MDDT proposal for eeDAP: evaluation environment for digital and analog pathology

23 January 2017

1 Authors and Affiliation

Brandon D. Gallas and Qi Gong

- FDA/CDRH Office of Science and Engineering Laboratories, Division of Imaging, Diagnostics, and Software Reliability
  Silver Spring, MD, 20903

2 Product Description

eeDAP is being proposed as an MDDT for Clinical Outcome Assessments (COA). eeDAP is an evaluation environment for digital and analog pathology. eeDAP is a software and hardware platform for designing and executing digital and analog (microscope) pathology studies where the digital scan of a glass slide, or whole slide image (WSI) is registered to the real-time view of the corresponding glass slide on the microscope. This registration allows for different pathologists to evaluate the same fields of view (FOVs) in digital mode or in microscope mode. Consequently, it is possible to reduce or eliminate a large source of variability in comparing these modalities in the hands of the pathologist: the FOVs (the tissue) being evaluated. In fact, the current registration precision of eeDAP allows for the evaluation of the same individual cell in both domains. As such, a study can be designed where pathologists are asked to evaluate a preselected list of individual cells or specific FOVs in the digital mode and with the microscope. Consequently, paired observations from co-registered FOVs are collected allowing for a tight comparison between WSI and optical microscopy.

A reader study with eeDAP is intended to evaluate the scanned image, not the clinical workflow of a pathologist or lab. Instead of recording a typical pathology report, eeDAP enables the collection of explicit evaluation responses (formatted data) from the pathologist corresponding to very narrow tasks. This approach removes the ambiguity related to the range of language and the scope that different pathologists use in their reports.

Reader studies utilizing eeDAP are meant to focus on tasks related to specific histopathology features. Since certain image features can challenge image quality properties (color fidelity, focus quality, and depth of field), reader studies with tasks based on features can provide valuable information for the assessment of WSI and its role in clinical practice. eeDAP allows for the formulation of different types of tasks, many of which are currently available in eeDAP: free-text, integer input for counting tasks, a slider in a predefined range for a confidence scoring task (ROC task, receiver operating characteristic task), check boxes of specific

categories for a classification task, and marking the image for a search task. These simple tasks can be customized with moderate MATLAB programming skills to suit the study's specific purposes.

In what follows, we provide a quick summary of the eeDAP hardware and software. More details can be found in a paper by Gallas et al. [1].

2.1 eeDAP Hardware

The eeDAP hardware includes a microscope with a camera (mounted for simultaneous viewing with the eyepiece), a motorized stage (programmable with a stage controller and joystick), and computer with a monitor as shown in Figure 2.1. There is also a reticle in the microscope eyepiece. The reticle in the eyepiece is also synthesized in software; virtual reticle marks are superimposed on the digital WSI image and on the camera image. The reticle serves two purposes. The reticle marks allow for the localization of specific features during image registration. The reticle marks are also used to help reduce the area of the FOV or to specify individual cells to evaluate (create a *study* FOV).

eeDAP supports usb and FireWire (IEEE 1394) cameras. The camera pixel sizes tested are smaller than 10um. At 40x magnification, this corresponds to camera pixels that cover areas smaller than 0.25 x 0.25 um^2; in the common parlance, we say that the camera resolution is 0.25 um/pixel. Regarding the motorized stage, eeDAP supports communications with Ludl and Prior stages. The step size for the stages tested are smaller than 0.1 um. The computer needs to be a Windows 7 operating system fast enough with memory and space to manage large WSI images (~10GB). Additionally, the computer must be able to communicate with the camera (usb or IEEE 1394) and with the stage (RS-232). There are no specific requirements of the display, and in fact, the display may be considered a component of the WSI system to be evaluated.

2.2 eeDAP Software

The eeDAP software is made up of graphical user interfaces (GUIs) written in MATLAB [2]. Using eeDAP does not require a full licensed version of MATLAB. It can be run as a precompiled stand-alone application. The precompiled stand-alone application requires that the free MATLAB compiler runtime (MCR) library be installed [2]. The software uses the Bio-formats library to read digital WSI images and extract FOVs [3]. Normally the resolution of 40x WSI images is about 0.25 um/pixel. The Bio-formats library supports several proprietary WSI image formats such as svs, ndpi and tiff. Please note that the Bio-formats library was not used in earlier versions of eeDAP; specifically, it was not used in the version described in the paper by Gallas et al. [1]. The change to Bio-formats was made because it is better supported, it works on 64-bit systems, and it is available under the GNU public "copyleft" licenses.

The eeDAP software is made up of three GUIs: **study initialization, global image registration, and data collection**. During **study initialization**, eeDAP reads a study-specific input file. The input file contains the filenames of the WSIs, hardware specifications, and the list of tasks with corresponding FOV locations that will be interpreted by the pathologist. At the end of the study initialization, eeDAP extracts all the WSI FOVs for fast access and transforms the colors so that the image viewed in eeDAP is the same as the image viewed on the scanner-specific viewer.

