CDRH Critical Path Application FY 2019

Submissions will be accepted no later than October 19, 2018

For content and technical questions regarding the application, please contact: Stanley.Liu@fda.hhs.gov or Mimi.Nguyen@fda.hhs.gov

Section I: Basic Information

Lead Investigator

Name	Brandon D. Gallas
Position Title	Research Physicist
Office	OSEL - Office of Science and Engineering Laboratories
Division/Lab	Division of Imaging, Diagnostics, and Software Reliability
Project Role *(350 characters)*	Project management and vision. Lead and oversee study design and data analysis. Communicate to collaborators and stakeholders. Leverage a digital pathology working group with over 145 members including all kinds of stakeholders: https://nciphub.org/groups/wsi_working_group

Office Phone	+1 (301) 796-2531	E-mail brandon.gallas@fda.hhs.gov	**PI Expected Time Commitment** *Expressed as % of total workload*
Supervisor	Kyle Myers		45% but probably more

Proposal Title *(500 characters)*

High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an FDA-qualified medical device development tool (MDDT)

Critical Path Status: ○ New Application   ● Ongoing CP Project

Was a similar proposal also submitted to the FY19 Office of the Chief Scientist Proposal Call? (Full proposals only)   ○ Yes   ● No

Plain Language Summary

Provide a summary of the project in lay language and explain how this proposal helps CDRH achieve its vision. (3100 characters)

* Artificial intelligence (AI) promises to reduce pathologists' burden searching and evaluating cells and features on the slides; let the computer do it. The regulatory question is then, "How well can the computer algorithms do the tasks?" The most practical ground-truth for evaluating algorithm performance is pathologists' assessments of the WSI images. The problem is that clinicians make mistakes and don't always agree. Furthermore, the scanners have limited spatial and color resolution and currently produce a 2D slice of a 3D specimen. In this work, we plan to conduct high-throughput truthing studies to qualify data (images and annotations) as a medical device development tool. The data can then be used by any algorithm developer as the validation set for an FDA submission.

* We have developed a hardware and software evaluation environment for digital and analog pathology (eeDAP). eeDAP allows us to automatically present pre-specified regions of interest or individual cells and features on a microscope for pathologist evaluation. This allows us to compare location-specific computer algorithm results to microscope-based pathologist evaluations. Last year, we installed eeDAP on a 14-head microscope and completed a data-collection session, collecting evaluations from 12 pathologists simultaneously in a single visit to Memorial Sloan Kettering (MSK). We loaned the eeDAP system to MSK and they proceeded to conduct a study to compare mitotic figure (MF) counting on a microscope and four different WSI scanners. We did the primary analysis of the results using statistical methods that we developed that account for reader and case variability, and we are coauthors on the paper under review at Diagnostic Pathology.

* Since our last submission, we have identified two other partners and challenges that are still in the development stage, giving us a better opportunity to get the glass slides and impact the challenge. The first group is led by DIDSR colleagues; they are designing a challenge for the SPIE Medical Imaging community. The second group is organized at www.tilsinbreastcancer.org. This highly motivated group has more than 140 pathologists, and the chair, Roberto Salgado, wants to leverage his group to create an MDDT dataset. He believes that a regulatory-grade FDA-qualified dataset for the community to use in a submission to the FDA would encourage algorithm developers to focus on the detection of lymphocytes to prognose cancer. We just piloted a lymphocyte evaluation study at the annual meeting of the American Society of Clinical Pathology and hope to collect pivotal data in partnership with that and other society conferences that have thousands of pathologists in attendance, enabling high throughput truthing. We also have several new partners that are offering support with hosting data-collection events, digital data infrastructure and technology, and providing/recruiting pathologists.

Research Strategy: Aims

Please list the specific objectives of the proposed research (3000 characters)

* Objective 1: Find partners and resources (status: largely complete). Partner #1 is Roberto Salgado, chair of www.tilsinbreastcancer.org. He brings pathologists, slides/images, and the TILS evaluation task (tumor infiltrating lymphocytes: a prognostic biomarker for cancer). Partner #2 is John Freymann, the director of the Cancer Imaging Program – Imaging Cancer Informatics unit and Technical Manager of The Cancer Imaging Archive (TCIA). He is providing cloud infrastructure to support hosting, viewing, and annotating images. The remaining partners (Refer to "VII: Additional Information") will provide/recruit pathologists, host data-collection events, and provide algorithm expertise to support the development of reader studies and performance evaluation.

* Objective 2: Write the protocol (status: advanced). With our MSK collaborators we have conducted studies to compare mitotic figure detection on a microscope to detection on four whole-slide scanners. These studies were built on the 14-head microscope study last year. A paper is under review (Refer to "VII: Additional Information") and contains a study design and analysis that is the template for the studies to support a data MDDT (images + annotations + algorithm performance analysis) for the TILS evaluation task. October 3-5, 2018 we piloted a data-collection event at the annual meeting for the American Society for Clinical Pathology (ASCP) largely following the protocol template. We intend to take the results and experience to our partners to refine the protocol for clinical relevance, high-throughput efficiency, and statistical rigor. In parallel, we will develop and submit an MDDT proposal.

* Objective 3: Conduct data-collection events (status: planning). The ASCP experience demonstrated the potential of running studies at conferences where thousands of pathologists go. We collected 30 minutes of data from 25 pathologists, and we can do better. With our partners support, we are planning data-collection events at conferences, at single sites (like the MSK study on the 14-head microscope), and over the web: crowd-sourcing pathologists. We plan to conduct a data-collection event at FDA's White Oak campus with three eeDAP systems (microscope mode), three calibrated digital displays, and several un-calibrated laptops. This event will feed the current project and DIDSR's display research.

* Objective 4: Submit MDDT package (status: not started). We have data and experience. While most of the energy in the coming year will be about collecting data, we are expecting our partner's clinical trial experts and statisticians to start analyzing and modeling the data we have already collected. We plan to evaluate the performance and agreement of pathologists and AI algorithms (conceptually interchangeable). We will submit statistical analysis plans for feedback from FDA/CDRH through the MDDT program.

Research Strategy: Approach and Methods

Please describe the study design including the overall strategy, methodology, and analyses to be used to accomplish the specific aims of the research. (3300 characters)

* The novel aspect of this project is to classify individual cells (or other features) in pre-specified regions of interest (ROIs) using glass slides on a microscope and WSIs on a digital display. We can navigate the microscope to pre-specified ROIs using a system we developed at CDRH called eeDAP: evaluation environment for digital and analog evaluation. eeDAP is a software and hardware platform for designing and executing digital and analog pathology studies where evaluation ROIs in the digital image are registered to the real-time view on the microscope (https://github.com/DIDSR/eeDAP and Gallas2014_J-Med-Img_v1p037501, attached).

* The data-collection protocol is best discussed by example. Last year we completed a study using a 14-head microscope at MSK. We installed our camera and computer-controlled stage onto a MSK microscope, an Olympus BX45 frame that has 14-heads mounted on it. There were 12 study pathologists who counted mitotic figures (MFs) in 40 ROIs on 4 glass slides. The study pathologists also classified 128 candidate MFs.

* The candidate MFs were determined during a preliminary image curation study where 4 pathologists counted and marked the locations of MFs arising from the same 40 ROIs. We call them "candidate" MFs because only 21 of 92 were unanimously identified. The distribution of agreement results across the candidates reflects a spectrum of candidates from obvious and likely MFs to subtle and improbable MFs. This spectrum allows an efficient and effective evaluation of an AI algorithm.

* The data-collection protocol can be used for the TILs evaluation task in a manner like above. There are several reasons we are considering the microscope instead of the digital WSIs. First, the WSIs have limited spatial and color resolution and currently produce a 2D slice of a 3D specimen. Next, most pathologists in practice have never used WSIs in clinical practice. Lastly, the annotations will then be scanner agnostic. We can scan the slides with any scanner and map the annotations onto that WSI. This allows evaluation of algorithms that were trained on images from different scanners. (Algorithms are not generally robust across scanners.)

* The key principle that will guide the analysis methods will be that the results and error bars will account for reader and case variability. The reader is the pathologist, and the case is the patient, a patient image, an ROI within a patient image, or an individual cell within a patient image. Dr. Gallas has developed, validated, and publicly shared such tools and source code (https://github.com/DIDSR/iMRMC/releases). The current tools treat data where the truth is known. Dr. Gallas has been developing analogous tools for data that do not have truth. Truth from a panel of pathologists is somewhere in between. Such truth, or reference, has a level of uncertainty (reader variability) that should be accounted for in analyses. Along with reader variability, we are exploring the establishment of the reference as a probability, or likelihood, instead of a binary true/false label. This project will support that development.

Other Project Details

Please check all that apply to this application

☒ Human Subjects, Sample or Data Use   ☐ Animal Specimen/Data Use   ☐ IT Infrastructure Support

For more information on RIHSC, please contact Dora Vega.   *For information on IACUC, please visit the IACUC intranet page.*

FDA

Research Timeline, Milestones, and Deliverables
*Please provide specific information regarding staff, time spent on project, and projected timeline.
Include a plan for interactive discussions with stakeholders. (4200 characters)*

Year one of this project was awarded on 3/15. As such, we are currently in the third quarter of year one (Y1Q3).

Y1Q3 – Y1Q4 (Objectives 1 & 2): Recently, we presented an outline of the proposal by email and webinar to the WSI working group (Refer to "IV: Non-CDRH Collaborators") as well as FDA colleagues in OSEL and OIR. Staff from OIR and OSB are on this proposal. We plan to share this proposal and progress updates with these two groups regularly. The purpose for sharing with these groups is to get feedback, build consensus, recruit additional partners for conducting the reader studies, and to disseminate methods/tools, expertise, and the entire experience. We need this project to be transparent and in the public domain so that FDA scientists may collaborate hand-in-glove with industry (and the broader community) to produce regulatory-grade truth data and analyses.

We plan to finalize the source(s) of samples and obtain needed IRBs with the collaborators to allow for the sharing of glass slides, digital images, and other resources. We plan to convene a committee meeting of project partners in November to lay the plans for the project. We plan to share the experiences of the last year, including the protocols used in the MSK studies (mitotic figure counting) and the protocols used at the ASCP conference (TILS evaluation). We plan to refine the protocol for clinical relevance, high-throughput efficiency, and statistical rigor. We plan to share the project and protocols with the TILs working group at their annual face-to-face meeting Dec. 4, 2018. We plan to submit the MDDT proposal. We plan to present an advanced version of the protocol to the WSI working group.

Y2Q1-Y2Q2 (Objective 3): The annual meeting of the United States and Canadian Academy of Pathologists is March 16, 2019 at the National Harbor, MD (30 miles from the FDA White Oak campus). We plan to target this conference as our first data-collection event to provide pivotal data (annotations) for the TILs data MDDT. We plan to collect data on two microscopes and two digital displays throughout the conference, following the model demonstrated at ASCP. We plan to host a data-collection event at the FDA White Oak campus (target date June, 2019). We plan to collect data on three microscopes, three calibrated digital displays, and the uncalibrated displays of computers and laptops of local and remote participants via the web. We also plan to conduct two data-collection events outside the US: at the University of Leeds (UK) with the support of Dr. Darren Treanor (Refer to his attached letter of support, "VII: Additional Information") and at the European Congress of Pathology September 7, 2019 in Nice, France. Arrangements and commitments are not fully secured, but the experience at ASCP and our partners give us confidence that we will be successful with more than one of our planned data-collection events. If need be, we can target the sites and pathologists of our partners and collaborator Anant Madabhushi has offered support for an event.

Y2Q3-Y2Q4 (Objective 4): We will bring together the reader study data and representative algorithm results to produce regulatory-grade reports (study description and analyses). We will present a draft of the analysis methods to the WSI WG to generate feedback for revisions, to invite statisticians and algorithm developers to participate, and to keep that community up to date. After appropriate revisions, the data, analysis methods, and representative algorithm results will constitute the main evidence to support the MDDT qualification. We will then submit this to the MDDT program for feedback and revise the report to satisfy FDA reviewers and all stakeholders.

Data collection for the Mitotic Figure and Tumor-Infiltrating Lymphocytes studies run in parallel at ASCP 2018 meeting in Baltimore. Local destination for low travel cost.
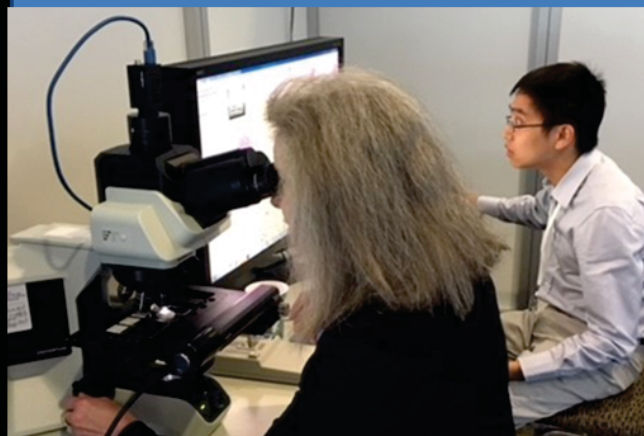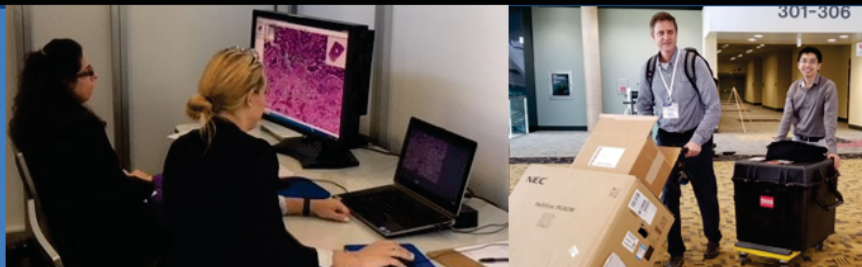
TILs study evaluated 80 regions of interest for a stromal TILs density task, and TIL marking task. Second study registered mitotic figures found on a mechanically driven microscope to a digital whole slide image. Programs ran with eeDAP software and embedded viewer system. Data from 26 pathologists; 13 collection hours total.

