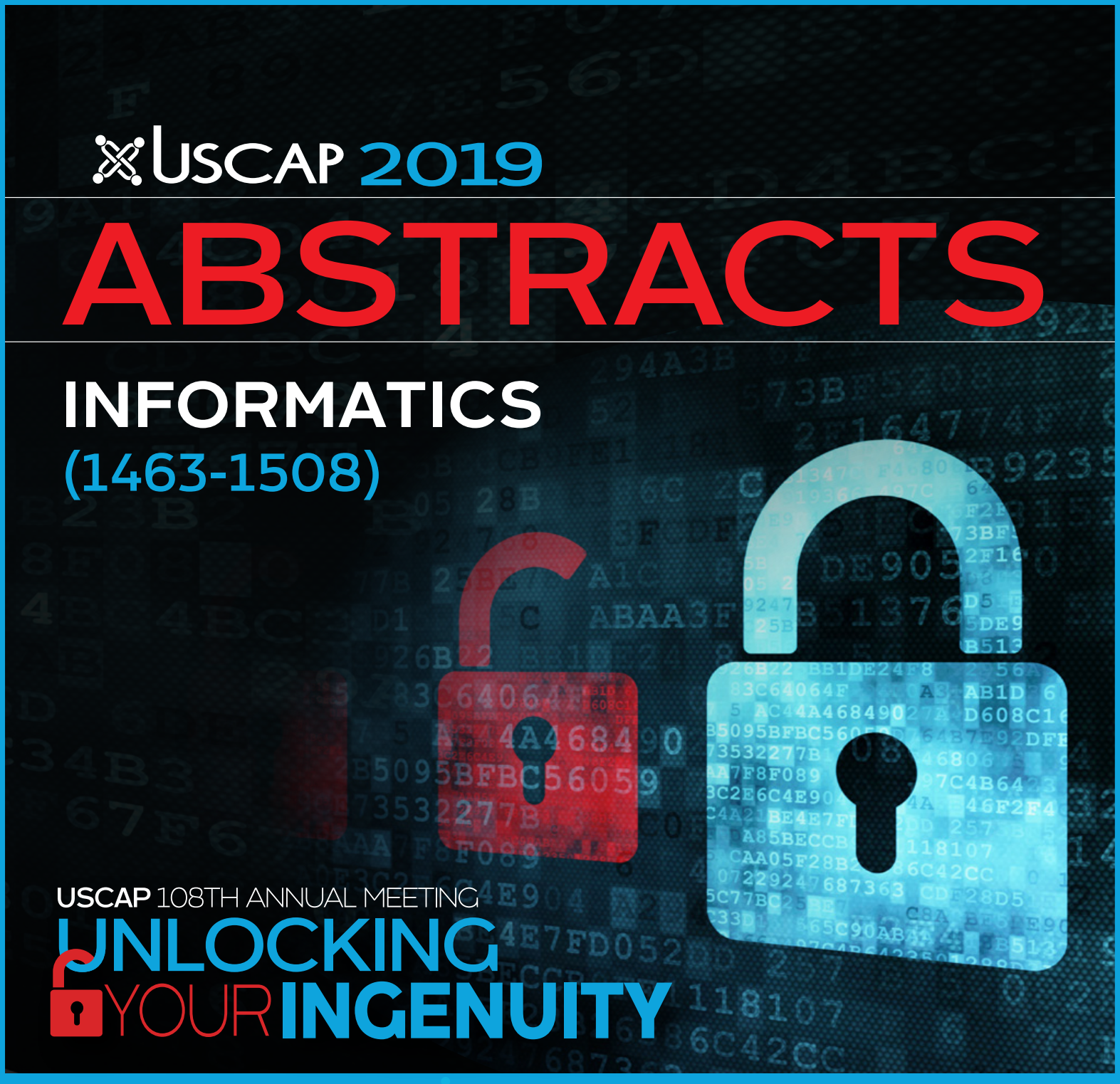
VOLUME 32 | SUPPLEMENT 2 | MARCH 2019

MODERN PATHOLOGY

USCAP 2019

ABSTRACTS

INFORMATICS
(1463-1508)



USCAP 108TH ANNUAL MEETING

UNLOCKING YOUR INGENUITY

MARCH 16-21, 2019

National Harbor, Maryland
Gaylord National Resort & Convention Center

Published by
SPRINGER NATURE
www.ModernPathology.org

USCAP Creating a Better Pathologist
AN OFFICIAL JOURNAL OF THE UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY

EDUCATION COMMITTEE

Jason L. Hornick, Chair
Rhonda K. Yantiss, Chair, Abstract Review Board and Assignment Committee
Laura W. Lamps, Chair, CME Subcommittee
Steven D. Billings, Interactive Microscopy Subcommittee
Shree G. Sharma, Informatics Subcommittee
Raja R. Seethala, Short Course Coordinator
Ilan Weinreb, Subcommittee for Unique Live Course Offerings
David B. Kaminsky (Ex-Officio)
Aleodor (Doru) Andea
Zubair Baloch
Olca Basturk
Gregory R. Bean, Pathologist-in-Training
Daniel J. Brat
Ashley M. Cimino-Mathews
James R. Cook
Sarah M. Dry
William C. Faquin
Carol F. Farver
Yuri Fedoriw
Meera R. Hameed
Michelle S. Hirsch
Lakshmi Priya Kunju
Anna Marie Mulligan
Rish Pai
Vinita Parkash
Anil Parwani
Deepa Patil
Kwun Wah Wen, Pathologist-in-Training

ABSTRACT REVIEW BOARD

Benjamin Adam
Michelle Afkhami
Narasimhan (Narsi) Agaram
Rouba Ali-Fehmi
Ghassan Allo
Isabel Alvarado-Cabrero
Christina Arnold
Rohit Bhargava
Justin Bishop
Jennifer Boland
Elena Brachtel
Marilyn Bui
Shelley Caltharp
Joanna Chan
Jennifer Chapman
Hui Chen
Yingbei Chen
Benjamin Chen
Rebecca Chernock
Beth Clark
James Conner
Alejandro Contreras
Claudiu Cotta
Timothy D'Alfonso
Farbod Darvishian
Jessica Davis
Heather Dawson
Elizabeth Demicco
Suzanne Dintzis
Michelle Downes
Daniel Dye
Andrew Evans
Michael Feely
Dennis Firchau
Larissa Furtado
Anthony Gill
Ryan Gill
Paula Ginter
Tamara Giorgadze
Raul Gonzalez
Purva Gopal
Anuradha Gopalan
Jennifer Gordetsky
Rondell Graham
Alejandro Gru
Nilesh Gupta
Mamta Gupta
Krisztina Hanley
Douglas Hartman
Yael Heher
Walter Henricks
John Higgins
Mai Hoang
Mojgan Hosseini
Aaron Huber
Peter Illei
Doina Ivan
Wei Jiang
Vickie Jo
Kirk Jones
Neerja Kambham
Chiah Sui (Sunny) Kao
Dipti Karamchandani
Darcy Kerr
Ashraf Khan
Rebecca King
Michael Kluk
Kristine Konopka
Gregor Krings
Asangi Kumarapeli
Alvaro Laga
Cheng-Han Lee
Zaibo Li
Haiyan Liu
Xiuli Liu
Yan-Chun Liu
Tamara Lotan
Anthony Magliocco
Kruti Maniar
Jonathan Marotti
Emily Mason
Jerri McLemore
Bruce McManus
David Meredith
Anne Mills
Neda Moatamed
Sara Monaco
Atis Muehlenbachs
Bita Naini
Dianna Ng
Tony Ng
Ericka Olgaard
Jacqueline Parai
Yan Peng
David Pisapia
Alexandros Polydorides
Sonam Prakash
Manju Prasad
Peter Pytel
Joseph Rabban
Stanley Radio
Emad Rakha
Preetha Ramalingam
Priya Rao
Robyn Reed
Michelle Reid
Natasha Rekhtman
Michael Rivera
Michael Roh
Andres Roma
Avi Rosenberg
Esther (Diana) Rossi
Peter Sadow
Safia Salaria
Steven Salvatore
Souzan Sanati
Sandro Santagata
Anjali Saqi
Frank Schneider
Jeanne Shen
Jiaqi Shi
Wun-Ju Shieh
Gabriel Sica
Deepika Sirohi
Kalliopi Siziopikou
Lauren Smith
Sara Szabo
Julie Teruya-Feldstein
Gaetano Thiene
Khin Thway
Rashmi Tondon
Jose Torrealba
Evi Vakiani
Christopher VandenBussche
Sonal Varma
Endi Wang
Christopher Weber
Olga Weinberg
Sara Wobker
Mina Xu
Shaofeng Yan
Anjana Yeldandi
Akihiko Yoshida
Gloria Young
Minghao Zhong
Yaolin Zhou
Hongfa Zhu
Debra Zynger

1463 Deploying Home Grown Applications of Deep Learning Utilizing Widely Available Hardware and Software: Interactive Demo

Alaa Alsadi[1], Lucy Fu[2], Nasma Majeed[1], Yash Dharmamer[1], Uwadia Edomwonyi[1], Charisse Liz Baste[1], John Groth[3], Manmeet Singh[4], Tushar Patel[1]
[1]University of Illinois at Chicago, Chicago, IL, [2]Northwestern University Feinberg School of Medicine, Chicago, IL, [3]University of Illinois Hospital & Health Sciences, Libertyville, IL, [4]Burr Ridge, IL

Disclosures: Alaa Alsadi: *Major Shareholder*, Diagnosis Protocol; Lucy Fu: None; Nasma Majeed: None; Yash Dharmamer: None; Uwadia Edomwonyi: None; Charisse Liz Baste: None; John Groth: None; Manmeet Singh: None; Tushar Patel: None

Background: One solution to increase utilization of deep learning (DL) applications in routine pathology operations is to further engage practicing pathologists. However, DL techniques are perceived to be of high complexity, as they still suffer from lack of mainstream exposure among practicing pathologists.

We explore DL methods to demystify the technical complexity of currently available DL platforms and reveal relatively simple approaches to prototype, experiment with, and deploy DL algorithms with practical applications within the department of pathology.

Through this platform presentation we plant to utilize a combination of screen shots, videos, and screen recordings to demo a re-creation of skills acquired and lessons learned, during the preparation of DL use cases submitted separately to USCAP.

Design: We leveraged publicly available DL platforms that use graphical user interfaces to develop an approach for achieving the above potentials with mainly intra-departmental resources.

We utilized image pre-processing tools (MacOS Preview), image sorting and labeling tools (Diagnosis Protocol), commercially available cloud-based DL platforms (Customvision.ai from Microsoft [figure 1], AutoMLvision from Google, and Rekognition from Amazon) as well as more advanced DL techniques (Python and DIGITS from NVIDIA [figure 2]) for bench marking. We designed a DL workflow to be independent from extra-departmental resources (e.g. engineering and computer science divisions).

Results: We develop the following use cases [figures 1 and 2]:

1. **Automated workflow for triaging** colon polyps: aims for automating report pre-population and triaging case review (abstact ID 1345).

2. Approach for **detecting and typing** electron-dense deposits in renal biopsies: aims to aid the renal pathologist and advance research interest in the field (abstact ID 2619).

3. **Teaching organ-systems** via prototyping deep learning algorithms: couples DL technical tasks with resident learning of intraductal proliferative lesions of the breast (abstact ID 2452).

Figure 1 - 1463

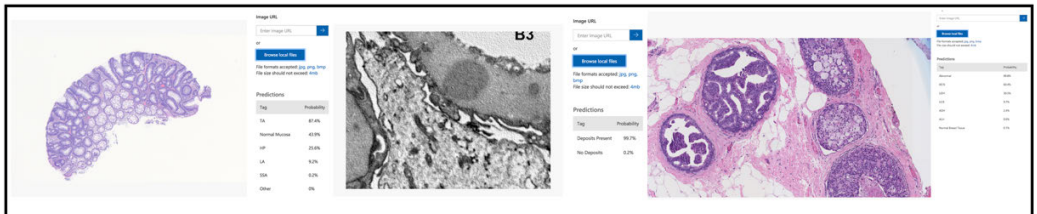


Figure 1. Predictions via user friendly interactive platform (CustomVision.ai from Microsoft) from our colon polyp biopsy screening service (left), electron microscopy deposit detection (middle), and the educational assignment for breast pathology teaching (right)

Figure 2 - 1463

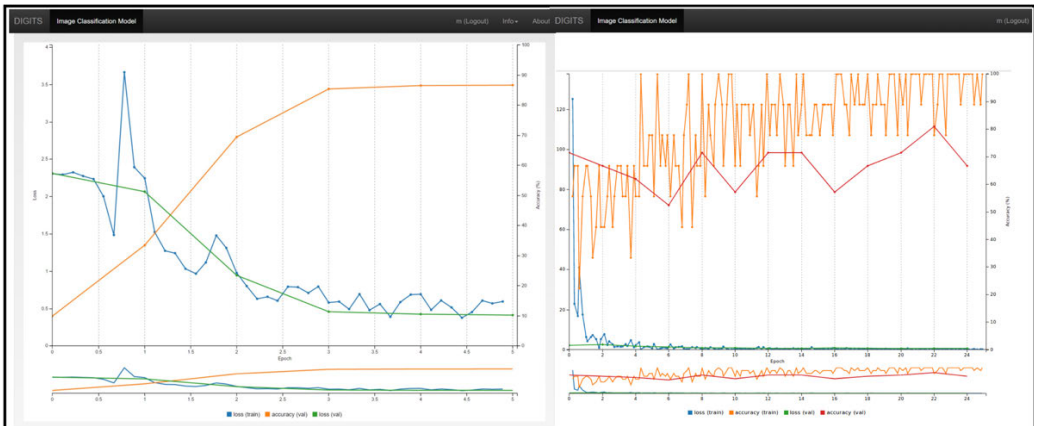


Figure 2. **Left**: A training graph from NVIDIA DIGITS of a model trained by an expert (NVIDIA Deep Learning Institute tutorial) on thousands of images (the MNIST dataset). Note the accuracy (Y axis) of the validation data (orange) climbing to higher 80s by the 3rd epoch (X axis) of cycling through the images. **Right**: Our own training graph from DIGITS utilizing limited amount of electron microscopy immune deposit images (hence the jagged appearance of the training accuracy [orange]). Note the acceptable validation accuracy (red) reaching low 80s at the 22nd epoch of training before it starts to overfit (Dissociation of training and validating).

Conclusions: Compared to published metrics, acceptable performance of the developed DL algorithms can be achieved by practicing pathologists. Using these resources with low barriers to entry, provides evidence, that as artificial intelligence technology continues to mature, accessibility to pathologists will increase. As a result, commercially available platforms could be as effective as expert level platforms for prototyping departmental home grown DL applications.

1464 Building an Anatomic Pathology Report Search Tool Using Business Intelligence Software

Simone Arvisais-Anhalt[1], Justin Bishop[1], Andrew Quinn[1], Ming Zhou[2], Jyoti Balani[3], Ellen Araj[1]
[1]University of Texas Southwestern Medical Center, Dallas, TX, [2]Clements University Hospital, Dallas, TX, [3]Irving, TX

Disclosures: Simone Arvisais-Anhalt: None; Justin Bishop: None; Andrew Quinn: None; Ming Zhou: None; Jyoti Balani: None; Ellen Araj: None

Background: Pathology generates rich and abundant data. This data is likely capable of improving patient outcomes, quality of care, and advancing medicine; however, it is largely inaccessible. Most of the data generated in anatomic pathology is stored as unstructured, free text reports. Our institution desired the ability to perform retrospective free-text searches on these reports with a cost effective, easy to use, and functional solution. We used existing business intelligence (BI) software to develop a tool to search free text pathology reports and structured patient data.

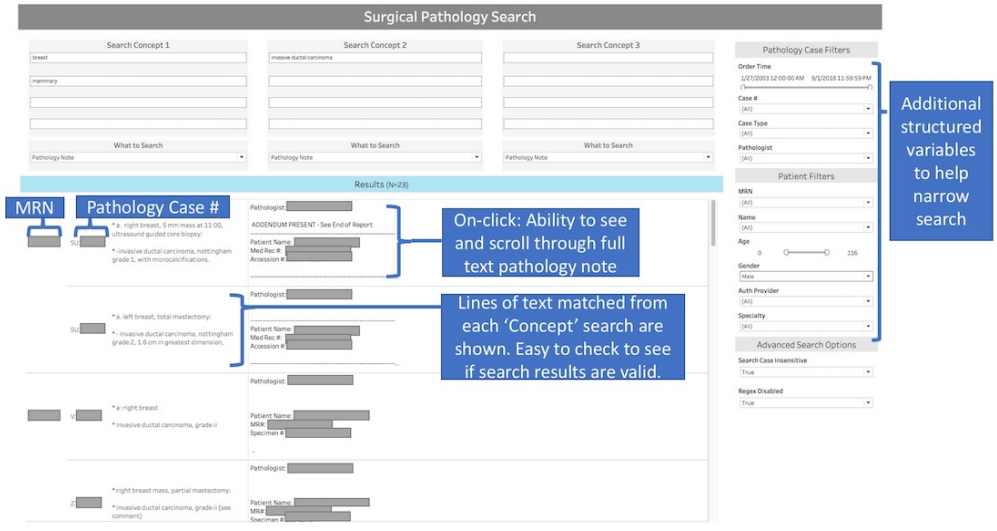
Design: We wrote a custom Structured Query Language (SQL) script to capture both text from anatomic pathology reports and structured patient data stored in our institution's clinical relational database. We uploaded the data into the BI tool, Tableau Version 10.5, and created a series of text parameters as user input boxes. The interface allowed the user to input up to three search concepts logically combined as "AND" in the user input boxes. For each of the three concepts, the user could input up to four synonyms logically combined as "OR". The search could then be made case sensitive and or use regular expression syntax. When the user performed the search, the input text parameters were taken through a series of algorithmic manipulations using Tableau's built-in regular expression functionality to search the pathology report one line at a time. The search output showed the full text reports and highlighted the line where each search concept was found (Figure 1).

Results: N=262,704 reports were populated from the clinical relational database between 1/27/2003 to 9/1/2018. We performed 20 searches using the Tableau tool and direct database SQL query to confirm the regex functionality performed similar to a SQL "like" operator and achieved 100% concordance. Table 1 shows example search times on a 2.9GHz processor against all records

Table 1: Example Search Criteria and Search Times

Description	Search Criteria	Search Features	Total Cases	Time
Breast carcinoma in a male	Text Search: ("breast" or mammary") and ("invasive ductal carcinoma") Discrete Field: Sex = Male	Search Case Insensitive = "true" Regex Disabled = "true"	N=23	15 sec
ICD 9 diagnosis of heart failure with pathology findings of amyloidosis	Text Search: (ICD9 search for "heart failure" or "428.0") and Text Search: (pathology case search for ("amyloid.*positive" or "positive.*amyloid" or "congo red.*positive" or "positive.*congo red"))	Search Case Insensitive = "true" Regex Disabled = "false"	N=62	14 sec
Diffuse large B cell lymphoma	Text Search: ("diffuse large b cell lymphoma" or "diffuse large b-cell lymphoma" or "dlbcl")	Search Case Insensitive = "true" Regex Disabled = "true"	N=704	20 sec

Figure 1 - 1464



Conclusions: We developed a cost effective, easy to maintain pathology report search tool using business intelligence software. Although regular expression search is one of the most inefficient search methodologies, our test users found the search times to be acceptable. Our department adopted this tool for operational, educational, and research use. Since many institutions already have subscriptions to a business intelligence software, we believe this solution could be easily reproduced.

1465 Digital Imaging for Systematic Validation of Spatially Annotated Mirror-Image Simultaneous Flash Frozen and FFPE Tissue Banking for Research

Umesh Bhanot[1], Maria Corazon Mariano[1], Emily Lin[1], Priscilla McNeil[1], Diana Carrillo[1], Jody-Ann Facey[1], Peter Sharpe[1], Nurul Alhamid[1], Peter Ruh[1], Aziza Barnett[1], Ucalene Harris[1], Joachim Silber[1], Michael H. A. Roehrl[2]
[1]Memorial Sloan Kettering Cancer Center, New York, NY, [2]Scarsdale, NY

Disclosures: Umesh Bhanot: None; Maria Corazon Mariano: None; Emily Lin: None; Priscilla McNeil: None; Diana Carrillo: None; Jody-Ann Facey: None; Peter Sharpe: None; Nurul Alhamid: None; Peter Ruh: None; Aziza Barnett: None; Ucalene Harris: None; Joachim Silber: None; Michael H. A. Roehrl: None

Background: The Precision Pathology Biobank at MSKCC handles the wide variety of tumor samples that are banked for clinical trials, early drug development, molecular characterization, and basic research. The comprehensive evaluation of tissues to assess tumor content, necrosis, inflammation, and fibrosis is necessary before samples are used for research. We decided to integrate digital pathology into the biobank workflow to assess our model of routinely collecting a spatial mirror image pair of liquid nitrogen fresh frozen and FFPE samples from every biobanked specimen.

Design: Tissues are routinely flash frozen in liquid nitrogen and stored in vapor phase LN2 (~8,000 patients/year). We also routinely prepare spatially annotated "mirror image" paraffin blocks from fresh tissue. An H&E stained slide from the "mirror image" paraffin block provides an indirect histopathologic assessment of the frozen tissue in LN2. All H&E slides are prospectively scanned on a Leica AT2. Because LN2-frozen samples should undergo a minimum of freeze-thaw cycles, one would like to use specimen content information from the mirror image FFPE to extrapolate to the frozen aliquot. Here, we devised a study to assess the validity of such approach. We utilized a fully digital image workflow in eSlideManager (Aperio).

Results: Our database of scanned H&Es from biobanked research FFPE blocks alone contains >15,000 scanned slides. H&Es prepared from 27 randomly selected frozen aliquots retrieved from LN2 were digitized for comparison with corresponding digital slides of mirror image FFPE blocks. The histopathologic findings were identical in 93% of cases (concordant cases). In 7% of cases, tumor was absent in the H&E from frozen tissue but present on the FFPE slide. The study is ongoing and currently including concordance assessment of 200 additional cases.

