Pathologists' viewing behaviors contribute to diagnostic accuracy of melanocytic skin lesions

Fatemeh Ghezloo

February 28, 2022

PAUL G. ALLEN SCHOOL

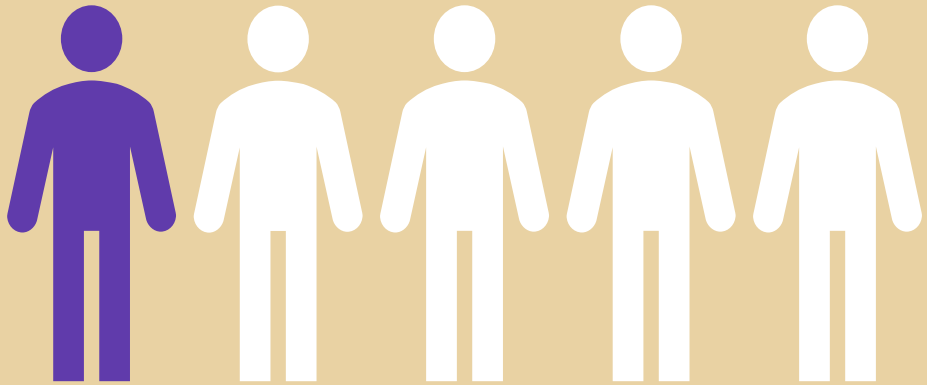
Outline

- Introduction
- Material and Methods
- Results
- Application
- Conclusions

Outline

- **Introduction**
- Material and Methods
- Results
- Application
- Conclusions

Skin cancer is the most commonly diagnosed cancer in the United States!

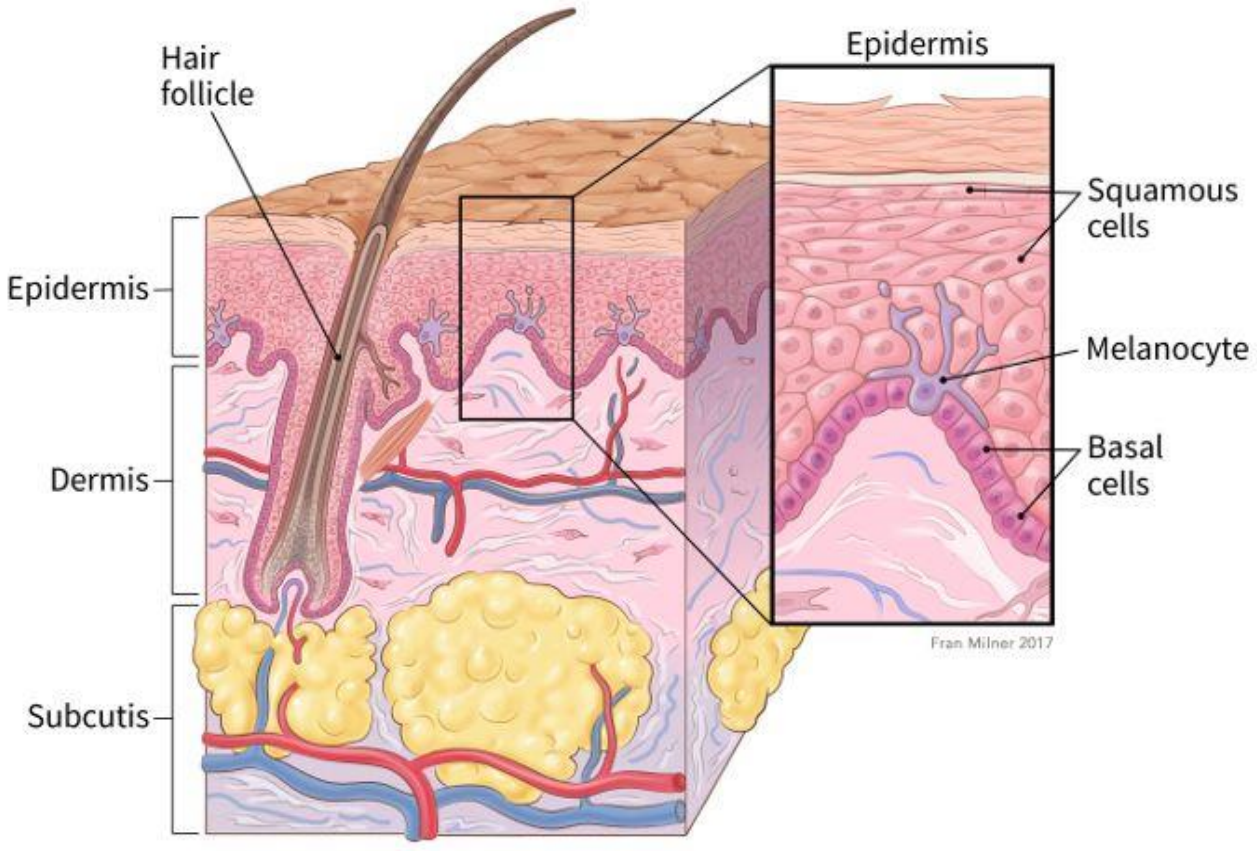


1 in 5 Americans will develop skin cancer by the age of 70.[1]

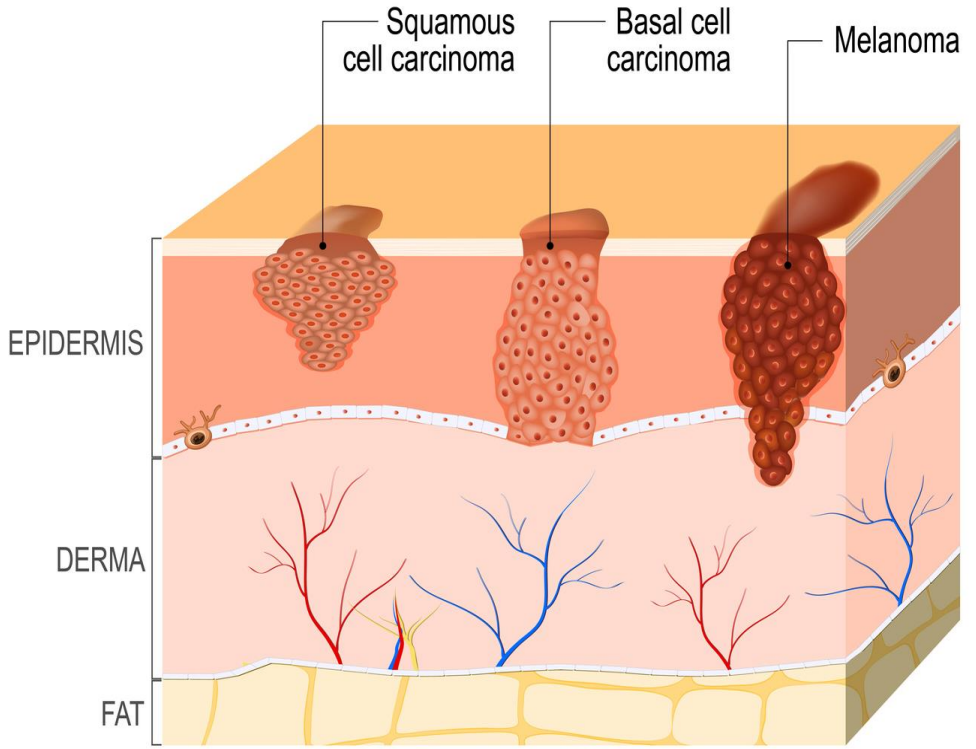
More than 2 people die of skin cancer in the U.S. every hour.[1]

1. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

Where do skin cancers start?



