

Eye-Tracking, Mouse Tracking, Stimulus Tracking, and Decision-Making Datasets in Digital Pathology

Veronica Thai^{1,†}, Rui Li^{1,†}, Meng Ling¹, Shuning Jiang¹, Jeremy Wolfe², Raghu Machiraju¹, Yan Hu³, Zaibo Li³, Anil Parwani^{3,*}, and Jian Chen^{1,*}

¹Computer Science and Engineering, The Ohio State University, Columbus, 43210, USA

²Departments of Ophthalmology and Radiology, Harvard Medical School, Department of Surgery, Brigham and Women’s Hospital, Boston, 02115, USA

³Department of Pathology, The Ohio State University Medical Center, Columbus, 43210, USA

*Corresponding authors: Anil Parwani (anil.parwani@osumc.edu) and Jian Chen (chen.8028@osu.edu)

†These authors contributed equally to this work.

ABSTRACT

Interpretation of giga-pixel whole-slide images (WSIs) is an important but difficult task for pathologists. Their diagnostic accuracy is estimated to average around 70%. Adding a second pathologist does not substantially improve decision consistency. The field lacks adequate behavioral data to explain diagnostic errors and inconsistencies. To fill in this gap, we present PathoGaze1.0, a comprehensive behavioral dataset capturing the dynamic visual search and decision-making processes of the full diagnostic workflow during cancer diagnosis. The dataset comprises 18.69 hours of eye-tracking, mouse interaction, stimulus tracking, viewport navigation, and diagnostic decision data (EMSVD) collected from 19 pathologists interpreting 397 WSIs. The data collection process emphasizes ecological validity through an application-grounded testbed, called PTAH. In total, we recorded 171,909 fixations, 263,320 saccades, and 1,867,362 mouse interaction events. In addition, such data could also be used to improve the training of both pathologists and AI systems that might support human experts. All experiments were preregistered at <https://osf.io/hj9a7>, and the complete dataset along with analysis code is available at <https://go.osu.edu/pathogaze>.

Background & Summary

Visual search expertise in digital pathology depends on how pathologists allocate attention across complex tissue landscapes¹⁻³. Collecting gaze data can provide insight into the causes of diagnostic inconsistency⁴, support decision-making⁵, model visual search errors⁶, and enable the development of machine learning methods to assist pathologists in daily diagnostic tasks⁷, or to create a digital twin of a pathologist’s diagnostic process⁸. However, multimodal behavior analyses remain scarce in digital pathology. Existing datasets such as CAMELYON16⁹, were designed for understanding neural network behaviors, but have not been analyzed together with pathologists’ visual search. Here, we captured comprehensive behavioral datasets that integrate gaze and diagnostic actions that can be used to explain human pathologists’ visual and cognitive processes.

Collecting gaze data can also contribute to the community’s efforts to compare, replicate, and extend findings on observer behaviors. Pathologists’ gaze emphasizes task-specific attentional behaviors, thus both bottom-up and top-down processes are involved¹⁰, where participants use both the stimuli and their domain knowledge to allocate regions that stand out. Together, this large data collection enables analyses of visual search behaviors¹⁰ such as “Looked But Failed to See” errors¹¹, failures to recognize², or being “satisfied” too early¹², behaviors that have also been reported in non-pathology medical imaging domains. Furthermore, our data can be used to validate whether mouse-action and eye-gaze are correlated as in other imaging modalities¹³, or help identify the interaction between image features, actions, individual differences, and tasks.

The primary contribution of this paper is a collection of eye-gaze behaviors, mouse tracking, image stimuli, and associated decisions (Figure 1). Our data collection uses application-grounded evaluation, where data collection is conducted within a replication of application tools: our testbed, called PTAH, has assembled features from the state-of-the-art clinical diagnosis platforms. We recruited 19 pathologists across two experiments (10 in Experiment I and 9 in Experiment II), designed to study complementary aspects of diagnostic reasoning. Experiment I (P10S60T600) captured the behavior of 10 pathologists viewing the same set of 60 unique WSIs (15 benign, 45 with metastases) and recorded their annotated region that led to their final decisions. As a result, we can measure behavior consistency between pathologists. In contrast, Experiment II, P9D397T540, asked participating pathologists to look at as many WSIs as possible for broad behavior monitoring. Nine pathologists read in 540 trials over 9.14 hours, which led to 85,605 fixations on 397 CAMELYON16 WSIs.

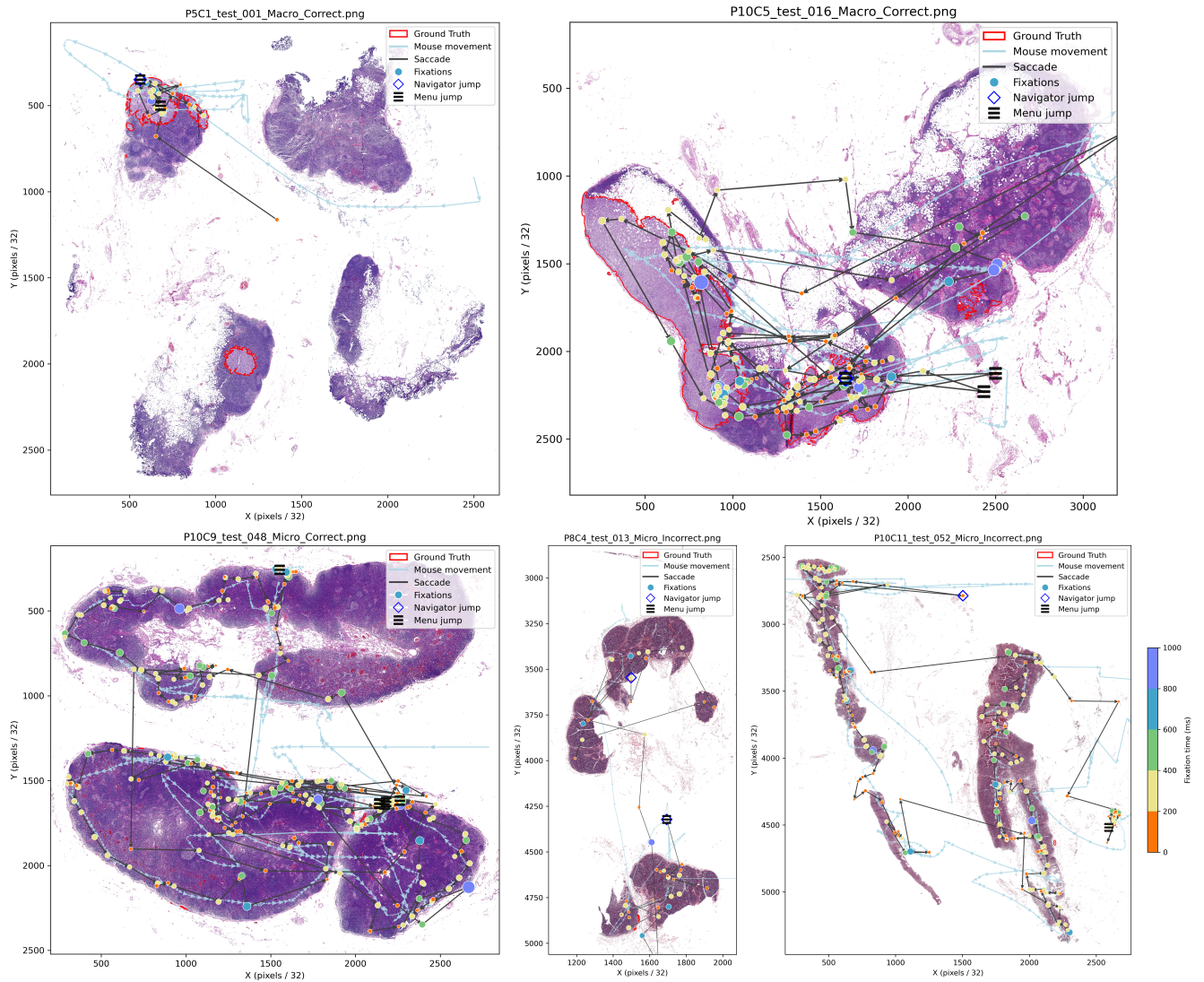


Figure 1. Examples of gaze behaviors from pathologists in PathoGaze1.0. Each whole-slide image (WSI) contains approximately one billion pixels (gigapixel scale). The axis tick labels are scaled down by a factor of 32×32 relative to the original image size. Dots represent fixation points, with color and size indicating fixation duration, while connecting lines denote saccades (scan paths). Both fixations and saccades are overlaid on the corresponding original WSIs. **Observations.** The behavior data captured distinct viewing strategies. For example, for cases with large tumor regions (top row), Participant P5 made a diagnostic decision without scanning all tumor areas, whereas Participant P10 examined most regions before responding. In challenging small-tumor cases, one participant correctly identified a tumor in one slide but misclassified another. Participant P8 exhibited a search error by failing to fixate on the tumor regions (highlighted in red). In general, we observe mouse movement (blue lines and arrows) aligns with gaze points, such as the examples from Participant P10. However, there are a few cases, such as Participant P5, where the mouse does not follow the gaze, in this case, spanning a wider area than the gaze.