Conclusions: Digital imaging can be applied in biobanking to confirm the lesional histology of biospecimens as shown here in a pilot cohort. Comparison of digital images from simultaneous flash frozen and FFPE tissues shows a fully concordant representation of lesional histology in 93% of samples. Spatially indexed "mirror image" paraffin blocks are highly effective for selection of tissues without having to sacrifice the integrity of LN2-frozen samples. Incorporation of digital pathology in research biobanking provides many additional advantages, such as real-time access to banked sample histology from anywhere and resource for computational algorithm development.

1466 Automated Mitotic Detection of Whole Slide Images in Pancreatic Neuroendocrine Tumors

Morgan Blakely[1], Levi Machado[2], Brandon Veremis[3], Marcel Prastawa[4], Abishek Sainath Madduri[3], Jack Zeineh[5], Rafael Cabal[3], Cynthia Harris[6], Michelle Kim[3], Michael Donovan[3], Gerardo Fernandez[3]
[1]New York, NY, [2]Mount Sinai Hospital, New York City, NY, [3]Mount Sinai Hospital Icahn School of Medicine, New York, NY, [4]Mount Sinai Hospital, New York, NY, [5]Newport Coast, CA, [6]Massachusetts General Hospital, Boston, MA

Disclosures: Morgan Blakely: None; Levi Machado: None; Brandon Veremis: None; Marcel Prastawa: None; Cynthia Harris: None; Michelle Kim: *Consultant*, Novartis; *Consultant*, Ipsen

Background: Mitotic figure counting is an important prognostic and grading parameter in numerous tumors across different organ systems. Considering its ubiquity, mitotic figures represent an ideal quantitative measure to automate through deep learning. Several groups have developed mitotic detection by training a convolutional neural network (CNN) on specific organ systems, i.e. breast cancer, with respectable accuracy. However, despite promising results, most groups have tested CNNs on static regions of interest (ROI) rather than applying the CNN across a whole slide image (WSI). We propose a novel approach for detection of mitotic figures in WSIs of pancreatic neuroendocrine tumors (pNET) by iteratively training and refining a CNN, utilizing a web-based remote multi-pathologist collaborative platform for interactively correcting detection errors.

Design: We trained a CNN that was then applied to whole slide images (WSI) of pNETs (n= 63; grade 1: 33; grade 2: 26; grade 3: 4 as determined by Ki-67 immunohistochemistry and mitotic indices). After initial training, two pathologists iteratively performed corrections on a subset of the pNET WSI using a web-based tool generating a total of 1111 corrections (1034 false positive; 77 false negative) and 652 confirmations. Dominant errors were false detections of hyperchromatic squeezed cells in pNETs, which are marked negative with relatively low effort. After final retraining and refining, we used the CNN to automatically compute maximum count of mitotic figures in areas corresponding to 10 high powered fields. We then measured predictive power of mitotic counts using Pearson correlation.

Results: The CNN was applied to pancreatic NET WSIs to generate maximum mitotic counts, which were found to have a strong correlation of 0.72 to Ki-67 staining, and a strong correlation of 0.81 to manual mitotic counts per 10 high powered fields.

Figure 1 - 1466

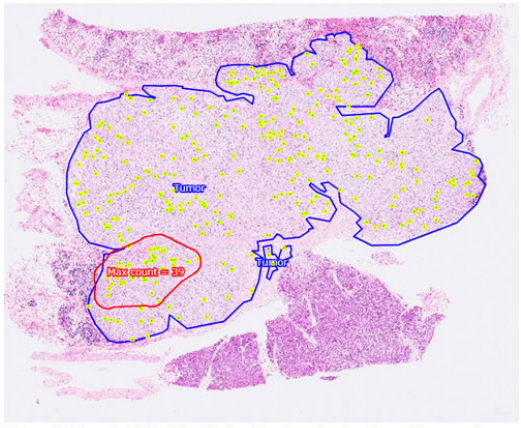


Figure 1: Example WSI with mitotic detections (yellow box with green box inside)

Figure 2 - 1466

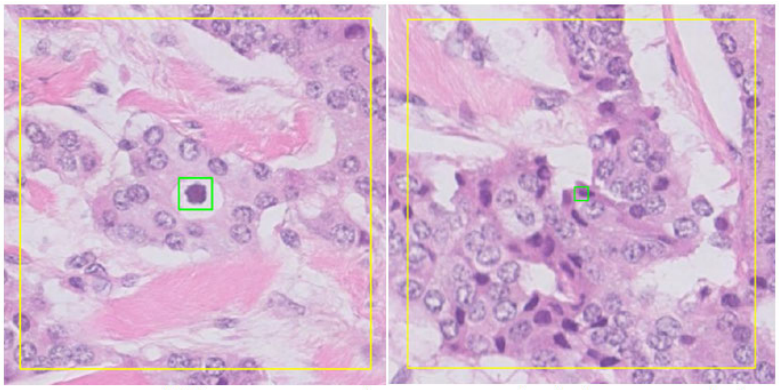


Figure 2: Image on left depicts a true positive detection. Image on the right depicts an example of a false positive detection that required minimal retraining of the CNN to achieve final correlation results.

Conclusions: We have trained and applied a CNN for mitotic figure detection in pNET, with strong correlation to Ki-67 IHC and manual mitotic figure counts. To our knowledge, this is the first development and application of a mitotic detector CNN for pNET, with reduced annotation effort by using iterative training and refinement. We have successfully applied the CNN to WSIs rather than traditional ROIs. Future efforts will explore incorporation within staging systems to ultimately improve outcome prediction.

1467 Automated Gleason Grading of Prostate Biopsies Using Deep Learning

Wouter Bulten[1], Hans Pinckaers[1], Christina Hulsbergen-van de Kaa[2], Geert Litjens[3]
[1]Radboud University Medical Center, Nijmegen, Netherlands, [2]Labpon, Hengelo, Netherlands, [3]Radboud University Nijmegen Medical Center, Nijmegen, Netherlands

Disclosures: Wouter Bulten: None; Hans Pinckaers: None; Christina Hulsbergen-van de Kaa: None; Geert Litjens: *Grant or Research Support*, Philips Medical Systems

Background: Grading prostate cancer is a time-consuming process and suffers from high inter- and intra-observer variability. Advances in computer-aided diagnosis have shown promise in improving histopathological diagnosis. We trained a deep learning system using data retrieved from the patients records to grade digitized prostate biopsies. Our system is the first that can automatically classify background, benign epithelium, Gleason 3, 4, and 5 on a gland-by-gland level in prostate biopsies.

Design: 532 glass slides containing 2162 prostate biopsies, evaluated by an experienced urogenital pathologist were collected and scanned. 596 biopsies were kept separate for evaluation, the remaining 1576 were used to train the deep learning algorithm (Table 1).

A single label denoting the Gleason score (e.g. 3+4=7) was available for each biopsy, without information on tumor location. To generate detailed annotations for training we used two previously trained deep learning networks to first segment the epithelium and, subsequently, to detect cancer. The Gleason grade from the patient record was assigned to the cancerous epithelium. These generated weakly annotated regions of tumor were then used to train a Gleason grading system.

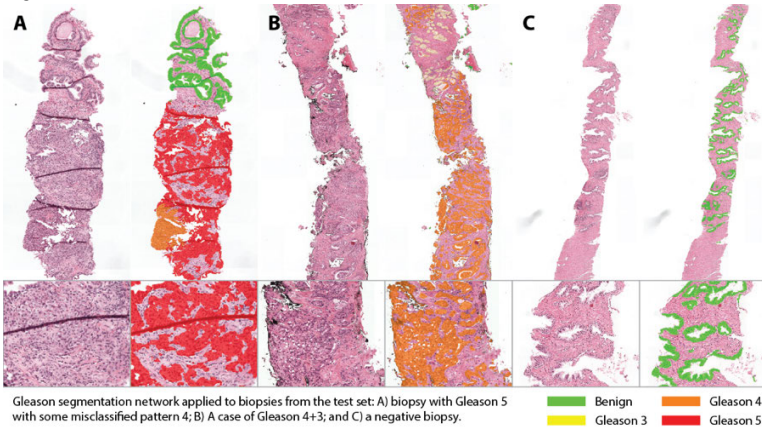
To evaluate, the system was applied to the biopsies in the test set. We used the total predicted surface area of each growth pattern to determine the Gleason score of the biopsy. Predicted tumor areas smaller than 15% of total epithelial tissue were considered unreliable (e.g. incomplete glands at the edges of the biopsy) and ignored for biopsy level classification. For predicted grades only areas larger than 5% of all epithelial tissue were considered, which is also common in clinical practice.

Results: Predicting whether a biopsy contains tumor resulted in an accuracy of 86% (linear weighted kappa (k) of 0.73, area under the ROC curve of 0.96). We compared the predicted primary Gleason grade to the one from the pathologists' report. Our system achieved an accuracy of 75% (k 0.64). On predicting the Grade Group, the system achieved an accuracy of 67% (k 0.57). Misclassifications of more than one grade are rare.

Overview of cases in the data set.

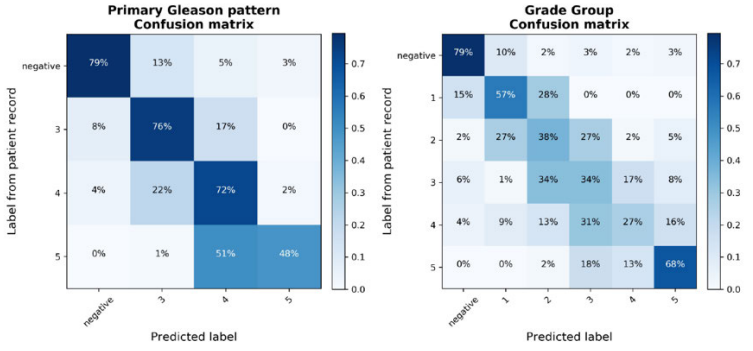
		Number of biopsies			
	Total	**Negative**	**Gleason 3**	**Gleason 4**	**Gleason 5**
Training set	1576	714	474	295	93
Test set	596	241	119	151	85
Total	2162	955	593	436	178

Figure 1 - 1467



Gleason segmentation network applied to biopsies from the test set: A) biopsy with Gleason 5 with some misclassified pattern 4; B) A case of Gleason 4+3; and C) a negative biopsy.

Figure 2 - 1467



Conclusions: Our deep learning system automatically identifies Gleason patterns and benign tissue on a gland-by-gland basis (Fig 1). This can be used to determine the biopsy-level Grade Group and Gleason score, and show which parts of the tissue contribute to this prediction. Improvements need to be made to decrease misclassifications, for example in areas with inflammation.

1468 Validation and Implementation of Leica Aperio LV1 Remote Digital Pathology System for Intraoperative Evaluation of Neurosurgical Specimens

Sandra Camelo-Piragua[1], Kyle Conway[1], Kathryn McFadden[1], Peter Ouillette[1], Joshua Jacques[1], Drew Pratt[1], Andrew Lieberman[1], Paul Mckeever[2], Sriram Venneti[3], Eric Jedynak[4], Stephen Eskesen[1], Lloyd Stoolman[1], David McClintock[1]
[1]University of Michigan, Ann Arbor, MI, [2]University of Michigan Medical Center, Ann Arbor, MI, [3]Ann Arbor, MI, [4]Michigan Medicine, Ann Arbor, MI

Disclosures: Sandra Camelo-Piragua: None; Kyle Conway: None; Kathryn McFadden: None; Peter Ouillette: None; Joshua Jacques: None; Drew Pratt: None; Andrew Lieberman: None; Paul Mckeever: None; Sriram Venneti: None; Eric Jedynak: None; Stephen Eskesen: None; Lloyd Stoolman: None; David McClintock: None

Background: Intraoperative evaluations (IOE) of neuropathologic specimens require the expertise of a trained neuropathologist. However, it is often not feasible to have continuous on-site staffing to cover IOE, due to relatively low specimen volumes and scarcity of available specialists. In preparation for the Pathology Department's move to a location three miles away from the main hospital, we validated and implemented the Leica Aperio LV1 telepathology system (LV1) as a solution to cover neurosurgical frozen sections remotely. The LV1 is a robotic microscope that when combined with remote desktop software allows a pathologist to control it from off-site. We have fully implemented remote neurosurgical IOE using the LV1 system.

Design: Five neuropathologists validated the LV1 using intra- and inter-pathologist concordance on 74 retrospective cases with a washout period of at least a year. The LV1 was launched for full clinical use on 7/23/18, with all five neuropathologists performing IOE remotely. We reviewed the results of the IOE performed since live implementation.

Results: A total of 64 IOE were performed remotely over a two-month period. Two discrepancies (3.1%) were identified between the LV1 and glass interpretation. Both were minor discrepancies that did not alter surgical management or affect patient care, attributed to the learning curve associated with using the LV1 in clinical practice. The mean turnaround time (TAT) using LV1 (n=64) was 24.6 minutes

(range = 16 to 42 minutes); compared to TAT of 20.1 minutes (range= 8-46 minutes) over a one-year period (n=277) using traditional IOE methods prior to implementation of the LV1 (p <0.01).

Conclusions: Remote, digital IOE performed using the LV1 have a high concordance rate with glass on-site evaluations. To date, TAT increased by 4.1 minutes. Our validation and implementation use of the LV1 system have identified several areas fundamental to the successful use of this technology: development of standard operating procedures for pathologist notification, slide preparation, LV1 loading and connection from remote site, communication between the central site and remote viewer, troubleshooting instrument failures in the validation stage, appropriate training of the LV1 operators, adequate training familiarizing the pathologist with LV1 software, and developing downtime procedures. We anticipate continuous improvement in discrepancy rates and TAT as we continue to improve our use of this technology.

1469 The California Telepathology Service: UCLA's Experience Deploying a Regional Digital Pathology Sub-Specialty Consultation Network

Thomas Chong[1], Miguel Palma Diaz[2], Craig Fisher[3], Dorina Gui[4], Nora Ostrzega[5], Geoffrey Sempa[6], Anthony Sisk[7], Mark Valasek[8], Beverly Wang[9], Jonathan Zuckerman[10], Chris Khacherian[1], Scott Binder[11], William Wallace[1]
[1]David Geffen School of Medicine at UCLA, Los Angeles, CA, [2]Los Angeles, CA, [3]University of California, San Diego, San Diego, CA, [4]University of California, Davis, Sacramento, CA, [5]UCLA Medical Center, Beverly Hills, CA, [6]University of California, Irvine, Orange, CA, [7]Tarzana, CA, [8]University of California, San Diego, La Jolla, CA, [9]UC Irvine Medical Center, Orange, CA, [10]University of California, Los Angeles, Los Angeles, CA, [11]Hermosa Beach, CA

Disclosures: Thomas Chong: None; Miguel Palma Diaz: None; Craig Fisher: None; Dorina Gui: None; Nora Ostrzega: None; Geoffrey Sempa: None; Anthony Sisk: None; Mark Valasek: None; Beverly Wang: None; Jonathan Zuckerman: *Consultant*, Leica Biosystems; Chris Khacherian: None; Scott Binder: None; William Wallace: *Advisory Board Member*, Leica Biosystems

Background: The need for extending pathology diagnostic expertise to more areas is now being met by the maturation of technology that can effectively and reliably deliver this level of care. The experience and lessons learned from our successfully deployed International Telepathology Service (ITS) to a hospital system in China guided us in starting a domestic telepathology network, the California Telepathology Service (CTS). Many of the lessons learned from the ITS project informed our decision-making for the CTS. New challenges were recognized and overcome, such as addressing the complexity and cost-benefit tradeoffs involved in setting up a digital consultation system that competes with an established conventional glass-slide system.

Design: The CTS is based on a Hub-and-Spoke telepathology system using Leica Biosystems Whole Slide Image (WSI) scanners and the eSlide Manager (ESM) digital image management software system. The service currently comprises 7 Spoke sites (UC San Diego (UCSD), UC Irvine (UCI), UC Davis (UCD), Northridge Medical Center (NMC), Olive View Medical Center (OVMC), Children's Hospital of Los Angeles (CHLA) and Camp Pendleton Navy Hospital (CPNH)) and 1 central Hub (UCLA Medical Center). So far, 5 sites have been validated for telepathology case consultations following established practice guidelines, and 3 sites (UCI, UCSD, and OVMC) have activated the service.

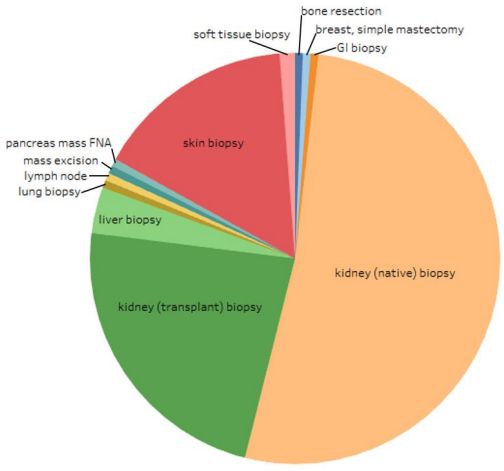
Results: For the active Spoke sites, we reviewed the volume, turn-around-time (TAT) and case types and evaluated for utility and value. From May 2017 to July 2018, 165 total cases were submitted. See Table 1 and Figure 1 for the breakdown of the different specimen types and TATs. Of note, digital consult was particularly of advantage for preliminary renal biopsy diagnoses (avg TAT 0.7 day).

Table 1. Telepathology Subspecialty Consultation Cases

	OVMC	UCI	UCSD
Mean number of whole slide images per case	3	16	9
Mean file transfer time [s]	34	1872	187
Mean image file size [Mb]	174	410*	253
Mean TAT to preliminary diagnosis [days**]	N/A	0.7	1.8
Mean TAT to final diagnosis [days**]	7.9	3.3	8.8
TAT - turn-around-time; OVMC - Olive View Medical Center; UCI - Univ. of California, Irvine; UCSD - Univ. of California, San Diego; * 16 slides were scanned at 40x; ** excludes non-working days			

Figure 1 - 1469

Figure 1. Digital Consult Case Types from UC Irvine, Olive View, and UC San Diego



Conclusions: For Spoke sites, telepathology provided shortened TAT and significant financial savings over hiring faculty with expertise to support a potentially low-volume service. For Hub sites, the value includes exposure to educationally valuable rare cases, additional caseload volume to support specialized services, and improved referring inter-facility communication versus traditional carrier mail. The creation of a Hub-and-Spoke telepathology network is an expensive undertaking and careful consideration needs to be given to the requirements needed to support the clinical services, the appropriate equipment, network requirements, scanner locations, and laboratory workflows to ensure an effective and cost-efficient system.

1470 Robust Automated Screening Algorithm for Acid Fast Bacilli in Surgical Pathology

Ryan Cristelli[1], Wenjin Chen[2], David Foran[2], Evita Sadimin[2]
[1]Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, [2]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ

Disclosures: Ryan Cristelli: None; Wenjin Chen: None; David Foran: None; Evita Sadimin: None

Background: In clinical practice, pathologists often have to perform manual search of Ziehl-Neelsen (ZN) stained tissue to identify acid fast bacilli (AFB), a very time-consuming process as the entire slide must be meticulously screened at high magnification for very small (2-4 um), frequently sparse, organisms. While several computer algorithms have been developed to identify AFB on sputum using traditional ZN or fluorescent staining, only a few attempted to detect them on ZN stained tissue, with highly technical methods. In this study, we present an algorithm which reliably screens for AFB on whole slide images.

Design: We searched our database and identified 218 ZN stained slides in the past two years. Out of these cases, 12 cases were positive for AFB. Whole slide images were scanned at 40x magnification on the Olympus VS120. Imaging analysis algorithm was developed utilizing Visiopharm software, consisting of three phases. Initially, the average percent of blue and red wavelengths across a field of pixels was generated, then each pixel that contained a greater amount of the sum of the percent of blue and red than the average was marked as potential candidates. The second phase eliminated these candidates by degree of saturation. The last phase further screened remaining candidates by size. The final list of candidates was then presented to a pathologist for definite verification.

Results: Our algorithm identified all 12 AFB slides as possible positive cases. The average time spent running the algorithm was 3-20 minutes, depending on the size of tissue. The pathologist then spent 30 seconds to 5 minutes to verify the candidates presented, depending on the difficulty of the cases. None of the slides previously identified by pathologists under the microscope as positive were missed by the algorithm.

Conclusions: We have developed a reliable screening tool for detection of AFB on ZN stained slides. To increase sensitivity the algorithm identified features that were red-shifted due to light refraction of objects during slide scanning and occasional staining artifacts, such as occasional red blood cells. Despite this limitation, we have demonstrated that it was possible to screen a ZN stained slide in seconds to minutes. For future studies we will gather a larger dataset in collaborations with other institutions to identify the robustness of the algorithm on other software.

1471 Computer Assisted Complete Biomarker Profile Extraction of Pathology Reports

Ryan Cristelli[1], Jorden Norin[2], Eitan August[3], Huiqi Chu[4], Evita Sadimin[4], Wenjin Chen[4], David Foran[4]
[1]Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, [2]Rutgers University, Somerset, NJ, [3]Duke University, Durham, NC, [4]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ

Disclosures: Ryan Cristelli: None; Jorden Norin: None; Eitan August: None; Huiqi Chu: None; Evita Sadimin: None; Wenjin Chen: None; David Foran: None

Background: Advances in precision medicine mandate that a patient's cancer treatment plan be personalized based on their tumor's specific molecular profile. Choices in the treatment plan is the result of experience gained and outcomes from previous studies of large numbers of cases exhibiting similar profiles. The Cancer Registry collects standard biomarker information on each cancer type, but their data set is not comprehensive and does not include supplemental biomarker testing results. These results are often buried in text-based pathology reports, and cannot be easily identified, abstracted, indexed or queried. Abstracting biomarker information from text reports is crucial to development of comprehensive enterprise-wide, clinical data warehouses (CDW), allowing them to contain data that is salient to patient care decisions.

Design: We assembled a set of 129 pathology reports of primary tumors with synoptic data in PDF, which were then processed through software that we have developed to allow semi-manual abstraction of biomarker data from pathology reports into a functional CDW. The software leveraged the open-source NOBLE natural language processing library to recognize NCI concepts on proteins, genes and immunological tests, including all variations. Each abstracted biomarker data point was recorded as a json expression, which included biomarker name, test method, results, and related NCI concept. The json expression linked back to text location in the original report. The gold standard was established by a pathology resident manually identifying relevant biomarker data on the reports.