If you would like to add 1 graphic to this page, please click in the space above and upload. It will adjust to the space provided.

Previous Sources of Project Funding
Please list the past three (3) funding sources for this project.

FY Funded and Project Title	Award Amount	Program Funding
Year 1 of this project	$124,000.00	Critical Path

Progress to Date
Describe the progress of your project to date. This may include milestones met, outcomes, input from stakeholders, and changes to the research strategy or aims. (3900 characters)

Year 1, Objective 1: We determined the TUPAC data doesn't fit our needs. The glass slides are unavailable. I have engaged a new partner as a replacement, the TILS in breast cancer working group. They have added me to a funding proposal to the Breast Cancer Research Foundation to pay for creating infrastructure and data collection. I was also invited to contribute to an overview paper, Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning," that was submitted to Seminars in Cancer Biology.

I met with the organizer of the Camelyon challenge at Pathology Informatics in May. He is still intent on sharing the slides and he is determining how many slides are from his hospital that he can share easily. These will likely be Camelyon17 cases instead of Camelyon16 cases as he has more authority over those. He invited me to present at the European Congress of Pathology on Sept. 9 on the topic of AI evaluation methods and our project plans. Our 348 sponsored travel request has been approved (small bureaucratic victory). The trip will also give us a chance to finalize plans for some glass slides from the Camelyon data set. We hope to have the MTA in the hands of the lawyers by then. The Camelyon IRB for the data can be added to our local IRB for the evaluation of anonymized images and their use in reader studies.

We also joined a proposal to the UK National Institute for Health Research with Darren Treanor from the UK NHS. Unfortunately, we just learned that we did not with the funding award.

Year 2, Objective 2: I presented the study on the 14-head microscope at Pathology Informatics in May. MSKCC collaborators drafted a manuscript on a reader study comparing mitotic figure detection on four different WSI scanners to the microscope. They invited me and my assistant Qi Gong (who is funded by the CP award) to be coauthors. We helped them with the analyses. The manuscript's submission is imminent. This work is the foundation of the evidence module and future study designs and analyses.

In addition to Qi Gong, we have hired Weizhe Li, finishing an MPH in biostatistics from UMD with an internship with us. Weizhe also has a PhD in Biophysics from the Institute of Biophysics, Chinese Academy of Sciences. Beijing, China. He is part time with us and the NIH. He has downloaded most of the Camelyon data (~600 MB for Camelyon16 and 3TB for Camelyon 17). He has set up his work environment and has started to code the building blocks for training an algorithm that we can use to test our evaluation methods.

Our Cold Spring Harbor collaboration has been moving forward with the design of studies using our eeDAP system. They first executed registration accuracy studies that provide evidence supporting the context of use of the eeDAP MDDT. Since then we have been designing new work flows for eeDAP that allow the pathologist to evaluate slides on the microscope in a manner close to their clinical behavior while continuously recording the eyepiece view, stage coordinates, voice, and evaluation results (entered into the GUI). The workflows are faster because we will be able to map the results to the WSI after the fact. This collaborator has also sent us the green light (from their VP of Business Development and Technology Transfer) to begin the creation of a CRADA that will govern our data and bring in $25K/year to the FDA (travel or equipment).

What would happen to the project if funding was not continued this year? *(1500 characters)*

The project is in progress with lots of momentum. Please read the letters of support. If the project doesn't get funded, Dr. Gallas will try to keep up with the responsibilities of managing all the collaborators and the other DIDSR collaborators will try to manage the IRBs and agreements instead of doing the science that is needed. The biggest challenge is that it is unclear who will be the lab maintaining and improving the hardware AND software of the eeDAP system, who will help to conduct the data-collection events (preparing workflows, setting up equipment, proctoring the studies etc.), and who will provided statistical programming support (especially data cleaning). DIDSR is stretched very thin because our workload has increased this year with many algorithm submissions.

Regulatory Science Priorities

Please identify at most two priorities that apply to your proposal.

Develop methods and tools to improve and streamline clinical trial design

Leverage precision medicine and biomarkers for predicting medical device performance, diagnostics and prognosis

Regulatory Impact/ Alignment with Regulatory Science Priorities

Discuss the regulatory impact of your proposal and how your proposal addresses and/or supports CDRH's regulatory science priorities. (1700 characters)

At the end of our project, we will have regulatory-grade annotation data for validating a digital pathology AI algorithm (bioinformatics, software as a device) suitable for submitting to the FDA/CDRH MDDT program. The protocol, study design, data, and analyses with code will have been developed, collected, and shared in the public domain. We plan to design a sharing plan to allow any algorithm developer to use the data to evaluate the performance of their algorithm in a device submission to the FDA. The MDDT will then be a model for the broader pathology community (societies and associations, hospitals and provider networks) to impact the regulatory process by creating data sets for other tasks, other use cases, including use cases in radiology. By collaborating with the community, we are leveraging data and expertise that are not traditionally involved in regulatory decision making but should be. We are training a whole new group of stakeholders on the regulatory process and best practices in clinical trials. This will be consensus-building with real data to improve and streamline clinical trials. One of the key elements of this work is that it honors the current expertise and the full information that is associated with a pathologist viewing the physical specimen on a microscope. We must build trust and comfort as we adopt these revolutionary technologies.

Regulatory Science Metrics

Select the metrics that can be used to measure the success and impact of this proposal. For information on how to complete this section, please visit the CDRH Metrics SharePoint.

Level 4: Research Imperatives (max 3)	Level 3: Internal Outputs (max 3)	Level 2: External Impacts (max 2)
Collaboration	Updated and Streamlined Review Process	CDRH Tools in Use by Industry
Effective Management	Communications Impact	
Public Health Need		

Outcomes for Evaluating Impact

Using the selected metrics, please describe how this work will change CDRH processes and impact the public health. Provide details that demonstrate how the research imparative metrics will be met, specifically addressing each metric chosen. Then provide plans to produce internal outputs and accomplish external impacts. (2100 characters)

Like everything else, the microscope is going digital. The volume of data is staggering. MSK alone is scanning 40,000 slides a month. This data is meaningless without pathologist annotations and, ultimately, the clinical outcomes. The pathologist annotations are fueling a huge computational pathology revolution. "Deep learning" is everywhere, and it is coming to the FDA.
Dissemination: The WSI WG is hosted on NCIPhub (National Cancer Informatics Program Hub). NCIPhub is a site for community research and collaboration in cancer research and informatics. Users around the world can share resources, host online communities, and use collaboration tools . Similarly we are using GitHub, an international code hosting platform for version control and collaboration. Finally, we are partnering with the Cancer Imaging Program. This will publicize the current project and the MDDT program; the subsequent tools and protocols will be discoverable and available to a much larger audience and future investigators.
We believe that the CDRH-developed eeDAP system can play an important role in training and validating the algorithms at the cell annotation level. We have proposed eeDAP as an MDDT. The proposal has been accepted into the MDDT program. In the past year we have evaluated performance characteristics and are preparing for our next interaction with the program administrators.
The WSI WG has convened four face-to-face meetings attached to annual meetings of the Association of Pathology Informatics and the Digital Pathology Association. We have also hosted many organizational T-cons and webinars. We use NCIPhub to share all the presentations given and plans being made. The infrastructure and audience are in place to receive updates, archive information, and become active project participants.

FDA

Public Health Impact

Discuss the public health need this proposal addresses (disease area, device type, etc) and how it addresses the Center's vision. (3300 characters)

Digital pathology systems have the potential to help pathologists be more effective, more efficient, and have more impact. Increased effectiveness may be possible with the use of image analysis software (enhancement filtering, quantitation, and AI algorithms), facilitated consulting, and better education. Increased efficiencies may be possible because of high-throughput whole-slide scanning, automated image analysis performing laborious tasks, and integration with the patient digital record (archiving, annotation, indexing, and retrieval). Impact can most definitely be increased with telepathology for underserved populations and the use of AI algorithms when there are no pathologists available.

The proposed project will shape the regulatory process with the engagement of a broad spectrum of stakeholders. In a community setting, we will discuss and collect the data that needs to be collected to support an MDDT qualification of data (images + annotations). In addition to algorithms, the analysis methods and study designs allow for the evaluation of the WSI scanning device themselves, and any of the other components in the imaging chain, like the monitor. The analysis methods and the study designs are also not specific to digital pathology. They can be used for radiology or any imaging modality.

If you would like to add 1 graphic to this page, please click in the space above and upload. It will adjust to the space provided.

Participating CDRH Collaborators

Name | Dr. Dandan Xu

Office/Division | OMPT/CDRH/OSB/DBS/DSBII **Supervisor** Meijuan Li

Role Technical Consult **FTE % Time Commitment** | 10%

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Dr. Xu is a statistician in the Diagnostics Statistics Branch II of the Division of Biostatistics in the Office of Surveillance and Biometrics. This group is responsible for providing statistical consulting on digital pathology scanner and algorithms (and other product areas). She will receive quarterly updates on the project progress and provide feedback regarding study designs and analyses. If her time allows, she will provide statistical support with method development, modeling and simulation.

Name | Dr. Cheng Cui

Office/Division | OMPT/CDRH/OIR/DMGP/MPCB **Supervisor** Soma Ghosh

Role Advisor **FTE % Time Commitment** | 10%

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Dr. Cui is a senior reviewer with expertise in pathology and cytology. He will receive quarterly updates on the project progress and provide informal, high-level feedback regarding study design and analyses. He will also help coordinate interactions between the project collaborators and his group and educate the collaborators on the regulatory submission processes, including the pre-submission program.

Name | Dr. Weijie Chen

Office/Division | OMPT/CDRH/OSEL/DIDSR **Supervisor** Kyle Myers

Role Technical Consult **FTE % Time Commitment** | 10%

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Dr. Chen is an expert on statistical assessment methodologies for diagnostics and biomarkers. He has lead a number of initiatives on performance assessment studies in imaging, investigation of sources of variability, and AI/ML in radiology and pathology.

Name | Marios Gavrielides

Office/Division | OMPT/CDRH/OSEL/DIDSR **Supervisor** Kyle Myers

Role Collaborator/Working Group **FTE % Time Commitment** | 10%

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Dr. Gavrielides has experience developing artificial intelligence algorithms and conducting reader studies, within the application area of digital pathology. He also has expertise in the IRB process at the FDA. He will help design and execute reader studies, including the development and organization of study materials (slides, images, training documents, IRBs).

Name — Please refer to the letters of support attached in "VII: Additional Information"

Organization

Category

Collaborative Mechanism

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Name — The Whole Slide Imaging Working Group

Organization — The Whole Slide Imaging Working Group

Category — To be determined

Collaborative Mechanism — Informal Collaboration

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500) — The WSI WG is a community for discussing topics related to technical and pathologist performance evaluation using WSI images (https://nciphub.org/groups/wsi_working_group). The WSI WG is led by Dr. Gallas and Dr. Gavrielides (FDA), Dr. Treanor (Leeds U.), and Dr. Hewitt (NCI). There are over 145 members including all kinds of stakeholders from industry, government (including FDA), academia, clinicians, hospital and health provider network department chairs, VPs, and CEOs.

Name

Organization

Category

Collaborative Mechanism

Contribution
Describe the expertise of this collaborator and how this contributes to the proposal. (500)

Estimated FY19 CP Budget

List the expected project expenses to be covered by ONE year of CP funding.

Budget Category	Item Description	Justification	Cost
ORISE	Masters level bioengineer	Maintain hardware, prepare study materials (slides and images), coordinate logistics.	
ORISE	Masters level statistical programmer	0.5 FTE: Starting with CDRH-developed tools for sizing and analyzing reader studies, expand them to analyze data when truth is not known (under direction of senior statistician).	
Contracts	Data-collection events	Data-collection events will have conference-like costs: set-up, AV, telecommunications, cloud services, electricity, furniture, and space	
		Total Requested Budget	

Out Year Costs

Project the expected expenses for continuing the project AFTER ONE year of CP funding.

Budget Category	Item Description	Justification	Cost
		Total Budget	

Technical Reviewers

Provide up to three (3) FDA employees who could be technical reviewers for your project. These reviewers must be familiar with the research need/area but not involved with this specific proposal.

Name and Title	Center/Office	E-mail
Baku Patel	CDRH/OCD	bakul.patel@fda.hhs.gov
Daniel Krainak, or Zivana Tezak	CDRH/OIR	daniel.krainak@fda.hhs.gov, or zivana.tezak@ +
Yun-Fu Hu	CDRH/OIR	yun-fu.hu@fda.hhs.gov

Section VII: Additional Information for Reviewers/Leadership

Is there any additional information you would like to share with the reviewers and leadership? You may upload a maximum of 5 documents to this proposal but it is not guaranteed that all files will be reviewed by reviewrs. This can include but is not limited to letters of support from external collaborators/stakeholders, CVs/Resumes, and recent publications.