Skin structure[1]

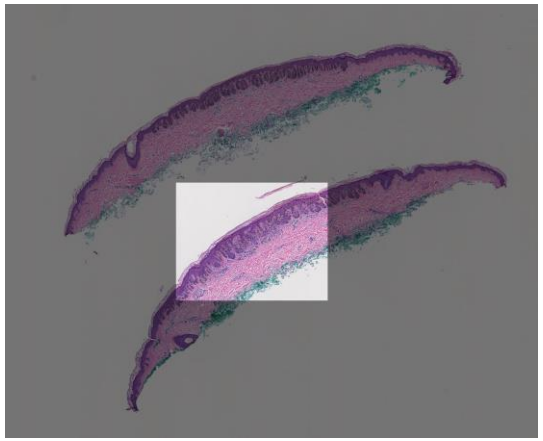


Skin cancer types[2]

1. https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html
2. https://smart.servier.com

Melanoma diagnosis:
Traditional Microscopy to Whole Slide Imaging

Traditional Microscopy



Region of Interest (ROI)

Whole Slide Imaging

FDA U.S. FOOD & DRUG ADMINISTRATION

Boyce, B., *An update on the validation of whole slide imaging systems following FDA approval of a system for a routine pathology diagnostic service in the United States.* Biotechnic & Histochemistry,2017.

Pathologists' Viewing behavior: Why is it important?

Digital pathology and Whole Slide Imaging

- More cases are being interpreted in digital format
- Computer-based technologies are being developed for diagnosis

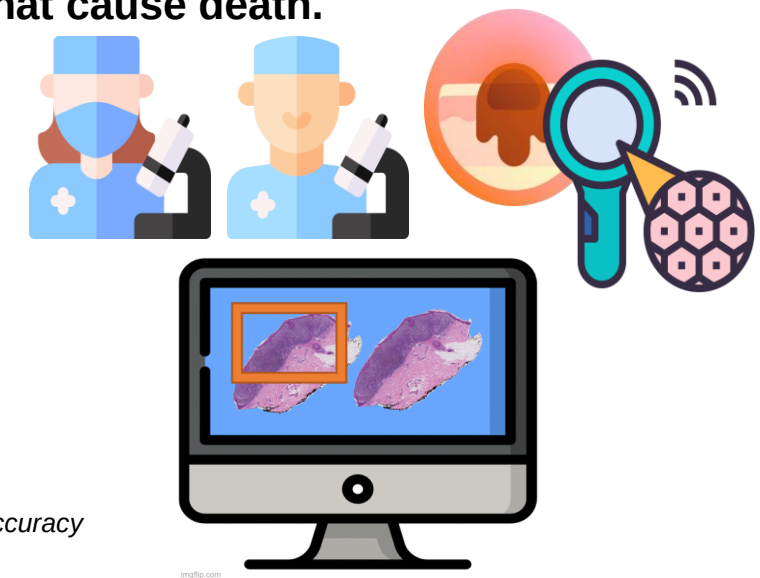
Pathological examination

However,

There is discordance among pathologists' diagnoses even when they observe similar features on a biopsy sample slide.
Errors made in cancer diagnosis are one of the main reasons that cause death.

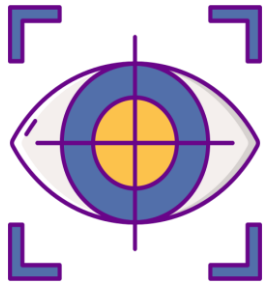
Diagnostic accuracy varies due to the interaction among

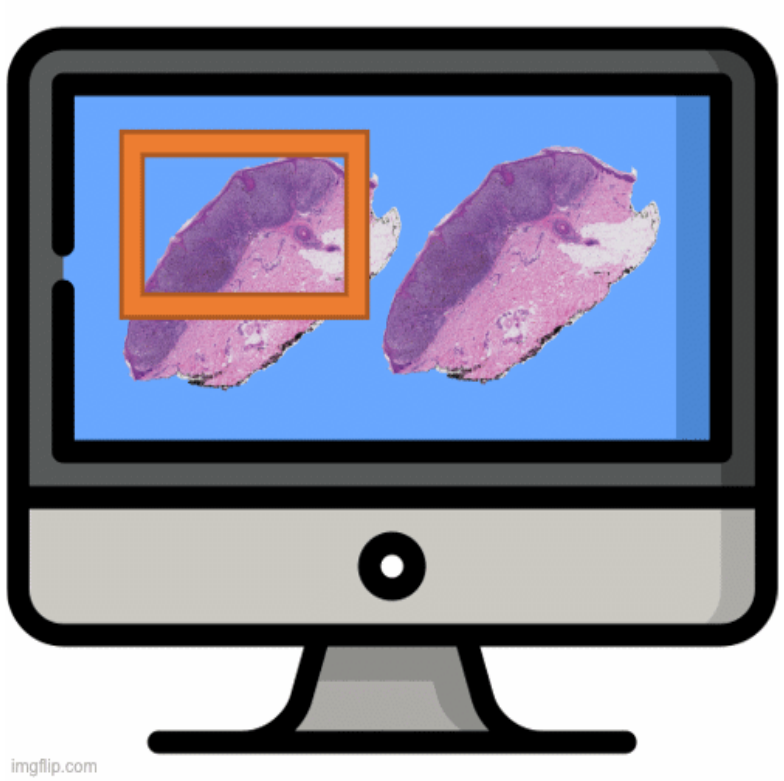
- Case characteristics
- Pathologist characteristics
- Visual search process[1]



1. Brunyé, T.T., et al., *Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images.* Journal of biomedical informatics, 2017. 66: p. 171-179.

Pathologists' Viewing behavior: Study workflow





- Eye fixation
- Saccade
- Timestamp
- Zoom level
- Mouse click
- Screen viewport

Tracking Devices

Interpretation session

Data Analysis

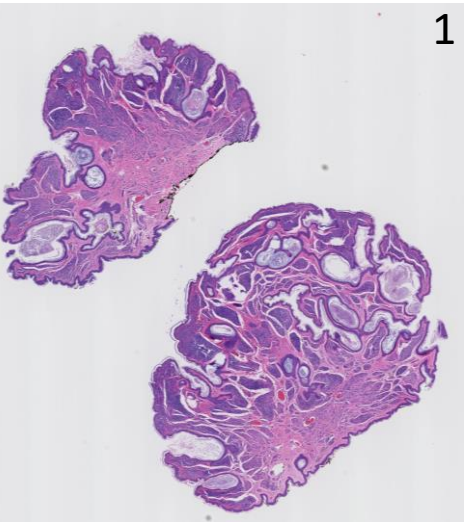
Study Goals

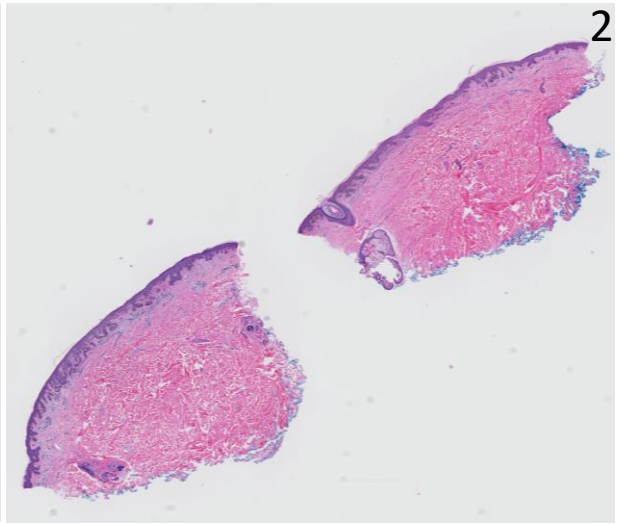
- We outline the types of data points that can be gathered to describe pathologists' viewing behavior using viewport data
- How do specific viewing behaviors contribute to diagnostic accuracy?
- How are pathologists' characteristics associated with specific viewing patterns?