Results: Of the 129 pathology reports reviewed, 97 contained biomarker information, with 6 reports containing 15 or more. A total of 444 biomarker data points encompassing 105 unique tests were detected. The software-assisted method incorrectly identified or missed 4% (18) of biomarker data.

Conclusions: We have shown our software's ability and reliability to help extraction of complete pathology biomarker profile from text reports. An open vocabulary set based on NCI standard accommodates emerging biomarkers to be collected. The current semi-manual process will also harvest contextual data of each biomarker mentioned so that the output of these functions and processes can be used to train more sophisticated algorithms such as deep-learning to optimize high efficiency.

1472 An Information Theory Approach as Compared to LASSO Regression in Feature Selection: The Simplest is Better

Leslie Dalton, St. David's South Austin Medical Center, West Lake Hills, TX

Disclosures: Leslie Dalton: None

Background:
In medical research the biostatisticians are front and center to perform data analysis while those in the computer science department are often many buildings away. Therefore, LASSO regression, and other regression methods, are more likely to be chosen for feature selection (FS) instead of a method with reliance on information theory (IT).

Design: 419 invasive breast cancers from the TCGA provisional dataset (cBioPortal.org) had both mRNA expression (15901 nucleotides) and corresponding whole slide images (WSI) of each tumor. Each tumor was graded. Studied was to build a mRNA-based model to predict grade. For variable importance LASSO regression (R library glmnet) was the recommended approach, but also evaluated was an in-house IT algorithm. The IT "greedily" searches for the attribute having highest mutual information (maxMI) with target. The next step searches for attribute providing highest conditional mutual information (conditioned on maxMI).

Results: From over 15,000 choices, and despite completely different formulaic construction, both algorithms selected TTK as the most important mRNA nucleotide. IT selected FAM161B as second, while LASSO had FAM161B as third, after KLHDC1. By a generalized linear model (glm), TTK, and FAM161B were joined to build a final model. By glm, both TTK and FAM161B were highly significant (both $p<0.00001$) in predicting grade. The IT approach to FS corresponded to an area under curve (AUC) of receiver operator curve (ROC) equal to 0.8804 with high grade as binary target. The LASSO model corresponded to AUC of 0.884. There was no significant difference in AUC ($p=0.84$) between the two models. If 214 proteins were evaluated as predictors (reverse phase protein assay) both algorithms chose ER_alpha, and Cyclin-B1 as the two most valuable. Similar comparability of IT and LASSO was found in other datasets.

Conclusions: The IT algorithm corresponds to two simple formulas: calculation of MI followed by CI. MI identifies the most valuable attribute and maxCI finds attribute with maximal relevance without redundancy. Often maxCI is anti-correlated with maxMI. Selection of ER_alpha after Cyclin-B1 is an example, and with mRNA little known FAM161B was anti-correlated with better known TTK. MI has a

curious connection whereby MI corresponds to maximal growth of utility (in terms of currency this is Kelly betting). This can provide a real-world sense of magnitude. If all else is roughly equal, the simplest is best, and proposed is IT as first in line for FS.

1473 Digital imaging and analysis of nuclear expression of GATA1 in erythroid precursors can distinguish erythroleukemic blasts from normal erythropoiesis

Rajan Dewar[1], Winston Lee[2], Geraldine Pinkus[3], Srikanth Ragothaman, Olga Weinberg[4]
[1]Ann Arbor, MI, [2]University of Michigan, Ann Arbor, MI, [3]Brigham and Women's Hospital, Boston, MA, [4]Children's Hospital Boston, Boston, MA

Disclosures: Rajan Dewar: None; Winston Lee: None; Geraldine Pinkus: None; Srikanth Ragothaman: None; Olga Weinberg: None

Background: GATA binding factor 1 (GATA1) is a zing finger transcription factor that is involved in erythropoiesis. Previous studies from our group showed that GATA1 is a sensitive marker for erythroid and megakaryocytic lineages. However, the practical utility was limited since light microscopic analysis is not sensitive enough to distinguish differences in nuclear expression of GATA1. Digital pathology tools are powerful to assist pathologists in distinguishing subtle differences and thus aid in histopathologic analysis and diagnosis. Here, we have developed a digital pathology algorithm that can distinguish normal erythroid colonies from erythroid precursors in erythroleukemia (M6b).

Design: Paraffin sections of 18 bone marrow core biopsies were retrieved and their diagnosis confirmed by repeat review. These included 6 cases of Pure erythroleukemia (M6b), 6 cases of MDS (formerly erythroid-myeloid leukemia M6b) and 6 cases of staging marrow. They were stained with GATA1 (D52H6 clone; Cell signaling Technology, Danvers, MA). The slides were scanned using Motic (Eazyscan) scanner. The whole slide images were viewed using an Aperio Imagescope. Using the 'Extract Region' tool, clusters of erythroids, comprising at least 10 cells were analyzed as individual erythroid colonies. A nuclear intensity segmentation algorithm was developed using Imagescope - defining finer areas of differential GATA1 expression within the nucleus (Figure 1). A colony cell count and maturation index (frequency distribution of intensity) was developed for each colony.

Results: GATA1 nuclear intensity distribution is different within the three groups. Normal erythropoiesis shows a high proportion of weak staining of GATA1. Pure erythroid leukemia shows a larger number of strong staining and very few weak staining intensity. The GATA1 staining is somewhat homogeneously distributed in erythroid colonies of Myelodysplastic syndrome (see frequency distribution in figure 2 and Table 1).

	3+	2+	1+	WEAK	SD
Normal	8.503412	6.705797	7.977998	76.81278	34.55009794
MDS	34.64242	20.36598	17.70563	27.28598	7.591308384
M6B	44.67422	22.49111	13.20947	19.61975	13.67885259

Figure 1 - 1473

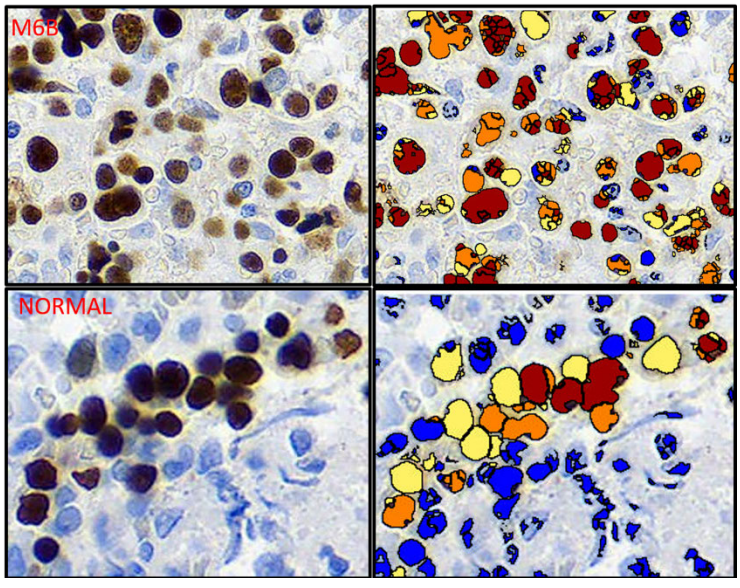
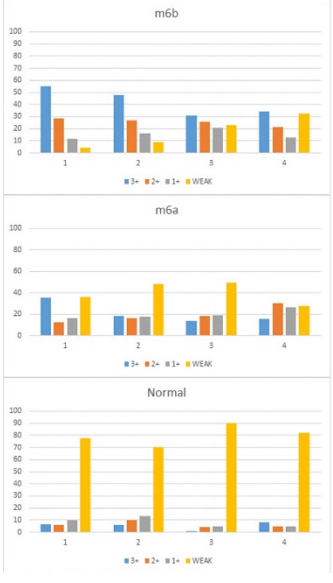


Figure 2 - 1473



Conclusions: The diagnosis of erythroleukemia remains a challenge by conventional morphology. Robust erythropoiesis, erythroid dominant MDS can morphologically be difficult to distinguish from pure erythroid leukemia. We use digital analysis of GATA1 expression to

delineate finer distribution of nuclear expression, which may not be normally apparent by routine histopathology. Thus, digital quantification and analysis can aid in the diagnosis of erythroleukemia.

1474 A Comparison of Image Analysis Algorithms Versus Conventional Methods of Analysis to Detect TTF-1 Positive Lung Cancer on Cell Blocks

Deepak Donthi[1], Gaurav Khullar[2], F. Zahra Aly[3], Mark Bowling, Bryan Dangott[4]
[1]Vidant Medical center/ East Carolina University, Greenville, NC, [2]Cedars Sinai Medical Center, Los Angeles, CA, [3]University of Florida College of Medicine, Gainesville, FL, [4]Greenville, NC

Disclosures: Deepak Donthi: None; Gaurav Khullar: None; F. Zahra Aly: None; Mark Bowling: *Consultant*, Medtronic; Bryan Dangott: None

Background: Manual visual scoring of immunohistochemistry (IHC) has high intra and interobserver variability. Interobserver variability is introduced due to different approaches for determining the positive cell (PC) ratio which may include: PC as percentage of tissue area, PC as a percentage of total cells, or PC as a percentage of tumor. In addition, intensity scores introduce another degree of variability. Image analysis algorithm's performance is challenged by image artifact, cell detection, and cell classification. This study evaluates manual reads versus various image analysis algorithms for assessment of image based IHC analysis on cytology cell blocks.

Design: The study was conducted at East Carolina University with the Department of Pathology and Pulmonology. The study was performed on 15 lung cancer patients with cytology sampling which included cell blocks stained with TTF-1. The IHC slides were manually analysed by 2 independent pathologists, image analysis software Aperio Nuclear Algorithm (Vista California, Aperio digital pathology) and an internally developed ECU algorithm (IDEA) of our center. The IDEA distinguished positive benign cells from positive tumor cells for calculations, while the vendor algorithm reported percentage of PC at various intensities (0, 1+, 2+, 3+) without discretely identifying cells as tumor. The IDEA addressed issues where benign cells may express TTF-1 or non-cellular material was counted as cells by the vendor algorithm.

Results: A regression analysis showed a very high correlation between the vendor algorithm and the IDEA which reported PC as a percentage of total cells (R^2 = 0.97) Figure 1. There was no significant difference between the means for the vendor algorithm and the IDEA (50.3, 51.7, p=0.19) Table 1. The IDEA algorithm reported tumor as a subset of total cells which averaged 17% less than the total positive cells (p= 2.23E-06). Comparison of means for reader 1 and 2 highlighted few challenges with interobserver variability (means 53.6, 68.7, p=0.046).

Figure 1 - 1474

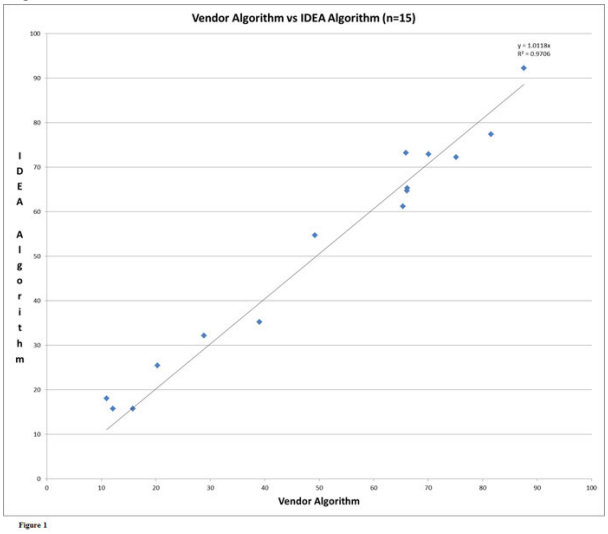


Figure 1

Figure 2 - 1474

Table of P values with Means

		Reader 1	Reader 2	Vendor Algo	IDEA All Pos	IDEA Tumor Only
	Mean	53.7	68.7	50.3	51.7	35
Reader 1	53.7		0.04594	0.40775	0.61405	0.00011
Reader 2	68.7	0.04594		0.01805	0.02092	0.00002
Vendor Algo	50.3	0.40775	0.01805		0.18749	0.00003
IDEA All Pos	51.7	0.61405	0.02092	0.18749		0.00000
IDEA Tumor Only	35	0.00011	0.00002	0.00003	0.00000	

Table 1: Average reported positive percentage for each category with associated p-values.

Conclusions: Use of the IDEA allowed separation of tumor positive cells from benign positive cells which is consistent with clinical guidelines. Reporting of positive tumor cells may provide better clinical relevance in comparison to generalized percentage scores. Further study and validation may allow the IDEA to be used to determine if enough tumor is present to perform additional refreference testing which may be relevant when considering cost of molecular tests.

1475 Assessing Hospital Capacity for Complex Data Analysis in Stage IV Colorectal Cancer Patients via Chart Review

Kevin Dowlatshahi[1], Annelisse Santiago Pintado[2], Carol Geary[1], James McClay[1], James Campbell[1], Walter Campbell[1]
[1]University of Nebraska Medical Center, Omaha, NE, [2]Omaha, NE

Disclosures: Kevin Dowlatshahi: None; Annelisse Santiago Pintado: None; Carol Geary: None; James McClay: None; James Campbell: None; Walter Campbell: None

Background: Histologic, proteomic and genomic assessment of tissue by pathologists remain the gold standard of cancer diagnosis and prognosis. In most instances, these assessments are recorded in textual, natural language formats rendering this rich source of data as "lost to follow up" and unsuitable for computational reuse and analytics. University of Nebraska Medical Center (UNMC) efforts to render pathology and clinical data in natural language and computable formats were assessed in the context of colorectal cancer (CRC) and gene targeted therapy use.

Design: UNMC records were reviewed for diagnoses of stage IV CRC or non-stage IV CRC with metastatic disease between 2013-2016. Records were assessed for mismatch repair testing (by immunohistochemistry (IHC) and microsatellite instability), BRAF,*RAS and gene variants. Gene targeted therapies (BRAF/EGFR inhibitors and immunotherapy) were also identified and analyzed against FDA guidelines. Computational analysis was validated by manual chart review.

Computational analysis of UNMC cancer registry data reported to NAACCR identified all patients seen at UNMC for CRC with staging data. Use of SNOMED CT encoded problem lists, ICD-10-CM encoded encounter data and LOINC encoded lab data for carcinoembryonic antigen (CEA) levels ensured identification of CRC patients who developed metastatic disease during the study period. Computer records of IHC and genomic test results were employed to determine the performance and results of genomic and proteomic assessments. EHR medication order/dispense events and encoded RxNorm and NDC codes provided data on gene targeted therapies.

Results: Computational analysis of NAACCR registry data indicated 400 patients were diagnosed with CRC during 2013-2016. Exclusion criteria were those who solely received surgical, palliative or radiologic services at UNMC, resulting in a patient cohort of 264. Computational phenotyping of patients with metastatic disease reduced the final cohort to 75 patients of which 14 received gene targeted therapy (panitumumab and cetuximab), as confirmed by chart review.

Conclusions: This study demonstrated the data "readiness" of UNMC EHR and clinical data warehouse data to support complex analysis of patient records (genomic and pathology data) and that such analyses are possible. Findings also indicate gaps in the information management and flow of data which, when filled, will promote more advanced and meaningful use and reuse of pathology data.

1476 Comparability of Databased Pathological Content, Pathology Report, and Diagnostic Whole Slide from the Cancer Genome Atlas: Experience from the Uveal Melanoma Cohort

Philippe Echelard[1], Leonardo Lando[2], Anne Xuan-Lan Nguyen[3], José-Mario Capo-Chichi[4], Vincent Quoc-Huy Trinh[5]
[1]Université de Sherbrooke, Sherbrooke, QC, [2]Federal University of Goias, Goiania, GO, Brazil, [3]McGill University, Montreal, QC, [4]University Health Network, University of Toronto, Toronto, ON, [5]Centre Hospitalier de l'University de Montreal, Montreal, QC

Disclosures: Philippe Echelard: None; Leonardo Lando: None; Anne Xuan-Lan Nguyen: None; José-Mario Capo-Chichi: None; Vincent Quoc-Huy Trinh: None

Background: Publicly accessible genomic databases represent a wealth of data. Digital diagnostic slides (DDS) have been used to discover novel biomarkers. However, there is a paucity of data concerning their representativity compared to encoded diagnostic data (EDD) and pathology reports (PR). Our goal was to correlate acquisition methods, and to determine the best method of data aggregation for uveal melanoma (UM).

Design: Eighty UM are available from The Cancer Genome Atlas (TCGA). Clinical, pathological and histological data were extracted using different programs in R and from web-based platforms: R x64 3.5.1, cBioPortal, TCGABiolinks 3.7, GDC Legacy Archive, GDC Data Transfer Tool v1.3.0. Digital diagnostic and frozen section slides were all reviewed by three observers in Aperio Imagescope v12.2.2.2.5015. Uncertain histology was negative. The following data were extracted from DDS: basal diameter, thickness, growth pattern, mitosis, lymphocytes, vascular pattern, pigmentation, and histologic variant. Seven datapoints were compared according to acquisition method (EDD, PR, DDS). The best method of aggregation of diagnostic data was then selected. The aggregated data was tested for microvascular loops (MVL) in progression-free survival (PFS). SPSS v23.0 was used.

Results: Data extraction from EDD generated 2632 datapoints, PR generated 948 and DDS generated 720 (figure 1). Comparison between different acquisition methods highlight discordances (table 1). Notably, DDS was accurate for basal diameter, mitoses, pigmentation and histologic variant, but underestimated thickness. DDS generated higher rates of datapoints for MVL (from 30 to 80), yet did not detect them in two EDD-positive cases. DDS detected higher rates of significant lymphocyte infiltration. Globally, the highest measurable value was chosen when aggregating the data. For example, since it could be visually measured, the highest mitotic rate of all

methods prevailed, just as the presence of MVL regardless of detection method was considered positive. When testing MVL, EDD and PR data did not show significant log-rank testing (P=0.843), but the aggregation of EDD, PR, and DDS data was significant (P=0.047).

Table 1. Paired sample testing and Spearman correlation testing of encoded databased data vs pathology reports vs histological review by three observers for uveal melanoma from the Cancer Genome Atlas (TCGA)

Comparison	TCGA data versus pathology report	TCGA data versus digital diagnostic slide	Pathology report versus digital diagnostic
Mean Difference in Basal Diameter (mm)*	1.85 (p<0.001)	2.41 (p=0.003)	0.68 (p=0.397)
Mean Difference in Thickness (mm)*	-0.18 (p=0.494)	2.49 (p<0.001)	2.60 (p<0.001)
Mean difference in Mitosis per 6.08 mm^{2}*	0.63 (p=0.322)	2.40 (p=0.145)	2.21 (p=0.116)
Presence of microvascular loops (Correlation coefficient)**	r = 0.64 (p<0.001)	r = 0.21 (p=0.057)	r = 0.14 (p=0.206)
Presence of lymphocytes (Correlation coefficient)**	r = 1.00 (p<0.001)	r = 0.16 (p=0.160)	r = 0.19 (p=0.360)
Pigmentation (Correlation coefficient)**	N/A	N/A	r = 0.41 (p=0.001)
Histologic variant (Correlation coefficient)**	r = 0.93 (p<0.001)	r = 0.50 (p<0.001)	r = 0.56 (p<0.001)

* Paired T-test.

**Spearman correlation.

Figure 1 - 1476

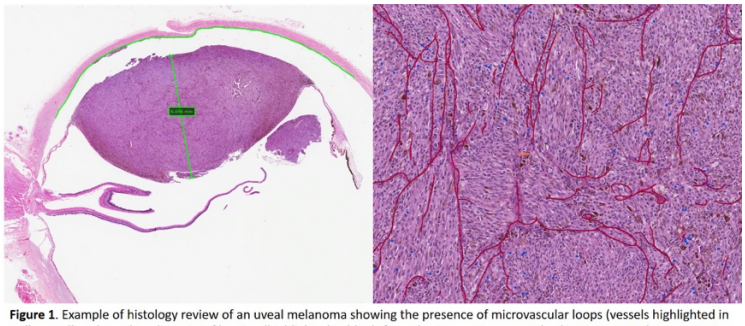


Figure 1. Example of histology review of an uveal melanoma showing the presence of microvascular loops (vessels highlighted in red) as well as dense lymphocytic infiltration (highlighted in blue), from the Cancer Genome Atlas (0.4X, inset 10X). One mitotic figure is identified in orange.

Figure 2 - 1476

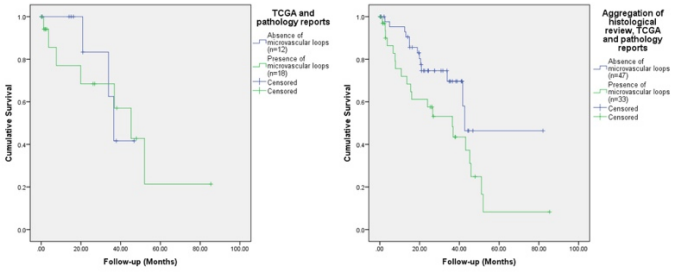


Figure 2. Comparison of progression-free survival curves in uveal melanomas with or without microvascular loops, from the Cancer Genome Atlas (TCGA). A) Survival curves were calculated with data extracted from the TCGA database and pathology reports (log-rank P=0.843). B) Survival curves were calculated with data extracted from the TCGA, pathology reports and histological review. The uveal melanomas was considered positive for microvascular loops if any of these sources noted their presence.

Conclusions: Our results show that DDS and PR data extraction generated large amounts of information that were either incomplete or absent from the TCGA database. Certain analyses were relevant after data aggregation. Our results serve as the base for other studies and explore the importance of histology in the TCGA.