	Name of File	Description
Attach	CPscorecardFY18-Q3.pdf	Update for project Y1Q1 & part Y1Q2 (7/13/2018)
Attach	supportBinder-GallasFY2019.pdf	Letters of support
Attach		
Attach	Gallas2018_J-Med-Img_v1p037501	Manuscript describing eeDAP: "Evaluation environment for digital and analog pathology: a platform for validation studies"
Attach		

References

Please use this space to list any references relevant to your proposal. (2600 characters)

Below is a list of collaborators for this project. LOS stands for letter of support. The letters demonstrate an enthusiastic endorsement of this project by the community and the huge resources offered. Furthermore, this project will offer opportunities, expertise, resources, and shared experiences for the other digital pathology and algorithm research in DIDSR and OSEL, especially the proposal of Dr. Sahiner, "Leveraging imperfect post-market reference indices for the evaluation of artificial intelligence and machine learning devices." **Laura Coombs** from the American College of Radiology (ACR) has provided an LOS because ACR has the same aims as this project, just a different imaging modality and community. Together, our work will empower the front-line clinicians to be owners, drivers, and judges of the imaging products and software. Our work will also clarify what the clinicians want from key related industries: algorithms, picture archiving and communication systems, and voice recognition software for dictating image evaluations. ACR will help us connect with the College of American Pathologists, the American Society of Clinical Pathology, and other associations and societies to facilitate data-collection events at conferences and meetings where all the pathologists (and radioloigsts!) go. **Pathologist collaborators**: Roberto Salgado: TILs in Breast Cancer Working Group (> 140 pathologists, LOS) ** Darren Treanor: University of Leeds (UK: > 45 pathologists … 70 in region, LOS) ** Jim Crawford: Hofstra/Northwell (> 100 pathologists) ** Yukako Yagi: Memorial Sloan Kettering (> 100 pathologists) ** John Tomaszewski: U. Buffalo (> 25 pathologists) **Technology collaborators**: John Freymann: NCI Cancer Imaging Program (TCIA and infrastructure, LOS) ** Joel Saltz: Stony Brook U. (LOS) ** Ashish Sharma: Emory U. (QuIP: Quantitative imaging in pathology software and caMicroscope viewing and annotation software, LOS) ** Jeroen van der Laak, Radboud U. (LOS) ** Anant Madabushi, Case Western Reserve U. (LOS) ** Partha Mitra: Cold Spring Harbor Laboratory (LOS) ** Gabe Siegel: Augmentiqs

For questions, please contact: *Stanley.Liu@fda.hhs.gov* or *Mimi.Nguyen@fda.hhs.gov*

Critical Path Score Card

FDA

Lead PI | Brandon Gallas

Project Title

High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of pathology slides: leveraging data collected in international "grand challenges"

What milestones have been accomplished so far? Also please provide any update or delays related to the project.
(Note: Text area expands and may run over the lines)

Objective 1: We determined the TUPAC data doesn't fit our needs. The glass slides are unavailable. I have engaged a new partner as a replacement, the TILS in breast cancer working group. They have added me to a funding proposal to the Breast Cancer Research Foundation to pay for creating infrastructure and data collection. I was also invited to contribute to an overview paper, Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning," that was submitted to Seminars in Cancer Biology.
I met with the organizer of the Camelyon challenge at Pathology Informatics in May. He is still intent on sharing the slides and he is determining how many slides are from his hospital that he can share easily. These will likely be Camelyon17 cases instead of Camelyon16 cases as he has more authority over those. He invited me to present at the European Congress of Pathology on Sept. 9 on the topic of AI evaluation methods and our project plans. Our 348 sponsored travel request has been approved (small bureaucratic victory). The trip will also give us a chance to finalize plans for some glass slides from the Camelyon data set. We hope to have the MTA in the hands of the lawyers by then. The Camelyon IRB for the data can be added to our local IRB for the evaluation of anonymized images and their use in reader studies.
We also joined a proposal to the UK National Institute for Health Research with Darren Treanor from the UK NHS. Unfortunately, we just learned that we did not with the funding award.

Objective 2: I presented the study on the 14-head microscope at Pathology Informatics in May. MSKCC collaborators drafted a manuscript on a reader study comparing mitotic figure detection on four different WSI scanners to the microscope. They invited me and my assistant Qi Gong (who is funded by the CP award) to be coauthors. We helped them with the analyses. The manuscript's submission is imminent. This work is the foundation of the evidence module and future study designs and analyses.
In addition to Qi Gong, we have hired Weizhe Li, finishing an MPH in biostatistics from UMD with an internship with us. Weizhe also has a PhD in Biophysics from the Institute of Biophysics, Chinese Academy of Sciences. Beijing, China. He is part time with us and the NIH. He has downloaded most of the Camelyon data (~600 MB for Camelyon16 and 3TB for Camelyon 17). He has set up his work environment and has started to code the building blocks for training an algorithm that we can use to test our evaluation methods. Our Cold Spring Harbor collaboration has been moving forward with the design of studies using our eeDAP system. They first executed registration accuracy studies that provide evidence supporting the context of use of the eeDAP MDDT. Since then we have been designing new work flows for eeDAP that allow the pathologist to evaluate slides on the microscope in a manner close to their clinical behavior while continuously recording the eyepiece view, stage coordinates, voice, and evaluation results (entered into the GUI). The workflows are faster because we will be able to map the results to the WSI after the fact. This collaborator has also sent us the green light (from their VP of Business Development and Technology Transfer) to begin the creation of a CRADA that will

How much funding has been utilized this year and on what? (Note: Text area expands and may run over the lines)

It's all be obligated to ORISE for Qi Gong and Weizhe Li.

If you have not utilized all your funding, how much do you have remaining?

How do you plan to use the remaining funds before the fiscal cut off?

Do you anticipate applying for FY19 CP funding?   ● Yes   ○ No   ○ Not Sure

CDRH Metrics Collection: Collaboration & Dissemination

Metric	Description	Status/Fulfillment
Standards, studies, and guidances cited in applications	This demonstrates a broad understanding and acceptance of research.	
Research is cited in an expedited PMA	Research contributes to approval of a breakthrough medical device.	
Guidances impacted	Research helps the development of a new guidance document (the guidance cites the research).	
New regulations	Research contributes to the development of a new regulation.	
Licensures	Research results in a licensing agreement.	
Data sharing	Data generated from the research is shared externally.	
Web traffic to research outputs	Public-facing research has a site that generates hits or visits, or is found in internet searches.	
Impact of internal presentations related to the research	Research results are internally integrated, and reviewers are trained on the research results.	
Media coverage	Research is relevant and communicated well enough to be covered by the media.	
Submissions impacted	Research is cited in submissions to the FDA.	
Consults	Researchers receive consults related to the research.	
Potential standards impacted	Research is likely to contribute to standard development in areas of need.	
Research collaborations with external stakeholders	There is documented collaboration with external stakeholders.	1 presentation, 1 paper submitted, 2 proposals submitted
Input from industry stakeholders.	Industry stakeholders have had direct input into the need for, design of, or use of the research.	
External funding	External stakeholders provide additional funding	1 sponsor travel, 1 CRADA green light
Documented collaboration with internal stakeholders (scientific reviewer, stats, etc.)	Research has input from potential internal users, or is conducted in collaboration with internal users.	Provided update to Chen Cui (OIR/DMGP/MPCB)
Patents	Research results in a patent.	
Public meetings/workshops	Research is presented at a public meeting or workshop.	
Publications	Research is published in peer-reviewed journals.	
Presentations- external	Research is presented in a public capacity (not limited to internal use).	1 presentation
Attending conferences	Research is disseminated at conferences.	1 presentation
Miscellaneous high-impact communication with stakeholder	Center leadership incorporates research in communications.	
Expert training provided to review staff	Review staff is effectively prepared to evaluate the results of the research.	

CDRH Metrics Collection: Science Impact

Metric	Description	Status/Fulfillment
External use of internally or collaboratively developed computational tools	The actual use of tools that are developed by CDRH. The tool both satisfies a need and is accessible.	eeDAP system, images/slides, and study design were used to collect the data for the MSKCC ⊞
Number of devices approved due to innovative computational models and clinical trials designs that enable a rebalancing of non- and post- clinical data	Research that facilitates the use of postmarket data, bench data, and animal data.	
Reliability and reproducibility of test and trial methods	Research protocol permits reproducibility; results are significant.	
Creation of or impact on postmarket registry	Research affects clinical data registries for devices on the market.	
Creation of test methods, tools, and computational models	Research affects the preclinical evaluation of device safety and effectiveness.	Data for eeDAP MDDT. New eeDAP workflows
Development of new or improved clinical endpoints; validation of biomarkers (MDDTs)	Research affects the clinical evaluation of device safety and effectiveness.	Data for eeDAP MDDT. New eeDAP workflows.
Creation of or impact on other surveillance systems	Research contributes to surveillance of medical device safety.	
Creation of infrastructure and methods for evidence generation, capture, synthesis, and analysis	Research contributes to better evidence for medical device evaluation and decision-making.	Data for eeDAP MDDT. New eeDAP workflows
Relevance to emerging technology forecast	Research builds internal capacity and expertise in novel, needed areas.	Obtaining Camelyon images and developing building blocks of code for training AI
Reverse translation	Taking a product or clinical insight and further developing the science behind it.	
Product code relevance	Research is tied to an existing or new product code.	
Relevance to common premarket deficiencies	Relevance to common premarket deficiencies	
Potential test methods developed (MDDTs)	Research is likely to contribute to the development of a medical device development tool, a preclinical test method that may be used in submissions.	
Potential development of new or improved clinical endpoints; validation of biomarkers (MDDTs)	Research is likely to contribute to the development of a medical device development tool, a clinical test method that may be used in submissions.	
Relevance of research to manufacturing processes (e.g. 3D printing)	Research contributes to better understanding of the safety considerations associated with a manufacturing process.	

CDRH Metrics Collection: Mission Relevance

Metric	Description	Status/Fulfillment
Postmarket requirements and commitment studies impacted	Research affects the design and conduct of postmarket studies.	
Interactions with Congress	Research addresses a congressional requirement or request.	
Impact on a public health issue panel meeting	Research contributes to an advisory committee meeting.	
Progress in adverse events reporting and analysis	Research helps FDA track adverse events and address them.	
QALY	Research quantifiably contributes to an increase in quality-adjust life years for the public, per economic evaluation.	
Meets division/office/center/FDA priorities	See FDA priorities page for reference.	
Relevance to postmarket surveillance signal (ad hoc)	Research is in direct response to a device safety concern.	
Relevance to a public health risk (warning letters, recalls, seizures, injunctions)	Research is in direct response to an emergent public health issue.	
Related to voluntary recall	Research is in direct response to a voluntary recall.	
Compliance actions impacted or resolved	Research contributed to the resolution of a postmarket compliance concern.	
Translating bench-to-biology	Research bridges the gap between bench testing and biological effects.	
Translations from reduced biological model system to human	Research bridges the gap between the results of animal testing and the therapeutic and safety ramifications for patients.	

Is there any additional information you'd like to share regarding your progress?

Digital pathology Images are critical to this project. Most of the images are hosted on box, google cloud, or similar cloud services. We are struggling getting these images into the FDA to do our research because the websites are blocked. I have an FDA box account, and I have to use my home computer to download the images from the blocked websites and then upload them to the FDA box. I asked for help from ERIC and said the sites were blocked and they could not help. Can you?

Also, my UK collaborator has forwarded an opportunity for us to host PhD students as interns ... pretty much free of charge. They are kicking off a center for doctoral training in artificial intelligence. The challenge, in case you didn't already think of it, is security. FDA security for foreign visitors has gotten very challenging. Could we get a blessing from high up the chain of command to maybe get some exceptions? Seems like a win-win.

Brandon D. Gallas, PhD
Division of Imaging, Diagnostics, and Software Reliability
Office of Science and Engineering Laboratories
Center for Devices and Radiological Health
10903 New Hampshire Ave
WO-62 Rm 4104
Silver Spring, MD 20993

October 18, 2018

Dear Brandon:

As the Senior Director of the American College of Radiology (ACR) Data Science Institute (DSI), I am writing this letter on behalf of the ACR in strong support of your proposal, High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an FDA-qualified medical device development tool (MDDT). The goal of the proposal is to qualify a set of pathology images and annotations as a medical device development tool (MDDT) so that algorithm developers can use the data as a validation data set for FDA submission. The goal is directly in keeping with the plans that the ACR DSI has for creating validation datasets for medical imaging. Our hope is that we can collaborate with you in this space to create radiology\pathology validation datasets for common use cases. Such a collaboration would provide a very powerful mechanism for advancing artificial intelligence algorithms in a clinical setting to improve patient care.

We hope that your proposal is funded, and we are looking forward to future collaborations on this topic with the ACR.