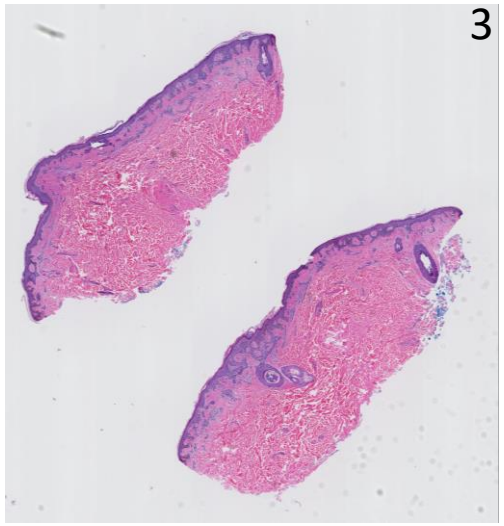
Outline

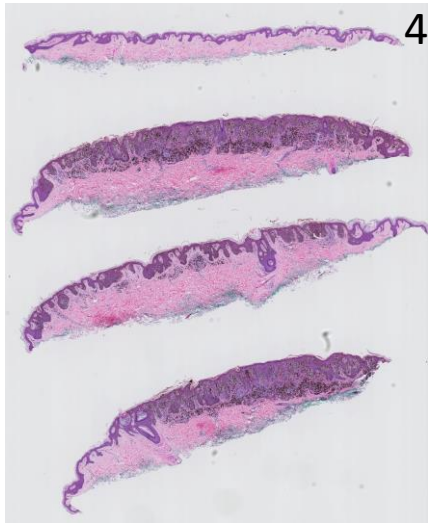
- Introduction
- **Material and Methods**
- Results
- Application
- Conclusions

M-PATH data



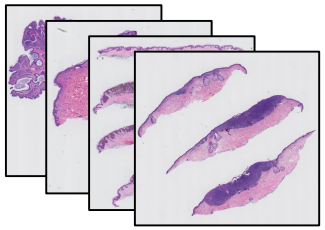






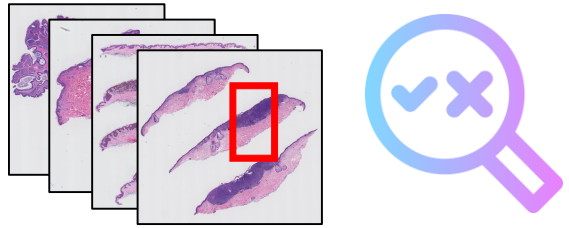
Class		Count
Class 1	Nevus	15
Class 2	Moderate atypia	30
class 3	Melanoma in situ	45
Class 4	Stage pT1a invasive melanoma	45
Class 5	Stage pT1b invasive melanoma	45

Data collection



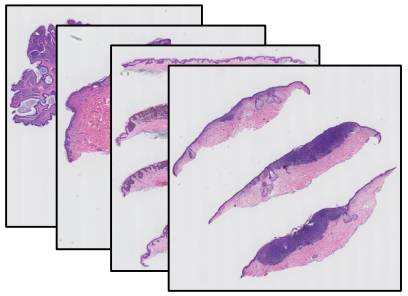
Digital skin biopsy images

3 Experts in dermatopathology



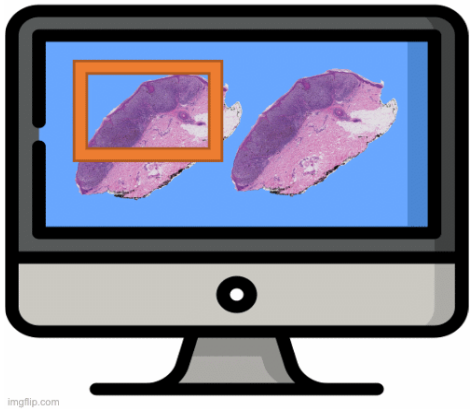
Consensus ROI and diagnosis

32 Pathologists

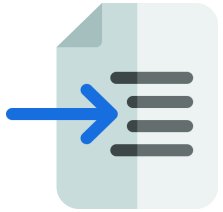


180 Cases
Divided into 5 groups of 36

Survey



Interpretation



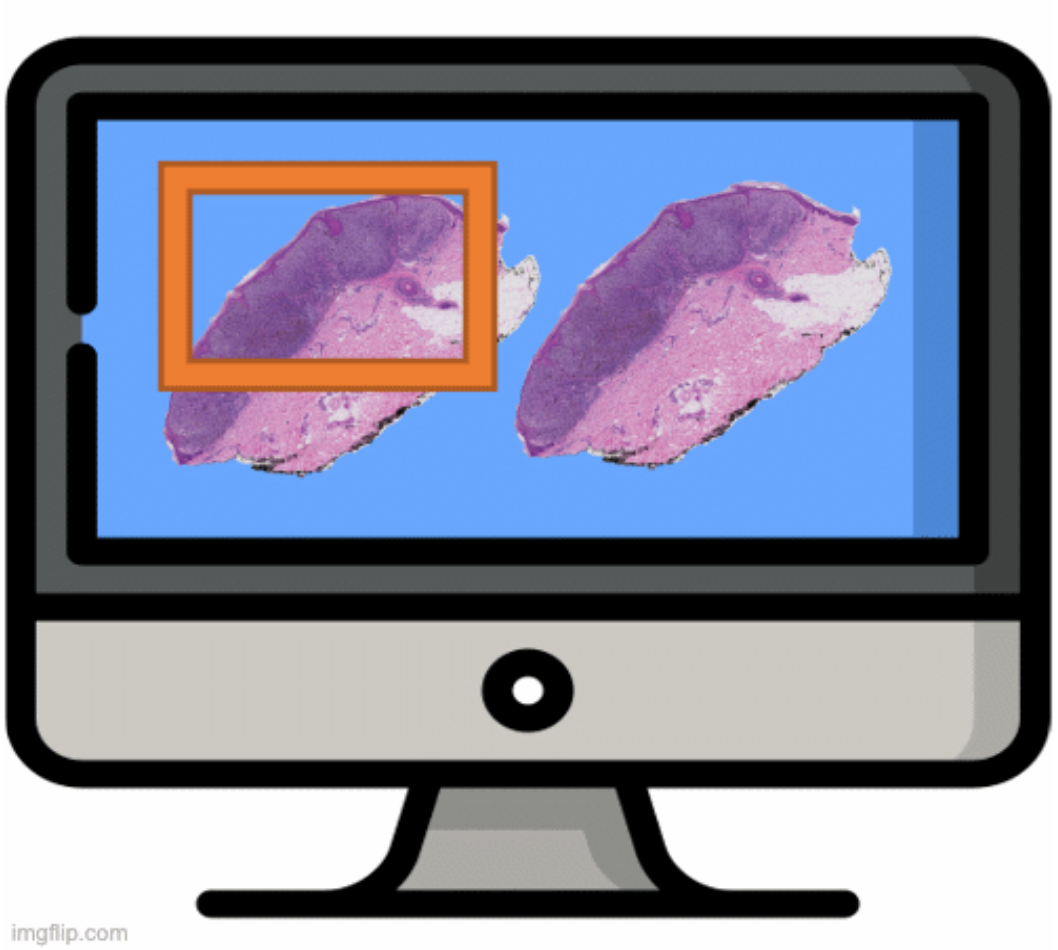
Viewport log

Diagnosis

Pathologists' characteristics

Pathologist Characteristics	Number (%)
Gender	
Male	13 (40.6%)
Female	19 (59.4%)
Age (years)	
20 – 49	12 (37.5%)
50 – 64	20 (62.5%)
Board certification/ Dermatopathology Fellowship training	
Yes	10 (31.3%)
No	22 (68.8%)
Experience with interpreting melanocytic skin lesions (years)	
< 5	3 (9.4%)
5 – 9	10 (31.3%)
10 – 19	9 (28.1%)
> 20	10 (31.3%)
Caseload of melanocytic skin lesions (%)	
< 10	14 (43.8%)
10 – 24	13 (40.6%)
25 – 49	5 (15.6%)

