Original Article

PathEdEx – Uncovering High-explanatory Visual Diagnostics Heuristics Using Digital Pathology and Multiscale Gaze Data

Dmitriy Shin[1,4], Mikhail Kovalenko[1,4], Ilker Ersoy[1,4], Yu Li[2], Donald Doll[3], Chi-Ren Shyu[4,2], Richard Hammer[1,4]

Departments of [1]Pathology and Anatomical Sciences, [2]Computer Science, [3]Medicine, and [4]MU Informatics Institute, University of Missouri, Columbia, Missouri, USA

Received: 30 March 2017 Accepted: 22 May 2017 Published: 25 July 2017

Abstract

Background: Visual heuristics of pathology diagnosis is a largely unexplored area where reported studies only provided a qualitative insight into the subject. Uncovering and quantifying pathology visual and nonvisual diagnostic patterns have great potential to improve clinical outcomes and avoid diagnostic pitfalls. **Methods:** Here, we present PathEdEx, an informatics computational framework that incorporates whole-slide digital pathology imaging with multiscale gaze-tracking technology to create web-based interactive pathology educational atlases and to datamine visual and nonvisual diagnostic heuristics. **Results:** We demonstrate the capabilities of PathEdEx for mining visual and nonvisual diagnostic heuristics using the first PathEdEx volume of a hematopathology atlas. We conducted a quantitative study on the time dynamics of zooming and panning operations utilized by experts and novices to come to the correct diagnosis. We then performed association rule mining to determine sets of diagnostic factors that consistently result in a correct diagnosis, and studied differences in diagnostic strategies across different levels of pathology expertise using Markov chain (MC) modeling and MC Monte Carlo simulations. To perform these studies, we translated raw gaze points to high-explanatory semantic labels that represent pathology diagnostic clues. Therefore, the outcome of these studies is readily transformed into narrative descriptors for direct use in pathology education and practice. **Conclusion:** PathEdEx framework can be used to capture best practices of pathology visual and nonvisual diagnostic heuristics that can be passed over to the next generation of pathologists and have potential to streamline implementation of precision diagnostics in precision medicine settings.

Keywords: Digital pathology, eye tracking, gaze tracking, pathology diagnosis, visual heuristics, visual knowledge, whole slide images

Introduction

Pathology diagnosis is a highly complex process where multiple clinical and diagnostic factors have to be taken into account in an iterative fashion to produce a plausible conclusion that most accurately explains these factors from biological standpoint.[1] Information from patient clinical history, morphological findings from microscopic evaluation of biopsies, aspirates, smears as well as data from flow cytometry, immunohistochemistry (IHC), and "omics" modalities, such as comparative hybridization arrays and next-generation sequencing, are used in the diagnostic process, which currently can be described more like a subjective exercise than a well-defined protocol. As such, it can frequently lead to diagnostic pitfalls, which may harmfully impact a patient case with a wrong diagnosis. The diagnostic pitfalls are most often encountered in the case of complex diseases such as cancer,[2] where multiple fine-grained clinical phenotypes may require a better understanding of genomic variations across patient populations and therefore require better protocols for pathology diagnosis. It is especially important in light of realizing precision medicine ideas.[3] A better understanding of pathology diagnosis, especially of heuristics of visual reasoning over microscopic slides, is crucial to develop means for new genomic-enabled precision diagnostics methods.

Over the last decade, digital pathology (DP) and whole slide imaging (WSI) have become a mature technology that allows reproduction of the histopathologic glass slide in its entirety.[4] For the first time in history of microscopy, diagnostic quality

Access this article online

Quick Response Code:

Website: www.jpathinformatics.org

DOI: 10.4103/jpi.jpi_29_17

Address for correspondence: Dr. Dmitriy Shin, Department of Pathology and Anatomical Sciences, University of Missouri, 1 Hospital Dr. M251 Pathology, Med Sci Bldg, Columbia, MO 65212, USA. E-mail: shindm@health.missouri.edu

How to cite this article: Shin D, Kovalenko M, Ersoy I, Li Y, Doll D, Shyu CR, *et al*. PathEdEx – Uncovering high-explanatory visual diagnostics heuristics using digital pathology and multiscale gaze data. J Pathol Inform 2017;8:29.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2017/8/1/29/211596

digital images can be stored electronically and analyzed using computer algorithms to assist primary diagnosis and streamline research in biomedical imaging informatics.[5] This has not been possible before the DP era when cropped image areas from pathology slides could be used only to seek second opinion or share the diagnostic details with clinicians. At the same time, gaze-capturing devices have undergone transformation from bulky systems to portable, mobile trackers that can be installed on laptops, and used to seamlessly collect user's gaze.

There have been a number of research studies where WSI was used to analyze visual patterns of regions of interests (ROIs) annotated without[6-9] and with gaze-tracking technology.[10] In the former case, ROIs were identified either manually or using a viewport analysis, and in the latter case, gaze fixation points were used for the same purpose. Gaze tracking along with mouse movement has also been used to study pathologists' attention while viewing WSI during pathology diagnosis[11] as well as to study visual and cognitive aspects of pathology expertise.[12] The underlying idea in majority of these studies was to analyze captured gaze data with respect to the ROIs marked in the pathology images, gaze time spent within the ROIs, and total number of fixations within the ROIs as measures of diagnostically relevant viewing behavior. However, it is challenging to use ROIs to articulate underlying biology to understand visual heuristics of specific diagnostic decisions. It is not trivial to encode morphological patterns of an ROI using narrative language. As such, ROI might not be an effective means to translate best practices findings from gaze-tracking studies into the pathology education and pave a road toward precision diagnostics.

In this paper, we present PathEdEx, a web-based, WSI- and gaze tracking-enabled informatics framework for training and mining of visual and nonvisual heuristics of pathology diagnosis. PathEdEx system enables the development of interactive online atlases that can be used not only for educational purposes but also have the potential of capturing best diagnostic strategies and representing them in high-explanatory format for pathology practice. We demonstrate the capabilities of PathEdEx for mining visual and nonvisual diagnostic heuristics by performing quantitative studies on the time dynamics of zooming and panning operations utilized by experts and novices to come to correct diagnosis, association rule (AR) mining[13] studies to determine sets of diagnostics factors that consistently result in a correct diagnosis and quantifying differences in diagnostic strategies across different levels of pathology expertise using Markov chain (MC) modeling[14,15] and MC Monte Carlo (MCMC) simulations.[16,17] These studies allowed us to understand effective and efficient practices of human diagnostic heuristics that can be passed over to the next generation of pathologists and streamline implementation of precision diagnostics in precision medicine settings. The studies are performed using the first PathEdEx interactive training atlas of hematopathological cancers.

Methods

Each PathEdEx training atlas is a web-based system and runs inside a browser window as a thin client in an in-house developed multitiered informatics ecosystem depicted in Figure 1. The systems tier represents a foundation of the PathEdEx, containing essential software modules for processing of imaging and nonimaging data.