Sincerely,

Laura P Coombs

Laura Coombs, PhD
Senior Director, Informatics and Data Science Institute
American College of Radiology

HEADQUARTERS
1891 Preston White Drive
Reston, VA 20191
703-648-8900

GOVERNMENT RELATIONS
505 Ninth St. N.W.
Suite 910
Washington, DC 20004-2173
202-223-1670

CLINICAL RESEARCH
1818 Market Street
Suite 1720
Philadelphia, PA 19103-3604
215-574-3150

AMERICAN INSTITUTE FOR
RADIOLOGIC PATHOLOGY
1100 Wayne Ave., Suite 1020
Silver Spring, MD 20910
703-648-8900

Wednesday 17th October 2018

Dear Brandon,

Here is a letter of support for your internal *Critical Path* funding proposal titled, "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT." I am enthusiastic about your plans to create regulatory grade annotation data (medical device development tool), and I am willing to participate and recruit readers from the consortium I lead (www.tilsinbreastcancer.org).

The evaluation of tumor infiltrating lymphocytes appears very promising for prognosing cancer, and the pathologist community needs training in this area. I chair a consortium of mainly expert-pathologists that was installed in 2014 and includes mainly pathologists with expertise and interest in the immuno-oncology biomarker area, including expert clinicians and statisticians. For a full member list, please consult www.tilsinbreastcancer.org. See also https://www.tilsinbreastcancer.org/researcher-resources/ for a recent Special Issue in Seminars in Cancer Biology, Impact Factor 10.198) this group has developed on Immuno-Oncological Biomarkers. You co-developed one of those papers on machine learning and TIL-evaluation (https://www.sciencedirect.com/science/article/pii/S1044579X1830066X).

The group has evolved to having >100 pathologists on board, from about 40 different laboratories in the US, EU, Asia and South-America, including the expert-pathologists from the Biomarker Groups of Merck, Roche-Genentech, Astra Zeneca, all having expertise on TIL-assessment. We do not ask for funding from the participating companies to avoid conflicts of interest. The continuous inclusion of experienced pathologists, assisted by expert clinicians like Sherene Loi, Sylvia Adams and Sybille Loibl, and expert statisticians like Stefan Michiels make the Working Group the leading expert group on TIL-assessment worldwide.
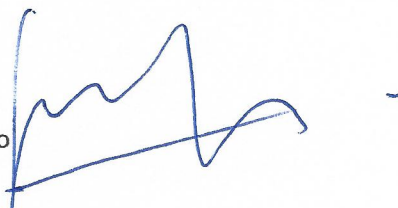
Relevant for the submission of this proposal, the Working Group, with your input has developed a manuscript entitled "Scoring of tumor-infiltrating lymphocytes: from manual estimation to machine learning", Klauschen et al. Seminars in Cancer Biology (Impact Factor 10.198). This is a perspective paper on TIL-assessment by machine learning methods and serves as the basis for this grant submission.

If I can, I will also use my network to introduce you to leadership people in pathologist societies and associations. I agree with you that these are excellent opportunities to get pathologists to participate in the studies FDA is conducting on quality assurance of machine learning applications in pathology. On a related note, I would like to invite you to speak at our meeting Dec. 6, which is attached to the San Antonio Breast Cancer Symposium. It will be a great opportunity to share the project, develop the data-collection protocol, and possibly collect some data!

I wish you success with your funding application and your aims and objectives. I think you are doing a great job leading the communities of pathologists, hardware manufacturers, and algorithm developers towards methods and tools that will have a positive impact on public health.

Sincerely,

Roberto Salgado

Histopathology Department
Level 5, Bexley Wing
St James University Hospital
Leeds LS9 7TF
United Kingdom

Tel. +44(0)113 20 67594
darrentreanor@nhs.net

18 Oct 2018

Dr Brandon Gallas

Re: Support for Critical Path Proposal

Dear Dr Gallas

I'm writing this letter in support of your Critical Path proposal entitled "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT".

As you know, at University of Leeds we have been researching digital pathology for over 15 years, with one of the largest digital pathology image repositories in the world (386,000 slides, 150 Terabytes of image data). In our hospital at Leeds Teaching Hospitals NHS Trust we have transitioned to a fully digital workflow for clinical pathology diagnosis, scanning over 1000 slides a day, and 6 of our pathologists working fully digitally for primary diagnosis, with the other 35+ following over the next 12 months. We prioritise an evidence based approach to digital pathology, with over 80 papers specifically on digital pathology including work on clinical validation, colour and human computer interaction. Our work has led to the UK national guidelines for digital pathology and significant partnerships with several vendors including Leica Biosystems, Roche and Sectra.

We think your proposal is timely and practical. The FDA/CDRH Medical Device Development Tool program seems ideal to demonstrate how the clinical care community can impact the evaluation of AI in digital pathology and gain trust in the use of AI tools. The pathologists can have ownership of their contributions – their annotations – by creating an FDA-qualified validation dataset to be used by any and all algorithm developers. We are particularly impressed by the leading position that FDA have taken in digital pathology and the coming AI revolution in healthcare, and your work furthers that advanced position for your organisation.

We would be pleased to support you in any way we can, including:
- Access to our images, data, and staff
- Input on experimental design

- Participation in your experimental work
- Supporting and co-organising an event at the UK Pathological Society in 2-5 July 2018 at which we could run a data-gathering exercise

We remain at your service for any areas where my expertise may be of value to your research.

Sincerely

_____
Darren Treanor BSc (Computing) PhD FRCPath
Consultant Pathologist/ Honorary Clinical Associate Professor

leidos

Leidos Biomedical Research, Inc.

October 15. 2018

Brandon D. Gallas, PhD
U.S. Food and Drug Administration
Center for Devices and Radiological Health
Division of Imaging, Diagnostics, and Software Reliability

Re: Letter of Support for MDDT Proposal

Dear Dr. Gallas,

I am writing to express my enthusiastic support for your funding proposal entitled "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT." This work will focus on the qualification of data (images + annotations) by the FDA/CDRH medical device development tool (MDDT) program and "crowdsourcing pathologists" at 1-2 conference "data-collection" events and via the web. The MDDT qualified data would be available to algorithm developers to be used to validate their algorithm performance in their submission to the FDA

I am the director of the Cancer Imaging Program – Imaging Cancer Informatics unit and Technical Manager of The Cancer Imaging Archive (TCIA) which is supported by the National Cancer Institute and provides a vast, freely accessible, open archive of cancer-specific medical images and metadata. The archive, including software, processes, data, and staff, is housed at University of Arkansas for Medical Sciences.

As the core mission of TCIA is to make highly valuable data sets available to the cancer research community, TCIA would be well suited to provide infrastructure to support the hosting of images and for viewing and to host annotation software to support the primary application of your proposal to evaluate TILs (tumor infiltrating lymphocytes).

Sincerely,

John Freymann (Contractor)
Director, CIP-ICI
Applied/Developmental Research Directorate
Frederick National Laboratory for Cancer Research
Leidos Biomedical Research, Inc.
Support to: Cancer Imaging Program/DCTD/NCI
9609 Medical Center Dr., Room 4W-322
Bethesda, MD 20892-9729

Joel H. Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics
Vice President for Clinical Informatics, Stony Brook Medicine
Associate Director, Stony Brook University Cancer Center

October 17, 2018

Brandon D. Gallas, PhD
U.S. Food and Drug Administration
Center for Devices and Radiological Health
Division of Imaging, Diagnostics and Software Reliability

Dear Dr. Gallas:

On behalf of the Department of Biomedical Informatics at Stony Brook University it is my pleasure to give support to your proposal entitled, "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT."

As a leader in research in Pathology Informatics, over the past twenty years, I have led the development of several highly impactful, innovative systems to support the analysis, visualization, and management of whole slide imaging data. My published whole slide virtual microscope system with concepts was ultimately adopted by many commercial vendors and embodied in the NCI supported caMicroscope system. I have also led a variety of multi-disciplinary efforts for creating cutting-edge machine learning algorithms for analysis of whole slide tissue imaging data and extraction and classification of imaging features. These algorithms include methods for classifying whole slide images and image regions, segmenting microanatomic structures and for carrying out analysis of IHC stained tissue. My overall research program focuses on the development of principles, techniques and tools that can be used in cancer research studies to assemble a coherent biomedical picture by integrating information from multiple complementary pathology, clinical, radiology and molecular data sources.

Much of this research has been in collaboration with Dr. Sharma's group at Emory University; one key Emory collaboration has been the development of high resolution segmentation of TILs in multiple tumor types, and high resolution classification of tumors in breast cancer with ongoing efforts to generalize to additional tumor types. Our current efforts in this area are supported by the National Cancer Institute Information Technology for Research (ITCR) program, CBIIT and SEER programs. I am PI of NCI ITCR U24 CA180924 and NCI SEER UG3 CA225021 both with focus on digital Pathology visualization, algorithms, data management and applications and MPI of ITCR U24 CA215109 (with Drs. Fred Prior PI and Ashish Sharma MPI) with work that includes integrating digital Pathology capabilities into TCIA. Finally I lead the Stony Brook subcontract for Dr. Sharma's CBITT funded digital pathology informatics project.

The Stony Brook Department of Biomedical Informatics looks forward to collaborating with the FDA in the management of whole slide imaging data including images and derived metadata as well as in the development and evaluation of pathology related deep learning algorithms. We are delighted to make available software and systems for management of annotation/markup data from pathologists and from algorithms. We will work with your team to support the data collection activity and will host images, annotations and derived metadata in TCIA. In collaboration with Emory, we will run our algorithms on WSIs submitted to the FDA and will also make the code available to the FDA. We are also delighted to offer help in the ground truth generation process; the Chair of the Department of Pathology at Stony Brook, Dr. Kenneth Shroyer, has offered the services of his pathologists to help support the crucial truthing effort.

Sincerely,

Joel H. Saltz

October 18, 2018

Brandon D. Gallas, PhD
U.S. Food and Drug Administration
Center for Devices and Radiological Health
Division of Imaging, Diagnostics, and Software Reliability

Dear Dr. Gallas

I am writing this letter to express my support to your proposal entitled, "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT." As you well know, I lead the caMicroscope project — a CBIIT funded project to support digital pathology informatics and I'm also part of the team that manages the Cancer Imaging Archive focusing on Digital Pathology capabilities for TCIA. I'm also the PI/MPI (together w/ Fred Prior – Chair, Dept. of Biomedical Informatics at UAMS, and Joel H. Saltz – Chair, Dept. of Biomedical Informatics at Stonybrook University) of an ITCR U24 (1U24CA215109) that is tasked with the Sustainment and Scalability of TCIA to support Quantitative Imaging Informatics in Precision Medicine. Through these and other related activities I am well aware of the value of high-quality labeled data and its usefulness in the development, training, and validation of AI algorithms for digital pathology. I was therefore very excited to learn about your plans to run a controlled data collection activity for characterizing TILs in breast cancer cases.

We will work with your team to support the data collection activity, incl., the hosting of images in TCIA; the development of annotation templates that accompany slide markup; as well as any automated TIL characterization efforts that may arise from this project. As you are aware, Dr. Joel Saltz and I have published methods to extract and characterize TILs in various cancer types. If interested, we'd make our pipelines available to your project. I wish you the best and look forward to working on this important and exciting activity.

Sincerely,

Ashish Sharma, PhD
Assistant Professor, Dept. of Biomedical Informatics
Vice Chair of Education and Training in Biomedical Informatics
Emory University

1639 Pierce Drive NE, Suite 4105, Dept. of Biomedical Informatics,   Tel: 404.654.0124
Woodruff Memorial Research Building, Atlanta, GA 30322   Email: ashish.sharma@emory.edu
The Robert W. Woodruff Health Sciences Center
An equal opportunity, affirmative action university

To Dr. Brandon Gallas and the FDA Critical Path Proposal Review Committee,

I support Dr. Brandon Gallas' proposal, "High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT." I believe he has an excellent plan for conducting studies to create regulatory grade annotation data for a medical device development tool (MDDT). I am willing to participate and recruit pathologists from my organization. The evaluation of tumor infiltrating lymphocytes (the primary application) appears very promising for prognosing cancer, and the pathologist community needs training in this area.

Additionally, I have a grant to organize a speaker series on AI in digital pathology and I'd like to invite Dr. Gallas to speak and conduct a data-collection event as part of that. I believe we can get a respectable turn out of pathologists to hear about the project and be study participants. My team is also very active in the development, investigation, and deployment of AI in digital pathology. I believe we can provide algorithm expertise and outputs to support the development of the reader studies and performance evaluation. If I can, I will also use my network to help the project conduct studies at conferences where the pathologists go. These are excellent opportunities to get pathologists to participate in the studies.

My CCIPD team has developed pioneering computer aided diagnosis, pattern recognition, image analysis tools for diagnosis and prognosis of different types of cancers (prostate, breast, medulloblastoma, oropharyngeal) based on quantitative and computerized histomorphometric image analysis of digitized histologic biopsy tissue specimens. This novel approach involves quantitatively mining the histologic image data for hundreds of image features via sophisticated image segmentation, feature extraction, machine learning and pattern recognition methods and then predicting the risk of disease recurrence and patient prognosis. We have also pioneered new ways of combining histomorphometric imaging features with "omics" derived biomarkers for improved and integrated prediction of prostate cancer outcome. We have published over 140 peer-reviewed journal papers, over 160 peer-reviewed conference papers, and have over 70 patents awarded or pending.

The aims of this project are what the community needs. It engages pathologists, giving them a voice and role in the regulatory process. If the community is involved in creating the validation data and evaluating algorithms, there will be confidence in the technology.