Global image registration is only done if eeDAP is run in MicroRT mode. The global image registration is equivalent to finding the mathematical relationship between the stage coordinates and WSI coordinates. It allows the stage to move to a location that is the same as the WSI location. The global image registration requires three anchors, three pairs of stage-WSI registered coordinates. Each anchor is generated by a local registration: an (x,y) stage coordinate and an (x,y) WSI coordinate that correspond to the same specimen location. The local registration is an interactive process with the study administrator. The study administrator

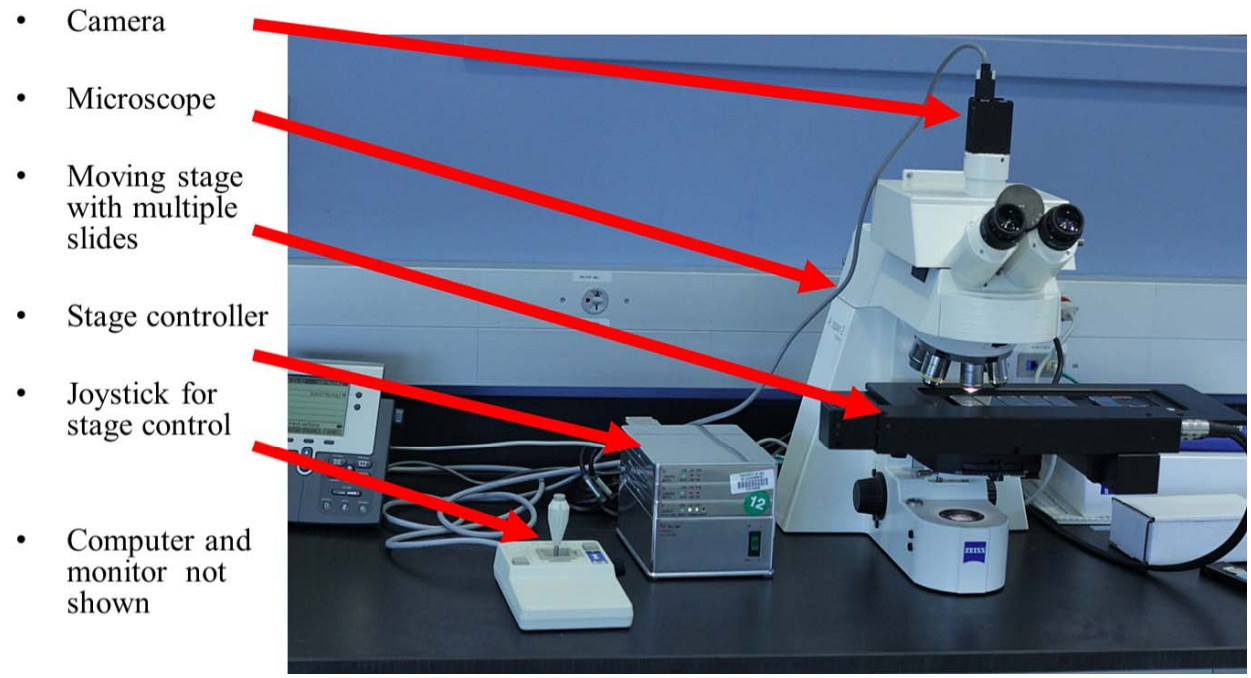

Figure 1: The eeDAP hardware: microscope, camera, computer-controlled stage with joystick, and a computer with monitor (not shown).

navigates the microscope to a landmark, takes a snapshot of the microscope FOV with the microscope mounted camera, and eeDAP records the stage coordinates. Then the study administrator clicks near the corresponding location in the WSI image displayed in the GUI. Figure 2.2 shows global registration GUI and the camera image of a landmark location chosen by the study administrator. It also shows the corresponding location in the WSI image; the location the study administrator should click on next. eeDAP then finds the local registration by maximizing the match of the camera image and the patch of WSI selected by the study administrator.

Once we have solved for the global registration between the WSI image and the stage (via the camera image), we have to register the camera image and the eyepiece. This is an interactive process with the study administrator. While looking through the microscope, the study administrator navigates the stage so that a reticle mark lands at a precise location. After clicking on the GUI to record that location, the study administrator navigates the stage so that the virtual reticle mark superimposed on the camera image lands on the same precise location and clicks on the GUI to record that location. The difference in the two positions determines the shift needed to register the eyepiece view and the camera view.

The **data collection** GUI is the same for digital and microscope modes. The GUI shows the WSI FOV and has interfaces for collecting the pathologist's responses. The difference between modes is that, when run in digital mode, the pathologist sits at the computer and interacts with the data collection GUI. In microscope mode, a study administrator sits at the computer and interacts with the data collection GUI. Meanwhile, the pathologist is engaged with the microscope. The pathologist speaks his or her responses for the study administrator to enter into the GUI. The study administrator also has the responsibility of verifying that the microscope is accurately registered.

Figure 2.2 shows the data collection GUI with the WSI patch and the corresponding camera image. In both of these images are virtual reticle marks, though they are difficult to see in the figure. The study administrator verifies accurate registration by comparing the images and the virtual reticle marks. If registration is not accurate, there are buttons to perform an automated local registration of the current FOV. Finally, the FOV that the pathologist sees in the microscope is round and it is larger than what they see on the GUI, but they are supposed to be the same. The mismatch comes from a slight misuse the term FOV. We should really say that eeDAP allows for pathologists to evaluate the exact same *study* FOVs in both modalities. A study FOV is based on the reticle.

For the cameras and stages tested, we are able to repeatably and reliably register the WSI image and the glass slide so that the pathologists can evaluate the same FOVs in both modes.

