Computer-Aided Diagnosis Using Whole Slide Images

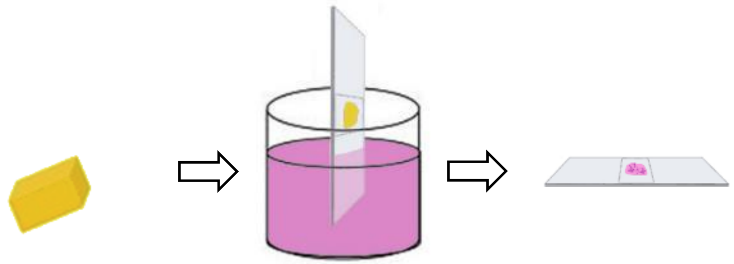
Linda G. Shapiro, Ph.D.
Paul G. Allen School of Computer Science & Enginnering
Department of Electrical and Computer Engineering
Department of Biomedical Informatics and Medical Education
University of Washington

Breast Cancer Research

- With PI Dr. Joann Elmore under the “Digipath” grant, R01-CA172343
- Most of the work performed by Dr. Ezgi Mercan, now at Children’s Hospital and Research Institute.
- Ezgi’s dissertation covered multiple aspects of our study including:

1. Region of interest detection
2. Semantic segmentation into histopathological classes
3. Feature extraction and diagnosis
4. Region of interest identification and diagnostic concordance
5. Characterization of diagnostic search patterns: drillers vs. scanners