1477 Validation of Multiple High-Throughput Whole Slide Imaging (WSI) Systems for Primary Diagnoses and Clinical Applications at a Large Academic Institution

Ogechukwu Eze[1], Scott Hammond[2], Xiaoyan Cui[2], Zaibo Li[1], Rulong Shen[3], Adrian Suarez[2], Wendy Frankel[1], Lynn Schoenfield[1], Anil Parwani[2]
[1]The Ohio State University Wexner Medical Center, Columbus, OH, [2]The Ohio State University, Columbus, OH, [3]Columbus, OH

Disclosures: Ogechukwu Eze: None; Scott Hammond: None; Xiaoyan Cui: None; Zaibo Li: None; Adrian Suarez: None; Wendy Frankel: None; Lynn Schoenfield: None; Anil Parwani: None

Background: Whole slide imaging (WSI) technology allows for high-speed, high-resolution digital acquisition of diagnostic quality images from glass slides. Images can be interpreted on a computer monitor with the same magnification and spatial maneuvering capabilities as conventional microscopy (CM). Digital images created by FDA WSI systems can be used for primary diagnosis however; a critical concern for clinical application is whether diagnoses by WSI are comparable to CM. We performed a multi-instrument clinical validation study to demonstrate that WSI can be safely and effectively used for primary clinical diagnoses of H&E, IHC and special stained (SS) slides.

Design: A non-inferiority study using intra-observer variability was performed. Diagnoses using the WSI (manual digital or MD) and conventional microscopy (manual optical or MO) were compared to a clinical reference standard (the main, original sign-out diagnoses), as determined by clinical arbitration process. To reflect clinical practice, 100 cases (1 slide/case) were included; 60 H&E, 20 IHC and 20 SS slides. Pathologists (n=5) representing a wide range of experience and subspecialties reviewed a test set by MO. Following a wash-out period of at least 2 weeks, the test sets were reviewed by MD. Tests were scored by a 3rd party and discrepancies reconciled based on the primary diagnosis. Intra-pathologist concordance was calculated as the percentage equivalent diagnoses rendered by manual optical and manual digital microscopy.

Results: The overall intra-pathologist concordance was high (95%) with values ranging from 90% to 99% (Table). This high rate is reflected in each subcategory (H&E, IHC and SS). Inter-pathologist concordance was 96%, with values ranging from 88% to 98% for H&E reads.

Test Name	H&E (n=60)	IHC (n=20)	Special Stains (n=20)	Intra-pathologist concordance
A	57	18	20	95%
D	59	20	17	96%
H	59	20	20	99%
C	59	16	18	93%
E	53	17	20	90%
Overall concordance				95%
SD				3%

Conclusions: To our knowledge, this is the single largest WSI scanning facility in the world for clinical applications and this is the first validation data set from a single institute with multiple WSI systems in a single implementation. The overall intra-pathologist concordance rates were high and in line with published studies. Some of the variability may be explained by the differences amongst pathologists such as years of sign-out experience, familiarity with interpretation of the test IHC and SS or variability in experience with digital sign-out. Future efforts are focused on continued validation of the remaining WSI systems as well as collection of post-digital signout data from pathologists using the systems for clinical applications.

1478 Explainable AI (XAI) in Computational Pathology Pipelines: Translating Machine Learning Features into Pathologist-Friendly Language

Jeffrey Fine[1], Akif Tosun[1], D. Lansing Taylor[1], Michael Becich[2], S. Chakra Chennubhotla[1]
[1]University of Pittsburgh, Pittsburgh, PA, [2]University of Pittsburgh School of Medicine, Pittsburgh, PA

Disclosures: Jeffrey Fine: *Major Shareholder*, SpIntellx, Inc.; Akif Tosun: None; D. Lansing Taylor: None; Michael Becich: *Major Shareholder*, SpIntellx; S. Chakra Chennubhotla: *Major Shareholder*, SpIntellx, Inc.

Background: An initial challenge in implementing machine learning in pathology is establishing a full understanding and trust in AI by pathologists who could potentially use the technology for computer-guided review of slides. A future trend toward increasingly automated analyses is likely, but more extensive validation and clinical trials would be needed with more details on how the algorithms work. Explainable AI (XAI) is AI that is transparent, with actions can be understood by humans. This goes beyond just understanding what AI is doing; it can also help define an AI system's limitations in order to predict or prevent pitfalls or outright failure. Herein we report work on

explaining an AI system that was used to review breast core biopsies for atypical ductal hyperplasia (ADH), a notoriously difficult lesion for pathologists to diagnose.

Design: Using spatial image statistics the computational pathology pipeline extracted then triaged breast duct regions of interest (ROIs) in breast core biopsies as atypical vs not atypical (n=46 cases). For this 269 deidentified whole slide images (WSIs) were created (20x, ScanScope XT, Leica Biosystems, Buffalo Grove IL USA). The AI used a combination of nuclear morphology and architectural arrangement to create 18 micro-spatial duct features that described 95% of the training set of 1009 ROIs (Fig. 1). The study was approved by an institutional review board.

Results: ROIs were reviewed in the context of the 18 features especially features 5, 9 and 15 which were differentially seen in atypical vs non-atypical ducts (Fig. 2). Feature 5 appeared to correlate with architectural rigidity and cobblestoning of cells; feature 9 appeared to represent stromal density immediately surrounding ducts; and feature 15 seemed to correlate with hyperplasticity of the duct. Review of these ROIs suggested several opportunities for creation of software tools specifically for XAI (e.g., an AI feedback overlay, automatic presentation of Feature examples, etc.).

Figure 1 - 1478

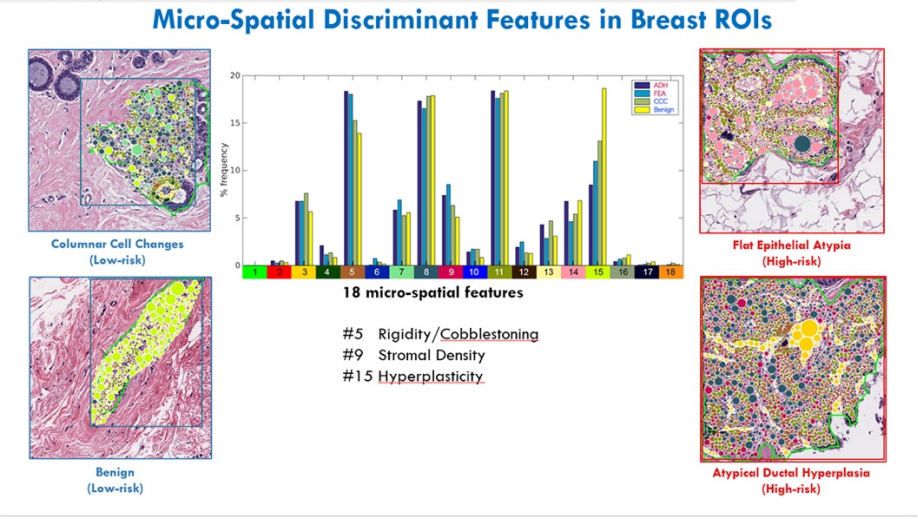
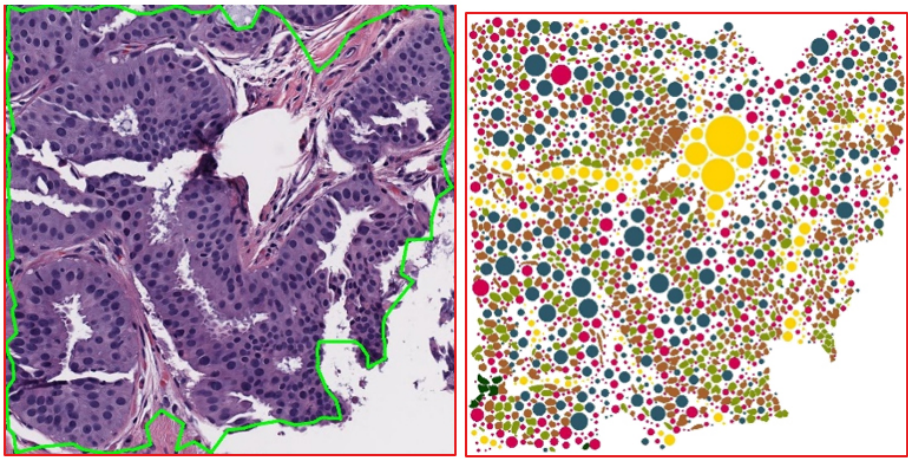


Figure 2 - 1478



Conclusions: AI is poised for disruptive growth in pathology, and stakeholders will need XAI tools to evaluate, improve, and ultimately trust machine learning. Our study took apart a previously reported computational pathology pipeline to critically look at its feature results; while ADH is normally described as “rigid” or “cobblestoned”, the finding of a stromal feature was surprising. Next steps include larger studies with more pathologists, and creation of standardized tools for evaluating computational pathology systems.

1479 Assessing HER2 Immunohistochemistry Using Digital Imaging Analysis and its Correlation with HER2 Fluorescence in Situ Hybridization Results

Ramon Hartage[1], Kurtis Yearsley[2], Scott Hammond[2], Anil Parwani[2], Zaibo Li[3]
[1]The Ohio State University Wexner Medical Center, Westerville, OH, [2]The Ohio State University, Columbus, OH, [3]The Ohio State University Wexner Medical Center, Columbus, OH

Disclosures: Ramon Hartage: None; Kurtis Yearsley: None; Scott Hammond: None; Anil Parwani: None; Zaibo Li: None

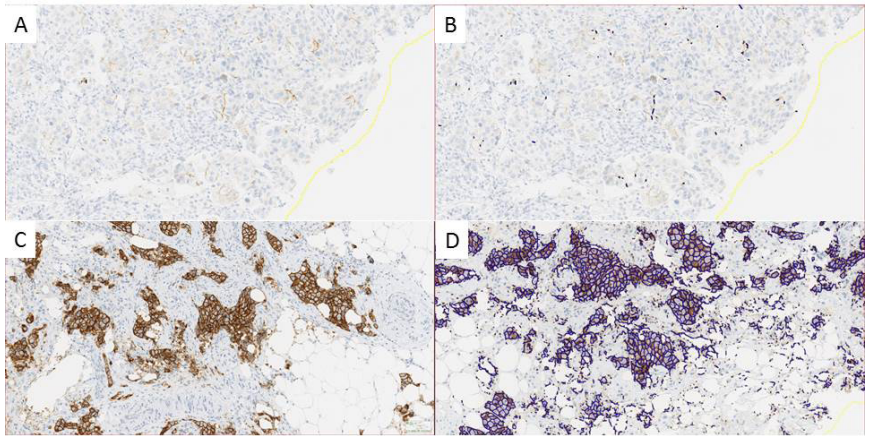
Background: The Visiopharm HER2 immunohistochemistry (IHC) algorithm evaluates cell membrane connectivity and the preliminary data have demonstrated accurate assessment of HER2 IHCs. We aimed to validate this algorithm for clinical use by comparing with pathologists' read and correlating with HER2 fluorescence in situ hybridization (FISH) results.

Design: The study cohort consisted of 213 invasive breast carcinoma specimens including 129 biopsies and 84 resections. HER2 IHC was performed (4B5, Ventana Medical Systems) and reported in accordance with the 2018 HER2 ASCO/CAP guidelines (0, 1+, 2+ and 3+). HER2 FISH utilized the dual-color Vysis FDA-approved PathVysion HER2 DNA Probe Kit (Abbott Molecular, Des Plaines, IL) and was interpreted using the 2018 HER2 ASCO/CAP guidelines. HER2 IHC slides were scanned using Philips IntelliSite Scanners and the digital images were analyzed using Visiopharm HER2-breast cancer App to obtain the connectivity values (0-1) and scores (0, 1+, 2+, and 3+). The conversion rules between connectivity and scores were defined as the following: score 0: connectivity = 0; score 1+: 0 < connectivity <= 0.1002 (Figure 1A and 1B); score 2+: 0.1002 < connectivity <= 0.49; score 3+: 0.49 < connectivity <= 1 (Figure 1C and 1D).

Results: The concordance between Visiopharm scores and pathologist scores was 82.6% (176/213). All discordant cases (n=37) were only one-step discordant (negative to equivocal/equivocal to positive, or vice versa) (Table 1). Twenty-four discordant cases had HER2 FISH results, which were re-interpreted using 2018 HER2 guidelines. The detailed FISH results from the 24 discordant cases were as follows: 15 pathologist negative/Visiopharm equivocal cases with 14 as FISH negative and 1 as FISH positive; 10 pathologist equivocal/Visiopharm negative cases with all as FISH negative; 1 pathologist equivocal/Visiopharm positive case as FISH positive; and 1 pathologist positive/Visiopharm equivocal case as FISH positive. Overall, one pathologist-IHC-negative case was FISH-positive, but none of Visiopharm-IHC-negative cases was FISH-positive.

		Visiopharm			
		negative (0/1+)	equivocal (2+)	positive (3+)	total
Pathologist	negative (0/1+)	106	18	0	124
	equivocal (2+)	11	32	1	44
	positive (3+)	0	7	38	45
	Total	117	57	39	213

Figure 1 - 1479



Conclusions: The HER2 IHC scores obtained by imaging analysis from Visiopharm HER2 algorithm demonstrate excellent concordance with pathologists' scores. Furthermore, Visiopharm HER2 algorithm discriminates between HER2 FISH positive and negative cases with high accuracy.

1480 Class Saliency Maps Reveal Computer Vision's Basis for Diagnosing Metastatic Carcinoma in Lymph Nodes

Nicholas Heller[1], Nikolaos Papanikolopoulos[1], Vassilios Morellas[1], Alexander Truskinovsky[2]
[1]University of Minnesota, Minneapolis, MN, [2]Roswell Park Cancer Institute, Buffalo, NY

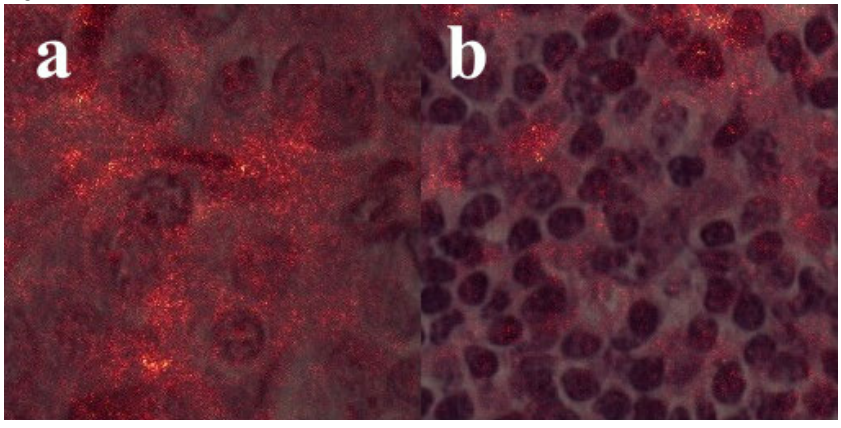
Disclosures: Nicholas Heller: None; Nikolaos Papanikolopoulos: None; Vassilios Morellas: None; Alexander Truskinovsky: None

Background: We and others have successfully applied computer vision to diagnosing a variety of malignant neoplasms in histopathologic images. Machine learning being an opaque process, little is known about the basis on which computer vision makes its diagnostic decisions in surgical pathology. Here, we use class saliency maps to determine which parts of the image the electronic classifier considers important for its diagnosis of carcinoma of the breast metastatic to lymph nodes.

Design: We extracted 32,027 benign and 22,560 malignant 256x256-pixel patches from the publicly available CAMELYON breast cancer lymph node metastasis dataset and classified them using a Deep Neural Network (DNN) model called DenseNet 201. The DNN was pretrained on the Imagenet dataset, a very large collection of natural images of everyday scenes. Only the final two layers (one convolutional, one dense) were trained for the CAMELYON recognition task. We trained the model at various magnifications and found that the 20x objective yielded the best results. To produce the visualizations, the logit corresponding to the correct class was differentiated with respect to the pixels in the input patch, and a Euclidean norm was computed over the three image channels. Therefore, the bright pixels in the heatmap may be interpreted as the pixels most likely to influence the model's decision if changed. These are called "Class Saliency Maps."

Results: A balanced diagnosis accuracy of 95% was achieved on the testing dataset, comparable to the accuracy of the best-performing previously reported classifiers. On the heatmaps of both malignant and benign patches diagnosed with high confidence, the bright pixels concentrated in the cytoplasm of the benign and malignant cells, avoiding the nuclei (Figure 1; a, malignant; b, benign).

Figure 1 - 1480



Conclusions: We used class saliency maps to see which parts of the image the electronic classifier considers important for its diagnosis of breast carcinoma metastatic to lymph nodes. To our great surprise, the pixels of highest diagnostic importance were concentrated in the cytoplasm, instead of the nucleus. Human pathologists mostly base their diagnosis of malignancy on the nuclear features, which we also teach to our trainees. It appears that computer vision bases its diagnostic decisions, at least in cases of nodal metastasis, on the cytoplasm, a part of the image that is harder for humans to evaluate than the nucleus. Whether this preference applies to other histopathologic settings remains to be determined.

1481 Predicting response to neoadjuvant chemotherapy in HER2-positive breast cancer using machine learning models with combined tissue imaging and clinical features

Zhi Huang[1], Zhi Han[2], Anil Parwani[3], Kun Huang[4], Zaibo Li[5]
[1]Purdue University, West Lafayette, IN, [2]Indiana University, Indianapolis, IN, [3]The Ohio State University, Columbus, OH, [4]Indiana University School of Medicine, Carmel, IN, [5]The Ohio State University Wexner Medical Center, Columbus, OH

Disclosures: Zhi Huang: None; Zhi Han: None; Anil Parwani: None; Kun Huang: None; Zaibo Li: None

Background: Pathologic complete response (pCR) to anti-HER2 neoadjuvant chemotherapy (NAC) is a presumptive surrogate for disease-free survival in patients with HER2 positive breast cancer (BC). Potential factors for achieving pCR have been investigated

separately. We aimed to develop a machine learning model with integrated information including tumor histology, checkpoint immune reaction and clinical features to predict the response to NAC.

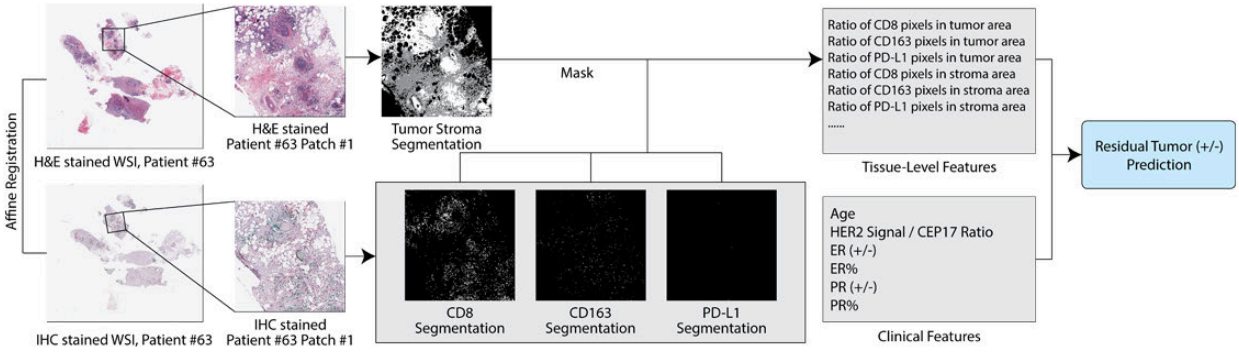
Design: A multi-color immunohistochemical (IHC) multiplex assay with PD-L1, CD8, and CD163 was performed on biopsy whole sections from 58 HER2-positive BCs treated with anti-HER2 NAC and subsequent resection. H&E and IHC slides were digitally scanned into whole slide images and 14 tissue-level histologic and immunostaining features were extracted and were combined with six clinical features (age, ER positivity/percentage, PR positivity/percentage and HER2/CEP 17 ratio) to predict response to NAC using machine learning models including logistic regression and neural network with LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge penalization, random forest classifier, AdaBoost classifier, GBM (Gradient Boosting classifier), and linear Support Vector Machine. Pipeline is shown in Figure 1.

Results: The data from all machine learning models demonstrated the tissue-level features extracted were well correlated with response to NAC and integrating clinical features further improved the prediction results. We performed 5-fold cross-validation on each dataset 10 times, and then took the average on each of them to find which dataset-method pair gives optimal prediction accuracy. The accuracy of predictions was measured with four different metrics [AUC (area under the ROC curve), F1-score, precision, and recall]. The detailed results were shown in Table 1 and the logistic regression with LASSO penalization outperformed other machine learning models. Finally, LASSO-regularized logistic regression model demonstrated that the top three features correlating with good response included high HER2/CEP17 ratio, high CD8 within tumor, and number of tumor pixels; while the top three features correlating with poor response included high CD163 in stroma, older age, and positive PR.

Table 1. Performance (%) of 5-fold cross-validation, averaged values from 10-times repeated experiments. AUC stands for Area under the ROC curve; F1 stands for F1-score; P and R stand for Precision and Recall, respectively. Bold texts indicate the best result across all methods in one measurement. L1, L2 stands for LASSO and Ridge regularization, respectively.

Methods	**Tissue-level features (14)**				**Clinical features (6)**				**Tissue-level + Clinical features (20)**			
	AUC	F1	P	R	AUC	F1	P	R	AUC	F1	P	R
Logistic Regression (L1)	78.43	**52.91**	58.55	**56.16**	**83.45**	**66.83**	**71.32**	**69.27**	**89.35**	**74.57**	**79.30**	**74.38**
Logistic Regression (L2)	79.66	50.85	61.40	51.63	80.53	53.05	62.36	51.72	88.19	72.23	77.01	73.16
SVM (Linear)	71.55	9.98	13.00	13.67	74.04	40.20	52.74	38.73	86.01	61.01	69.33	59.46
Random Forest	72.82	49.63	59.47	50.15	80.41	59.35	66.30	59.83	80.93	62.34	75.63	59.60
AdaBoost	65.66	40.78	48.12	41.75	72.13	53.94	56.18	59.05	75.92	57.42	66.48	57.49
GBM	68.02	45.48	52.18	47.85	78.90	63.98	70.64	63.98	75.33	54.43	65.10	54.44
Neural Network (L2)	**80.04**	51.84	**69.63**	48.90	76.99	41.68	56.30	38.55	83.80	59.93	74.07	54.87

Figure 1 - 1481



Conclusions: Our data demonstrated tissue-level features derived from H&E and IHC images provided rich informations which were well correlated with response to NAC. More importantly, integrating biomarker results into the imaging-based algorithm significantly improved its predictive power.