3 Context of Use

3.1 COU Statement

eeDAP is a Clinical Outcome Assessment used in reader studies for whole slide imaging premarket submissions (PMA or 510k deNovo) to compare the accuracy or reproducibility of pathologist evaluations of digital images on a display to those of glass slides on a microscope. The pathologist evaluations of patient tissue are the clinical outcomes. The accuracy or reproducibility is the clinical outcome assessment; this assessment reflects image quality.

3.2 The device or product area for which the MDDT is to be qualified.

eeDAP is to be qualified for the evaluation of WSI systems, also known as virtual microscopy systems, which can digitize whole slides at microscopic resolution in a short period of time [4]. WSI systems are part

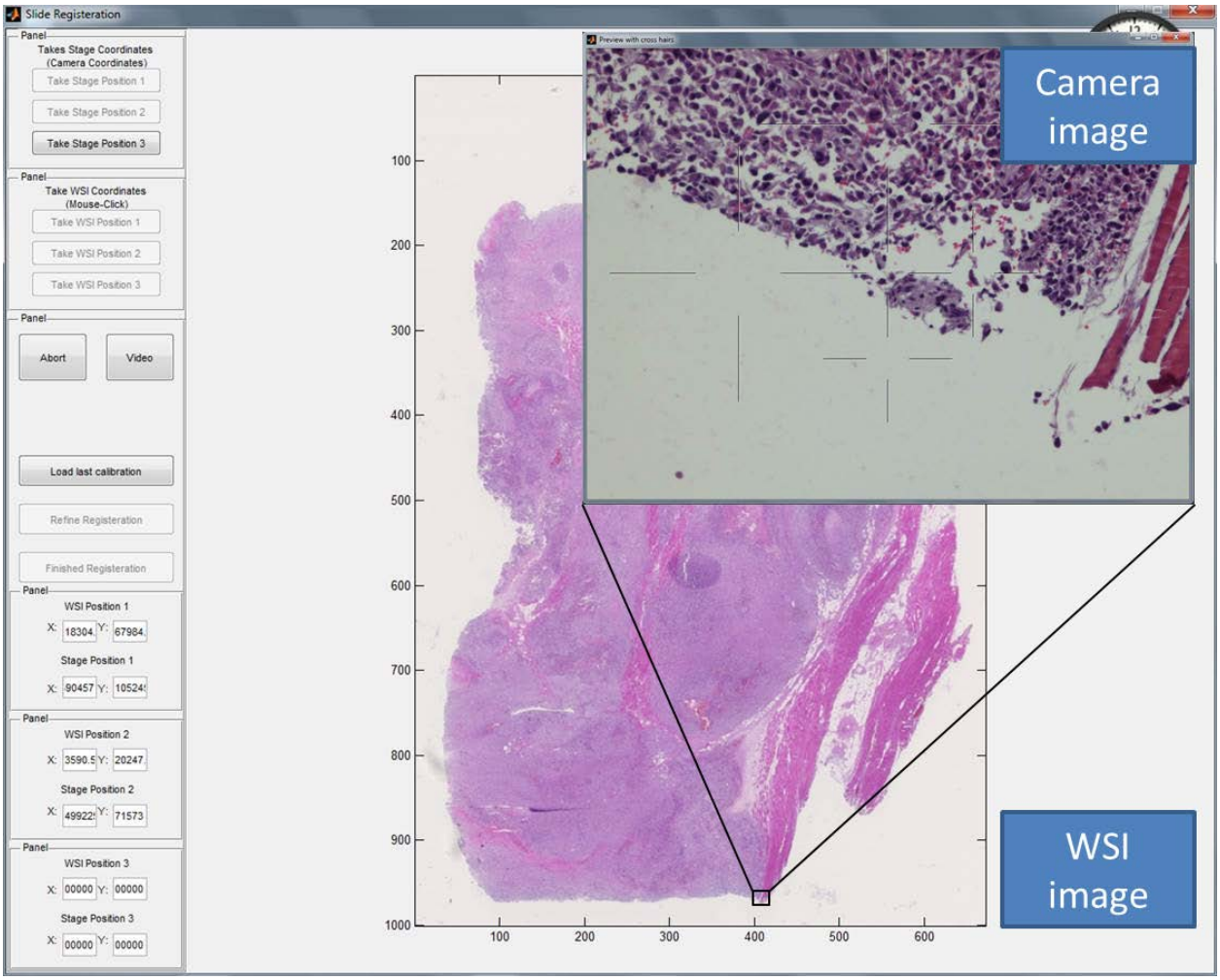

Figure 2: Screen shot of the image registration GUI, including the window showng the camera image.