Viewport data

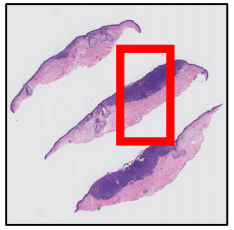


ID	X	Y	Width	Height	Zoom	Time
1	4	6	50	100	2	1:20:30
2	20	6	50	100	2	1:20:33
3	20	13	50	100	2	1:20:34
4	23	15	30	80	4	1:20:38
5	23	4	25	70	5	1:21:10
6	40	4	25	70	5	1:21:50
7	44	4	25	70	5	1:22:30
8	5	10	60	110	2	1:23:40
9	5	11	60	110	2	1:24:10
10	5	12	60	110	2	1:24:56

Viewing behaviors

- Total interpretation time — Duration of Interpretation

- Average zoom level — Average of zoom levels used during an interpretation
- Maximum zoom level — Maximum of zoom levels used during an interpretation
- Zoom level variance — Variance of zoom levels used during an interpretation
- Magnification — Percentage of viewports associated with consecutive zooming in



- ROI time Percentage — Percentage of time spent viewing consensus ROI to the total time

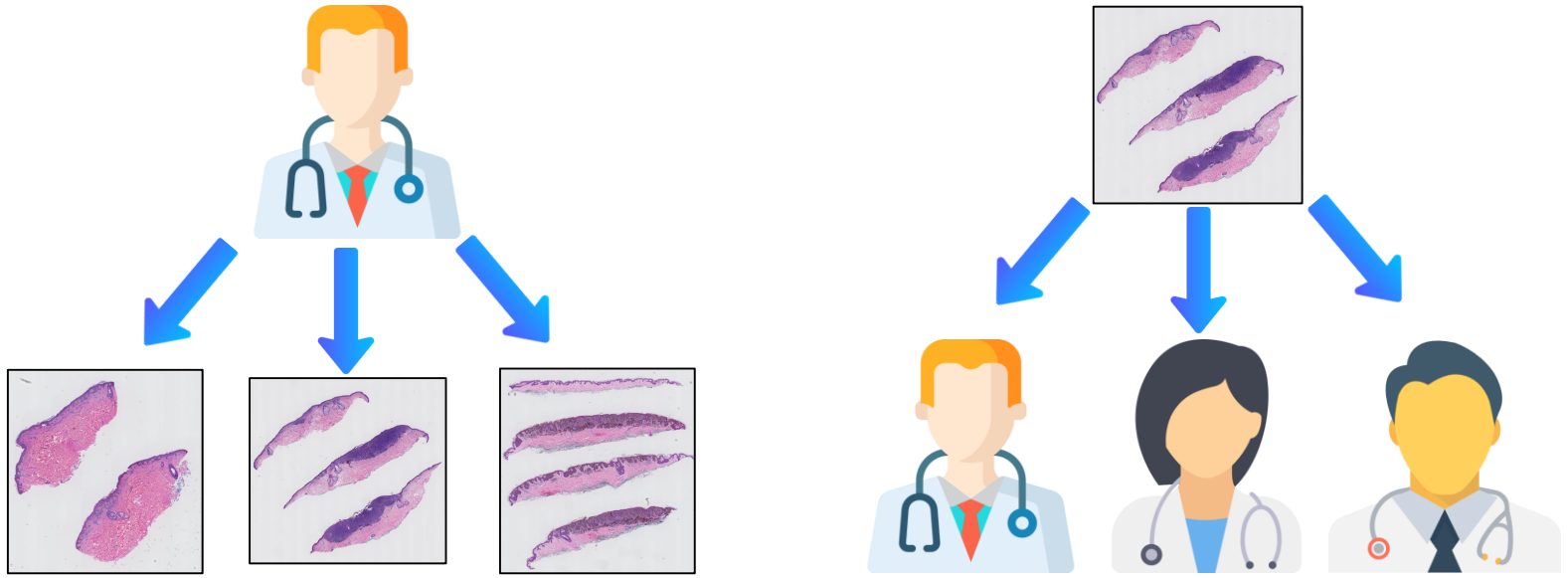
- Scanning Percentage — Percentage of viewports associated with fixed zoom level and panning

Statistical analysis

- Crossed-level structure of cases and pathologists
- Both case and pathologist contribute to the variation of the outcome