Sincerely,

Anant Madabhushi, PhD
F. Alex Nason Professor II
Department of Biomedical Engineering,
Director, Center of Computational Imaging and Personalized Diagnostics
Case Western Reserve University

Department of Biomedical Engineering
Case School of Engineering / School of Medicine

Mailing Address:
Case Western Reserve University
Wickenden Building, Room 340
10900 Euclid Avenue
Cleveland, Ohio 44106-7207

Phone   (216) 368-4063
Fax   (216) 368-4969

Cold Spring Harbor Laboratory

Partha Mitra, Ph.D., Professor

Dr Brandon Gallas, Ph. D
Food and Drug Administration
Friday Oct 19, 2018

Dear Brandon,

I write in strong support of your internal proposal for a study on "*High-throughput truthing of microscope slides to validate artificial intelligence algorithms analyzing digital scans of same slides: data (images + annotations) as an MDDT*". I am looking forward to collaborative interactions and we have set up your eeDAP device at CSHL to aid the study in question. Over the last year, as you know we have made considerable progress, with your help, in modifying eeDAP to enable the recording stage position, so that we can study the behavior of an AP physician performing diagnoses *in situ*.

As part of this study we plan to compare the performance of human Anatomic Pathologists with automated or semi-automated algorithms on tasks involving detection of relevant objects in Whole Slide Imaging data of cancer tissues. A staff computer scientist as well as a neuroscientist will participate in this study. I am collaborating with Dr Adam Kepecs, a Professor at CSHL, and Dr Jim Crawford, Professor and Chair of Laboratory Medicine at Hofstra/Northwell. Dr Crawford will be providing the slides with tissues which we will scan using Nanozoomer WSI scanners in my lab, as well as expertise in Anatomic Pathology. My laboratory has considerable expertise in AI/Machine Vision methods as applicable to tissue images, developed previously in the context of Neuroanatomical image data. I am also collaborating in this context with Prof. Mohanasankar Sivaprakasan of the Health Care Technologies Innovation Center at IIT Madras, Chennai, India.

I do not see us at CSHL making a marketing submission directly in relation to the eeDAP study (the subject of this letter), which is being conducted in the public domain like the rest of our academic research. The focus of this research project is scientific and algorithmic. However, I am interested in ultimately developing algorithms that have practical clinical utility. I cannot rule out the possibility that I will at some future date participate in a submission to the FDA for a digital pathology software application. I do not expect however that this will impact our current study in any way.

I wish you success with your funding application and the collaborative research that this will enable.

Sincerely,

Partha P Mitra
Crick-clay Professor of Biomathematics
Cold Spring Harbor Laboratory

Radboud umc

824

Date	Our reference	Page	Dutch Chamber of Commerce
18 oktober 2018	18.10.2018/MBal	1 of 1	trade register 41055629/4

Your reference	Contact		

Subject:
Proposal Dr Brandon D. Gallas

Dear reviewers,

With this letter I want to express my support for the proposal submitted by Dr Brandon Gallas, aimed at collecting evidence for use of computational pathology algorithms. My own research focuses on the development of such algorithms, and I am therefore very interested in being involved in this project.

My research focuses on the development of algorithms that may aid anatomic pathology practice. Last few years, we have witnessed a significant increase in the possibilities of machine learning (ML). Combined with whole slide imaging, ML has the potential to automate certain tasks of pathologists. As part of this pursuit, we organized two so-called 'grand challenges' (CAMELYON16 and 17) in which computational researchers were invited to work on a diagnostically relevant problem in histopathology. An important result of CAMELYON16 is that we now have algorithms that are capable of performing one specific task (detection of lymph node metastases) at the same level as trained pathologists.

To support Dr Gallas' project, we may provide glass slides from our challenges as well as expertise in the areas of algorithm development, organization of challenges and use of algorithms in Pathology practice in general. We hope you will be able to advise positively on this project, as I am confident that it will further the field of computational Pathology, which in turn may offer large benefits to the health care system.

Best regards,

Dr Jeroen van der Laak, PhD
Associate Professor in Computational Pathology

Evaluation environment for digital and analog pathology: a platform for validation studies

Brandon D. Gallas
Marios A. Gavrielides
Catherine M. Conway
Adam Ivansky
Tyler C. Keay
Wei-Chung Cheng
Jason Hipp
Stephen M. Hewitt

SPIE.

Evaluation environment for digital and analog pathology: a platform for validation studies

Brandon D. Gallas,[a,*] **Marios A. Gavrielides,**[a] **Catherine M. Conway,**[b] **Adam Ivansky,**[a] **Tyler C. Keay,**[a] **Wei-Chung Cheng,**[a] **Jason Hipp,**[b] **and Stephen M. Hewitt**[b]

[a]FDA/CDRH/OSEL, Division of Imaging, Diagnostics, and Software Reliability, 10903 New Hampshire Avenue, Building 62, Room 3124, Silver Spring, Maryland 20993-0002, United States
[b]National Cancer Institute, National Institutes of Health, Center for Cancer Research, Laboratory of Pathology, 10 Center Drive, MSC 1500, Bethesda, Maryland 20892, United States

Abstract. We present a platform for designing and executing studies that compare pathologists interpreting histopathology of whole slide images (WSIs) on a computer display to pathologists interpreting glass slides on an optical microscope. eeDAP is an evaluation environment for digital and analog pathology. The key element in eeDAP is the registration of the WSI to the glass slide. Registration is accomplished through computer control of the microscope stage and a camera mounted on the microscope that acquires real-time images of the microscope field of view (FOV). Registration allows for the evaluation of the same regions of interest (ROIs) in both domains. This can reduce or eliminate disagreements that arise from pathologists interpreting different areas and focuses on the comparison of image quality. We reduced the pathologist interpretation area from an entire glass slide (10 to 30 mm^2) to small ROIs (<50 μm^2). We also made possible the evaluation of individual cells. We summarize eeDAP's software and hardware and provide calculations and corresponding images of the microscope FOV and the ROIs extracted from the WSIs. The eeDAP software can be downloaded from the Google code website (project: eeDAP) as a MATLAB source or as a precompiled stand-alone license-free application. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JMI.1.3.037501]

1 Introduction

Digital pathology (DP) incorporates the acquisition, management, and interpretation of pathology information generated from a digitized glass slide. DP is enabled by technological advances in whole slide imaging (WSI) systems, also known as virtual microscopy systems, which can digitize whole slides at microscopic resolution in a short period of time. The potential opportunities for DP 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.[1–3] 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 Gu and Ogilvie[4] There are currently numerous commercially available WSI systems as reviewed by Rojo et al. in terms of technical characteristics.[5]

A number of studies (many cited in Refs. 6 and 7) have focused on the validation of WSI systems for primary diagnosis, with findings generally showing high concordance between glass slide and digital slide diagnoses. A common drawback of current validation studies of WSI systems is that they sometimes combine diagnoses from multiple pathology tasks performed on multiple tissue types. Pooling cases can lead to the undersampling of clinical tasks as discussed in the study by Gavrielides et al.[8] It can also dilute differences in reader performance that might be specific to certain tasks. Another issue from current validation studies is that agreement was typically determined by an adjudication panel comparing pathology reports from the WSI and microscope reads head-to-head. Guidelines are sometimes developed for defining major and minor discrepancies, but there is a considerable amount of interpretation and judgment required of the adjudication panel as the pathology reports are collected as real-world, sign-out reports (free text). Additionally, the focus of most validation studies is on primary diagnosis, with minor emphasis on related histopathology features that might be affected by image quality. The quantitative assessment of a pathologist's ability to evaluate histopathology features in WSI compared to the microscope would be useful in identifying possible limitations of DP for specific tasks. Related work includes the study of Velez et al.[9] where discordances in the diagnosis of melanocytic skin lesions were attributed to difficulty in identifying minute details such as inflammatory cells, apoptosis, organisms, and nuclear atypia. Finally, studies focusing on primary diagnosis do not typically account for differences in search patterns or FOV reviewed by observers. The selection of different areas to be assessed by different observers has been identified as a source of interobserver variability.[10]

In this paper, we present an evaluation environment for digital and analog pathology that we refer to as eeDAP. eeDAP is a software and hardware platform for designing and executing digital and analog pathology studies where the digital image

*Address all correspondence to: Brandon D. Gallas, E-mail: brandon.gallas@fda.hhs.gov

is registered to the real-time view on the microscope. This registration allows for the same regions of interest (ROIs) to be evaluated 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 FOV (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 groups of cells in the digital mode and with the microscope. Consequently, paired observations from coregistered FOV 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. At the same time, this approach requires the study designer to narrow the criteria for cases (slides, ROIs, cells) to be included in the study set.

Reader studies utilizing eeDAP can focus on the evaluation of specific histopathology features. Since certain features challenge image quality properties such as color fidelity, focus quality, and depth of field, such reader studies can provide valuable information for the assessment of WSI and its role in clinical practice. The presented framework allows for the formulation of different types of tasks, many of which are currently available and customizable 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. Figure 1 shows the examples of the GUI presentation for two scoring tasks that we have explored: on the left, the reader would be asked to provide a score between 1 and 100 reflecting their confidence that the cell within the reticle is a plasma cell [in hematoxylin and eosin (H&E) stained, formalin-fixed, paraffin-embedded colon tissue], whereas on the right, the reader would provide a score reflecting their confidence that the cell within the reticle is a mitotic figure (in H&E stained, formalin-fixed, paraffin-embedded sarcoma).

In this paper, we outline the key software and hardware elements of eeDAP. First, we discuss the eeDAP software as a publicly available resource and describe software specifications and requirements. We next talk about the tone reproduction curves that characterize eeDAP and the native viewers: the curves showing the lightness in the output image given the transmittance of the input slide. In Sec. 2.3, we summarize the local and global registration methods that are key to pairing ROIs across the digital and microscope domains. In Sec. 2.4, we provide the key hardware specifications that eeDAP requires and then demonstrate the differences in FOVs and image sizes between the two domains: the digital image and the glass slide. These calculations and corresponding representative images help to provide a sense of scale across the digital and analog domains. Finally, we talk about reticles and their important role in narrowing the evaluation area to a small ROI or an individual cell.

2 Methods

In this section, we summarize the key elements of the eeDAP software and hardware. The eeDAP software is made up of three graphical user interfaces (GUIs) written in MATLAB (Mathworks, Natick, Massachusetts).

The first interface establishes the evaluation mode (Digital or MicroRT) and reads in the study input file. The input file contains the file names of the WSIs, hardware specifications, and the list of tasks with corresponding ROI locations that will be interpreted by the pathologist. Each ROI is defined by a location, width, and height in pixel coordinates of the WSI, and all are automatically extracted on the fly from the WSIs named.
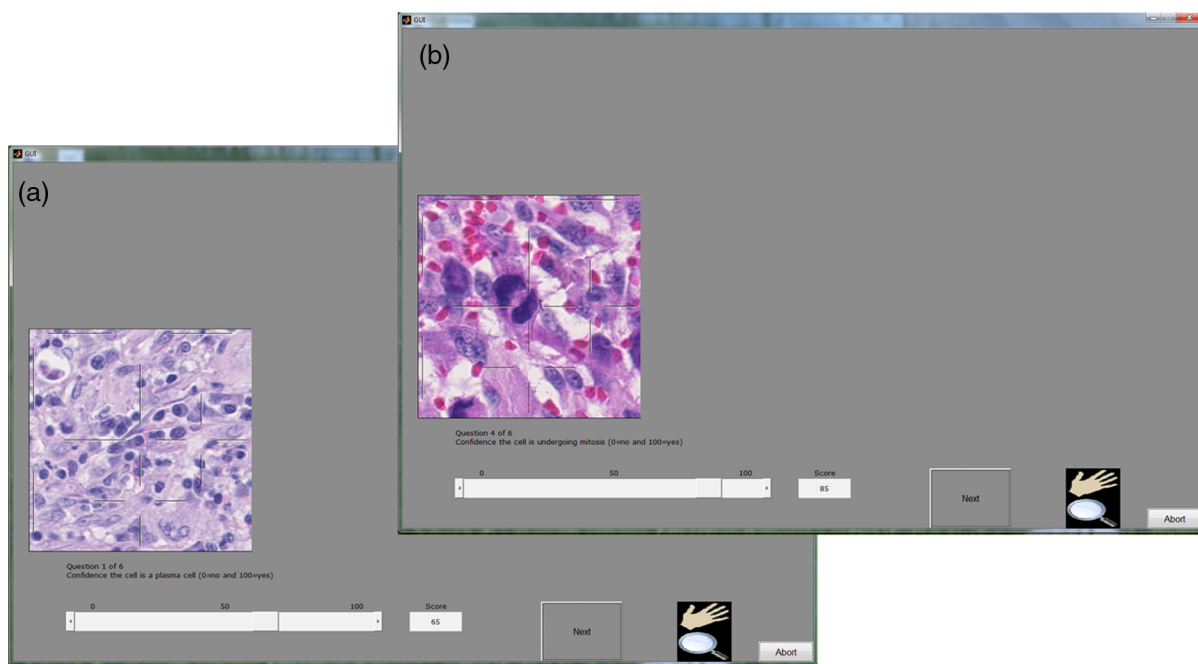
Fig. 1 Here are the two windows, each showing the eeDAP presentation of a slider task: the image on the left is of colon tissue, the image on the right is of sarcoma.
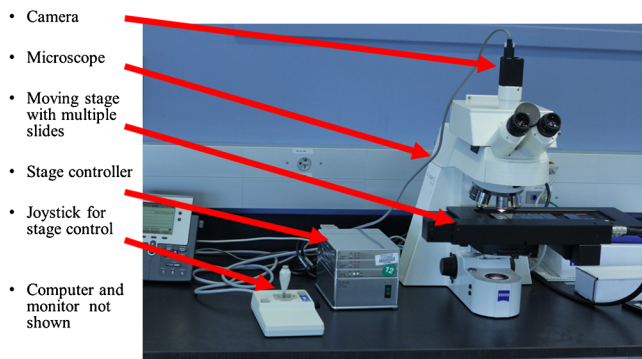
Fig. 2 The evaluation environment for digital and analog pathology (eeDAP) hardware: microscope, camera, computer-controlled stage with joystick, and a computer with monitor (not shown).