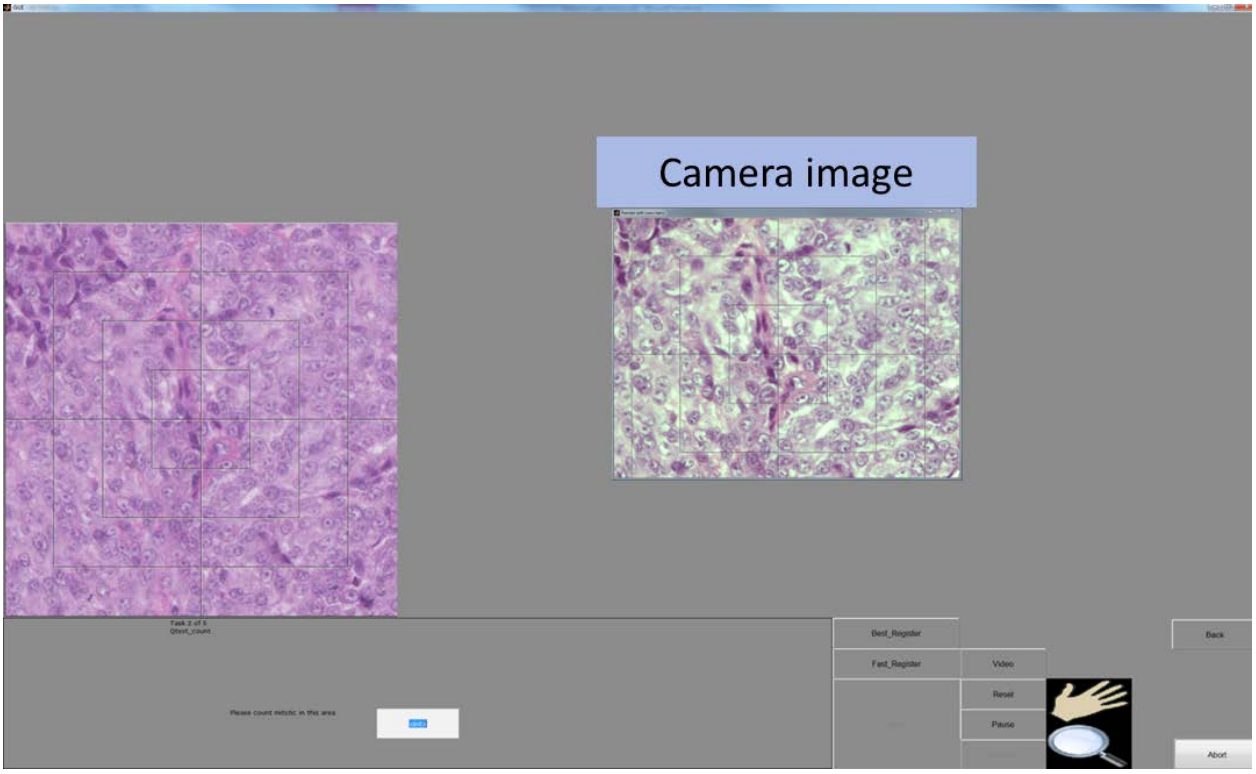

Figure 3: Screenshot of eeDAP data collection GUI with camera image of microscope view. In both of these images are virtual reticle marks, though they are difficult to see in the figure. The reticle marks help verify registration and to define a study FOV.

of the vision for digital pathology (DP). DP incorporates the acquisition, management, and interpretation of pathology information, including WSIs. The imaging chain of a WSI system consists of multiple components including the light source, optics, motorized stage, and a sensor for image acquisition. WSI systems also have embedded software for identifying tissue on the slide, auto-focusing, selecting and combining different fields of view (FOVs) in a composite image, and image processing (color management, image compression, etc.). Details regarding the components of WSI systems can be found in a paper by Gu and Ogilvie [5].

The potential public impact of WSI systems are well documented and include telepathology, digital consultation and slide sharing, pathology education, indexing and retrieval of cases, and the use of automated image analysis [6, 7, 8].

In addition to WSI systems, eeDAP could be used in the development and evaluation of WSI image analysis tools and CAD algorithms (computer aided diagnosis, detection, prognosis, etc.). Examples of such tools include:

- Color standardization/normalization
- Segmentation of tumors and other tissue structures
- Automated detection of different cell types and biomarkers
- Image retrieval
- Decomposing images containing multiple biomarkers (multiplex images)
- Diagnosis and prognosis, subtyping, and staging of cancer and non-cancer diseases

We do not plan to qualify eeDAP for the development of WSI systems or the development and evaluation of these adjunct products in this proposal in order to keep the COU focused. However, if eeDAP is qualified for the evaluation of WSI systems, it should also have value for the additional uses. A future MDDT proposal can consider expanding the COU if needed.

The stage of WSI development is advanced. WSI devices have received approval elsewhere in the world (European Union, UK and Ireland, and Canada.) In the United States, however, WSI devices are available for research and education, but they have not been approved for clinical use. The challenge for FDA approval is that there is not a clear pathway to approval. There is guidance for the technical assessment [9], but there is not guidance for the reproducibility/feature studies or the clinical trial. However, the FDA has communicated an outline for the clinical trial [10].

The stage of development of WSI image analysis tools and CAD algorithms is broad. There are likely applications that are ready for clinical use and applications that are still in the early stages of development. For the applications early in the development cycle, it is expected that the approval of WSI systems will lead to a substantial increase in data (images + annotations + outcomes). Consequently, it is expected that the pace of image analysis development will increase and more applications will be ready for clinical use in the near future.

