


Characterizing Diagnostic Search Patterns in Digital Breast Pathology: Scanners and Drillers

Ezgi Mercan¹  · Linda G. Shapiro¹ · Tad T. Brunyé² · Donald L. Weaver³ · Joann G. Elmore⁴

© Society for Imaging Informatics in Medicine 2017

Abstract Following a baseline demographic survey, 87 pathologists interpreted 240 digital whole slide images of breast biopsy specimens representing a range of diagnostic categories from benign to atypia, ductal carcinoma in situ, and invasive cancer. A web-based viewer recorded pathologists' behaviors while interpreting a subset of 60 randomly selected and randomly ordered slides. To characterize diagnostic search patterns, we used the viewport location, time stamp, and zoom level data to calculate four variables: average zoom level, maximum zoom level, zoom level variance, and scanning percentage. Two distinct search strategies were confirmed: *scanning* is characterized by panning at a constant zoom level, while *drilling* involves zooming in and out at various locations. Statistical analysis was applied to examine the associations of different visual interpretive strategies with pathologist characteristics, diagnostic accuracy, and efficiency. We found that females scanned more than males, and age was positively correlated with scanning percentage, while the

facility size was negatively correlated. Throughout 60 cases, the scanning percentage and total interpretation time per slide decreased, and these two variables were positively correlated. The scanning percentage was not predictive of diagnostic accuracy. Increasing average zoom level, maximum zoom level, and zoom variance were correlated with over-interpretation.

Keywords Digital pathology · Diagnostic decision-making · Breast cancer · Breast histopathology · Whole slide imaging · Diagnostic interpretation

Background and Significance

Digital imaging technologies have revolutionized clinical medicine, particularly within diagnostic radiology. In pathology, digital whole slide images (WSIs) are well-established; have proven efficient and reliable for research [1], education [2–5], and archiving [6]; and are now being utilized in pathologic diagnosis [7, 8]. Although they are not approved by the US FDA for primary pathologic diagnosis, digital WSIs are increasingly used to obtain second opinions remotely. In addition to their advantages in clinical settings, the use of computers to interpret digital WSI provides a unique opportunity to study pathologists' viewing behaviors and better understand how their interpretive *strategies* relate to diagnostic accuracy and efficiency.

Pathologic diagnosis is a complex process characterized by visual search and interpretation strategies. Previous research concerning the visual search patterns of physicians has focused on volumetric lung images [9–11], mammography [12], and breast pathology [13, 14]. The method of investigation has usually included eye tracking or video recordings of physicians interpreting medical images in a setting controlled by the experimenter. Three outcomes from published research

Electronic supplementary material The online version of this article (doi:10.1007/s10278-017-9990-5) contains supplementary material, which is available to authorized users.

✉ Ezgi Mercan
ezgi@cs.washington.edu

- ¹ Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA
- ² Department of Psychology, Tufts University, Medford, MA, USA
- ³ Department of Pathology and UVM Cancer Center, University of Vermont, Burlington, VT, USA
- ⁴ Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA

are relevant to the present study. First, physicians reviewing medical images tend to adopt one of two search strategies: *drilling* versus *scanning*. Drilling involves restricting a search to a region of interest and zooming in to high magnification levels. Conversely, scanning involves maintaining a particular zoom level while searching relatively broad regions of interest [11]. Second, search strategies change as a function of acquired experience in an expert domain [9, 15] and prior experience with novel review formats [16]. Third, certain visual search strategies have been associated with greater diagnostic accuracy and efficiency. In radiology, physicians who use a drilling search pattern tend to show higher accuracy and efficiency when detecting lung nodules in volumetric images [11, 17], though no research has explored drilling and scanning strategies by pathologists reviewing non-volumetric images.

To address this knowledge gap, our study attempts to provide an initial understanding of the interpretative strategies pathologists use when reviewing digital slides of breast biopsy specimens. In this study, we investigated three aims. First, we considered how various pathologist characteristics are associated with the two image review strategies (drilling and scanning) identified in the extant cognitive science literature [11]. Second, we tracked how these image review strategies may change as pathologists gain experience with the digital imaging format. Finally, we examined the extent to which each interpretive strategy is associated with diagnostic accuracy and efficiency.

While digital slides are becoming a powerful adjunct tool for breast pathology, understanding the diagnostic processes used by pathologists as they interpret cases may provide insight to improve the education and training of pathologists and lead to the development of computational tools that can aid in the diagnostic decision-making process.

Materials and Methods

Data were collected as part of the Breast Pathology (B-Path) and Digital Pathology (digiPATH) studies. The detailed explanation of methods used for test case development and recruitment of participant pathologists has been previously described [18, 19] and is briefly summarized below.

Case Selection

The 240 excisional ($N = 102$) and core ($N = 138$) breast biopsy specimens were selected from pathology registries in Vermont and New Hampshire using a random sampling stratified by woman's age, breast density ($N = 118$ low density and $N = 122$ high density), and initial diagnosis. New glass slides were prepared from the selected tissue blocks.

The newly prepared glass slides were scanned at $\times 40$ magnification (iScan Coreo, Ventana Medical Systems, Tuscon,

AZ, USA) to create digital WSIs, which were then reviewed by a research technician and a breast pathologist to ensure consistency and quality. A web-based digital viewer, which was developed specifically for this study, allowed users to pan the image and zoom in or out (up to $\times 40$ actual and $\times 60$ digital magnification), providing an interface similar to industry-sponsored WSI viewers but enhanced with study-specific data collection capabilities.