There are installation requirements that make the ROI extraction possible from the proprietary WSI file formats. We also discuss a color gamut comparison between eeDAP and a native WSI viewer (a viewer designed by a WSI scanner manufacturer).

The second interface is executed only for studies run in the MicroRT mode. This interface globally registers each WSI to its corresponding glass slide. For each global registration of each WSI, a study administrator must interactively perform three local registrations. The local and global registration methods are described in Sec. 2.3.

The third interface runs the study in accordance to the list of tasks given in the input file. If the study is run in the Digital mode, the pathologist views the ROIs on the computer display in the GUI and enters the evaluations therein. If the study is run in MicroRT *mode*, the pathologist views the ROIs through the microscope (calibrated for Köhler illumination) and is responsible for any focusing in the *z*-plane. Although the pathologist is engaged with the microscope in the MicroRT mode, the study administrator is viewing the ROIs on the computer display in the GUI and enters the evaluations there as dictated by the pathologist. The study administrator also monitors a live camera image of what the pathologist sees through the microscope. This allows the study administrator to confirm and maintain a high level of registration precision in MicroRT mode.

The eeDAP hardware consists of an upright brightfield microscope, a digital camera, a computer-controlled stage with a joystick, a computer monitor, and a computer (see Fig. 2). The microscope requires a port for mounting the camera that allows for simultaneous viewing with the eyepiece. eeDAP currently supports a Ludl controller and compatible *xy* stage, and an IEEE 1394 FireWire camera communicating according to a DCAM interface (RGB, 8 bits-per-channel, minimum width 640, minimum height 480). Setup instructions and example specifications can be found in the user manual.

Below we summarize how these components are used in registration and how the WSI and real microscope image appear to the pathologist. We also identify an important part of the microscope, the reticle. The reticle is housed in the microscope eyepiece. One reticle that we use identifies ROIs in the microscope FOV and another points at individual cells.

2.1 eeDAP Availability and Technical Requirements

The software component of eeDAP is publicly available as MATLAB source code or as a precompiled stand-alone license-free MATLAB application.[11] Running eeDAP source code

requires the MATLAB image acquisition toolbox and the installation of third party software to extract ROIs from WSIs. WSIs are often extremely large (several GB) and are stored as large layered TIFF files embedded in proprietary WSI file formats. eeDAP uses ImageScope, a product of Aperio (a Leica Biosystems Division) to read images scanned with WSI scanners from Aperio (.svs) and other formats, including .ndpi (Hamamatsu). ImageScope contains an ActiveX control named TIFFcomp that allows for the extraction and scaling of ROIs. A consequence of using TIFFcomp is that the MATLAB version must be 32 bits.

The precompiled stand-alone application requires that the MATLAB compiler runtime (MCR) library be installed. It is important that the version of the MCR correspond exactly to that used for the stand-alone application (refer to the user manual).

2.2 Tone Reproduction Curves

Manufacturers of WSI scanners typically provide software for viewing their proprietary file formats. These viewers may include specialized color management functions. In fact, we observed color differences when we viewed .ndpi images with the native Hamamatsu viewer (NDP.view) side-by-side with the Aperio viewer (ImageScope) and MATLAB (with the Aperio ImageScope Active X component TIFFcomp). In an attempt to understand the native viewer and correct for these differences (so that we can show the images as they would be seen in the native viewer), we considered the image adjustments that may have caused them. From these, we observed that the images appeared the same in the three viewers when we adjusted the gamma setting. To confirm our observations, we measured the tone reproduction curves of NDP.view (gamma = 1.8 and gamma = 1.0) and ImageScope (no adjustments made; equivalent to the MATLAB).

Following the work of Cheng et al.,[12] we measured the transmittance of the 42 color patches of a color phantom (film on a glass slide, see Fig. 3). Using an FPGA board, we then retrieved the sRGB values of a Hamamatsu scanned image of the color phantom from the NDP.view with gamma set to 1.8 (default), gamma set to 1.0 (turning off the gamma adjustment), and ImageScope (default, no gamma correction). We then converted the sRGB values to the CIELAB color space and plotted the normalized lightness L^* channel against the normalized transmittance. The results (Fig. 3) supported our visual observations:

- There is good agreement between the tone reproduction curves of NDP.view with gamma = 1.0 and ImageScope.
- The tone reproduction curve of NDP.view with gamma = 1.0 appears to be linearly related to transmittance.
- The tone reproduction curve of NDP.view with gamma = 1.8 appears to be 1/1.8 gamma transformation of transmittance.
- The default images displayed by NDP.view and ImageScope (and MATLAB by equivalence) differ only in the gamma setting.

2.3 Registration

eeDAP uses registration to link the stage (glass slide) coordinates to the WSI coordinates. eeDAP has two levels of registration: global and local. The global registration is equivalent to
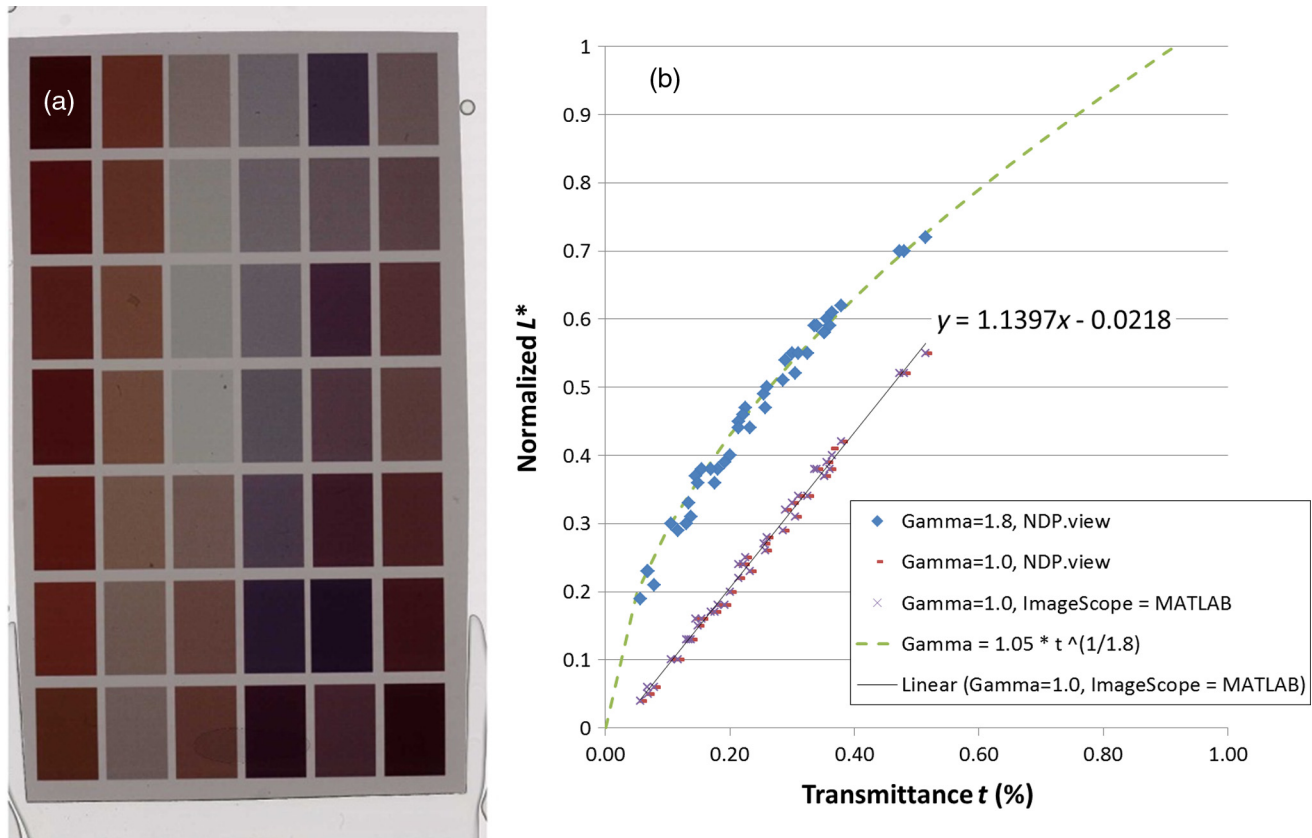
Fig. 3 (a) Hamamatsu scanned image of a color phantom (film on a glass slide). (b) The transmittance of the 42 color patches plotted against the normalized lightness $L*$ in the CIELAB color space (derived from the average sRGB values in a patch).

finding the transformation between the stage and WSI coordinates. The global registration requires three anchors, three pairs of stage-WSI registered coordinates. Each anchor is generated by a local registration: a (x, y) stage coordinate and a (x, y) WSI coordinate that correspond to the same specimen location.

eeDAP conducts two levels of global registration: low and high resolutions. Low resolution corresponds to microscope magnifications such as 2×, 5×, and 10×; the entire WSI image is scaled to fit in the GUI. High resolution registration corresponds to the microscope magnifications such as 20× and 40×; the low-resolution registration results are used to limit the amount of the WSI shown in the GUI, sequentially zooming in on the location of the low-resolution anchors.

eeDAP uses local registration for two purposes. The first purpose is to support global registration as discussed. The second purpose is to maintain a high level of registration precision throughout data collection. During our pilot studies, we observed that the precision of the global registration was deteriorating as the stage moved throughout the study. Therefore, we implemented a button that could be pressed during data collection that could register the current microscope view to the current task-specific ROI. The current level of precision appears to allow for the reliable evaluation of individual cells. Technical details of local and global registrations are provided below.

2.3.1 Local registration

A local registration is accomplished by taking a snapshot of the microscope FOV with the microscope-mounted camera and by

finding a search region containing the corresponding location in the WSI (see Fig. 4). The search region is identified by the study administrator and avoids searching the entire (very large) image. A local registration yields a (x, y) coordinate in the WSI and a (x, y) coordinate on the microscope stage that identify the same location.

The camera image contains some amount of specimen on the glass slide and is labeled by the (x, y) coordinate of the current stage position. See, for example, the "Preview with cross hairs" window labeled "Camera image" depicted in Fig. 4. The camera image has three channels (RGB) and must be at least 640×480. The physical size of a (square) camera pixel is given by the manufacturer specifications. This size divided by any magnification by the microscope (objective × camera adapter) determines the camera's spatial sampling period in units of the specimen.

We extract a patch of the WSI image (RGB) that is larger than and contains the same content as captured by the camera. See, for example, the image labeled "WSI image" depicted in Fig. 4. The WSI's spatial sampling period (often referred to as the WSI resolution) is given by the manufacturer specifications in units of the specimen and is often recorded in the WSI image.

An ROI extracted from a WSI image can be rescaled (interpolated) to have the same sampling period as the camera image using the ratio of the sampling periods. In other words, the number of pixels before and after rescaling is determined by
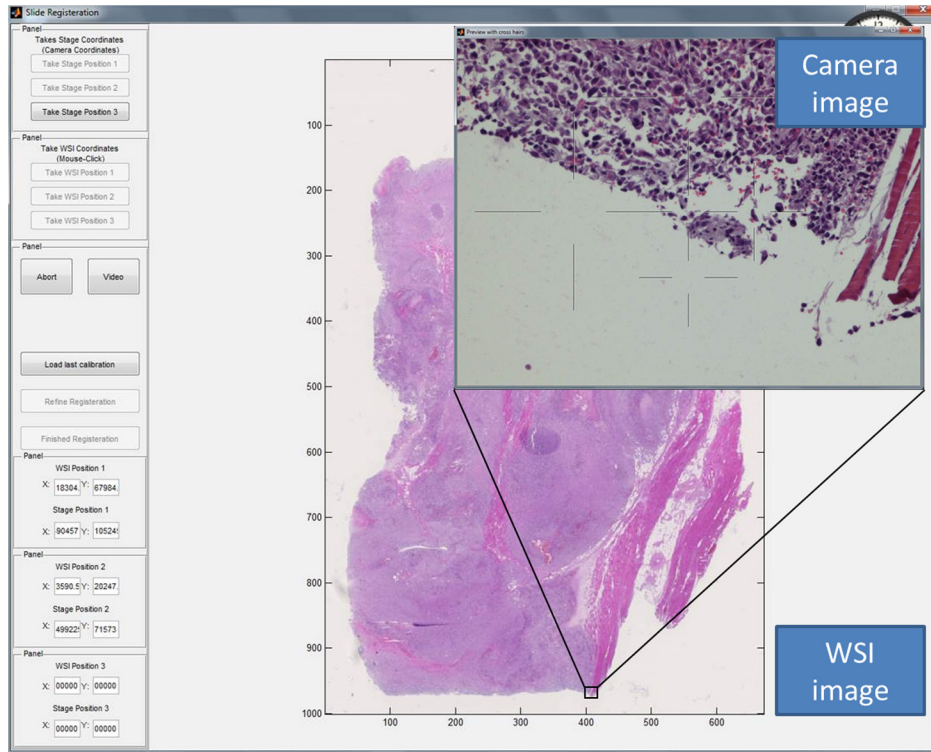
Fig. 4 Screen shot of the registration interface including the real-time microscope field of view (FOV) as seen with the mounted camera ("Camera image").