Methods

Experimental Design

We recorded *perception-centered* gaze behaviors, *action-centered* navigation behaviors, and subsequent diagnostic decisions in both experiments. In Experiment I, participants also marked tumor regions. The *perception-centered* data captured gaze behaviors that informed search, selection, filtering, and decision processes, while the *action-centered* data tracked pathologists' interactive behaviors such as mouse activities reflecting zooming, panning, dragging, and changes in viewport position.

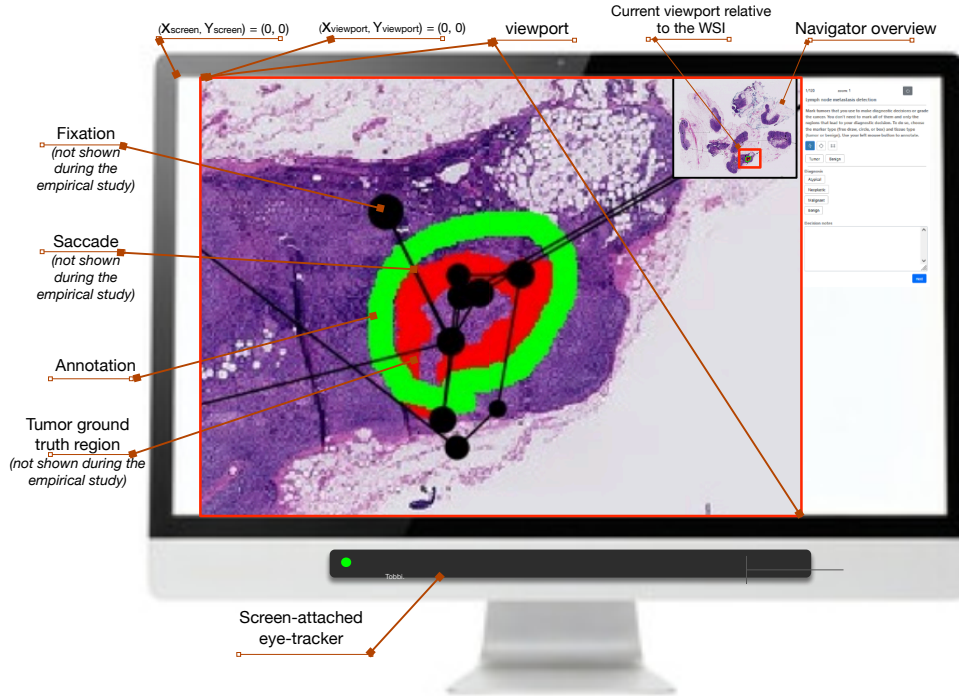


Figure 2. Viewport and screen coordinates. All data are calibrated in the WSI image coordinates, where the upper-left corner of the WSI is (0, 0) and bottom right corner is largest (x_{image} , y_{image}) of the WSI in its full pixel resolution. A viewport is the rectangular area on the screen where the WSI is actually rendered and displayed. It excludes surrounding user interface elements. Each viewport was recorded as its corresponding position on the full WSI. Specifically, we stored the pixel coordinates of the viewport’s upper-left corner in the WSI coordinate system, as illustrated in the navigator overview. Fixations were captured in the screen coordinates and were subsequently transformed to the WSI image coordinates.

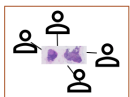
Data Collection Goals and Generic Data Choices

We chose to use the CAMELYON16 competition data⁹, a standard benchmark for digital pathology of breast cancer lymph node samples taken from patients at two medical centers in the Netherlands (University Medical Center Utrecht and Radboud University Medical Center).

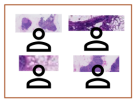
We used it for several reasons: it was carefully curated and contained cases of varying sizes and difficulty levels. Specifically, among the 399 cases, 239 were benign and 160 were cancer metastases with ground-truth locations. Moreover, the dataset includes multiple zoom levels and tumor sizes, allowing us to compare search behaviors across stimuli of different scales. These WSIs were scanned at either 20× magnification for a pixel size of $0.243\mu\text{m} \times 0.243\mu\text{m}$, or 40× for a pixel size of $0.226\mu\text{m} \times 0.226\mu\text{m}$. To label tumor sizes, we followed clinical practice to categorize slides based on tumor diameter and identified eight *isolated* (tumor regions < 0.2mm in diameter), 93 *micrometastases* (tumor regions $\in [0.2, 2)$ mm), and 59 *macrometastases* (tumor regions ≥ 2 mm). The CAMELYON16 dataset is also among the most widely used benchmarks in machine learning for digital pathology. A review of lymph node machine learning papers involving detection, characterization, and segmentation found that of the 41 studies that used public datasets, 26 used CAMELYON16, equivalent to 63% of the studies¹⁴. We next describe the experimental design details for the two data collection experiments.

Experiment I (P10S60T600) Data Selection, Tasks, and Participants.

Experiment I focused on behavior consistency between pathologists. We refer to this dataset as P10S60T600: 10 pathologists (P10) observing the same 60 (S60) breast cancer lymph node slides for a total of 600 trials (T600). These 60 WSIs consist of 15 benign and 45 malignant slides with tumor sizes: 3 isolated, 25 micrometastases, and 17 macrometastases tumors. Participants were invited to perform three tasks grounded on common diagnostic tasks: (1) label the cancerous tissue regions that lead to their diagnosis, (2) rate on a tumor scale, (3) describe their rationale briefly. There was no time limit to finish these tasks. We also instructed pathologist participants not to exhaustively mark all tumors, but the tumor region that influenced their final decision. For non-benign diagnoses, they were required to mark at least one region. Participants were told that annotation was optional if they believed the WSI to be benign.



Experiment II (P9D397T540) Data Selection, Tasks, and Participants.



Experiment II focused on broad coverage and thus no more than two participants viewed the same WSI. We refer to this dataset as P9D397T540. Nine pathologists (P9) each examined a distinct subset of 397 CAMELYON16 breast cancer WSIs (D397), for a total of 540 trials (T540). Two slides were used for training, and each participant viewed a unique subset of 60 slides comprising 36 benign and 24 malignant cases. Benign slides were randomly chosen from the 239 samples and randomly assigned to participants. Malignant slides were assigned using stratified random sampling to balance the tumor-to-tissue ratio across slides. Both sampling procedures were performed without replacement, ensuring that each slide was viewed by no more than two participants. Participants performed the same diagnosis task as in step 2 of Experiment I. However, to preserve fidelity to real clinical practice, they neither marked tumor regions nor verbalized their reasoning (steps 1 and 3 in Experiment I).

Data Acquisition

Digital Pathology Application Grounded Data Collection Testbed (PTAH)

We designed and implemented our application-grounded testbed, PTAH, to record eye-gaze, mouse and keyboard events, and button clicks, so that we could analyze where pathologists looked and the decisions they made. The Tobii Pro Fusion eye-tracker was mounted on the screen, and participants were not stabilized with a chin rest. They sat naturally and were asked to behave as they would in their usual clinical practice. We loaded WSIs using the OpenSeadragon 5.0 library¹⁵ that outputs the current viewport, which was subsequently recorded and converted to the WSI image coordinates (Figure 2). We implemented a broad range of interaction techniques to facilitate navigation in the WSI.

First, participants could hold the left mouse button to pan across the WSI or use the arrow keys to navigate the slide tile-by-tile, ensuring that no region of the tissue was overlooked. Second, we implemented the annotation interface using Annotorious¹⁶, providing three shape options: circle, rectangle, and freehand drawing with the mouse. The green curve illustrates the annotation made by a participant in our experiment. Participants could also edit their annotations by clicking to move or delete them. Each free-hand annotation was recorded as a polyline, a single connected straight-line segments passing through the drawing path. Circular annotations were recorded by their center coordinates and radius, while rectangular annotations were recorded by their position, width, and height. In addition, benign annotation markers were available for participants who wished to label regions they identified as either benign or tumorous. Finally, to capture the diagnostic decisions, a task menu was displayed on the right side of the interface, allowing participants to record their final diagnosis and rationale. During the experiments, these actions and corresponding viewports and zoom-levels were logged.

Gaze and Action-based Data

We recorded gaze using a Tobii Pro Fusion eye-tracker (120Hz for Experiment I and 60Hz for Experiment II) and our testbed. We used the Tobii I-VT Fixation Filter¹⁷ included with Tobii Pro Lab to classify saccades and fixations with the good default configurations (Gap fill-in: disabled, eye selection: average, noise reduction: moving median with a window size of 3 samples, velocity calculator: window length 20ms, I-VT fixation classifier: threshold 30°/s, Merge adjacent fixations: enabled with a max time between fixations of 75ms and a max angle of 0.5°, and discard short fixations: enabled, with a minimum fixation duration of 60ms).

Data Collection Procedure

Pathologists were asked to behave as they would normally in clinical diagnostic settings. All participants completed the study in the same room as they performed clinical diagnosis. The room had natural light and the display was a 27" Philips Barco monitor (MDPC-8127) with a resolution of 2,560 × 1,440, which is a monitor used in pathologists' every day diagnosis. The eye-tracker was calibrated for each participant and re-calibrated after each break. A secondary monitor displays the eye-tracking images for experimenter to monitor the recording process. The monitor was under the desk and was not visible to the participant. Each participant went through training and practice sessions to familiarize themselves with the testbed functions before proceeding to the data collection. They could ask questions during the training and practice but were not allowed during the formal study. Each pathologist performed a tumor rating task from the assigned set of slides in random order. Selecting a diagnostic decision was mandatory before they could proceed to the next slide, and they were not allowed to revisit previously viewed slides. They were also asked to explain their decisions. They took a mandatory two-minute break every 30 minutes. After completing the viewing tasks, participants were interviewed about their subjective viewing experience and attitudes towards eye-tracking.