Expert Consensus Diagnosis

The digital WSIs were independently interpreted by three experienced breast pathologists to determine independent diagnoses and representative regions of interest (ROIs); these pathologists then established a consensus diagnosis for each case following a modified Delphi approach in subsequent webinars and in-person meetings [19, 20]. Cases spanned a wide range of diagnostic categories: benign without atypia ($N = 60$), atypia ($N = 80$), ductal carcinoma in situ (DCIS) ($N = 78$), and invasive cancer ($N = 22$). See Supplementary Table 1 for details.

Participants

More than 200 pathologists from across the USA (Alaska, Maine, Minnesota, New Hampshire, New Mexico, Oregon, Vermont, and Washington), who regularly interpret breast biopsy specimens in their clinical practices, were invited to participate in the study. Each participant completed a baseline survey that included demographic data, experience with breast pathology, and perceptions about breast cancer interpretation.

Each participant was randomly assigned to interpret the cases in glass or digital format. A small portion of the participants did not complete the study. In this work, we are using the data collected from 87 pathologists who were assigned to digital format.

Data Collection on Interpretations

The 240 cases were arranged into four sets of 60 cases each that preserved the distribution of diagnostic categories and breast densities of the overall case set. Participants were randomly assigned to interpret one of the four test sets. The order of the 60 cases was randomized for each participant, and they interpreted each case independently, considering histopathological features and accompanying information regarding patient age and biopsy type. After viewing each case, participants were instructed to select all applicable diagnoses on an electronic histology form listing 14 possible diagnostic interpretations. The same categorical mapping scheme was used for participant diagnoses as was used for the expert consensus diagnoses (see Supplementary Table 1).

The study was conducted in two phases so that each participant interpreted the same test set twice, either in glass slide format or in digital slide format or both. The study is explained in detail in [21]. Participants were not informed that they were seeing the same cases in phase II, and the cases were presented in a different order for each participant and also were randomly reordered in phase II.

Detailed tracking data were automatically logged by the web-based digital viewer. As pathologists navigated each slide, the viewer software logged their coordinate positions in the digital WSI, their magnification (zoom) levels, and time stamps.

Tracking data were collected only for those interpreting the cases in digital format in phase II. Half of the participants in phase II were then asked to electronically annotate the digital WSI with an ROI supporting the highest-order (most severe) diagnosis while the other half were not asked to mark an ROI on the digital image. This was done to control for any potential impact of the ROI placement task on the diagnostic decision-making process. The participants randomized to mark the ROI used a tool in the web-based viewer to draw a rectangular ROI following their diagnostic interpretation. The relationship between ROI identification and diagnostic concordance was explored in [20].

Tracking Data Analysis

A viewport scene is a rectangular part of the image that is visible on the pathologist's computer monitor at any time during an interpretation. The time spent on each viewport scene was calculated using logged timestamps. If an entry exceeded a total duration of 1 min, it was excluded under the assumption that the pathologist was not actively interpreting during that time. From the tracking logs, several variables were calculated to characterize the viewing behaviors of each participant, as described in the succeeding sections.

Average Zoom Level, Maximum Zoom Level, and Zoom Level Variance The web-based viewer allowed zoom levels from $\times 1$ to $\times 60$. For each interpretation, viewport tracking logs provided a variable number of zoom level values depending on pathologists' interpretive behavior; for this reason, summary statistics were used to describe zoom level behavior during each interpretation. Average and maximum zoom levels, as well as zoom level variance, were calculated for each interpretation. For each interpretation, we calculated the average zoom level by summing the zoom level values of all viewport scenes and dividing by the number of viewport scenes. Similarly, we calculated the maximum zoom level of

each interpretation and the standard deviation of the zoom level variable as the zoom level variance.

Scanning Percentage We quantified scanning behavior by calculating the percentage of log entries associated with panning behavior (i.e., changing viewport scene coordinates) in each interpretation. Unlike average zoom level, maximum zoom level, and zoom level variance, scanning percentage considers the changes of zoom level in consecutive log entries, regardless of the zoom level itself. In other words, scanning percentage quantifies a behavior that can manifest at different zoom levels. Scanning percentage approaches 100% when the pathologist pans across different areas of the digital image at a constant zoom, and it approaches 0% when zooming in and out at different locations, with less panning or infrequent but long distance pans at a low zoom magnification. For analysis, the scanning percentages were grouped into five categorical variables (0–20, 20–40, 40–60, 60–80, and 80–100%).

Analysis

To assess how pathologist demography influenced interpretive strategy, we modeled our data using repeated-measures regressions, implementing the generalized estimating equation (GEE) approach. The model included ten categorical predictors (factors), as detailed in Table 1. The model used scanning percentage as a linear dependent variable (outcome).

To assess how case order within each set of 60 cases influenced viewing behaviors, we again modeled our data using repeated-measures regressions, implementing the GEE approach. We implemented two models, both including interpretation order as the continuous predictor. We used a linear dependent variable (outcome) for both models: scanning percentage for the first model and total interpretation time per case for the second model.

To assess how interpretive strategy influenced diagnostic outcome, we conducted four separate repeated-measures analyses of variance (ANOVA) with four variables that describe the interpretative behaviors. Each model included one of four continuous variables (average zoom, maximum zoom level, zoom level variance, or scanning percentage) and one of three categorical dependent variables for diagnostic outcome (over-interpretation compared to the expert consensus diagnosis, concordance with the expert consensus diagnosis, and under-interpretation compared to the expert consensus diagnosis). To assess the effect of interpretative behaviors on diagnostic efficiency, we used a repeated-measures ANOVA with a continuous dependent variable (time) and one of five independent categorical variables (scanning percentages 0–20, 20–40, 40–60, 60–80, or 80–100%).