Medical Diagnosis of Cancer

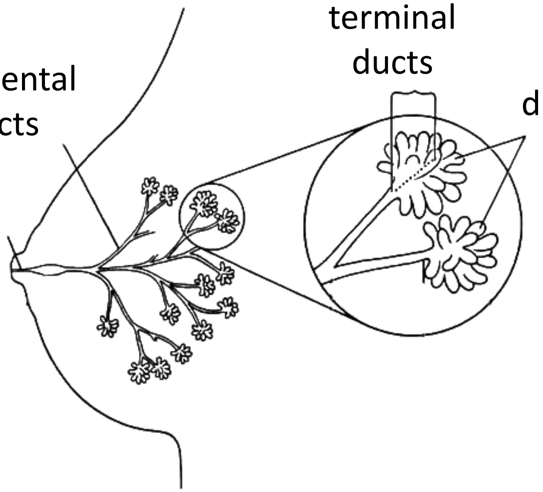


tissue specimen *H&E staining* *glass slide*

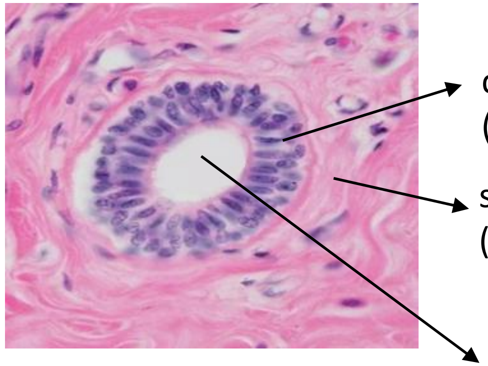
light microscope

whole slide imaging

Breast Histopathology

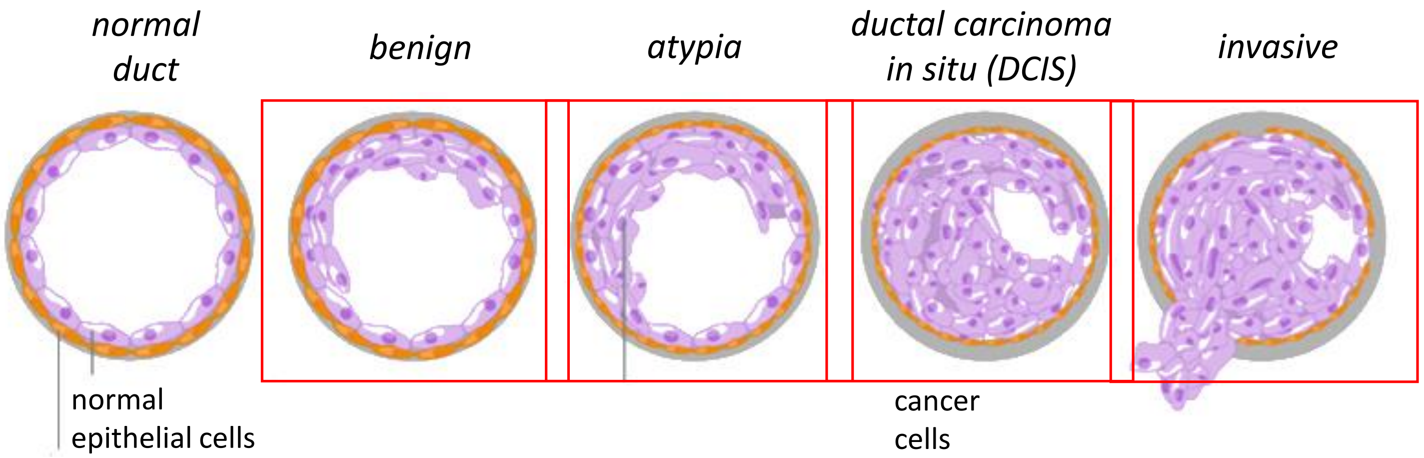


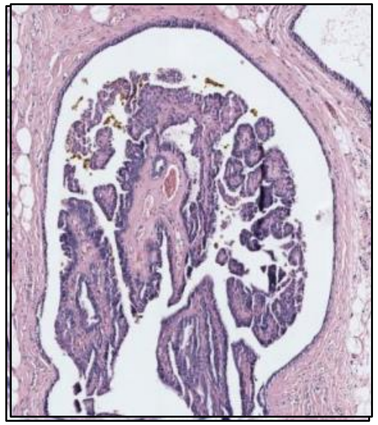
female breast



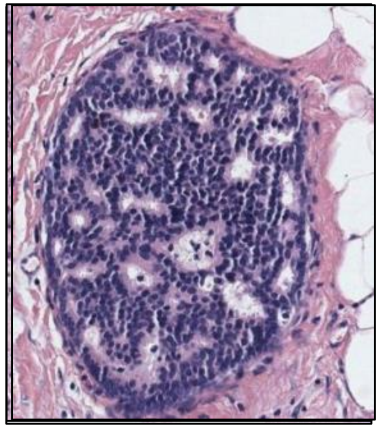
a duct

Diagnostic Categories

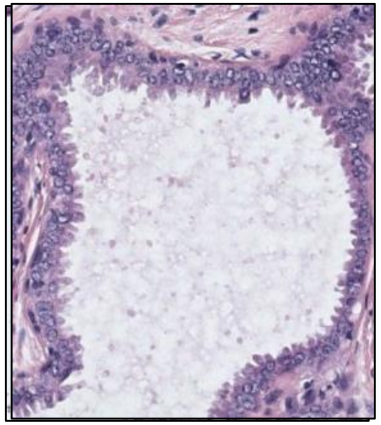




papilloma w/atypia / *columnar cell change*



ductal carcinoma in situ (DCIS) / *atypical ductal hyperplasia*



flat epithelial atypia / *intraductal papilloma*

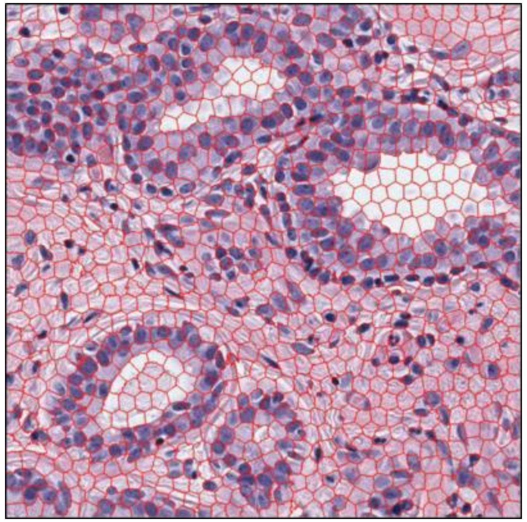
Tissue Label Segmentation

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

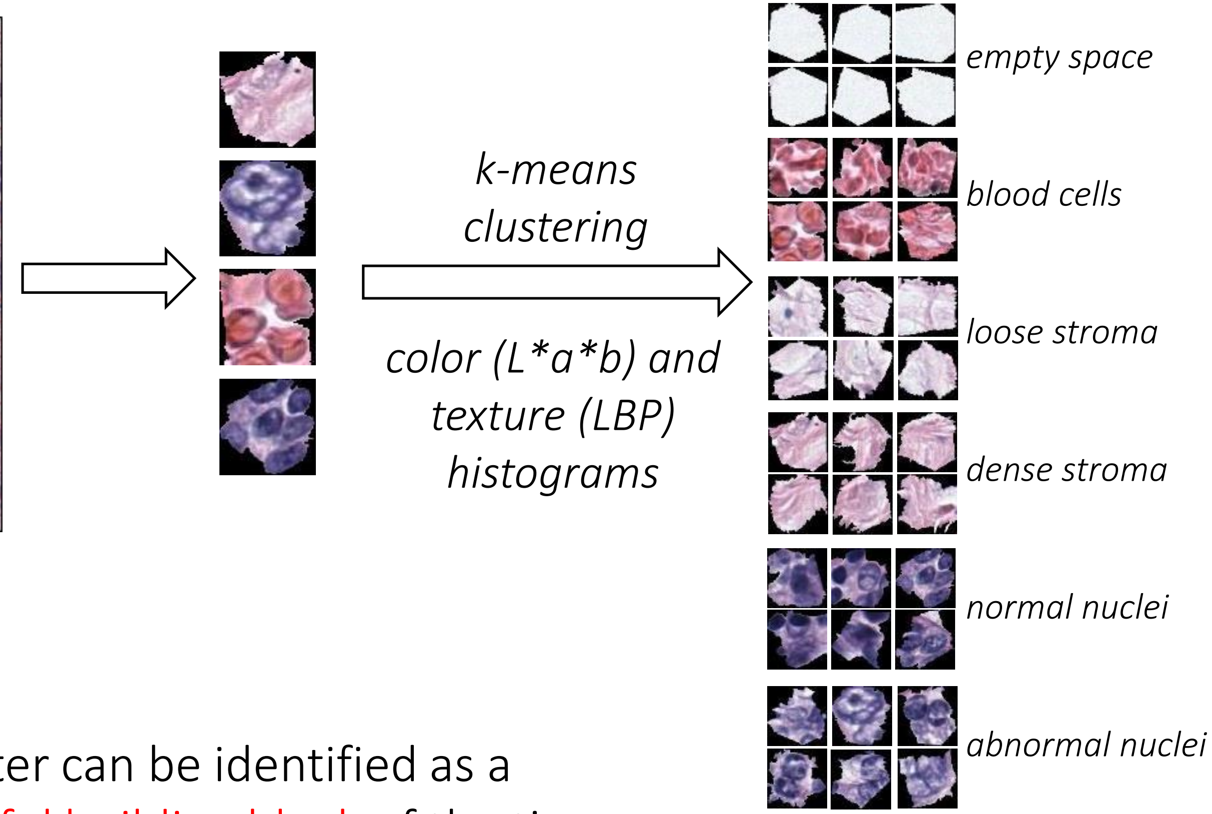
- Segmentation is a powerful that provides information about the distribution and arrangement of different tissue types.