Whole slide imaging management module

This module is responsible for processing of WSIs. The client side of this module is implemented in JavaScript using Bootstrap[18] and JQuery[19] libraries and the WSI online viewer is based on the OpenSeaDragon JavaScript library[20] that enables the interactive viewing of large WSIs over internet. At the server side, WSIs are served by IIPImage server which is an open-source tiled image server.[21] The WSI images of the patient cases are obtained by scanning the slides with the Aperio ScanScope CS (Leica Biosystems)[22] slide scanner at × 40; and they are converted to tiled pyramid TIFF images using the open-source VIPS library.[23] Each TIFF image contains 256 × 256 pixel tiles of the slide at seven different zoom levels. This enables the efficient communication of the whole slide data since only the visible part of the WSI is served to the client at any time. The WSI online viewer allows users zooming and panning of the image to provide a virtual microscope experience.

The multiscale gaze processing module

This module is responsible for capturing and processing of gaze data. The module consists of the gaze data server manager that works on the server side, the gaze data client manager that works on the client side inside of a browser, and the driver-connector that provides connectivity to a gaze capturing software supplied by a gaze tracker manufacturer. The data exchange between the WSI online viewer, gaze data server manager, and the gaze data client manager, and the driver-connector is implemented using web sockets technology. This enables a fast and real-time communication. When a user starts viewing

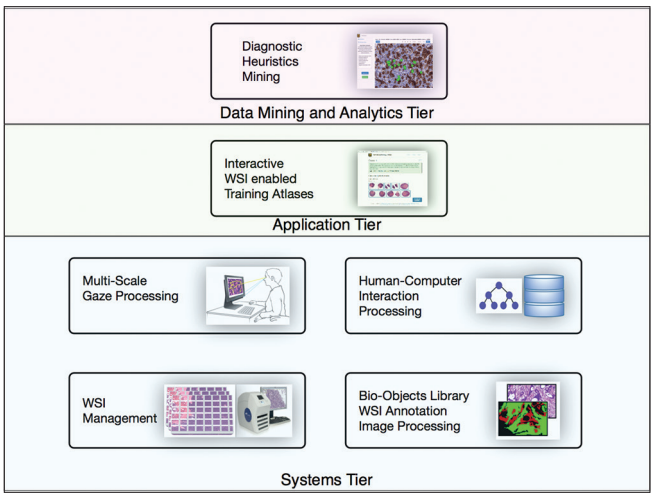

Figure 1: PathEdEx is a multitiered informatics ecosystem that enables education and research in digital pathology and pathology informatics

a WSI, the gaze data client manager is notified by WSI online viewer [blue-dotted section in Figure 2] about a WSI viewing session and sends a request to the driver-connector that in turn communicates with a vendor-specific gaze capturing software to start collecting gaze data [left side of Figure 2]. The collected data flow back through driver-connector to the gaze data client manager. The gaze data client manager uses viewport data from WSI online viewer such as viewport position, zooming, panning, and mouse position information to compute in real-time the coordinates of each gaze point relative to the origin of the WSI and sends them to the gaze data server manager [bottom of Figure 2]. The gaze data server manager saves the coordinates in a database for further processing and analysis.

Human-computer interaction processing module

This module of the PathEdEx systems tier provides functionality to managing user sessions for specific applications. On the basic level, it provides authentication and authorization functionality for PathEdEx users and groups. On the higher level, this module manages data related to various diagnostic algorithms as well as users' navigational data as they use such PathEdEx resources as WSI. PathEdEx provides a set of software tools with graphical user interface to help pathology experts define reference diagnostic algorithms for different types of cases. These algorithms as well as users PathEdEx navigational data are stored in form of graphs in a Neo4J database, which has superior capabilities for structured queries over relational databases.[24]

Bio-objects library, whole slide imaging annotation, and image processing module

This module is responsible for capturing and storing information about microanatomical structures on histopathological WSIs. For instance, different types of cells and tissues are treated as "biological objects" by the PathEdEx system and stored in form of WSI annotations. At present, majority of such bio-objects are defined semi-automatically by experts using in-house developed software tools. However, the work in underway to use advanced WSI image processing algorithms to aid WSI annotation. For instance, we have been developing novel methods for follicle and nucleus detection on WSI[25,26] and plan on incorporating these algorithms into the PathEdEx ecosystem.

Interactive whole slide imaging-enabled training atlases

The application tier of PathEdEx informatics ecosystem [middle panel in Figure 1] represents interactive, WSI-enabled pathology atlases that assist pathology trainee in learning various pathology subspecialty areas using simulated WSI-rich diagnostic environment. Here, we present the first such PathEdEx Atlas that consists of 55 patient cases of hematopathological cancers from University of Missouri Ellis Fischel Cancer Center. Each case includes a complete set of information such as clinical notes, radiological imaging studies, WSI of histopathology slides, IHC, and molecular tests and any additional laboratory data that are available for the case. For each patient case, PathEdEx includes a reference diagnostic workflow that is defined by an expert hematopathologist. The trainee reviews images and makes decisions on selection of additional tests. Any unjustified selection of a test in PathEdEx is penalized by a negative score, which reflects dollar amount of the corresponding test. For instance, if after reviewing morphological aspects of a hematoxylin and eosin (H and E) WSI slide, a trainee decides to view a specific IHC WSI slide but that IHC test is considered to be irrelevant to the case, he or she will receive a penalty. Every attempt to diagnose

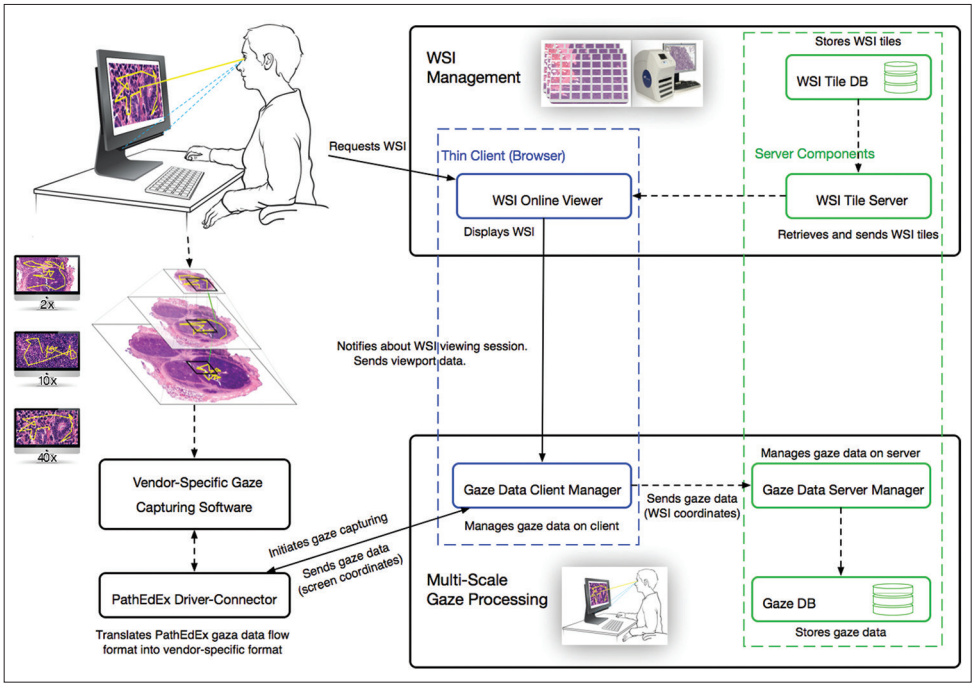

Figure 2: Processing of multiscale gaze data over whole slide imaging in PathEdEx

similar cases is logged and the overall progress of a trainee during a period is reported along with a detailed analysis of the diagnostics decisions. The trainee's diagnostic workflow is then stored for the analysis and progress assessment purposes. An example of a trainee's diagnostic workflow is presented in Figure 3.