Table 1 Characteristics and average scanning percentages of pathologists ($N = 87$)

Variable	Number of pathologists	Average scanning percentage	p value	Wald chi square
Age at survey (years)				
30–39	10 (11%)	69	0.041	8.251
40–49	25 (29%)	77		
50–59	36 (41%)	75		
60+	16 (18%)	70		
Gender				
Male	57 (66%)	70	0.035	4.439
Female	30 (34%)	82		
Affiliation with academic medical center				
Yes	19 (22%)	77	0.642	0.216
No	68 (78%)	73		
Facility size				
<10 pathologists	55 (63%)	76	0.019	5.484
≥10 pathologists	32 (37%)	69		
Fellowship training in surgical or breast pathology				
No	41 (47%)	75	0.076	3.141
Yes	46 (53%)	73		
Do your colleagues consider you an expert in breast pathology?				
No	70 (80%)	73	0.103	2.666
Yes	17 (20%)	79		
Breast pathology experience (years)				
<20	65 (75%)	76	0.073	3.210
≥20	22 (25%)	68		
Number of breast cases per week				
<5	19 (22%)	73	0.490	1.426
5–9	36 (41%)	75		
≥10	32 (37%)	72		
Marked an ROI				
Yes	44 (51%)	71	0.565	0.330
No	43 (49%)	77		
How confident are you in your assessments of breast cases?				
1 (very confident)	13 (15%)	67	0.100	7.783
2	43 (49%)	75		
3	21 (24%)	75		
4	8 (9%)	77		
5 (not confident at all)	2 (2%)	83		

Results

Viewport tracking data from 87 pathologists, who completed 60 cases in the digital format, were analyzed, producing a total of 5220 interpretations and approximately 1.03 million viewport log entries. Nine hundred seven entries were excluded because they exceeded 1 min in total duration.

Tracking logs were visualized and analyzed to summarize the interpretive strategy of each pathologist. Figure 1 contrasts visualizations representing two different pathologists. The pathologist represented on the left, a *scanner*, chose a consistent zoom level and systematically panned to investigate the whole image. The scanner pathologist

used the same zoom level on the majority of their cases. In contrast, the pathologist represented on the right, a *driller*, zoomed out periodically, selected a new area to view, then zoomed in again. The driller pathologist zoomed in and out on different regions throughout their interpretations. It could be argued that the driller scanned the image with eye movements (rather than screen pans) at a lower resolution to determine areas for drilling. Some of the scanning versus drilling strategies may reflect the pathologist's comfort level when scanning with eye movements at lower magnifications. The scanning percentage for the visualization on the left is close to 100%, while it is closer to 0% for the visualization on the right.

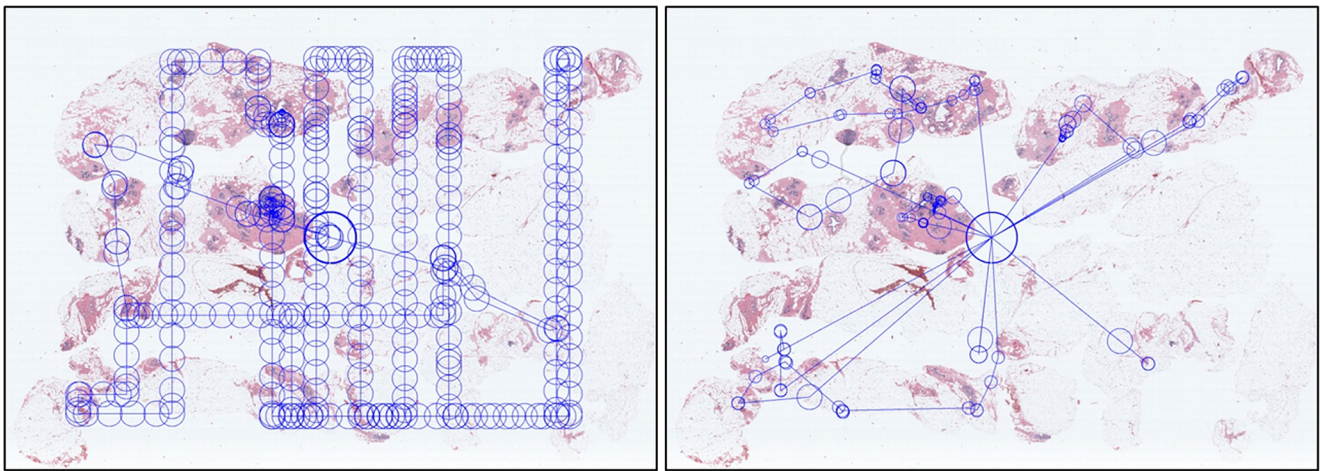


Fig. 1 Visualization of viewport tracking logs, a scanner (*left*) and a driller (*right*), on the same image. Each participant starts at the center of the image with a zoom level of $\times 1$. The *rings* indicate the center of each

viewport, the size of the *rings* indicate the zoom level (the larger the ring, the lower the zoom level), the thickness of the *rings* indicate the time spent at that viewport, and the *lines* connect consecutive viewports

Pathologist Demographics and Viewing Behaviors

Overall, pathologists tended to show scanning percentages exceeding 50% ($\mu = 74\%$, $\sigma = 16\%$), demonstrating a disproportionate trend toward scanning rather than drilling. This pattern was confirmed with a one-sample *t* test comparing to 50%, $t(86) = 13.53$, $p < 0.001$. However, this pattern also varied significantly as a function of certain pathologist demographics.