Participants

Pathologists from The Ohio State University Wexner Medical Center with different levels of experience volunteered to participate in the data collection experiments. Their experience was categorized into four levels: resident, < 5 years, (5–10) years, and > 10 years. The first experiment consisted of 10 participants: three residents, three < 5 years, two (5–10) years, and two > 10 years. The second experiment consisted of nine pathologists: two residents, two < 5 years, two (5–10) years, and three > 10 years.

Computational Data Processing

Coordinate system transformation

The mouse, eye-tracking device attached to the screen, graphical interface, and WSI each operated in distinct coordinate systems (Figure 2). Thus, we need to calibrate these hardware and software reported numbers in a common coordinate, in the WSI image space so the numbers are comparable. Here image coordinates in the WSI image space define pixel positions within the WSI itself, with the origin (0,0) located at the upper-left corner and the coordinate range being the same as the WSI pixel resolution. Screen coordinates, by contrast, are used by the eye-tracking and mouse-tracking devices, with the origin located at the upper-left corner of the computer monitor and the data domain is the monitor resolution of [2,560, 1,440] pixels. The viewport coordinates specify the boundaries of the WSI region currently visible, extracted from the OpenSeadragon. Adjusting the viewport corresponds to panning or zooming operations that determine which portion of the WSI is displayed. In PTAH, the red rectangle shown in the navigation window represented the viewport in image coordinates. Viewers can drag to change the viewport interactively.

Coordinate transformation from screen space to image space. In addition to providing the raw eye-tracker outputs in screen coordinates, we also include the transformed data in which fixation positions are mapped to image coordinates. This transformation is achieved by rescaling the screen-space values and adjusting for the current viewport offset relative to the screen and WSI. The coordinate conversion from screen position to image position was computed as follows:

$$\begin{bmatrix} x_{\text{image}} \\ y_{\text{image}} \end{bmatrix} = \begin{bmatrix} \frac{w_{\text{viewport}}}{w_{\text{window}}} & 0 \\ 0 & \frac{h_{\text{viewport}}}{h_{\text{window}}} \end{bmatrix} \begin{bmatrix} x_{\text{screen}} \\ y_{\text{screen}} \end{bmatrix} + \begin{bmatrix} x_{\text{viewport}} \\ y_{\text{viewport}} \end{bmatrix} \quad (1)$$

where $(x_{\text{image}}, y_{\text{image}})$ denote the image-space coordinates, $(x_{\text{screen}}, y_{\text{screen}})$ the raw screen-space coordinates, and $(x_{\text{viewport}}, y_{\text{viewport}})$ the viewport offset. w_{viewport} and h_{viewport} represent the viewport width and height reported by Seadragon, while w_{window} and h_{window} correspond to the monitor width (2,560 px) and height (1,440 px), respectively.

The image zoom level (*zoom*), representing the displayed slide size relative to the full-resolution WSI, was computed as:

$$\text{zoom} = \frac{w_{\text{displayed}}}{w_{\text{slide}}}. \quad (2)$$

Synchronization between mouse and eye-gaze. After concluding data collection, the eye-tracking data from the Tobii eye-tracker and all other tracking data (mouse, viewport, image, zoom, trial) from our testbed were synchronized for each participant so that each fixation recorded by Tobii Pro Lab was associated with the zoom, viewport, image, and trial recorded by the testbed. We first aggregated gaze samples into single fixation or saccade points (fixation and saccade identification done by the Tobii I-VT filter). The testbed recorded the timestamps from the eye-tracker, allowing synchronization between these eye-tracker fixations and testbed events, so that fixations could be associated with testbed data such as viewport position, image coordinate, and zoom level. Because the testbed and eye-tracker were recording data on different intervals, there were instances where a fixation from the eye-tracker occurred “between” the testbed’s data points. In these cases, the fixation cannot be matched with specific zoom or viewport information. For these fixations, these data were forward-filled, as changes in these data are recorded as a testbed event. Relative timestamps were calculated for each trial based upon the frame the image was loaded. Calculation of peak saccade velocity was done by calculating the peak velocity between adjacent gaze samples within a saccade. The average gaze position between the left and right eye was used.

Each trial was defined as the period from the moment the WSI image finished loading to the time the participant clicked the “Next” button to proceed to the next trial. After synchronizing all data streams, we removed gaze samples recorded outside of these trial intervals, such as when participants were viewing an empty screen prior to loading the image. We also labeled the gaze when

Data Records

In accordance with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles¹⁸, PathoGaze1.0 is structured and documented to maximize transparency, accessibility, and long-term usability. All data files are accompanied by metadata, standardized naming conventions, and consistent formatting to ensure interoperability across analysis platforms. Each file includes descriptive headers and variable definitions, facilitating reuse by both computational researchers and domain experts. Persistent identifiers and public repository hosting further ensure the dataset’s discoverability and accessibility in alignment with open science standards.

Figure 3 illustrates the hierarchical organization of files and metadata in our online database. To facilitate subsequent analyses, we generated separate eye-tracking and mouse-tracking data files for independent examination of each modality, as well as a combined, time-synchronized version. The collected dataset includes the following data files: ImageMetrics (shared



Figure 3. Data directory structure. Shared resources such as common processing scripts and image metric data are placed at the root level. Experiment specific data are organized into subdirectories named after the first experiment, P10S60T600, and the second, P9D397T540. Both subdirectories contain the same type of data with the same organization: experiment specific code, raw testbed and eye-tracker data, screen recordings, and processed data. Publicly available via <https://go.osu.edu/pathoems>.

CAMELYONImageName	experimentImageName	width(px)	height(px)	tissueArea(px)	nonTissueArea(px)	diagnosis	tumorDiameter(mm)	tumorSize
test_001	C1	8.60×10^4	8.96×10^4	1.17×10^6	7.71×10^9	Malignant	2.29	Macro
test_002	C41	9.78×10^4	2.21×10^5	1.45×10^6	2.16×10^{10}	Malignant	1.34	Micro
test_004	C42	9.83×10^4	1.04×10^5	1.60×10^6	1.02×10^{10}	Malignant	0.16	Isolated
test_014	C43	1.11×10^5	1.00×10^5	2.17×10^6	1.11×10^{10}	Benign	0	Benign
test_053	C12	9.78×10^4	2.19×10^5	3.39×10^6	2.14×10^{10}	Benign	0	Benign
test_099	C30	9.83×10^4	8.60×10^4	1.58×10^6	8.45×10^9	Malignant	0.08	Isolated
test_105	C33	1.27×10^5	9.32×10^4	5.16×10^6	1.18×10^{10}	Malignant	23.50	Macro

Table 1. Excerpt of the WSI information. We tabled information for each WSI. Images consisted of different diagnoses (benign and malignant) and tumor sizes. Tumor regions were classified as isolated tumor cells (<0.2 mm), micrometastases ($[0.2, 2]$ mm), or macrometastases (>2 mm) based on tumor diameter, while non-tumorous regions were labeled as benign.

between both experiments and stored in the root directory), and within each experiment’s subdirectory, several CSV files, GazeOnlyData, GazeMetrics, MouseOnlyData, CombinedGazeMouseData, and Metadata. A Readme file is also provided, describing the structure and content of each data file.

Tissue Regions in WSIs

While both experiments were performed on the original WSIs, we isolated tissue regions (foreground) from the background for subsequent data analyses. Defining tissue regions enables the quantification of gaze positions within or outside tumor areas.

We identified the tissue background using Otsu’s thresholding method¹⁹. For each WSI, this algorithm returns a single intensity threshold that separate pixels into two classes of foreground and background; and the threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance. It analyzed the image’s histogram of pixel intensities, assuming a bimodal (two-peaked) distribution, and then calculating the variance from all possible threshold splits to identify the best separating threshold. This process results in a binary image (Figure 4). The lower bounds for the H and S channels were determined automatically by Otsu’s method, while the lower bound for the V channel was fixed at 70 to retain well-saturated tissue regions and exclude shadows or dark areas. The upper bounds for H and S were set to 180 and 255, respectively, and the upper bound for V was determined by Otsu’s method to exclude overly bright background.

Image Metrics Data

Located in the root directory (ImageMetrics.csv), this file consists of data regarding the images themselves. Columns A and J were provided by CAMELYON16. The remainder of the columns were calculated by us (Table 1).

