Digital Pathology:
Diagnostic Errors, Viewing Behavior and Image Characteristics

Ezgi Mercan

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Our Aim

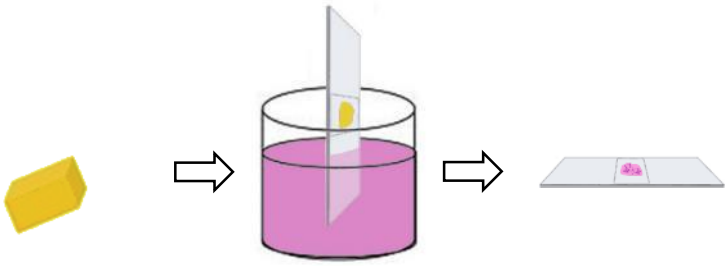
We propose novel image features and viewing behavior analysis methods

- to investigate the causes of diagnostic errors and
- to discover accurate and efficient interpretation strategies for pathologists.

Outline

1. Introduction
2. ROI Localization in WSIs
3. Tissue Label Segmentation
4. Automated Diagnosis of ROIs
5. Viewing Behavior Analysis
6. Conclusions

Medical Diagnosis of Cancer

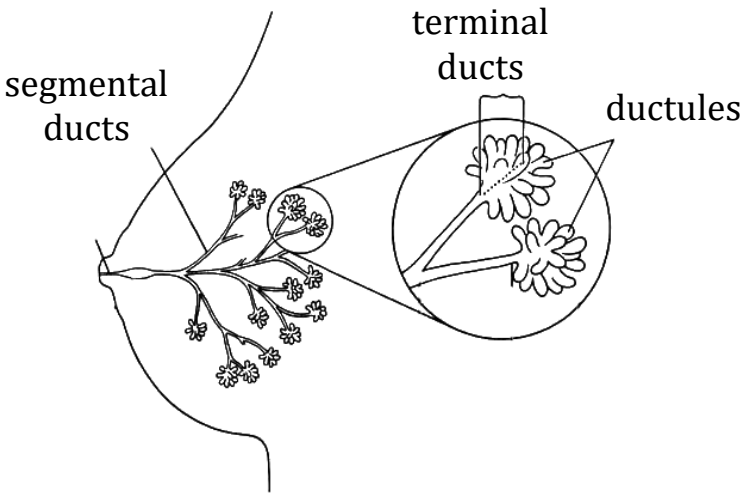


tissue specimen *H&E staining* *glass slide*

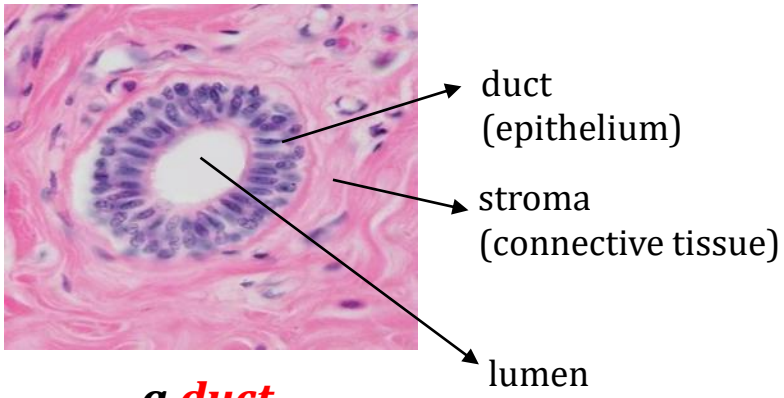
light microscope

whole slide imaging

Breast Histopathology

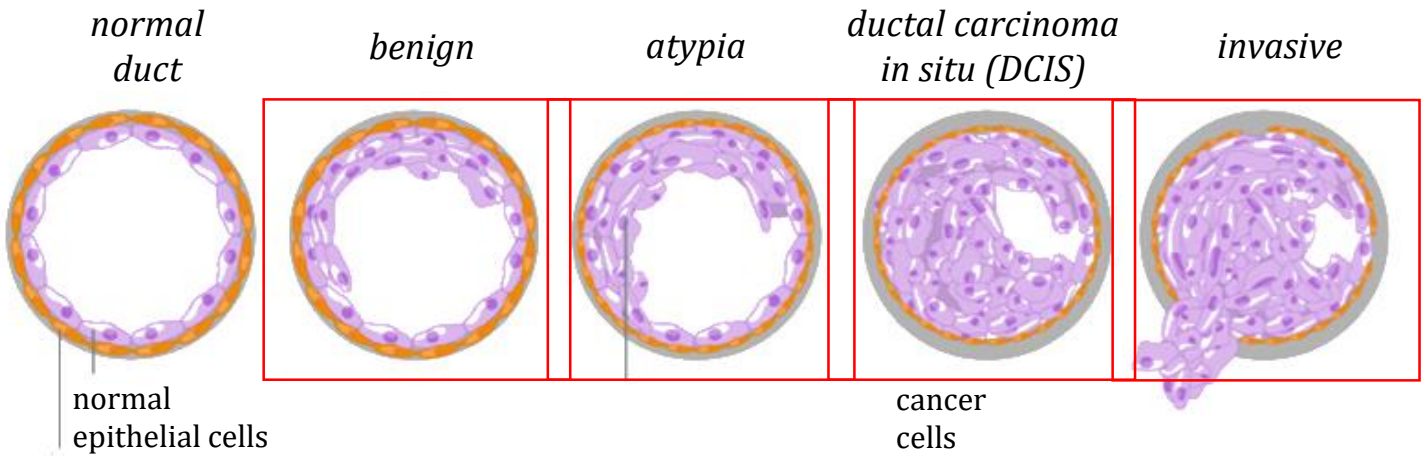


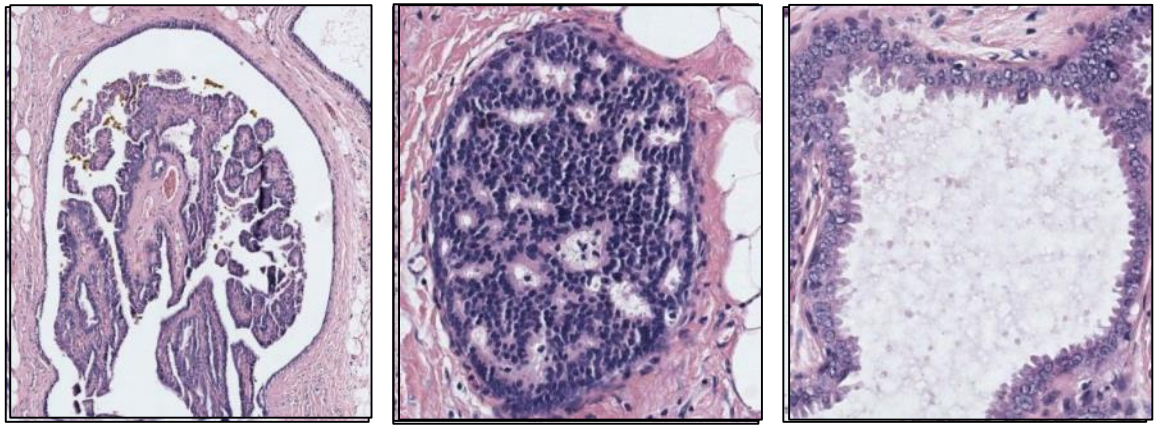
female breast



a duct

Diagnostic Categories



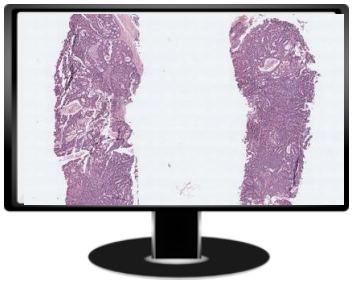


digiPATH Dataset

pathologist
(n=87)

age
years in practice
cases per week
affiliation
lab size
education
pathology perception

assessment
(n=5,220)



participant ROI*
participant diagnosis
viewport log

confidence score
difficulty score
borderline diagnoses

diagnostic outcome

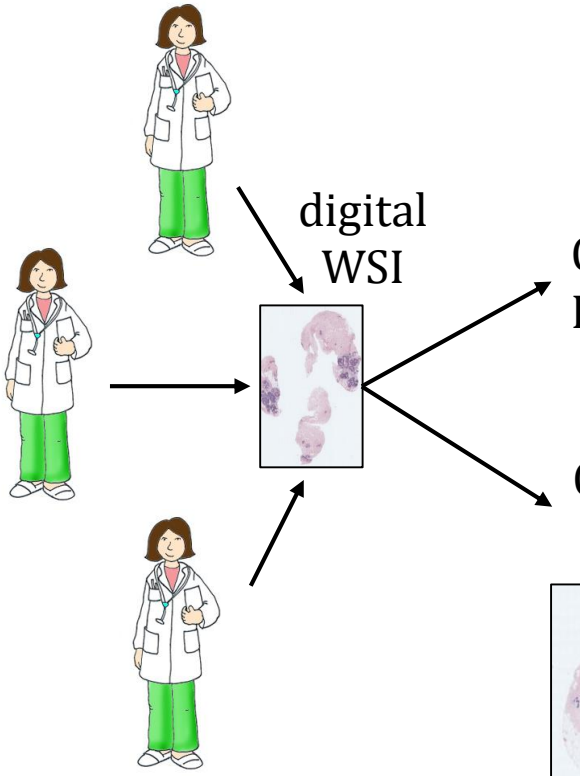
ROI identification*
diagnostic agreement
efficiency

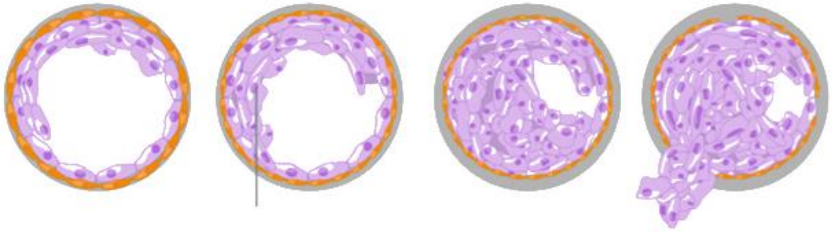
case
(n=240)

digital WSI
consensus ROI
consensus diagnosis

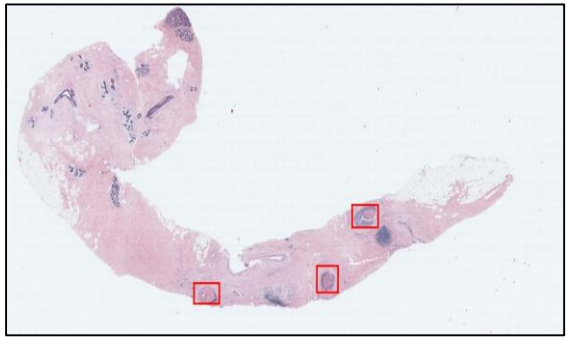
biopsy type
breast density
patient age

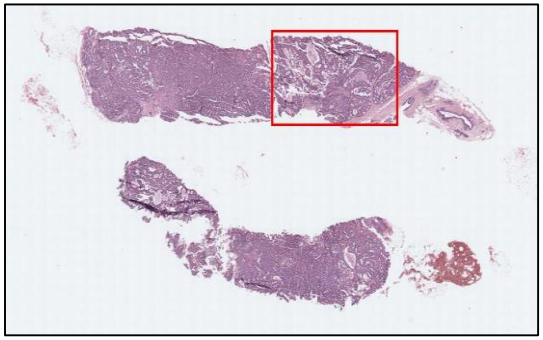
Expert Consensus Data Collection





Consensus Diagnosis	*benign*	*atypia*	*(DCIS)*	*invasive*
# Cases	*60*	*80*	*78*	*22*
Consensus ROIs	*102*	*128*	*162*	*36*





digiPATH Dataset

pathologist
(n=87)

assessment
(n=5,220)

case
(n=240)

age
years in practice
cases per week
affiliation
lab size
education
pathology perception

participant ROI*
participant diagnosis
viewport log

confidence score
difficulty score
borderline diagnoses

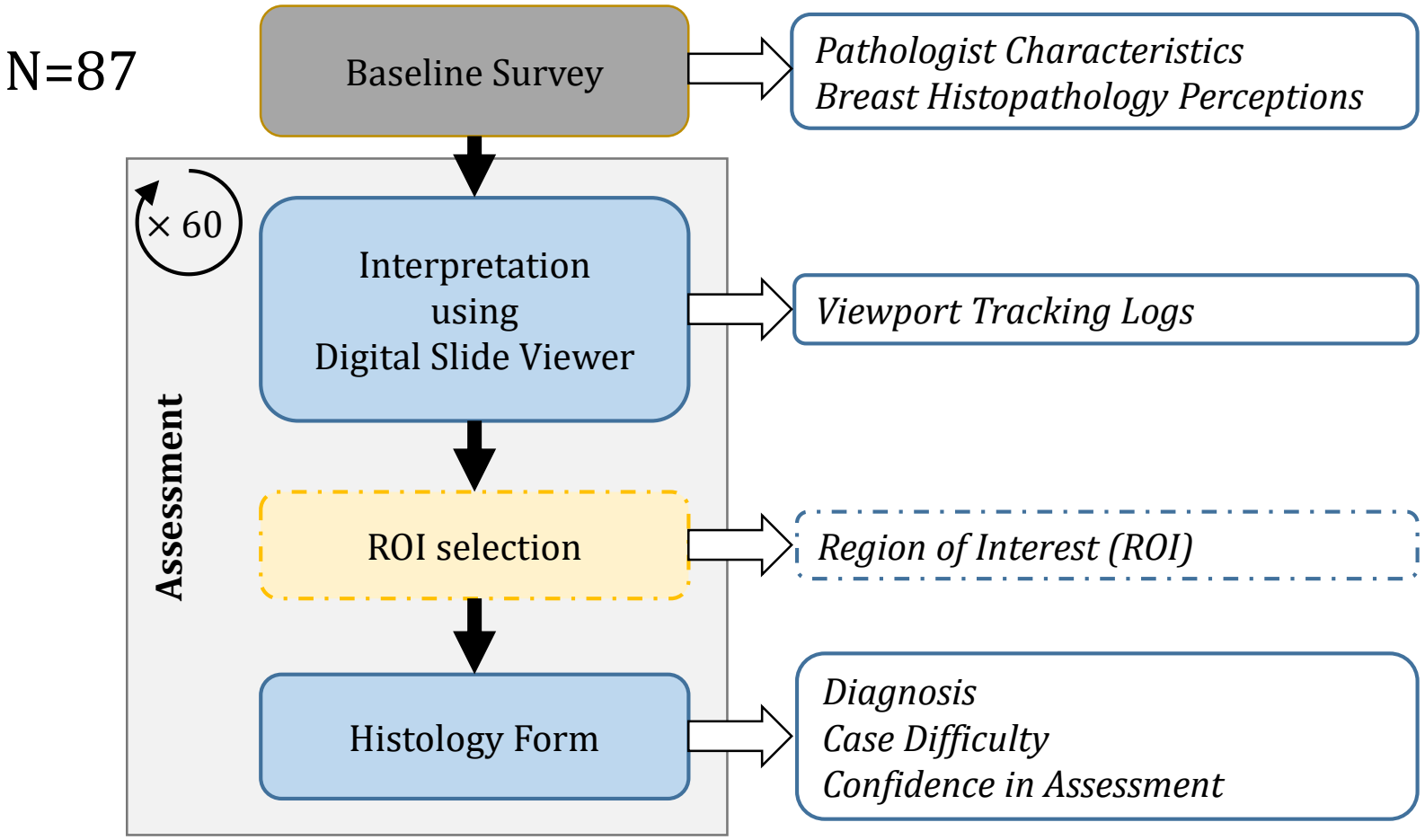
digital WSI
consensus ROI
consensus diagnosis

biopsy type
breast density
patient age

diagnostic outcome

ROI identification*
diagnostic agreement
efficiency

Participant Data Collection



Outline

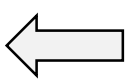
1. Introduction
2. **ROI Localization in WSIs**
3. Tissue Label Segmentation
4. Automated Diagnosis of ROIs
5. Viewing Behavior Analysis
6. Conclusions

digiPATH Dataset

assessment
(n=5,220)

participant ROI*
participant diagnosis
viewport log

confidence score
difficulty score
borderline diagnoses



case
(n=240)

digital WSI
consensus ROI
consensus diagnosis

biopsy type
breast density
patient age

Localization of Diagnostically Relevant Regions of Interest in Whole Slide Images

- Whole slide images are big (50,000 x 20,000 pixels).
- The diagnostic decision making is a complex cognitive process that includes visual search and interpretation tasks.
- **Region of Interest (ROI):** Parts of the whole slide to which the pathologists paid attention.

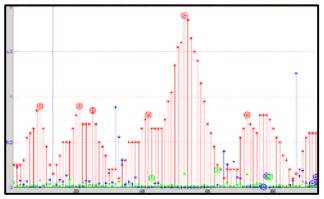
- Our aims are:
  - To understand what attracts' pathologists attention, and
  - To model diagnostically relevant regions computationally using image features.

E. Mercan, *et al.*, International Conference on Pattern Recognition, 2014.
E. Mercan, *et al.*, Journal of Digital Imaging, 2016.

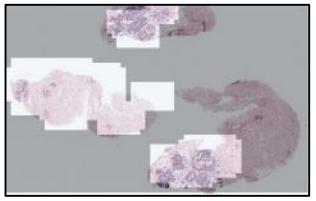
ROI Localization in WSIs

Tracking Data Analysis to Discover ROIs

tracking data

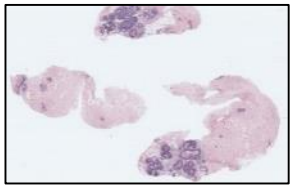


our viewport analysis

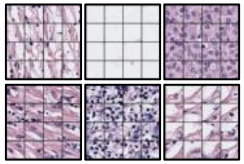


ROIs to which pathologist actually paid attention

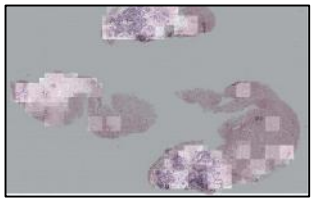
A Model based on Color and Texture to Predict ROIs



an unseen WSI



our computational model based on color and texture

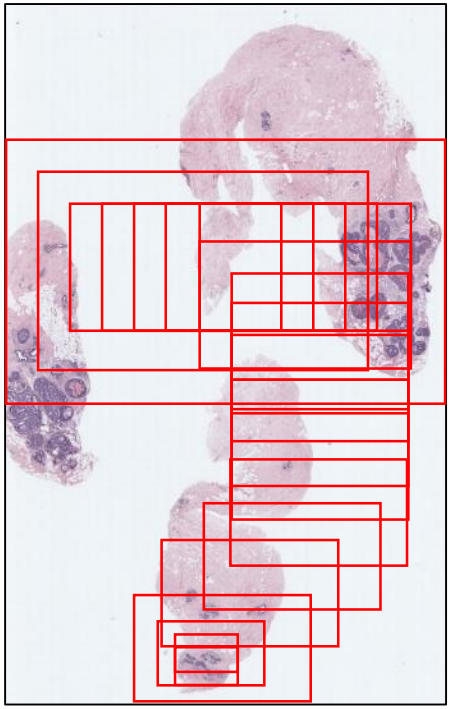


our predicted ROIs

Viewport Log Analysis

- *Where did pathologists really look?*
- **viewport**: a rectangular part of the image pathologist sees on the screen
- **viewport logs:** records all viewports
  - *panning:* changes the location of viewport
  - *zooming in/out:* changes the size of viewport

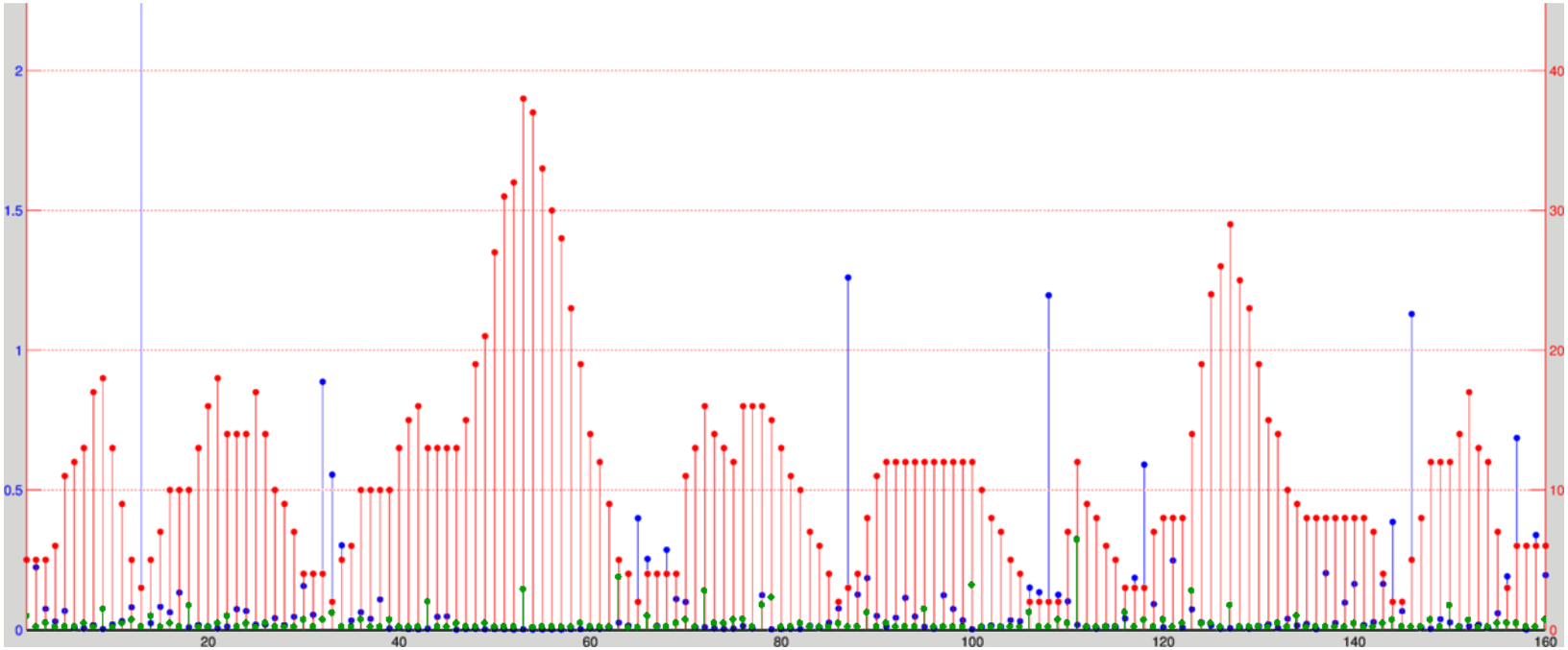
Viewport Log ID	Participant ID	Case ID	Log Time	Position X	Position Y	Width	Height	Zoom Level	Control Width	Control Height	Viewer Version
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
277	406	1883	4/3/13 14:20:39.167	13434	10907	19660	16000	2	983	800	17039872
278	406	1883	4/3/13 14:20:42.463	10154	15527	19660	16000	2	983	800	17039872
279	406	1883	4/3/13 14:20:47.573	15254	17667	19660	16000	2	983	800	17039872
280	406	1883	4/3/13 14:20:47.823	17294	18767	19660	16000	2	983	800	17039872
281	406	1883	4/3/13 14:20:48.087	17814	18767	19660	16000	2	983	800	17039872
282	406	1883	4/3/13 14:20:49.120	17574	19767	19660	16000	2	983	800	17039872
283	406	1883	4/3/13 14:20:49.387	17434	20647	19660	16000	2	983	800	17039872
284	406	1883	4/3/13 14:20:50.683	17474	20647	19660	16000	2	983	800	17039872
285	406	1883	4/3/13 14:20:50.947	21014	21107	19660	16000	2	983	800	17039872
286	406	1883	4/3/13 14:20:51.230	22854	21827	19660	16000	2	983	800	17039872
287	406	1883	4/3/13 14:20:51.510	22874	21827	19660	16000	2	983	800	17039872
288	406	1883	4/3/13 14:20:52.807	27789	25827	9830	8000	4	983	800	17039872
289	406	1883	4/3/13 14:20:54.900	27789	25837	9830	8000	4	983	800	17039872
290	406	1883	4/3/13 14:20:55.167	27799	27717	9830	8000	4	983	800	17039872
291	406	1883	4/3/13 14:20:55.447	27929	28457	9830	8000	4	983	800	17039872
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

