads. Conversely, bacteria such as *H pylori*, spirochetes, and acid-fast bacilli are smaller and associated with higher false negative reads. Organisms with unique shapes were easier to differentiate from background elements and artifacts on the digital images. The recognition of these organisms on the digital images were associated with a lower number of false positives, higher accuracy and specificity, and a lower number of equivocal reads.

<table>
<thead>
<tr>
<th>Stain</th>
<th>Color</th>
<th>Organism</th>
<th>Background</th>
<th>Contrast Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>IHC (Helicobacter pylori, herpes simplex virus, CMV, spirochete)</td>
<td>Brown (strong staining)</td>
<td>White (stroma or no tissue)</td>
<td>7.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gray (nuclei)</td>
<td></td>
<td>6.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Light Brown (weak staining)</td>
<td>White (stroma or no tissue)</td>
<td>2.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gray (nuclei)</td>
<td></td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Fungi (GMS)</td>
<td>Black</td>
<td>Light green</td>
<td>12.52</td>
<td></td>
</tr>
<tr>
<td>Giemsa (<em>H pylori</em>)</td>
<td>Blue</td>
<td>Pale blue</td>
<td>5.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>Blue</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cryptococcus spp (Mucicarmine)</td>
<td>Bright pink</td>
<td>3.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acid-fast bacilli (Ziehl-Neelsen)</td>
<td>Pink</td>
<td>2.7</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: GMS, Grocott methenamine silver; IHC, immunohistochemistry.

Figure 3. The diagnostic concordance between glass and digital slides for each category among different pathologists.
For example, the pink-stained Cryptococcus spp in mucicar- mine stain has the lowest contrast ratio, yet its conspicuous morphology (eg, large size, distinctive capsule) is not easily mistaken for some nonmicrobial tissue element. On the other hand, the fungal hyphae highlighted by the GMS had the highest contrast ratio, but the fragmented hyphae can be easily mistaken for mucus strands, elastic fibers, or other background elements. In general, organisms with round shapes (eg, Cryptococcus spp, Pneumocystis, yeast) were detected on digital images with higher level of confidence than linear (eg, rod-shaped, spiral, filamentous) forms that can resemble mucus strands, fibrin, or staining artifacts, resulting in high number of false positives. Understandably the number of microorganisms impacted on the ease of detection.

Technical aspects of the digital scanner may also contribute to false negative cases. The UFS scanner uses a \( \times 40 \) objective for scanning; therefore, the produced digital images have a resolution of \( \times 40 \). The users could zoom in on scanned images with the image viewer application up to \( \times 100 \). This digital zoom expands and stretches the pixels using pixel replication and zooming algorithms, which subsequently decreases the final resolution and to some degree compromises image quality. Indeed, some studies have shown that higher resolution scanning (\( \times 60 \) objective) results in improved concordance between review of digital images and glass slides in the interpretation of microorganism stains.\(^{17-19}\) Z-plane focusing may be another important tool for enhancing the visibility of microorganisms on digital images. The UFS scanner scans the slides at a single plane. Some next-generation scanners use z-stacking or focus-stacking technology to combine multiple images taken at different focal distances to create 3-dimensional images that offer depth of field visualization. Kalinski et al\(^{20}\) reported that 9 focused z-planes could achieve diagnostic values equivalent to the glass slide. In effect, the sensitivity of detecting acid-fast bacilli on digital images could be improved with newer technology, but at some cost, including increased scan times and greater storage demands. Finally, the inability to adjust light intensity and image contrast in an effort to enhance microorganism detection is a limiting characteristic of all current scanning technology.

The skill set of the pathologist can also have a significant impact on performance. Pathologist factors affecting diagnostic concordance include not just experience and level of concentrated attention, but also familiarity with operating the digital viewer. Activation of the automated navigator mouse while adjusting the movement to an optimal speed can make the viewing process easier and more systematic. Another helpful feature to enhance thorough image review is the tracking indicator (included in the Philips viewer), which highlights uninspected areas of the image. A remarkably high degree of diagnostic concordance was achieved by observer 4 (Figure 3), pointing to something approaching the upper limit of digital pathology performance when shortcomings directly related to the pathologist are minimized, and emphasizes the need for adequate training and continuous feedback to optimize pathologist performance. Some have advocated for a 2-month transition period to improve user performance in digital image interpretation.\(^{10}\)

In summary, our study showed the examination of digital images of herpes simplex virus immunostain offers the highest diagnostic value among the organisms’ stains. To detect the \( H \) pylori in gastrointestinal specimens, the immunostain provides superior diagnostic values and better digital performance compared to Giemsa stain. Our findings also indicated that the interpretation of digital ZN and mucin stains was associated with low interrater reliability, accuracy, sensitivity, and NPV, and should not be used as an alternative to conventional glass slides.
The 2021 CAP guideline\(^1\) indicates that a concordance rate of at least 95% is needed for the digital platform to be used for primary diagnostic purposes. Using this threshold, only \textit{H pylori} IHC was acceptable for digital interpretation. While the performance of digital pathology for ZN in the detection of acid-fast bacilli and mucicarmine stains for the detection of \textit{Cryptococcus} spp is simply too poor to currently be used as an alternative to conventional glass slides, other stains may achieve concordance thresholds following adequate training. Technologic advances including z-plane scanning and artificial intelligence applications may further expand the role of digital pathology in the evaluation of microorganism stains.

References