3.3 The stage(s) of device development (e.g., early feasibility study, pivotal study, etc.) the MDDT will support

eeDAP is to be qualified to provide data for the analytic validation of a WSI system as is expected in a premarket submission. The specific regulatory path has not been established yet, but it is likely PMA or

510k deNovo. For analytic validation, the manufacturer must demonstrate that the WSI system creates digital images accurately and reliably for interpretation in the hands of the pathologist. Analytic validation complements the clinical performance studies and the technical performance assessment [9]: the technical performance of the WSI system and the components in the imaging chain, from image acquisition to image display. As the WSI technology matures, analytic validation has the potential to play a larger role in the approval process. Specifically, it is possible that analytic performance could be shown to be effective surrogate for clinical performance. If true, analytic performance could support the approval or clearance of WSI devices with a clinical study. Furthermore, it is possible that technical performance could predict analytic performance in the future as we understand both better. If true, technical performance could ultimately support the approval or clearance of WSI devices without analytic or clinical studies.

3.4 The specific role of the MDDT (for clinical uses this includes the study population or disease characteristics, as well as specific use – diagnosis, patient selection, clinical endpoints).

The rold of eeDAP is to be a platform for the design and execution of studies of pathologists performing task-based evaluations of tissue and cell features that are critical to diagnosis or differential diagnosis of disease. Such studies characterize image quality and the accuracy and reliability of the digital WSI images in the hands of the pathologist. When there is a reference (truth) result for the task, the endpoints of these studies characterize pathologist accuracy. When a reference result is not available (too burdensome or destroys the tissue), the endpoints of these studies measure the precision or reproducibility of pathologist evaluations: reader agreement from a single scan can provide a measure of precision, and reader agreement across multiple scans on the same or different WSI scanners can provide measures of reproducibility. Bias, correlation (or concordance), and percent agreement are typical agreement measures that can be used.

It is not always clear what level of accuracy and reproducibility is "good enough" for the analytic validation. This is why it is extremely useful to compare the accuracy and reproducibility of WSI evaluations to those from the glass slide on the microscope. This comparison controls for task difficulty and case selection. eeDAP allows for such a comparison, and it does so in a way that tightly reduces pathologist variability and correlates the results across the modalities (WSI and microscope), making for an efficient study in statistical and practical terms. Pathologist variability is reduced because all the pathologists evaluate the same FOVs. The results across the modalities are correlated because the pathologists are evaluating the same FOVs *in both modalities*. Reducing pathologist variability clearly improves statistical precision, and better precision allows studies with fewer resources. Regarding correlations, consider the following property of variances: $\mathrm{var}\,(A-B) = \mathrm{var}\,(A) + \mathrm{var}\,(B) - 2 \times \rho\sqrt{\mathrm{var}\,(A)\,\mathrm{var}\,(B)}$. Here we see that as the correlation ρ increases, the variance of the difference decreases, improving statistical precision and allowing smaller studies.

4 Advantages of eeDAP

The advantage of using eeDAP compared to typical clinical evaluation protocols is that eeDAP studies are fast and customizable, use fewer resources, and yield performance results that are more precise and reproducible.

The speed comes from the fact that FOVs are queued up and automatically presented to the pathologist; there is no need to search an entire slide. Additionally, the pathologist evaluations are recorded electronically in lock-step with the presentation of the FOVs. As such, there is little concern for transcribing errors.

Fewer resources (slides) are needed because the sampling unit is the FOV instead of the slide. Thanks to biological variability, the FOVs can often sample a diverse set of presentations of the feature (detectability, counts, orientations, morphologies, or classes). Precision and reproducibility are improved because pathologists are basing their evaluations on the same tissue.

Regarding precision and reproducibility, current practices lead to extremely noisy data. Specifically, let's consider the clinical protocol for mitotic counting as recommended by Smedly et al. [11]. The mitotic index, "should be determined by counting the number of mitotic figures in 10 consecutive hpf [sic. high-powered FOVs] commencing in the area of highest mitotic activity for oral and lip neoplasms and in random fields for cutaneous neoplasms." This protocol was followed in a study by collaborators at the NIH (Mark Simpson PI). As part of the study, the pathologists saved annotations of the FOVs that they used while evaluating the WSI. Figure 4 shows FOVs selected by different pathologists. There is very little overlap in the FOVs; pathologists are counting different mitotic figures from different tissue. Figure 4 shows the within-reader correlation of counts made using the microscope compared to the counts made using the WSI. Each point shows the counts from the same reader evaluating the same case in the two modes, and each count is the sum of 10 FOVs. It is clear that this data yields a lot of variability. The highlighted points show some extreme differences, and the highlighted band qualitatively identifies the variation in counts from H&E 40X WSI images when 10 counts are observed on the microscope.We expect to show that eeDAP will significantly reduce this variability by forcing pathologists to count mitotic figures in the same tissue.

5 Disadvantages of eeDAP

There are two disadvantages of eeDAP worth mentioning. First, when collecting data in digital mode, the image is displayed with Matlab. As such, eeDAP doesn't evaluate the native image viewer's human factors and workflow components. Also, there is currently no panning or zooming.

The other main disadvantage is that eeDAP is primarily designed for tasks that can be done on a finite FOV. The finite FOV may be at different magnifications, but eeDAP is not designed for complicated tasks that explore the WSI and aggregate different features. This disadvantage, however, keeps the focus of eeDAP on evaluating image quality.

6 Strength of Evidence

Tool Validity: Does the available data adequately support the validity of the measurement? Does the MDDT measure reliably and accurately? Depending on the tool type, this may include analytical, clinical, and construct validity, sensitivity, specificity, accuracy, precision, repeatability, external validity, reduction of bias, verification of the constitutive model, uncertainty quantification, numerical convergence, etc.