[corresponding number of camera pixels]

$$= \frac{\text{WSI sampling period}}{\text{camera sampling period}} [\text{number of WSI pixels}].$$

Given the camera image $c(x, y)$ and the WSI image $d(x, y)$ at the same scale, we perform normalized cross-correlation to find the Δx, Δy shift that best registers the two images. In other words, we find the Δx, Δy that maximize the following sum:

$$\frac{1}{n} \sum_{x,y} \frac{(c(x, y) - \bar{c})(d(x + \Delta x, y + \Delta y) - \bar{d})}{\sigma_c \sigma_d},$$

where the sum is over the n pixels in the camera image, (x, y) indexes the pixels in the image, Δx, Δy is the shift in pixels, and $\bar{c}$, σ_c and $\bar{d}$, σ_d are the average and standard deviation of the elements of $c(x, y)$ and $d(x, y)$ considered in the sum.

2.3.2 Global registration

Global registration is done for each WSI in the input file and corresponding glass slide on the microscope stage. Each global registration is built on three local registrations. The three local registrations yield three pairs of coordinates that define the transformation (the change of basis) between the coordinate system of the WSI (image pixels) and the coordinate system of the stage (stage pixels).

Let the three pairs of coordinates be given by $(x_i^{\text{WSI}}, y_i^{\text{WSI}})$, $(x_i^{\text{stage}}, y_i^{\text{stage}})$ for $i = 1, 2, 3$. Given these pairs, we define the two coordinate systems with the following matrices:

$$M_{\text{WSI}} = \begin{bmatrix} x_2^{\text{WSI}} - x_1^{\text{WSI}}, & x_3^{\text{WSI}} - x_1^{\text{WSI}} \\ y_2^{\text{WSI}} - y_1^{\text{WSI}}, & y_3^{\text{WSI}} - y_1^{\text{WSI}} \end{bmatrix},$$

$$M_{\text{stage}} = \begin{bmatrix} x_2^{\text{stage}} - x_1^{\text{stage}}, & x_3^{\text{stage}} - x_1^{\text{stage}} \\ y_2^{\text{stage}} - y_1^{\text{stage}}, & y_3^{\text{stage}} - y_1^{\text{stage}} \end{bmatrix}.$$

Then given a new location in the WSI coordinate system $[x_{\text{new}}^{\text{WSI}}, y_{\text{new}}^{\text{WSI}}]^t$, we can determine the corresponding location in the stage coordinate system with the following transformation:

$$\begin{bmatrix} x_{\text{new}}^{\text{stage}} \\ y_{\text{new}}^{\text{stage}} \end{bmatrix} = M_{\text{stage}} M_{\text{WSI}}^{-1} \left(\begin{bmatrix} x_{\text{new}}^{\text{WSI}} \\ y_{\text{new}}^{\text{WSI}} \end{bmatrix} - \begin{bmatrix} x_1^{\text{WSI}} \\ y_1^{\text{WSI}} \end{bmatrix} \right) + \begin{bmatrix} x_1^{\text{stage}} \\ y_1^{\text{stage}} \end{bmatrix}.$$

In words, we first shift the new point according to the origin in the WSI coordinate system $(x_1^{\text{WSI}}, y_1^{\text{WSI}})$. Next, we map the point from the WSI coordinate system to the standard one with M_{WSI}^{-1} and then map it to the stage coordinate system with M_{stage}. Finally, we shift the point according to the origin in the stage coordinate system $(x_1^{\text{stage}}, y_1^{\text{stage}})$. Consequently, the location of each ROI for each task given in the input file can be accessed in the WSI coordinate system or the stage coordinate system.

The study administrator determines each local registration by navigating the microscope with the joystick to an appropriate anchor, taking the camera image, and then approximately identifying the corresponding anchor in the WSI. An appropriate anchor is one that can be recognized in the WSI image and is surrounded by one or more salient features. Salient features increase the likelihood of a successful registration; repetitive features and homogeneous regions do not. Additionally, global registration is better when the set of three anchors are widely separated; encompassing the entirety of the tissue is best. The most challenging aspect in finding the appropriate anchors is navigating the microscope stage

with the joystick, focusing on the specimen, and determining the corresponding location in the WSI image.

In Fig. 4, we see the "Camera image" and the "WSI image." The study administrator has clicked on the WSI image to indicate where in the WSI to search for the camera image. A patch of the WSI image is extracted from the WSI at the full scanning resolution, the patch is scaled to the resolution of the camera, and a local registration produces the shift that identifies the corresponding WSI location to pair with the current stage location.

2.4 Comparing FOV and Image Sizes

In the following, we provide the key hardware specifications that eeDAP requires and demonstrates the calculation of different FOVs and image sizes. These calculations provide the relationships regarding scale across the digital and analog domains.

2.4.1 Microscope FOV

An important parameter of an optical microscope is the field number (FN); it is the diameter of the view field in millimeters at the intermediate image plane, which is located in the eyepiece. The FN is a function of the entire light path of the microscope starting with the glass slide, through the objective, and ending at the intermediate image plane in the eyepiece; the FN is often inscribed on the eyepiece. To get the FOV in units of the specimen being viewed, we divide the FN by the objective magnification. We currently have an Olympus BX43 microscope (FN = 22 mm) and a Zeiss Axioplan2 Imaging microscope (FN = 23 mm). At 40× magnification due to the objective, the FOV covered in the specimen plane is given by

- Olympus FOV at 40×

    - diameter = 22/40 = 0.550 mm,
    - area = 0.2376 mm,

- Zeiss FOV at 40×

    - diameter = 23/40 = 0.575 mm,
    - area = 0.2597 mm.

The FN can also be used to determine the perceived size of the microscope image at an effective viewing distance of 25 cm. The 25 cm effective viewing distance is a design convention[13] that is not well documented or well known. The perceived size is then simply the FN times the eyepiece magnification. Since the eyepieces on both microscopes above have 10× magnification, the perceived diameters of the intermediate images are 22 cm (Olympus) and 23 cm (Zeiss) at the effective viewing distance of 25. This corresponds to a visual angle (subtended angle of object at the eye) equal to $2 \times \arctan(23/(2*25)) \approx 50$ deg. In Fig. 5, we show what the microscope FOV looks like for the sarcoma slide scaled to fit the page.
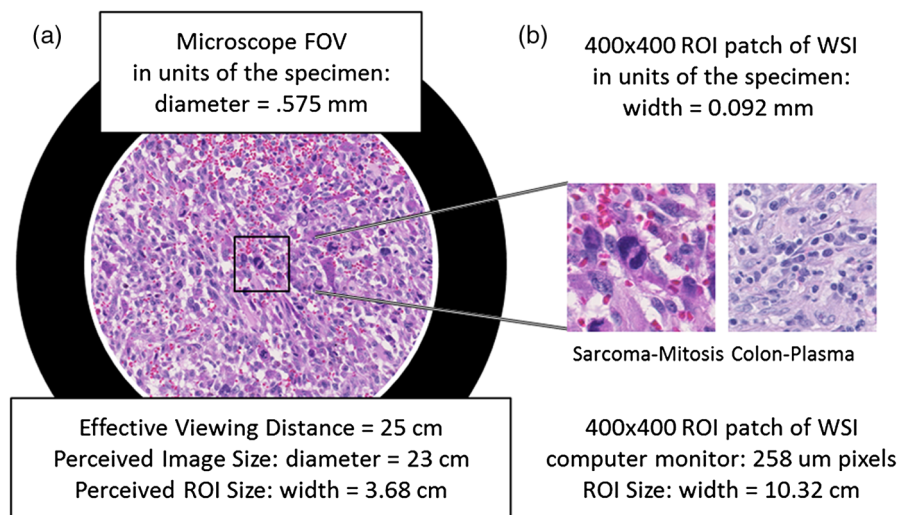
2.4.2 Size of scanner images

We have access to two WSI scanners: a Hamamatsu Nanozoomer 2.0HT and an Aperio CS. They both operate at 20× and 40× magnification equivalent settings with similar sampling periods:

- 0.4558 μm at 20× and 0.2279 μm at 40× (Hamamatsu);

- 0.5000 μm at 20× and 0.2500 μm at 40× (Aperio).

The 40× Hamamatsu scanned images we have been using for pilot studies have $123,008 \times 82,688$ pixels (10 GB) and $39,680 \times 51,200$ pixels (2 GB). By multiplying the number of pixels by the sampling period, we get the size of the images in units of the specimen on the glass slide. These images correspond to the image areas of 28.0 mm × 18.8 mm and 9.0 mm × 11.7 mm. We have been extracting 400 × 400 ROI patches that show 0.092 mm × 0.092 mm patches of the specimen (area ≈ 0.0084 mm^2) for our most recent pilot study, which is 3.2% of the microscope FOV.

The size of a patch seen by a pathologist depends on the computer monitor pixel pitch (distance between pixels). For a computer monitor with a 258-μm pixel pitch, the display size of a 400 × 400 patch is 10.32 cm × 10.32 cm (area ≈ 106 cm^2). If we assume a viewing distance of 25 cm from the computer monitors (to match the effective viewing distance in the microscope), we can compare the image size of the ROI on the computer

(a) Microscope FOV in units of the specimen: diameter = .575 mm

Effective Viewing Distance = 25 cm
Perceived Image Size: diameter = 23 cm
Perceived ROI Size: width = 3.68 cm

(b) 400x400 ROI patch of WSI in units of the specimen: width = 0.092 mm

Sarcoma-Mitosis Colon-Plasma

400x400 ROI patch of WSI computer monitor: 258 um pixels ROI Size: width = 10.32 cm

Fig. 5 The two images in this figure depict the relative sizes of the microscope image as seen through the eyepiece at 40× (a) and a 400 × 400 regions of interest patch from a whole slide image as seen on a computer monitor with 258-μm pixels at a viewing distance of 25 cm (b).
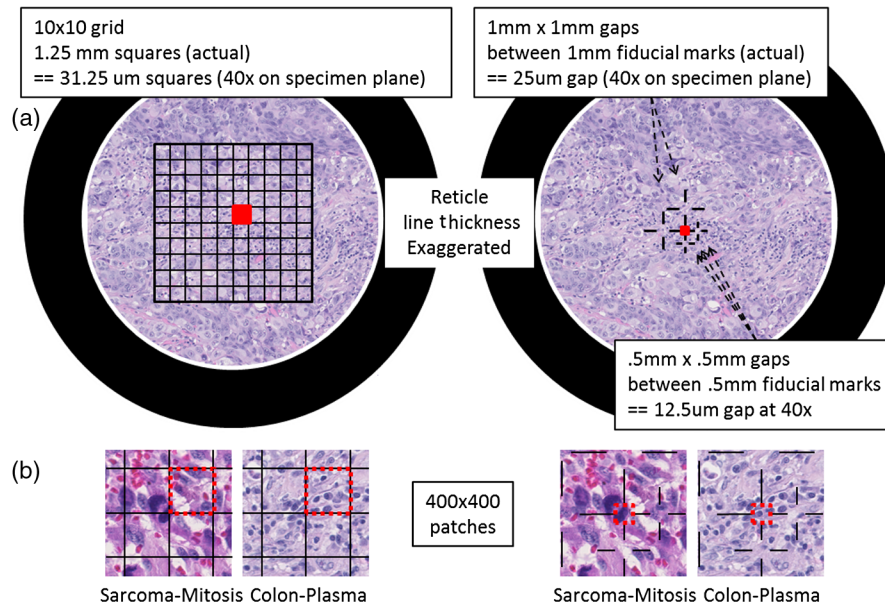
Fig. 6 Reticles (line thicknesses exaggerated) as seen through simulated microscope FOV (a) and 400×400 patches as they appear in eeDAP (b). The red squares in the simulated microscope FOVs and the red-dash boundary squares in the 400×400 patches indicate the "evaluation" areas. In the 400×400 sarcoma patch on the left, a majority of the obvious mitotic figure falls in the grid square to the left of the evaluation square. In the 400×400 sarcoma patch on the right, the central cross hairs point to the obvious mitotic figure. In the 400×400 colon patch on the left, there are several plasma-cell candidates in the "evaluation" square. In the 400×400 colon patch on the right, the cross hairs point to a single plasma-cell candidate to be evaluated.

monitor to the microscope perceived image size. Figure 5 shows the relative sizes of the two views side by side, demonstrating the apparent magnification of the specimen area in the displayed 400×400 patches.

2.4.3 *Size of camera images*

We currently have a Point Grey Flea2 color camera (FL2G-50S5C-C) that has a default output format of 1024×768 with 6.9-μm pixels. This format corresponds to 2×2 binning of a camera with a native pixel size of 3.45 μm. At 20× magnification (40× objective times 0.5× camera adapter), the spatial sampling period in units of the specimen plane corresponds to 0.345 μm (6.9/20) and the camera FOV is 0.353 mm × 0.265 mm (area = 0.0234 mm^2), which is about 36% of the microscope FOV.