Diagnostic heuristic mining module

This module of PathEdEx of data mining and analytics tier is responsible for the essential functionality that supports diagnostic heuristic mining studies. Specifically, it provides data structures and representational models for manipulation and quantification of visual and nonvisual diagnostic entities. To do that, PathEdEx introduces a number of novel informatics approaches. Unlike ROI-based gaze tracking studies that are based on low-level, pixel-based features, mining of visual diagnostic heuristics in PathEdEx is based on high-level, semantic labels that represent visual diagnostic clues (VDC) related to biological entities such as, for instance, type of cells (e.g., centroblasts, Reed–Sternberg cells), and diagnostics factors (e.g., mitotic rate). Therefore, the results of heuristics mining in PathEdEx are much easier for understanding by a human expert and for drawing quantitative conclusions about most impactful factors that lead to correct diagnosis. The annotation of gaze data with semantic labels is done on a semiautomatic fashion and is followed by a fully computerized processing of visual and nonvisual heuristics analytics. Because VDCs play a central role in mining of pathology diagnostic heuristics, we describe it here in greater details.

The human eye goes through a series of fixation points and saccades during processing of a visual scene. Since the human ability to discriminate fine detail drops off outside of the fovea, eye movements help process the visual field through a series of fixation points separated by saccades (the actual eye movement) from one fixation point to other. The typical mean duration of fixation is about 180–275 ms.[27] The vision is suppressed during the saccade, and visual information is acquired during the fixation. Hence, a procedure to compute fixation points is required to analyze the collected eye gaze

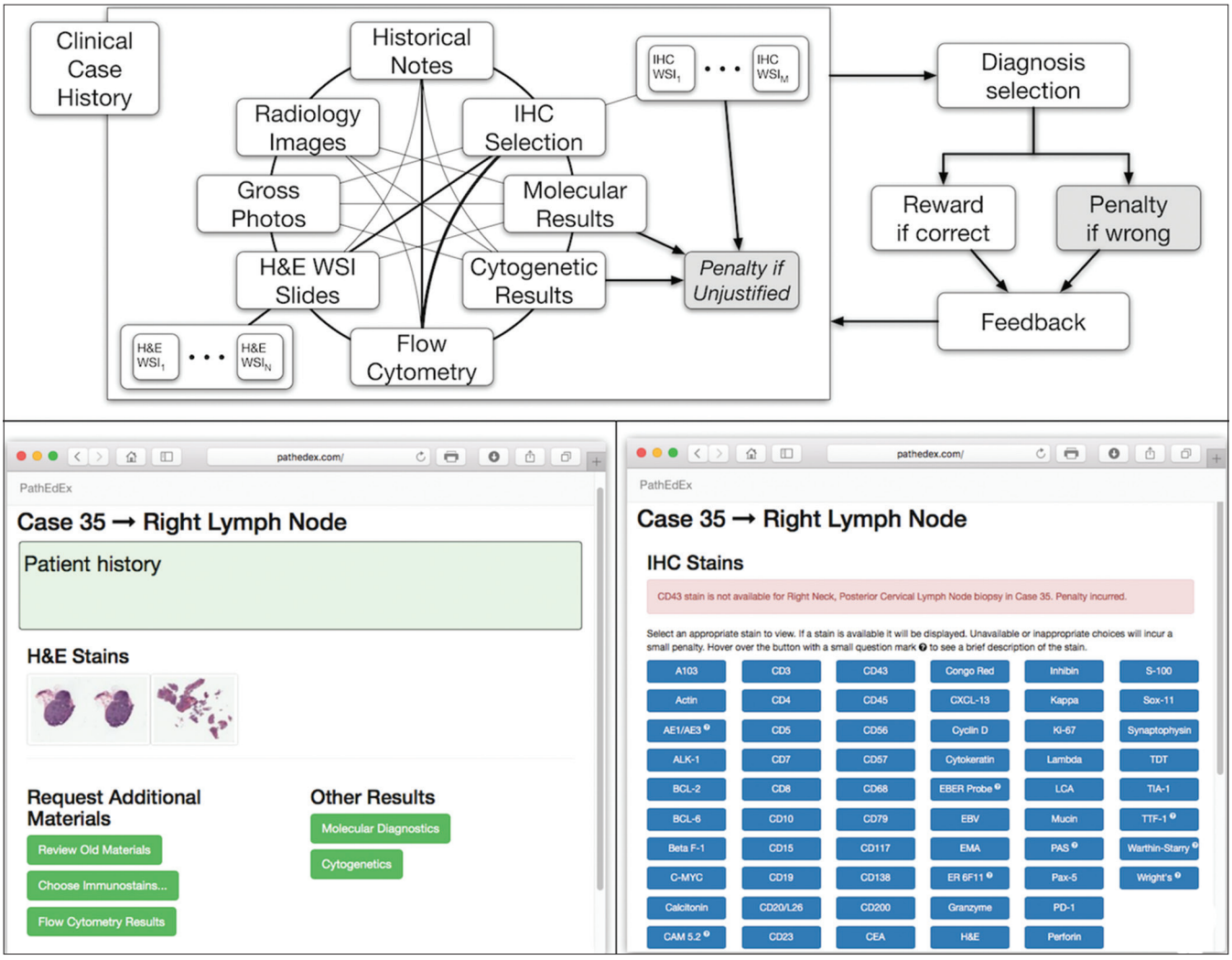

Figure 3: Example of diagnostic worklfow. A trainee selects a case and reads the patient history, then goes through radiology images, H and E whole slide imaging images, selects immunohistochemistry tests, and views immunohistochemistry whole slide imaging, and other available material (flow cytometry, etc.), and chooses a diagnosis

data. In case of histopathologic images, these fixation points can be grouped into visual diagnostic focus areas (VDFAs) and annotated. That way instead of an analysis of "low-level individual" fixation points, a more biologically relevant analysis of possibly diagnostically relevant VDFAs can be performed. Unlike commercial off-the-shelf fixation points algorithms, the PathEdEx's VDFA algorithms are specifically designed to take into account morphological structures of histopathologic sections. We have been developing a number of novel algorithms that are designed to identify and annotate VDFAs on histopathologic images such as tissue-, cell-, and sub-cellular level microanatomical structures of different morphology. However, in the context of this paper, we describe only one such method that is based on mean shift clustering.

Mean shift algorithm is a nonparametric clustering technique that does not require prior knowledge of the number of clusters and does not constrain the shape of the clusters. Mean shift considers the set of points as discrete samples from an underlying probability density function. For each data point, mean shift defines a window around it and computes the mean of the data point. Then, it shifts the center of the window to the mean and repeats the algorithm until convergence. The set of all locations that converge to the same mode defines the basin of attraction of that mode. The points that are in the same basin of attraction are associated with the same cluster. Hence, the mean shift clustering algorithm is a practical application of the mode finding procedure. Since mean shift does not make assumptions about the number of clusters, it can find naturally occurring clusters in the data. Mean shift clustering requires a bandwidth parameter that defines the scale of clusters. Smaller bandwidth leads to larger number of smaller scale clusters whereas larger bandwidth leads to smaller number of larger scale clusters. We represent the collected eye gaze data associated with a WSI slide as a set of four-dimensional points (x, y, z, t) where dimensions correspond to coordinates of the eye gaze (x, y) and zoom level of the slide (z) at time (t). Sets of points are separated by zoom levels, and clusters are found at each zoom level. The cluster centers (modes) represent the fixation points of user's eye. Discovered in this fashion, clusters will correspond to VDFAs, which are then annotated in semiautomatic setting. An example of the result of identification of VDFAs is shown in Figure 4. The VDFAs are then annotated during a diagnostic heuristic mining study to represent VDCs as will be demonstrated in the results and discussion section.