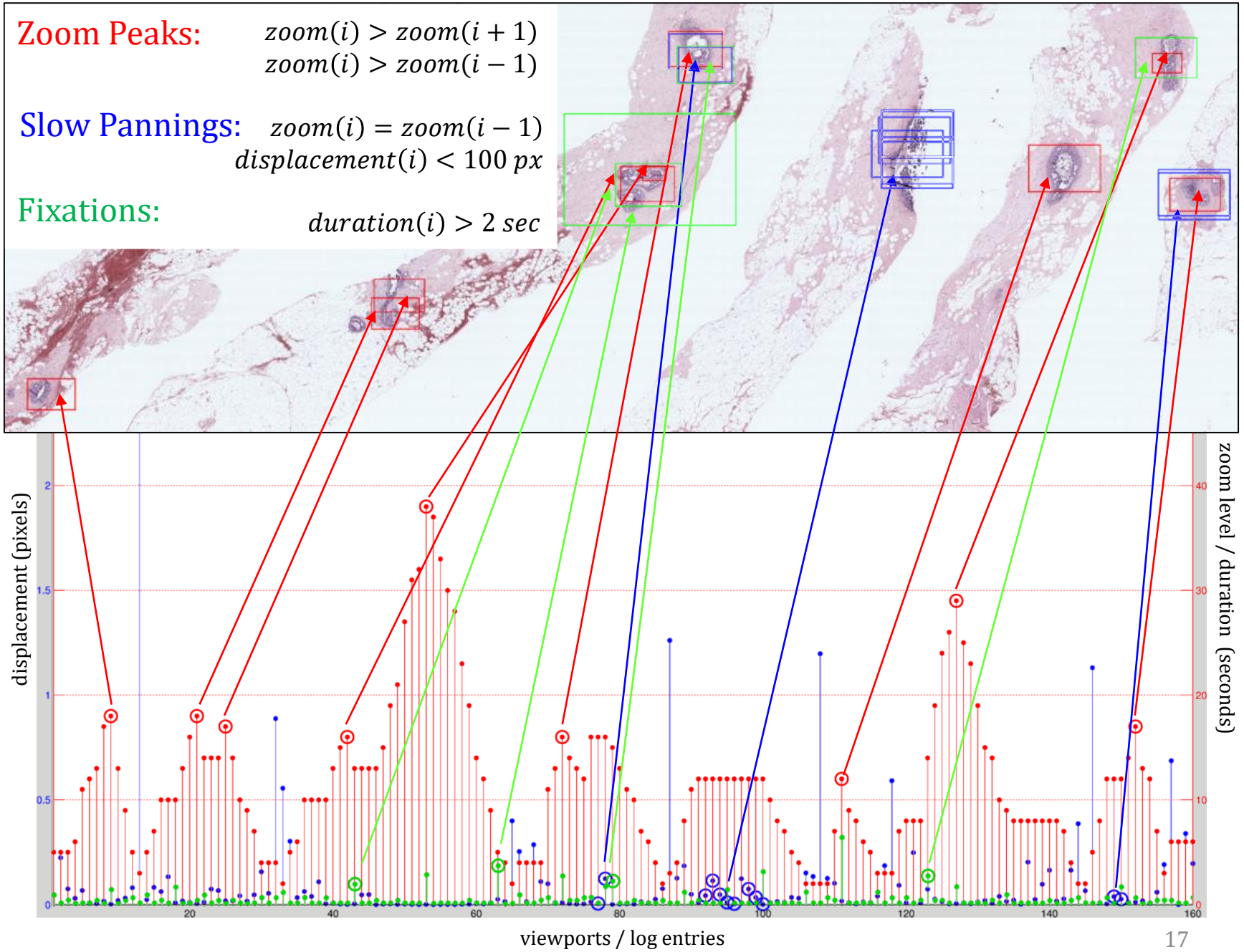


Zoom Level - The value of the **magnification** per each log entry. Higher the zoom, the smaller the rectangle.

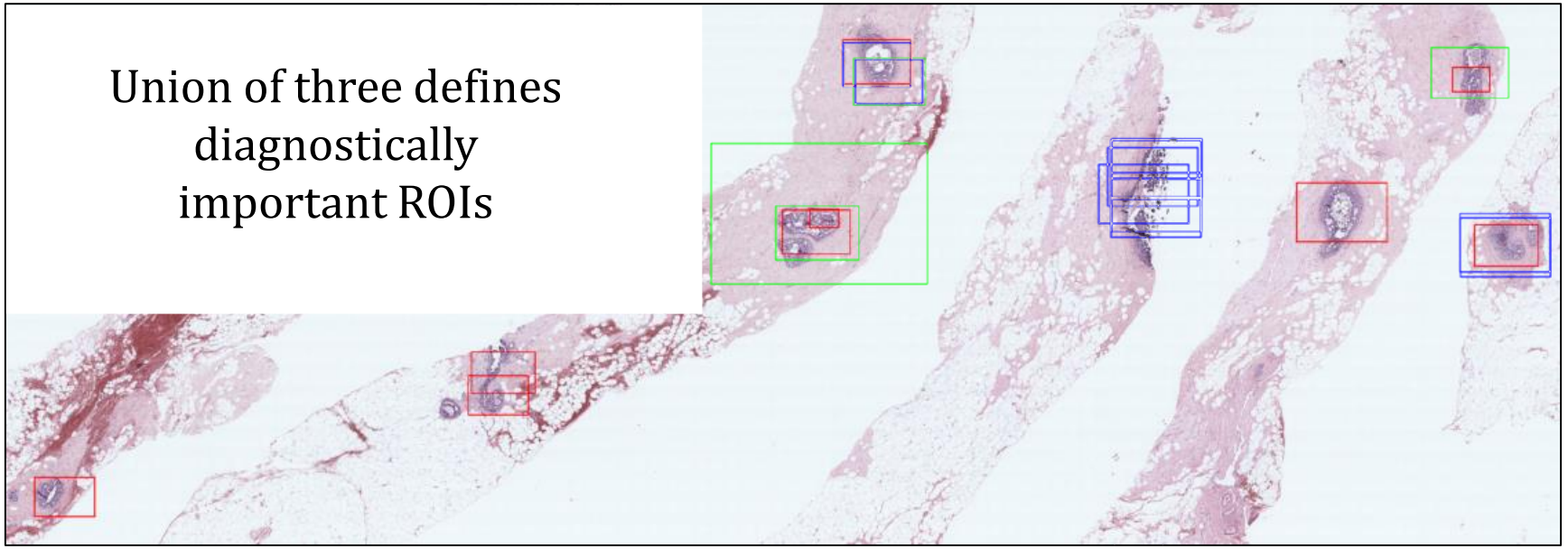
Displacement - The **distance** between the two consecutive viewport rectangles **in pixels.**

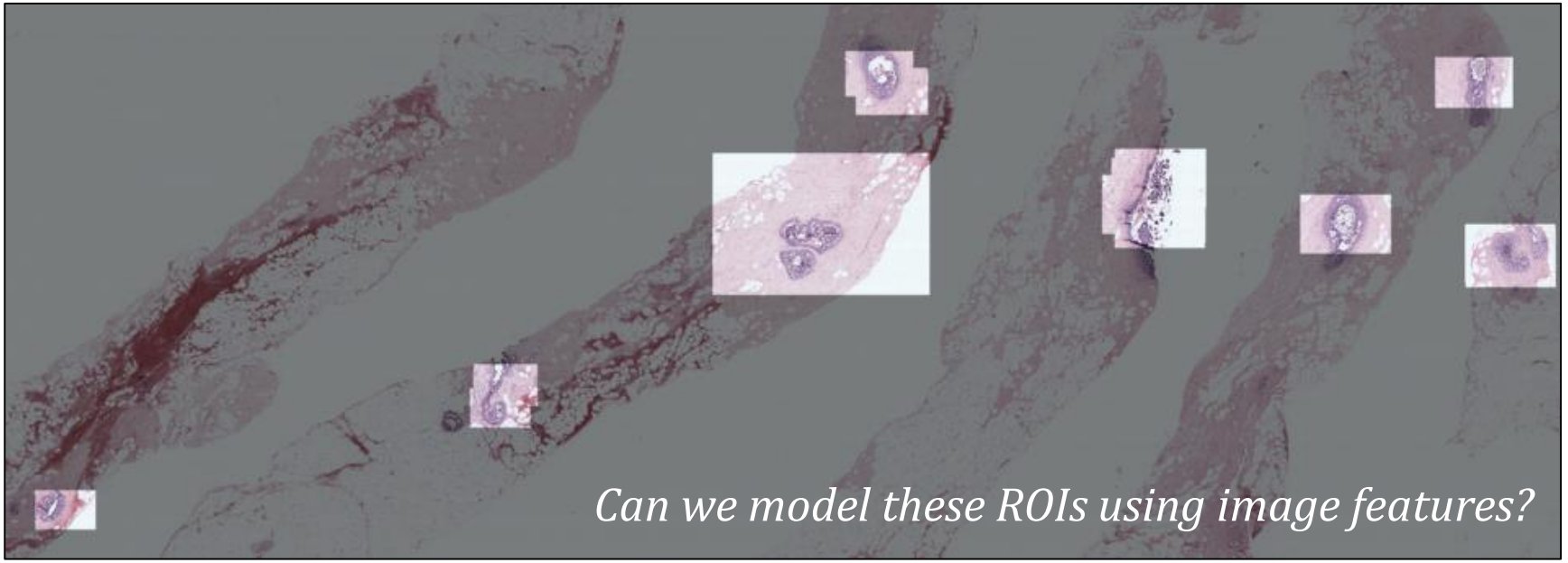
Duration - The **amount of time** a pathologist spent at each viewport rectangle.




Zoom Peaks:
$zoom(i) > zoom(i+1)$
$zoom(i) > zoom(i-1)$
Slow Pannings:
$zoom(i) = zoom(i-1)$
$displacement(i) < 100\ px$
Fixations:
$duration(i) > 2\ sec$
displacement (pixels)
zoom level / duration (seconds)
viewports / log entries

Union of three defines diagnostically important ROIs

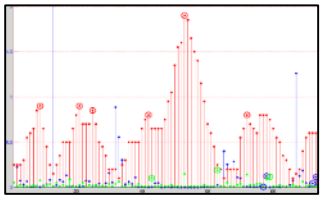




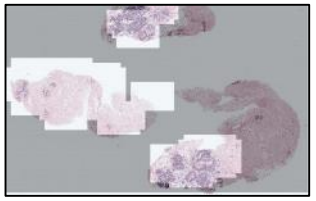
ROI Localization in WSIs

Tracking Data Analysis to Discover ROIs

tracking data

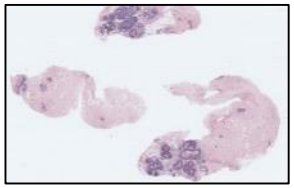


our viewport analysis

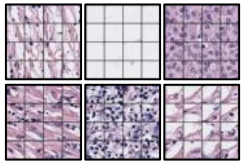


ROIs to which pathologist actually paid attention

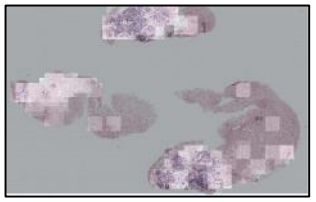
A Model based on Color and Texture to Predict ROIs



an unseen WSI

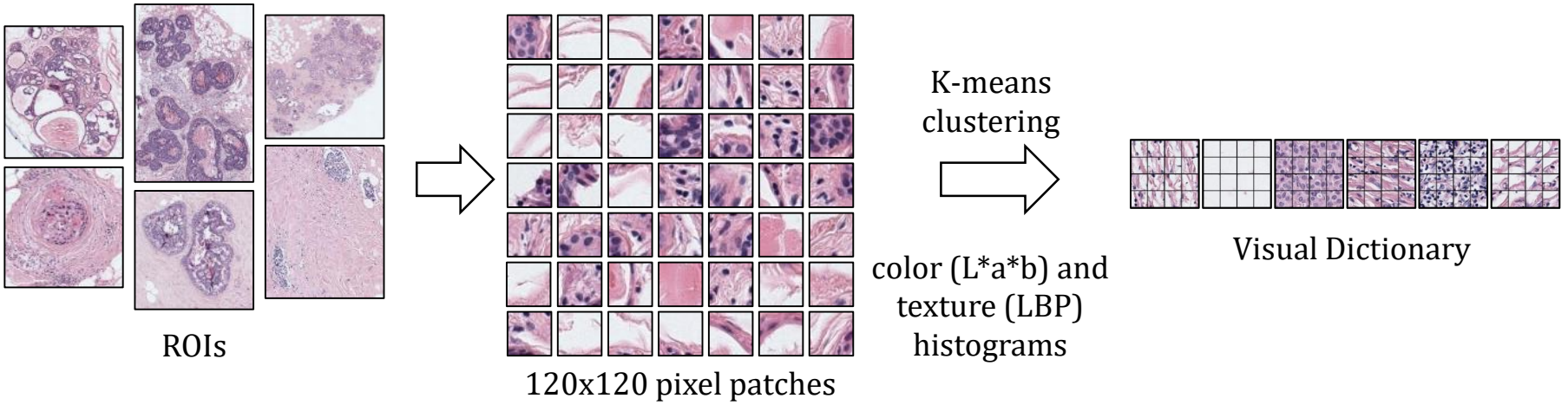


our computational model based on color and texture



our predicted ROIs

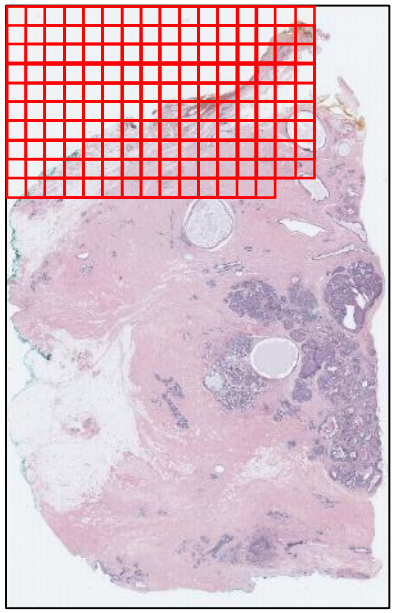
Learning Visual Dictionary



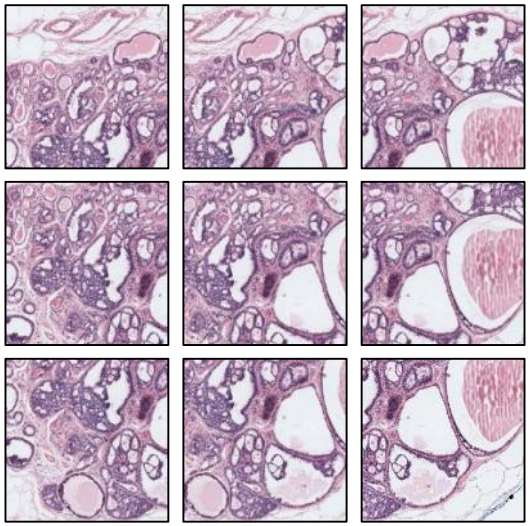
- Words are obtained from ROIs.
- Each pixel patch is a visual *word.*
- We calculated color and texture histograms from each visual word.
- Using k-means clustering we obtained a visual dictionary.

Training: Sliding Window

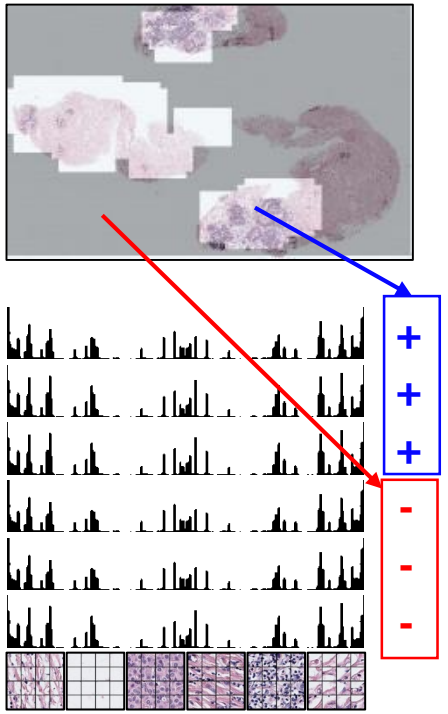
- Visual bag : overlapping sliding windows
- Each visual bag has 900 visual words.
- Bag-of-words: a histogram of *K* visual words from the visual dictionary
- Bags inside the ROIs are positive samples.
  Bags outside the ROIs are negative samples.



whole slide images



3600x3600 pixel
overlapping sliding windows

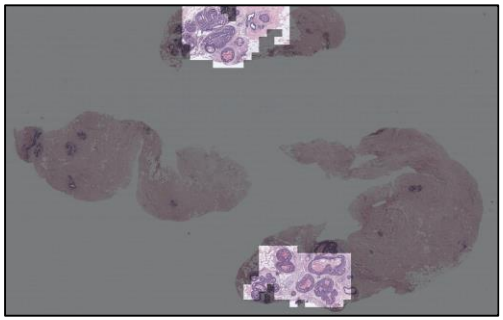


a histogram (bag of words)
for each sliding window

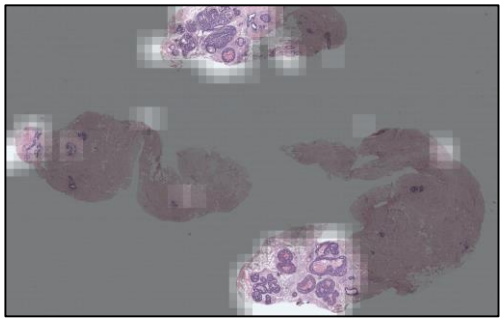
Evaluation

- Each sliding window is a sample.
- Evaluation metric: Accuracy

$$accuracy = \frac{\#\ correctly\ classified\ sliding\ windows}{total\ \#\ sliding\ windows}$$



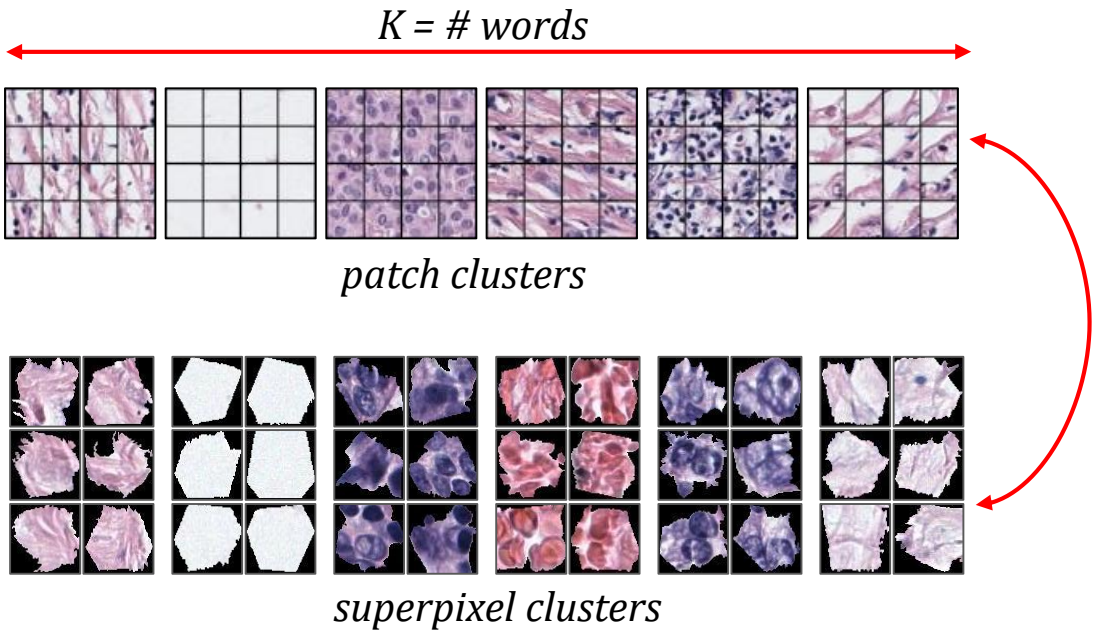
Ground Truth
(from viewport analysis)



Prediction
(from visual bag-of-words model)

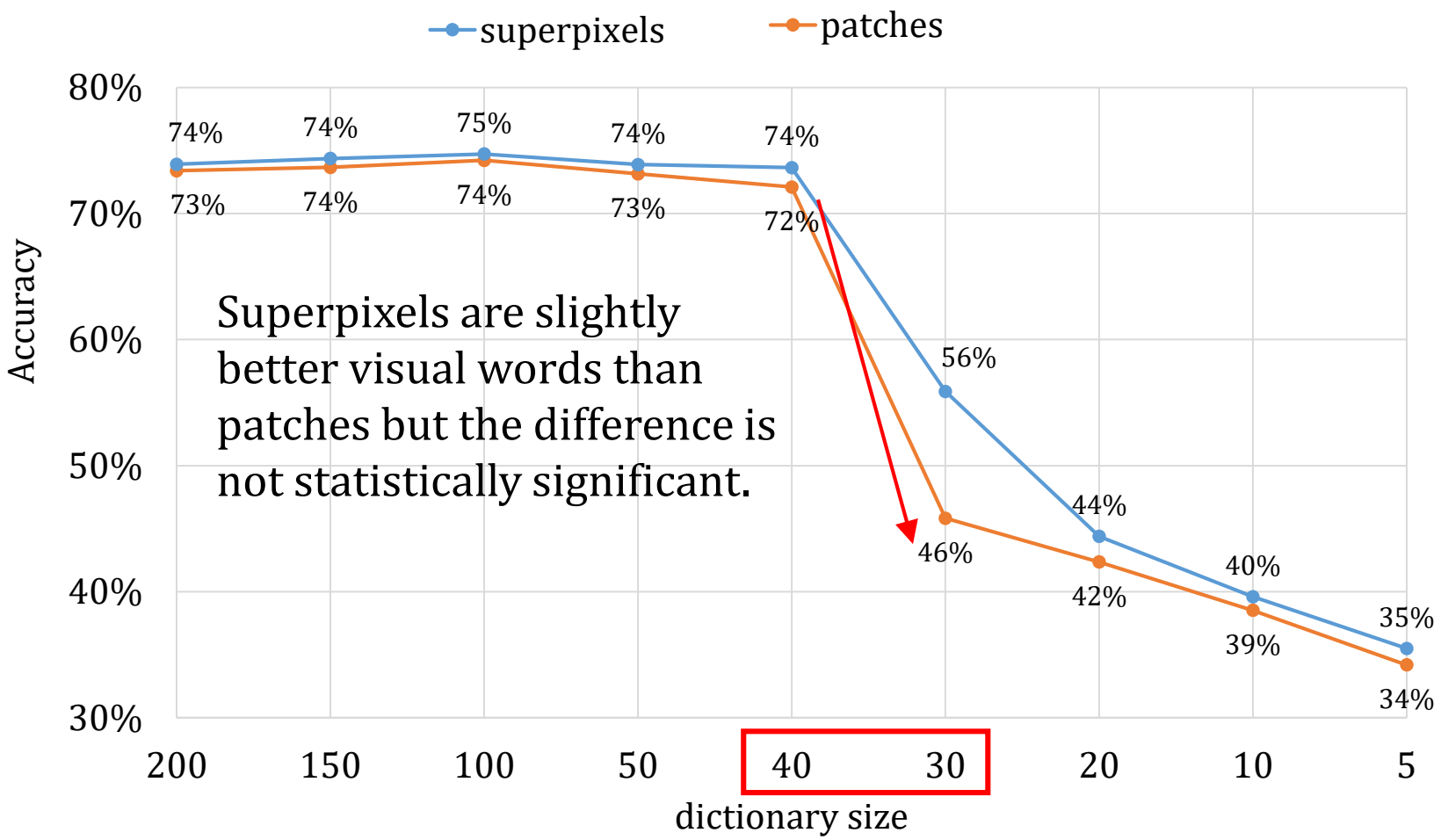
Experiments

- Dictionary size:
  - K = 200 ⟶ 5
- Visual words:
  - patches
  - superpixels



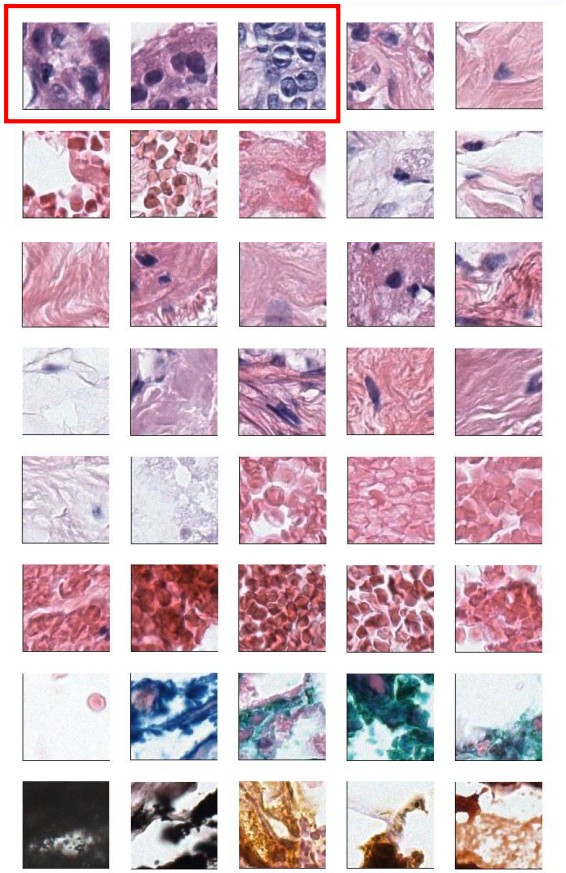
- We ran 10-fold cross-validation experiments using Logistic Regression.
- Dataset:
  - N=240 whole slide images
  - Expert viewport logs.