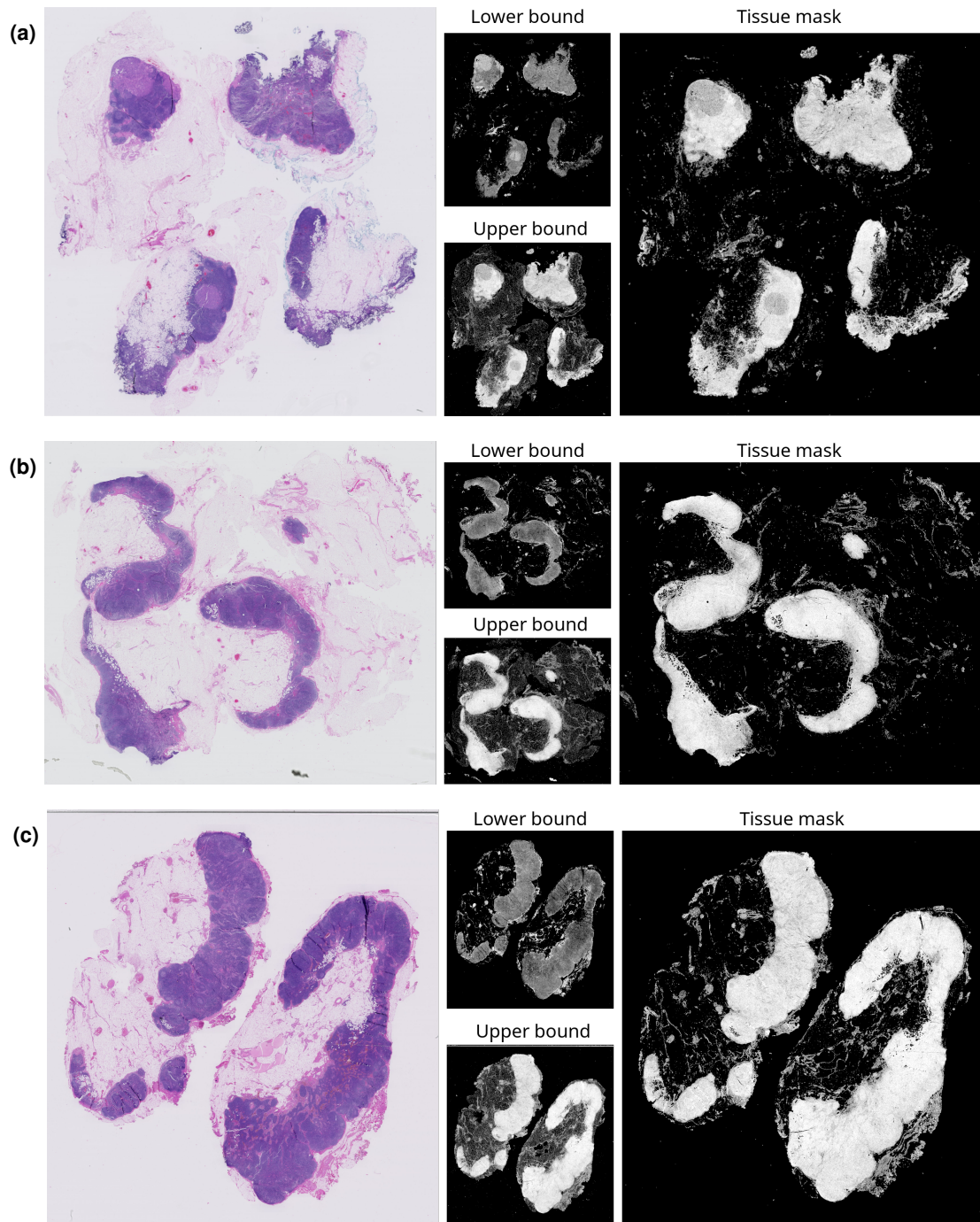


Figure 4. Tissue background processing results. We use Otsu's method to threshold the HSV values of the original images (left-column) and use these to create upper and lower color bounds for the image (mid-column). The lower bound recognizes areas of high saturation, and ignores darker areas, or shadows. The upper bound ignores bright areas, representing the white background. These bounds are combined (right-column) to create the final tissue mask.

experimentImageName	participantID	trialID	taskCompletionTime(ms)	diagnosis	decisionNotes	annotations	correct
C48	P1	1	36,139	Benign	Benign	{ }	FALSE
C4	P1	2	37,117	Benign	Benign	{ }	FALSE
C58	P1	3	65,717	Malignant	Malignant	{ "Tumor": "(35,995.179..."	TRUE
...	...	...	...	...	...	...	...
C14	P9	5	456,833	Benign	Dense fibr...	{ "Benign": "(40,382.11..."	TRUE

Table 2. Excerpt of an Experiment I metadata file. These files contain information about the trials themselves. Participants answers are provided in the diagnosis, diagnosisNotes, and annotations columns. The diagnosis is selected from the following options: “Benign”, “Atypical”, “Neoplastic”, and “Malignant”. Annotations are provided in JSON format, with an annotation type of “Tumor” or “Benign” and a list of the annotation coordinates in counterclockwise direction.

A. The **experimentImageName** is the name used by the CAMELYON16 dataset (test_001, normal_001, etc.). **B.** The **CAMELYONImageName** is a shortened ID identifying the slide used for the testbed. **C. width(px)** and **D. height(px)** are the width and height of the image in pixels, respectively. **E. tissueArea(px)** and **F. nonTissueArea(px)** are the areas of the foreground and background, respectively, calculated by Otsu’s method. **G. diagnosis** is the ground truth diagnosis of the slide. **H. tumorDiameter(mm)** is the diameter of the largest tumor region. **I. tumorSize** denotes the slide-level classification based on the diameter of the tumor region: Benign, Isolated Tumor Cells, Micrometastasis, or Macrometastasis. Isolated Tumor Cells measure less than 0.2 mm in diameter, Micrometastases are greater than 0.2 mm but less than 2 mm, and Macrometastases are larger than 2 mm. **J. tumorMaskFile** is the link to the ground truth tumor mask for this image. **K. tumorMaskThumbnail** is a thumbnail image of the ground truth tumor mask.

Participant Experience Data

Located in the root directory, (`Experience.csv`) contains the experience level grouping of the participants.

A. The **participantID** is a pseudonymous ID identifying the participant. **B.** The **experience** is the experience level, one of these values: resident, < 5, 5–10, or > 10, to indicate a resident, pathologist with less than 5 years of experience, 5–10 years, or more than 10 years of experience, respectively.

Eye-, Mouse-, Viewport, WSI Region-Tracking, and Decision Dataset

Experiment I (P10S60T600) and Experiment II (P9D397T540) used the same data storage format.

Metadata

This file contains data from participants’ trials, including summary information about their gaze data and trial-level attributes such as completion time and responses. This file is named `Metadata.csv` and appears under the `/P10S60T600` and `/P9D397T540` subdirectories. Its columns **A-H** are (Table 2):

A. The **experimentImageName** is a shortened ID identifying the slide used for the testbed. **B.** The **participantID** is a pseudonymous ID identifying the participant. **C.** The **trialID** is the trial number for this participant. The first image a participant sees will have a **trialID** of 1, the second image a **trialID** of 2, and so on, up to 60. **D.** The **taskCompletionTime** is the time in milliseconds for the participant to complete their tasks. **E. diagnosis** records each participant’s response to the diagnosis task, which can be Benign, Atypical, Neoplastic, or Malignant. **F. decisionNotes** contain notes provided by the pathologist explaining their diagnostic reasoning. **G. annotations** list the pathologists’ annotations in JSON format, with the annotation type as “Benign” or “Tumor” with a list of points in image-coordinates. **H. correct** represents whether the participants’ diagnosis was correct or not, with TRUE representing correct, and FALSE incorrect.

Eye-Tracking Data

This file contains recorded temporal gaze data along with viewport information. It includes all participants’ fixations and saccades and the viewports (or the corresponding visible areas of the WSI images). We provide both separated and combined gaze-tracking and mouse-tracking data files, to facilitate both gaze-specific or mouse-specific analyses. Each row in the gaze data file corresponds to a single event: either a fixation, a saccade, or the beginning or end of a trial. The trial start and end entries include values only for the `imageID`, `participantID`, `trialID`, and `relativeTimestamp` columns. We provide the data in two formats: the original screen (viewport) coordinates and the corresponding coordinates transformed into the WSI image coordinate system (Table 3). This file, named `GazeOnlyData.csv`, is located within the `/P10S60T600` and `/P9D397T540` subdirectories. The columns **A-X** are:

A. The **experimentImageName** is a shortened ID identifying the slide used for the testbed, ranging from “C1” to “C397”. **B.** The **participantID** is a pseudonymous ID identifying the participant, ranging from “P1” to “P10” for the first experiment and “P9” for the second. **C.** The **trialID** is the trial number for this participant. The first image a participant sees will have a **trialID** of 1, the second image a **trialID** of 2, and so on, up to 60. **D. EyeMovementType** indicates whether this row is

experimentImageName	participantID	trialID	EyeMovementType	FixationPointX(px)	FixationPointY(px)
C48	P1	1	Saccade Fixation	1,861	331
C48	P1	1			
C48	P1	1			
.....					
C48	P1	1			
Continued					
ImageFixationPointX(px)	ImageFixationPointY(px)	EventEyesPositionX(mm)	EventEyesPositionY(mm)	EventEyesPositionZ(mm)	peakVelocity(deg/s)
171,996.33	50,832.50	266.02	96.74	622.39	107.91
.....		267.83	95.6	624.58	
103,128.42	81,090.34	314.49	84.53	669.18	
Continued					
ViewportUpperLeftX(px)	ViewportUpperLeftY(px)	ViewportWidth(px)	ViewportHeight(px)	relativeTimestamp(ms)	GazeEventDuration(ms)
-113,829.89	0	32,5451.78	221,184	0	58
-113,829.89	0	325,451.78	221,184	148	125
.....				206	
				36,139	
Continued					
outOfScreen	inMenu	inImage	inNavigator	zoomScale	zoomMagnification
.....			0.0065	1	
.....					