The GEE model goodness of fit was 1140.98 (QIC), with three significant main effects. First, age positively predicted increasing scanning percentage ($\chi^2 = 8.25$, $p < 0.05$), with higher age groups showing increasingly higher scanning percentages. Second, there were higher scanning percentages among female versus male pathologists ($\chi^2 = 4.44$, $p < 0.05$). Finally, facility group size negatively predicted scanning percentage ($\chi^2 = 5.48$, $p < 0.05$), with pathologists working in larger facility groups

showing lower scanning percentages. No other patterns reached traditional ($\alpha = 0.05$) significance levels.

Interpretation Order

The GEE model showed a significant negative relationship between case position and scanning percentage ($\chi^2 = 16.01$, $p < 0.001$), with scanning percentage decreasing over the course of the 60 cases (see Fig. 2). The total time spent on an interpretation of each case also decreased on average with interpretation order. The participants interpreted later cases in less time compared to earlier cases ($\chi^2 = 67.36$, $p < 0.001$). In a previous study, we found that the diagnostic concordance with the expert panel does not change significantly over the 60 cases interpreted digitally [21].

Fig. 2 Average scanning percentage of 87 pathologists during the interpretation of 60 test cases. The order of the 60 cases was randomized for each pathologist so that the *n*th case included a random sampling of cases from all diagnostic categories

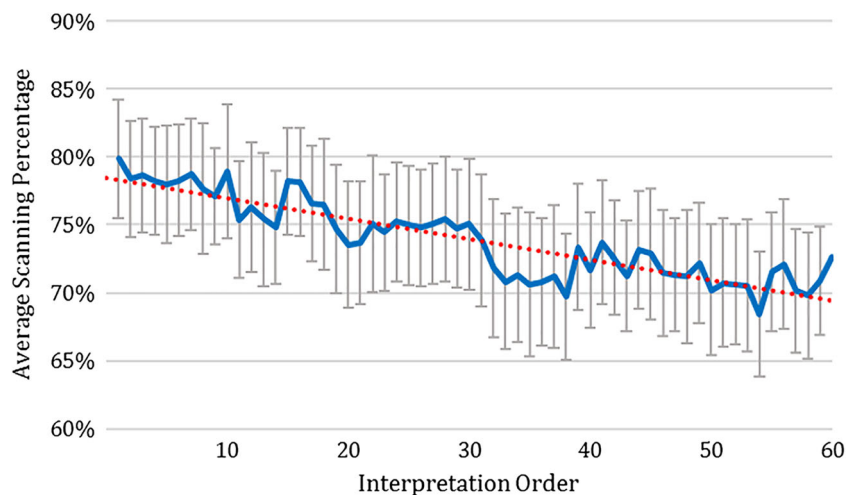


Table 2 Zoom and scanning variables by concordance with expert consensus diagnosis

Consensus diagnosis	Concordance with consensus	Number of interpretations	Average zoom level	<i>p</i> value	Maximum zoom level	<i>p</i> value	Zoom level variance	<i>p</i> value	Scanning percentage	<i>p</i> value
All	Under	760	7.89	≤0.001	24.93	≤0.001	6.22	≤0.001	75%	0.574
	Agree	3672	8.86		27.29		6.98		74%	
	Over	788	9.94		31.87		8.10		73%	
Benign	Under	–	–	≤0.001	–	≤0.001	–	≤0.001	–	0.278
	Agree	933	6.64		22.10		5.28		75%	
	Over	348	9.71		31.49		7.98		72%	
Atypia	Under	492	7.39	≤0.001	23.21	≤0.001	5.78	≤0.001	76%	0.276
	Agree	882	8.88		27.51		7.03		72%	
	Over	384	10.00		31.79		8.13		73%	
DCIS	Under	259	8.74	≤0.001	28.14	0.001	7.03	0.003	73%	0.365
	Agree	1386	8.36		27.66		6.88		74%	
	Over	56	10.95		34.82		8.64		78%	
Invasive	Under	9	10.40	0.312	26.67	0.79	7.73	0.763	74%	0.652
	Agree	471	14.72		36.10		10.57		75%	
	Over	–	–		–		–		–	

Diagnostic Concordance with Expert Consensus Diagnosis

The mean values of the average zoom level, maximum zoom level, zoom level variance, and scanning percentage variables for interpretations are shown by expert consensus diagnosis and concordance with expert consensus diagnosis in Table 2. Supplementary Table 2 provides detailed results of ANOVA tests.

Over-interpretation was associated with increased drilling (average zoom level, maximum zoom level, and zoom level variance). Average zoom level, maximum zoom level, and zoom level variance were higher than the expert consensus diagnosis for over-interpretations and were lower than the expert consensus diagnosis for under-interpretations. The trend was replicated in benign, atypia, and invasive cases. For DCIS cases, both over-interpretation and under-interpretation were associated with higher zoom values. All associations except those for invasive cases were statistically significant.