Prediction Results

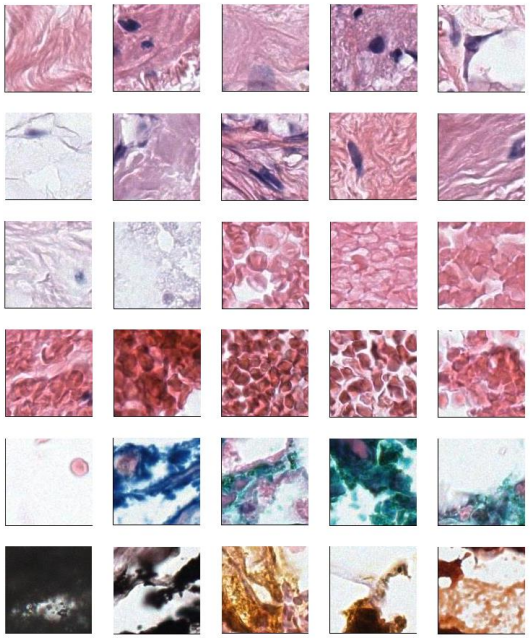


Dictionary Size

K = 40



K = 30

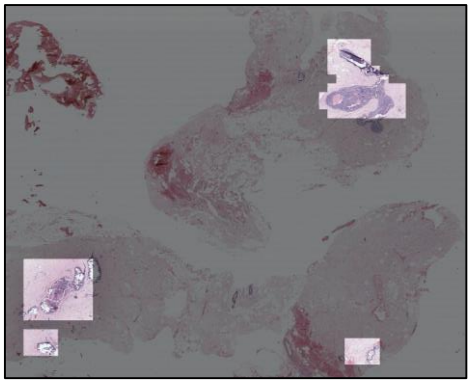


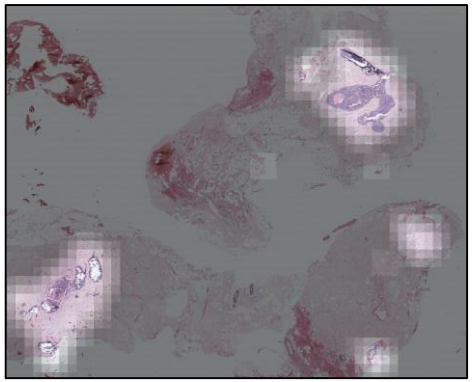
Removing the visual words that describe the **epithelium** reduces the accuracy significantly.

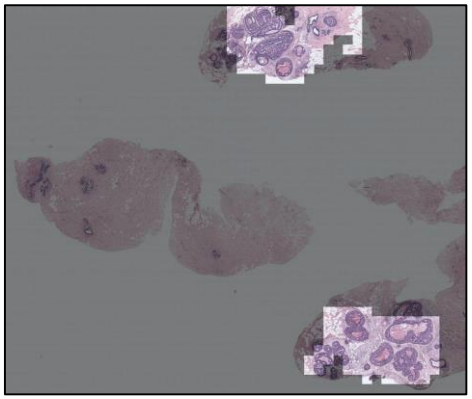
Results

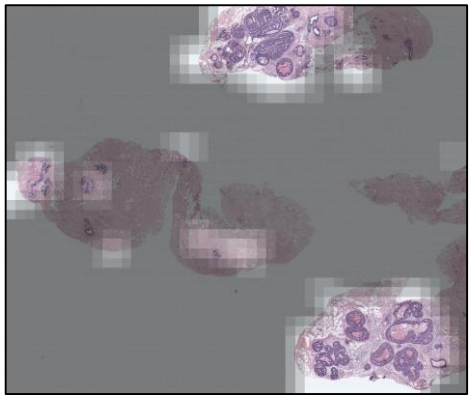
Ground Truth
(from viewport analysis)

Prediction









Summary

- We presented
  - a novel viewport log analysis to detect diagnostically relevant ROIs from pathologists' actions, and
  - a visual bag-of-words model based on color and texture features to predict ROIs in unseen WSIs.
- 74% prediction accuracy in 240 WSIs.

Outline

1. Introduction
2. ROI Localization in WSIs
3. **Tissue Label Segmentation**
4. Automated Diagnosis of ROIs
5. Viewing Behavior Analysis
6. Conclusions

digiPATH Dataset

case
(n=240)

digital WSI
consensus ROI
consensus diagnosis

biopsy type
breast density
patient age

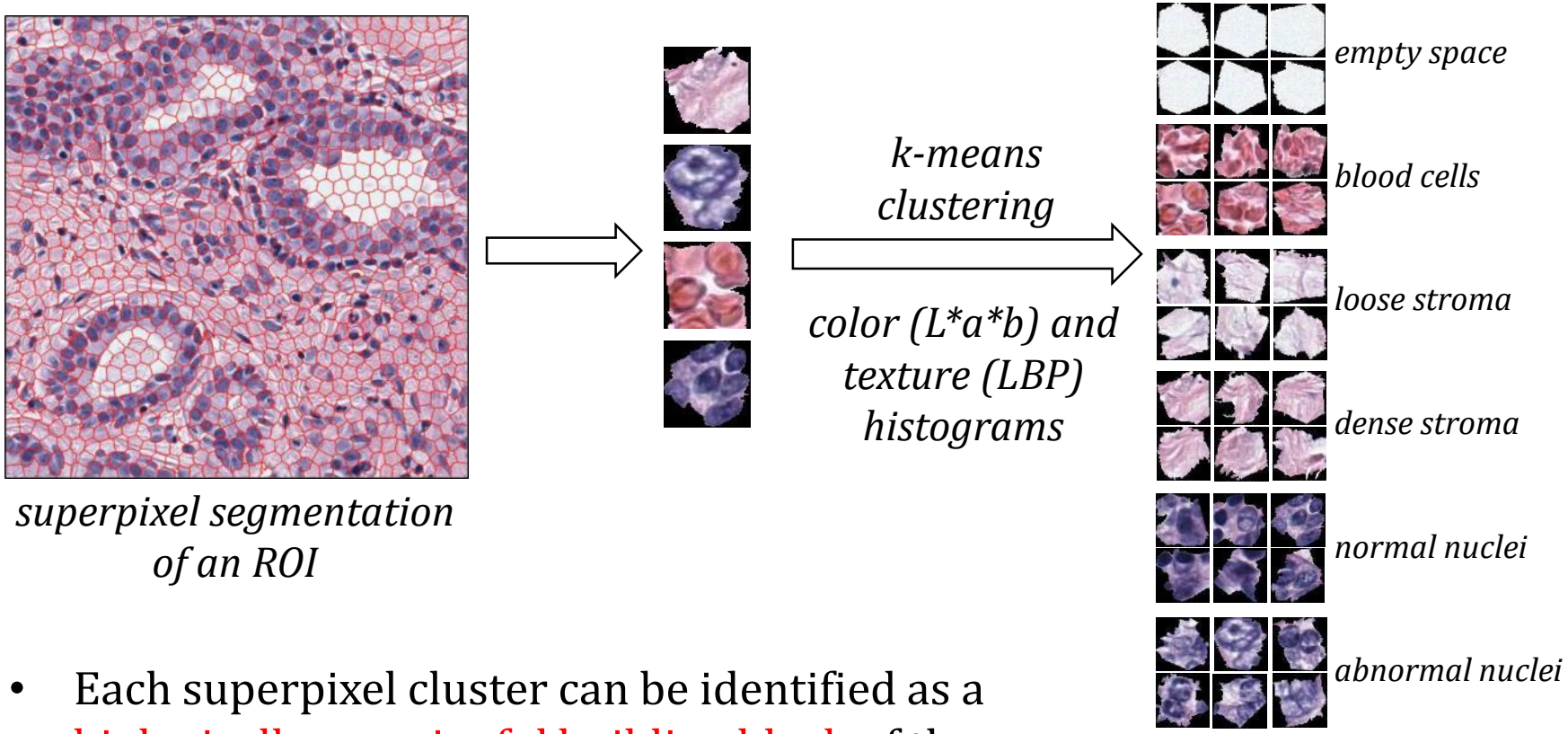
Tissue Label Segmentation

- For an automated diagnosis system, we need to describe the structural changes that lead to cancer.

- Segmentation is a powerful that provides information about the distribution and arrangement of different tissue types.

S. Mehta, E. Mercan, *et al.*, British Machine Vision Conference, 2017 (submitted).

Superpixel Clustering

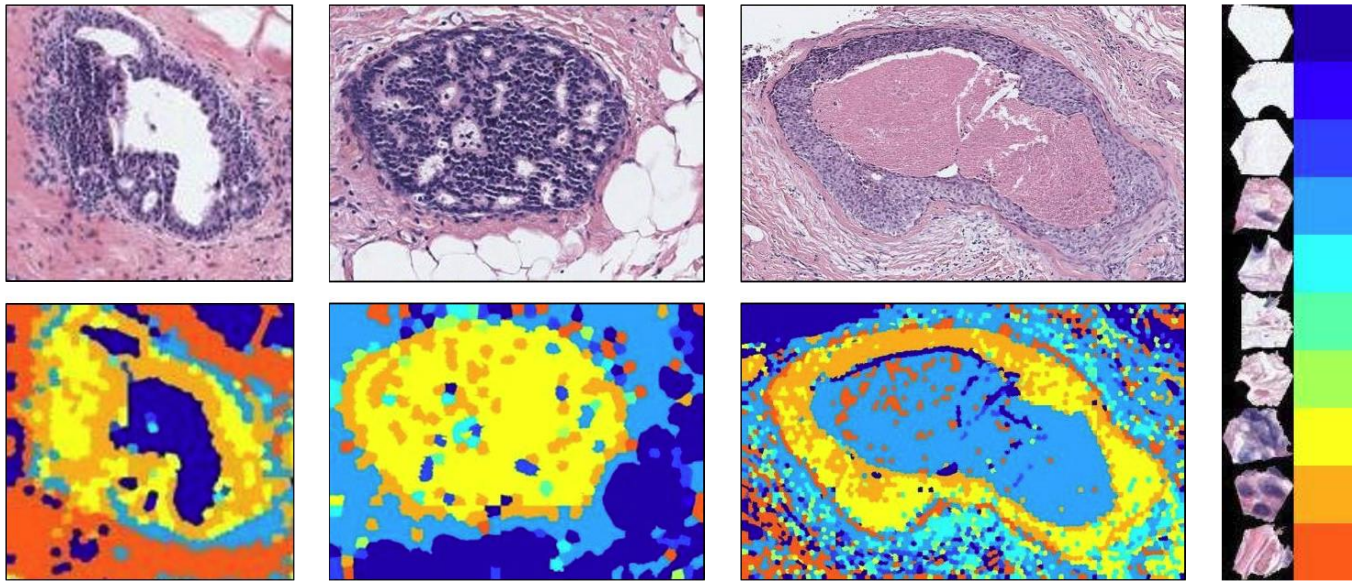


superpixel segmentation of an ROI

superpixel clusters

- Each superpixel cluster can be identified as a biologically meaningful building block of the tissue.

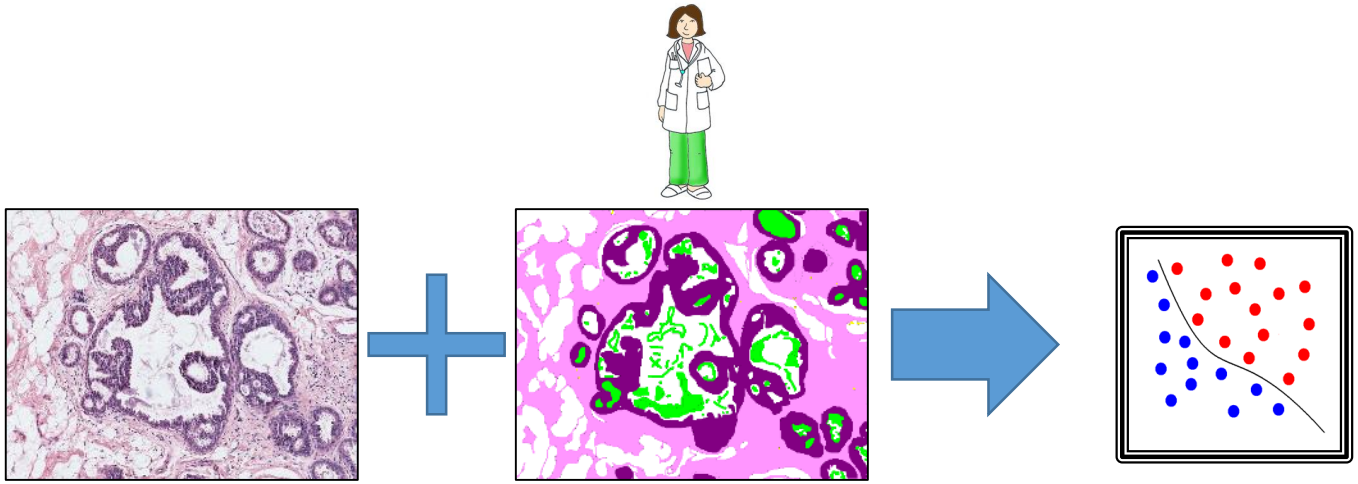
Superpixel Clustering



benign *atypia* *DCIS*

- Patterns emerge when we label the superpixels in an unsupervised manner.

Supervised Tissue Label Segmentation



original image

ground truth label image

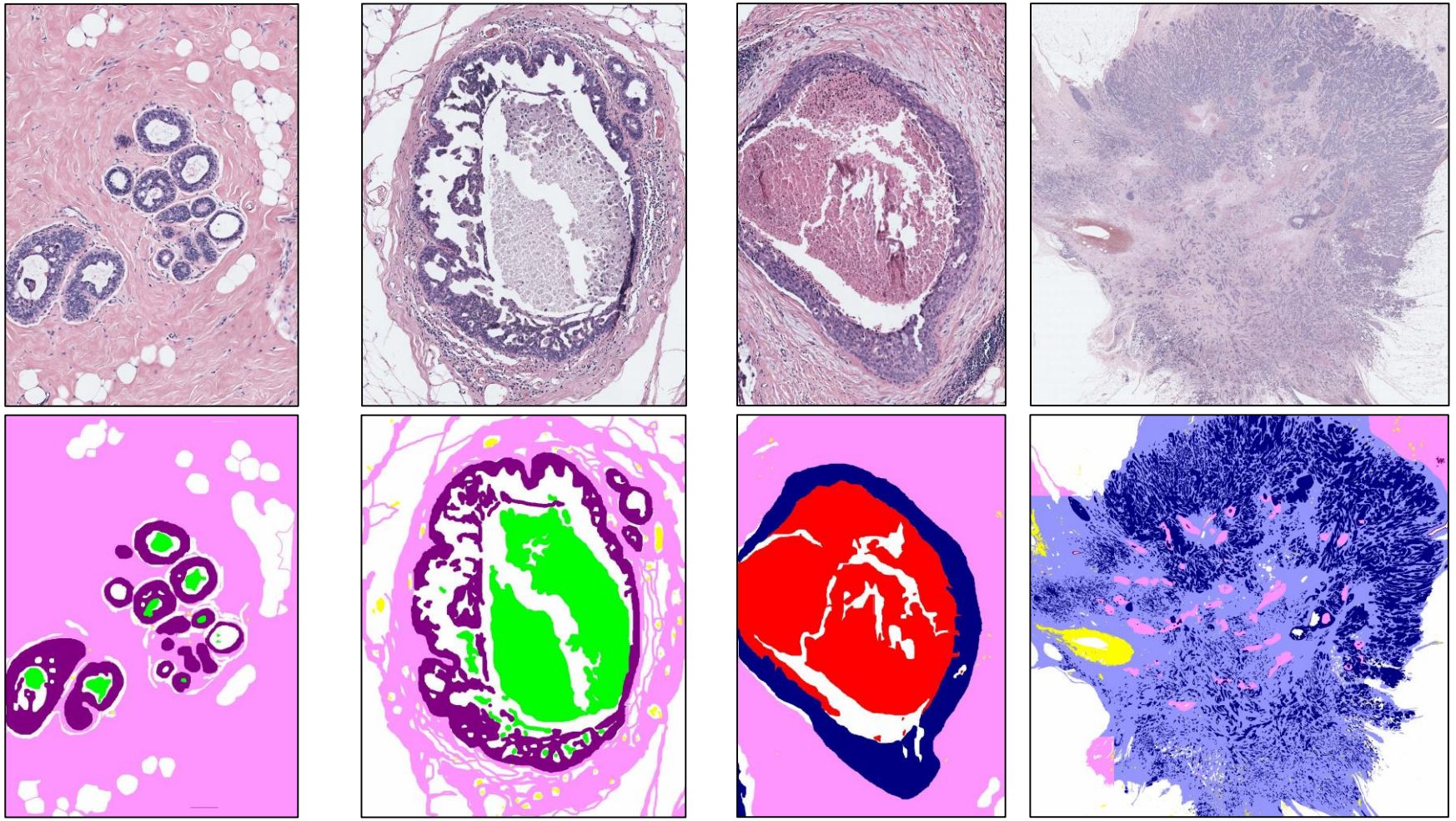
classifier

We tested two models using a subset of ROIs (N=58):

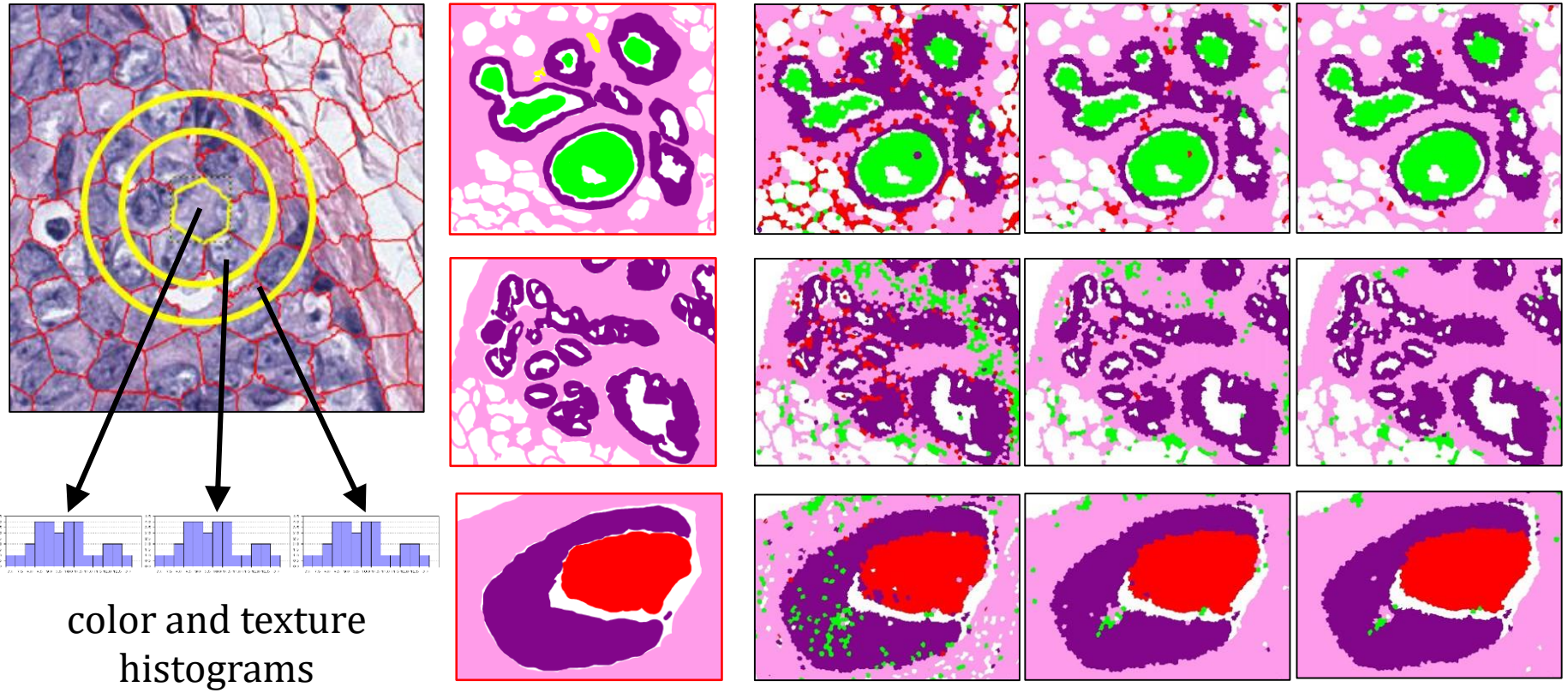
- Support Vector Machines (SVM)
- Convolutional Neural Nets (CNN)

Training Labels

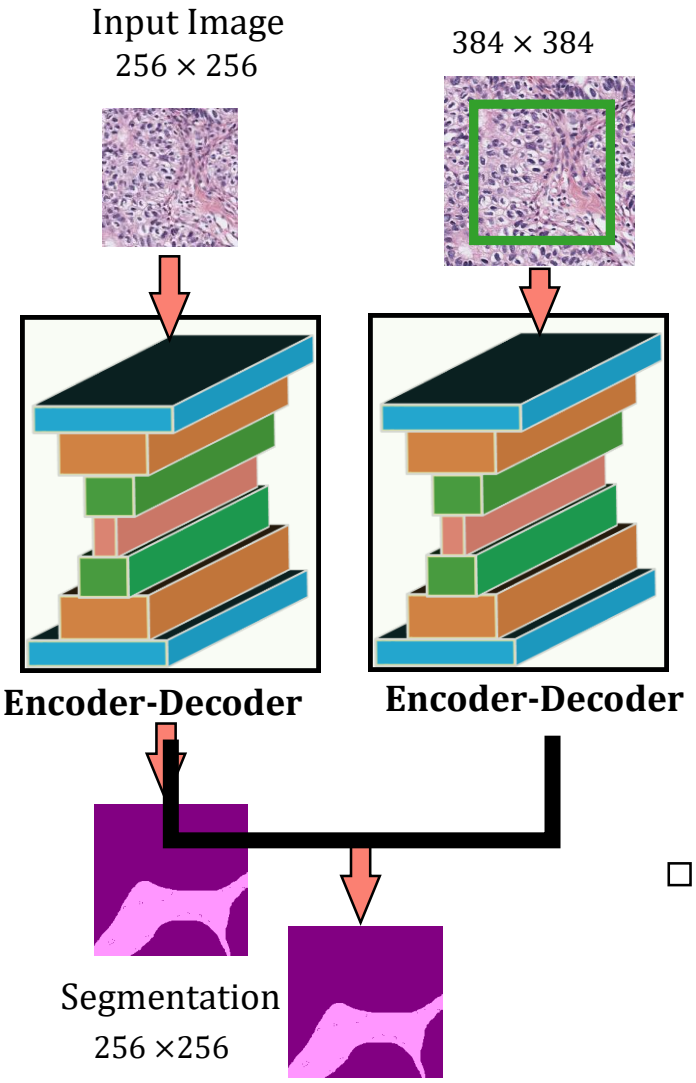
background benign epithelium normal stroma secretion necrosis
malignant epithelium desmoplastic stroma blood