Results and Discussion

Mining visual and nonvisual diagnostic heuristics

We have conducted three data mining studies of visual and nonvisual diagnostic heuristics. The main purpose of these studies was to test the capabilities of PathEdEx informatics framework. We were not focused on comprehensiveness of the studies setup and the scientific validity of the results and conclusions. Instead, we aimed to demonstrate the utility of the PathEdEx as a framework to uncover and quantify pathology diagnostic heuristics with high-explanatory results over a complete set of diagnostic materials from real patient cases, which is to our best knowledge has not been previously reported.

Study 1: Quantification of diagnostic workflow

A total of 9 users navigated through four selected cases in the developed PathEdEx online atlas of hematopathology cases. The users were given numerical ranks to indicate their level of experience starting with rank one for expert pathologist up to rank six for postsophomore fellows (PSFs). They were also sorted into three groups of three users for a less granular ranking. Table 1 shows the distribution of experience levels and their ranks.

Their navigation and gaze data were collected as described in the methods section. The selected four patient cases had a total of 29 WSIs of H and E and IHC slides. Since each user has a freedom to order different number of IHC tests as they deem necessary, not every user went through all 29 images. Total number of image views was 149, averaging about 16 WSI views per user. The collected gaze data were analyzed using PathEdEx's diagnostic heuristic mining module. Gaze data for each image were first separated by their corresponding zoom levels, and then, the mean shift algorithm was applied to gaze data at each zoom level to

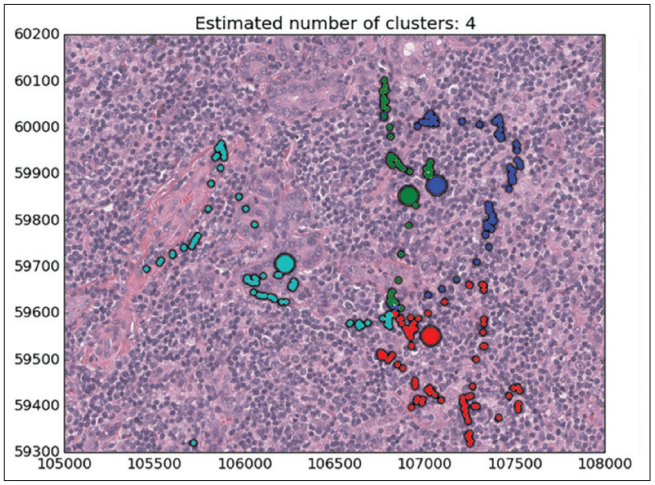

Figure 4: Computational method to identify visual diagnostic focus areas using sliding window (SW) over zoom/time dimensions (250 ms cutoff) with consequent application of mean shift clustering over positional dimensions X and Y with fixed bandwidth parameter per each zoom level

Table 1: Experience levels and given ranks for users in the quantification of diagnostic workflow study

Experience level	Number of users	Ranks	Groups
Expert pathologist	2	1, 2	A
4th year resident	1	3	A
2nd year resident	1	4	B
1st year resident	2	5	B
Post-sophomore fellow	3	6	C

compute fixation points. A set of 17 features that had the potential to capture viewing behavior was computed for each user and each WSI using the collected and computed data. The computed features were as follows:

- 1 - Number of zoom levels with user's fixations in the image
- 2 - Number of fixation points in the image
- 3 - View time of the image
- 4 - Total duration of fixations in the image
- 5 - Mean zoom level in the image weighted by number of fixations
- 6 - Total distance of eye gaze scan in the whole image
- 7 - Total distance of eye gaze scan within fixation clusters
- 8 - Focus ratio: Total distance (6) divided by end to end gaze distance in the image
- 9 - Dispersion: Within-fixation distance (7) divided by total distance (6)
- 10 - End-to-end gaze distance in the image
- 11 - Dispersion: Average of dispersions computed at each separate zoom level
- 12 - Mean zoom level (over viewing time) of the image
- 13 - Standard deviation of zoom level (over viewing time) of the image
- 14–17 - Means and standard deviations of zoom level differences in time with and without zeroes.

Figure 5 shows an example of eye gaze data superposed on a WSI. Since users can pan and zoom, the trajectory of the eye gaze represents all zoom levels converted to the global image coordinates.

Table 2 shows the total time spent on different aspects of the cases while Figure 6 presents a comparison between an expert and a PSF in terms of the distribution of the fixation points among zoom levels. PSF navigates through a variety of zoom levels whereas the expert spends time concentrated on a small number of zoom levels. Figure 7 shows the zoom levels as a function of time for the same users.

We computed Pearson's correlation coefficient and Spearman's rank correlation coefficient between these features and the given ranks overall image views to explore the relations between features capturing visual behavior and the level of experience.

Figure 8 presents the Pearson's coefficients computed overall images for each feature versus three groups of users. The best correlations are obtained for 9th and 11th features (dispersion features) at about $r = 0.2$. Spearman's coefficients also show similar results. Computing with more granular ranks (1 through 6) did not change these results significantly. The low correlation values suggest that these features do not successfully capture the general viewing behavior overall WSIs. Since these images belong to separate patient cases with different levels of diagnostic challenges, we computed the same set of correlations per case basis. Figure 9 shows the correlations and *P* values for each of the four patient cases.

Table 2: Total navigation time (min) for all cases

Time spent (min)	Experts		Senior research		Junior research		PSFs
Reading patient history	0:58	3:5	1:39	2:30	2:23	2:32	1:59
Viewing H & E slides	5:34	18:32	11:30	22:10	9:46	15:12	13:16
Viewing IHC slides	3:53	9:12	6:00	16:55	5:41	2:4	6:31
Looking at other materials	0:39	12:18	3:23	13:32	2:58	7:7	6:34
Choosing diagnosis	0:29	3:32	2:4	5:10	3:23	2:39	2:35
Total case time	11:33	46:39	24:36	60:17	24:11	29:34	30:55

Data are summarized from detailed logs of user navigation through the case pages and slides. PSFs: Postsophomore fellows, IHC: Immunohistochemistry, H&E: Hematoxylin and eosin

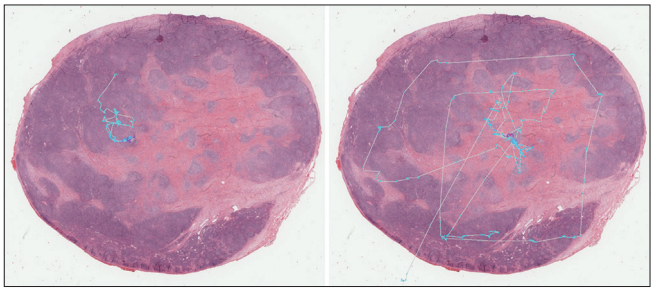

Figure 5: Examples of captured gaze data of an expert (left) versus post-sophomore fellow (right). Colors indicate different zoom levels of the image where the eye gaze is captured. Expert localizes and finds diagnostic clues quickly; postsophomore fellow roams the image at different zoom levels