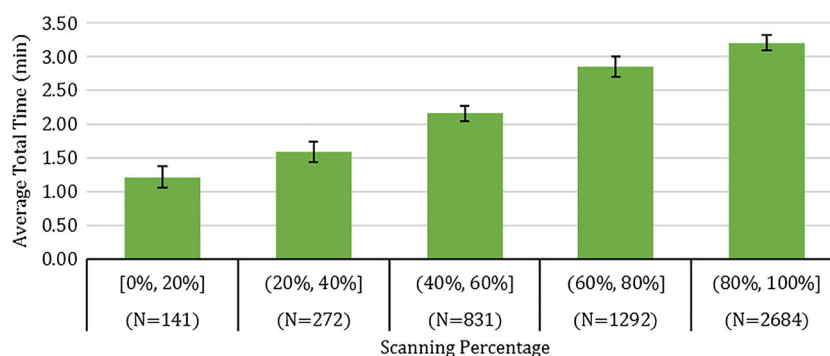
No association was noted between scanning percentage and accuracy (Table 2). Supplementary Fig. 1 shows the average

over-interpretation and under-interpretation rates within different scanning percentage groups.

Diagnostic Efficiency

Efficiency to arrive at an accurate diagnosis was negatively predicted by the extent to which pathologists followed a scanning strategy; in other words, higher scanning percentage was associated with lower efficiency. A repeated-measures ANOVA revealed a main effect of scanning percentage category, $F(4, 52) = 6.72$, $p < 0.001$, demonstrating significantly higher case review times as a function of increased scanning percentage. This pattern is depicted in Fig. 3. Follow-up paired *t* tests demonstrated significant differences between all pairwise category comparisons, with the exception of the first (0–20%) versus second (20–40%) categories and fourth (60–80%) versus fifth (80–100%) categories. In contrast, rates of diagnostic concordance with the expert consensus diagnosis showed no significant difference across scanning percentage groups.

Fig. 3 Average total time of interpretation in five categories of scanning percentage



Discussion

The field of pathology has begun adopting the digital WSI format as it offers great potential for teaching [2–5] and research [1], as well as archival purposes [6] and gathering second opinions [7, 8]. To better understand the visual search patterns used in breast pathology, 87 pathologists across the USA interpreted 60 digital WSIs of breast biopsies representing a range of diagnostic categories, amounting to 5220 individual independent interpretations for analysis.

A web-based viewer tracked and recorded the interpretive behaviors of pathologists as they viewed each digital WSI. The viewer provided pathologists with two possible actions: zooming and panning. Zooming in to an area allowed pathologists to examine cytological, cellular, and nuclear structural details, thereby revealing those that are not as visible to the human eye at lower magnification, but also limiting the portion of the whole slide image viewable on the screen. The panning action allowed pathologists to view neighboring areas of the whole slide image that were not viewable on the screen at higher magnifications.

Combinations of both actions were used by all pathologists to interpret the digital WSI, but interpretive patterns emerged when we analyzed the tracking logs. Specifically, we found that participants varied in their extent of panning and zooming behaviors over time and across cases. *Drilling* behavior showed a relative tendency to zoom in on a particular region, use panning actions sparingly to examine that region, and then zoom out to a lower magnification. In contrast, *scanning* behavior showed a relative tendency to use panning actions to systematically explore the complete image at a constant, and relatively low, magnification. We conceptualize drilling and scanning behavior as two complementary strategies falling at the ends of a bipolar continuum. To quantify image review behavior along this continuum, we calculated the proportion of case review behavior indicative of scanning (i.e., scanning percentage). We wanted to explore potential explanations for the interpretative strategies through their correlation with diagnostic accuracy and efficiency, as well as determining if these patterns change over time.

A number of pathologist demographic characteristics were associated with changes in scanning percentage, including age, gender, and facility size. Higher age was positively correlated with increased scanning percentage, females scanned more than males, and pathologists from smaller facility sizes had higher scanning percentages. It may have been the case that younger participants had relatively more prior experience with similar computer interfaces or image manipulation tools (e.g., mapping software, digital slide viewers, image editing software), thereby making them more comfortable with image drilling behaviors [22]. Although not a statistically significant trend, pathologists with higher scanning percentages also reported lower baseline confidence in their breast pathology

skills. This finding suggests that increased scanning may be related to personality-level (e.g., neuroticism [23]) and/or situation-level (e.g., anxiety [24]) factors. The scanning percentage and total time per slide decreased as pathologists gained experience throughout the set of 60 cases. This suggests a learning curve where participants who started with a scanning-based strategy adopted a more hybrid approach of scanning and drilling as they interpreted through the digital images. This learning curve may be due to prior inexperience with digital slides and computer-based viewing systems that pathologists began to overcome through their experience in this study. Previous research shows a learning curve for interpreting mammograms before and after residency, suggesting a correlation between interpretive behavior and experience [25]. The participants who marked an ROI at the end of their interpretations had lower scanning percentages than those who did not mark an ROI, but the difference was not statistically significant. In other words, we could not find a link between the additional task of looking for a region of interest and the visual search strategy of the participants.

We also noted a pattern of over-interpretation at higher zoom levels. For all diagnostic categories except invasive cancer, the cases that were over-interpreted based on the expert consensus diagnosis had higher values of average zoom level, maximum zoom level, and zoom level variance. This relationship aligns with some research in the cognitive science and visual search literature. Specifically, when observers repeatedly examine a visual scene in detail, the probability of making an erroneous “guess” increases [26]. These inaccurate interpretations likely result from a failed match between perceived image features and stored histopathological features in their memory.