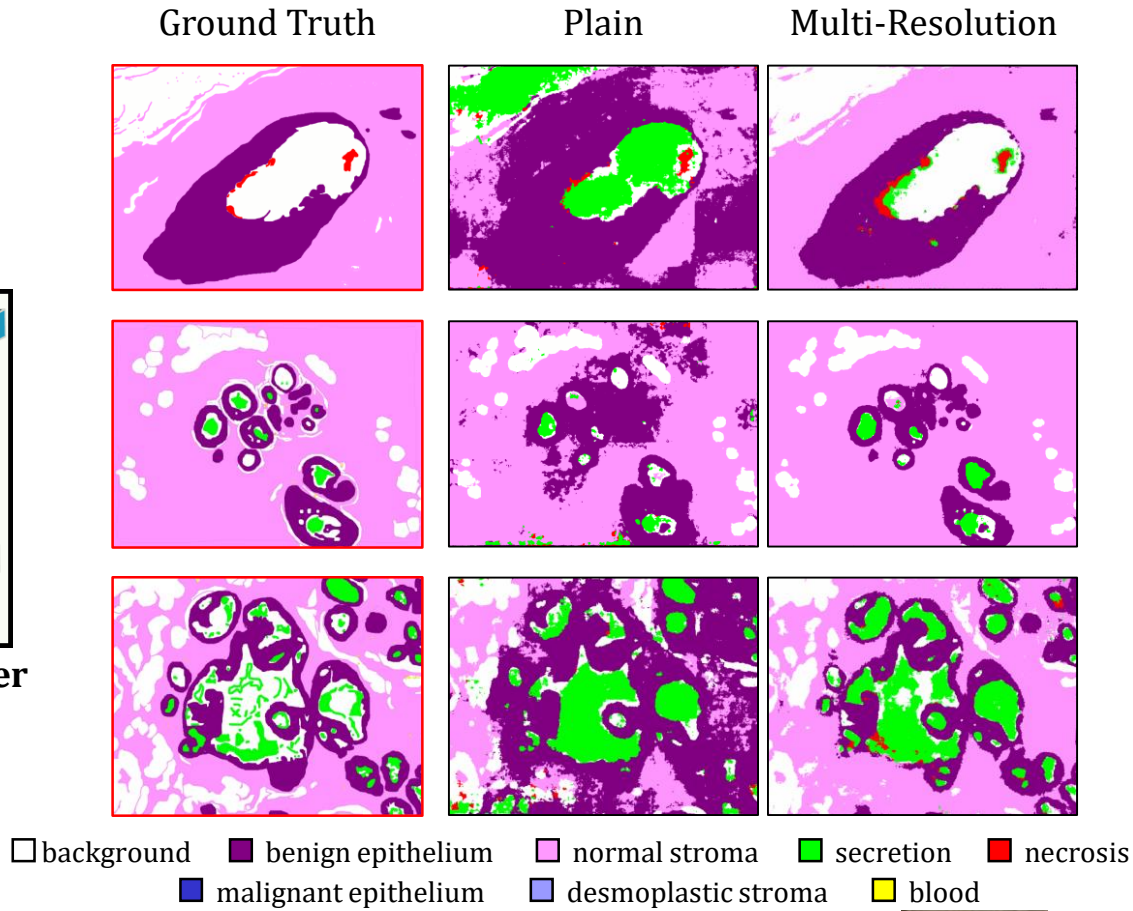


Superpixel + SVM-based Segmentation



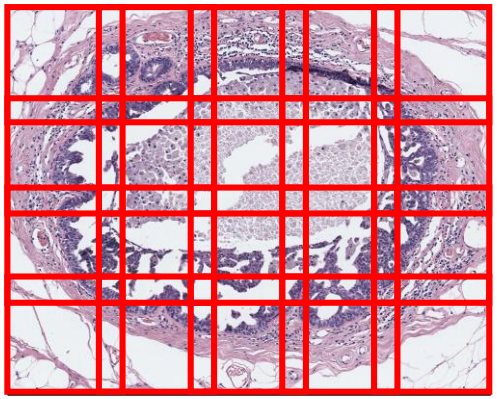
CNN-based Segmentation





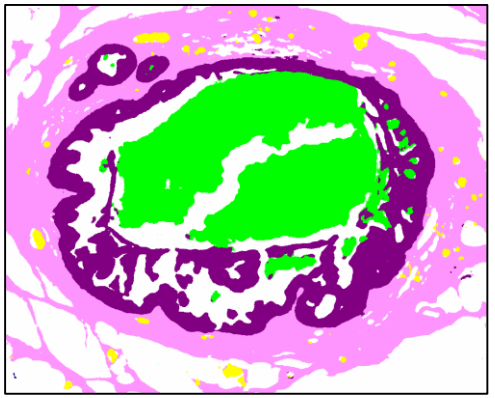
joint work with Sachin Mehta

CNN-based Segmentation

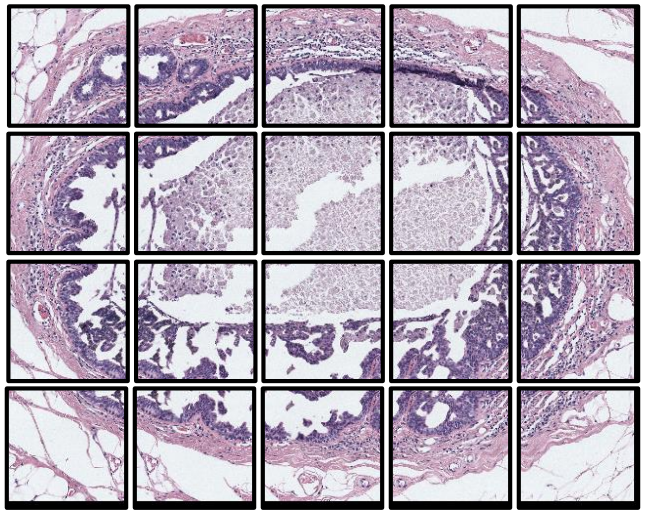


ROI

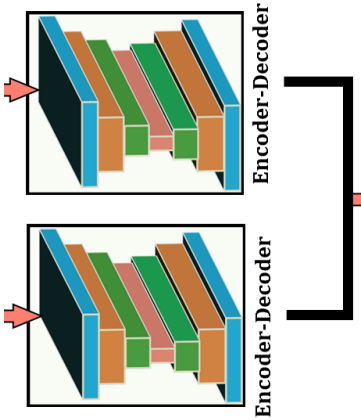
Training set: 38 ROIs
Test set: 20 ROIs



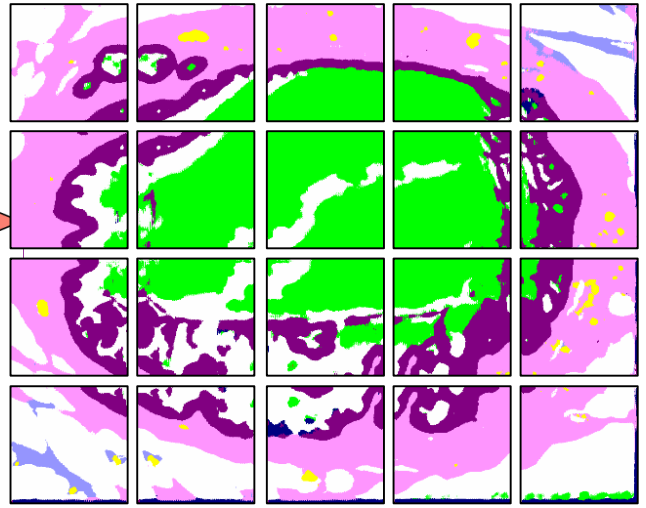
Segmented ROI



Overlapping Patches
256x256 pixel



CNN

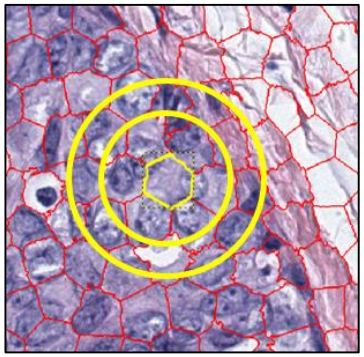


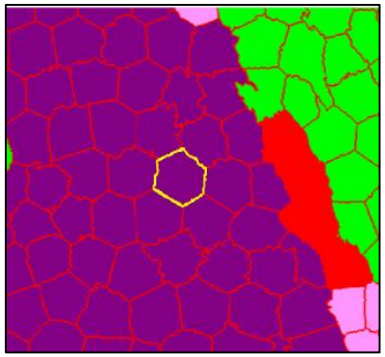
Segmented Patches

Supervised Tissue Label Segmentation

Superpixel + SVM

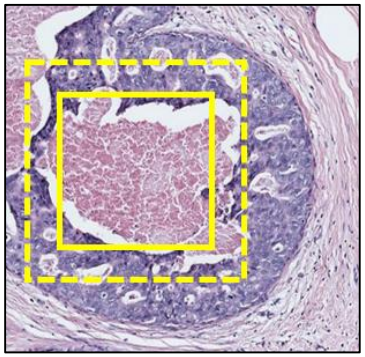
- Each superpixel is assigned a class label.
- Context: Two circular neighborhoods
- Relatively simple model
- Faster to train (~3 hours)

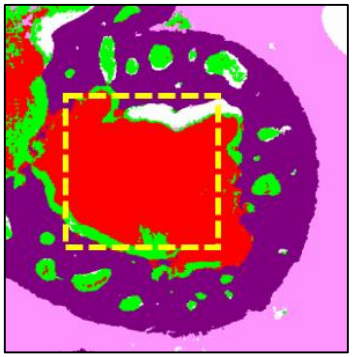




CNN

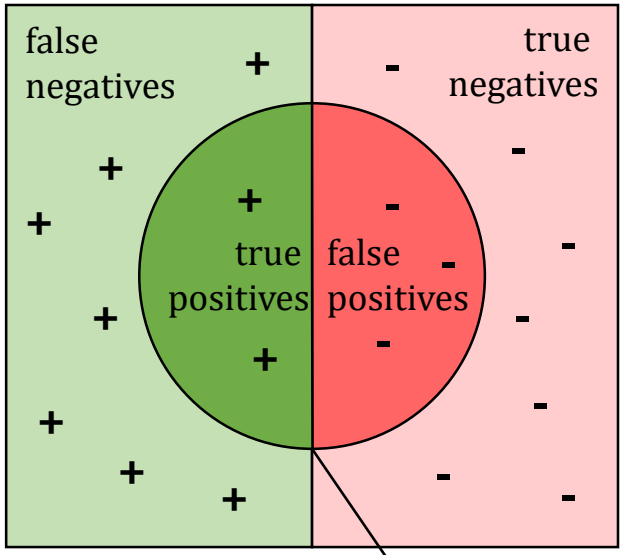
- Each pixel is assigned a class label.
- Context: 256x256 and 384x384 pixel patches
- More complex model
- ~1 week to train on special hardware





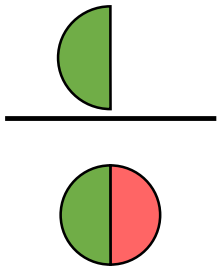
Evaluation

Each pixel is a sample.



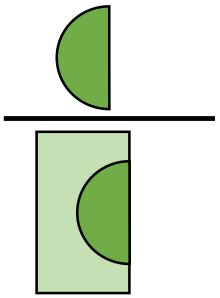
$\# tp = true\ positive$
$\# tn = true\ negative$
$\# fp = false\ positive$
$\# fn = false\ negative$

precision



$$\frac{\# tp}{\# tp + \# fp}$$

recall



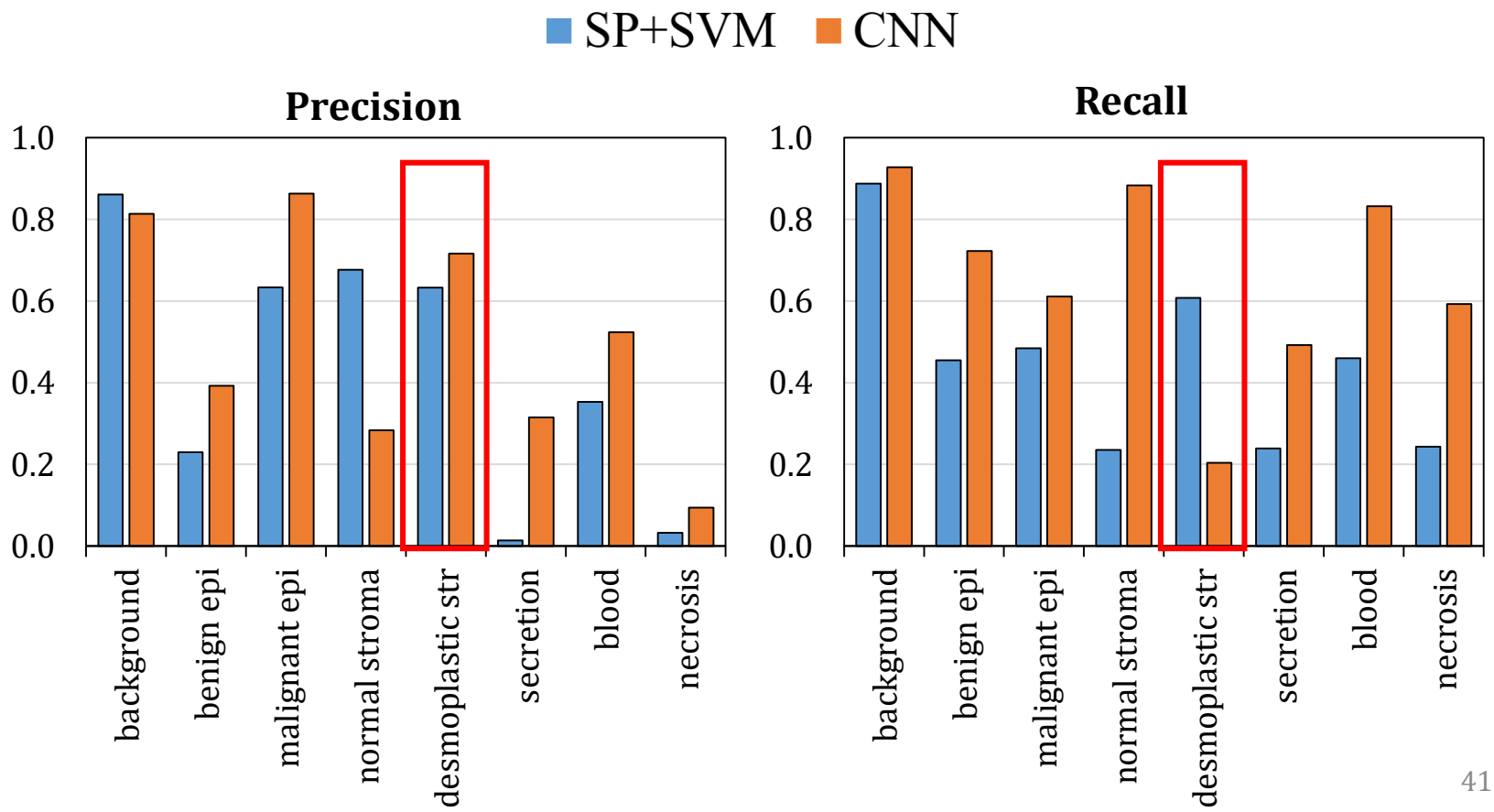
$$\frac{\# tp}{\# tp + \# fn}$$

$$F_1\ score = 2 \times \frac{precision \times recall}{precision + recall}$$

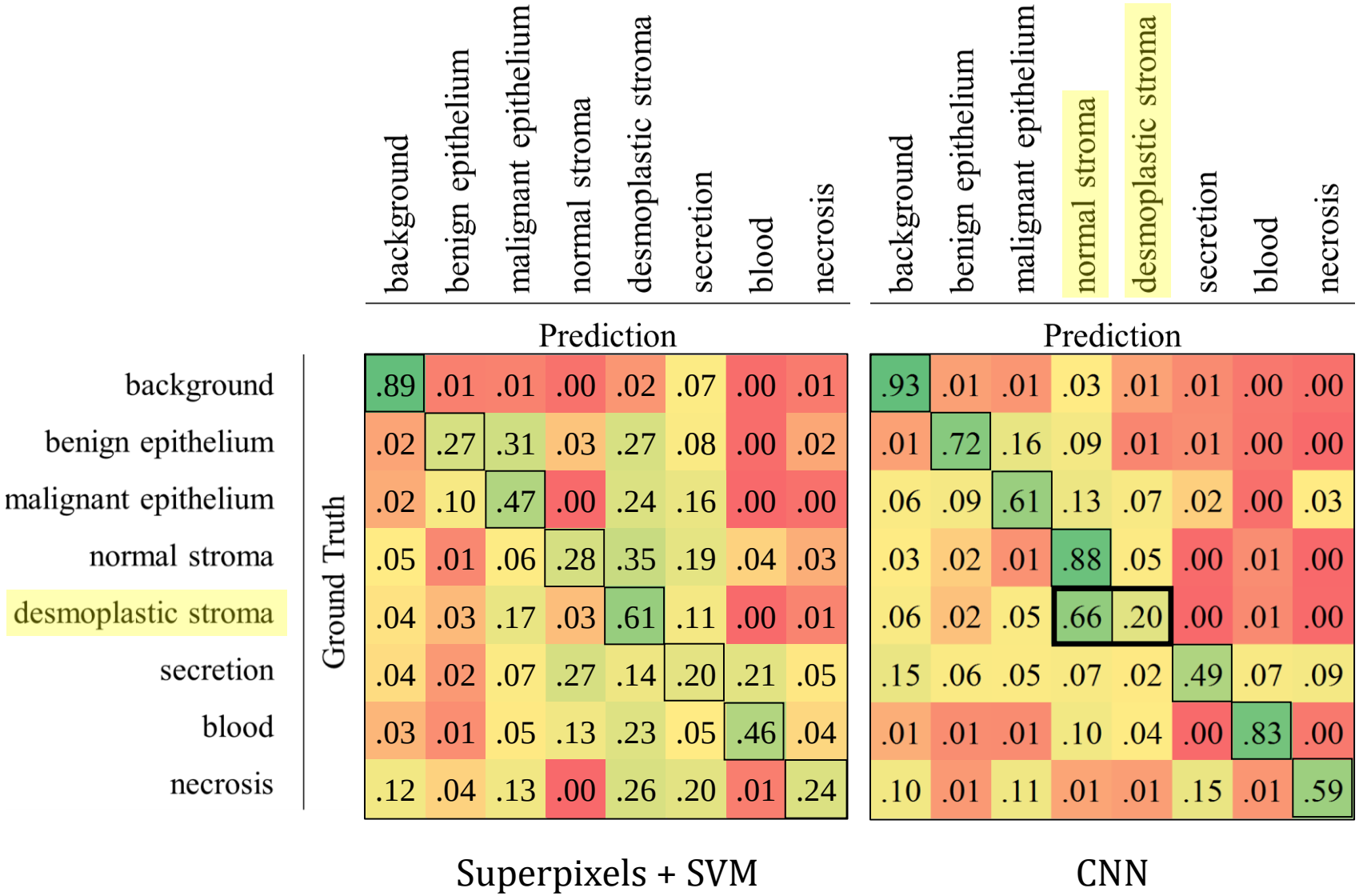
Results

Mean F_1-score

SP+SVM	0.40
CNN	0.50

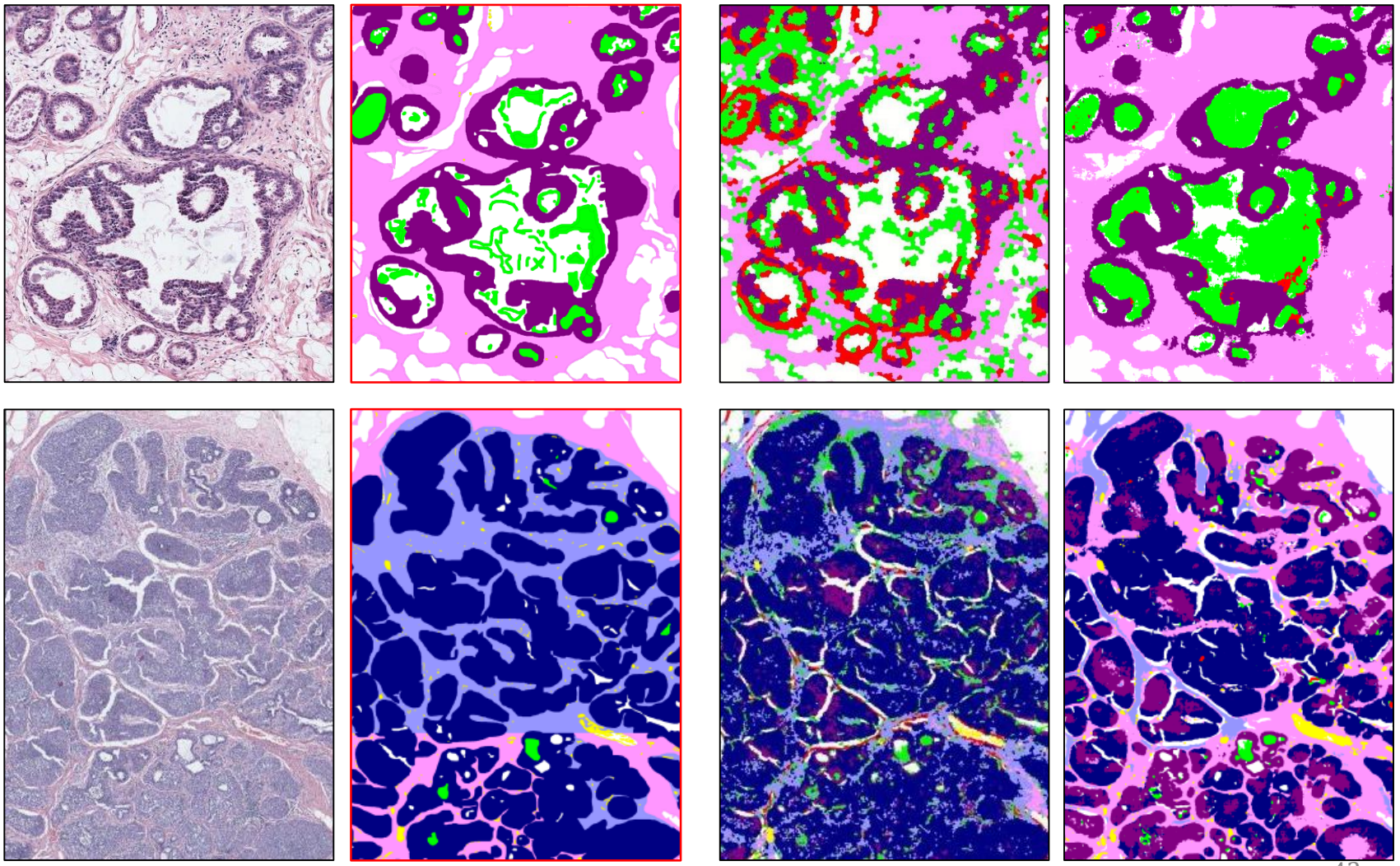


Confusion Matrices



Segmentation Results

RGB | Ground Truth Labels | SVM Predictions | CNN Predictions



background | benign epi | normal stroma | secretion | necrosis | malignant epi | desmoplastic stroma | blood

Summary

- Tissue-label segmentation is a useful abstraction.
- We developed a set of 8 tissue labels and collected pixel-label data from a pathologist on 58 ROIs.
- We trained two models: SVM and CNN
- CNNs performed significantly better than SVMs both quantitatively and qualitatively.