1482 An online Platform to Facilitate Stratification and Ongoing Quality Assurance of Tumor Infiltrating Lymphocytes (TILs) Evaluation in Immunotherapy-Trials

Jan Hudecek[1], Leonie Voorwerk[1], Jose van den Berg[1], Koen Van de Vijver[2], Marleen Kok[1], Roberto Salgado[3], Hugo Horlings[4]
[1]Netherlands Cancer Institute, Amsterdam, Netherlands, [2]University Hospital Ghent, Ghent, Belgium, [3]GZA-ZNA Ziekenhuizen, Antwerp, Belgium, [4]Antoni van Leeuwenhoek Hospital, Amsterdam, Netherlands

Disclosures: Jan Hudecek: *Major Shareholder*, Slide Score BV; Leonie Voorwerk: None; Jose van den Berg: None; Koen Van de Vijver: None; Marleen Kok: None; Roberto Salgado: None; Hugo Horlings: None

Background: The extent of stromal tumor infiltrating lymphocytes (TILs) have been shown to be a promising predictive biomarker of response to immune-checkpoint blockade. However, despite successful standardization efforts for the scoring of TILs, the implementation in daily practice and clinical trials is still lacking. Digital pathology allows quick evaluation and studying TILs and their spatial distribution.

Here, we present a web platform (Slide Score) that enables the assessment of morphological characteristics (e.g. TILs), percentages and intensities of stainings on whole slides images (WSI) or tissue microarray (TMA) cores in a structured manner. In addition, we demonstrate its use as a platform for evaluation of TILs in a phase 2 TONIC trial (NCT02499367), where a rapid evaluation of TILs is needed for patient stratification according to predefined cut off levels.

Design: A web-based viewer for manual scoring of WSI and TMAs was developed. We illustrate its use in the TONIC trial; adaptive phase II randomized non-comparative trial of nivolumab after induction treatment in triple-negative breast cancer patients. Hematoxylin and Eosin staining of a metastatic breast cancer lesion was scanned at 40x magnification. TILs were assessed by 4 pathologists from 3 institutions. After TILs evaluation, eligible patients were stratified by the extent of TILs in the low (< 5%) and high (> 5%) categories.

Results: Within the TONIC-trial, 25 biopsies of 23 patients, screened between April 2018 and October 2018, were assessed. All biopsies were scored by at least 3 pathologists within 48 hours. Median TILs for all 22 patients was 5%, with 10 patients in the TILs low and 12 patients in the TILs high category. Furthermore, the viewer was used in 25 research projects between December 2017 and September 2018. In total 5,821 whole slides and 1,248 TMAs were uploaded in the viewer. 45 international pathologists have used the system during this period to submit 251,924 answers - structured data points.

Conclusions: We developed a novel web platform to facilitate the fast assessment of both WSI and TMA cores by pathologists. Here, we demonstrate its feasibility as a tool for rapid stratification of patients in a clinical trial. Ongoing quality assurance in a clinical trial between pathologists is crucial whenever biomarkers are used for stratification of patients.

1483 Artificial Intelligence Predicts the Genetic Information of The Integrated Diagnosis of Brain Tumors

Yusuke Ishida[1], Masumi Tsuda[2], Jun Suzuka[2], Lei Wang[2], Satoshi Tanikawa[2], Hirokazu Sugino[3], Shinya Tanaka[2]
[1]Hokkaido University Faculty of Medicine, Sapporo, Japan, [2]Hokkaido University, Sapporo, Japan, [3]Hookaido University Graduate School of Medicine, Sapporo, Japan

Disclosures: Yusuke Ishida: None; Masumi Tsuda: None; Jun Suzuka: None; Lei Wang: None; Satoshi Tanikawa: None; Hirokazu Sugino: None; Shinya Tanaka: None

Background: The surgical and post-operative management of the central nervous system (CNS) tumors, varies depending on their histologic subtypes or gene informations. Distinguishing glial or non-glial tumor, especially high grade glioma or lymphoma is very difficult to differentiate. And if the tumor is high grade glioma, integrated diagnosis will be required for estimating the prognosis.

Our group used deep neural learning software with GPU accelerated PC, to predict the brain tumor integrated diagnosis without IHCs nor genetic information.

Design: We prepared HE slides of the brain tumors submitted to our faculty for diagnosis, gliomas and CNS lymphomas, diagnosed by one or more well-trained pathologists, using FFPE-HE and IHC (GFAP and/or CD20) slides, most of gliomas had information about gene sequencing for IDH1/2 mutation and 1p-19q codeletion by FISH.

Whole slide imaging (WSI) data, were scanned by NanoZoomer of Hamamatsu Photonics, Shizuoka, Japan.

To build those learning dataset, 256 pixel square JPEG image tiles, were automatically extracted by our software. Those images are directly transmitted to the machine learning system, Linux based PC with nVidia GPUs, and deep learning software.

Results: Our system classified brain tumor tile images, not used for teaching deep learning system dataset, with very high accuracy rates, glioma as 93% and lymphoma as 99%.

And also, AI predicts the gene information, IDH1/2 mutation as 60%, and 1p-19q codeletion as 80%.

Figure 1 - 1483

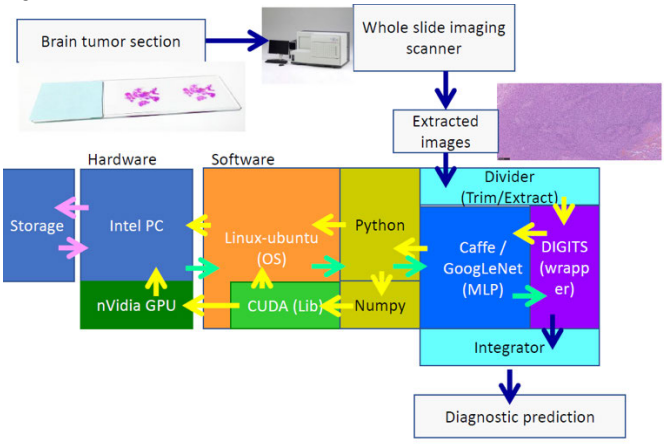
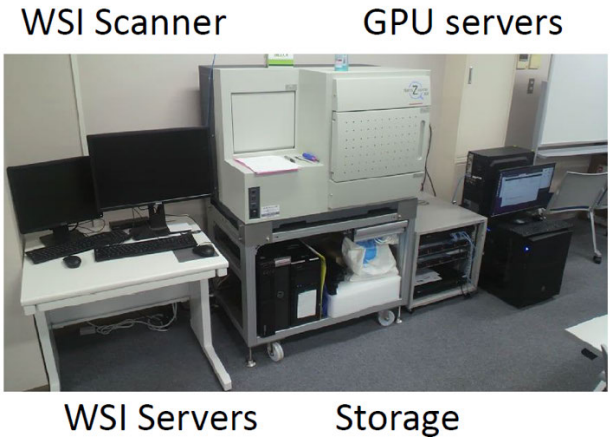


Figure 2 - 1483



Conclusions: If the pathologic laboratory of the faculty already has a WSI scanner for telepathology, additional cost is only for these GPU and installation, AI-based system will be the most effective aid for the brain tumor diagnosis as telepathology or consultation.

1484 A cognitive model of whole-slide image viewing and interpretation

Sebastian Jofre[1], Callan Powell[2], David Breen[2], Fernando Garcia[3], Mark Zarella[4]
[1]Drexel/Hahnemann University Hospital, Philadelphia, PA, [2]Drexel University, Philadelphia, PA, [3]Eastern Regional Medical Center/CTCA, Philadelphia, PA, [4]Drexel University College of Medicine, Philadelphia, PA

Disclosures: Sebastian Jofre: None; Fernando Garcia: None; Mark Zarella: None

Background: Pathology relies on the accurate interpretation of microscopic images to correctly diagnose patients and guide therapy. Methods to assess pathologist (and trainee) performance are in their infancy due to the inherent limitation that microscopes place in the ability to track a pathologist's gaze. With the advent of whole-slide imaging (WSI), however, pathologists have begun to analyze microscopic slides using computer monitors. Previous observations reveal that expert pathologists adopt distinct strategies to efficiently examine whole-slide images. Understanding how pathologists navigate whole-slide images has important implications on (1) optimization of WSI systems for efficient navigation and viewing, (2) development of histology-specific computer-assisted tools to reduce misdiagnosis, and (3) evaluation of viewing strategies to direct training in residents, fellows, and technicians.

Design: We examined mechanisms by which subjects extract meaningful information from histologic images. Based on decision theory, we developed a cognitive model designed to parse the factors that ultimately contribute to diagnosis. With the aid of eye tracking and ideal observer analysis, we tested the model as subjects determined histologic grade of invasive breast carcinoma from whole-slide images.

Results: We parsed whole-slide viewing into a set of quantifiable variables that include interpretation accuracy, low magnification acuity, decision threshold, and bias. Together, these variables formed an equation that predicted the gaze and navigation patterns that contribute to histologic grading. Modeling revealed that modulating these variables from an ideal observer led to predictable impacts in performance. When subjects were constrained to only two magnifications of viewing (0.25x and 5x magnification) while grading breast cancer images, their performance and gaze patterns were consistent with the model. By artificially manipulating images and task, we independently measured the variables in our model in subjects. We found that experts and novices exhibited stereotypical differences in all four variables, suggesting that a quantitative model of whole-slide viewing can distinguish expert from novice performance and can identify the areas in which they may differ.

Conclusions: Our results not only demonstrate a paradigm that can be used to monitor training outcomes in pathology residents, but also provide a pathway for optimizing computer-aided diagnostic algorithms in a human behavior-inspired manner.

1485 Data Mining and Visualization Solution to Satisfy CAP Requirements for Breast IHC Benchmark Comparison

David Joseph[1], Simone Arvisais-Anhalt[2], Helena Hwang[3], Yisheng Fang[4], Yan Peng[4], Katja Gwin[4], Venetia Sarode[3], Sunati Sahoo[2], Ellen Araj[2]
[1]The University of Texas at Southwestern, Plano, TX, [2]University of Texas Southwestern Medical Center, Dallas, TX, [3]University of Texas Southwestern, Dallas, TX, [4]Dallas, TX

Disclosures: David Joseph: None; Simone Arvisais-Anhalt: None; Helena Hwang: None; Yisheng Fang: None; Yan Peng: None; Katja Gwin: None; Sunati Sahoo: None; Ellen Araj: None

Background: The College of American Pathologists (CAP) requires laboratories to compare immunohistochemical (IHC) and in situ hybridization (ISH) test results that provide independent predictive information to published benchmarks. This allows laboratories to track their testing validity. Estrogen receptor (ER) IHC, progesterone receptor (PR) IHC, and HER2 IHC and ISH testing provide important breast cancer diagnostic and prognostic information. The need arose to develop a computational solution to assist in tracking ER, PR, and Her2 results from our laboratory as the historic pathology reports at our institution are stored in a non-discrete, free-text format.

Design: We designed a custom structured query language (SQL) script to search anatomic pathology free text reports stored in our institution's clinical relational database for the following variables: case type (biopsy or mastectomy), type of tumor (invasive, in-situ, and/or metastatic disease), and ER, PR, and HER2 IHC and ISH results (positive/negative/equivocal, % positive, intensity). We applied additional filters to record IHC/ISH results for the patient's first biopsy with invasive carcinoma and exclude IHC/ISH results for metastasis and mastectomies. We then created a dashboard to visualize the data using the business intelligence tool, Tableau. The dashboard enables users to quickly check the variables extracted from the free text reports for accuracy and exclude incorrect data.

Results: We queried over two years of anatomic pathology reports with n=83,501 reports. After we applied filters via our custom algorithm, it took a pathologist an hour of hand cleaning. The final number of testing results recorded were ER IHC n=304, PR IHC n=301, Her2 IHC n=290, and Her2 ISH n=13.

Conclusions: This solution allowed us to quickly extract a high-quality, representative sampling of our laboratory's breast IHC/IHC results from free text reports to satisfy the CAP requirement. The dashboard also allowed us to stratify testing results based on pathologist to determine whether there were practice differences. This computational solution satisfied our needs and can be easily and quickly reproduced at other institutions.

1486 Machine Learning-Based Natural Language Processing for Automated Extraction and Standardized Annotation of IHC Results from Free Text Pathology Reports

Young Suk Kim[1], Michael H. A. Roehrl[2]
[1]Memorial Sloan Kettering Cancer Center, New York, NY, [2]Scarsdale, NY

Disclosures: Young Suk Kim: None; Michael H. A. Roehrl: None

Background: Natural language processing (NLP) is a subfield of computer science that combines Artificial Intelligence (AI) and other Machine Learning (ML) disciplines. Here, we extend our previously validated NLP work to automatically extract meaningful information from real-world unstructured and narrative pathology reports. We extract key standardized information about tumors as measured by immunohistochemistry (IHC), specials stains, or molecular tests such as FISH.

Design: We used various NLP methods such as multiple cTAKES (clinical Text Analysis and Knowledge Extraction System; Mayo) components, word2vec (Google), as well as our large cancer center in-house Application Programming Interfaces (API). To get optimal ground truth data, we manually encoded free-text reported IHC results, special stains, and FISH results in a standardized manner (comprising >130 individual test types). We pre-processed raw free-text datasets to make our input more uniform and to improve performance with unknown tokens such as abbreviation, acronyms, or ambiguous words. Pre-processed datasets enhance the classifier outcomes when using ML and NLP analysis tools including CoreNLP (Stanford), NLTK, scikit-learn, and other text analysis tools. We compared several standard supervised classifiers using k-fold cross validation.

Results: We processed surgical pathology reports (36,431) from a two-year period and added additional data (8,174) from another year to assess whether our model works well with unseen data. We evaluated different supervised classifiers including SVM (Support Vector Machine), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and K-Nearest Neighbors (KNN) to compare ML accuracy. We observed that a SVM has the highest accuracy score for most stains. We show that our NLP approach is able to automatically transform virtually any free-text resulted stain into structured quantitative data (e.g., overall positive/negative, intensity (weak/moderate/strong), % staining, heterogeneity (uniform, focal, scattered), cellular localization (membranous, nuclear, cytoplasmic), and tissue compartment (tumor, inflammatory, stromal, etc.)).

Conclusions: We developed a set of NLP tools for automated stain result annotation from narrative free-text pathology reports. Our NLP software can find large-scale IHC protein expression patterns in tumors, drive global biomarker discovery, and guide practicing pathologists in optimal diagnostic test ordering.

1487 Smart Glass Technology in Surgical Pathology

Ozlem Kulak[1], Anastasia Drobysheva[2], Neda Wick[3], Simone Arvisais-Anhalt[3], Charles Timmons[4], Jason Park[5]
[1]UT Southwestern Medical Center, Dallas, TX, [2]Office of Chief Medical Examiner of the City of New York, New York, NY, [3]University of Texas Southwestern Medical Center, Dallas, TX, [4]Children's Medical Center, Dallas, TX, [5]Dallas, TX

Disclosures: Ozlem Kulak: None; Anastasia Drobysheva: None; Neda Wick: None; Simone Arvisais-Anhalt: None; Charles Timmons: None; Jason Park: None

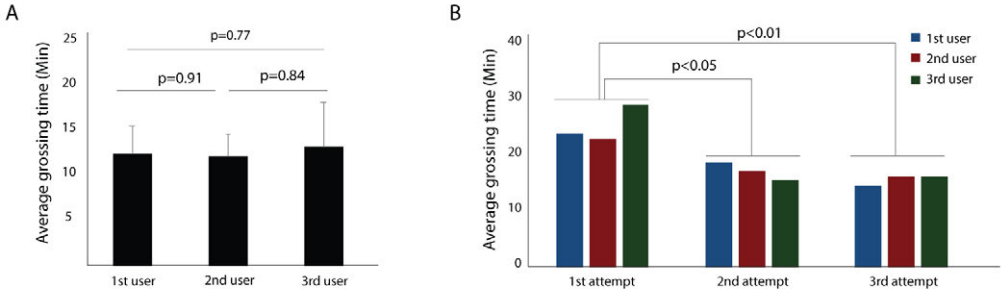
Background: Smart glasses are a wearable technology that provide hands-free access to communication and productivity tools. This technology may be useful in surgical pathology for improving access to information such as grossing manuals and medical records as well as providing data recording such as still-pictures and movies.

Design: A digital surgical pathology grossing application was developed in a logistics software application (Proceedix, Belgium) for use on the Google Glass Enterprise (Mountain View, CA) smart glass platform. The application prompts key steps in the grossing procedure as well as hands-free measurement and photo documentation. All procedures are initiated by voice command and are automatically timed and recorded. The application on Google Glass Enterprise was tested by three users dissecting lamb hearts.

Results: The total heart grossing time for each user averaged 17:34 min (+/- 1:25), 17:10 min (+/- 1:07), and 18:35 min (+/- 2:19) for the 3 users, respectively (difference between users is not significant, p>0.05) (Fig 1A). When grossing times were organized by the first, second and third organ grossing attempt, the average time per attempt was 23:13 min (+/-1:00), 15:55 min (+/- 0:23), and 14:10 min (+/-0:13) for the 3 users, respectively (difference between first and repeated attempts are significant, p<0.05) (Fig 1B). The first to last attempt of the protocol led to a significant reduction in grossing time for each user. In addition to measuring the total time of the protocol performance, each sub-step within the protocol was automatically recorded. The step which took the longest time for each user was measurement of the valves and wall thickness (average of 4:22 min +/-0:25, 5:42 min +/-0:31, 4:47 +/-0:55) for the 3 users, respectively (difference between first and second users was significant p=0.03; difference between first and third users (p=0.52) , and the difference between second and third users are not significant (p=0.21).

NA

Figure 1 - 1487



Conclusions: Smart glasses are a hands-free platform for data retrieval and recording. An additional benefit of smart glasses is the passive information gathered for workflow management. Workflow timing is useful to document the relative efficiency of each user's performance. This proof-of-concept demonstrates a gross application for smart glasses in the pathology laboratory.

1488 Digital Telepathologic Diagnosis, A Practical Substitute of Pathological Service for Community Hospitals that Lack of Pathologists Base upon Analysis of 27,113 Digital Pathological Cases

Jing Li[1], Pifu Luo[2], Zhikui Zhang[1], Xiangdong Ding[1], Tao Wu[1]
[1]KingMed Diagnostics, Guangzhou, China, [2]KingMed Diagnostics, Central Point, OR

Disclosures: Jing Li: None; Pifu Luo: None; Tao Wu: None

Background: Due to a significant short of pathologists in China, a digital telepathological practice (DTP), as a new diagnostic mode in Guangzhou Kingmed Diagnostics (GKD), has been implemented and practiced since 2012 to provide pathological service for 42 remote community hospitals. This study is to summarize quality data and experience of DTP from January 2012 to June 2018.

Design: H&E-Stained glass sections of all cases were firstly scanned into digital slides with scanners. Pathologists at GKD opened files of the digital slides, viewed the digital images, yielded the diagnosis, and issued an electronic report and/or communicated with clinicians.

Results: A total of 27113 cases were completed using DTP from January 2012 to June 2018, including 17319 routine cases, 6682 cases of frozen sections diagnosis, 2664 quality assurance cases and 448 consultation cases. For the routine diagnosis, 84.6% (14652/17319) of cases were completed in less than 3 days and 98.1% (16990/17319) in less than 5 days. Average turnaround time (TAT) of frozen sections diagnosis was 35.7 minutes, in which, TAT was 24 minutes for the single block cases. Concurrence rate of diagnoses between frozen sections and permanent sections was 99.31%, and majority of discrenpcy cases were the thyroid papillary microcarcinoma due to inappropriate sampling. Among 2664 of quality-assurance cases, concurrence rate was 71.5% of (1906/2664), and a total discrepancy rate was 28.5% (758/2664), in which the major discrepancy rate was 3.9% (103/2664). After the quality-assurance process, the diagnoses were revised before sign-out to avoid clinical impaction to the patients. Average TAT of the consultation cases was 47.5 hours, in which, 55% (246/448) of cases were completed within 24 hours, 89% (399/448) cases were completed within 3 days.

Conclusions: This study and analysis have shown that DTP is a valuable, practical and satisfactory pathologic service mode for the community hospitals. DTP provides a reliable and accountable way to meet a diagnostic need, and It is a significant substitute of pathologic service for the rural community hospitals that lack of pathologists. The diagnostic accuracy and TAT can be further improved by strictly following a standardization of Procedures (SOP), including a sufficient tissue sampling and submission, a high quality of H&E slide preparation, personnel education and training.

1489 Using Computer Vision to Diagnose Non-Neoplastic Pulmonary Diseases

Xinyan Li[1], Juan Manivel[2], Jose Jessurun[3], Nikolaos Papanikolopoulos[1], Vassilios Morellas[1], Alexander Truskinovsky[4]
[1]University of Minnesota, Minneapolis, MN, [2]Minneapolis Veterans Administration Med Center, Minneapolis, MN, [3]New York Presbyterian Hospital, New York, NY, [4]Roswell Park Cancer Institute, Buffalo, NY

Disclosures: Xinyan Li: None; Jose Jessurun: None; Nikolaos Papanikolopoulos: None; Vassilios Morellas: None; Alexander Truskinovsky: None

Background: Having successfully used computer vision to diagnose malignant and benign neoplasms of the endometrium, prostate, breast, ovary and myometrium, squamous intraepithelial lesions of the lower female genital tract, and metastatic tumors in lymph nodes, we now turn away from neoplasia and apply computer vision to histopathologic diagnosis of non-neoplastic pulmonary diseases.

Design: We used 19 H&E-stained images, digitally scanned at x50 magnification, 5 each of hypersensitivity pneumonitis (HP), usual interstitial pneumonia (UIP), and normal lung, and 4 of cystic fibrosis (CF). HP and UIP represent interstitial pulmonary diseases, and CF, as bronchiectasis, represents airway diseases. Images were annotated using the special GUI we had developed, and sliced into a total of over 20,000 usable 32x32-pixel blocks. HP having the lowest number of blocks, to avoid classifier bias, we randomly selected equal numbers of samples from other classes for a balanced dataset of 1,920 samples. We applied K-nearest neighbor classification using region covariance features, focusing on texture features, and selected I (intensity of the image), Ix and Iy (gradient of the image along x and y axis), $\sqrt{(I_x^2 + I_y^2)}$ (its magnitude), and second-order terms Ixx, Iyy and Ixy, producing a smaller feature space and accelerating analysis, while maintaining the high quality. A 10-fold cross-validation tested the performance of the classifier. One-tenth of the samples was used as test samples, leaving the rest to build the database. Iterating from 1 to 25 nearest neighbors, we found that the k of 5 nearest neighbors performed the best.