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

Superpixel Clustering



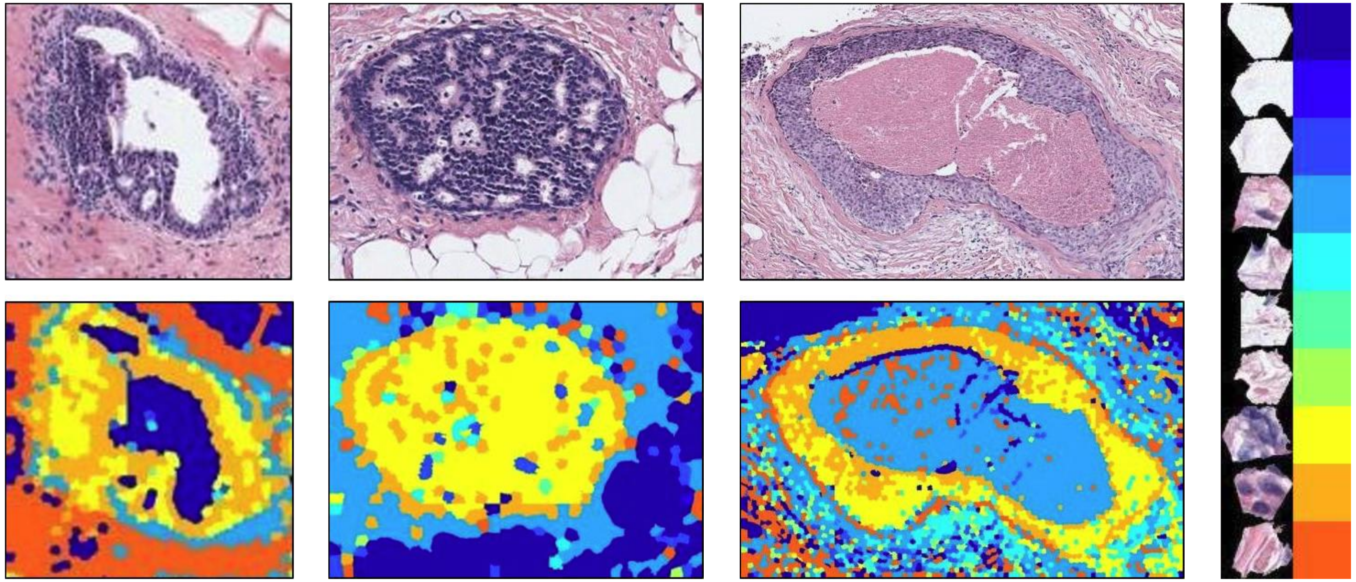
superpixel segmentation of an ROI



superpixel clusters

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

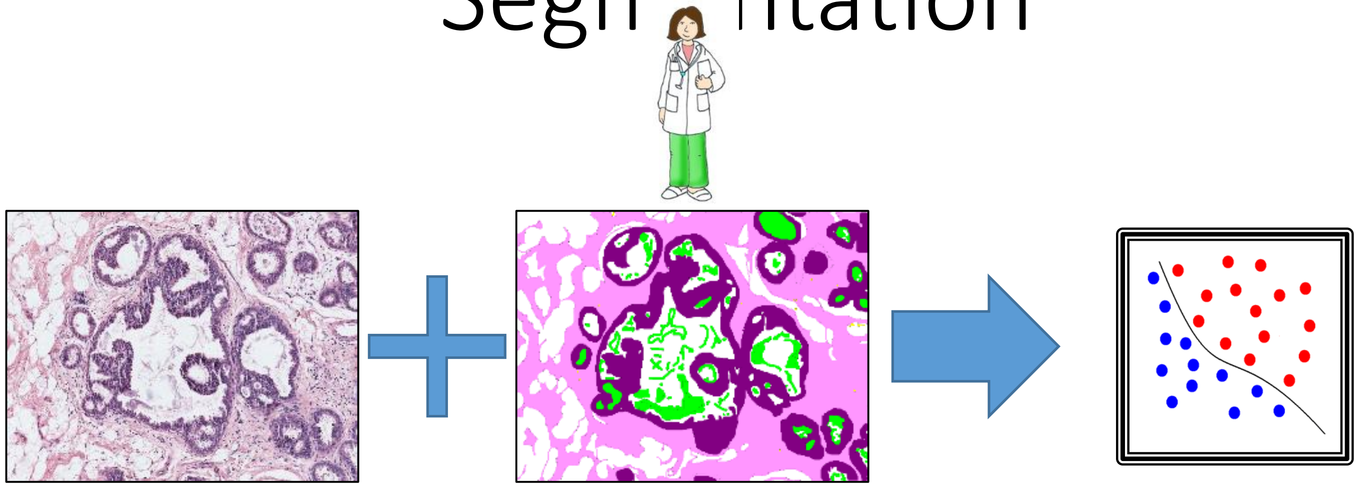
Superpixel Clustering



benign *atypia* *DCIS*

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

Supervised Tissue Label Segmentation



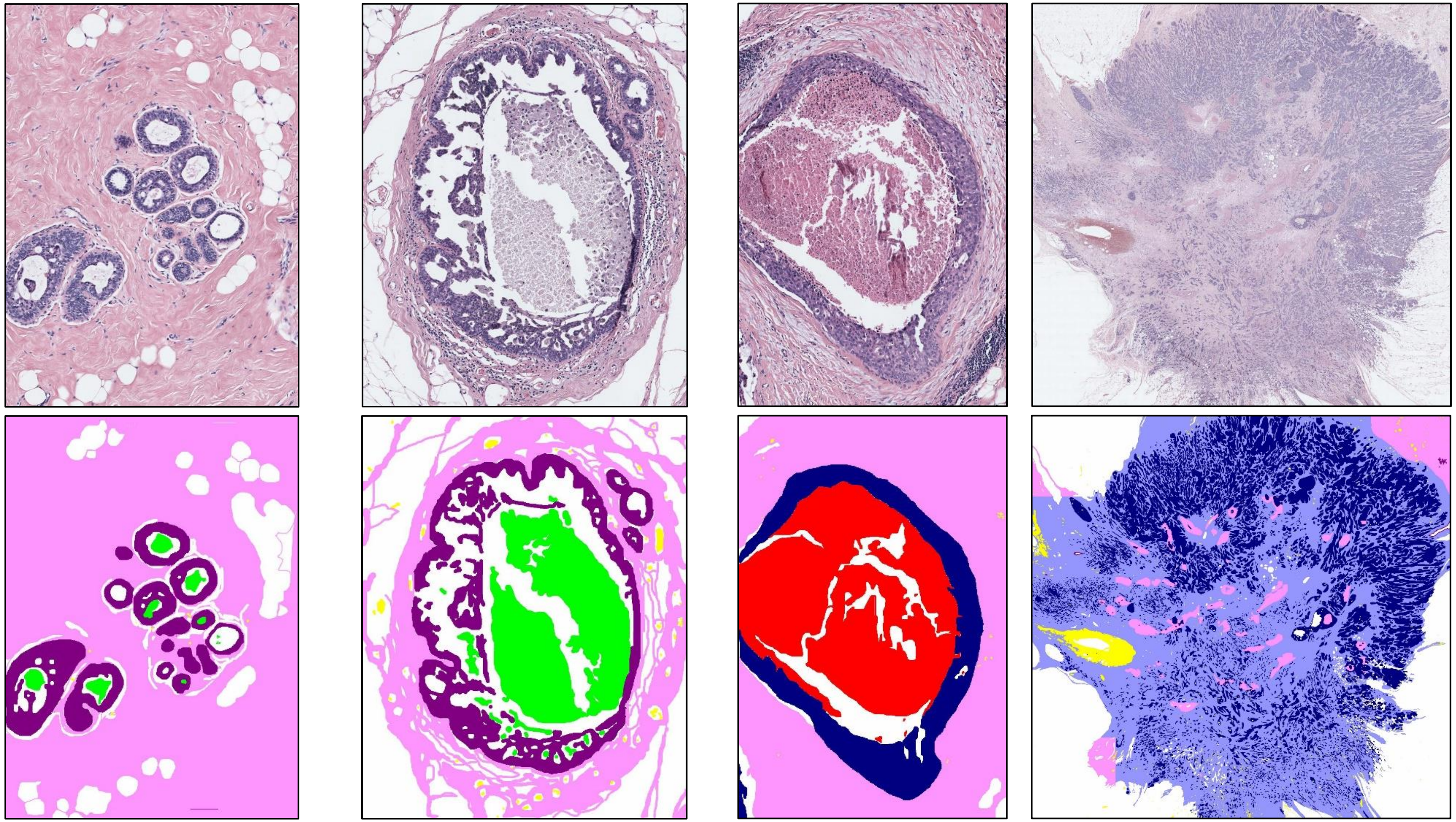
original image | ground truth label image | classifier