Outline

1. Introduction
2. ROI Localization in WSIs
3. Tissue Label Segmentation
4. **Automated Diagnosis of ROIs**
5. Viewing Behavior Analysis
6. Conclusions

digiPATH Dataset

case
(n=240)

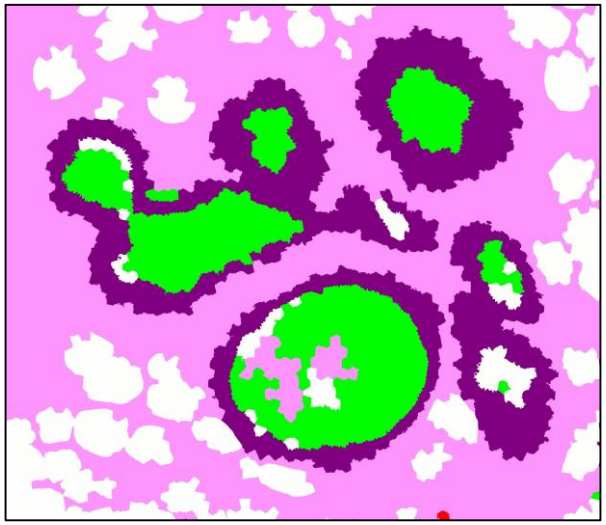
digital WSI
consensus ROI
consensus diagnosis

biopsy type
breast density
patient age

Automated Diagnosis of ROIs

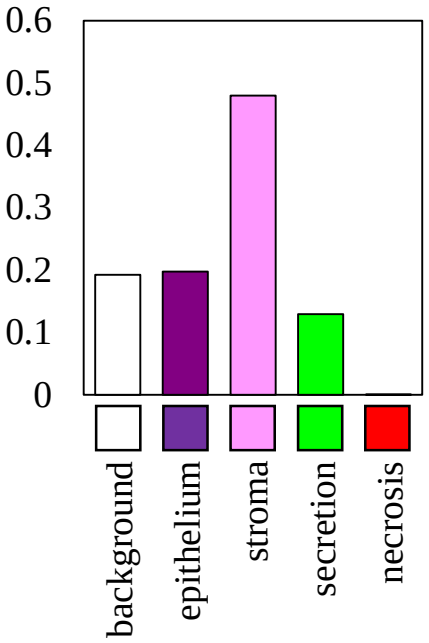
- Diagnostic errors are alarmingly high for pre-invasive lesions of the breast.
- In the digiPATH study, the agreement between pathologists and experts for the atypia cases is only 48%.

- Novel image features for diagnosis can help
  - develop computer aided diagnosis systems, and
  - study the reasons for diagnostic errors.

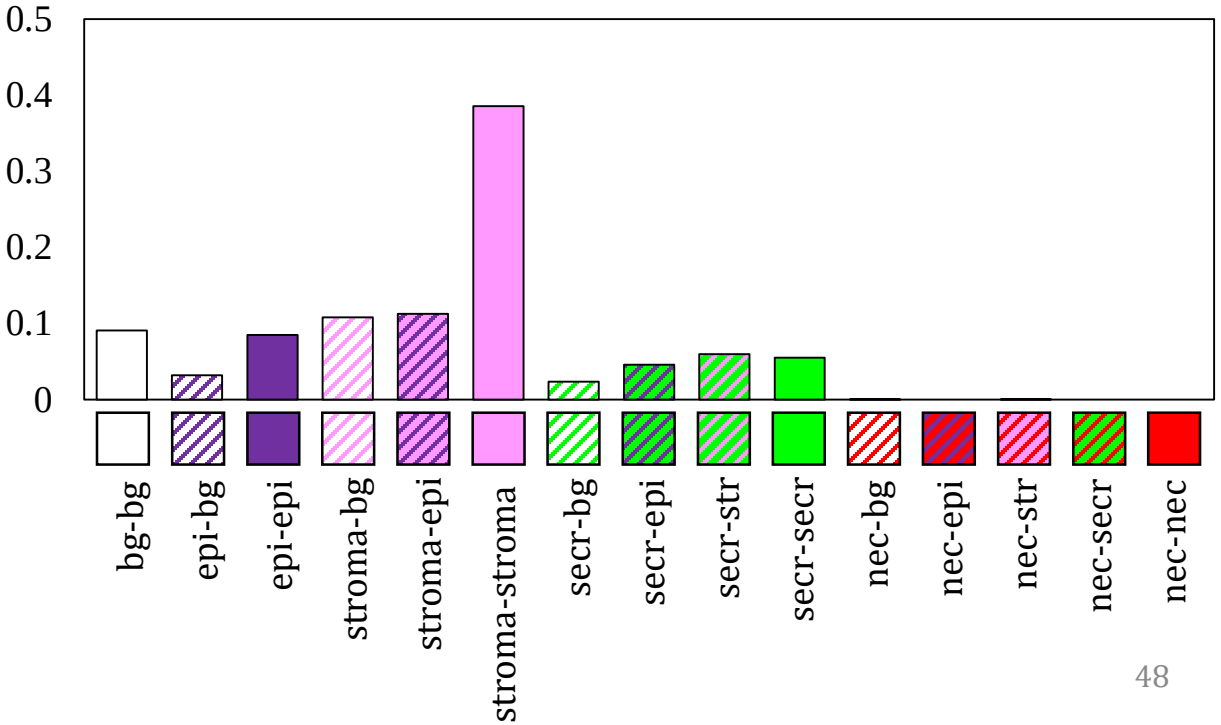


Superpixel Label Frequency and Co-occurrence Histograms

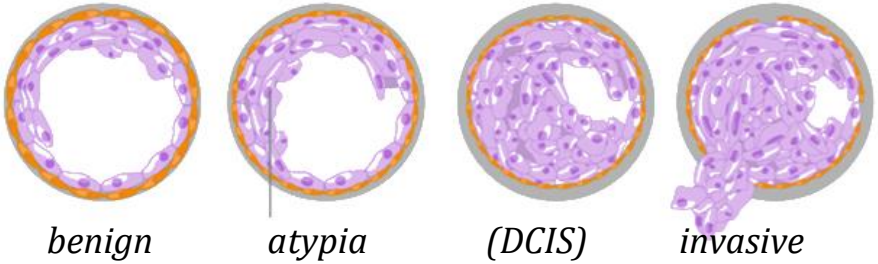
Frequency Histogram

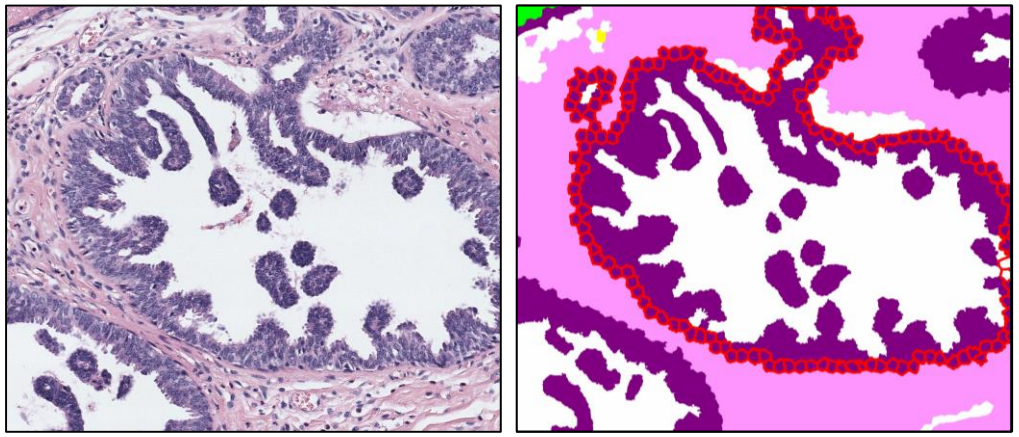


Co-occurrence Histogram

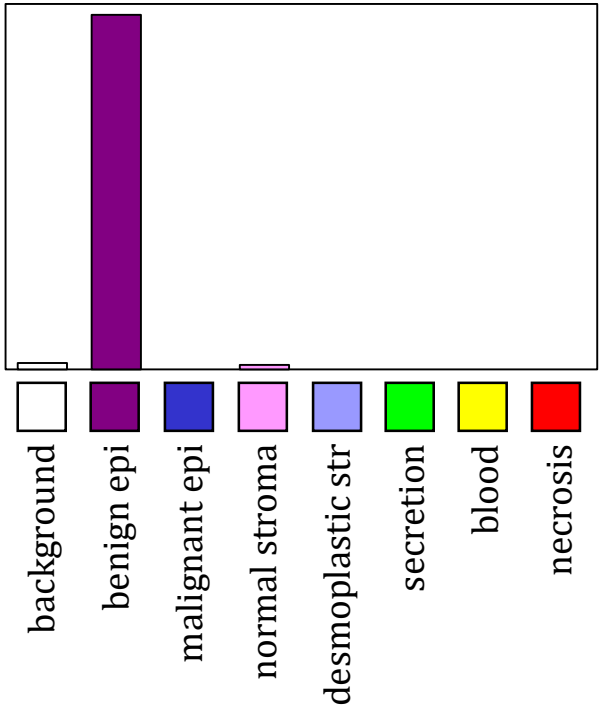


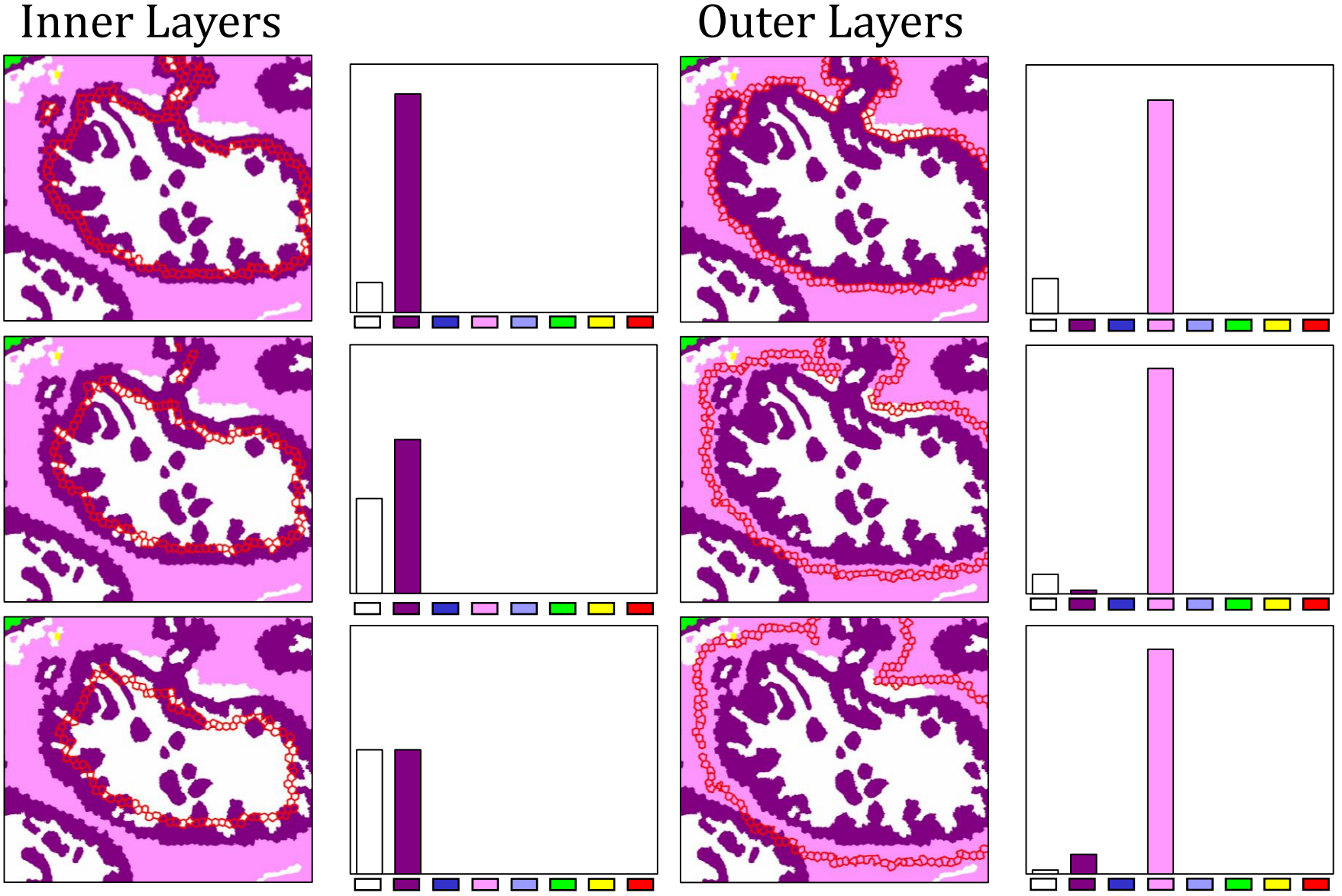
Structure Feature



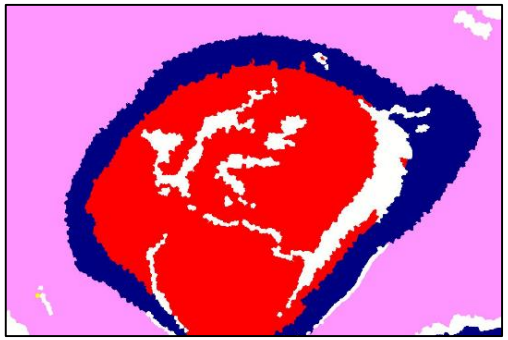


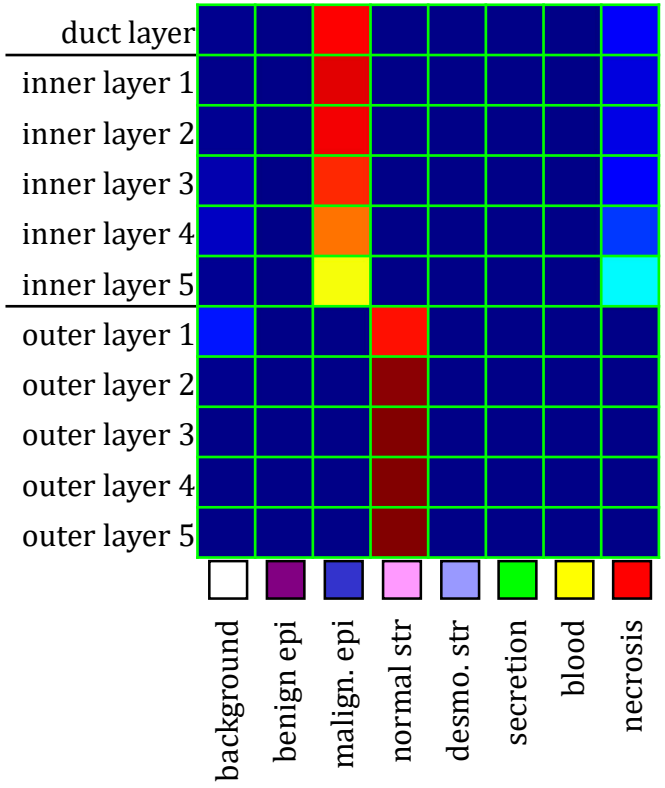
duct layer

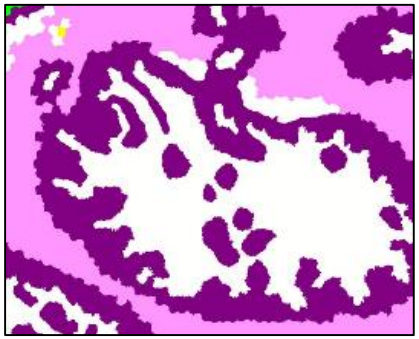


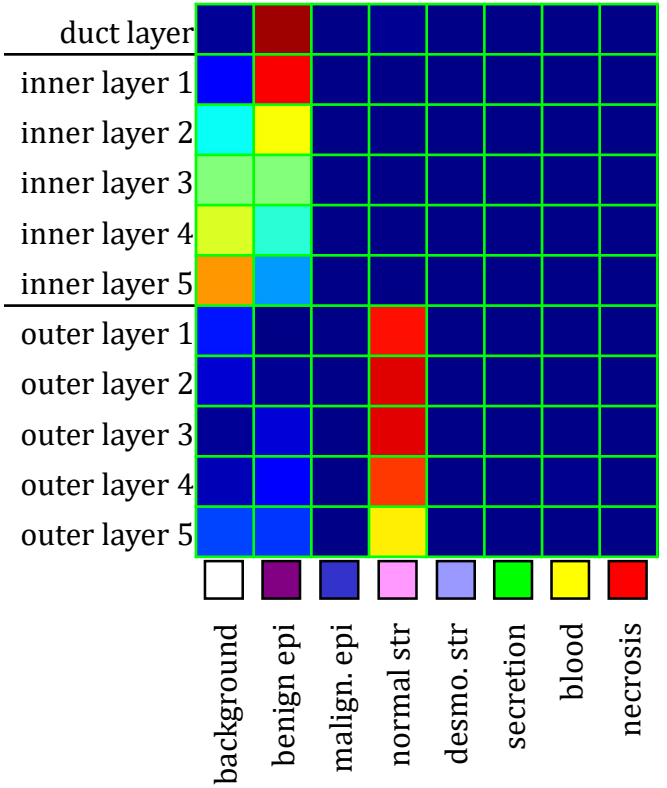

Inner Layers
Outer Layers
background
benign epithelium
malignant epithelium
normal stroma
desmoplastic stroma
secretion
blood
necrosis

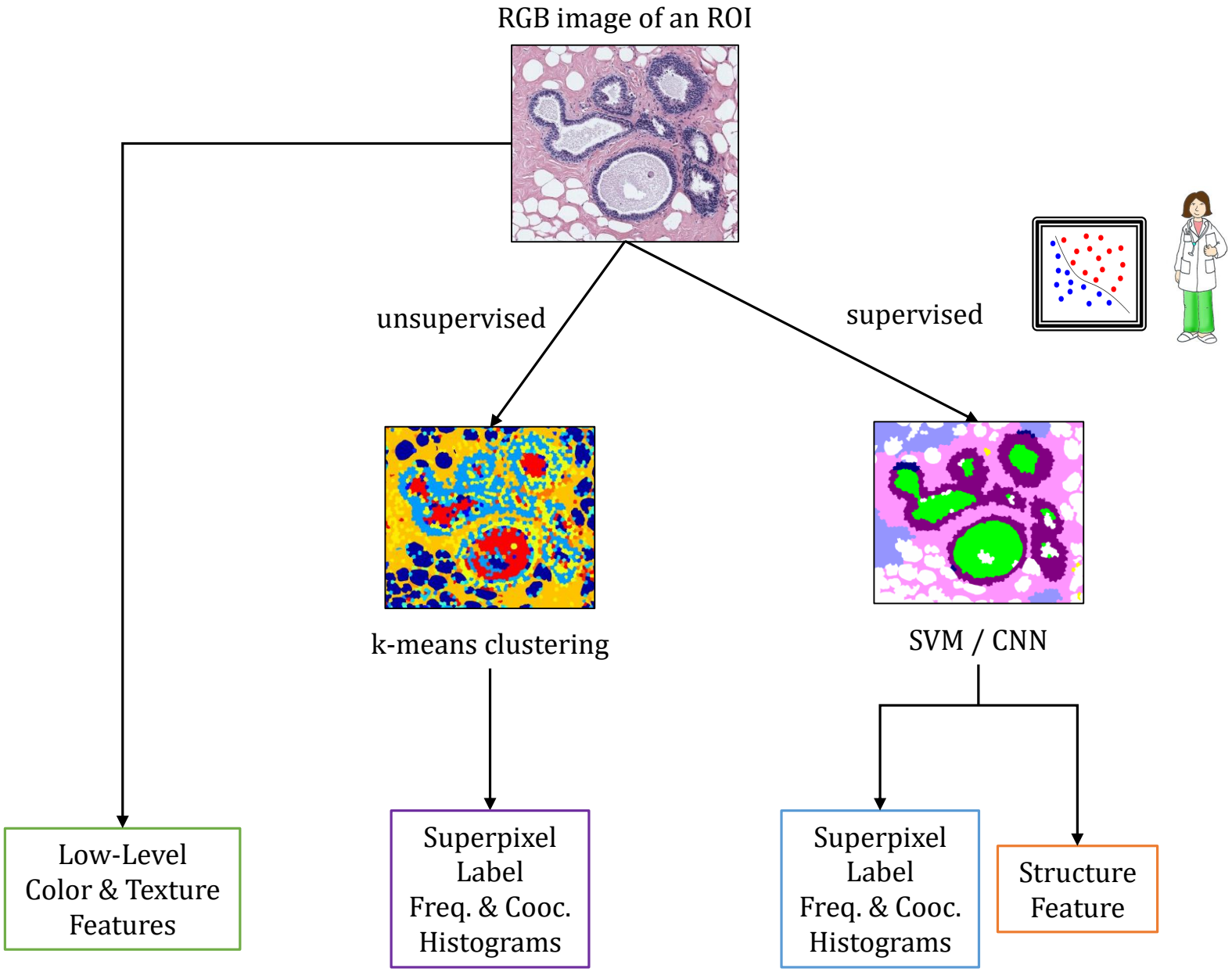
Structure Feature



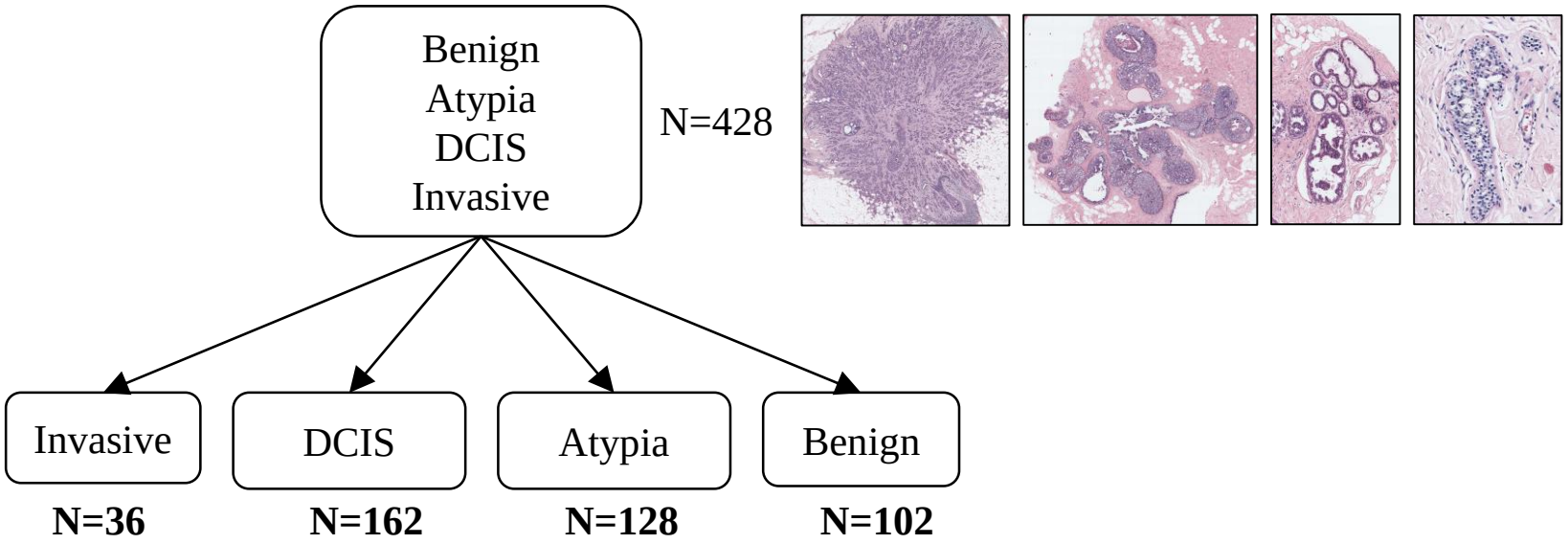






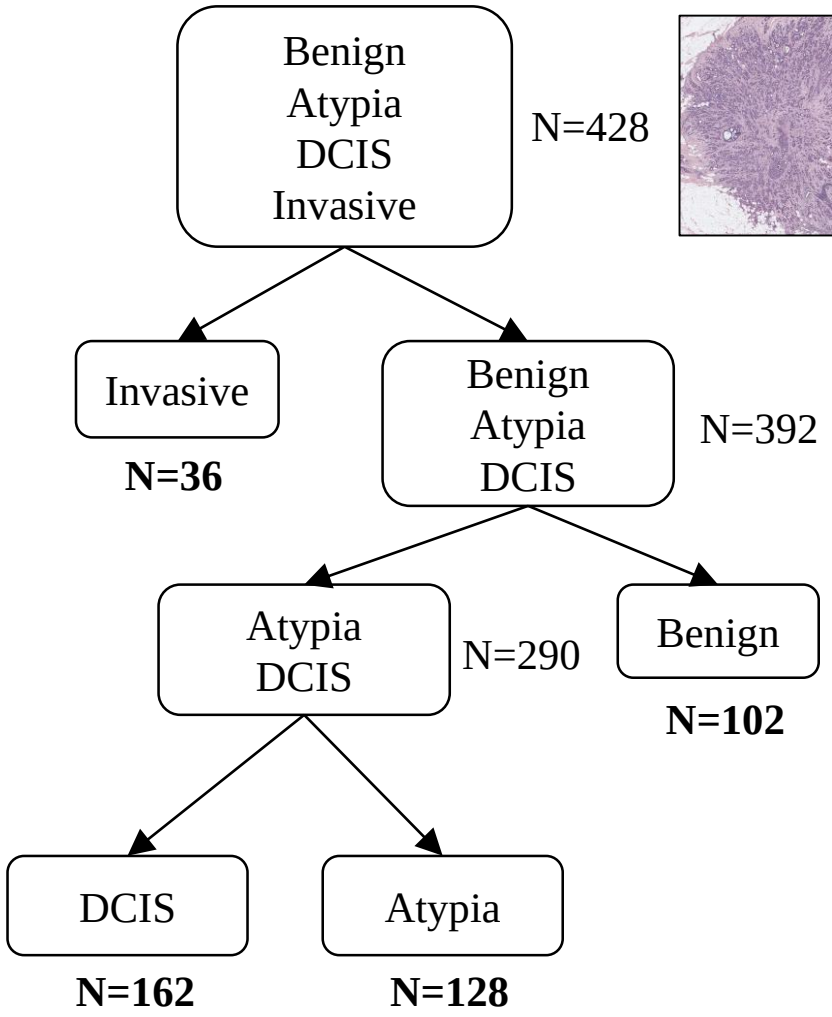

RGB image of an ROI
unsupervised
supervised
k-means clustering
SVM / CNN
Low-Level
Color & Texture
Features
Superpixel
Label
Freq. & Cooc.
Histograms
Superpixel
Label
Freq. & Cooc.
Histograms
Structure
Feature