2.5 *Reticles*

Reticles are pieces of glass that are inserted at the intermediate image plane in the eyepiece. They contain fine lines and grids that appear superimposed on the specimen. Reticles help to measure features or help to locate objects. The current version of eeDAP uses them to narrow tasks to very small regions and individual cells, allowing for an expansion of capabilities. In Fig. 6, we depict reticles as seen through the microscope (line thickness exaggerated) and as they appear in eeDAP (400×400 patches). These reticles are described below and were studied in two feasibility reader studies for their functionality.

In the first feasibility study, we used a reticle containing a 10×10 grid with squares that are 1.25 mm on a side (Klarmann Rulings: KR-429). At 40×, these squares are 31.25 μm on a side in the specimen plane. When running in Digital mode, eeDAP

digitally creates a reticle mask to create the same effect as the real reticle in the microscope. The instructions for this study were to score the reticle square that was immediately above and to the right of the center cross (red squares in Fig. 6). Identifying the center cross in the 10×10 grid in MicroRT mode is challenging; it is accomplished by rotating the eyepiece as the center cross remains fixed. Additionally, the instructions to score a square were to score the cell that was most likely the target (mitotic figure or plasma cell as shown in Fig. 1), considering cells with at least half their nuclei in the square.

In the second (similar) feasibility study, we used a custom reticle that has fiducials that point to gaps (Klarmann Rulings: KR-32536). Two gaps are 1 mm × 1 mm and three gaps are 0.5 mm × 0.5 mm. At 40×, these gaps are 25.0 μm and 12.5 μm on a side. The instructions for this study were much more direct: score the cell at the center of the center fiducials (red squares in Fig. 6).

3 Results and Discussion

We have been using pilot studies to identify weaknesses and future improvements needed for eeDAP and the general study design. The main weakness that we identified was that the registration precision throughout data collection was not good enough: pathologists were not evaluating identical ROIs. We have addressed this in the current generation of eeDAP by incorporating the ability to do a local registration for every ROI during data collection. We have also created a custom reticle that allows us to point at individual cells. This reduces ambiguity and disagreements due to evaluations based on multiple different cells within an ROI.

We also observed that the .ndpi WSI images appeared darker when viewing with eeDAP (and ImageScope) compared to

viewing with the native viewer, NDP.view. Through observation and subsequent measurement, we determined that the difference was a simple gamma adjustment, and we implemented a color look-up table to make this and any other color adjustment possible with eeDAP.

Our pilot studies emphasized the need for reader training. We found that pathologists needed to develop a level of comfort in scoring individual candidate cells, as this is not a part of a pathologist's typical clinical work flow. This is especially true when we asked for a 101-point confidence rating instead of a yes–no decision. Consequently, we are focusing our efforts to creating training on the cell types and scorings. Training on cell types may include Power Point slides that contain verbal descriptions of typical features and sample images. Training may also include eeDAP training modules: the training modules may elicit scores of the typical features as well as the overall score and then provide feedback in the form of the scores of experts.

As we move beyond pilot studies to pivotal studies, we need to investigate and establish several methods and protocols to reduce the variability between the pilot study and the pivotal study, to reduce variability during a pivotal study, and to allow for a study to be replicated as closely as possible. Methods and protocols are needed on the following issues:

- Computer monitor QA/QC and calibration, including color

  - It is understood in radiology that poor-quality displays can result in misdiagnosis, eye-strain, and fatigue.[14] As such, it is common in the practice and evaluation of radiology to control, characterize, and document the display conditions. This culture has led to the creation of standards that treat displays.[15] This issue is not yet fully enabled and appreciated within the culture of DP practice or evaluation. Study reports do very little to describe the display characteristics and calibration, with recent work being the exception.[8] However some groups, including the International Color Consortium, are filling the void and addressing the challenging issue of display and color calibration.[16–18]

- Slide preparation.

  - It is well known that there is significant variability in tissue appearances based on processing, sectioning, and staining differences and this variability leads to variability in diagnosis.[19] Protocols for slide preparation are a part of standard lab practice and are changing with increased automation, driving standards in this space.[20]

- Tissue inclusion/exclusion criteria, including methods to objectively identify candidate cells for the evaluation task.

  - Identifying inclusion/exclusion criteria for study patients (or in the current context, their tissue) is needed to convey the spectrum of the tissue being used, and thus the trial's generalizability and relevance.[21,22] Given the tissue, when the task is to evaluate individual cells, it is important to not bias the selection process. For our work, we intend to first

identify the entire spectrum of presentations, not just presentations that are easy in one modality or another (as might result from pathologist identified candidates). Once the entire spectrum of presentations is identified, there may be reasons to subsample within to stress the imaging system evaluation and comparison. Future work may include the incorporation of algorithms for the automated identification of candidate cells to be classified or histopathological features to be evaluated. Such algorithms may be less biased and more objective in creating the study sets.

Finally, a coherent analysis method is needed that does not require a gold standard, since one is typically not available for the tasks being considered. To address this need, we are investigating agreement measures, such as concordance, that compare pathologist performance with WSI to conventional optical microscopy. The goal is to develop methods and tools for multi-reader, multicase analysis of agreement measures, similar to the methods and tools for the area under the ROC curve[23] and the rate of agreement.[24]

4 Conclusions

In this paper, we presented the key software and hardware elements of eeDAP, a framework that allows for the registration and display of corresponding tissue regions between the glass slide and the WSI. The goal was to remove search as a source of observer variability that might dilute differences between modalities. The software part of eeDAP can be downloaded from the google code website (project: eeDAP) as a MATLAB source or as a precompiled stand-alone license-free application.[11] This software can be paired with the required hardware (microscope, automated stage, and camera) and used to design and execute reader studies to compare DP to traditional optical microscopy.

References

1. L. Pantanowitz et al., "Review of the current state of whole slide imaging in pathology," *J. Pathol. Inf.* **2**, 36 (2011).
2. R. S. Weinstein et al., "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," *Hum. Pathol.* **40**(8), 1057–1069 (2009).
3. S. Al-Janabi, A. Huisman, and P. J. Van Diest, "Digital pathology: current status and future perspectives," *Histopathology* **61**(1), 1–9 (2012).
4. J. Gu and R. W. Ogilvie, Eds., *Virtual Microscopy and Virtual Slides in Teaching, Diagnosis, and Research*, CRC Press, Boca Raton, Florida (2005).
5. M. G. Rojo et al., "Critical comparison of 31 commercially available digital slide systems in pathology," *Int. J. Surg. Pathol.* **14**(4), 285–305 (2006).
6. L. Pantanowitz et al., "Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American

Pathologists Pathology and Laboratory Quality Center," *Arch. Pathol. Lab. Med.* **137**(12), 1710–1722 (2013).

7. A. R. Jara-Lazaro et al., "Digital pathology: exploring its applications in diagnostic surgical pathology practice," *Pathology* **42**(6), 512–518 (2010).

8. M. A. Gavrielides et al., "Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers," *Anal. Cell. Pathol.*, Epub ahead of print (2014).

9. N. Velez, D. Jukic, and J. Ho, "Evaluation of 2 whole-slide imaging applications in dermatopathology," *Hum. Pathol.* **39**(9), 1341–1349 (2008).

10. P. J. van Diest et al., "A scoring system for immunohistochemical staining: consensus report of the task force for basic research of the EORTC-GCCG. European Organization for Research and Treatment of Cancer-Gynaecological Cancer Cooperative Group," *J. Clin. Pathol.* **50**(10), 801–804 (1997).

11. B. D. Gallas, *eeDAP v2.0: Evaluation Environment for Digital and Analog Pathology (code.google.com/p/eedap)*, Division of Imaging and Applied Mathematics, CDRH, FDA, Silver Spring, Maryland (2013).

12. W.-C. Cheng et al., "Assessing color reproducibility of whole-slide imaging scanners," *Proc. SPIE* **8676**, 86760O (2013).

13. H. G. Kapitza *Microscopy from the Very Beginning*, 2nd ed., Carl Zeiss, Jena, Germany (1997).

14. D. Hirschorn, E. Krupinski, and M. Flynn, "Displays," in *IT Reference Guide for the Practicing Radiologist* (2013).

15. J. T. Norweck et al., "ACR-AAPM-SIIM technical standard for electronic practice of medical imaging," *J. Digit Imaging* **26**(1), 38–52 (2013).

16. ICC, "ICC White Paper 7: The Role of ICC Profiles in a Colour Reproduction System, International Color Consortium," Reston, VA (2004), http://www.color.org/ICC_white_paper_7_role_of_ICC_profiles .pdf (27 10 2014); P. Green, Ed., *Color Management: Understanding and Using ICC Profiles*, Wiley, New York, NY (2010).

17. E. A. Krupinski et al., "Observer performance using virtual pathology slides: impact of LCD color reproduction accuracy," *J. Digit Imaging* **25**(6), 738–743 (2012).

18. A. Badano et al., "Consistency and standardization of color in medical imaging: a consensus report," *J. Digit Imaging*, Epub ahead of print (2014).

19. K. H. Allison et al., "Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel," *Histopathology* **65**(2), 240–251 (2014).

20. *I/LA28-A2: Quality Assurance for Design Control and Implementation of Immunohistochemistry Assays; Approved Guideline*, 2nd ed., Clinical and Laboratory Standards Institute, Wayne, Pennsylvania (2011).

21. P. M. Bossuyt et al., "Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative," *Am. J. Clin. Pathol.* **119**(1), 18–22 (2003).

22. K. F. Schulz et al., "Consort 2010 statement: updated guidelines for reporting parallel group randomised trials," *BMJ* **340**, c332 (2010).

23. B. D. Gallas et al., "A framework for random-effects ROC analysis: biases with the bootstrap and other variance estimators," *Commun. Stat. A* **38**(15), 2586–2603 (2009).

24. W. Chen and A. Wunderlich, *iMRMC_Binary v1.0_Beta1: A Matlab Software Package for Simulating, Analyzing, and Sizing a Multi-Reader Multi-Case Reader Study with Binary Assessments*, Division of Imaging and Applied Mathematics, CDRH, FDA, Silver Spring, Maryland (2014).

Brandon D. Gallas provides mathematical, statistical, and modeling expertise to the evaluation of medical imaging devices at the FDA. His main areas of contribution are in the design and statistical analysis of reader studies, image quality, computer-aided diagnosis, and imaging physics. Before working at the FDA, he was in Dr. Harrison Barrett's radiology research group at the University of Arizona earning his PhD degree in applied mathematics from the Graduate Interdisciplinary Program.

Marios A. Gavrielides received his PhD degree in biomedical engineering from Duke University in 2001. From 2002 to 2004, he was a postdoctoral researcher at the Aristotle University of Thessaloniki, Greece. Since 2004, he has been a NIBIB/CDRH fellow and then a staff scientist with the Division of Imaging, Diagnostics, and Software Reliability, OSEL/CDRH at the Food and Drug Administration. His research interests include the development and assessment of quantitative imaging methodologies.

Catherine M. Conway provides expertise in digital pathology evaluation, validation, and study design. Her main areas of contribution are in histology study design and prep, digital pathology evaluation design, and computer-aided diagnosis. Before working at the NIH, she worked in industry in the area of digital pathology. She earned a PhD degree from Dublin City University in the subject area of development of image analysis software for the evaluation of IHC across digital slides.

Adam Ivansky is a PhD student at the University of Rochester. His main research interests include classical microscopy, nonlinear microscopy, and image processing. He is a member of a research group under Dr. Wayne H. Knox that focuses on development of non-invasive refractive laser eye surgery. His main contribution to the eeDAP project was coding of a portion of the software.

Tyler C. Keay received his BS degree in bioengineering from University of Maryland's Fischell School of Engineering in 2011. After graduating, he was an ORISE fellow at the Food and Drug Administration, where he began his work in digital pathology under the guidance of Marios Gavrielides, Wei-Chung Cheng, and Brandon Gallas. Since 2013, he has been a clinical affairs scientist at Omnyx, LLC. His research interests include quantitative imaging analysis and color science.

Wei-Chung Cheng is a research scientist at the FDA. Since 2009, he has been developing assessment methods for medical imaging devices in various modalities including digital pathology, endoscopy, and radiology. He received his PhD degree in electrical engineering from the University of Southern California. Before joining the FDA, he was an assistant professor in the Department of Photonics and Display Institute, National Chiao-Tung University, Taiwan. His research interests include color science, applied human vision, and display technology.

Jason Hipp served as a Rabson fellow in the Laboratory of Pathology at the National Cancer Institute (NCI), National Institutes of Health (NIH). He earned his MD/PhD degree from Wake Forest University School of Medicine. He then completed his residency in anatomic pathology at the NCI/NIH, and subsequently completed a 2-year pathology informatics fellowship at the University of Michigan, School of Medicine.

Stephen M. Hewitt is a clinical investigator in the Laboratory of Pathology, National Cancer Institute. His research interests include tissue-based biomarkers and whole-slide imaging. He is a treasurer of the Histochemical Society, Program Committee Chair for the Association for Pathology Informatics and a consultant and collaborator to the Center for Devices and Radiological Health, Food and Drug Administration. He has coauthored over 200 articles and serves on the editorial board of five peer-reviewed journals.