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

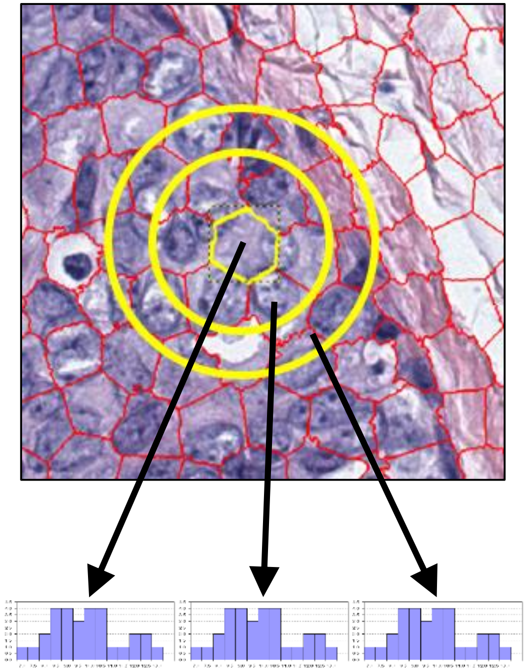
- Support Vector Machines (SVM)
- Convolutional Neural Nets (CNN)

Training Labels

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

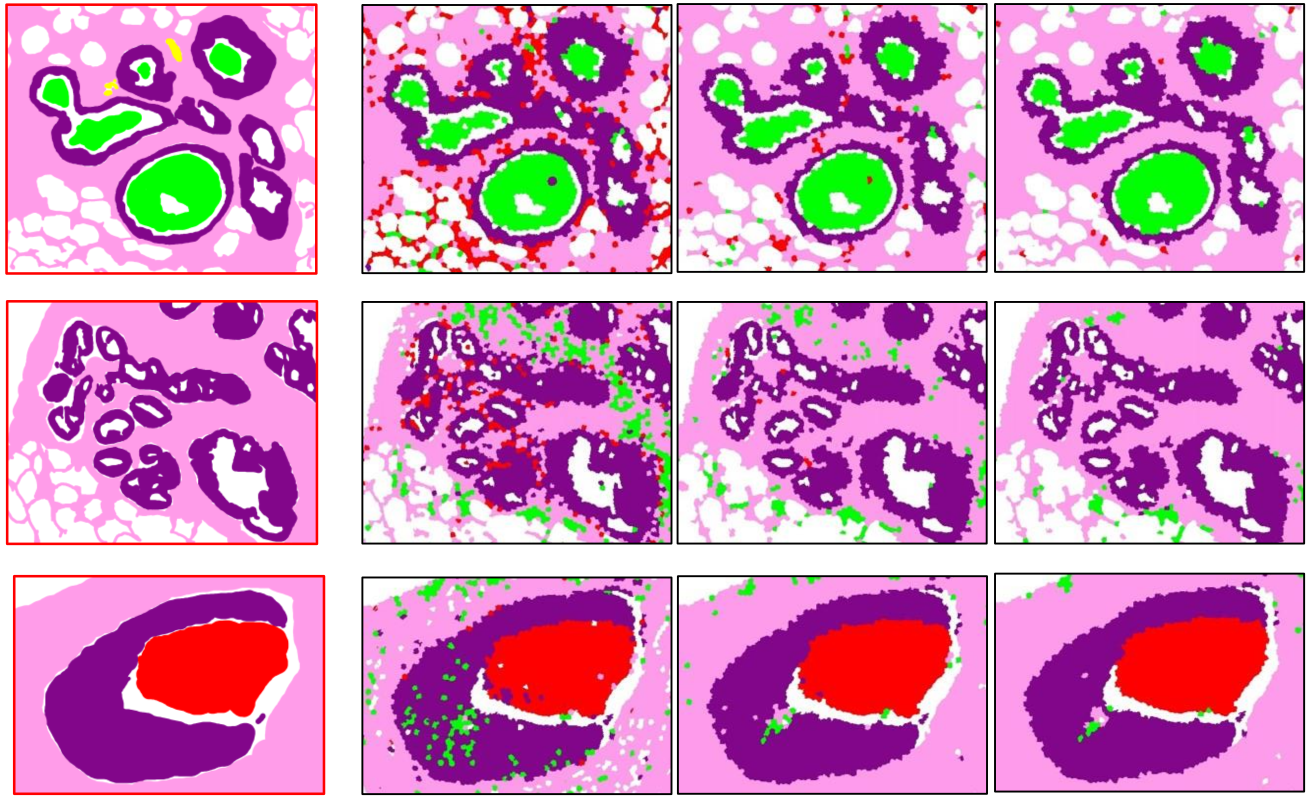


Superpixel + SVM-based Segmentation



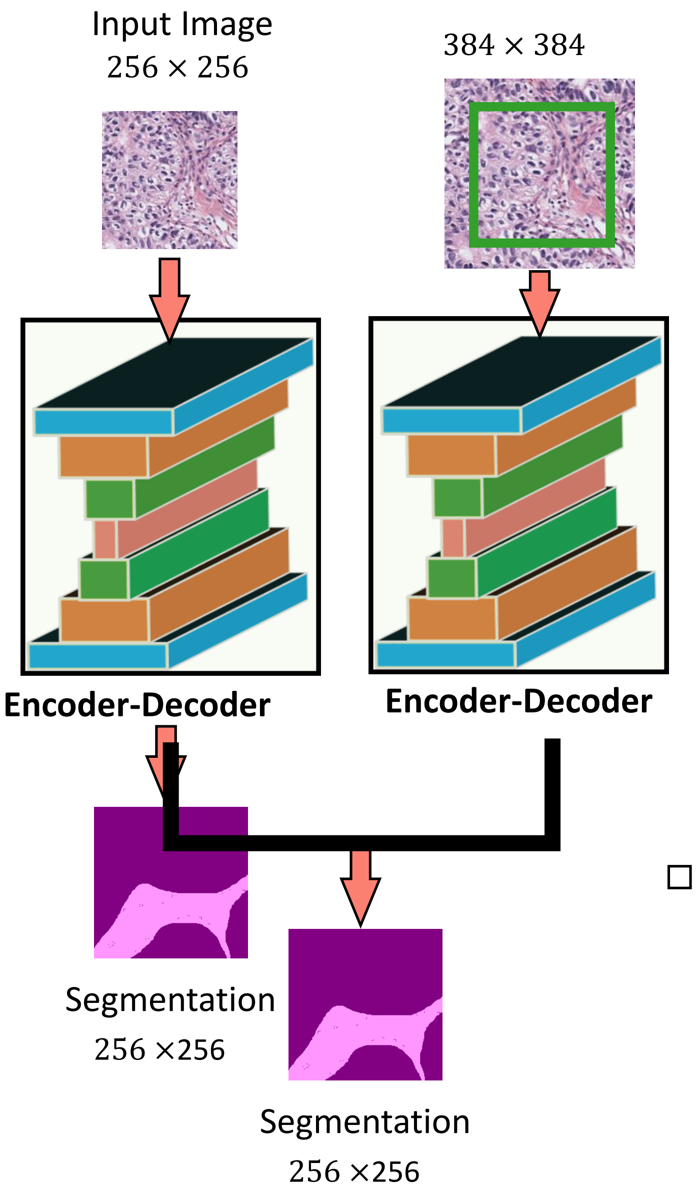
color and texture histograms

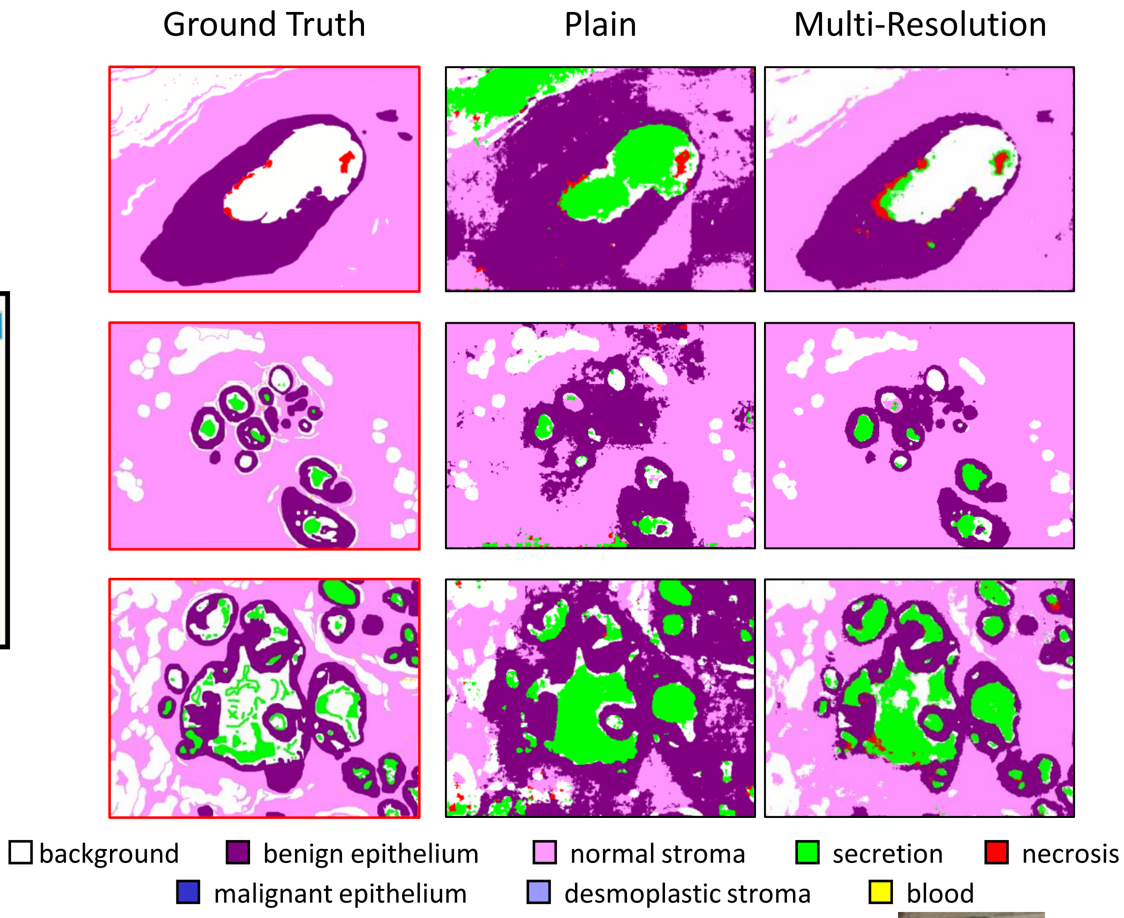
Ground Truth | no neighborhood | 1 neighborhood | 2 neighborhoods