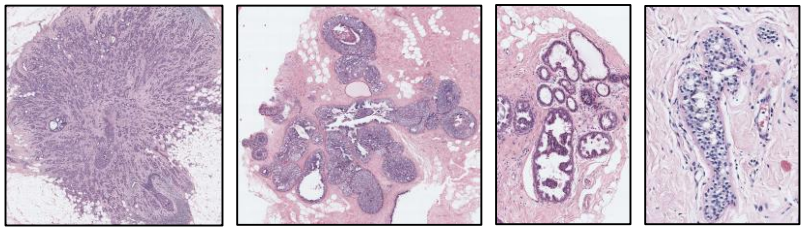
Diagnostic Classification



4-class classification

Diagnostic Classification



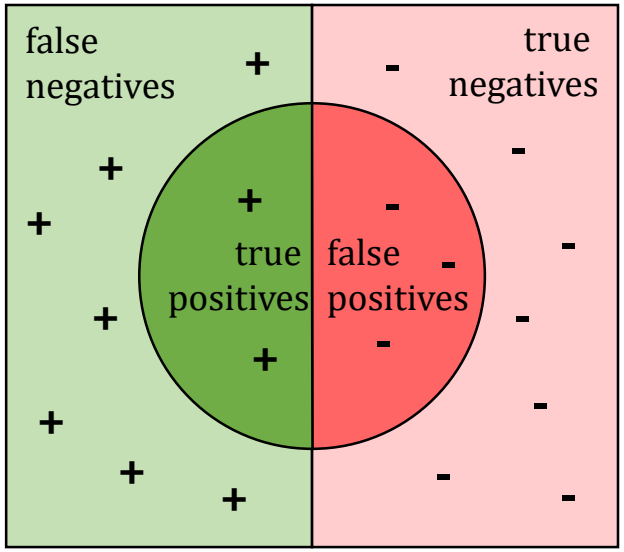


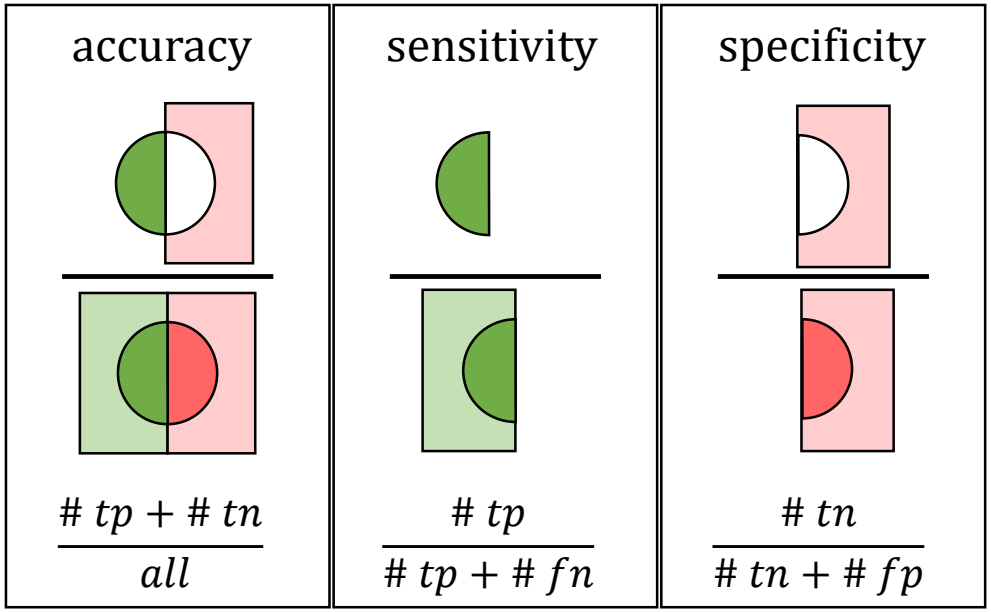
1-diagnosis-at-a-time

mimicking pathologists

Evaluation

Each ROI is a sample.

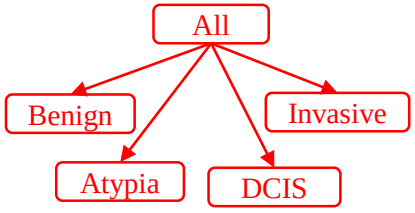


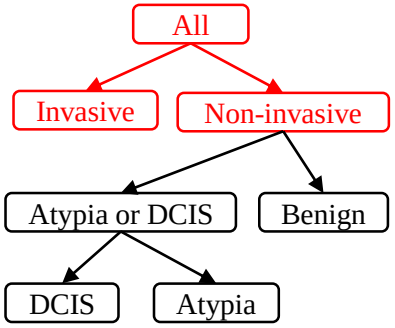


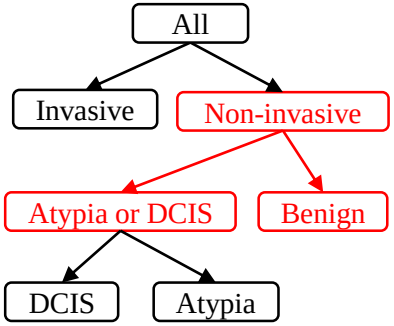
$\# tp = true\ positive$
$\# tn = true\ negative$
$\# fp = false\ positive$
$\# fn = false\ negative$

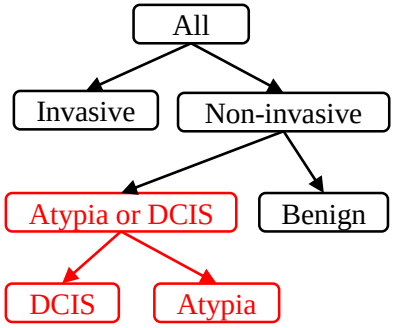
Experiments

- We subsampled the training data for a uniform distribution of all classes.
- We trained SVMs with different features
- We ran 10-fold cross-validation experiments for the 4 classification tasks:

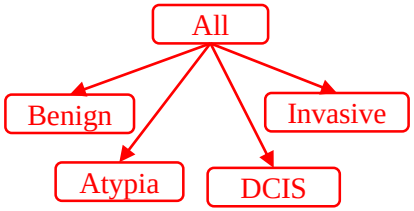


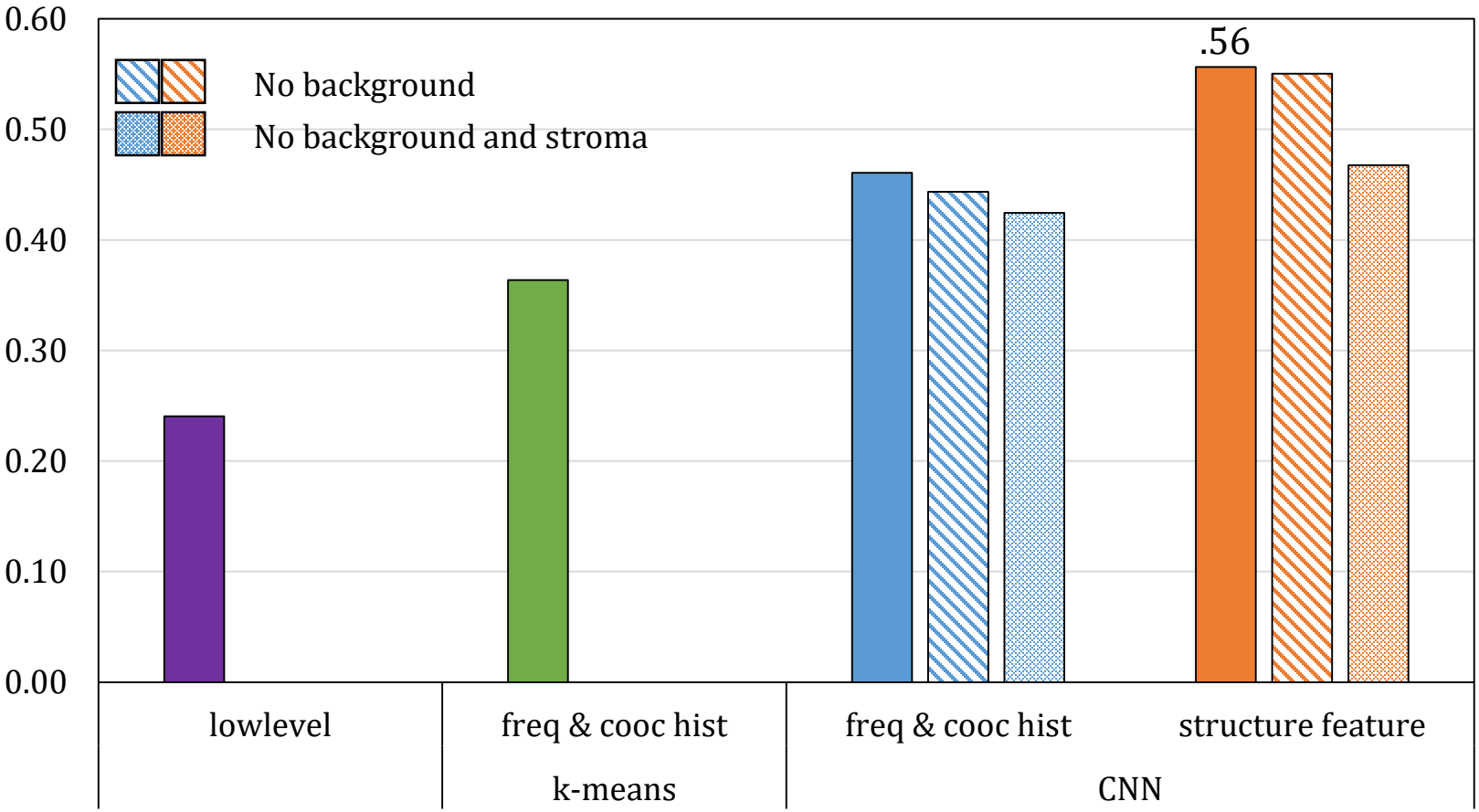




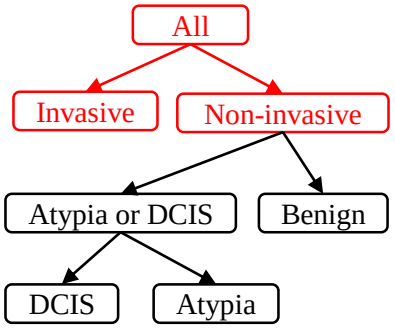


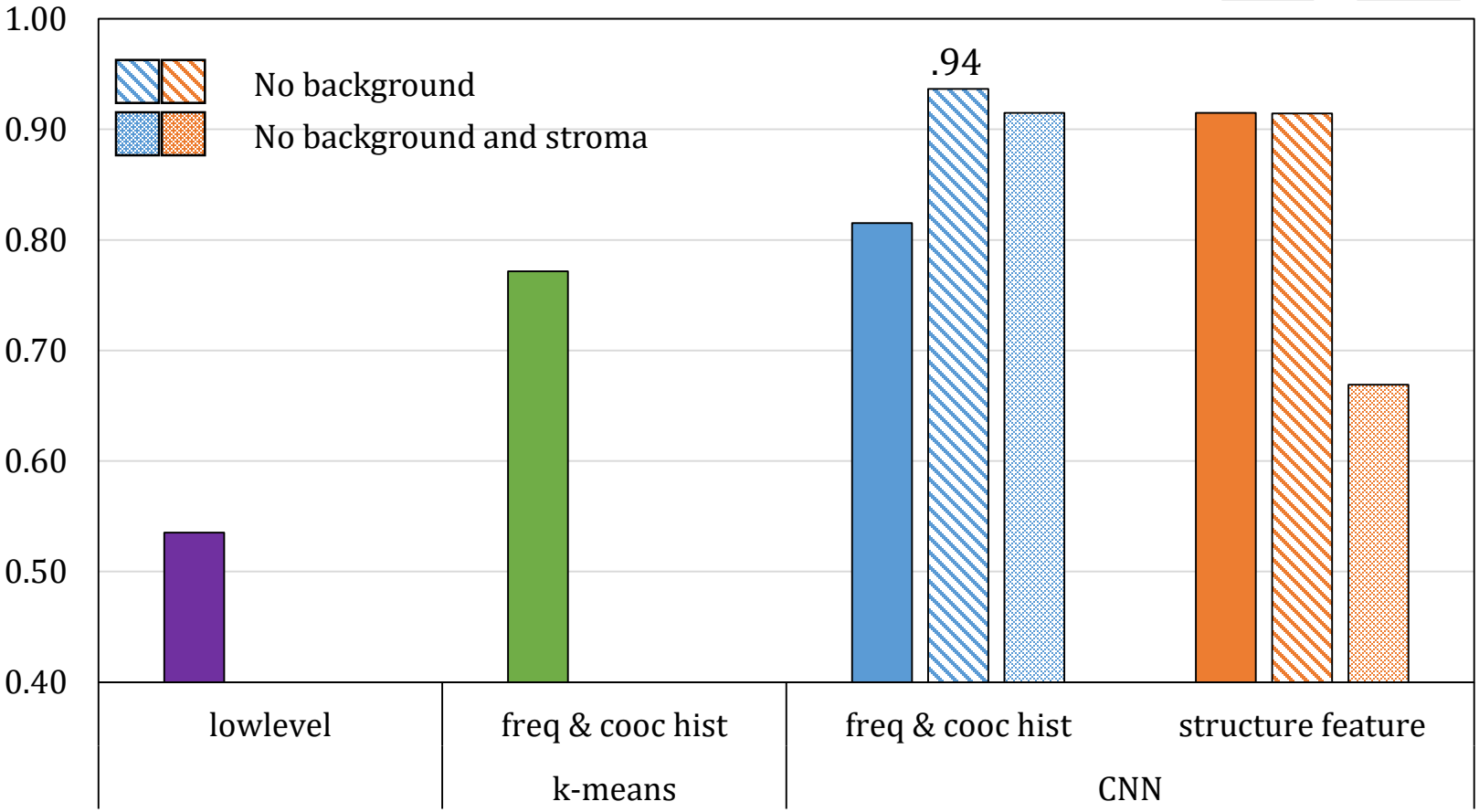
Accuracy for 4-class classification



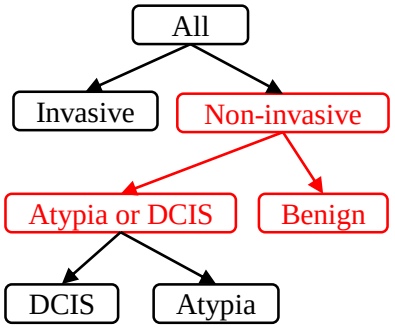


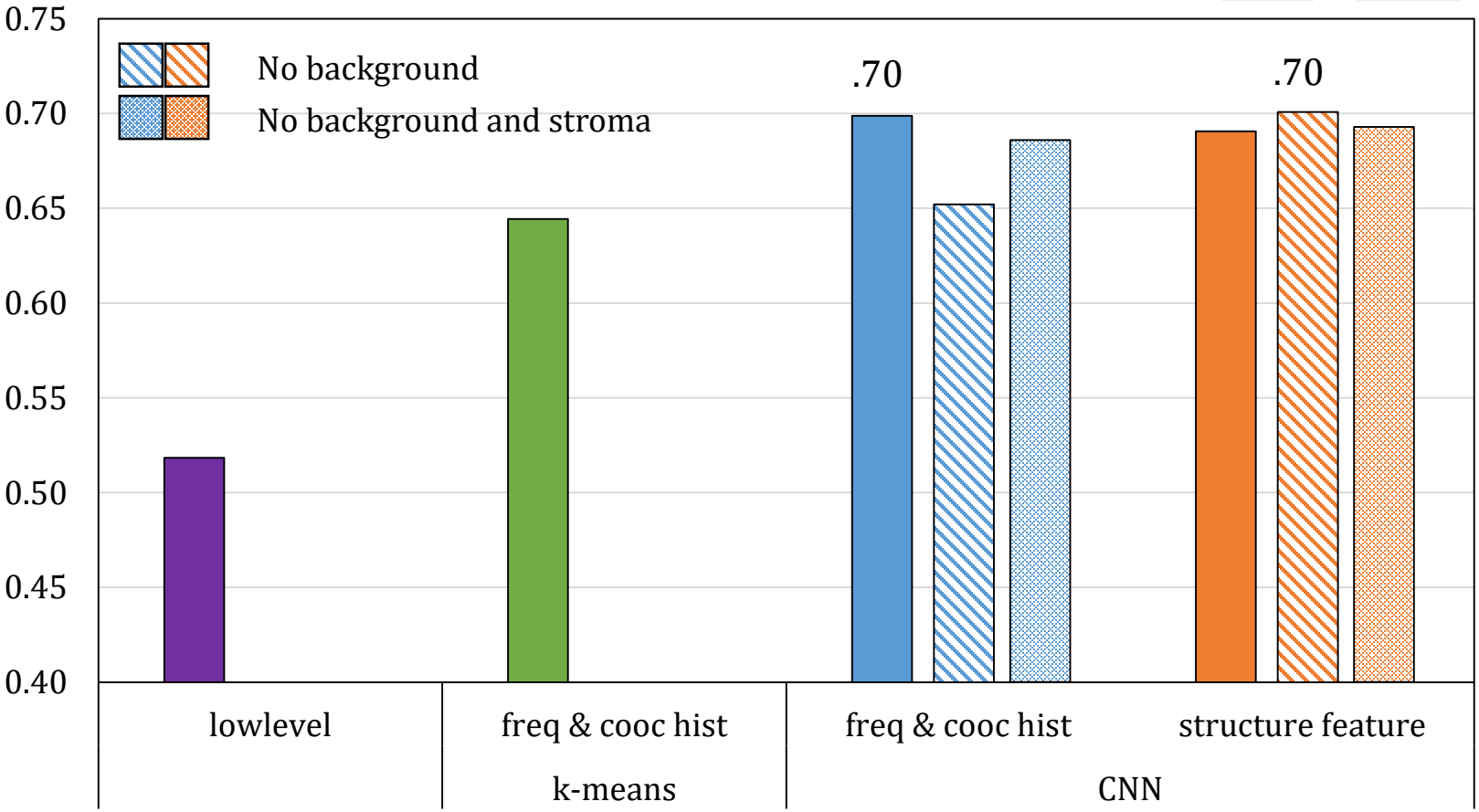
Accuracy for invasive vs. benign & atypia & DCIS