Plausibility: Is it scientifically plausible that the measurements obtained through use of the MDDT are related to the true outcome of interest? Is there a causal path or mechanistic explanation to connect the MDDT to the outcome?

Extent of Prediction: What data are available to demonstrate a predictive relationship between the MDDT and the true outcome of interest? What is the strength of that predictive relationship? Is the prediction repeatedly demonstrated in multiple studies or as a class effect? If relevant, is the conclusion (that the effect of treatment on the measurement obtained using the MDDT predicts the outcome of interest) supported by credible information?

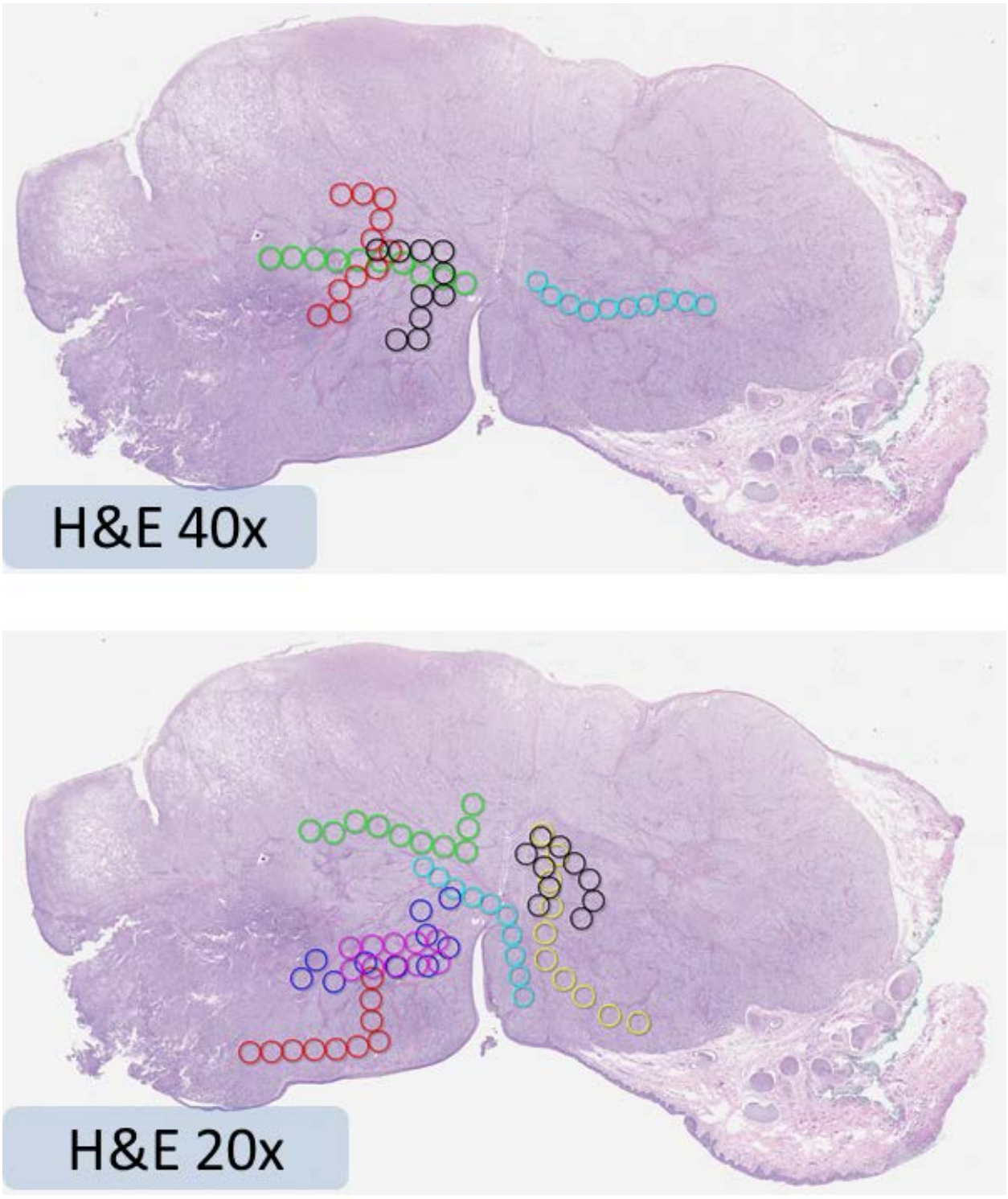

Figure 4: These two images show 10 FOVs selected by different readers while performing a mitotic counting task. The top image corresponds to counts taken at 20x and the bottom image corresponds to counts taken at 40x.