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

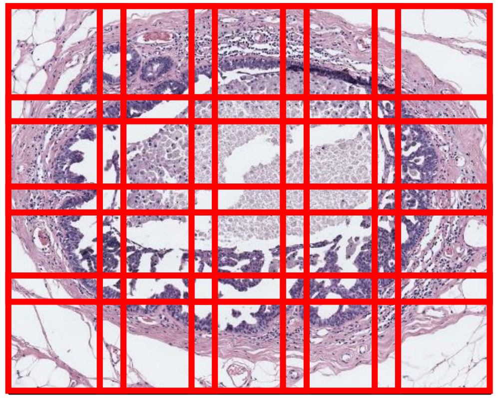
CNN-based Segmentation





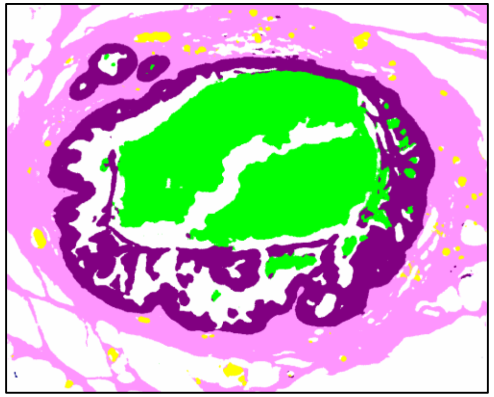
joint work with Sachin Mehta

CNN-based Segmentation

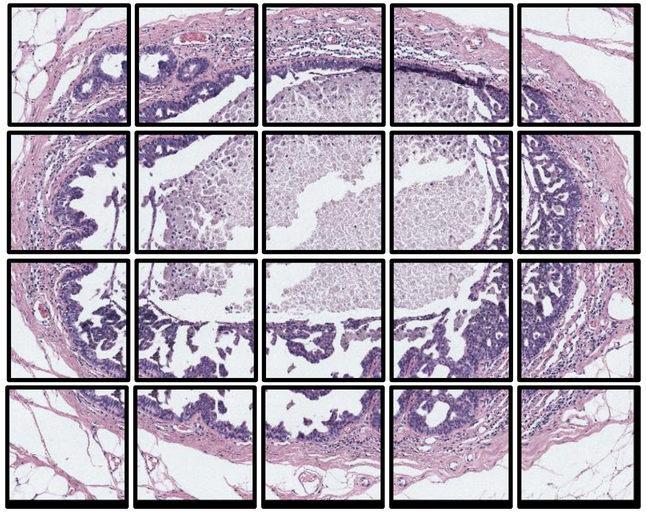


ROI

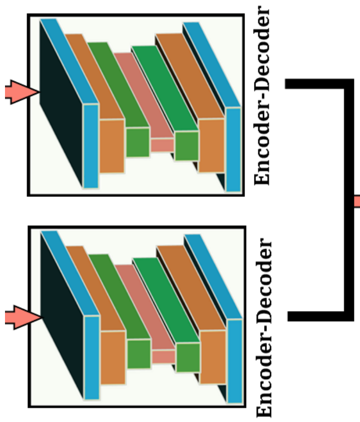
Training set: 38 ROIs
Test set: 20 ROIs



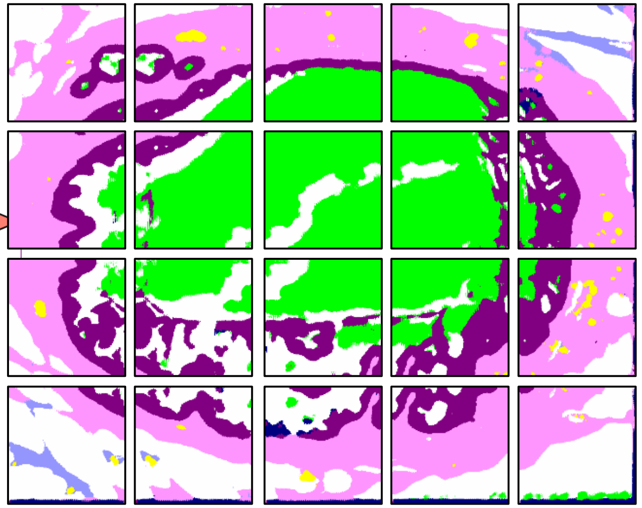
Segmented ROI



Overlapping Patches