Results: The overall accuracy of distinguishing all four classes was 87.6%. CF led to the most mistakes, being sometimes confused with HP and normal lung. UIP was diagnosed with the greatest accuracy. When CF was removed from the analysis, the resulting accuracy of differentiating HP, UIP and normal lung was 96%.

Conclusions: To our knowledge, no one has yet attempted to diagnose non-neoplastic lung diseases in histopathologic images by computer vision. Non-neoplastic pulmonary pathology is one of the hardest fields of surgical pathology, requiring high expertise, with a fairly low reported interobserver agreement of diagnosis. We have shown that computer vision can diagnose non-neoplastic lung diseases,

especially interstitial diseases, with high accuracy, although airway disorders can rarely be confused with normal lung and other conditions that have areas of normal alveolate parenchyma.

1490 The Development and Validation of an Automated Tool for Ki67 Assessment of Neuroendocrine Tumors

Wen-Yih Liang[1], Yu-An Chen[2], Anna Li[1], Woei-Chyn Chu[3]
[1]Taipei Veterans General Hospital, Taipei, Taiwan, [2]National Yang-Ming University, Taipei, Taiwan, [3]National Yang Ming University, Taipei, Taiwan

Disclosures: Wen-Yih Liang: None; Yu-An Chen: None; Anna Li: None; Woei-Chyn Chu: None

Background: Digital image analysis is repeatedly proposed for the assessment of Ki67 proliferative index of neuroendocrine tumors recently. Unfortunately, distinguishing tumor from non-tumor is regarded as an obstacle for digital image analysis, which usually makes it over-estimation on Ki67 proliferative index. We develop and evaluate an automated analysis program as well as an algorithm to deal with digital images of synaptophysin-Ki67 double stained slides which may solve the problem.

Design: A total of 27 cases diagnosed as neuroendocrine tumor in gastrointestinal system were retrieved from the archive of a single medical center from 2010 to 2017. Ki67 along and synaptophysin-Ki67 double immunohistochemical stains were performed, The hotspot images of each case were taken after viewing the whole double stain slide with all three participated pathologists' agreement. The Ki67 index of each case in three pathologists using eyeball estimation (EE) in glass slide of Ki67-only stained slide, synaptophysin-Ki67 double stained slides and eye-counting of pre-captured synaptophysin-Ki67 double stained slides images were recorded.

The Ki67 index of hotspot images were measured by our self-developed program which can automatically detect the Ki67 positive and negative tumor cells in the synaptophysin-positive area (Digital image analysis, DIA) and modify the result through the embedded graphic user interface (Computer-aided assessment, CAA) as Gold standard. (Figure 1)

The intraclass correlation coefficient (ICC) and Cohen's kappa coefficient (κ) between different groups were calculated and analyzed.

Results: The outcomes made by our program (DIA) showed high agreement (ICC: 0.99, 95% CI: 0.99-1.00) with CAA. And it's significantly higher than those assessed by eyeball estimation (EE) of Ki67-only stained slides (ICC: 0.60, 95% CI: 0.45-0.73), by EE of synaptophysin-Ki67 double stained slides (ICC: 0.81, 95% CI: 0.72-0.87), and by eye-counting of pre-captured synaptophysin-Ki67 double stained slides images (ICC: 0.81, 95% CI: 0.72-0.87). (Figure 2) As for the accuracy of single-cell detection, the sensitivities of Ki67-positive and -negative tumor cells are 0.9819 and 0.9552, while only 0.11% of false alarm toward non-tumor tissue.

Figure 1 - 1490

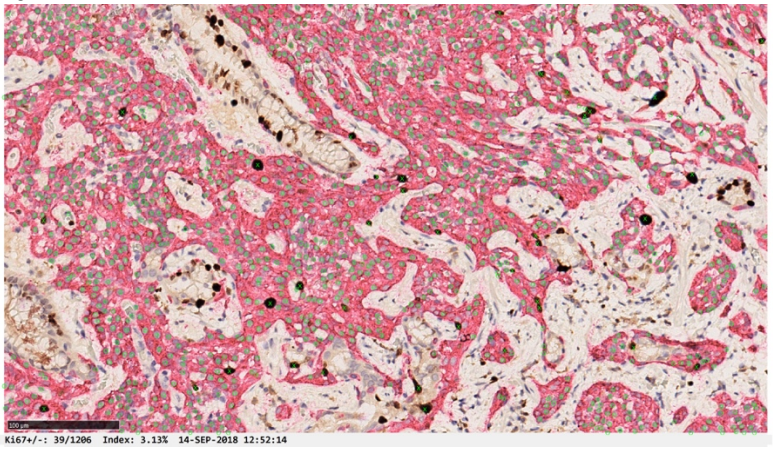
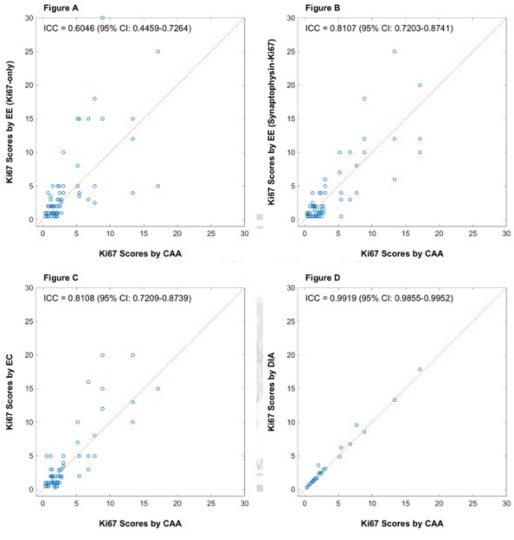


Figure 2 - 1490



Conclusions: The automated analysis program proposed in this study show high agreement with professional clinical opinion, and was proved to be beneficial for assisting in clinical work of Ki67 assessment of neuroendocrine tumors.

1491 Automated Identification of Mitoses using A Hybrid Approach: Combining Deep Learning with Classical Domain-Based Detection

Mark Lloyd[1], David Harding[1], Nishant Verma[1], Amir Mohammadi[2], James Monaco[1], Gary Tozbikian[3], Zaibo Li[3], Anil Parwani[4]
[1]Inspirata, Inc., Tampa, FL, [2]Inspira, Inc., Tampa, FL, [3]The Ohio State University Wexner Medical Center, Columbus, OH, [4]The Ohio State University, Columbus, OH

Disclosures: Mark Lloyd: *Employee*, Inspirata Inc.; David Harding: *Employee*, Inspirata, Inc.; Gary Tozbikian: None; Zaibo Li: None; Anil Parwani: None

Background: Measures of mitotic activity are highly prognostic of survival times in patients with node-negative breast cancers. However, identifying individual mitotic figures can be a time-consuming task for pathologists and may be subject to significant inter-observer variability. Adoption of digital pathology allows the development of automated image analysis algorithms to assist pathologists in efficiently and reliably identifying mitotic figures.

Recent major advances in performance of automated image analysis systems have been seen due to the application of machine learning techniques such as deep learning to these problems. These deep learning systems do not directly embed any of the collective domain knowledge of pathologists in their design, instead finding features in the image data directly through an optimization process.

We combined a traditional system using domain-specific features with a modern deep image segmentation network to see if this hybrid system could outperform them both.

Design: The deep image segmentation network we applied was Retinanet (Facebook AI Research, Menlo Park, CA). The traditional system and hybrid systems were implemented and evaluated using MATLAB 2016b (Natick, MA). The digitized image and annotation data was made available by the AMIDA13 MICCAI Grand Challenge and was digitized on an Aperio ScanScope XT at the University Medical Center Utrecht (Utrecht, Netherlands).

Each of our systems was individually trained on the training subset of the AMIDA13 data. The mitotic figures detected by each system were combined and reclassified using the confidence values reported by both systems. Performance data was gathered over the disjoint testing subset of the AMIDA13 data.

Results: As expected, the deep learning system outperformed the traditional system over this homogeneous challenge dataset. However, when operating over a range of specificities, we found sensitivity improve by 10% by combining the two systems. Specifically, at a rate of .25 false detections per true positive, the traditional, deep learning, and hybrid systems' sensitivities were 38%, 48%, and 53%, respectively.

Conclusions: We find that there is information captured by the domain-specific features of the traditional system that is not represented in the deep learning system. A combination of the two techniques outperformed either system individually.

1492 Optimizing Clinical Scanning Throughput in a Single Institution's Central Pathology Whole Slide Imaging Facility: Lessons Learnt

Mark Lloyd[1], David Kellough[1], Trina Shanks[2], Anil Parwani[3]
[1]Inspirata, Inc., Tampa, FL, [2]Inspirata, Inc., New Albany, OH, [3]The Ohio State University, Columbus, OH

Disclosures: Mark Lloyd: *Employee*, Inspirata Inc.; David Kellough: *Employee*, Inspirata Inc.; Trina Shanks: None; Anil Parwani: None

Background: By harnessing the value of a quantifiable digital pathology data modality through deep learning and similar computational advancements, it is reasonable that we can uncover morphological signatures occurring within tumors and their surrounding microenvironment to provide better clinically actionable insights. To facilitate meaningful artificial intelligence, large data sets are required. Our group is using process excellence to maximize the number of whole slide scans which can be acquired rapidly in the world's largest throughput pathology imaging facility.

Design: Our group has scanned over 800,000 slides in just 16 months. In order to accomplish this, we installed a bank of seven (7) high-throughput scanners and designed the initiative planning, the layout of the dedicated facility, SOPs, project governance, workflows, staffing, communication and stakeholder engagement activities, system integrations, file storage and IT support.

Results: Now we have clear metrics regarding the results of this initiative. We have scanned 800,000 slides at the time of this submission. We run three shifts of slide scanning technicians and operate the facility 24 hours a day, five (5) days a week. We have scanned over 3 years of retrospective cancer cases. Since our go-live date we have begun a follow-up process excellence review to enhance the following processes and increase our scanning volume: lab layout (storage and labeling procedures); slide touches (>2,400/day), slide inspection and cleaning procedures (>30sec/slide), scan error rate reduction (>2%) and QC reviews at the scanner and

for the WSI (>3%). We have run two (2) Process Excellence exercises to continue to increase our throughput by using LEAN and Six Sigma methods.

Our next phase of this project is primary diagnosis. We rendered the first primary diagnosis in the U.S. in March and have signed out over 1,500 cases in the first 6 months.

Conclusions: The promise of computational pathology is growing. It will require large data sets. In order to create large data sets, it is clear that production facilities will be needed. Our high-throughput facility is the first to generate a massive pipeline of image data has not previously been achieved at this scale. We are the first organization to create a repository of WSIs, to be linked with other relevant deidentified patient information, to be used to enable and accelerate these activities.

1493 Tumor Identification and Grading on Histopathology Images Using Deep Learning

Shachi Mittal[1], Andre Balla[2], Luke Pfister[3], Catalin Stoean[4], Kianoush Falahkheirkhah[3], Rohit Bhargava[3]
[1]University of Illinois at Urbana-Champaign, Champaign, IL, [2]University of Illinois, Chicago, IL, [3]University of Illinois at Urbana-Champaign, Urbana, IL, [4]University of Craiova, Craiova, Romania

Disclosures: Shachi Mittal: None; Andre Balla: None; Luke Pfister: None; Catalin Stoean: None; Kianoush Falahkheirkhah: None; Rohit Bhargava: None

Background: Current histopathological techniques rely on structural features of malignancy to identify patterns indicative of the diseased state. Although there has been significant progress in using immunohistochemistry for molecular characterization, simultaneous evaluation of tumor and its microenvironment is difficult. Automated pattern recognition can help identify global discriminative features across different tissue entities by using large datasets. This study includes the use of conventional stained images and infrared (IR) images towards an integrated diagnostic approach. Infrared imaging captures the molecular signature of the tissue as the vibrational modes of different molecular bonds lie in the mid-infrared spectrum of light along with the spatial detail offered by optimal microscopy.

Design: A large cohort of tumor microarrays and biopsy slides with a spectrum of usual ductal hyperplasia, atypical hyperplasia, DCIS and invasive tumors are analyzed using deep learning models for comprehensive tumor characterization. First, existing neural networks are used for feature extraction followed by machine learning to classify tumor regions. Second, a separate neural network is trained for multimodal image analysis. Both these models not only detect the diseased state but can also be used for determining tumor garde. Model performance will be tested based on the following two metrics: a) Spatial agreement with pathologist annotations and b) Receiver operating characteristic curves that illustrate the sensitivity and specificity of the diagnostic model.

Results: First, clustering methods were developed for H&E stained images to identify disease signatures in the epithelial and stromal regions. It can be seen in figure 1, zoom (i) that the model identifies two different stromal signatures indicative of normal stroma (in blue) and desmoplastic reaction (in red). Further, multiple signatures are identified in the epithelial domain. The model will then be extended to supervised identification of disease and prognostic markers.

Figure 1 - 1493

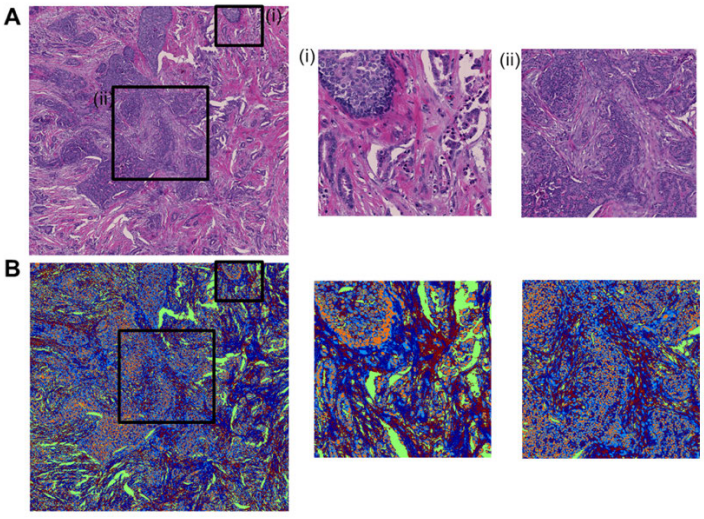


Figure 1. Segmentation of H&E stained images using K-means. A.H&E stained image with areas of interest shown in boxes. B. The corresponding clustered image along with the regions of interest.

Conclusions: The hotspots identified by the deep learning model will help improve diagnostic accuracy. This could also assist pathologists in borderline cases or differential diagnosis such as distinguishing between usual ductal hyperplasia and low grade ductal carcinoma in-situ (DCIS).

1494 Deep Learning Networks to Segment and Detect Breast Terminal Duct Lobular Units, Acini, and Adipose Tissue: A Step Toward the Automated Analysis of Lobular Involution as a Marker for Breast Cancer Risk

Allison Onken[1], Suzanne Wetstein[2], Michael Pyle[3], Josien Pluim[2], Stuart Schnitt[4], Gabrielle Baker[1], Laura Collins[1], Rulla Tamimi[5], Mitko Veta[2], Yujing Jan Heng[1]
[1]Beth Israel Deaconess Medical Center, Boston, MA, [2]Eindhoven University of Technology, Eindhoven, Netherlands, [3]Beth Israel Deaconess Medical Center, Watertown, MA, [4]Brigham and Women's Hospital, Boston, MA, [5]Brigham and Women's Hospital, West Newton, MA

Disclosures: Allison Onken: None; Suzanne Wetstein: None; Michael Pyle: None; Josien Pluim: None; Stuart Schnitt: None; Gabrielle Baker: None; Laura Collins: None; Rulla Tamimi: None; Mitko Veta: None; Yujing Jan Heng: None

Background: Terminal duct lobular unit (TDLU) involution is the physiological process whereby Type 2 and 3 lobules revert to Type 1 after child-bearing years. TDLU involution (quantitatively assessed by TDLU count per mm2, TDLU span, and acini count per TDLU) is inversely associated with breast cancer risk. The manual assessment of involution is time-consuming and subjective, making it impractical to perform on large epidemiological studies. Deep learning algorithms such as convolutional neural networks (CNNs) could be utilized for rapid and automated assessment of TDLU involution. We designed two CNNs to segment TDLUs and detect acini as the first step toward large-scale assessment of TDLU involution, and a third CNN to segment adipose tissue.

Design: Whole slide images (WSIs; *n*=50) were obtained from the Nurses' Health Study Incident Benign Breast Disease Study. For each WSI, TDLUs, acini, and adipose tissue were annotated within a region of interest comprising approximately 10% of the total tissue area. In order to assess involution in histologically normal breast parenchyma only, TDLUs with proliferative or metaplastic changes were excluded from manual evaluation. CNNs were engineered to recognize TDLUs, acini, and adipose tissue using 60% of the WSIs for training, 20% as a test set, and 20% for validation. F1 and Dice scores were calculated as accuracy measures to compare CNN segmentation to manual assessment.

Results: Our CNNs detected acini, segmented TDLUs, and segmented adipose tissue with accuracy measures of 0.73, 0.84, and 0.86, respectively. Two primary causes of discordance with manual assessment were identified: 1) complex clustering of TDLUs where our CNN had difficulty predicting TDLU boundaries and 2) acini with proliferative or metaplastic changes which our CNN frequently detected as acini but which were intentionally excluded from manual annotation.

Conclusions: We have developed a series of deep learning networks to segment and detect TDLUs, acini, and adipose tissue on WSIs. With accuracy measures of >0.7, our CNNs are sufficiently robust to be integrated into a computational pipeline for automated assessment of the quantitative features of TDLU involution, and will be further refined to address sources of discordance with manual assessment. This is the first step toward the large-scale quantification of TDLU involution which, when applied to patient samples, could be used to better determine the breast cancer risk associated with lobule type and degree of involution.

1495 Whole Slide Scanner and Anatomic Pathology Laboratory Information System Integration with HL7 interface to Support Digital Pathology Sign-out Workflow

Yeon Park[1], Wendy Frankel[1], Keith Mullins[1], Alexander Fitzthum[2], Zaibo Li[1], Anil Parwani[3]
[1]The Ohio State University Wexner Medical Center, Columbus, OH, [2]The Ohio State University Wexner Medical Center, Columbus, Ohio, [3]The Ohio State University, Columbus, OH

Disclosures: Yeon Park: None; Wendy Frankel: None; Alexander Fitzthum: None; Zaibo Li: None; Anil Parwani: None

Background: Recent advancements in the field of digital pathology allow us to utilize new systems and workflows that save time for pathologists and minimize errors in matching slides. One of the barriers to adopting digital pathology systems (DPS) is the task of integrating them with the systemwide anatomical pathology laboratory information systems (APLIS) already in place. A DPS integrated with the APLIS at our institution was recently implemented. We discuss the methodology and benefits of such a system in the context of signing out anatomical pathology cases.

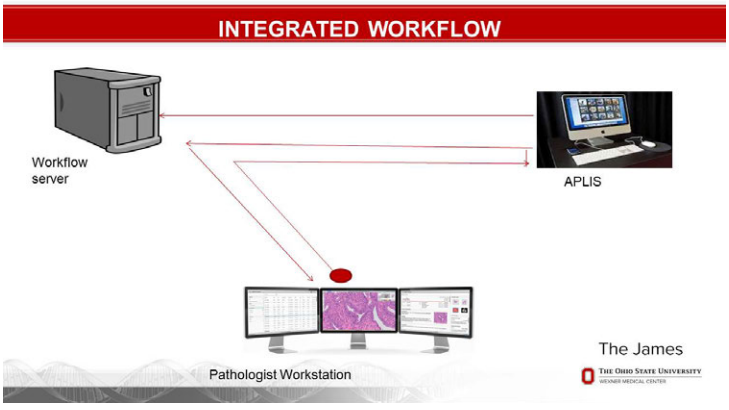
Design: Cases enter the workflow when added to the APLIS (Sunquest Co-Path). The scanner (Phillips IntelliSite Pathology Ultra-Fast Scanner) digitizes the glass slides from the cases and automatically uploads them to the DPS (Phillips IntelliSite Pathology Solution version 3.2.1). The slides are all uploaded with barcode labels encoded with case-specific information. The APLIS sends corresponding case data through a web server to the DPS. The DPS uses this information to automatically match the slides with the correct cases from the APLIS. This also allows for APLIS data, such as prior cases, to be accessed through the DPS when pathologists remotely open a case.

Results: A robust HL7 interface was built that allowed for orders to pass from the laboratory information system to image management system (Figure 1). The interface was unidirectional but allowed for automation of the signout process, enabling accessioned cases to appear in the pathologist workstation in real time with an integrated workflow where cases flowed from APLIS to the workflow server and to the pathologist workstation (Figure 2). Since implementing the new DPS system at our institution, more than 1,250 surgical pathology cases have been signed out. The addition of the DPS to our workflow led to the major benefit of allowing pathologists to review case slides while simultaneously looking at relevant case data pulled from the APLIS.

Figure 1 - 1495

Date In: 3/14/2017 11:01:23 **Status:** ok **Date Processed:**
3/14/2017 11:01:34 **Sequential Number:** 100000158610
MSH|^~\&|SQ|SQTUCSON|COPATH|DADD1|20170314110123||O
RR^O02|170390001839|P|2.2|
PID|||410000731||||||||||||||||
PV1||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ORC|NA||S17-27^CoPath C/S|
OBR||9921727|S17-27^CoPath C/S|

Figure 2 - 1495



Conclusions: Integration whole slide scanners with our APLIS was necessary to form the new digital workflow at our institution. This new workflow for pathology sign-out limited potential human error by automating the slide matching and delivering process, saved time in slide delivery, and improved image sharing and review for pathologists. Adoption of digital slide review is contingent upon successful APLIS integration with WSI systems.

1496 Data extraction in oral cancer pathology using text mining

Paromita Roy[1], Indranil Mallick[1]
[1]Tata Medical Center, Kolkata, India

Disclosures: Paromita Roy: None; Indranil Mallick: None

Background: Extraction of clinically useful data from pathology reports has traditionally been a challenge. Text mining with the use of software toolkits is being increasing used, but has met with only partial success. We aimed to extract pathological parameters from biopsy and operative pathology specimens of oral tongue primaries using data analysis and natural language toolkits.