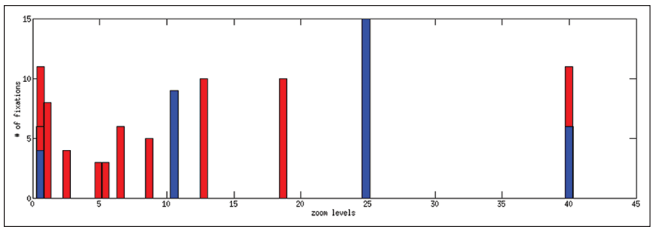

Figure 6: Number of fixation points per zoom level for an expert (blue) versus postsophomore fellow (red) viewing a whole slide image. Expert has a total of 34 eye fixations; postsophomore fellow has a total of 77 fixations

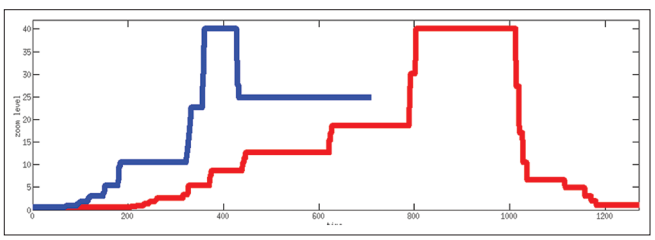

Figure 7: Zoom levels versus time for an expert (blue) versus postsophomore fellow (red) viewing a whole slide image. Expert views it for 19 s; postsophomore fellow views it for 32 s. Expert quickly zooms in on diagnostic clues whereas postsophomore fellows roams the image at several zoom levels

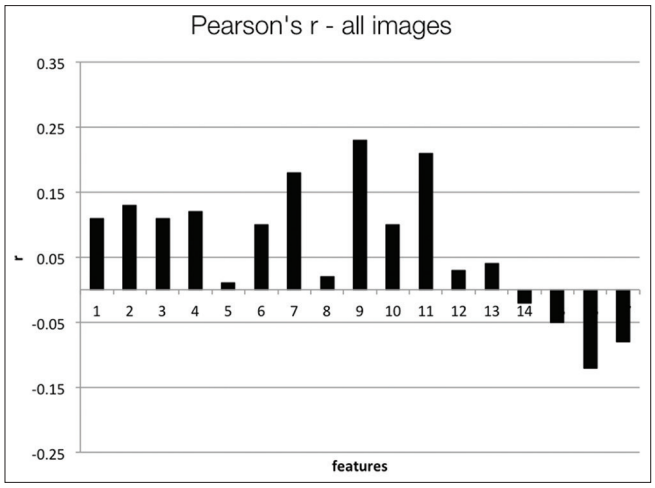

Figure 8: Pearson's correlation coefficients computed for each feature versus the group ranks over all image views

The users in three groups show higher correlations with some of the features depending on the case. For Case 1, 2nd feature ($r = 0.63$, $P = 0.07$) and 9th feature ($r = 0.69$, $P = 0.04$) have the best correlations. For Case 2, 7th feature ($r = 0.63$, $P = 0.07$) and 9th feature ($r = 0.58$, $P = 0.1$) have best correlations. For Case 3, 6th feature ($r = 0.58$, $P = 0.1$), and for the Case 4, 9th feature ($r = 0.63$, $P = 0.068$) have the best correlations. The correlations improve slightly for the more granular ranks of 1–6 and 11th feature becomes one of the highest correlated features in this category. In general, the dispersion features and number of fixations and their scan distance show higher correlation with levels of experience. Even though the *P* values are not low enough to reject the null hypothesis, these results suggest that these features have the potential to capture viewing behavior on a case-by-case basis.

As we discussed elsewhere in this paper, gaze tracking data analysis is becoming more available in radiology and pathology. Many studies analyze the captured gaze data with respect to the ROI marked in the medical images; gaze time spent within the ROI, and total number of fixations within the ROI are computed as measures of diagnostically relevant viewing behavior. In this study, we analyzed a variety of features computed from the gaze data on their own merit without utilizing ROIs to explore possibility of capturing viewing behavior directly from the gaze data. Utilizing ROIs may not be feasible for every case in pathology, especially for cases where diagnosis can be reached by analyzing any number of cells in the WSI without resorting to a particular ROI. Our study has shown that some features have the potential to capture viewing behavior on a case-by-case basis. This suggests that cases with different levels of diagnostic challenges influence the viewing behavior of the users that are captured by different features. Grouping patient cases by the diagnosis and other relevant information can elucidate the diagnostic heuristics of users.

This study has several limitations. The number of users per experience level is low even when grouped into three levels. Any user with sufficiently different viewing behavior in comparison to her group can skew the statistics. Another limitation is the accuracy of initial rankings. We rank all users in the same year of residency as equal which may not be realistic. And finally, the computed features may not be ideally capturing the viewing behavior. Even though there are a variety of features that can be borrowed from other domains to represent spatiotemporal data (x, y, t) we do not want to lose the high explanatory power of simpler features as they can be used as feedback to the participants. We are planning to conduct a more comprehensive study with more participants and patient cases grouped adequately for their diagnostic challenges to pursue high explanatory features that can capture viewing behavior and be used as a supplement to the rest of the numerical scoring in PathEdEx to keep track of user's progress. We are also currently working on automated and semiautomated detection and classification of cells and other relevant morphologies, as a part of the PathEdEx's bio-object library WSI annotation and image analytics module, to enable statistics computed over the types of cells and other relevant objects (follicles, etc.) the users viewed.[26]

Study 2: Uncovering association of visual diagnostic clues with diagnostic decisions

In this proof-of-concept study, we set the goal to uncover associations between sets of VDCs utilized by trainees most often and corresponding diagnoses that the trainees selected during PathEdEx training sessions. To do that, we annotated each VDFA with a set of VDCs. The set of VDCs used in hematopathology cases of the first volume of PathEdEx is shown in Table 3. The annotation was performed in a semiautomatic fashion. After processing raw gaze data and generating VDFAs, VDFAs were displayed one at a time to an expert hematopathologist, who assigned a set of VDCs that the expert believed might have been the goal of a trainee's visual investigation of a particular VDFA [Figure 10]. Each VDC had an associated confidence level assigned by the expert. VDCs with confidence level above the selected threshold (0.75 chosen for initial experiments) were selected to AR study. Treating each trainee's diagnostic session as an AR "market basket transaction" and corresponding VDCs and trainee's diagnostic choice as "basket items," we computed frequent item sets and induced ARs of the form:

$$VDC_1, VDC_2, \ldots, VDC_N \geq \text{Diagnosis}$$

Figure 11 presents an example of ARs inferred from PathEdEx hematopathology atlas's diagnostic training sessions. The primary goal of this preliminary AR study was only to test the capabilities of PathEdEx of manipulating VDCs. At present, more comprehensive AR studies are underway. In one such study, we compare the choice of VDCs used by experts and novices as well as identify VDCs that frequently lead to right and wrong diagnoses. In another AR study, we set a goal to quantify the contribution of various VDCs to diagnostic accuracy if they are added to a set of "routinely" used VDCs in specific diseases. Furthermore, in the future AR studies, we plan to extend a set of VDCs by other diagnostic clues,