256x256 pixel



CNN

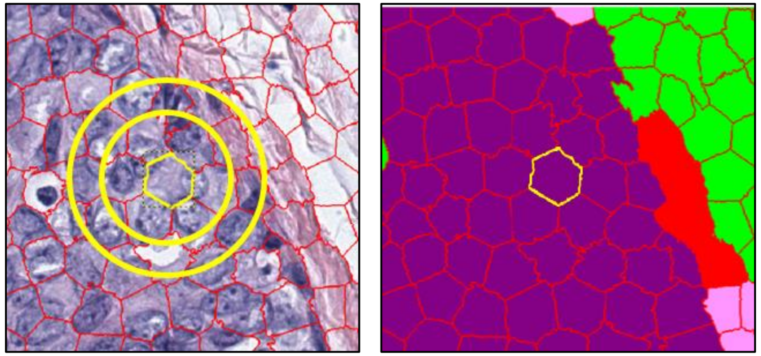


Segmented Patches

Supervised Tissue Label Segmentation

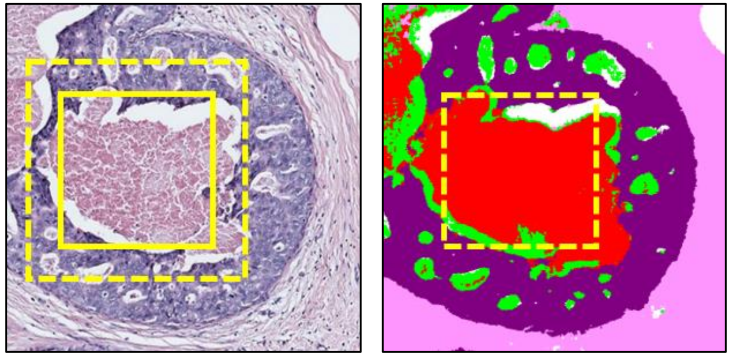
Superpixel + SVM

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



CNN

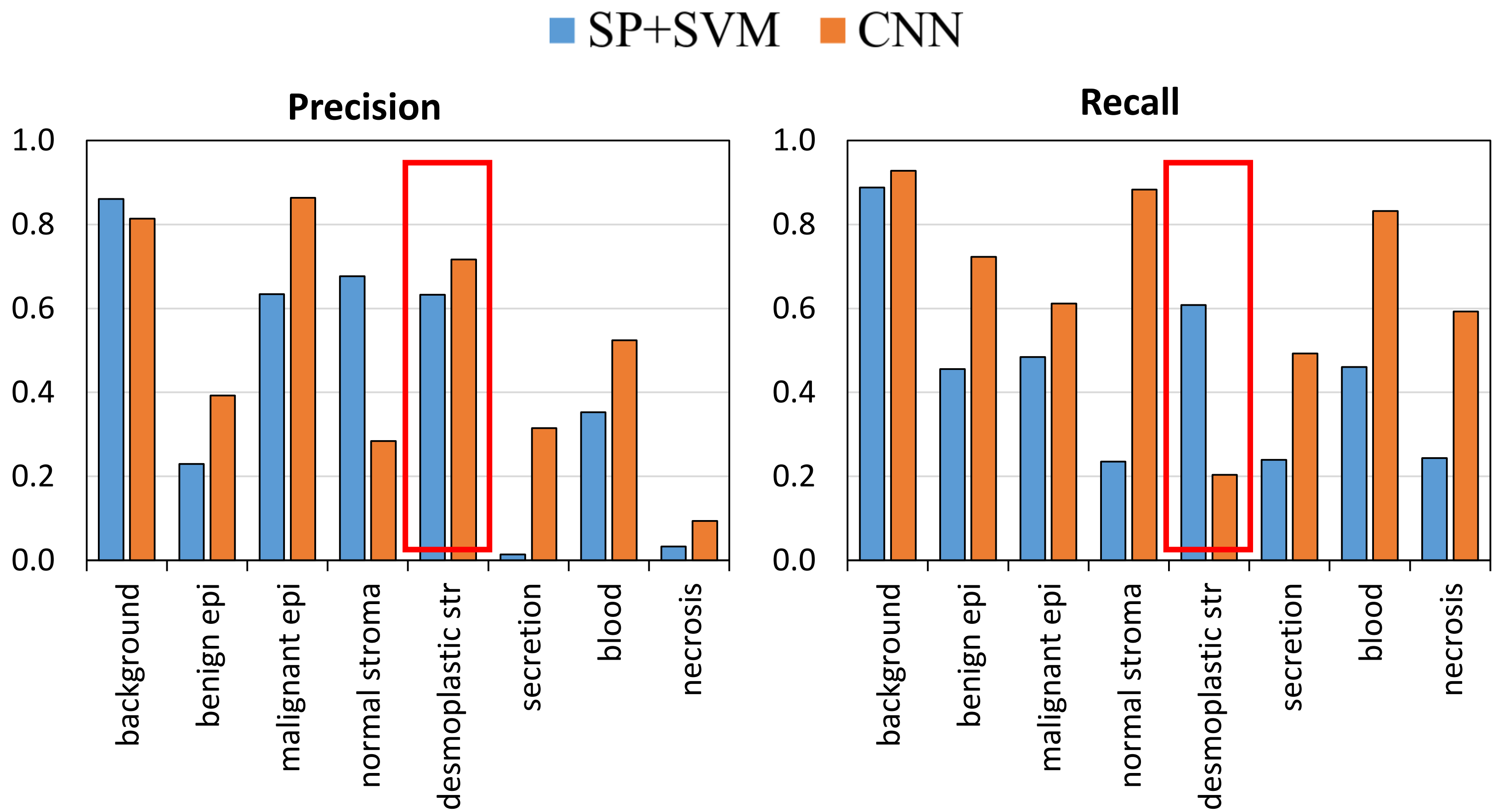
- Each pixel is assigned a class label.
- Context: 256x256 and 384x384 pixel patches
- More complex model
- ~1 week to train on special hardware