Table 3. Excerpt of an eye-tracking data file. The eye-tracking data files provide information for each fixation and saccade. Additionally, these data files have entries for the start and end of each trial for timing purposes. These entries represent events associated with participants’ fixations we have calculated in this study.

a fixation, a saccade, or a start or end row. It takes the value “Fixation” for fixation rows and “Saccade” for saccade rows. **E. FixationPointX(px)** and **F. FixationPointY(px)** are the x- and y-coordinates in pixels, respectively, of each fixation on the computer screen. **G. ImageFixationPointX(px)** and **H. ImageFixationPointY(px)** represent the x- and y-coordinates, respectively, of each fixation on the WSI image, measured in pixels. **I. EyesPositionX(mm)**, **J. EyesPositionY(mm)**, and **K. EyesPositionZ(mm)** represent the average eye position during this event, in millimeters, with the upper-left corner of the screen representing (0, 0, 0). **L. peakVelocity(deg/s)** is the peak velocity of the saccade in degrees of visual field per second. This column only has a value for saccade rows. **M. ViewportUpperLeftX(px)** and **N. ViewportAreaUpperLeftY(px)** are the x- and y-coordinates of the upper left corner of the PTAH viewport on the WSI image. **O. ViewportWidth(px)** and **P. ViewportHeight(px)** are the width and height of the viewport in teh WSI image space. **Q. relativeTimestamp(ms)** contains the time, in milliseconds, that the fixation occurred relative to the moment the image first appeared on the screen. **R. GazeEventDuration(ms)** is the fixation duration in milliseconds. **S. outOfScreen** is a TRUE or FALSE value indicating whether this fixation fell outside of the monitor screen. **T. inMenu**, **U. inImage**, and **V. inNavigator** are TRUE or FALSE values indicating whether this fixation fell in the task menu, in the displayed image, and in the navigator window, respectively. **W. zoomScale** contains the zoom level of the image relative to the full size WSI image resolution. **X. zoomMagnification** gives the zoom level of the image relative to the fully zoomed-out WSI image.

Action Data

We recorded all participants’ mouse movements and state of the PTAH viewports. Each row represents a distinct mouse interaction or the beginning/end of a trial (Table 4). This file, named `MouseOnlyData.csv`, is located in the folders `P10S60T600` and `P9D397T540` subdirectories. Its columns **A-R** are: **A. experimentImageName** is a shortened ID identifying the slide used for the testbed, ranging from “C1” to “C397”. **B. participantID** is a pseudonymous ID identifying the participant, ranging from “P1” to “P10” for the first experiment and “P9” for the second. **C. trialID** is the trial number for this participant. The first image a participant sees will have a **trialID** of 1, the second image a **trialID** of 2, and so on, up to 60. **D. MousePositionX(px)** and **E. MousePositionY(px)** are the x- and y-coordinates of the mouse cursor, respectively, on the screen. **F. ImageMousePositionX(px)** and **G. ImageMousePositionY(px)** are the x- and y-coordinates of the mouse cursor, respectively, in the WSI image coordinates. **H. ViewportUpperLeftX(px)** and **I. ViewportUpperLeftY(px)** are the x- and y-coordinates of the upper left corner of the PTAH viewport in the WSI image coordinates. **J. ViewportWidth(px)** and **K. ViewportHeight(px)** are the width and height of the viewport in the WSI image coordinates. The **L. relativeTimestamp(px)** gives the time the fixation occurred relative to the moment the image first appeared on the screen. **M. eventType** indicates the

experimentImageName	participantID	trialID	MousePositionX(px)	MousePositionY(px)	ImageMousePositionX(px)
C48	P1	1			
C48	P1	1	1,301	718	85,990.4
C48	P1	1	1,308	718	87,065.6
.....					
C48	P1	1			
<i>Continued</i>					
ImageMousePositionY(px)	ViewportUpperLeftX(px)	ViewportUpperLeftY(px)	ViewportWidth(px)	ViewportHeight(px)	relativeTimestamp(ms)
	-113,829.89	0	325,451.78	221,184	0
110,284.8	-113,829.89	0	325,451.78	221,184	288
110,284.8	-113,829.89	0	325,451.78	221,184	304
.....					
	33,542.66	79,301.60	64,249.34	43,708.77	36,139
<i>Continued</i>					
eventType	inMenu	inImage	inNavigator	zoomScale	zoomMagnification
image_loaded				0.01	0.96
mouse_move	FALSE	TRUE	FALSE	0.01	1
mouse_move	FALSE	TRUE	FALSE	0.01	1
.....					
next_btn_clicked				0.03	5.06

Table 4. Excerpt of a mouse-tracking data file. These mouse events are temporally aligned to the eye-tracking data and are recorded for the same start and end of each trial as the eye-gaze data.

Metric name	Definition	Rationale	Source
Fixation count	Number of fixations in the stimulus	Fewer fixations indicates less efficiency	Poole and Ball ²⁰
Orienting fixation count	Number of “orienting” fixations, before mouse interaction with the image	The orienting fixations reflect the participant familiarizing themselves with the image, not active search	This paper
Fixation time	Total time of all fixations for a stimulus	Compares amount of attention on different AOIs or stimulus	Poole and Ball ²⁰
Average fixation duration	Average duration of fixations in a stimulus	Longer indicates more time spent analyzing and interpreting the content, or more mental effort	Poole and Ball ²⁰
Saccade count	Total number of saccades	More saccades indicate more searching, related to mental workload and cognitive processes	Fritz et al. ²¹
Saccade duration	Total duration of all saccades	Related to mental workload and cognitive processes	Fritz et al. ²¹
Saccade amplitude	Degrees of visual field covered by the saccade	Indicates meaningful load cues, higher amplitude indicates lower effort	Poole and Ball ²⁰
Saccade peak velocity	Maximum speed within a saccade (deg/s)	Related to physiological arousal, mental workload, or predicted value of info	Brunyé et al. ²²

Table 5. Eye-tracking metrics. Metrics we have calculated using eye-tracking data, categorized by fixation-related metrics, duration-relation metrics, and saccade-related metrics.

type of mouse event that occurred. Its possible values are “mouse_move”, “mouse_drag”, and “mouse_scroll.” **N. inMenu**, **O. inImage**, and **P. inNavigator** are TRUE or FALSE values indicating whether this fixation fell in the task menu, the displayed image, and the navigator window, respectively. **Q. zoomScale** contains the zoom level of the WSI image, relative to its full size resolution. **R. zoomMagnification** stores the zoom level, relative to the fully zoomed-out WSI.

Merged Perception-Action Data

The previous two datasets provided separate gaze and mouse events. This merged perception-action data file integrates both gaze and mouse data for combined analysis, with all events sorted by timestamp. It is created by concatenating the eye-tracking and mouse-tracking data rows into a single file. All files have been synchronized to the start of the trial. Eye-tracking data rows do not contain entries for mouse-specific columns (e.g., MousePositionX(px) and MousePositionY(px)), and vice versa. This file, named `CombinedGazeMouseData.csv`, is located in the `P10S60T600` and `P9D397T540` subdirectories.

Overall Eye-Tracking Metrics Data

This file contains trial-level eye-tracking metrics, with the exception of saccade peak velocity, which is a per-saccade metric and is instead included in the Eye-Tracking Data file (Table 5). This file is named `GazeMetrics.csv` and appears under the `P10S60T600` and `P9D397T540` subdirectories. The columns are: **A.** The **experimentImageName** is a shortened ID identifying the slide used for the testbed, ranging from “C1” to “C397”. **B.** The **participantID** is a pseudonymous ID identifying the participant, ranging from “P1” to “P10” for the first protocol and “P9” for the second. **C.** The **trialID** is the trial number for this participant. The first image a participant sees will have a **trialID** of 1, the second image a **trialID** of 2, and so on, up to 60. **D. totalFixations** is the total number of fixations recorded for this participant for this trial/image viewing. Fewer fixations in

experimentImageName	participantID	trialID	totalFixations	totalOrientingFixations
C48	P1	1	100	15
C4	P1	2	111	13
C58	P1	3	188	10
.....				
<i>Continued</i>				
totalFixationDuration(ms)	averageFixationDuration(ms)	totalSaccades	totalSaccadeDuration(ms)	averageSaccadeAmplitude(deg)
28,742	287.42	110	3,914	5.76
28,987	256.52	138	4,160	5.02
52,125	277.26	214	7,388	5.40
.....				

Table 6. Excerpt of an eye-tracking metrics file. The eye-tracking metrics files provide trial-level eye-tracking metrics.

the AOI relative to total fixations indicates this search was less efficient. **E. totalOrientingFixations** are the fixations before the user began mouse navigation, that is, before the first zoom. It reflects the participant familiarizing themselves with the image before actively searching. **F. totalFixationDuration(ms)** and **G. averageFixationDuration(ms)** are the total and average duration, respectively, of all fixations for this trial in milliseconds. These allow comparison of attention on different stimuli, with longer times suggesting more time analyzing and interpreting the image, or more mental effort. **H. totalSaccades** is the total number of saccades recorded for this participant for this trial/image viewing. More saccades indicate more searching and are related to mental workload and cognitive processing. **I. totalSaccadeDuration(ms)** is the total duration of all saccades, also related to mental workload and cognitive processing. **J. averageSaccadeDuration(ms)** is the average duration of all saccades. **K. averageSaccadeAmplitude(deg)** is the average amplitude of saccades in degrees of visual field, related to physiological arousal and mental workload.