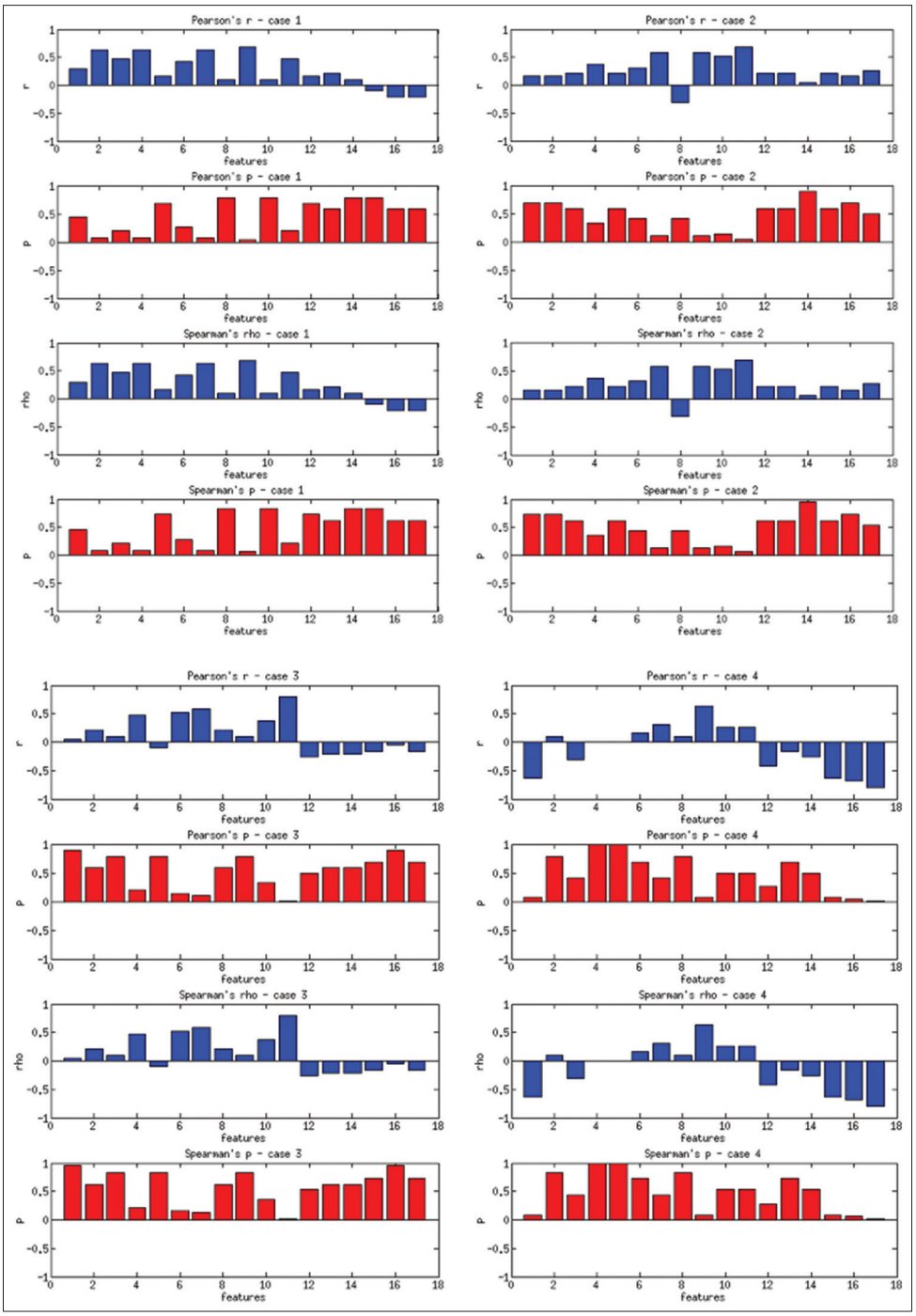

Figure 9: Correlations computed on a case-by-case basis for users in three groups

such as IHC and molecular tests and other units of diagnostic information.

Study 3: Uncovering and quantification of diagnostic strategies

In addition to performance quantification for pathologists, the PathEdEx platform provides capabilities for capturing their reasoning processes. Although pathologists receive similar training during medical school, experience with real cases allows them to develop heuristics that enables experts to make faster and more reliable judgments. With navigation data being recorded, the PathEdEx platform enables data analytics to analyze the strategies that pathologists adopt to approach and diagnose each case.

In this study, we adopt Markov models to quantify the transition probabilities among visual and nonVDC. In other words, we aimed to quantify probabilities of choosing specific tests for specific diagnostic clues, for example, IHC tests, at each point of iterative diagnostic process. For instance, after observing morphology and obtaining results for a set of specific diagnostic tests, which can serve as diagnostic clues,

Table 3: Visual diagnostic clues

Clear cells
Endothelial cells
Eosinophil
Epithelial cells
Fibrosis
Germinal center
Giant cell
Histiocyte
Immunoblast
Irregular lymphocyte
Large centroblast
Large centrocyte
Mitotic figure
Monocyte
Monocytoid B-cell
Negative immunostain
Neutrophil
Paraimmunoblast
Plasma cell
Platelet
Positive immunostain
Prominent vessels
Red cells
RS cell
RS-like cell
Small centroblast
Small centrocyte
Small round lymphocyte
Smudge cell
Tingible body macrophage
Unknown cell

RS: Reed-Sternberg

we computed probabilities of selecting other diagnostic tests. These probabilities can serve as a representation of a diagnostic strategy by an individual pathologist (even though it can still vary for the same pathologist).

Markov models abstract the behavior of a system as a series of transitions among possible states. The probability of next state depends only on the current state and a finite number of transition histories. In this study, we define that the current state contains all the information needed for the next transition. During trainee's simulated diagnostic process, PathEdEx platform records navigation of web pages and selection of diagnostic clues (e.g., radiology images, IHC stains, cytogenetics), which are modeled as states. The recorded navigational data from one page to another is mapped into a sequence of state transition. Then, we compute the transition matrix by calculating the transition probabilities from each state to all other states. Depending on the case, generated models normally have a state set size around 80.

Figure 12 shows two examples of Markov models generated using the navigational data from two pathologists diagnosing a classic Hodgkin's lymphoma case. Due to the large size of all possible states, transitions with 0 probability (states which

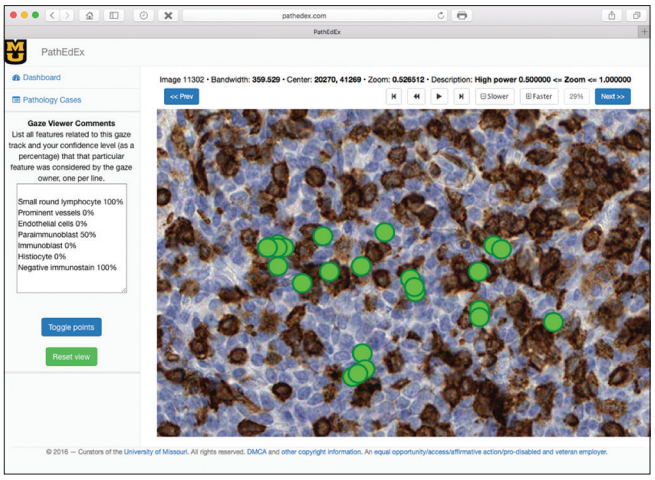

Figure 10: Semi-automatic annotation of visual diagnostic focus areas with a set of visual diagnostic clues. One visual diagnostic focus area (cluster of gaze points) is shown on the right and a corresponding set of visual diagnostic clues assigned by an expert hematopathologist with confidence levels is shown on the left

are never visited) are omitted in the graph. We can see that the expert not only spends less time and visits fewer states but also exhibits more deterministic behavior possibly manifesting a presence of a clear strategy.