In order to analyze the association of scanning with accuracy and total interpretation time, we divided image interpretations into five categories based on scanning percentage. As scanning percentage increased, so did interpretation time, though rates of over- and under-interpretations were not affected. Scanning was found to be a less efficient strategy for diagnostic interpretation, and the results with the learning curve indicated that pathologists adopted a more balanced and efficient strategy as they progressed through the set of 60 cases.

There are a few reasons why scanning may prove a less efficient, and sometimes less effective [11], method for searching visual images. Scanning at a moderate magnification level involves constantly monitoring and updating past and current positions relative to the entire image space, when only small portions of the overall image can be seen at a time. As demonstrated in prior literature, this type of constant monitoring can be very intensive for working memory, particularly when it is done simultaneously to a more important (primary) task (i.e., identifying malignancy) [27, 28]. In contrast, drilling enables a pathologist to focus attention on a single well-

defined region at a time: examining a single region of interest in great depth and detail, and then iteratively returning to low magnification and examining the next region. In this manner, the searcher need only remember which salient region(s) they have or have not already “drilled into,” which involves monitoring and updating only a representation of salient regions in the low-magnification space. The present results speak to the relative efficiency of drilling, suggesting support for this possibility; however, no research has specifically examined the relative memory cost of employing drilling versus scanning search strategies.

Recent research on volumetric lung images revealed that radiologists adopt distinct visual search strategies during interpretation [11]. Though this earlier research used eye tracking to monitor and interpret visual search patterns, our findings suggest that similar distinctions can be ascertained by recording zooming and panning behaviors. We expect that this is specifically the case with 2D digital pathology images. Indeed, these images require pathologists to zoom in and out dramatically in order to magnify breast tissue and reveal specific structural and cellular features. This process results in high-density zooming and panning data, which is likely uncharacteristic of viewing behavior with narrow slices of volumetric images. The unique characteristics of these breast biopsy digital WSIs may explain why our data did not suggest any influence of visual search strategy (i.e., scanning percentage) on diagnostic accuracy, unlike earlier research with volumetric lung radiographs [11, 17]. Of course, when attempting to identify specific structural or cellular features that were viewed or neglected during the interpretive process, eye tracking is an invaluable technique.

Several notable works in pathology studied the diagnostic search patterns on digital slides [12, 13, 15, 29–31]. The work of Krupinski et al. on breast pathology suggested a link between expertise and search patterns [12, 13, 15, 31]. In our study, the participants with more than 20 years of experience had lower scanning percentages yet the correlation was not significant ($p = 0.1$). However, our participant cohort included only practicing pathologists in comparison to the studies by Krupinski et al. which recruited trainees and experts to examine the changes in the search patterns. The work of Treanor et al. compared the localization errors with interpretation errors in esophageal biopsies and found a trend in which lower zoom levels are correlated with inaccurate diagnosis [29]. Our findings suggest that an opposite trend exists in breast pathology, where higher zoom levels are correlated with over-diagnoses of the pre-invasive lesions, and there seems to be a “happy medium” of magnification for an accurate diagnosis. The existence of a link between magnification and diagnostic accuracy is an important insight, but the nature of the relationship depends on the biopsy type and the visual characteristics of the tissue. Finally, Mello-Thoms et al. described a “focused and efficient” strategy that correlated with the correct outcome

in dermatopathology [30]. In our study, we found that drilling is the more efficient strategy in terms of interpretation time but we did not find any links to the diagnostic outcome. Putting the differences in breast and skin biopsies aside, the selection of the diagnostic categories, the difficulty of the cases, and the demographics of the pathologists are all important factors when comparing two studies. To the best of our knowledge, our work is the first to use an objective quantification of the viewing behavior in a large study.

Limitations and Strengths

This study was limited to one slide per case, which does not reflect actual clinical practice—a factor that may influence diagnostic accuracy but does not preclude evaluation of interpretive strategies. However, the one-slide-per-case study design reduced the workload of participants and allowed them to interpret more images representing a variety of tissue characteristics. This study also diverged from clinical practice in the distribution of diagnostic categories among the cases participants interpreted. Atypia and DCIS cases were over-sampled compared to actual clinical prevalence, with the purpose of better understanding the interpretation of these diagnostically difficult non-invasive cases. Previous research shows that atypia and DCIS cases are more likely to be over-interpreted or misinterpreted, so it is crucial to understand interpretive behaviors on these diagnoses [18]. A possible limitation was the participants’ prior inexperience with the digital format or digital viewers. Although the field of pathology has begun to incorporate digital WSI, most US pathologists are still inexperienced with software for digital WSI interpretation, making it difficult to dissociate the relative contributions of experience with the digital format versus expertise in breast pathology on drilling versus scanning. Similarly, some variation between pathologists may be attributable to participants using their own computer monitors; it is therefore possible that identical monitors may standardize the pathologists’ experience in viewing digital WSI. However, identical monitors do not reflect actual clinical practice, where monitors vary at the level of the practice and, often, between pathologists at the same facility.

Limitations aside, this study is a timely and unique investigation of pathologists’ interpretive strategies with digital media. Strengths of this study include the large sample size of breast biopsy cases ($N = 240$) representing a full spectrum of diagnostic categories from benign and atypia to DCIS and invasive cancer, interpreted and diagnosed by three expert pathologists to define diagnostic accuracy. Another strength is the large number of practicing pathologists ($N = 87$) from across the USA while the previous studies in the literature had recruited medical students and residents in small numbers, i.e., 4 to 11 pathologists. The use of a web-based viewer allowed participants to use their own computers in their own time,

which is as close to the real-world practice of digital pathology as possible. Furthermore, the order of the 60 cases was random for each participant, which allowed us to see a learning curve for the digital slide viewer without case biases attributable to interpretive difficulty or severity of diagnosis.