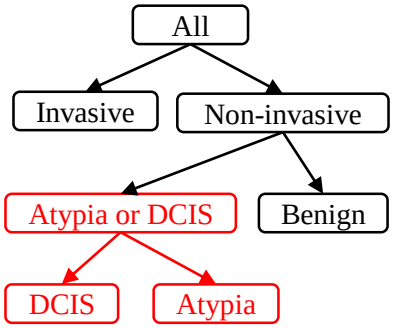


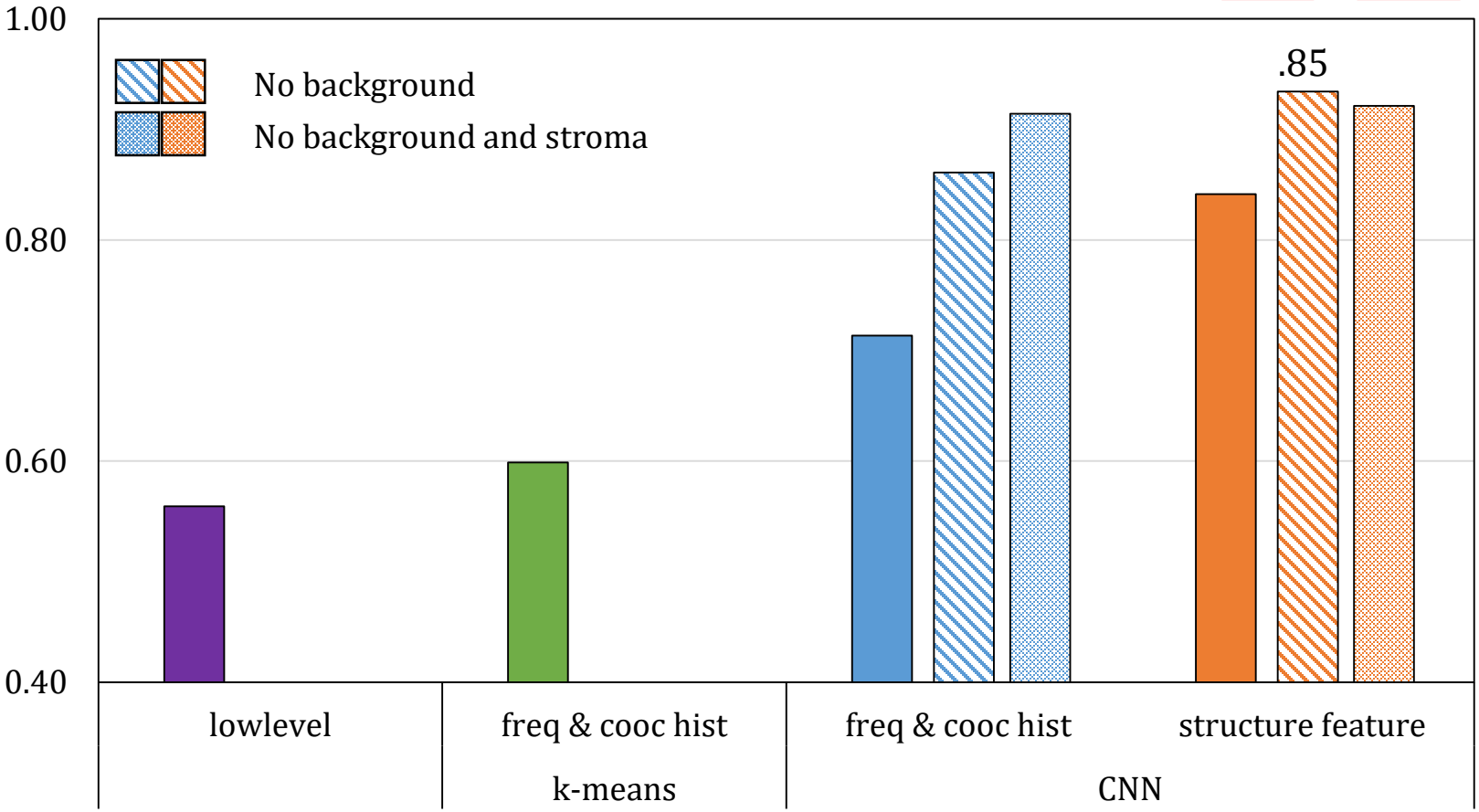
Accuracy for atypia & DCIS vs. benign





Accuracy for DCIS & atypia

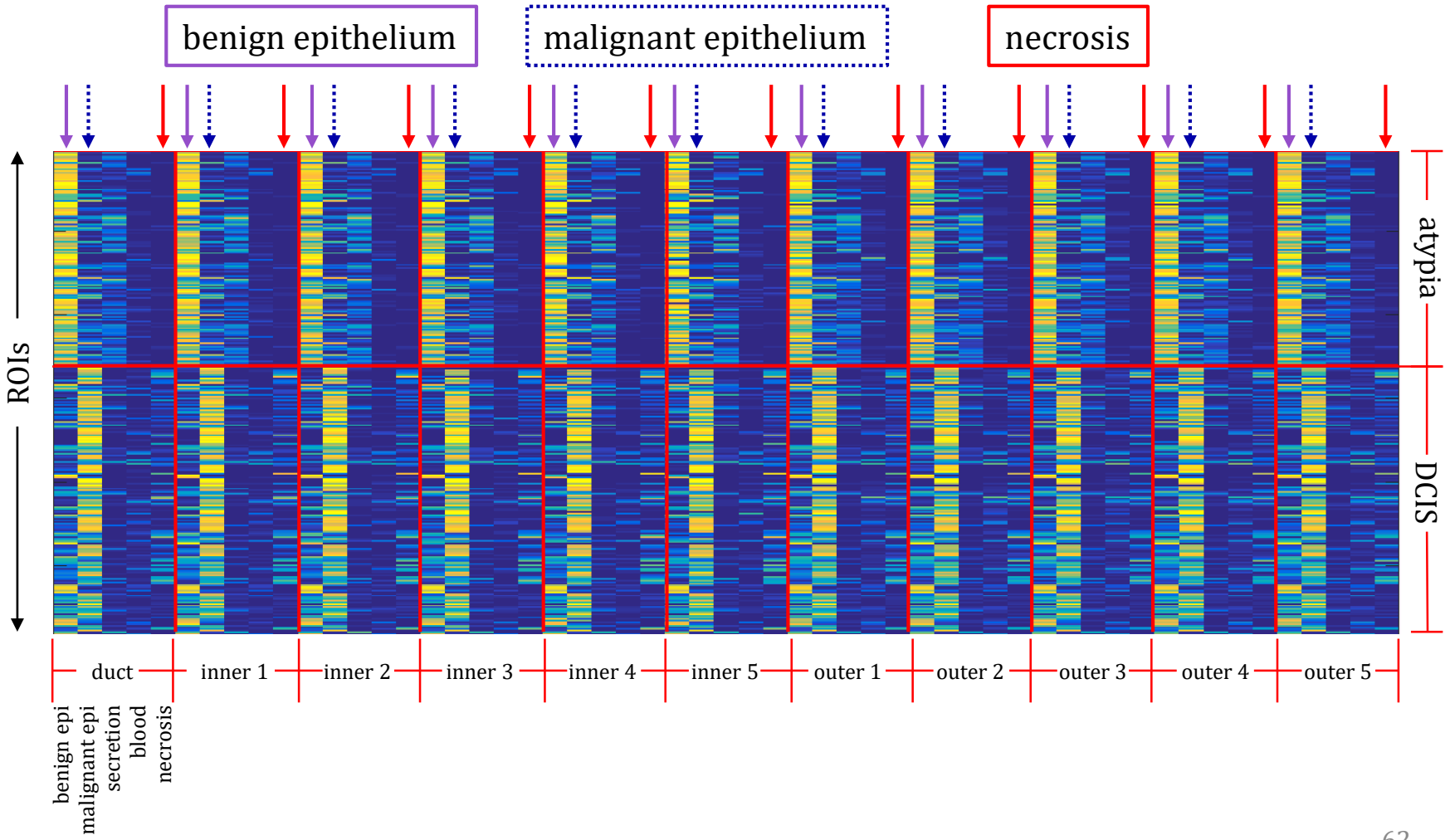




Results – Accuracies

		Average Accuracy			
		Invasive vs. Non-invasive	*Atypia & DCIS vs. Benign*	*DCIS vs. Atypia*	*4-class*
	sensitivity	.84	.72	.70	
	specificity	.99	.62	.82	
Participant Pathologists		.98	.81	.80	.70
	sensitivity	.70	.83		
	specificity	.95	.42		
Freq. and Cooc. Hist.		**.94**	**.70**	.83	.46
	sensitivity		.85	.89	
	specificity		.45	.80	
Structure Feature		.91	**.70**	**.85**	.56

Structure Feature



Summary

- Image features to describe the diagnostically important visual characteristics:
  - superpixel label frequency and co-occurrence histograms
  - structure feature
- Different features are informative for different diagnoses:
  - Superpixel label frequency and co-occurrence histograms for invasive cancer (0.94 accuracy).
  - Structure features for benign, atypia and DCIS (0.70 and 0.85 accuracies).

Outline

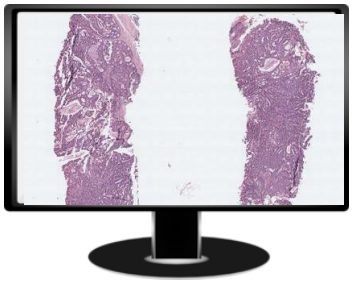
1. Introduction
2. ROI Localization in WSIs
3. Tissue Label Segmentation
4. Automated Diagnosis of ROIs
5. **Viewing Behavior Analysis**
6. Conclusions

digiPATH Dataset

pathologist
(n=87)

age
years in practice
cases per week
affiliation
lab size
education
pathology perception

assessment
(n=5,220)



participant ROI*
participant diagnosis
viewport log

confidence score
difficulty score
borderline diagnoses

Diagnostic Outcome
Accuracy & Efficiency

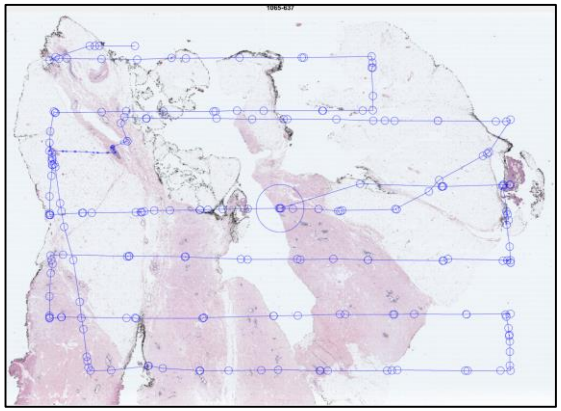
Diagnostic Search Patterns: Scanners and Drillers

- Histopathological diagnosis is a visual search.
- Prior work in radiology showed that physicians tend to adopt different search strategies.
- Our aim is:
  - To find efficient and accurate visual search strategies for diagnosis.
  - To understand the factors affecting visual search patterns.

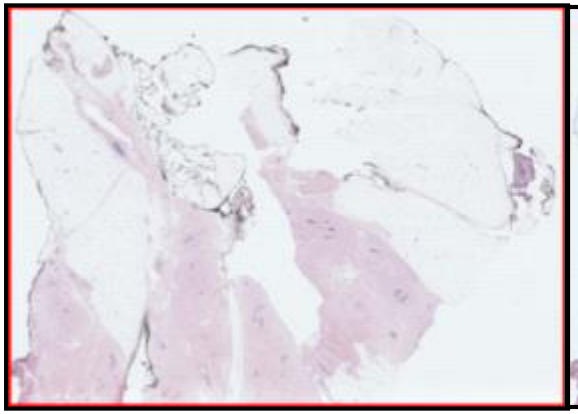
E. Mercan, *et al.*, Journal of Digital Imaging, 2017 (accepted).

Scanners and Drillers

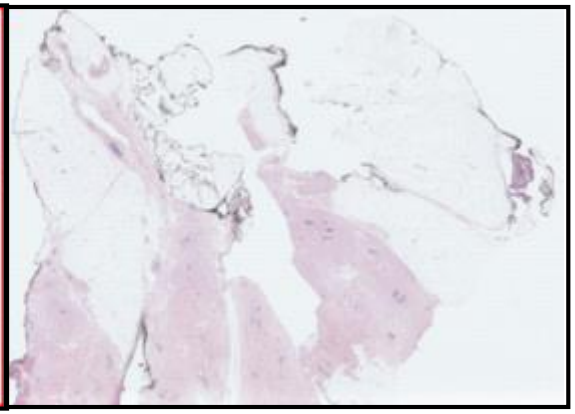
Scanner



Viewport Graph



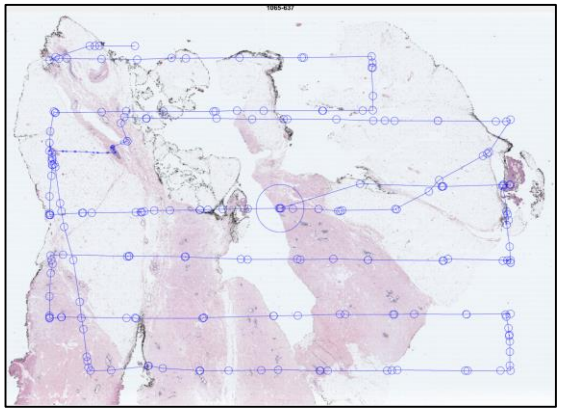
Viewport



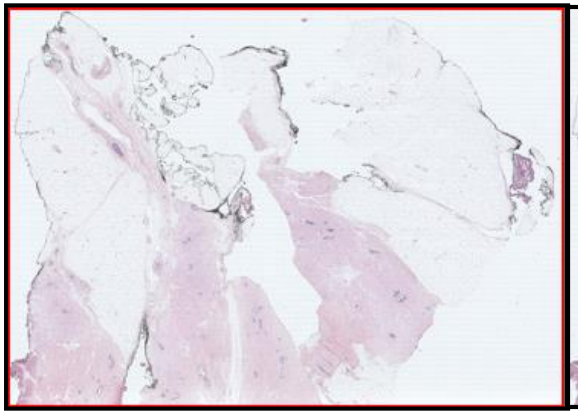
Pathologist's Screen

Scanners and Drillers

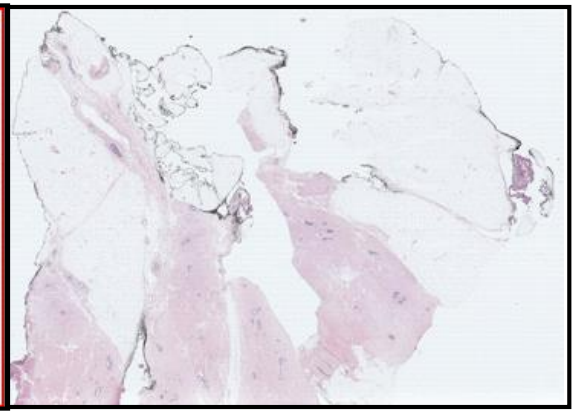
Scanner



Viewport Graph

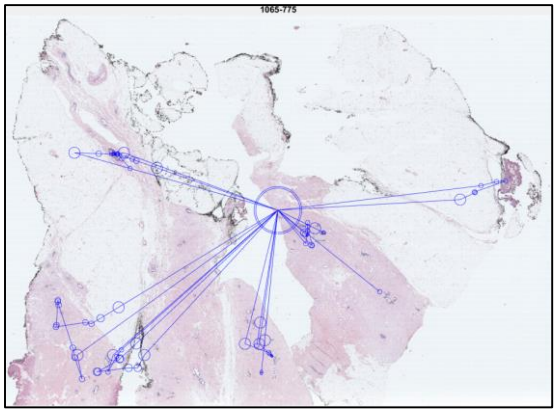


Viewport

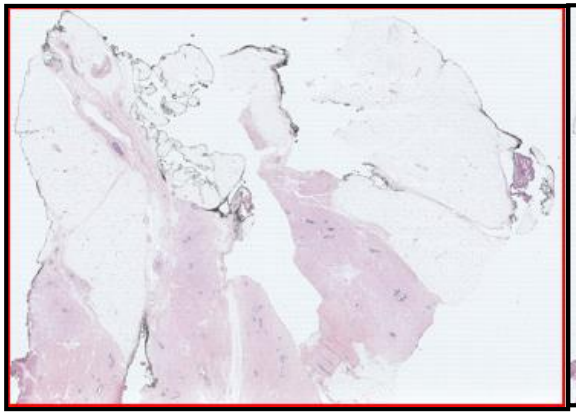


Pathologist's Screen

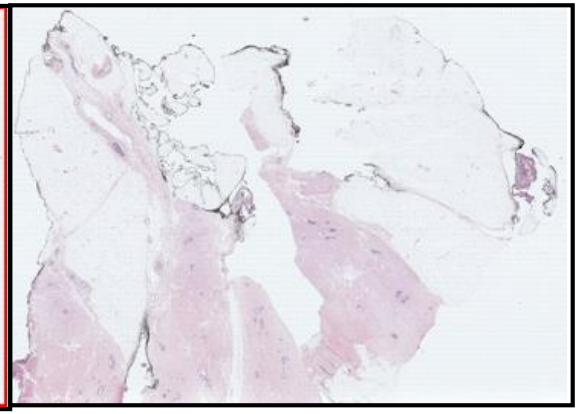
Driller



Viewport Graph

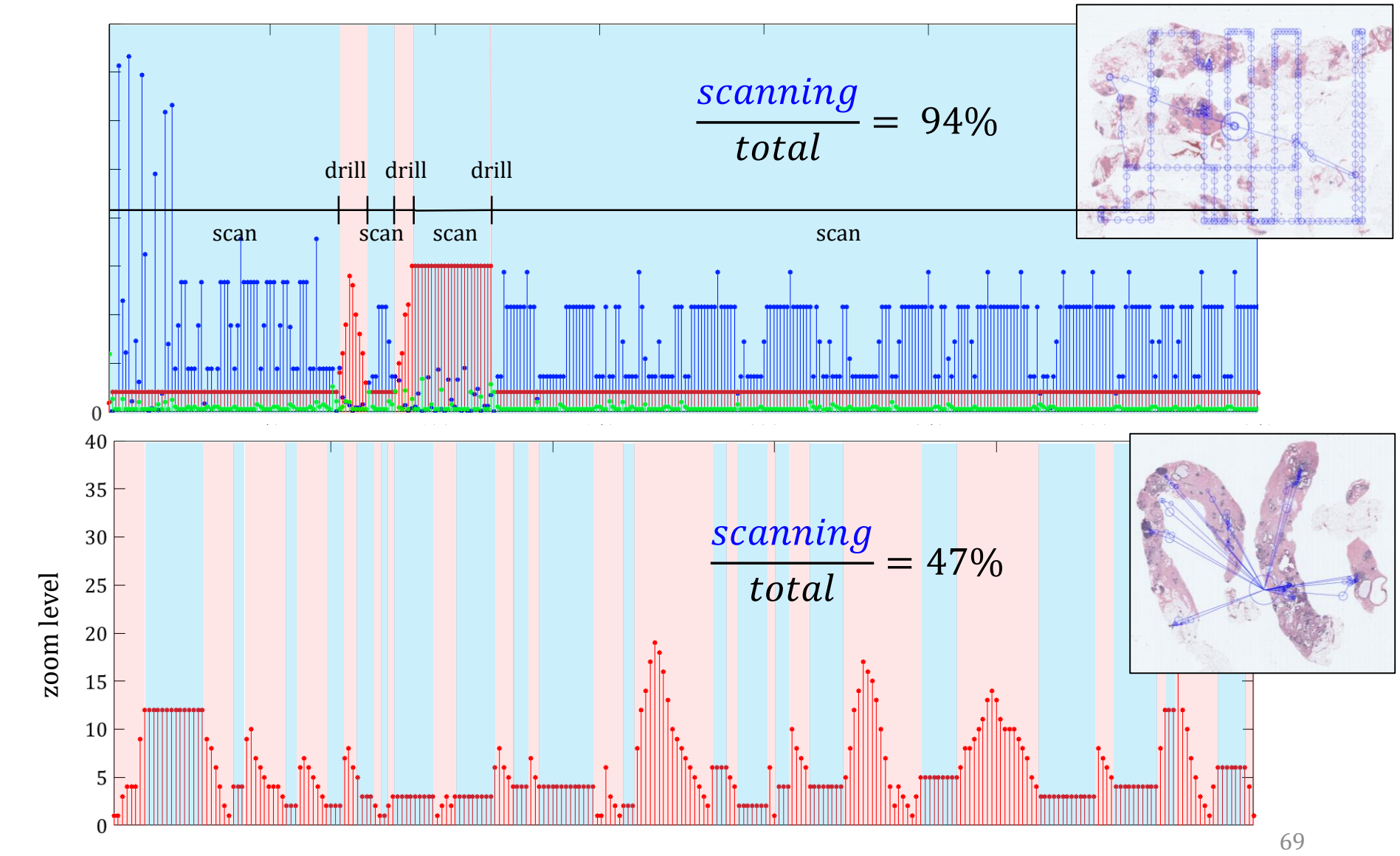


Viewport



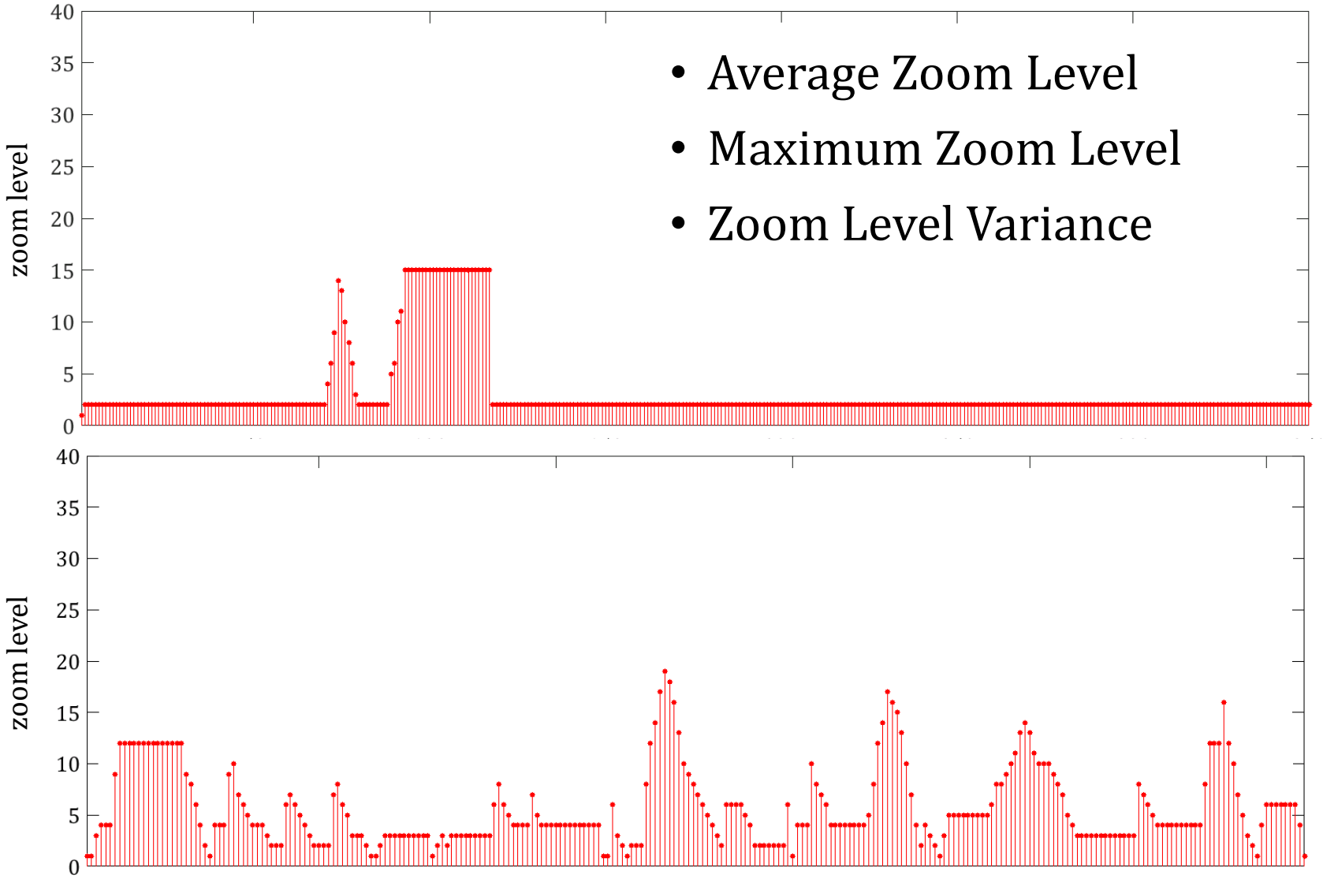
Pathologist's Screen

Scanning Percentage



Zoom Level Statistics

- Average Zoom Level
- Maximum Zoom Level
- Zoom Level Variance



Statistical Analysis

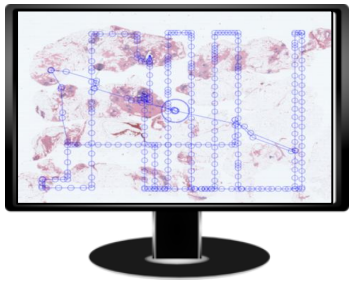
What factors affect interpretative strategy?