As a preliminary study, we used two-way ANOVA with replication to establish that the Markov models do capture the reasoning of pathologists, i.e., there are significant statistical differences between the Markov models generated by a group of pathologists and a group of Markov models with random transitions (the null distribution). The group of random Markov models is generated using MCMC method. Each group contains five individual models of the same pathology case. The ANOVA result with λ significance level of 0.05 is shown in Table 4. It shows that there are statistically significant interactions of groups and transitions. However, the difference between the two groups is not as significant. We therefore hypothesized that the difference among pathologists' Markov models (diagnostic strategies) is statistically significant to an extent that it diminishes the difference of group of pathologist to group of random Markov models. To investigate this hypothesis, we grouped our models according to pathologists. Each pathologist group contains the same five cases. The two-way ANOVA result for five groups (pathologists) is shown in Table 5. Here, we see that there are statistically significant differences among pathologists in addition to the significant interactions. Considering the five pathologists vary in the level of expertise, the results suggest that pathologists might develop different strategies as they get more specialized and experienced. They also suggest that the strategy differences directly influence how they consider each material in a pathology case. Similar studies were performed by Treanor *et al*.[28] However, more in-depth studies are needed to be performed to provide statistical significance for this conjecture.

Visual Diagnostic Clues		Diagnosis	Support (%)	Confidence (%)	Lift (%)
{Endothelial cells, Immunoblast, Irregular lymphocyte, Monocytoid B cell, Prominent vessels, Small round lymphocyte}	=>	{Diffuse large B-cell lymphoma}	1.08%	100.00%	18.53
{Endothelial cells, Paraimmunoblast, Prominent vessels, Small round lymphocyte}	=>	{Small lymphocytic lymphoma / Chronic lymphocytic leukemia}	1.80%	100.00%	5.56
{Histiocyte, Irregular lymphocyte, Small round lymphocyte}	=>	{Follicular lymphoma}	2.52%	87.50%	3.33
{Endothelial cells, Irregular lymphocyte, Prominent vessels, Small round lymphocyte}	=>	{Diffuse large B-cell lymphoma}	1.80%	71.43%	13.24
{Irregular lymphocyte, Monocytoid B cell, Small round lymphocyte}	=>	{Diffuse large B-cell lymphoma}	2.16%	54.55%	10.11

Figure 11: Example of association rules induced using PathEdEx with support, confidence, and lift parameters

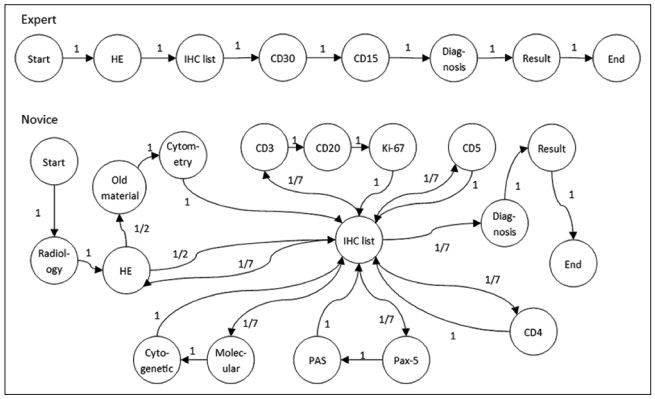

Figure 12: Markov models generated with navigational data by expert (top) and novice (bottom) pathologists diagnosing the same case. Transitions with 0 probability are omitted

At each iterative stage of a diagnostic process, a pathologist attempts to find an evidence to narrow down the possible outcomes. These underlying tasks can also be included in the model as a latent variables or hidden states using hidden Markov model (HMM).[28,29] With HMM, each trace that the PathEdEx platform records is a sequence of observations, and the underlying tasks are hidden states. We can infer the question that the pathologist is trying to answer by observing what material is being reviewed. In addition to a sequence of diagnostic clues modeled by Markov model, HMM can also generate the unobserved reasoning sequence of pathologists so we can quantifiably model their reasoning processes for each case.

This study suggests that experienced pathologists do better than trainees. Future in-depth studies can possibly statistically prove this logical assumption. However, due to different training and practices, pathologists can develop their own diagnostic strategies, which can lead to equal diagnostic conclusions. Another issue is related to diagnostic and biological isomorphism. For instance, different pathologists can utilize different antibodies to determine cell lineage in hematopoietic cancer cases. Furthermore, our previous work on selection of antibody tests suggests that there is even an intrapathologist variability (i.e., the same pathologist can use different tests

Table 4: Result of two-way ANOVA with replication comparing random model to model by pathologist

Source of variation	SS	df	MS	*F*	*P*	F criteria
Groups	0.003716	1	0.003716	2.608858	0.106274	3.841628
Transitions	17.3481	6888	0.002519	1.768176	8.4E-253	1.029971
Interaction	16.34651	6888	0.002373	1.66609	2.9E-200	1.029971
Within	78.50178	55,112	0.001424			
Total	112.2001	68,889				

SS: Sums of square, MS: Mean of square

Table 5: Result of two-way ANOVA with replication for 5 pathologists with 5 cases each

Source of variation	SS	df	MS	*F*	*P*	F criteria
Pathologists	0.012196	3	0.004065	4.171269	0.005817	2.605
Transitions	81.30678	6083	0.013366	13.71462	0	1.030973
Interaction	27.22417	18,249	0.001492	1.530701	0	1.018871
Within	94.87121	97,344	0.000975			
Total	203.4144	121,679				

SS: Sums of square, MS: Mean of square

for the same goal on different days).[1] We are in process of enhancing PathEdEx to identify equivalence classes for visual and nonVDC that have the same diagnostic and biological purpose. To do that, we are building an ontology-driven informatics framework to support diagnostic and biological isomorphism.

Conclusion and Future Work

Here, we have presented PathEdEx, a WSI and gaze-tracking-enabled framework. The PathEdEx informatics framework can be used not only to build comprehensive interactive online pathology training atlases that utilize cutting edge WSI imaging technology but also leverage multiscale gaze-tracking technology to capture trainees' diagnostic workflow patterns of using nonvisual and visual diagnostic materials. We have also demonstrated the potential of using

PathEdEx and such computational models as ARs data mining techniques and MC models and MCMC simulations to uncover and quantify trainees' diagnostic heuristics, identify diagnostic patterns (e.g., AR rules and transition probabilities that often lead to correct/wrong diagnoses) and represent them as high-explanatory narrative descriptors that can be readily used in pathology education.