→ **Cross-classified multilevel model[1]**

$$y_i = \beta_0 + \beta_1 x_i + u_{pathologist(i)} + u_{case(i)} + e_i$$

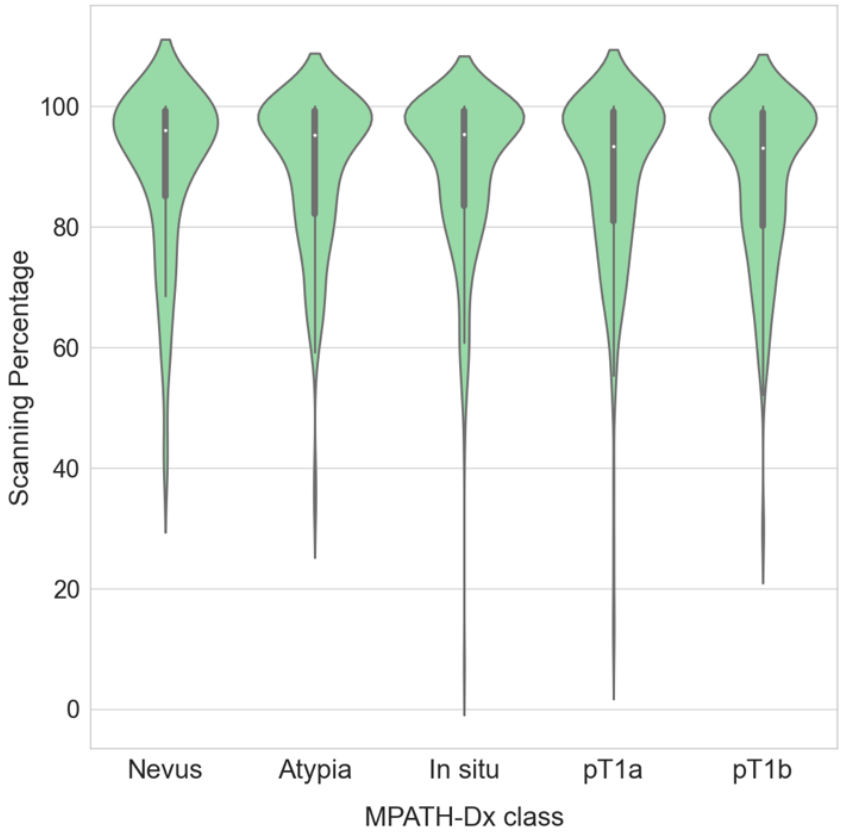


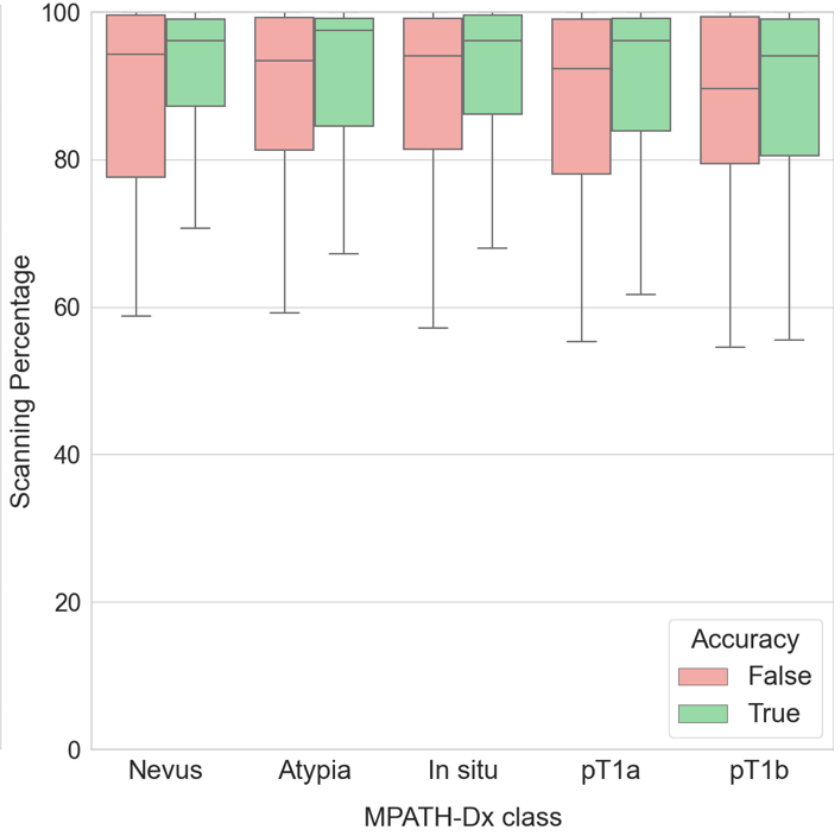
1. Browne, W.J., H. Goldstein, and J. Rasbash, *Multiple membership multiple classification (MMMC) models.* Statistical Modelling, 2001. **1**(2): p. 103-124.

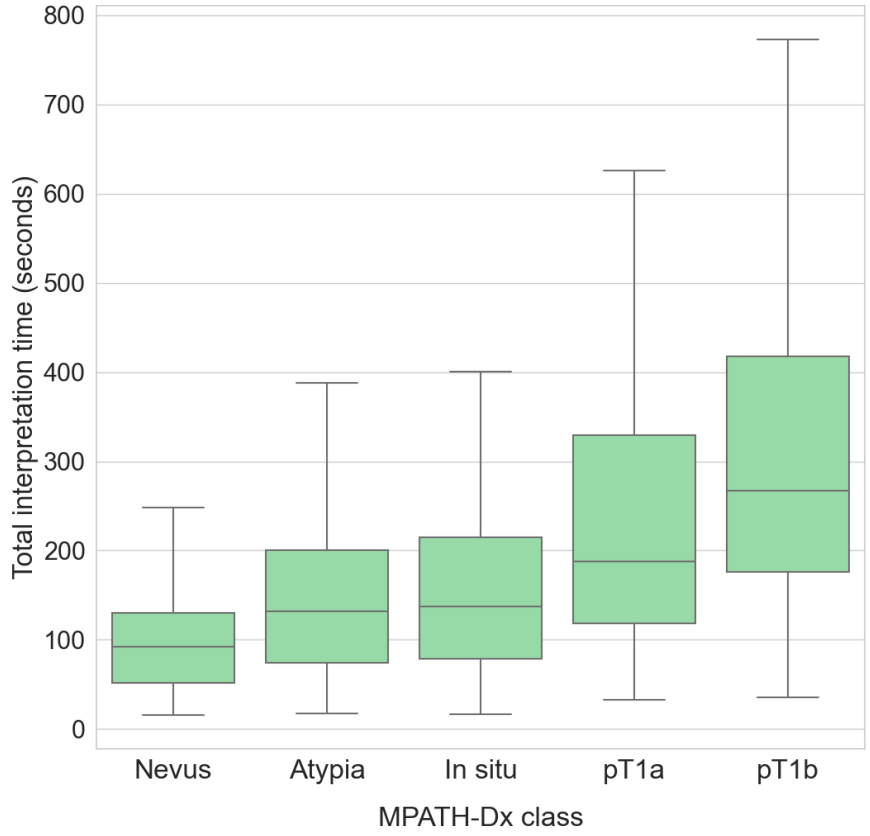
Outline

- Introduction
- Material and Methods
- **Results**
- Application
- Conclusions

Viewing behaviors visualization







Statistical analysis: Diagnostic accuracy

- How do specific viewing behaviors contribute to diagnostic accuracy?
  - ✓ Diagnostic accuracy is the agreement of a pathologist's diagnosis with the consensus diagnosis
  - ✓ Separate models with each viewing behavior as the predictor variable
  - ✓ Diagnostic accuracy is the outcome
  - ✓ All models are adjusted for pathologists' experience level and fellowship training

Predictor Variable	OR (95% CI)	P-value
Total time	1.33 (1.09, 1.62)	0.005
Average zoom	1.26 (1.03, 1.54)	0.023
Maximum zoom	1.24 (1.03, 1.50)	0.026
Zoom variance	1.37 (1.11, 1.68)	0.003
Magnification	0.76 (0.63, 0.92)	0.006
ROI time percentage	1.35 (1.07, 1.69)	0.011
Scanning percentage	1.21 (1.00, 1.47)	0.054

P-value < 0.05
OR > 1

P-value < 0.05
OR < 1

0.05 < P-value < 0.1
OR > 1

Statistical analysis:
Diagnostic accuracy - Multivariate model

PCA →

Viewing Behaviors
Total time
Average zoom
Maximum zoom
Zoom variance
Magnification
ROI time percentage
Scanning percentage

- ✓ Diagnostic accuracy is the outcome
- ✓ Model is adjusted for pathologists' experience level and fellowship training

Predictor Variables	OR (95% CI)	P-value
Total time	1.25 (1.01, 1.54)	0.0360
Zoom variance	1.22 (0.98, 1.53)	0.0786
ROI time percentage	1.38 (1.10, 1.73)	0.0058
Scanning percentage	1.20 (0.98, 1.47)	0.0716

Statistical analysis: Pathologists' characteristics

- How are pathologists' characteristics associated with specific viewing patterns?
  - ✓ Separate models with each characteristic as the predictor variable
  - ✓ Pathologists' viewing behaviors as the outcome

Pathologist's characteristic, clinical experience and ratings of difficulty and confidence on melanocytic skin lesions	**Average zoom**		**Maximum zoom**		**Zoom variance**	
	Contrast	P-value	Contrast	P-value	Contrast	P-value
Clinical characteristics						
Gender	0.03	0.878	0.05	0.825	0.22	0.290
Age	0.29	0.192	0.39	0.068	**0.41**	**0.038**
Clinical Experience Level						
Board certification/ Fellowship training	**-0.62**	**0.003**	**-0.45**	**0.037**	**-0.45**	**0.030**
Experience with melanocytic skin lesions	0.12	0.286	0.12	0.266	0.16	0.100
Caseload of melanocytic skin lesions	**-0.35**	**0.015**	**-0.29**	**0.039**	**-0.31**	**0.017**
Ratings on melanocytic skin lesions						
Difficulty level	-0.05	0.765	-0.03	0.849	-0.09	0.534
Confidence level	**0.21**	**0.044**	**0.19**	**0.059**	**0.17**	**0.075**