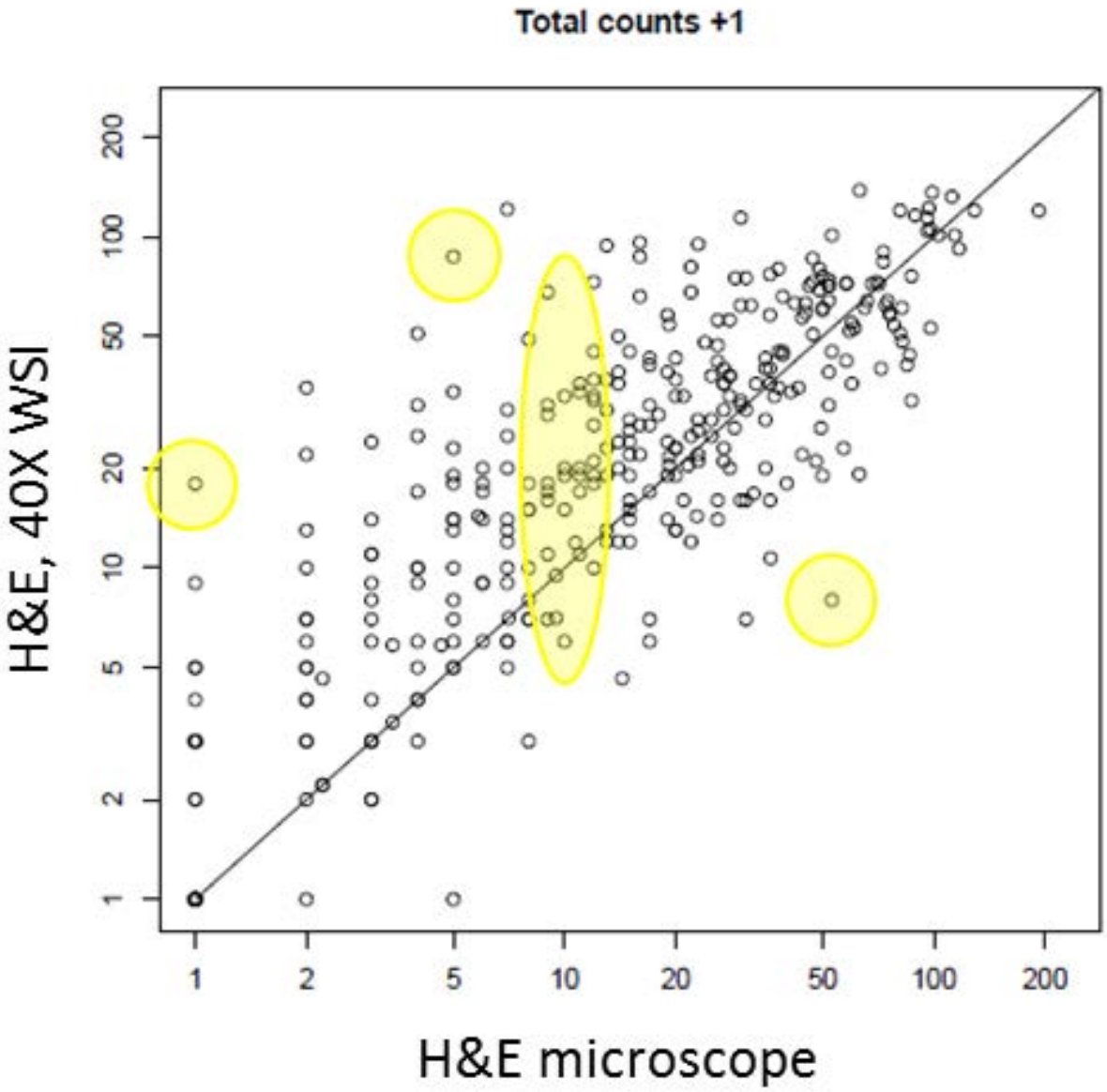

Figure 5: This figure shows the within-reader correlation of counts of mitotic figures made on the microscope and made on the WSI. Each point shows the counts from the same reader evaluating the same case on the two modalities. Each count is the sum of 10 FOVs, following the clinical protocol. Note that this plot is a log-log plot, and the subsequent coefficient of variation is very large. The highlighted points show some extreme differences, and the highlighted band qualitatively identifies the variation in counts from H&E 40X WSI images when 10 counts are observed on the microscope.

Capture: Does the MDDT fully capture the aggregate effect of the intervention on the true outcome of interest? Does the MDDT account for every major effect of the intervention? Are there available data which call this into question?

If our MDDT proposal is accepted, we will develop a full submission with supporting evidence of benefits to using eeDAP for the context of use.

eeDAP can be used to evaluate the performance of pathologists to find, classify, enumerate, and otherwise perform a feature-based task. The validity of the task is not disputable, but the relevance (extent of prediction and capture) to the device evaluation depends on whether or not the task is relevant to the clinical task or somehow stresses relevant imaging characteristics.

In some cases, the extent of prediction will be very strong. For example, an image analysis algorithm may be designed to find mitotic figures. In this case, a study with eeDAP could be created to explicitly evaluate the performance of the algorithm by collecting pathologist classifications on candidate cells and evaluating the algorithm's true-positive and true-negative rates.

In other cases, the extent of prediction will not be as strong or direct. A single eeDAP feature study may only demonstrate adequate image quality related to one task or one imaging characteristic and additional tasks may be needed to demonstrate adequate image quality related to other tasks or other imaging characteristics. We believe that a few different feature studies could demonstrate effectiveness that would be expected to generalize to a full and comprehensive intended use statement. For example, counting mitotic figures is a task that stresses the resolution of the WSI system, as fine details of the nucleus and chromosomes are needed to find and classify cells as mitotic figures. Also, identifying lesion tissue at low magnification is a task that stresses color resolution and dynamic range and could be designed for lower magnification. Tasks such as these may be considered as surrogates to *all* clinical tasks, and it will be the aggregate of evidence on the performance on such tasks that will adequately demonstrate effectiveness of a WSI system, an image analysis tool, or a CAD.