Conclusions

We identified two distinct interpretive strategies as pathologists viewed digital whole slides of breast biopsy specimens: *scanning*, where the pathologist pans at a constant zoom level, and *drilling*, where the pathologist zooms in and out repeatedly. Our analysis of pathologist characteristics indicated that scanning was more common among women and older pathologists. The facility size, defined as the number of pathologists who worked in the same facility as the participant, was also a significant predictor of the scanning percentage with the participants from smaller facilities scanning more. One possible explanation for this correlation is that the participants from larger facilities could share cases with their colleagues, obtain second opinions, and learn from each other; they would have more experience interpreting breast biopsies. Those who reported less confidence in their interpretation of breast tissue tended to spend more time scanning, but the correlation was not statistically significant.

Regarding accuracy and efficiency, we found that scanning is associated with longer interpretation time on average, yet scanners and drillers had similar levels of accuracy compared to the consensus reference diagnoses. Through our unique study design that randomized the order of cases, we also observed that scanning may be more common at the beginning of a pathologist's experience in interpreting cases in the digital format, while a more balanced strategy of both scanning and drilling is adopted by the end of the 60 cases. We found that average zoom level, maximum zoom level, and zoom level variance for an interpretation increased from under-interpretation to concordance and from concordance to over-interpretation. In other words, when participants under-interpreted a case, they used lower magnifications and changed the zoom level less, as compared to concordant interpretations. Similarly, when they over-interpreted a case, they used higher magnifications and changed the zoom level more, as compared to concordant interpretations. This trend was preserved for all diagnostic categories of breast tissue.

In conclusion, this study demonstrates that two different search strategies are employed by pathologists and these strategies can be explained by a pathologist's demographics, breast pathology perceptions, and prior experience in viewing the digital format. The interpretive strategy can affect the diagnostic outcome and the efficiency of the diagnostic process. These findings motivate further research in medical decision-making

and computerized decision support systems as digital pathology is adopted more widely.

Acknowledgements Research reported in this publication was supported by the National Cancer Institute awards R01 CA172343, R01 CA140560, and KO5 CA104699. The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health. We thank Ventana Medical Systems, Inc. (Tucson, AZ, USA), a member of the Roche Group, for the use of iScan Coreo Au™ whole slide imaging system, and HD View SL for the source code used to build our digital viewer. For a full description of HD View SL, please see <http://hdviews.codeplex.com/>.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Irshad H, Veillard A, Roux L et al.: Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential. *IEEE Rev Biomed Eng* 7:97–114, 2014. doi: [10.1109/RBME.2013.2295804](https://doi.org/10.1109/RBME.2013.2295804)
2. Yin F, Han G, Bui MM et al.: Educational value of digital whole slides accompanying published online pathology journal articles: a multi-institutional study. *Arch Pathol Lab Med* 140(7):694–697, 2016. doi: [10.5858/arpa.2015-0366-OA](https://doi.org/10.5858/arpa.2015-0366-OA)
3. Saco A, Bombi JA, Garcia A et al.: Current status of whole-slide imaging in education. *Pathobiology* 83(2-3):79–88, 2016. doi: [10.1159/000442391](https://doi.org/10.1159/000442391)
4. Kumar RK, Freeman B, Velan GM et al.: Integrating histology and histopathology teaching in practical classes using virtual slides. *Anat Rec - Part B New Anat* 289(4):128–133, 2006. doi: [10.1002/ar.b.20105](https://doi.org/10.1002/ar.b.20105)
5. Bruch LA, De Young BR, Kreiter CD et al.: Competency assessment of residents in surgical pathology using virtual microscopy. *Hum Pathol* 40(8):1122–1128, 2009. doi: [10.1016/j.humpath.2009.04.009](https://doi.org/10.1016/j.humpath.2009.04.009)
6. Gutman D, Cobb J, Somanna D: Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *Med Informatics* 20(6):1091–1098, 2013. doi: [10.1136/amiainl-2012-001469](https://doi.org/10.1136/amiainl-2012-001469)
7. Al-Janabi S, Huisman A, Van Diest PJ: Digital pathology: current status and future perspectives. *Histopathology*. 61(1):1–9, 2012. doi: [10.1111/j.1365-2559.2011.03814.x](https://doi.org/10.1111/j.1365-2559.2011.03814.x)
8. Pantanowitz L, Valenstein PN, Evans AJ et al.: Review of the current state of whole slide imaging in pathology. *J Pathol Inform* 2:36, 2011. doi: [10.4103/2153-3539.83746](https://doi.org/10.4103/2153-3539.83746)
9. Brunyé TT, Carney PA, Allison KH et al.: Eye movements as an index of pathologist visual expertise: a pilot study. *PLoS One* 9(8): e103447, 2014. doi: [10.1371/journal.pone.0103447](https://doi.org/10.1371/journal.pone.0103447)
10. Bahlmann C, Patel A, Johnson J et al.: Automated detection of diagnostically relevant regions in H&E stained digital pathology slides. *Proc. SPIE, Med. Imaging* 8315:831504, 2012. doi: [10.1117/12.912484](https://doi.org/10.1117/12.912484)
11. Drew T, Vo ML, Olwal A et al.: Scanners and drillers: characterizing expert visual search through volumetric images. *J Vis* 13(10): pii: 3., 2013. doi: [10.1167/13.10.3](https://doi.org/10.1167/13.10.3)
12. Tourassi G, Voisin S, Paquit V et al.: Investigating the link between radiologists' gaze, diagnostic decision, and image content. *J Am*