Discussion

- How do specific viewing behaviors and search patterns contribute to diagnostic accuracy?
  - ✓ More total time, zoom variables, and ROI time, and scanning percentage
  - ✓ Less magnification

  → Higher accuracy

- How are pathologists' characteristics associated with specific viewing patterns?
  - ✓ Higher age category → Higher zoom variance
  - ✓ Lower confidence level → Higher average zoom, maximum zoom, and zoom variance
- How do viewing patterns change as pathologists gain more expertise in diagnosing melanocytic lesions?
  - ✓ Having Board certification and/or fellowship training
  - ✓ Higher caseload of melanocytic skin lesions

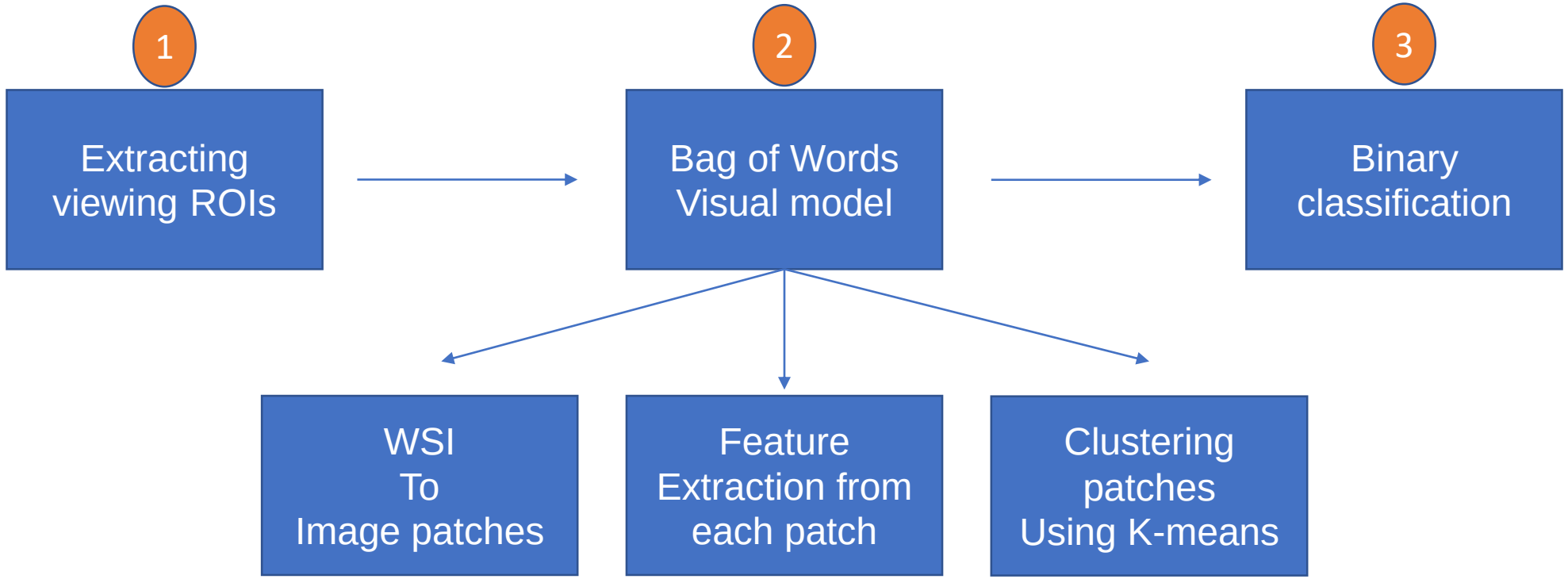
  → Lower average zoom, maximum zoom, and zoom variance

Outline

- Introduction
- Material and Methods
- Results
- **Application**
- Conclusions

ROI localization[1]

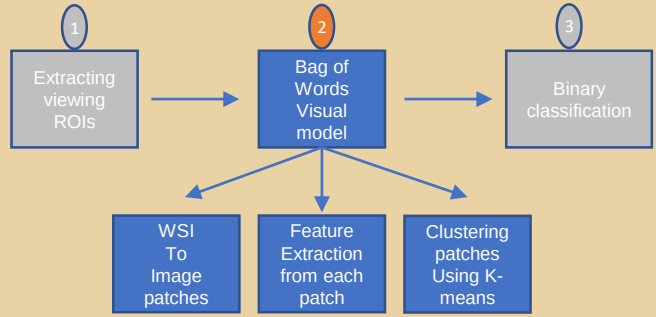
➢**Problem**: Predicting and Localizing diagnostically relevant Regions of Interest (ROIs)

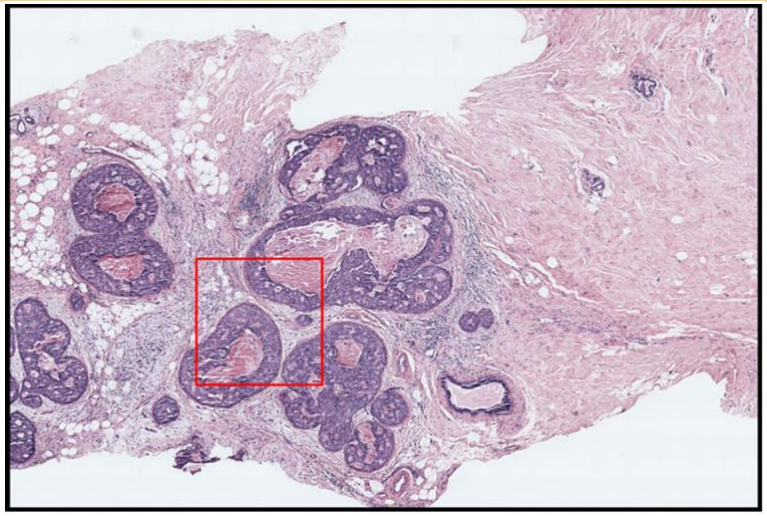


1. Mercan, Ezgi, et al. "Localization of diagnostically relevant regions of interest in whole slide images." *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014.

ROI localization

Bag of Words





Whole slide image with a red Sliding window (3600 x 3600)