Technical Validation

We manually checked all trials against the screen-recording in order to verify that the fixations in our processed data matched those directly recorded by Tobii Pro Lab. Eye-tracking and mouse-tracking data were synced by aligning timestamps recorded by the eye-tracker in the Tobii data output with those logged by the testbed, allowing mouse events to be mapped to the eye-tracker's clock. Additionally, we conducted three selected analyses on the data to understand the specificity of pathologists' search in large WSIs. Since none of the studies have been conducted using the same dataset, these validations allow for cross-study comparisons to better understand the unique challenges of the pathology domain.

Compare Medical Images Functional Visual Field

Wu and Wolfe examined functional visual fields (FVFs) in various visual search tasks (T among L search and conjunction) to examine how search targets may be missed²³. They plotted the FVFs at different task stages (search, targeting, and post-target), showing the distribution of saccade start points relative to their end positions. These plots represent the regions from which viewers tend to initiate their next eye movement.

In our FVFs, we observed a clustering of saccades around the origin, with larger amplitudes along the horizontal axis compared to the vertical. A greater proportion of saccades occurred in the cardinal directions (up, down, left, and right), exhibiting the same pattern reported by Wu and Wolfe (Figure 6).

Compare Generic Saccade Sequence Behaviors

Gibaldi and Sabatini modeled the saccadic eye movement as an indicator of oculomotor performance²⁴ for two tasks: following a cross-shaped target on the screen, and free exploration of a natural scene. The results support that the main sequence model, i.e., the predictable relationship between the amplitude (size) of a saccade and its other properties (e.g., duration and peak velocity), can be used to model saccade.

Here, we plotted the main sequences relationship between saccade amplitude and peak saccade velocity (Figure 7). We observed the same trend reported by Gibaldi and Sabatini: a generally increasing trend in which larger saccades exhibit higher velocities, leveling off around a saccade amplitude of approximately 10°. We fitted an exponential curve, consistent with the model that Gibaldi and Sabatini reported as having the highest goodness of fit²⁴. We then calculated the root mean squared error (RMSE) and R^2 . The results showed moderate RMSE values, ranging from 30.06°/s to 75.87°/s. We compared our R^2 values with those reported by Gibaldi and Sabatini, who generally observed $R^2 > 0.9$ ²⁴. Our R^2 values were lower, with highest being $R^2 = 0.64$. This difference is likely attributable to the inherently challenging visual search within WSIs, where pathologists must continuously zoom and pan across a large image space.

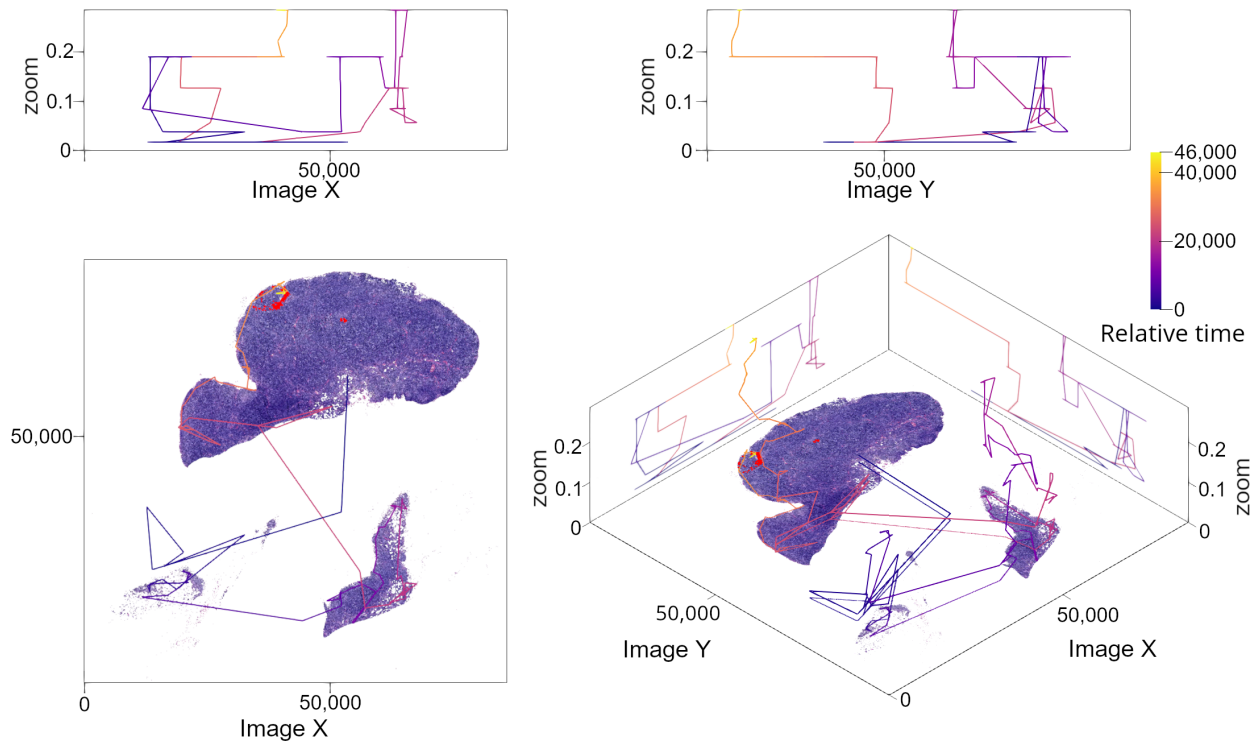


Figure 5. Example 3D scanpaths behaviors in orthogonal views. The top-down, front-left, front-right, and 3D view of a participant's behaviors. The scanpath is colored by the relative timestamp and the ground truth tumor region is marked in green. **Observations.** This participant started at the large tissue area at the top of the slide, then viewed the areas below at low magnification. On the right region, they zoomed in to view that area in higher detail before zooming back out. When they viewed the tumor area, they zoomed in again before finishing the trial.

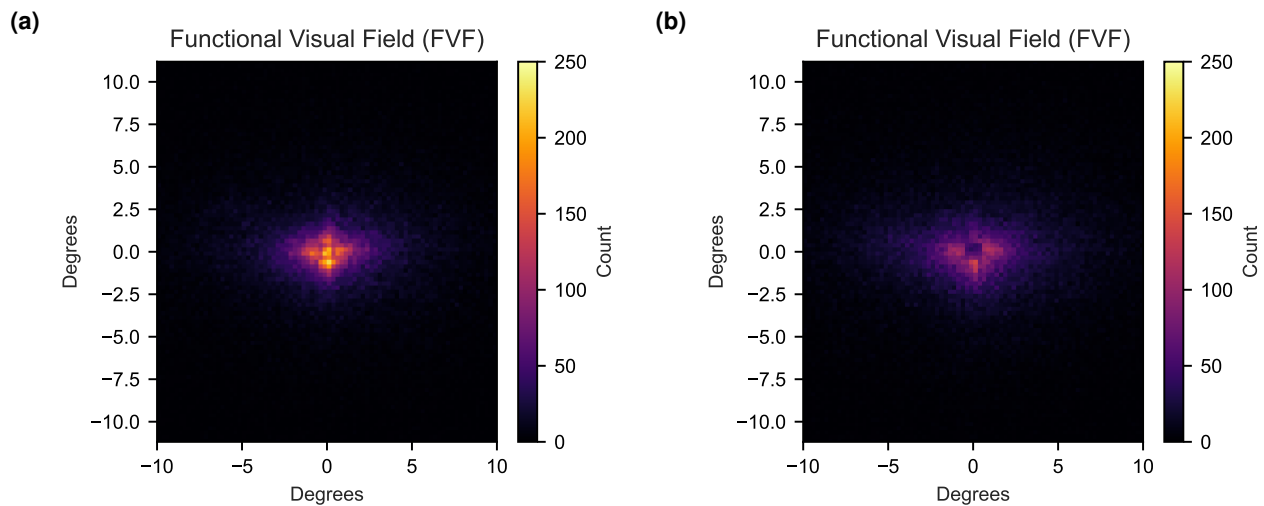


Figure 6. Technical Validation I: Functional visual fields. We plot the functional visual field for Experiment I (a) and Experiment II (b), showing the saccade origin points relative to the end point for the combination of all participants and their trials. **Observations.** (1) In general, greater amplitudes in the horizontal direction compared to the vertical. (2) A more disperse field appeared in the second experiment.