The purpose of the performed heuristic mining studies was only to demonstrate the capabilities of the PathEdEx framework. We plan to conduct a set of comprehensive studies to uncover specific heuristic patterns that frequently lead to accurate diagnoses as well as those that lead to diagnostic pitfalls. For these studies, we are planning a greater number of pathology trainees of various levels of expertise. We are also planning to improve identification of VDFAs by incorporating the morphology of histologic section to build a solid foundation for running probabilistic computational models. In these future studies, we will aim to define and quantify the deviation or distance of trainees' diagnostic heuristics from reference diagnostic workflows utilized by pathology experts. We will also aim to study how much biology (vs. simple morphological patterns) is leveraged and articulated by pathologists for diagnostic and prognostic decisions. Many more PathEdEx interactive atlases are currently in development, including blood smears, head and neck pathology, and neuropathology. The future volumes of PathEdEx atlases will be expanded to include: (i) A "display of diagnostic clues mode" that circles and highlights diagnostic cells and patterns and provides pertinent ancillary studies for "search and identify" function, (ii) review video of expert gaze data for educational purposes, and (iii) expanded answers to include key points, which will satisfy the knowledge-based problems for novices and improve application of knowledge to visual clues. Furthermore, since assumption that gaze fixation always means diagnostically interesting spot can prove to be wrong, we are in process of enhancing identification of VDFAs using image information at locations corresponding to the gaze fixations. Such "content-aware" identification of VDFAs, in our view, can result in reduction of false positive VDFAs through an additional "feedback" loop to justify specific gaze fixations.

We are currently in process of establishing of a PathEdEx academic consortium. Members of the consortium will be able to utilize the PathEdEx platform for educational and research activities, get technical support as well as to share case materials for larger multi-institutional and cross-disciplinary studies.

All in all, the PathEdEx informatics tools have great potential to uncover, quantify, and study pathology diagnostic heuristics and can pave a path for precision diagnostics in precision medicine era.

Acknowledgments

The authors acknowledge funding from NIH NLM grant 5T32LM012410-02 that supported Mikhail Kovalenko through his NIH T32 fellowship.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Shin D, Arthur G, Caldwell C, Popescu M, Petruc M, Diaz-Arias A, *et al.* A pathologist-in-the-loop IHC antibody test selection using the entropy-based probabilistic method. J Pathol Inform 2012;3:1.
2. Higgins RA, Blankenship JE, Kinney MC. Application of immunohistochemistry in the diagnosis of non-Hodgkin and Hodgkin lymphoma. Arch Pathol Lab Med 2008;132:441-61.
3. Tenenbaum JD, Avillach P, Benham-Hutchins M, Breitenstein MK, Crowgey EL, Hoffman MA, *et al.* An informatics research agenda to support precision medicine: Seven key areas. J Am Med Inform Assoc 2016;23:791-5.
4. Pantanowitz L. Digital images and the future of digital pathology. J Pathol Inform 2010;1. pii: 15.
5. Chen Z, Shin D, Chen S, Mikhail K, Hadass O, Tomlison BN, *et al.* Histological quantitation of brain injury using whole slide imaging: A pilot validation study in mice. PLoS One 2014;9:e92133.
6. Reder NP, Glasser D, Dintzis SM, Rendi MH, Garcia RL, Henriksen JC, *et al.* NDER: A novel web application using annotated whole slide images for rapid improvements in human pattern recognition. J Pathol Inform 2016;7:31.
7. Mercan E, Aksoy S, Shapiro LG, Weaver DL, Brunyé TT, Elmore JG. Localization of diagnostically relevant regions of interest in whole slide images: A comparative study. J Digit Imaging 2016;29:496-506.
8. Mercan C, Mercan E, Aksoy S, Shapiro LG, Weaver DL, Elmore JG. Multi-instance multi-label learning for whole slide breast histopathology. In: Gurcan MN, Madabhushi A, editors. SPIE. California, United States: Proc. SPIE 9791, Medical Imaging Conference 2016 on Digital Pathology; 2016. p. 979108-11.
9. Roa-Peña L, Gómez F, Romero E. An experimental study of pathologist's navigation patterns in virtual microscopy. Diagn Pathol 2010;5:71.
10. Brunyé TT, Carney PA, Allison KH, Shapiro LG, Weaver DL, Elmore JG. Eye movements as an index of pathologist visual expertise: A pilot study. PLoS One 2014;9:e103447.
11. Raghunath V, Braxton MO, Gagnon SA, Brunyé TT, Allison KH, Reisch LM, *et al.* Mouse cursor movement and eye tracking data as an indicator of pathologists' attention when viewing digital whole slide images. J Pathol Inform 2012;3:43.
12. Jaarsma T, Jarodzka H, Nap M, van Merriënboer JJ, Boshuizen HP. Expertise in clinical pathology: Combining the visual and cognitive perspective. Adv Health Sci Educ Theory Pract 2015;20:1089-106.
13. Agrawal R, Imielinski T, Swami AN. Mining Association Rules between Sets of Items in Large Databases. SIGMOD Record (ACM Special Interest Group on Management of Data). ACM. Vol. 22. 1993. p. 207-16.
14. Extension of the limit theorems of probability theory to a sum of variables connected in a chain", reprinted in Appendix B of Howard RA Dynamic Probabilistic Systems, (Volume 1: Markov models), Dover Publ. 2007.
15. Markov A. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. In: Dynamic Probabilistic Systems (Volume I: Markov Models). California; United States: 1971. p. 552-77.
16. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-92.
17. Robert C, Casella G. A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. Statistical Science. Vol. 26. Institute of Mathematical Statistics; 2011. p. 102-15.
18. Otto M, Thornton J. Bootstrap Framework. Available from: http://www.getbootstrap.com. [Last accessed on 2017 Jun 23].

19. jQuery Foundation. jQuery JavaScript Library. Available from: http://www.jquery.com. [Last accessed on 2017 Jun 23].
20. Contributors O, CodePlex Foundation. OpenSeadragon JavaScript Image Viewer. Available from:https://www.openseadragon.github.io. [Last accessed on 2017 Jun 23].
21. Pillay R. IIPImage Image Server. Available from:http://www.iipimage.sourceforge.net. [Last accessed on 2017 Jun 23].
22. Aperio L. Aperio ScanScope CS. Available from: http://www.leicabiosystems.com. [Last accessed on 2017 Jun 23].
23. Cupitt J, Martinez K, Padfield J. VIPS. Available from: http://www.vips.ecs.soton.ac.uk. [Last accessed on 2017 Jun 23].
24. Vicknair C, Macias M, Zhao Z, Nan X, Chen Y, Wilkins D. A comparison of a graph database and a relational database. New York, USA: ACM Press; 2010. p. 1.
25. Han J, Shin DV, Arthur GL, Shyu C-R. Multi-Resolution Tile-Based Follicle Detection using Color and Textural Information of Follicular Lymphoma IHC Slides. IEEE; 2010. p. 866-7.
26. Alzubaidi L, Ersoy I, Shin D, Bunyak F. Nucleus Detection in H&E Images with Fully Convolutional Regression Networks. Cancun, Mexico; 2016. Available from: http://www.dlpr2016.csp.escience.cn/dct/page/70005. [Last accessed on 2017 Jun 23].
27. Findlay JM, Gilchrist ID. Active Vision: The Psychology of Looking and Seeing. UK:Oxford University Press; 2008. p. 240.
28. Treanor D, Lim CH, Magee D, Bulpitt A, Quirke P. Tracking with virtual slides: A tool to study diagnostic error in histopathology. Histopathology 2009;55:37-45.
29. Baum LE, Petrie T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics; 1966.