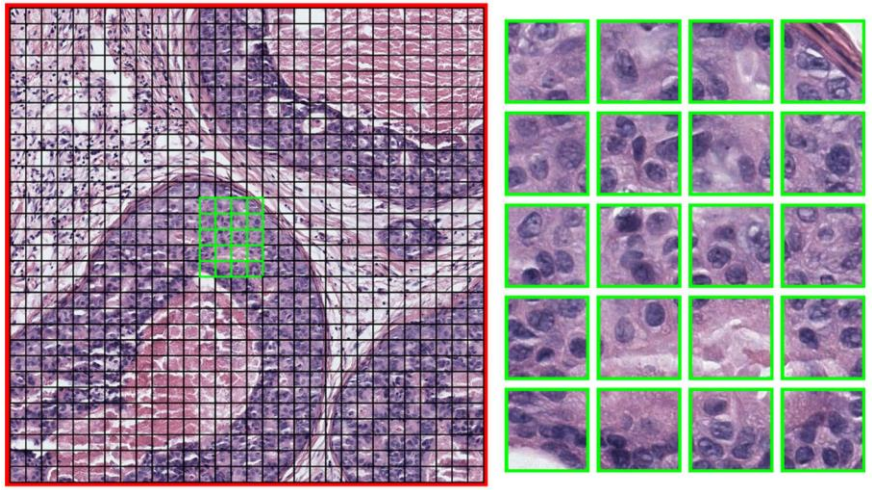
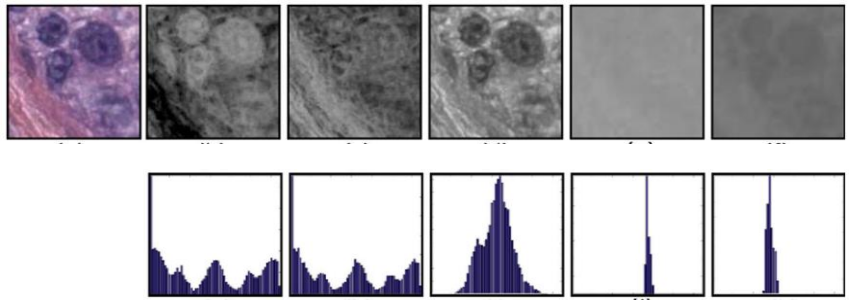
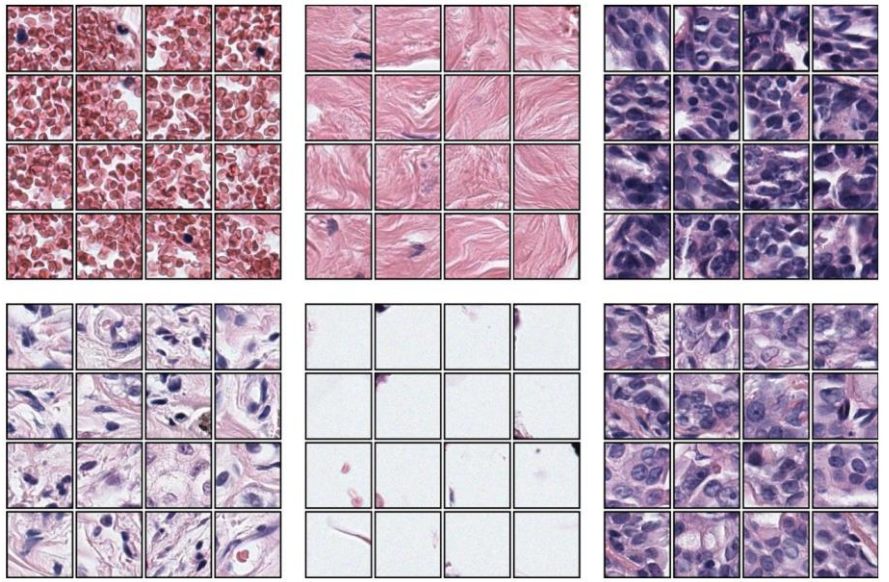


Image Patches (120 * 120)



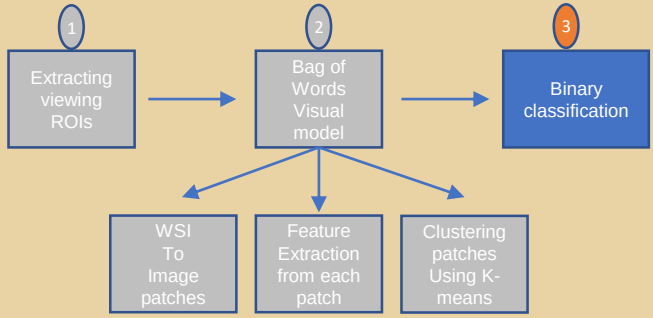
L*a*b color and LBP texture histograms



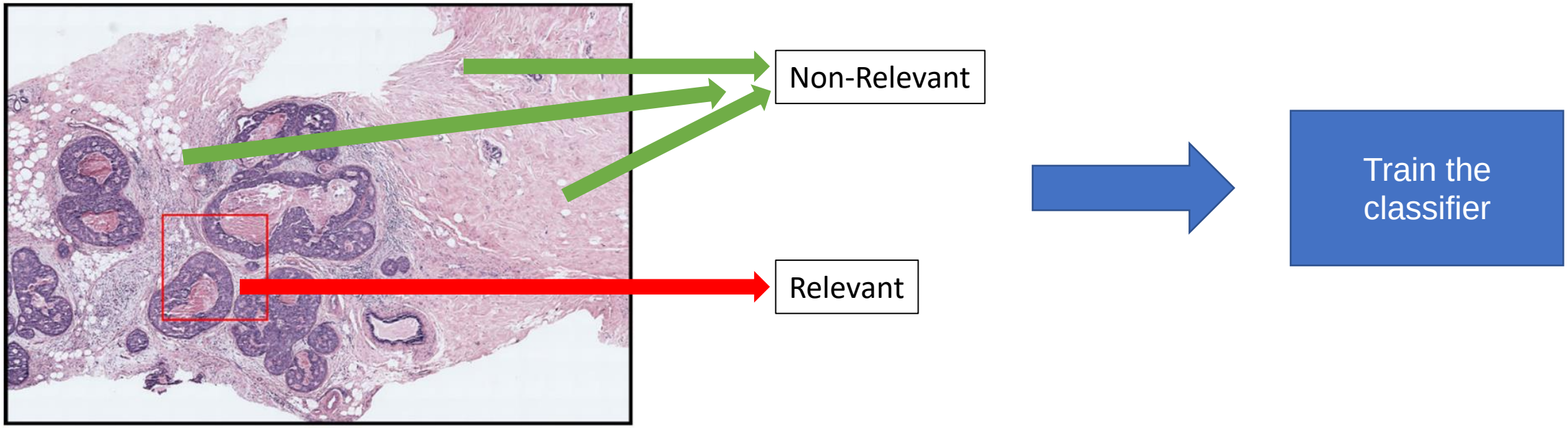
Clusters of Patches

ROI localization

Training classifier



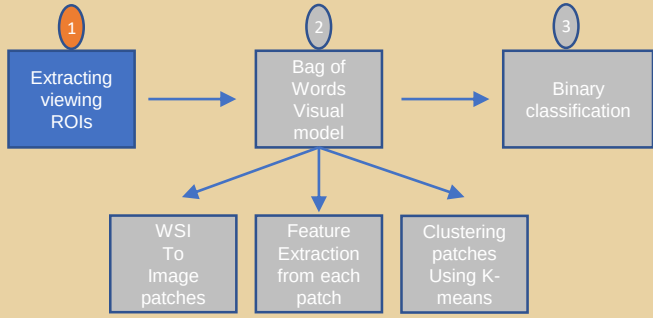
- Binary classification (relevant vs non-relevant)
  - Logistic regression
  - SVM