Compare Domain Specific Zooming and Scanning Behaviors

Drew et al.²⁵ studied pathologists' diagnostic accuracy in relation to their zooming/panning behaviors. They found that increased panning (i.e., greater movement across the image) was associated with higher accuracy, indicating a greater likelihood

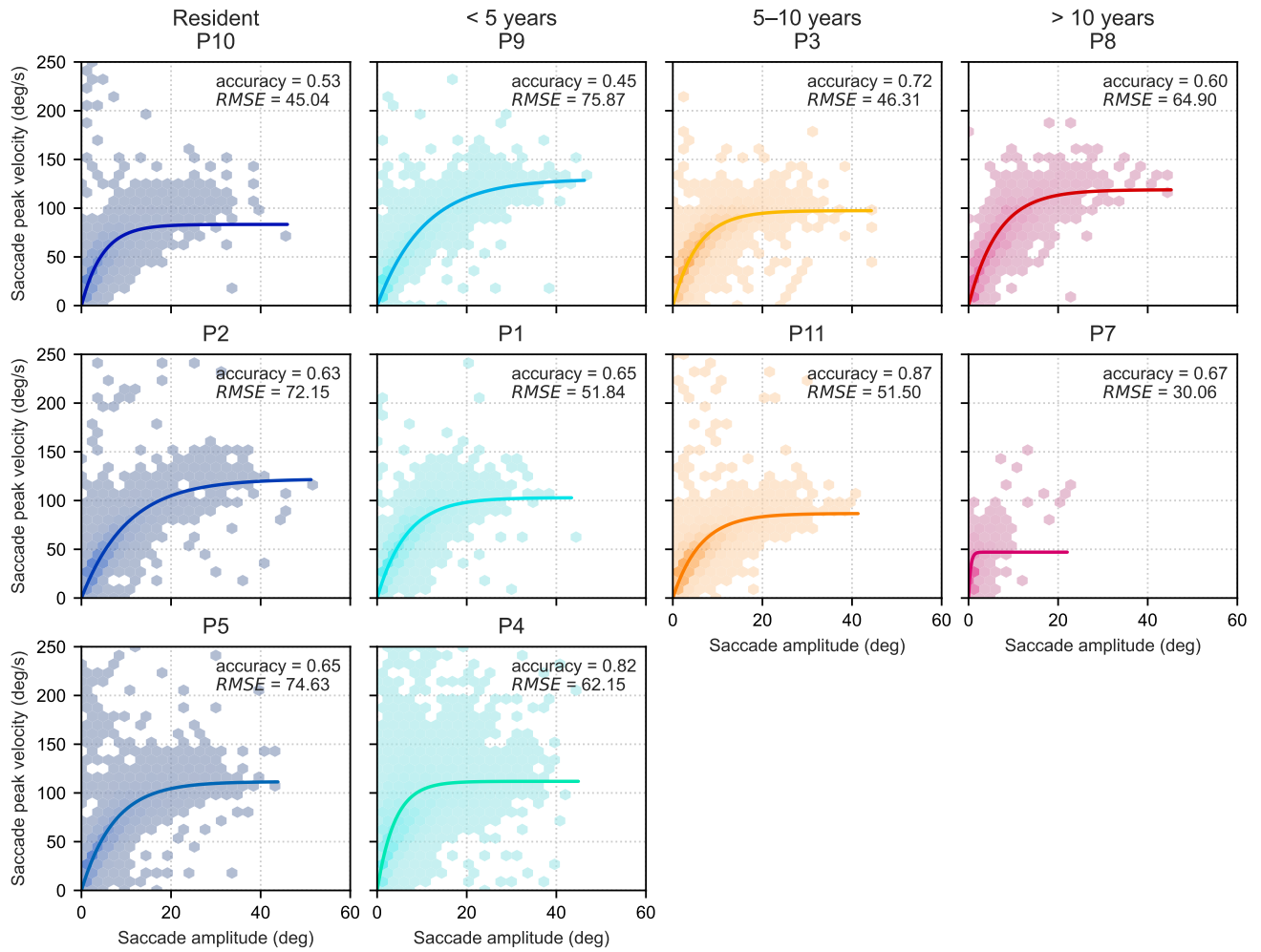


Figure 7. Technical Validation II: Saccade amplitude analysis results from Experiment I data collection. We grouped participants by experience: resident, < 5 years, (5–10) years, and > 10 years of experience. Within each group, participants have been ordered top to bottom by their overall decision accuracy. We fit an exponential curve to each participant’s saccade amplitudes and calculated the root mean squared error (RMSE). **Observations.** An increasing trend levels out around 10° amplitude. Our RMSE were of moderate size (30–80°/second, with the mean under 150°/second).

of reaching the correct diagnosis. However, the extent of panning did not necessarily correspond to the amount of zooming, and the two behaviors were not reliable opposites in pathologists’ search strategies.

In our eye-tracking experiment, the participant with the highest accuracy exhibited rapid, tile-by-tile panning, and noted that this helps avoid missing small tissue regions. Overall, we examine action and perception jointly by plotting zooming behavior over time, with fixation positions indicated along the timeline (Figure 8). We did not observe a significant correlation between panning or zooming behaviors; instead, tumor size emerges as the primary influencing factor (Figure 9).

Participants were substantially more accurate on macrometastasis slides (91.0%). We fit a logistic regression model to predict diagnostic accuracy, which showed an excellent fit (McFadden’s pseudo $R^2 = 0.25$, $p < 0.001$), according to McFadden’s R^2 range (0.2–0.4)²⁶. The model revealed a significant relationship between tumor–tissue ratio and diagnostic accuracy: among slides containing tumors (isolated tumor cells, micrometastases, and macrometastases), tumor size was a significant main effect influencing diagnostic accuracy ($F_{3,1,136} = 299.25$, $p < 0.01$).

Limitations of Our Data Collection

Our dataset has limitations. Although it is the largest collection of behavior data, the observed strategies and diagnostic accuracy inevitably reflect the individual choices and experience levels of the participating pathologists. For example, the sample size may still be too small to fully generalize visual search behaviors in complex 2.5D whole-slide images.

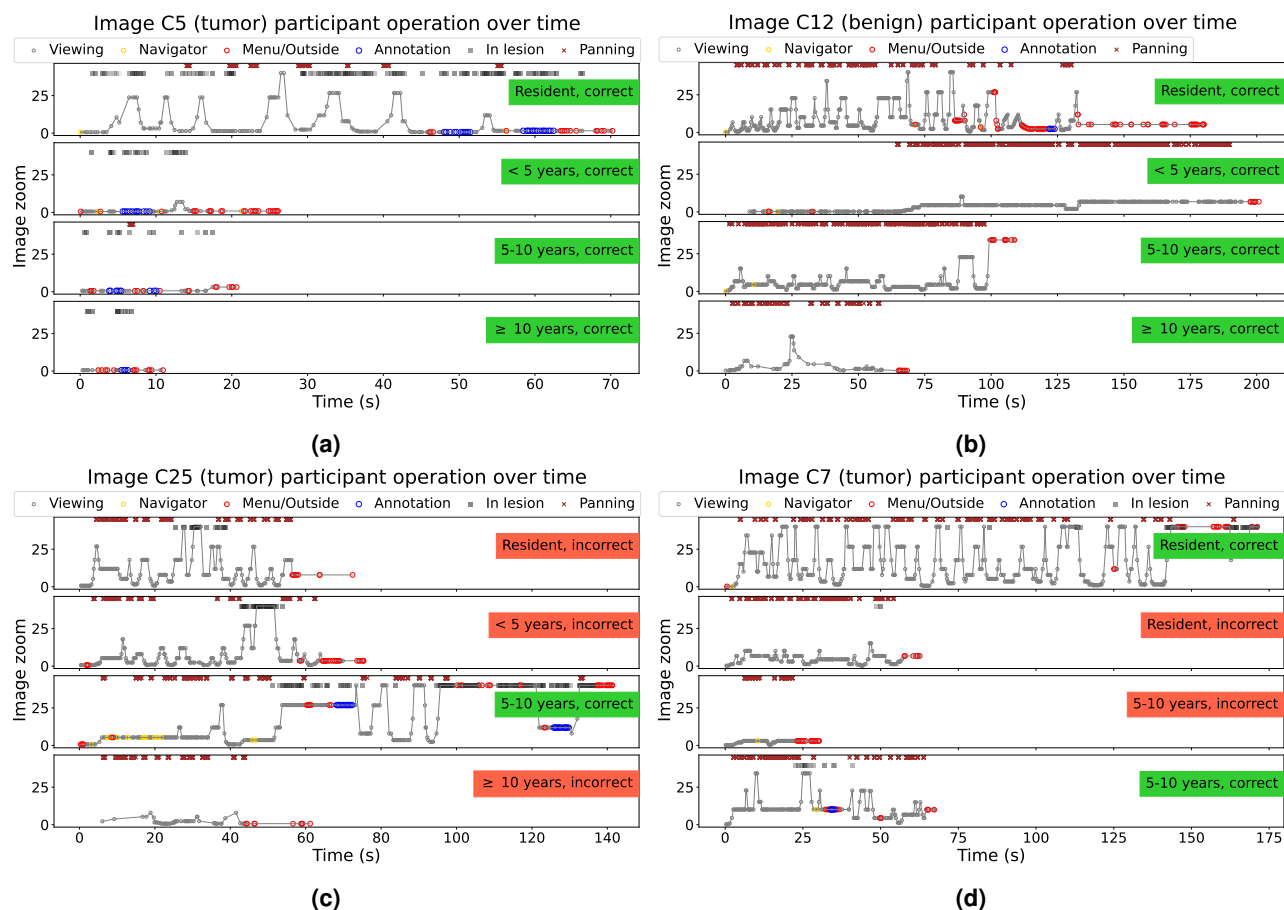


Figure 8. Participants navigation behavior. Each of the four sets of plots shows four participants' navigation behavior while viewing a slide. Zoom level is plotted on the y-axes, and panning is notated with a red cross on the top spine of each plot. **Observations.** Participants tended to have more “spiky” zooming behavior, characterized by frequent changes in zoom level and higher overall magnifications. There was no consistent pattern of increased zooming or panning relative to tumor size. This variability was observed across different case types in our data: (a) an easy tumor case, (b) an easy benign case, (c) a hard tumor case, and (d) a mid-range tumor case.