At a technical level, we need to validate that the registration process is accurate and the color representation in eeDAP is the same as would be in the proprietary WSI system-specific viewer.

7 Plan to get data

The main evidence that we plan to collect will be in the form of task-based studies of tissue features. At least one study will compare a mitotic counting study with eeDAP to the mitotic counting study following the clinical protocol mentioned above. We plan to demonstrate that pathologist evaluations taken with eeDAP are more reproducible than pathologist evaluations taken following a standard clinical method. In another study, additional evidence that we plan to collect will demonstrate that differences in imaging characteristics can be identified using eeDAP. For example, we intend to artificially degrade an imaging characteristic (resolution, focus, color) and conduct a study using eeDAP to show that the performance of pathologists is worse on the degraded system than on the baseline system.

For the evidence of technical validity we plan to design a study to characterize the accuracy of the image registration process and the color fidelity. For the color fidelity evaluation, we plan to follow the methods outlined previously [1].

8 Consent to Public Disclosure and Use

We hereby authorize FDA to make public sufficient information to support use of the qualified MDDT and for the general public to use and rely on data generated using the MDDT in gaining FDA clearance or

approval of other devices.

In fact, we are publicly sharing the submission process (documents and drafts, and communications with CDRH program staff) with the WSI working group at their NCIPhub group page https://nciphub.org/groups/wsi_working_group. The group page is open to the public and membership is open to anyone that requests it. Additionally, the compiled eeDAP software and all source code are shared on Github with HTML and PDF versions of user manuals [12]. The only prerequisite software for the compiled eeDAP software is the MATLAB runtime compiler [2], which is free to download and use.

References

[1] Brandon D. Gallas, Marios A. Gavrielides, Catherine Conway, Adam Ivansky, Tyler Keay, Wei-Chung Cheng, Jason Hipp, and Stephen M. Hewitt. Evaluation environment for digital and analog pathology (eedap): a platform for validation studies. *J Med Img*, 1(3):037501, 2014. doi: 10.1117/1.JMI.1.3.037501. URL http://spie.org/Publications/Journal/10.1117/1.JMI.1.3.037501.

[2] MathWorks. *MATLAB Runtime Comiler*. Natick, MA, US, Accessed 12/20/2016. URL https://www.mathworks.com/products/compiler/mcr.html.

[3] The Open Microscopy Environment. *Bio-Formats Documentation*. Dundee, UK, Accessed 12/20/2016. URL https://www.openmicroscopy.org/site/support/bio-formats5.3/.

[4] Marcial García Rojo, Gloria Bueno García, Carlos Peces Mateos, Jesús González García, and Manuel Carbajo Vicente. Critical comparison of 31 commercially available digital slide systems in pathology. *Int J Surg Pathol*, 14(4):285–305, Oct 2006. doi: 10.1177/1066896906292274. URL http://dx.doi.org/10.1177/1066896906292274.

[5] Jiang Gu and Robert W. Ogilvie, editors. *Virtual microscopy and virtual slides in teaching, diagnosis, and research*. CRC Press, Boca Raton, FL, 2005.

[6] Liron Pantanowitz, Paul N Valenstein, Andrew J Evans, Keith J Kaplan, John D Pfeifer, David C Wilbur, Laura C Collins, and Terence J Colgan. Review of the current state of whole slide imaging in pathology. *J Pathol Inform*, 2:36, 2011. doi: 10.4103/2153-3539.83746. URL http://dx.doi.org/10.4103/2153-3539.83746.

[7] Ronald S. Weinstein, Anna R. Graham, Lynne C. Richter, Gail P. Barker, Elizabeth A. Krupinski, Ana Maria Lopez, Kristine A. Erps, Achyut K. Bhattacharyya, Yukako Yagi, and John R. Gilbertson. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol*, 40(8):1057–1069, Aug 2009. doi: 10.1016/j.humpath.2009.04.006. URL http://dx.doi.org/10.1016/j.humpath.2009.04.006.

[8] Shaimaa Al-Janabi, André Huisman, and Paul J. Van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61(1):1–9, Jul 2012. doi: 10.1111/j.1365-2559.2011.03814.x. URL http://dx.doi.org/10.1111/j.1365-2559.2011.03814.x.

[9] FDA CDRH. *Guidance for Industry and FDA Staff - Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices*. Maryland, 2016. URL http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM435355.pdf.

[10] Nicholas Anderson. *Regulatory Challenges and Opportunities for Digital Pathology*, Accessed 12/20/2016. URL https://nciphub.org/resources/1927.

[11] R. C. Smedley, J. Lamoureux, D. G. Sledge, and M. Kiupel. Immunohistochemical diagnosis of canine oral amelanotic melanocytic neoplasms. *Vet Pathol*, 48(1):32–40, Jan 2011. doi: 10.1177/0300985810387447. URL http://dx.doi.org/10.1177/0300985810387447.

[12] Brandon D Gallas. *eeDAP v4p0: Evaluation Environment for Digital and Analog Pathology*. Division of Imaging and Applied Mathematics, CDRH, FDA, Silver Spring, MD, 2016. URL https://github.com/DIDSR/eeDAP/releases.