Pathologists
(n=87)

Pathologist Characteristics
Breast Histopathology Perceptions

Generalized Estimating Equations (GEE) Model

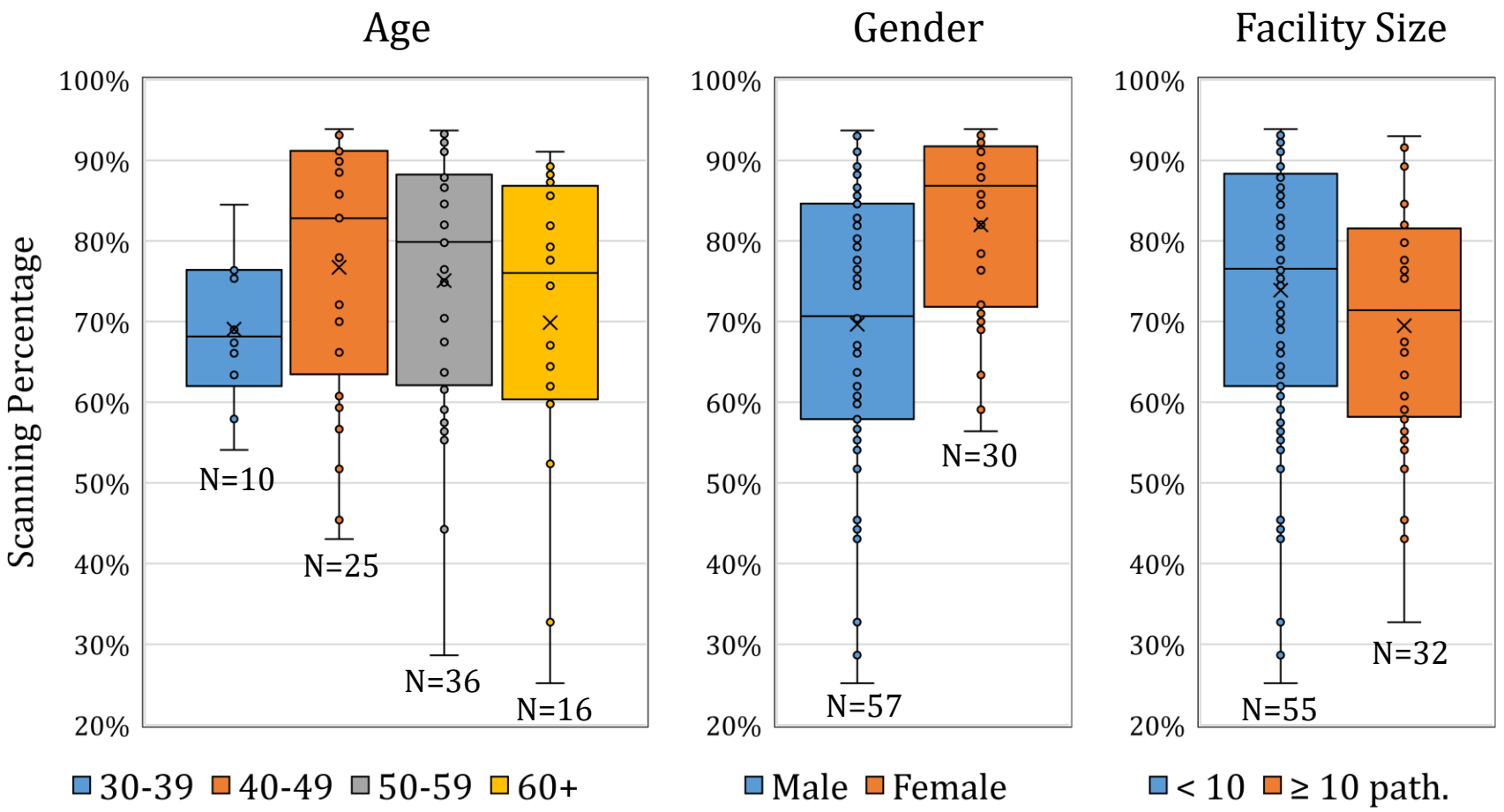
Assessments
(n=5,220)



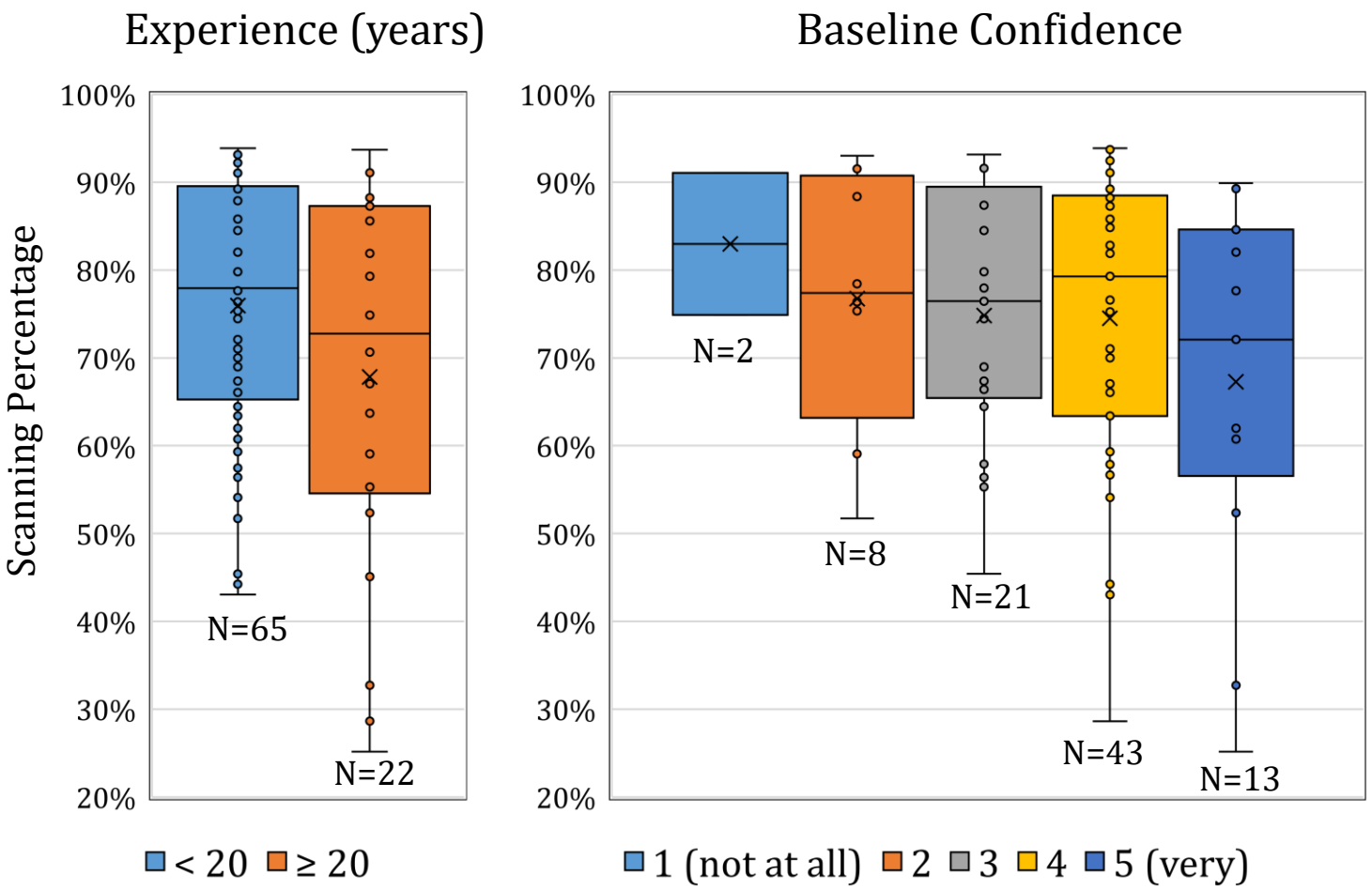
Scanning Percentage

joint work with Prof. Tad Brunye

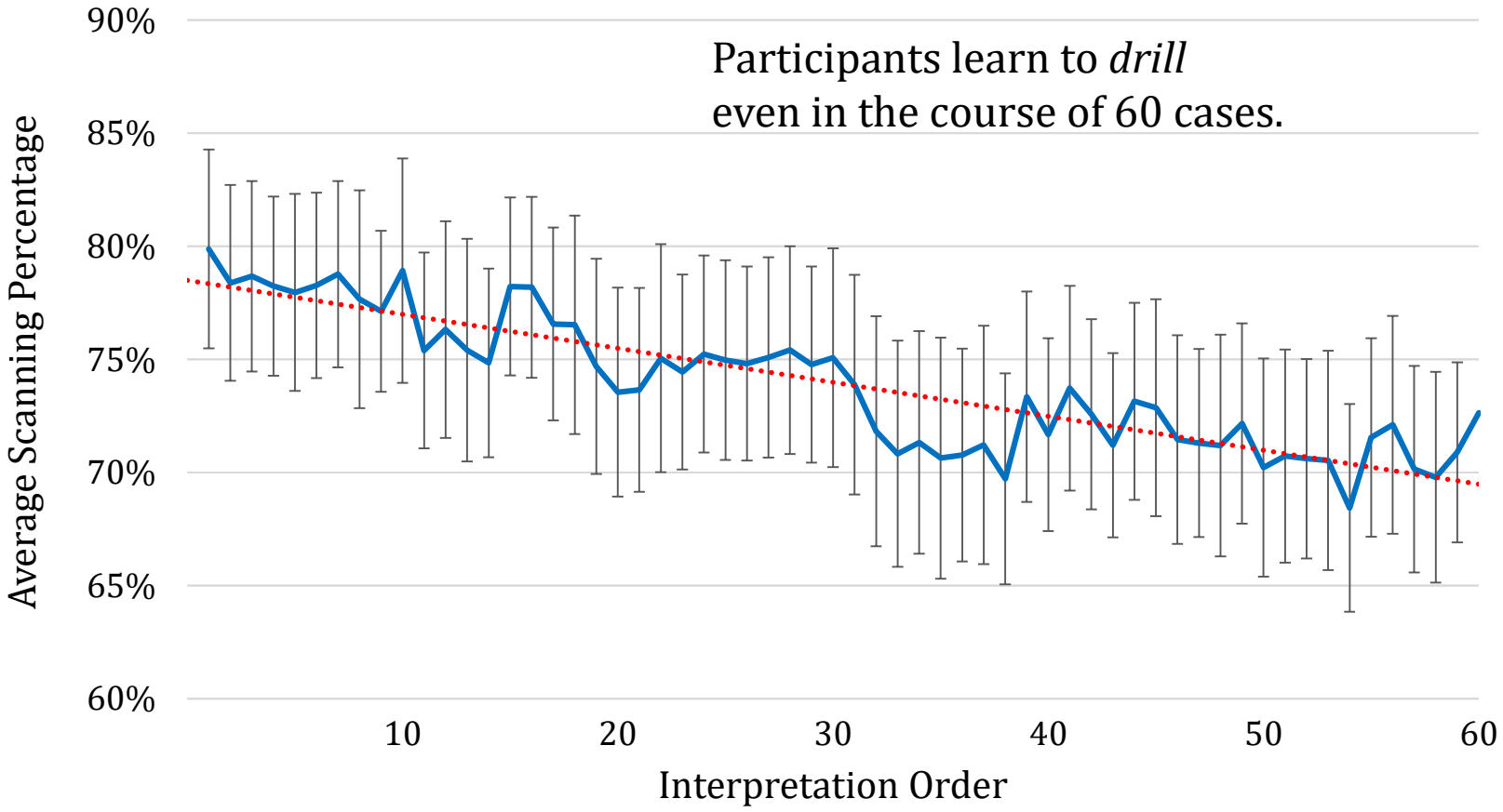
Pathologist Demographics



Pathologist Demographics



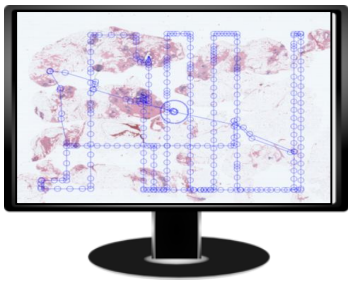
Interpretation Order



Statistical Analysis

How does interpretative strategy influence diagnostic outcome?

Assessments
(n=5,220)



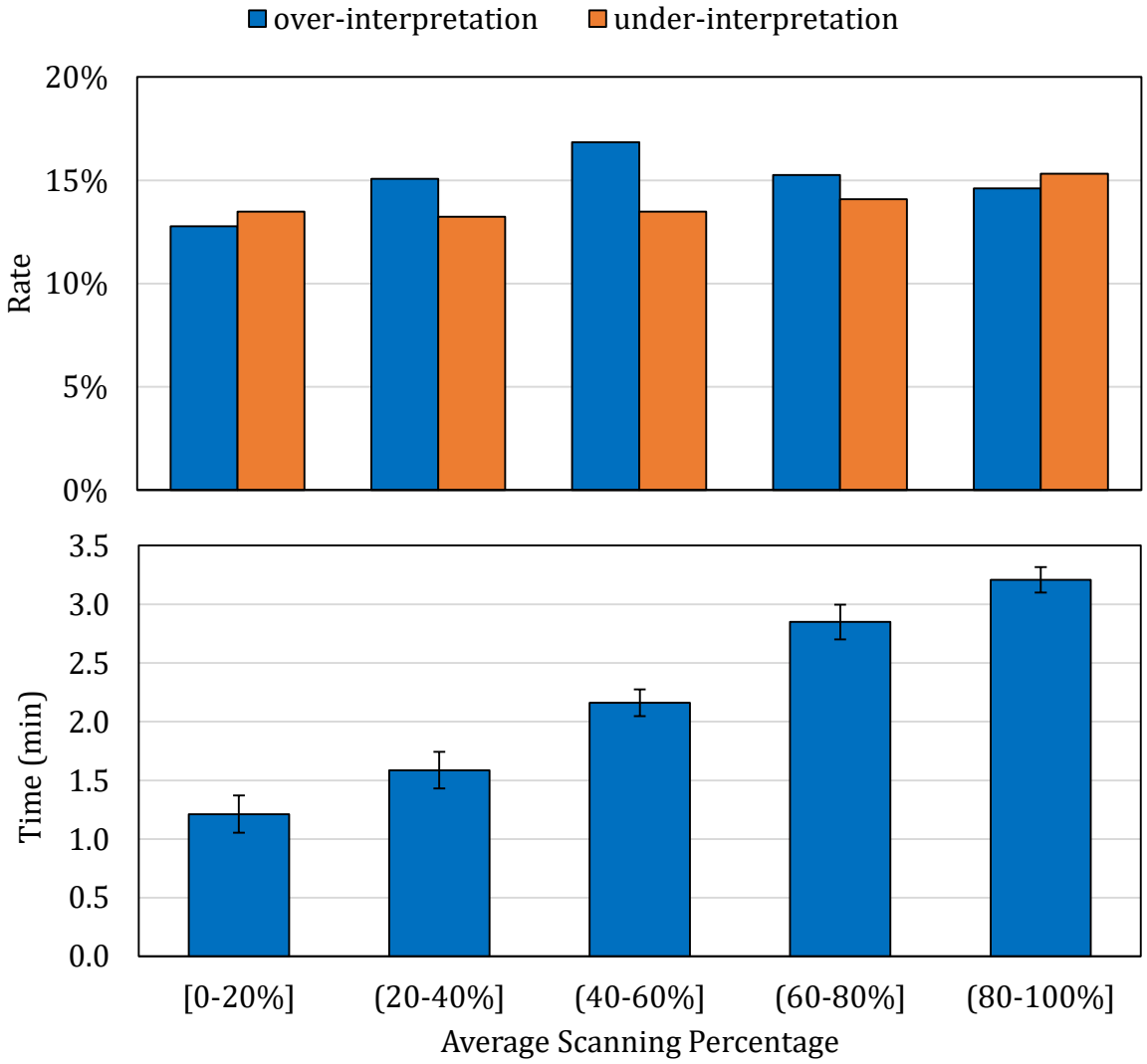
Average Zoom Level
Maximum Zoom Level
Zoom Level Variance
Scanning Percentage

ANOVA

Diagnostic
Accuracy & Efficiency

Under-interpretation
Over-interpretation
Interpretation Time

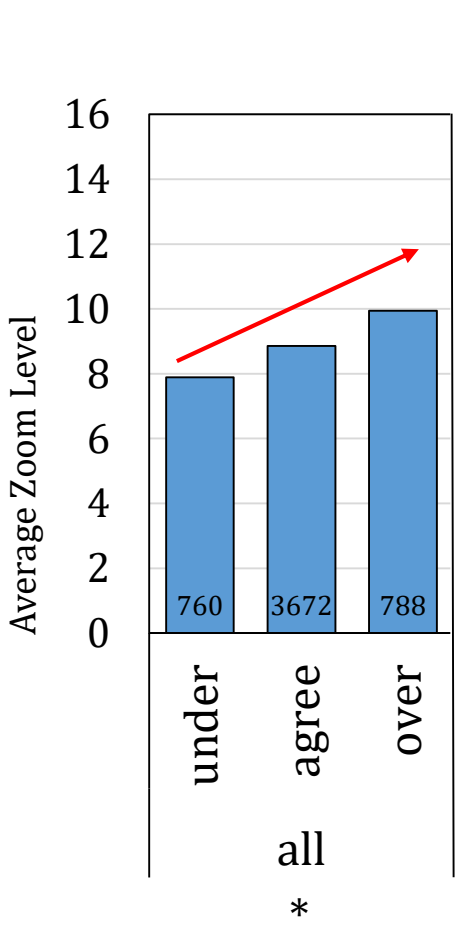
Diagnostic Accuracy and Efficiency



Scanning does not affect diagnostic accuracy

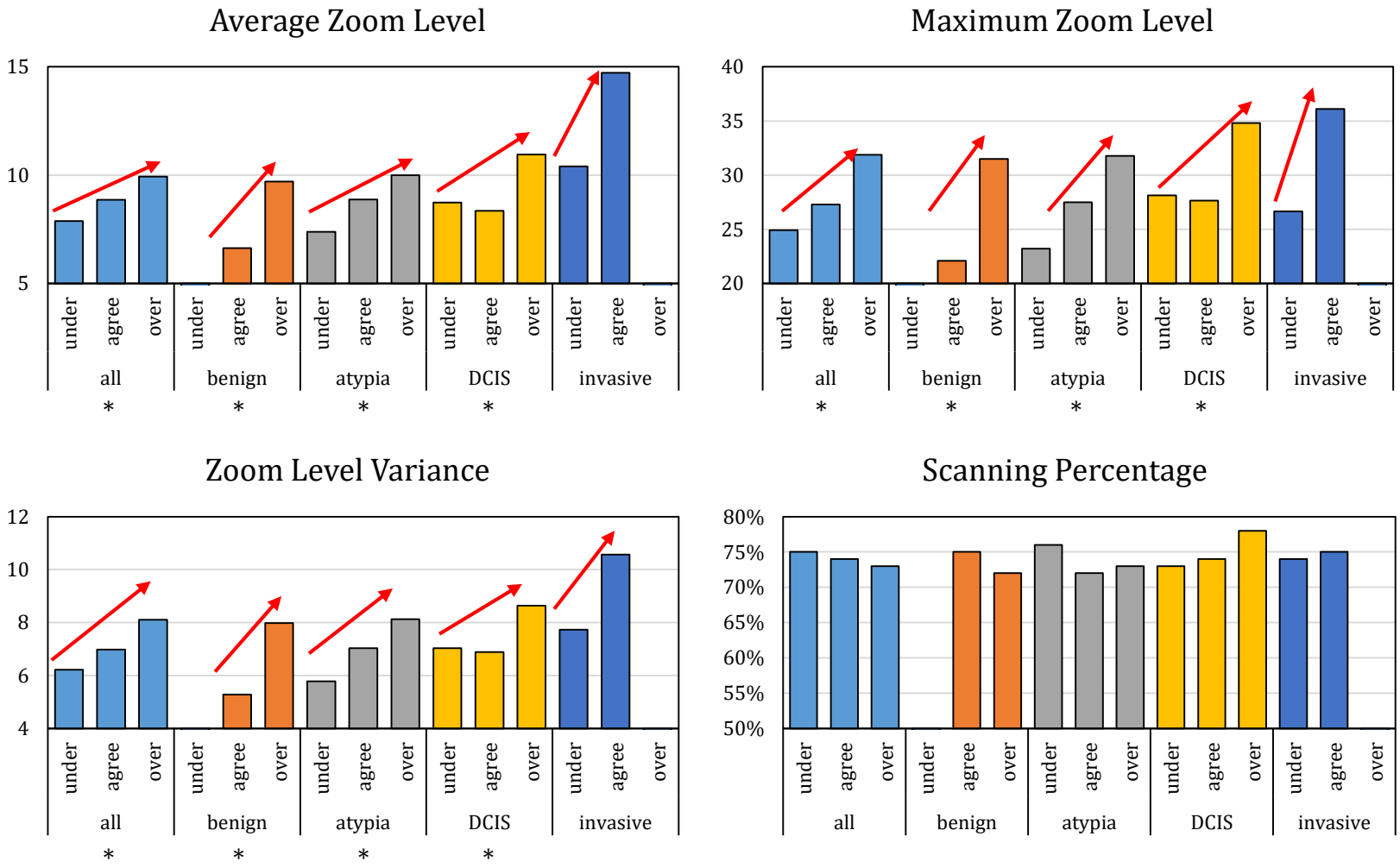
but it costs time.

Diagnostic Concordance



* p<0.02 (ANOVA)

Diagnostic Concordance



* p<0.02 (ANOVA)

Summary

- Pathologists exhibit two viewing patterns:
  - scanning and drilling

- We developed four objective measures to quantify the viewing behavior:
  - maximum zoom level
  - average zoom level
  - zoom level variance
  - scanning percentage

Summary

- Scanning is correlated with
  - sex *(females > males)*, age *(+)* , facility size *(-)*, experience *(-)* and confidence *(-)*.
- Pathologists learn to drill in the course of 60 cases.
- Scanning is not predictive of diagnostic accuracy but it is inefficient in terms of time.
- Increasing average and maximum zoom levels, and variance correlates with over-interpretation.

Outline

1. Introduction
2. ROI Localization in WSIs
3. Tissue Label Segmentation
4. Automated Diagnosis of ROIs
5. Viewing Behavior Analysis
6. **Conclusions**

Our Aim

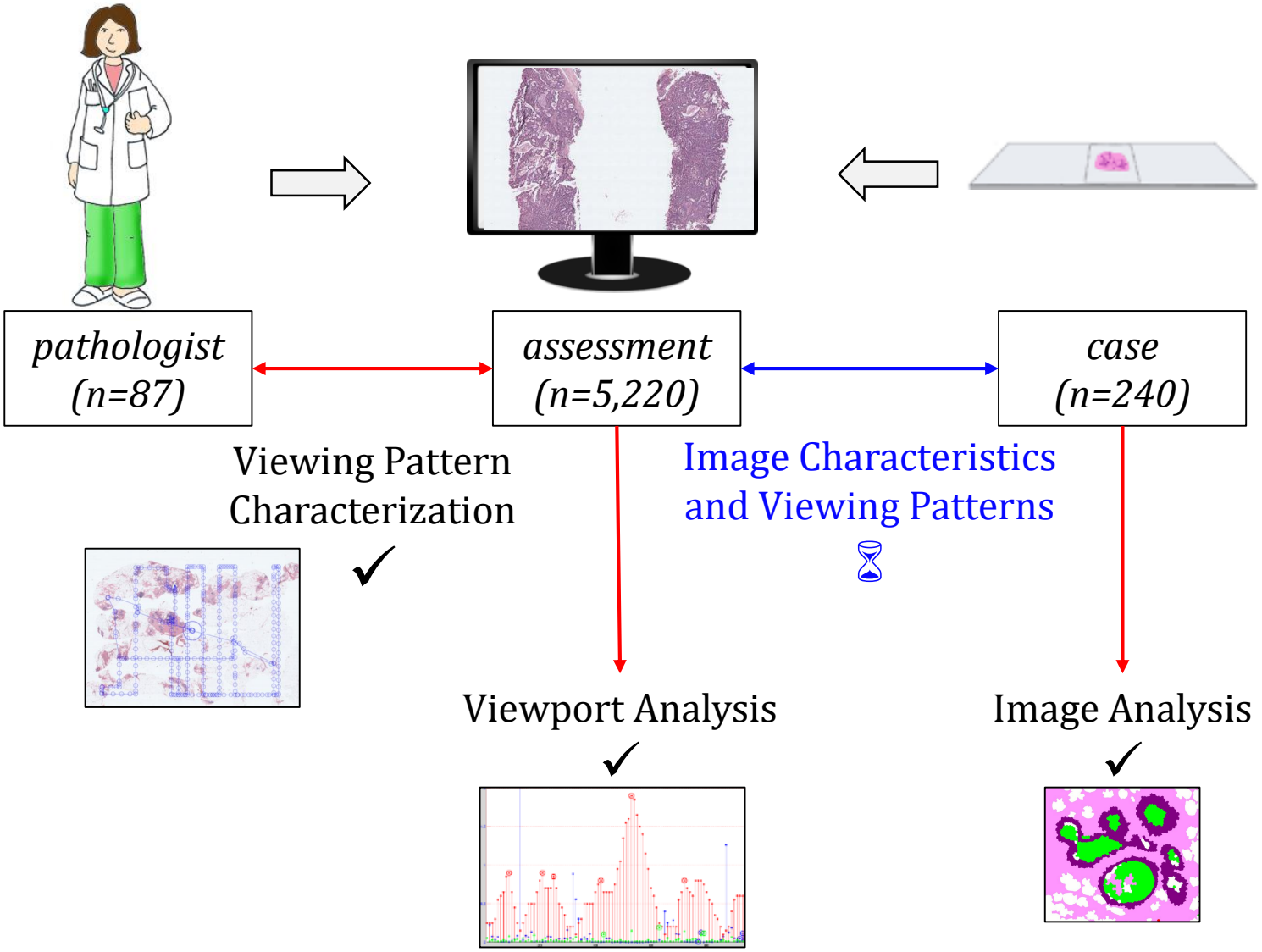
We propose novel image features and viewing behavior analysis methods

- to investigate the causes of diagnostic errors and
- to discover accurate and efficient interpretation strategies for pathologists.

Contributions

- Novel analysis methods for the viewing behavior:
  - Analysis methods for the tracking logs to detect diagnostically relevant regions
  - Objective measures to quantify the interpretation patterns
- Novel image features for the pre-invasive and invasive lesions of the breast:
  - An 8-label supervised segmentation
  - Structure feature to describe the changes associated with pre-invasive lesions

digiPATH


pathologist
(n=87)
assessment
(n=5,220)
case
(n=240)
Viewing Pattern
Characterization
Image Characteristics
and Viewing Patterns
Viewport Analysis
Image Analysis

Acknowledgments


Linda Shapiro


Su-In Lee


Joann Elmore


Don Weaver


Tad Brunye


Selim Aksoy

Bilge Soran
Sara Rolfe
Shu Liang
Deepali Aneja
Yao Lu
Yuguang Lee
Sachin Mehta
Shima Nofallah
Sean Yang
Bindita Chaudhuri
Nathan Sjoquist
Shenqi Tang

Irma Lam
Mabel Raza
Nicola Dell
Raven Travillian
Jia Wu
Shulin Yang
Alfred Gui

Safiye Celik
Scott Lundberg
Nao Hiranuma
Javad Hosseini
Gabriel Schubiner
Hugh Chen

Mara Rendi
Jamen Bartlett
Dilip Nagarkar
Tom Morgan
Paul Frederick
Andrea Radick
Ross Lambert
Natalia Oster
Hannah Shucard

Parmita Mehta
Mike Chung
Robert Gens
Kivanc Muslu
Paul Pham
Jinna Lei

NATIONAL CANCER INSTITUTE
NIH National Institutes of Health

R01 CA172343, R01 CA140560, and KO5 CA104699

FRED HUTCH CURES START HERE®
PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING
HARBORVIEW MEDICAL CENTER UW Medicine