Usage Notes

PathoGaze1.0 is a rich data set that can support many other analyses. For example, there are more than 50% decision errors for micrometastases, and higher than 75% decision errors for the isolated tumor WSIs in this dataset. How can we classify those errors? Borrowing the taxonomy used in radiology, we can ask if (and for how long) the target was fixated. In this dataset, we have the additional factor of the zoom at which it was viewed. We can ask if there are patterns of panning and zooming that appear to be related to errors. Finally, we have also carefully annotated the first and last fixations and recorded the corresponding decision outcomes in both experiments. We can investigate the onset of the first fixation and its relationship to the development of visual expertise. In line with Herbert Simon's model of bounded rationality, which frames decision-making as a heuristic rather than fully optimized process, these measures can offer insights into how experts allocate attention under cognitive constraints and how the subsequent scan paths and fixation durations may be linked to the quality of the final fixation and the accuracy of the resulting diagnosis.

Data Availability

The datasets are temporarily hosted on Google Drive and will be deposited in the repository recommended by *Nature Scientific Data*, upon publication recommendations.

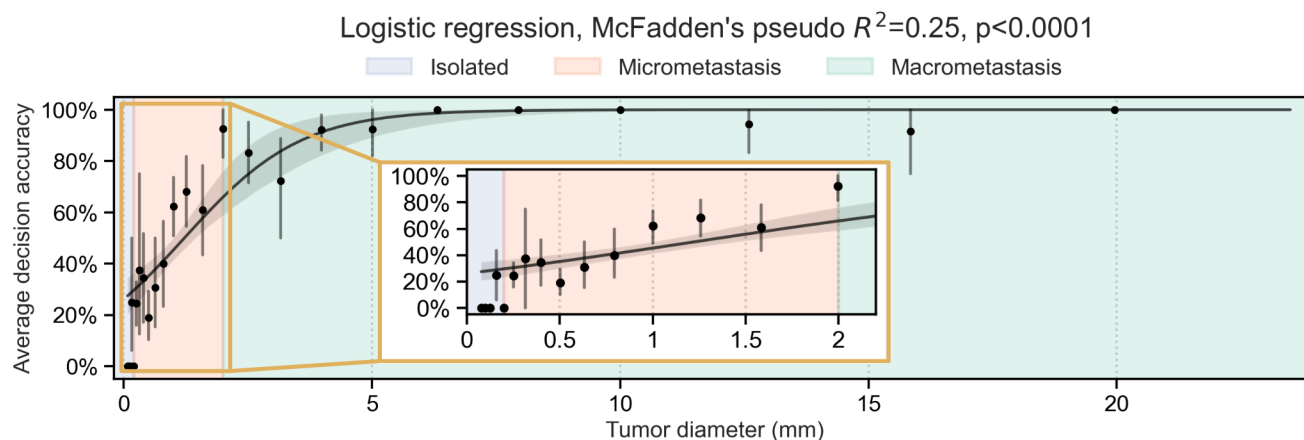


Figure 9. Technical Validation III. Tumor size is a significant main effect on diagnostic accuracy. Observations. We observed a correlation between diagnostic accuracy and tumor size. In general, participants were accurate in the macrometastasis cases and the errors increase when the tumor sizes get smaller. Thus, errors are largely search rather than diagnostic errors.

Code Availability

We released all data processing code online at <https://go.osu.edu/pathogaze>. The code include the eye-tracking and mouse tracking event alignment code, coordinate transforms, as well as statistical analyses and figure reproduction programs.

References

1. Nan, T. *et al.* Deep learning quantifies pathologists' visual patterns for whole slide image diagnosis. *Nat. Commun.* **16**, 5493, [10/p274](#) (2025).
2. Brunyé, T. T. *et al.* From image to diagnosis: characterizing sources of error in histopathologic interpretation. *Mod. Pathol.* **36**, 100162, [10/pz5b](#) (2023).
3. Wolfe, J. M., Lyu, W., Dong, J. & Wu, C.-C. What eye tracking can tell us about how radiologists use automated breast ultrasound. *J. Med. Imaging* **9**, 045502–045502, [10/p278](#) (2022).
4. Elston, C. *et al.* Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast. *Eur. J. Cancer* **36**, 1769–1772, [10/c8hj75](#) (2000).
5. Chakraborty, S. *et al.* Measuring and predicting where and when pathologists focus their visual attention while grading whole slide images of cancer. *Med. Image Anal.* 103752, [10/p8tc](#) (2025).
6. Anikina, A., Ibragimova, D., Mustafaev, T., Mello-Thoms, C. & Ibragimov, B. Longitudinal Anatomical Attention Maps for Recognizing Diagnostic Errors from Radiologists' Eye Movements. In *Med. Image Comput. Comput. Assisted Intervention*, 315–325, [10/p8s9](#) (Springer, 2025).
7. Ibragimov, B. & Mello-Thoms, C. The use of machine learning in eye tracking studies in medical imaging: a review. *J. Biomed. Heal. Inf.* **28**, 3597–3612, [10/p8tb](#) (2024).
8. Abadi, E. *et al.* Toward widespread use of virtual trials in medical imaging innovation and regulatory science. *Med. Phys.* **51**, 9394–9404, [10/g9wghk](#) (2024).
9. Bejnordi, B. E. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *J. Am. Med. Assoc.* **318**, 2199–2210, [10/gcn2kf](#) (2017).
10. Wolfe, J. M. Guided Search 6.0: An updated model of visual search. *Psychon. Bull. Rev.* **28**, 1060–1092, [10/gh2s45](#) (2021).
11. Wolfe, J. M., Kosovicheva, A. & Wolfe, B. Normal blindness: When we look but fail to see. *Trends Cogn. Sci.* **26**, 809–819, [10/gtwm3x](#) (2022).
12. Fleck, M. S., Samei, E. & Mitroff, S. R. Generalized “satisfaction of search”: adverse influences on dual-target search accuracy. *J. Exp. Psychol.: Appl.* **16**, 60, [10/fmrgws](#) (2010).

13. Raghunath, V. *et al.* Mouse cursor movement and eye tracking data as an indicator of pathologists' attention when viewing digital whole slide images. *J. Pathol. Inf.* **3**, 43, [10/pz5d](#) (2012).
14. Budginaite, E., Magee, D. R., Kloft, M., Woodruff, H. C. & Grabsch, H. I. Computational methods for metastasis detection in lymph nodes and characterization of the metastasis-free lymph node microarchitecture: A systematic-narrative hybrid review. *J. Pathol. Inf.* **15**, 100367, [10/pz4n](#) (2024).
15. OpenSeadragon contributors. OpenSeadragon. <https://openseadragon.github.io/> (2022). Version 3.1.0.
16. Annotorious contributors. Annotorious. <https://annotorious.github.io/> (2022).
17. Olsen, A. The Tobii I-VT fixation filter. <https://go.tobii.com/Tobii-I-VT-fixation-filter-white-paper> (2012).
18. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data.* **3**, 1–9, [10/bdd4](#) (2016).
19. Otsu, N. *et al.* A threshold selection method from gray-level histograms. *Automatica* **11**, 23–27, [10/bfq5pv](#) (1975).
20. Poole, A. & Ball, L. J. Eye tracking in HCI and usability research. In *Encyclopedia of Human Computer Interaction*, 211–219 (IGI Global Scientific Publishing, 2006).
21. Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S. & Züger, M. Using psycho-physiological measures to assess task difficulty in software development. In *Proc. Int. Conf. Software Eng.*, 402–413, [10/gpbdt8](#) (2014).
22. Brunyé, T. T., Drew, T., Weaver, D. L. & Elmore, J. G. A review of eye tracking for understanding and improving diagnostic interpretation. *Cogn. Res.: Princ. Implic.* **4**, 1–16, [10/gg5srt](#) (2019).
23. Wu, C.-C. & Wolfe, J. M. The functional visual field (s) in simple visual search. *Vis. Res.* **190**, 107965, [10/gnvwg2](#) (2022).
24. Gibaldi, A. & Sabatini, S. P. The saccade main sequence revised: A fast and repeatable tool for oculomotor analysis. *Behav. Res. Methods* **53**, 167–187, [10/gm5wxr](#) (2021).
25. Drew, T. *et al.* More scanning, but not zooming, is associated with diagnostic accuracy in evaluating digital breast pathology slides. *J. Vis.* **21**, 1–17, [10/pz48](#) (2021).
26. McFadden, D. Quantitative methods for analysing travel behaviour of individuals: some recent developments. *Behavioural Travel Modelling* 279–318, [10/pz49](#) (2021).

Acknowledgments

The authors grateful thank all the participants and professional pathologists and residents contributing to the empirical study data collection. This project was funded by The Ohio State University (OSU) Translational Data Analysis Institute (TDAI) and The OSU Wexner Cancer Center.

Author contributions statement

All authors had full access to all the data in the study and accept responsibility to submit for publication. V.T. collected the data. R.L. and V.T. developed the testbed, PTAH. Pathologists A.P., Z.L., and Y.H., together with J.C., J.W., and R.M., supervised the testbed design. Vision scientist J.W. supervised the attention-capturing and visual attention aspects of the research and suggested critical aspects of alignment. V.T., S.J., M.L., and J.C. analyzed the data. V.T. and J.C. wrote the manuscript. J.C. conceptualized and designed the study and supervised the first four student co-authors. All authors contributed to and approved the final manuscript.

Competing interests

The authors declare no competing interests.