Whole slide image with ROI selected by expert pathologists marked in red

ROI localization

Extracting Viewing ROIs

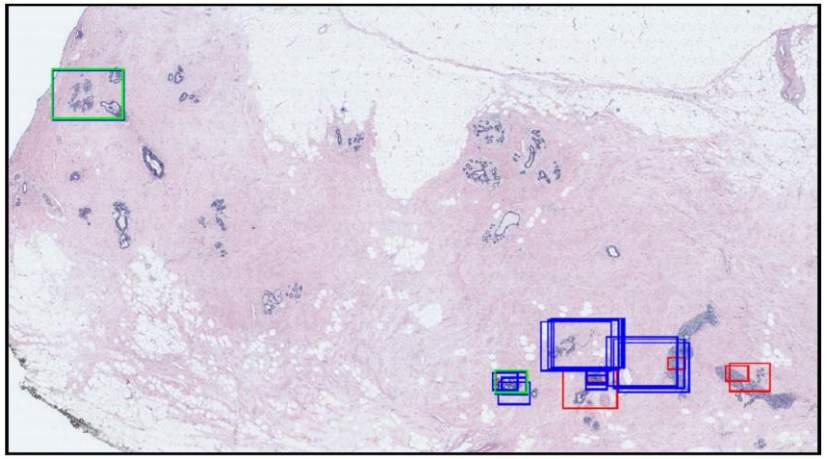


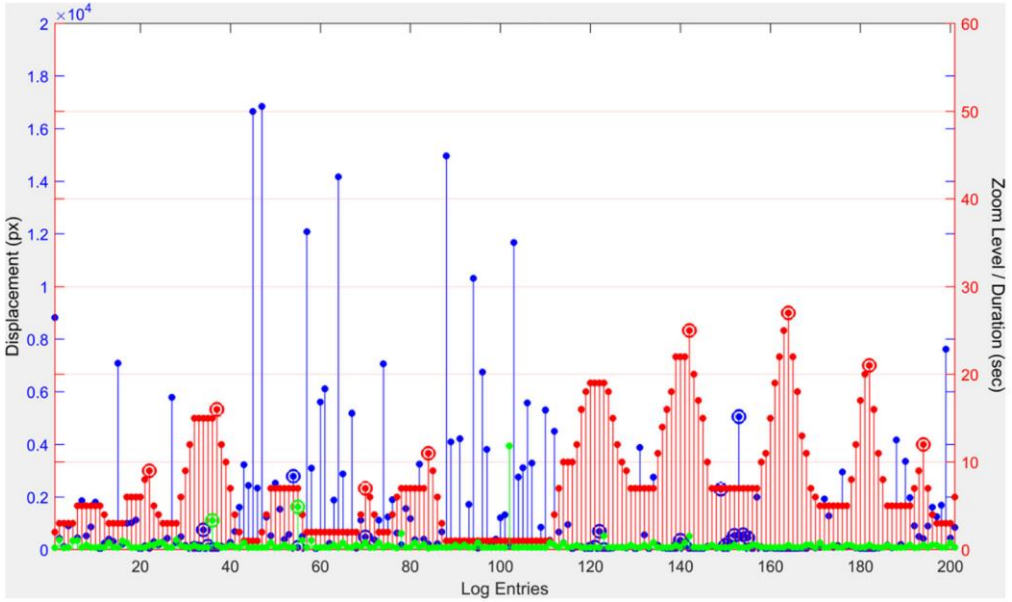
- Use the viewing data from pathologists to extract regions
- Use these regions as ground truth for testing the classifier

1. **Zoom peaks** are the log entries where the zoom level is higher than the previous and the next entries. A zoom peak identifies a region where the pathologist intentionally zoomed to look at a higher magnification.

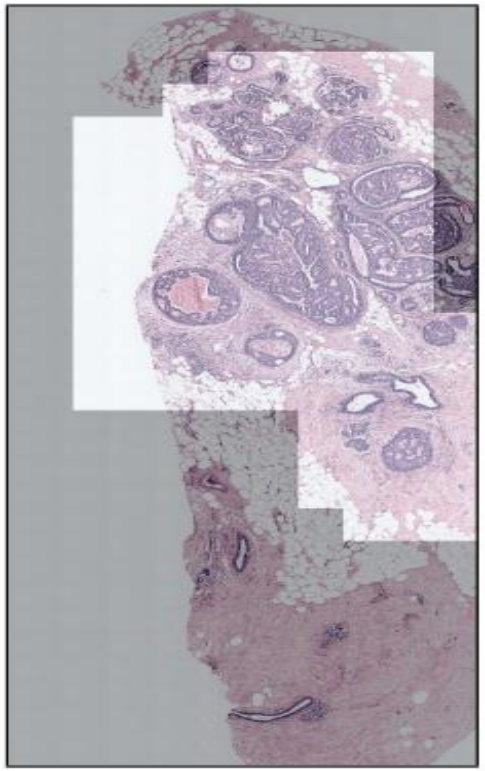
2. **Slow pannings** are the log entries where the zoom level is the same as the previous entry, and the displacement is small. Slow pannings are intended for investigating a slightly larger and closer area without completely moving the viewport.

3. **Fixations** are the log entries where the duration is longer than 2 seconds. Fixations identify the areas to which a pathologist focused extra attention by looking at them longer.

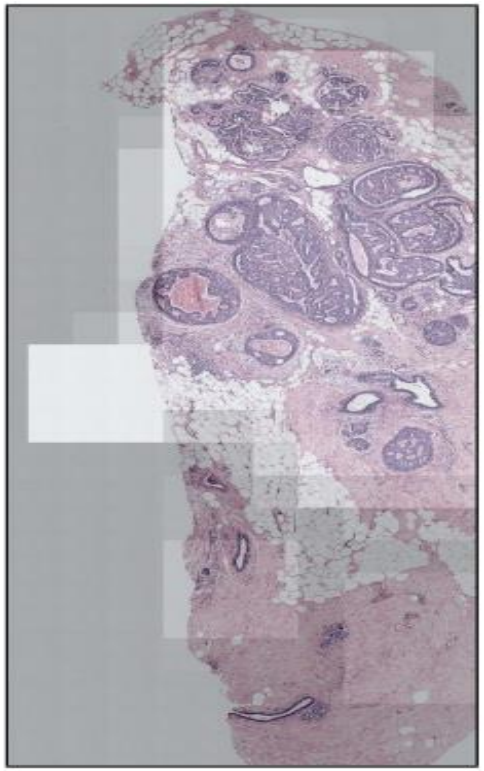




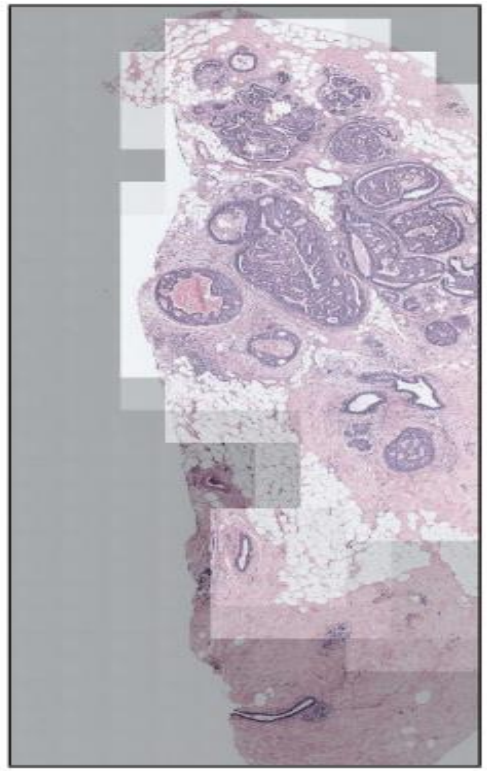
ROI localization - Results



Ground Truth
Viewing ROIs



Logistic Regression



SVM

Outline

- Introduction
- Material and Methods
- Results
- Application
- **Conclusions**

Limitations

High quality digital slide preparation is costly.

Ethical issues must be resolved when sharing patient information across larger platforms.

Experimental sample sizes in pathology studies are often small.

A thorough understanding of a variety of statistical principles is required.

Access to a multidisciplinary team of professionals, including statisticians and pathologists.

Conclusions

- Diagnosis of pathology slides is a complex task and requires years of training
- It is essential to study pathologists' viewing behaviors
- Digital pathology has made it possible to record and study these viewing behaviors
- We showed various ways of quantifying these behaviors
- We investigated the association of these viewing behaviors with accuracy
- The results of such behavioral studies can be beneficial in various areas
  - ✓ Improving the training and education of younger pathologists
  - ✓ Determining the reasons for diagnostic errors
  - ✓ Assisting with the development of computer-aided tools for diagnosis purposes

All icons taken from https://www.flaticon.com/