Design: Pathology reports from the departmental database were extracted with the search terms of 'tongue' and 'squamous'. Records were searched for category of differentiation (well, moderate, or poor); perineural invasion (present or absent); lymphovascular invasion (present or absent); maximum dimension of the tumor (in cm), maximum depth (in cm), closest margin (in cm) and status of margins (free, close or involved). The text blocks were submitted for mining using the python 3.6 programming language using the open source libraries for data analysis ('pandas'), and natural language processing ('nltk'). Both rule based and simple machine learning algorithms were used (Naïve Bayes Classification). The domain expert was consulted during development of the algorithm. Results were checked by the pathologist.

Results: Records of 351 patients with oral tongue cancer were identified (biopsy = 14, operative histopathology = 337). Of the 1970 parameters available in the set of reports 1968 were identified by the program. The accuracy in identification was 98.3% for differentiation, 99.7% for perineural and lymphovascular invasion, 99.6% for the maximum dimension, 98.5% for depth of invasion, 94.7% for the closest margin, and 99.4% for the margin classification. For binary classification variables, the precision, recall and F1 score was 1, 0.99 and 0.99 for both perineural and lymphovascular invasion. Errors mostly related to spelling and units used in the text reports. Closest margin assessment was more challenging than other features in view of margin descriptions spanning several complexly framed sentences.

Conclusions: This tested instance of text mining using data analysis and natural language processing was highly successful in extracting key features from biopsy and operative pathology reports. This has the potential for large scale implementation for clinical audit and research using limited resources with sparing of cost and time.

1497 Prostate Cancer: Predictive Models for Outcome and Treatment Strategy Using Artificial Neural Networks in Comparison with Logistic Regression

Paulo Guilherme Salles[1], Cesar Macieira[1], Thiago Leal[2], Gilderlanio Araujo[3], Marcos Santos[2], Kenneth Gollob[4], Astaruth Froede[1], Andy Petroianu[2], Wagner Magalhães[1]
[1]Instituto Mario Penna, Belo Horizonte, MG, Brazil, [2]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, [3]Universidade Federal do Pará, Belém, PA, Brazil, [4]A.C. Camargo Cancer Center, São Paulo, SP, Brazil

Disclosures: Paulo Guilherme Salles: None; Thiago Leal: None; Gilderlanio Araujo: None

Background: Prostate cancer is the second most common cancer in men. Here we evaluated how pathological findings of prostate biopsies and radical prostatectomy specimens perform to prognosticate outcome and treatment strategy using artificial neural networks and logistic regression. With this purpose in mind, we proposed a model to predict organ confined disease.

Design: A set of 3,858 patients was used to construct, validate and test two classification models based in Artificial Neural Networks (ANN) and Logistic Regression (LR). Our set of predictive variables includes patient's age (AGE), PSA value (PSA), primary (G1B) and secondary (G2B) Gleason grade and Gleason score (GSB) at prostate biopsy, and our independent variable is extra prostatic extension (present or absent). The models were evaluated by performing 10-fold cross-validations and model metrics such as harmonic mean, accuracy, sensibility, specificity and area under curve (AUC). After model construction we predicted disease outcome for 867 external samples.

Results: In all scenarios, both methods performed similar by observing the harmonic mean, accuracy, sensibility, specificity and area under curve. The ANN reached accuracies of around 0.8 in the 10-fold cross-validations. The ANN model that was trained with AGE, PSA, G1B and G2B performed slightly better with AUC = 0.85 and harmonic mean of 0.78. This model highlighted the need for some means of feature selection by removing the dependent variable of GSB. By predicting external data, the ANN presented and accuracy of 76.8%, class 1 error 8.3% and class 2 error of 14.9%.

Conclusions: We used a large dataset, in contrast to other studies in the literature, and find that the ANN and LR performed similarly in the prediction of organ-confined prostate cancer. By removing the dependent variables, the ANN performs better than the LR. Additional studies with different features and tuning parameters for model training and test are essential to improve predictions using ANN.

1498 Building Explainable Histology Models with Deep Learning Methods

Arunima Srivastava[1], Chaitanya Kulkarni[1], Anil Parwani[1], Raghu Machiraju[1]
[1]*The Ohio State University, Columbus, OH*

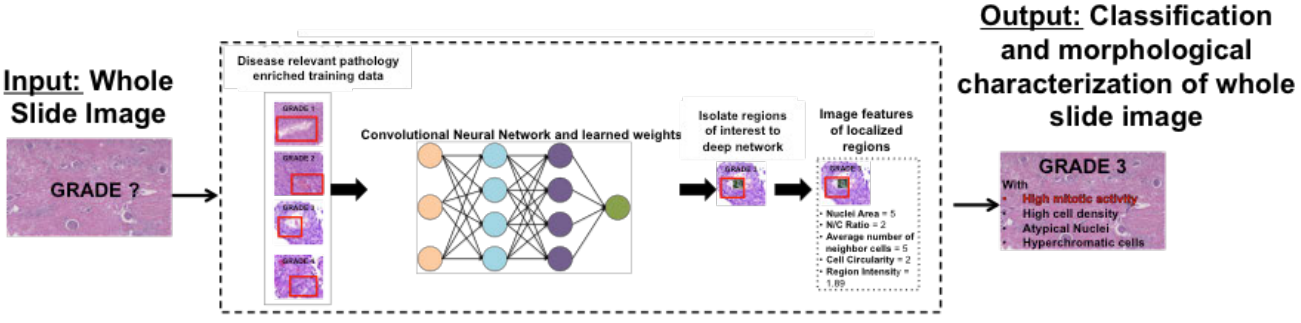
Disclosures: Arunima Srivastava: None; Chaitanya Kulkarni: None; Anil Parwani: None; Raghu Machiraju: None

Background: The utilization of Deep Convolutional Neural Networks (CNNs) illustrates steady success in cancer patient stratification and classification of tumor progression and severity. The drawbacks of using a deep learning approach remains the lack of biological context introduced while building deep models. While CNNs have proven to be powerful classifiers for disease subtypes, they fail to explain this classification, as the network engineered features used for modeling and classification are ONLY interpretable by the CNNs themselves. This work aims to change the traditional paradigm of deep network application by (a) enriching the training data by using selected pathological indicators relevant to the disease to model and (b) visualizing the CNN regions of interest and translating them to interpretable morphological features.

Design: The data set consists of The Cancer Genome Atlas's digitized histology image compendium for breast invasive carcinoma. It contains over 1000 histology images (SVS Aperio format, images from formalin-fixed paraffin-embedded tissues and frozen tissue sections) at 40x magnification, with corresponding patient attributes. In addition, regions annotating areas of high tubular, mitotic and nuclear index are recorded, which allow for training data enrichment. The model architecture utilized for building the subsequent deep learning models follows the traditional "AlexNet" architecture paradigm, which has previously been shown to be successful at modeling histology images.

Results: Our "enhanced" CNN approach, which utilizes training data, enriched with relevant pathology indicators (mitosis for the initial analysis) proved to build a more accurate model for predicting patient phenotype (2% more accurate disease stage and 10% more accurate node stage). The regions of interest to this model presented distinct morphological characteristics, including aberrant cell size and shape, hyper chromaticity and high granularity and low uniformity in texture.

Figure 1 - 1498



Conclusions: Introducing biological context in deep modeling of histology images aims to improve the quality of classification and explain the factors on which the model bases the classification. Our results prove that more accurate models are built by enriching the training data by known pathology indicators, such as mitosis for breast invasive carcinoma. Further, we are able to explain the model classification by evaluating the morphological characteristics of the regions that are of interest to the deep network.

1499 Regional Convolutional Neural Network for Multi-Organ Nuclei Segmentation in H&E Pathology Images

Peng Sun[1], Pak Hei Yeung[2], Shuoyu Xu[1]
[1]*Sun Yat-sen University Cancer Center, Guangzhou, China,* [2]*The University of Hong Kong, Hong Kong, China*

Disclosures: Peng Sun: None; Pak Hei Yeung: None; Shuoyu Xu: *Major Shareholder*, Bio-totem Pte Ltd

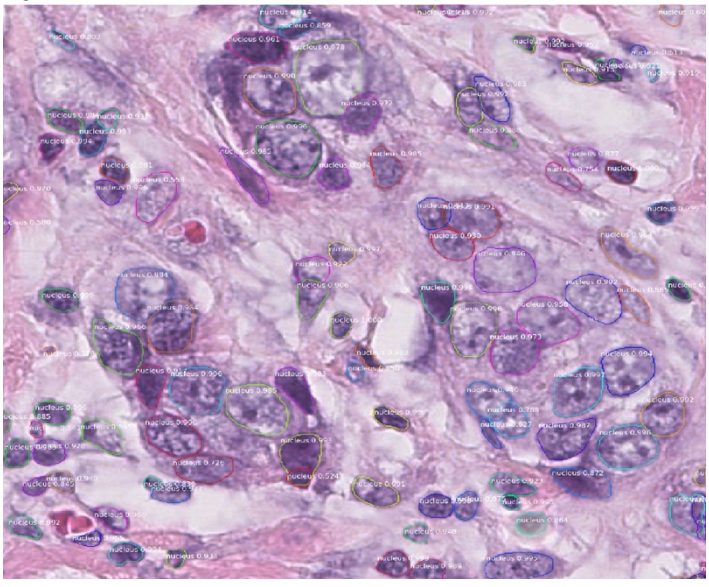
Background: Computational pathology is an emerging field which incorporates big data extracted from pathology images to establish mathematical models for more precise diagnosis and prognosis. Nuclei segmentation is one of the most important techniques to be deployed in computational pathology. Various measurements could be obtained once the nuclei are accurately segmented. Manual segmentation is tedious and prone to high observer variations. We see a great success of regional convolutional neural networks (R-CNN) for object detection and segmentation tasks in natural images, thus we believe similar methods could be suitable for segmenting nuclei in the pathology images.

Design: In this abstract, we present a nuclei segmentation method based on instance segmentation framework, i.e. object detection followed by semantic detection, using Mask R-CNN which is the newest generation of R-CNN. A ResNet 101 network as adopted as the convolutional backbone of Mask R-CNN for training and the output of the network was every segmented nucleus with the corresponding instance nucleus probability. Test image augmentation based post-processing approach was developed to further eliminate the segmented false positive nuclei. Aggregated Jaccard Index (AJI) was used to evaluate the segmentation accuracy.

Results: The dataset was provided from MICCAI 2018 multi-organ nuclei segmentation challenge which includes 30 H&E images from different organs. 23 of them were used for training and the remaining 7 for testing. A segmentation example is shown in Figure 1. AJI scores of segmentations obtained from different organs vary between 0.615 to 0.774 on training images and 0.524 to 0.678 on testing images. The overall segmentation performance after test augmentation improves especially on testing set from 0.602 to 0.614. The detailed AJI scores were shown in Table 1.

	Training		**Testing**	
Organ	Original	Test augmentation	Original	Test augmentation
Kidney	0.641±0.070	0.635±0.060	0.570±0.024	0.579±0.018
Breast	0.658±0.034	0.660±0.042	0.664	0.691
Liver	0.615±0.060	0.610±0.060	0.631	0.633
Prostate	0.684±0.053	0.683±0.047	0.678	0.678
Colon	0.645	0.642	0.524	0.541
Bladder	0.760	0.750	0.574	0.599
Stomach	0.774±0.007	0.768±0.005		
Overall	0.665±0.066	0.662±0.064	0.602±0.058	0.614±0.056

Figure 1 - 1499



Conclusions: We have developed a Mask R-CNN based nuclei segmentation method with test image augmentation, which outperforms the existing method on multi-organ H&E pathology images. The method could be adapted to nuclei segmentation in other types of pathology images, such as cytology or immunohistochemistry stained images.

1500 Telepathology Validation: Is There A Recipe? How Many Cases Do We Really Need?

Levent Trabzonlu[1], Guliz A. Barkan[2], Swati Mehrotra[3], Grazina Chatt[3], Reza Eshraghi[3], Patrick McIntire[3], Mohammed Atieh[3], Eva Wojcik[3], Stefan Pambuccian[3], Razvan Lapadat[3]

[1]Johns Hopkins University, Baltimore, MD, [2]Loyola University Healthcare System, Maywood, IL, [3]Loyola University Medical Center, Maywood, IL

Disclosures: Levent Trabzonlu: None; Guliz A. Barkan: None; Swati Mehrotra: None; Grazina Chatt: None; Reza Eshraghi: None; Patrick McIntire: None; Mohammed Atieh: None; Stefan Pambuccian: None; Razvan Lapadat: None

Background: The role of "rapid on?site evaluation" (ROSE) is indisputable in optimizing the adequacy and quality of fine needle aspirations (FNA). Using a telepathology system, pathologists are able to perform ROSE in a most time-efficient manner. The aim of this study is to establish validation of telepathology using operational metrics and benchmark data at a large academic institution.

Design: A retrospective review of institutional files was performed and consecutive cases from 6 months prior were chosen to reflect a case mix comparable to the real life. The Diff Quik slides were de-coverslipped and evaluated. The remote sites (Interventional Radiology, Endoscopy, Bronchoscopy and Ultrasonography suites) were all using LC30 USB cameras and the CellSens Standard software with the NetCam plugin (Olympus). The live image stream was visualized in a web browser using the VLC plugin. A fellow assessed the slides at the ROSE site while 6 cytopathology faculty convened in a conference room with a TV screen and without discussion amongst themselves, noted the adequacy (adequate/inadequate), diagnostic category (negative, atypical, suspicious, positive), and specific diagnosis. All participants were blinded to the original adequacy assessment and final diagnoses. For each case, evaluation time and the slides counts were noted. R studio was used for statistical analysis.

Results: FNA specimens from 52 patients (123 slides) were included in the study (20 thyroid, 12 lymph node, 4 lung, 5 liver, 6 pancreas, 5 abdominal mass). Observers were 6 board-certified cytopathologists (CYP), but not all were able to review all cases. Majority of the cases (51/52) were reviewed by 4CYP. There was a 4-way (out of 4) agreement on adequacy on 39/51 cases (76.5%), and a 6-way agreement (out of 6) on 22/28 cases (78.6%). After excluding the inadequate cases, in 26/32 cases (81.2%) there was a 4-way (out of 4) concordance on the diagnostic category, and in 20/20 (100%) there was a 5-way (out of 6) concordance on the diagnostic category. Calculations based on selecting 20/52 random cases two times, the concordance percentage on adequacy was not altered, raising the possibility that 20 cases may have been sufficient for validation. [Table]

Table. Concordance Telepathology Validation and Final Diagnosis

	No cases	% Adequacy concordance	Adequacy concordance on two runs of 20 cases	*p* value (Adequacy concordance and random selection)	% Diagnostic category concordance	Diagnostic category concordance on two runs of 20 cases	*p* value (Diagnostic category concordance and random selection)
CYP1	28	96.43%	Not enough cases for evaluation	NA	90.91%	Not enough cases for evaluation	NA
CYP2	51	88.24%	85%, 85%	>0.05	85.71%	87.5%, 87.5%	>0.05
CYP3	51	80.39%	80%, 80%	>0.05	86.84%	86.7%, 93.3%	>0.05
CYP4	51	78.43%	80%, 80%	>0.05	94.29%	86.7%, 100%	>0.05
CYP5	52	76.92%	70%, 75%	>0.05	92.11%	92.3%, 100%	>0.05
CYP6	34	85.29%	Not enough cases for evaluation	NA	92.31%	Not enough cases for evaluation	NA

Conclusions: The telepathology system was validated. The mean concordance rate for adequacy and diagnostic category compared to the final diagnosis was 83.2% and 90.1%, respectively. We postulate that 20 cases may be sufficient for telepathology validation.

1501 Toward an Automated Scoring Algorithms of Fluorescence in Situ Hybridization (FISH) on Formalin-Fixed Paraffin-Embedded (FFPE) tissues Using a Confocal Whole Slide Image Scanner and Image Analysis Software

Naohiro Uraoka[1], Mamta Rao[2], Yanming Zhang[1], Meera Hameed[1], Yukako Yagi[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY, [2]Memorial Sloan Kettering Cancer Center, Hackensack, NJ

Disclosures: Naohiro Uraoka: None; Mamta Rao: None; Yanming Zhang: None; Meera Hameed: None; Yukako Yagi: None

Background: The standard, manual scoring for fluorescence in situ hybridization (FISH) analysis of formalin-fixed paraffin-embedded (FFPE) tissues is labor-intensive and time-consuming. Confocal imaging which eliminates out-of-focus noise offers higher resolution and more spatial information than conventional widefield imaging. The purpose of this study was to establish an automated FISH scoring method using a confocal whole slide image (WSI) scanner and image analysis software which is commercially available.

Design: Six de-identified archival FFPE tissue of malignant lymphoma and Ewing's sarcoma with break-apart FISH assays were prepared for the current study (including four negative cases and two positive cases). For each slide, several regions of interest (ROIs) of tumor areas were selected for confocal scanning according to the corresponding H&E slides. FISH slides were digitized by a Pannoramic Confocal WSI scanner (3DHistech Ltd., Budapest, Hungary) with a 40× water immersion objective (0.1625 micrometer/pixel). We scanned seven layers at 0.6 µm intervals based on the optimization results of the previous study. The images were viewed in CaseViewer (3DHistech), and ROIs for image analysis were defined. The image analysis for FISH scoring was performed with Imaris (Bitplane, Zurich, Switzerland), in which we created algorithms for nuclear segmentation (Figure 1), spot quantification and detection of spot co-localization (Figure 2). The accuracy of the analysis was compared with the results by manual scoring in clinical reports.

Results: Confocal scanning of FISH slides provided sharp images with spatial information of spot signals. By our semi-automated algorithms in Imaris, nuclear segmentation and spot detection were successfully performed. The ratio of rearrangement signal patterns correlated with the clinical results when we used a 10% cutoff (Table).

Case #	1	2	3	4	5	6
Clinical results	-	-	+	-	+	-
Scoring by Imaris (%)						
2F	74.7	50.7	5.9	60.1	24.8	60.1
>2F	6.8	12.9	0.9	5.3	3.9	10.7
Rearrangement	1.0	3.5	51.6	5.3	23.5	2.4
1F1G1R	0.7	1.6	24.2	1.8	11.2	0.6
≥1F, ≥1G, and/or ≥1R	0.3	0.2	15.2	0.7	4.7	0.3
≥1F, 1G, ≥1R	0.0	0.0	4.7	0.0	0.8	0.0
≥1F, ≥1G, 1R	0.0	0.0	5.5	0.0	0.8	0.0
≥1F,1G,≥1R ≥1F,≥1G,1R	0.3	0.2	3.0	0.7	2.6	0.3
≥1F,≥1G,and/or ≥1R only	0.0	0.0	2.0	0.0	0.5	0.0
Other Rearrangement	0.0	1.6	12.2	2.8	7.6	1.5
Others	17.5	32.9	41.6	29.3	47.8	26.8
Total	100	100	100	100	100	100

(F: fusion, G: green, R: red)

Figure 1 - 1501

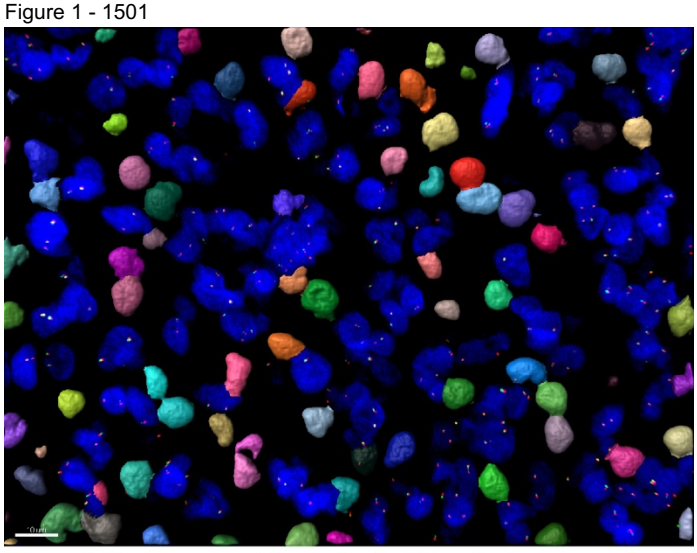


Figure 1: Nuclear segmentation by Imaris. Different colors were assigned to each segmented nucleus.

Figure 2 - 1501

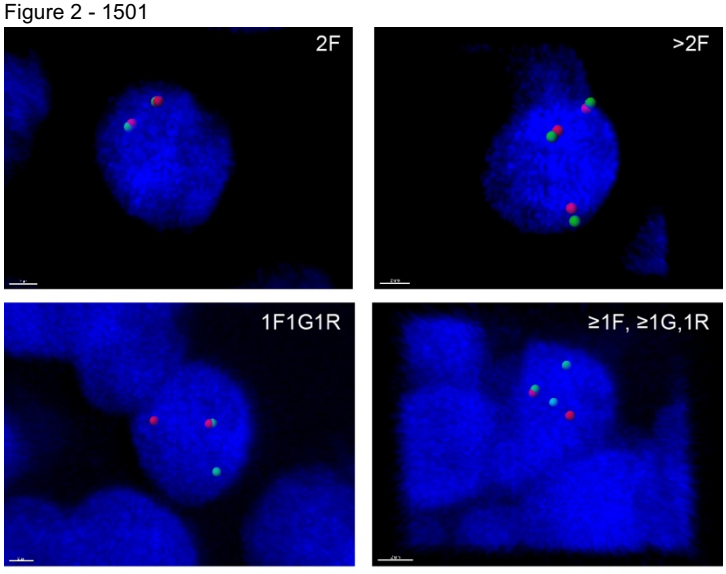


Figure 2: Spot detection by Imaris.

Conclusions: We established a semi-automated method of FISH scoring for FFPE tissue by combining with confocal scanning. This method was helpful to obtain accurate, plentiful information of FISH more efficiently than conventional manual scoring methods. The next step is to enrol more clinical cases for evaluation and validation of other FISH assays, such as for deletion, gain, amplification, and translocation assays. Furthermore, according to the protocol and the data of this study, we are currently developing in-house software for a fully automated FISH scoring system also considering deep learning.