- Med Inform Assoc 20(6):1067–1075, 2013. doi: [10.1136/amiajn-2012-001503](https://doi.org/10.1136/amiajn-2012-001503)
13. Krupinski EA, Graham AR, Weinstein RS: Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Hum Pathol* 44(3):357–64, 2013. doi: [10.1016/j.humpath.2012.05.024](https://doi.org/10.1016/j.humpath.2012.05.024)
 14. Crowley RS, Naus GJ, Stewart J et al.: Development of visual diagnostic expertise in pathology: an information-processing study. *J Am Med Informatics Assoc* 10(1): 39–51, 2003. doi: [10.1197/jamia.M1123](https://doi.org/10.1197/jamia.M1123)
 15. Krupinski EA, Weinstein RS: Changes in visual search patterns of pathology residents as they gain experience. In: *Proceedings of SPIE:79660P*, 2011. doi: [10.1117/12.877735](https://doi.org/10.1117/12.877735)
 16. Velez N, Jukic D, Ho J: Evaluation of 2 whole-slide imaging applications in dermatopathology. *Hum Pathol* 39 (9), 1341–1349, 2008. doi: [10.1016/j.humpath.2008.01.006](https://doi.org/10.1016/j.humpath.2008.01.006)
 17. Wen G, Drew T, Wolfe JM et al.: Computational assessment of visual search strategies in volumetric medical images strategies in volumetric medical images. *J Med Imaging* 3(1):015501, 2016. doi: [10.1117/1.JMI.3.1.015501](https://doi.org/10.1117/1.JMI.3.1.015501)
 18. Elmore JG, Longton GM, Carney PA et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* 313(11):1122–1132., 2015. doi: [10.1001/jama.2015.1405](https://doi.org/10.1001/jama.2015.1405)
 19. Oster NV, Carney PA, Allison KH et al.: Development of a diagnostic test set to assess agreement in breast pathology: practical application of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *BMC Womens Health* 13:3, 2013. doi: [10.1186/1472-6874-13-3](https://doi.org/10.1186/1472-6874-13-3)
 20. Nagarkar DB, Mercan E, Weaver DL et al.: Region of interest identification and diagnostic agreement in breast pathology. *Mod. Pathol.* 29(9):1004–1011, 2016. doi: [10.1038/modpathol.2016.85](https://doi.org/10.1038/modpathol.2016.85)
 21. Elmore J, Longton G, Pepe M et al.: A randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. *J Pathol Inform* 8:12, 2017. doi: [10.4103/2153-3539.201920](https://doi.org/10.4103/2153-3539.201920)
 22. Elias SM, Smith WL, Barney CE: Age as a moderator of attitude towards technology in the workplace: work motivation and overall job satisfaction. *Behav Inf Technol* 31(5):453–467, 2012. doi: [10.1080/0144929X.2010.513419](https://doi.org/10.1080/0144929X.2010.513419)
 23. Newton T, Slade P, Butler NM et al.: Personality and performance on a simple visual search task. *Pers Individ Dif* 13(3):381–382, 1992. doi: [10.1016/0191-8869\(92\)90119-A](https://doi.org/10.1016/0191-8869(92)90119-A)
 24. Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimed Tools Appl*, Published Online First: 15 February 2017. doi:[10.1007/s11042-017-4440-4](https://doi.org/10.1007/s11042-017-4440-4)
 25. Miglioretti DL, Gard CC, Carney PA et al.: When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 253(3):632–640, 2009. doi: [10.1148/radiol.2533090070](https://doi.org/10.1148/radiol.2533090070)
 26. Chun MM, Wolfe JM: Just say no: how are visual searches terminated when there is no target present? *Cogn Psychol* 30(1):39–78, 1996. doi: [10.1006/cogp.1996.0002](https://doi.org/10.1006/cogp.1996.0002)
 27. Miyake A, Friedman NP, Emerson MJ et al.: The unity and diversity of executive functions and their contributions to complex ‘frontal lobe’ tasks: a latent variable analysis. *Cogn Psychol* 41(1):49–100, 2000. doi: [10.1006/cogp.1999.0734](https://doi.org/10.1006/cogp.1999.0734)
 28. Turner ML, Engle RW: Is working memory capacity task dependent? *J Mem Lang* 28(2):127–154, 1989. doi: [10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
 29. Treanor D, Lim CH, Magee D et al.: Tracking with virtual slides: a tool to study diagnostic error in histopathology. *Histopathology* 55(1):37–45, 2009. doi: [10.1111/j.1365-2559.2009.03325.x](https://doi.org/10.1111/j.1365-2559.2009.03325.x)
 30. Mello-thoms C, Mello CAB, Medvedeva O et al.: Perceptual analysis of the reading of dermatopathology virtual slides by pathology residents. *Arch Pathol Lab Med.* 136(5):551–562, 2012. doi: [10.5858/arpa.2010-0697-OA](https://doi.org/10.5858/arpa.2010-0697-OA)
 31. Krupinski EA, Tillack AA, Richter L et al.: Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Hum Pathol* 37(12):1543–1556, 2006. doi: [10.1016/j.humpath.2006.08.024](https://doi.org/10.1016/j.humpath.2006.08.024)