1502 Pathologist Opinions About Epic Beaker AP: A Multi-Institutional Survey of Early Adopters

Mandy VanSandt[1], Kevin Turner[1], Raj Dash[2], Dorina Gui[3], Madalina Tuluc[4], Steven Hart[5], Phil Raess[1]
[1]Oregon Health & Science University, Portland, OR, [2]Duke University Medical Center, Durham, NC, [3]University of California, Davis, Sacramento, CA, [4]Philadelphia, PA, [5]Los Angeles, CA

Disclosures: Mandy VanSandt: None; Kevin Turner: None; Dorina Gui: None; Steven Hart: None; Phil Raess: None

Background: Epic Systems Corporation provides a widely used electronic medical record among healthcare organizations. Beaker Anatomic Pathology (AP) is a newly developed laboratory information system (LIS) gaining marketshare amongst academic pathology departments. Herein we describe survey results about Beaker AP from five academic institutions.

Design: A 37-question survey was electronically administered to pathologists and pathology trainees at five academic pathology departments. The majority of questions assessed participant agreement with statements concerning overall satisfaction and efficiency of Beaker AP. Additional questions recorded data concerning experience, sign-out responsibilities, daily usage, and legacy LIS. The survey was IRB approved and results were anonymous.

Results: Survey respondents included 51 faculty and 23 residents. The majority of respondents reported using Beaker AP for >4 hours daily (78%) and reported that they were not involved in the planning or transition to Beaker AP (69%). There was high inter-institutional variability regarding overall pathologist satisfaction with Beaker AP, however institutions with modern LISs (Powerpath, CERNER) had a generally neutral to negative assessment of Beaker AP. Additionally, the majority of respondents disagreed with the statements "my job satisfaction is higher with Beaker AP than the previous non-Beaker LIS" and "Beaker AP is easy to use and designed for my needs". Overall, pathologists had a neutral opinion of whether generating and signing out a complete report was faster in Beaker AP, again with marked inter-institutional variation. Opinions of Beaker AP were not different between pathologists with greater than 10 years experience and those with less than 10 years experience.

Survey Statements with Strongest Concordance	
Strongly Agree	Reviewing clinical information from patient charts is easier in Beaker AP than the previous non-Beaker LIS Complete and accurate billing is easier in Beaker AP than with the previous non-Beaker LIS The training I received helps me do my job
Strongly Disagree	Frozen turnaround time has improved with Beaker AP Searching for prior cases is easier in Beaker AP than the previous non-Beaker LIS Grossing is more efficient in Beaker AP than the previous non-Beaker LIS

Conclusions: Pathologist opinions about Beaker AP are generally neutral to negative, however high inter-institutional variability was noted in responses which is likely due to a combination of the efficacy of the legacy LIS, familiarity with Beaker AP at the time of the survey, and institution-specific optimization efforts. Pathologists felt that Beaker AP was useful for reviewing clinical information and billing; areas of weakness included frozen turnaround time, searching for prior cases, and grossing efficiency. Pathologists agreed slightly with the statement that Beaker AP is more functional overall than the previous non-Beaker LIS and disagreed slightly with the statement that Beaker AP is easy to use and designed for their needs.

1503 Computerized Nuclear Morphometric features from H&E Slide Images are prognostic of recurrence and predictive of added benefit of adjuvant chemotherapy in early stage non-small cell lung cancer

Xiangxue Wang[1], Cristian Barrera[2], Cheng Lu[3], Michael Yang[4], Vamsidhar Velcheti[5], Anant Madabhushi[3]
[1]Cleveland, OH, [2]Case Western Reserve University, Bogota, Colombia, [3]Case Western Reserve University, Cleveland, OH, [4]University Hospitals Cleveland Medical Center, Cleveland, OH, [5]New York University School of Medicine, New York, NY

Disclosures: Xiangxue Wang: None; Cristian Barrera: None; Cheng Lu: None; Michael Yang: None; Anant Madabhushi: *Consultant*, Inspirata Inc.; *Advisory Board Member*, Inspirata Inc.

Background: A significant number of early-stage non-small cell lung cancer (ES-NSCLC) patients will develop recurrence with 5-year survival rates less than 50%. Adjuvant chemotherapy has shown benefits in resected early-stage lung cancer patients, but there is no validated biomarker or companion diagnostic (CDx) tests to predict which of these patients will benefit from this cytotoxic treatment.

Design: This study consisted of 431 patients and 104 patients developed recurrence after treatment. All of them underwent surgery and 83 of them received adjuvant chemotherapy. Patients are randomly divided into training set (n=100) and validation set (n=331). A total of 281 features were extracted from both nuclei and peri-nuclei region to characterize the shape, texture, and architecture of tumor microenvironment for each patient. Elastic net-Cox proportional hazards model was built to select the most predictive image features and computed the continuous risk scores from the training set. Based on the threshold from the training set, the patients were separated into high-risk and low-risk survival groups. The univariate and multivariate analysis was employed to evaluate the performance of imaging features.

Results: The elastic model selected 17 metrics (nuclei shape, arrangement, and peri-nuclei area texture) as the most stable and prognostic features to build a risk model. The median of the risks from the training set could statistical significantly separate the validation set into low and high-risk groups (n=331, CI=0.72, HR = 1.9, P=0.046). The same model was also shown be predictive by subset analysis. The stratification by recurrence risk of the patients who received the surgery alone shown different survival hazards, which potentially could indicate the was patient should receive adjuvant chemotherapy for improving survival (n=264, CI=0.71, HR = 2.57, P = 0.025). However, the patients who received adjuvant chemotherapy after surgery didn't show any significant difference between high and, groups (n=67, CI=0.69, HR = 0.75, P = 0.93).

Figure 1 - 1503

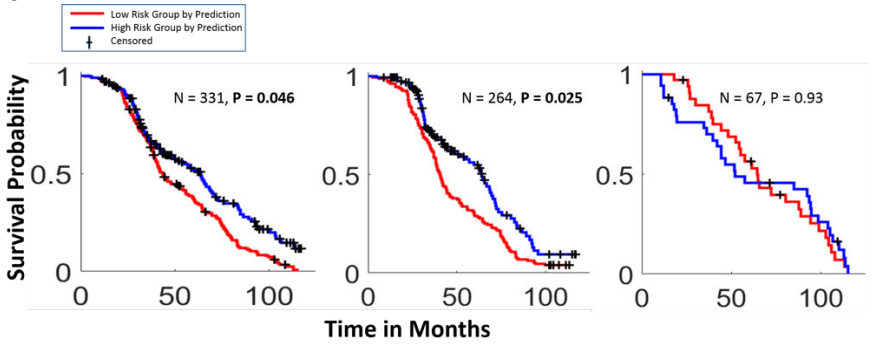


Figure 1. Group of patients stratified by risks of recurrence: a) low and high risk patients of the entire validation set; b) low and high risk patients of the surgery only patients; 3) low and high risk patients of the surgery plus chemotherapy patients.

Conclusions: A combination of computer-extracted features relating to nuclear morphology was found to be both prognostic of recurrence and predictive of the added benefit of adjuvant chemotherapy in early-stage non-small cell lung cancer.

1504 Machine learning digital image analysis for histotype classification of ovarian carcinoma

Nicholas Wiebe[1], Fady Mina[1], Thomas Kryton[2], Yves Pauchard[1], Martin Kobel[1]
[1]University of Calgary, Calgary, AB, [2]Calgary Laboratory Services, Calgary, AB

Disclosures: Nicholas Wiebe: None; Fady Mina: None; Thomas Kryton: None; Yves Pauchard: None; Martin Kobel: None

Background: Machine learning tools are increasingly accessible, and their application to digital image analysis for pathology has generated considerable interest. Histotype determination for ovarian carcinomas is a challenging task that nevertheless shows good interobserver reproducibility based on morphology alone. We applied previously published machine learning techniques to perform automated histotype classification of whole slide images of ovarian carcinomas.

Design: In order to train the classifier, we extracted image patches from a set of slides from 58 cases of ovarian carcinomas from the OVAL BC study, which were previously classified into one of five diagnostic categories based on consensus expert opinion with integration of immunohistochemistry: high grade serous carcinoma (HGSC), low-grade serous carcinoma (LGSC), clear cell carcinoma (CCC), mucinous carcinoma (MC), and endometrioid carcinoma (EC). The slides were digitized and partially annotated to indicate areas of known neoplastic and non-neoplastic tissue. We use data augmentation techniques to diversify the training set images. We trained a classifier based on the Inception v3 model, a model widely used for a variety of image classification tasks.

Results: The trained classifier was tested on an independent set of 29 annotated whole slide images.

It achieves an overall accuracy of 86.1% on image patches sampled from the testing set slides, with highly accurate classification of non-neoplastic patches (99.7%), and moderate accuracy for classifying carcinoma patches, ranging from 71.3% for MC to 89.2% for EC.

Conclusions: Overall, there was a highly accurate separation of neoplastic from non-neoplastic tissue. The trained classifier performs modestly well regarding histotype classification. Reviewing the cases in which the classifier failed to predict the correct histotype shows that in many cases, the morphology of the case is somewhat atypical for the class (e.g. mucinous carcinoma with areas lacking clear mucinous differentiation). We expect that the performance of the classifier can be improved by increasing the number of whole slide images used for training and optimizing the model parameters, but that there may be an upper limit to the performance that can be achieved without integrating information from immunohistochemistry. These techniques can also be easily applied to other classification problems in pathology.

1505 Seven-year experience and value of international telepathology consultation between KingMed Diagnostics and University of Pittsburgh Medical Center

Tao Wu[1], Pifu Luo[2], Xiangdong Ding[1], Zhikui Zhang[1], Jing Li[1], Dong Liu[1], Chengquan Zhao[3]
[1]KingMed Diagnostics, Guangzhou, China, [2]KingMed Diagnostics, Central Point, OR, [3]Pittsburgh, PA

Disclosures: Tao Wu: None; Pifu Luo: None; Chengquan Zhao: None

Background: Telepathology is the practice of pathology by interpreting digitalized images by pathologists at a long distance. International telepathology consultation (ITC) is a critical part of precision medicine for its accurate diagnosis. Since 2012, KM has established an ITC platform for Chinese patients by collaborating with University of Pittsburgh Medical Center (UPMC).

Design: This study summarizes our experience and quality of ITC service and shine the light of ITC model for pathology consultation. A total of 4173 difficult cases from January 2012 to August 2018 were consulted via this platform. All H&Es and immunostains were scanned into whole slide images (WSI) and transferred to UPMC. The hardwares, softwares and website platform were validated bilaterally.

Results: Of these cases, 61.7% were requested by primary pathologists, 35.3% by primary clinicians and 3% by the patients. Gynecologic/breast pathology received the most cases (21.9%), followed by bone and soft tissue pathology (20.7%) and hematopathology (19.3%). Turnaround time (TAT) varies depending on subspecialities (Table 1). Average TAT for all cases was 5.9 days. The longest TAT was on hematopathology cases due to their difficulties and requirement of more than one round of immunostains. 2141 cases (51.3%) were submitted without any preliminary diagnosis. Other 2032 cases were submitted with a primary diagnosis/interpretation rendered by the referring hospitals. Among these cases, only 25.2% of cases reached a complete agreement between the referring hospitals and UPMC pathologists, 23.1% with mild discrepancy and 51.7% with major discrepancy that may impact treatments. For the legal responsibility, KM pathologists reviewed all cases and rendered a diagnosis for each case before submitting to UPMC. It is noted that the diagnosis between KM and UPMC pathologists reached a very high concordance rate.

Table 1. Case Distribution and Turnaround Time (January 2012 – August 2018)

Subspecialities	Case Number	Percentage (%)	Turnaround Time (working days)
Gyn/Breast	914	21.9	3.8
Bone/Soft tissue	865	20.7	5.5
Hemato	805	19.3	9.1
Head/Neck/Endocrine	359	8.6	6
Derm/Melanoma	335	8	4.5
GI	323	7.7	4.8
GU	199	4.8	8
Thoracic	159	3.8	4.5
Neuro	92	2.2	5.5
Liver	58	1.4	8
Pediatric	32	0.8	8.8
Others	32	0.8	5.2
Total	4173	100	5.9

Conclusions: Our experience of 7 years' implementation has indicated ITC is a very successful model for pathology consultation. KM provided a valuable platform so that through it, the community hospitals could directly reach world well-known pathology institutions, such as UPMC. The results demonstrate that over a half of the diagnoses from referring hospitals were corrected by ITC service. It is very important for the patients' safety, treatment, management. It also indicates that KM with its pathologists is a great avenue and resource of pathology consultation for many community and tertiary hospitals in China.

1506 The Morphometric Analysis of Cervical Biopsy and the Correlation with Pathologist Interpretations Versus High-Risk HPV Results

Keluo Yao[1], Stephanie Skala[2], Emily Mcmullen[2], Girume Degu[2], Heather Walline[2], Ulysses Balis[2], Richard Lieberman[3]
[1]Ann Arbor, MI, [2]University of Michigan, Ann Arbor, MI, [3]Michigan Medicine, Ann Arbor, MI

Disclosures: Keluo Yao: None; Stephanie Skala: None; Emily Mcmullen: None; Girume Degu: None; Heather Walline: None; Ulysses Balis: *Advisory Board Member*, Inspirata, Inc; Richard Lieberman: None

Background: The management of patients with abnormal cervical cytology and/or high-risk HPV (hrHPV) relies on colposcopic directed biopsies and morphologic examinations by the pathologist. However, the biopsies are less reliable between the benign category versus low-grade squamous intraepithelial lesions (LSIL). Numerous publications have shown a Cohen's kappa coefficient of <0.5 for diagnostic reproducibility. The aim of this study is to see if morphometric analysis assisted by digital image analysis and machine learning algorithm maybe a solution.

Design: From electronic medical records, we selected thirty-six individual biopsies that were diagnosed as benign (n=16), LSIL (n=17), or suspicious for LSIL (n=3). All biopsies had hrHPV results either from primary HPV screening or HPV multiplex PCR-MassArray and L1 consensus PCR with sequencing. All biopsy glass slides are scanned into whole slide images at 40x and a representative 1000x1000 pixel image extracted from the superficial squamous epithelium for each biopsy. We designed a digital image analysis (DIA) algorithm for nuclear feature extraction using ImageJ v1.52 (Figure 1). The extracted features were analyzed using a decision tree classifier (DTC) (scikit-learn v0.19.2) using either the pathologist's routine diagnoses or the HPV results as the ground truth. The cases were split 3:2 into training and validation set and the composition was randomly shuffled 5 times.

Results: Using the pathologist's routine diagnoses as the ground truth, the DTC achieved the average precision of 0.466 (0.33 to 0.62). By switching the ground truth to hrHPV results, the DTC achieve the average precision of 0.746 (0.56 to 0.9).

Figure 1 - 1506

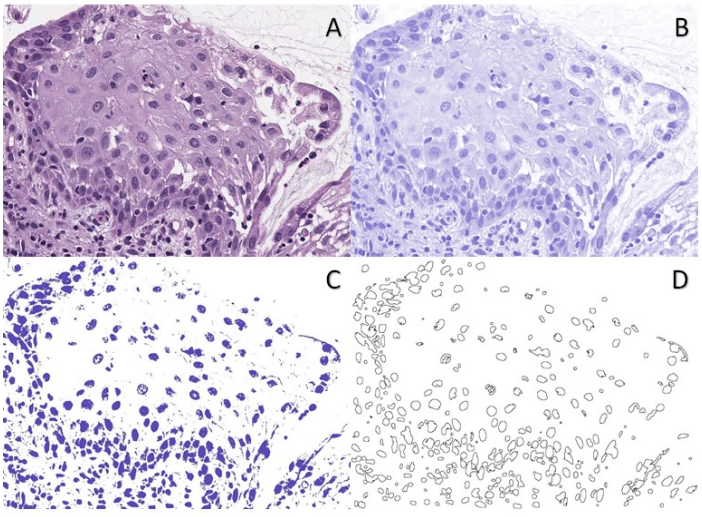


Figure 1: The extraction of nuclear features from the cervical biopsy superficial squamous epithelium using ImageJ. The original H&E images are normalized by background subtraction (A), followed by H&E color deconvolution for hematoxylin (B). The nuclear features are parsed out using a threshold method (C) and the final results are denosied and sent for analysis (D)

Conclusions: Our data suggest that the DIA algorithm and the subsequent DTC performs better with the hrHPV results as the ground truth than the diagnoses from routine examination by pathologists. The preliminary findings indicate that algorithmically extracted LSIL morphologic features have a strong correlation with hrHPV status which indicates a DIA algorithm may help us to achieve better diagnostic reducibility for colposcopic directed biopsies. In the next phase of the study, we will expand the sample size and build a more sophisticated DIA algorithm.

1507 Machine Learning Models Improve the Diagnostic Yield of Peripheral Blood Flow Cytometry

Mingjuan Zhang[1], Alan Guo[2], Stephan Kadauke[3], Anand Dighe[4], Jason Baron[1], Aliyah Sohani[1]
[1]Massachusetts General Hospital, Boston, MA, [2]Independent Researcher, Boston, MA, [3]Children's Hospital of Philadelphia, Philadelphia, PA, [4]Massachusetts General Hospital, Newton Center, MA

Disclosures: Mingjuan Zhang: None; Alan Guo: None; Stephan Kadauke: None; Anand Dighe: None; Jason Baron: *Consultant*, Roche Diagnostics; Aliyah Sohani: None

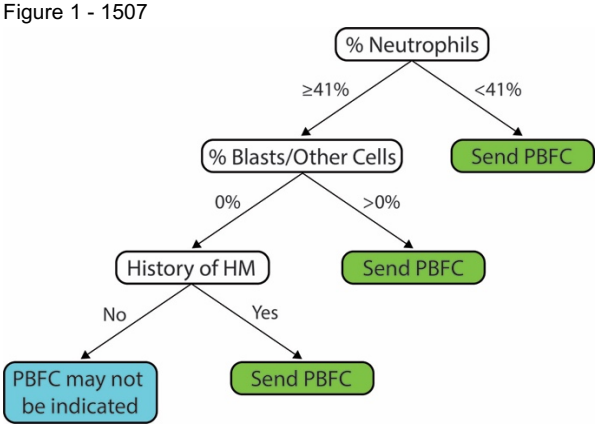
Background: Peripheral blood flow cytometry (PBFC) is a useful tool for evaluating circulating hematologic malignancies (HM), but has poor diagnostic yield when used as a screening test. In this study, we used machine learning to evaluate whether CBC/differential parameters could inform PBFC test ordering decisions and provide a practical strategy to improve PBFC utilization.

Design: 785 PBFC samples (744 patients) from 2017 with a concurrent/recent (within prior 6 weeks) CBC/differential were identified. The samples were randomly split 80%/20% into training (n=626) and test (n=159) cohorts. PBFC results with abnormal blast or lymphoid populations were labeled as positive; those that were negative or with findings of limited clinical significance were labeled as negative. Recursive partitioning and logistic regression models were used to predict whether each PBFC would be positive or negative based on the patient's CBC/differential results.

Results: Positive PBFC results were seen in 149/230 (64.8%) and 103/396 (26.0%) of cases with and without a history of HM, respectively ($p<0.001$). The mean white blood cell (WBC), hematocrit (HCT), platelet count (PLT), % Neutrophils, % Lymphocytes, and % Blasts/Other Cells were significantly different between the positive and negative PBFC groups. For model selection, we opted for a decision tree model to facilitate generation of a practical clinical algorithm and a logistic regression model for comparison. On the test cohort, using model

cutoffs that optimized primarily for sensitivity, the single decision tree model (Figure 1) achieved a sensitivity of 98% and specificity of 63% (area under the ROC curve [AUC] = 0.911). The logistic regression model revealed that WBC, PLT, % Neutrophils, % Blasts/Other Cells, and history of HM were significant predictors of a positive PBFC result; this model achieved a sensitivity of 98% and specificity of 58% (AUC = 0.893). The one case classified as a false negative in both models was an asymptomatic patient with persistent mild leukocytosis for many years, for whom PBFC showed 0.1% blasts; a follow-up PBFC in 2018 was negative.

Figure 1 - 1507



Conclusions: Our models show triaging strategies that could potentially defer ~60% of PBFC tests without missing any clinically-significant outcomes, thus dramatically improving the diagnostic yield of PBFC and decreasing unnecessary clinical utilization of this test. Preliminary implementation of these models prospectively in a test environment is underway.

1508 Diagnostic Assessment of Recalibrated Multi-instance Deep Learning Network for Gastric Dysplasia and Adenocarcinoma in Biopsy

Yaxi Zhu[1], Yan Huang[1], Shujun Wang[2], Hao Chen[3], Xinjuan Fan[4]
[1]The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2]The Chinese University of Hong Kong, Hong Kong, China, [3]The Chinese University of Hong Kong, Hong Kong, China, [4]Guangzhou, China

Disclosures: Yaxi Zhu: None; Yan Huang: None; Shujun Wang: None; Hao Chen: None; Xinjuan Fan: None

Background: Histological morphology is still the basis for pathological diagnosis. Whereas, the diagnosis of gastric biopsies with hematoxylin and eosin-stained sections is labor intensive and error-prone due to the morphological diversity and inter or intra-observer variability. Here, we trained a recalibrated multi-instance deep learning (RMDL) network to address this issue.

Design: In our study, a total of 765 gastric biopsy sections were collected and scanned at 40x magnification. Of which, 318 were gastric adenocarcinoma, 149 were dysplasia and 298 were normal. The dataset was randomly divided into training set (75%) and testing set (25%). RMDL network, including local-global feature fusion and instance recalibration, was proposed for diagnosing gastric dysplasia and adenocarcinoma. An adaptive weight initialization strategy was presented to make full use of the hidden layer information to improve the performance of our RMDL network.

Results: The sensitivity, specificity, false negative rate and false positive rate of RMDL network for diagnosing gastric dysplasia and adenocarcinoma were 85.3%, 93.4%, 14.7%, 6.6% and 92.4%, 91.1%, 7.6%, 8.9%, respectively. Importantly the overall accuracy were 91.7% and 91.7% for dysplasia and adenocarcinoma, respectively, indicating RMDL network could serve as a more effective and objective diagnostic strategy.

Conclusions: Our RMDL network displayed a strong diagnostic power in classifying dysplasia and adenocarcinoma in gastric neoplasm, providing an important approach to assist the pathologist. Furthermore, the network could be used for grading dysplasia and adenocarcinoma more objectively.