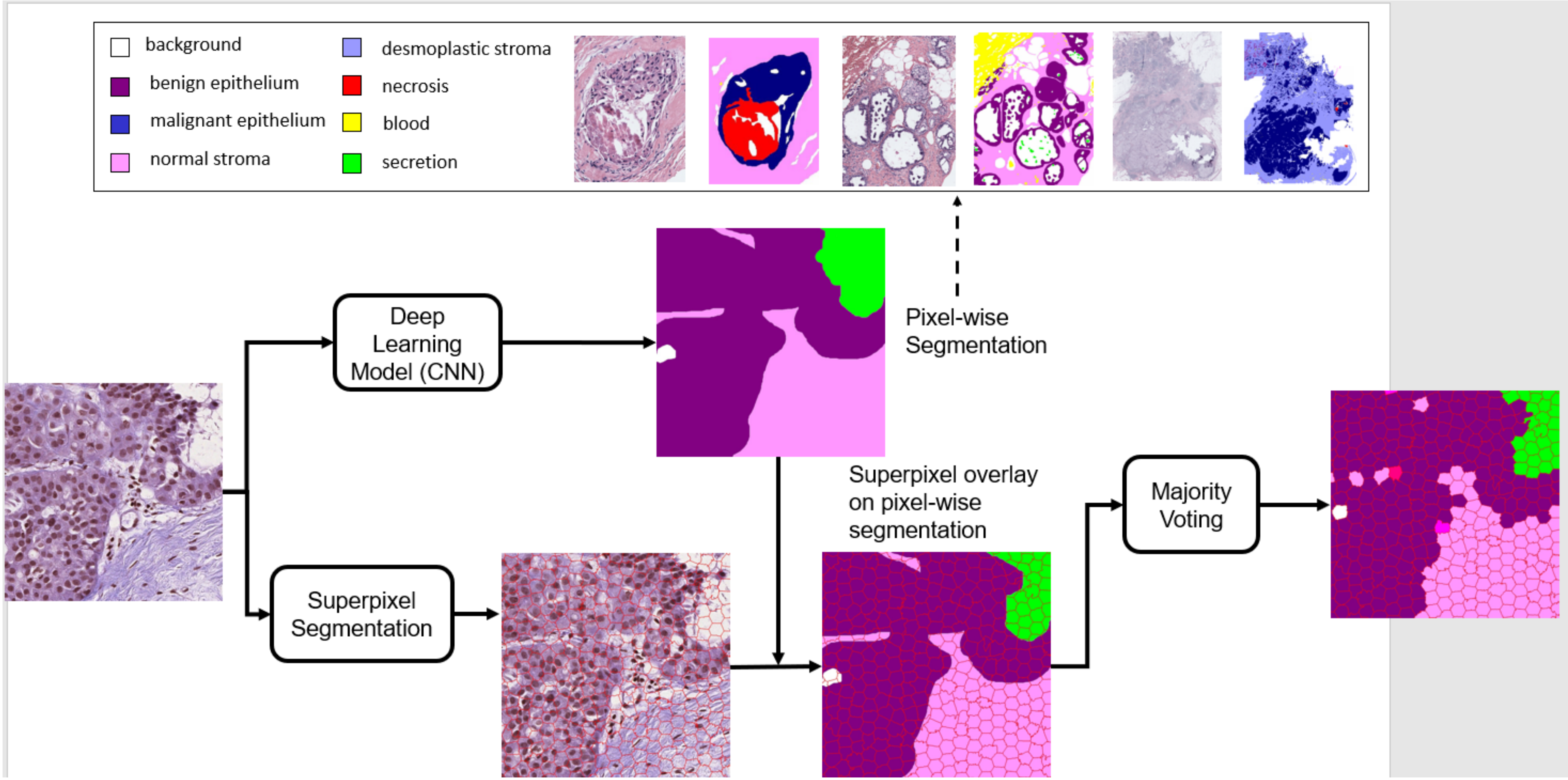
Results

Mean F_1-score

SP+SVM	0.40
CNN	0.50



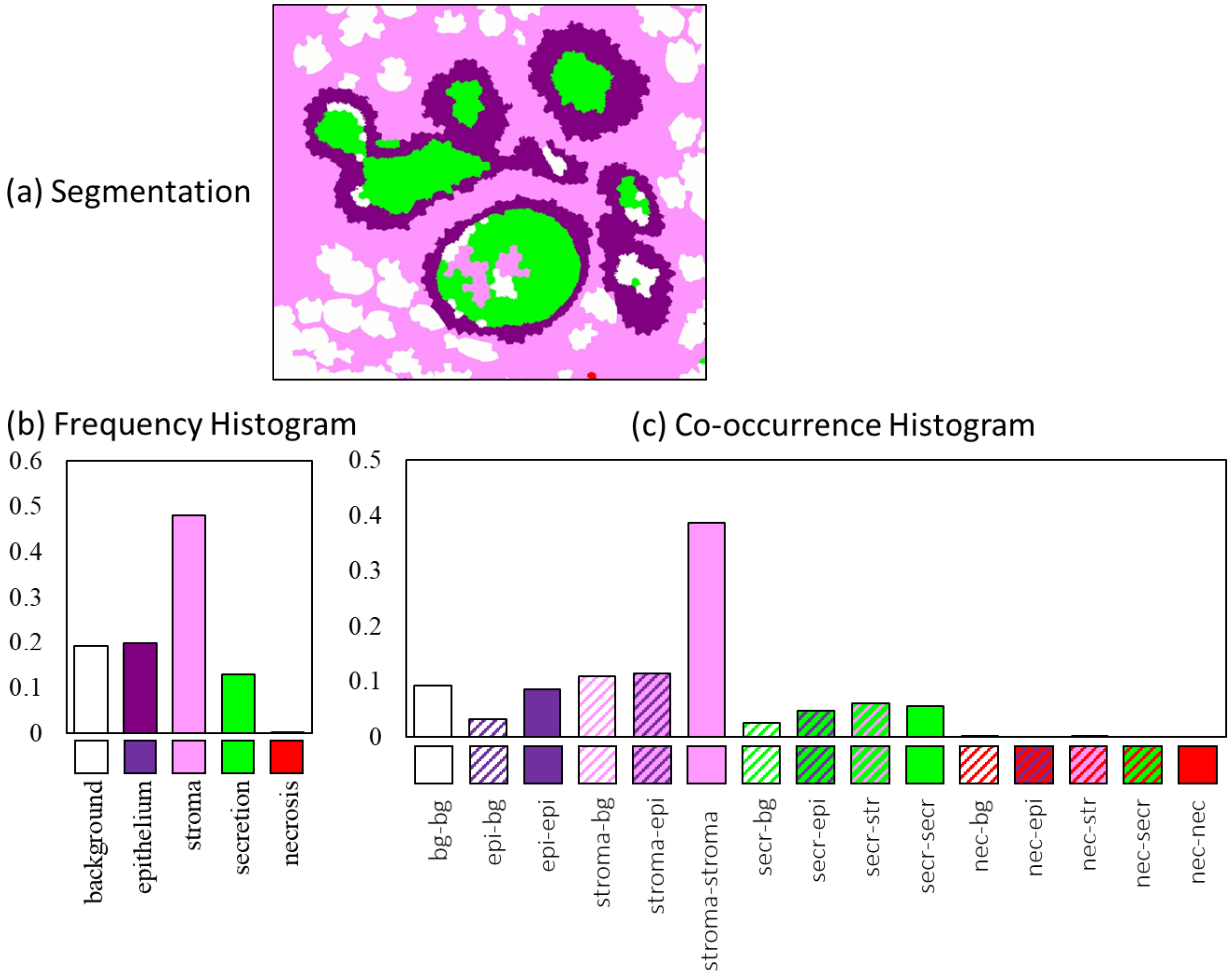
Semantic Segmentation of Tissue Classes



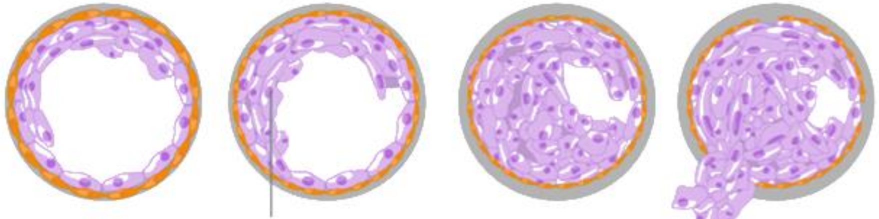
Automated Diagnosis of ROIs

- 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.

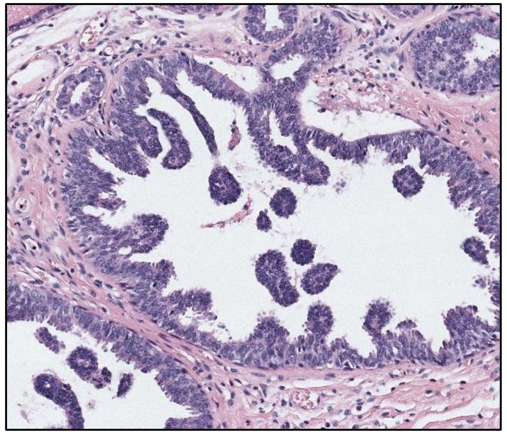
Features of the whole ROI: Frequency and Co-occurrence of tissue labels

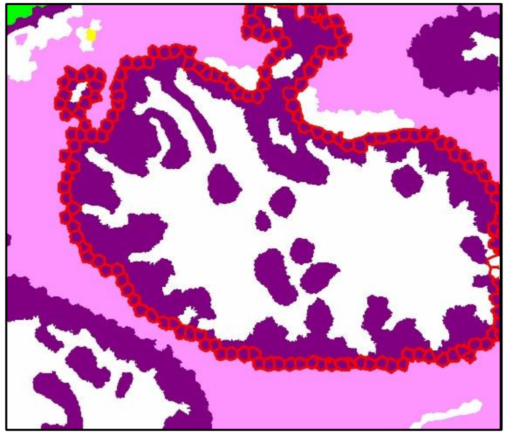


Structure Feature

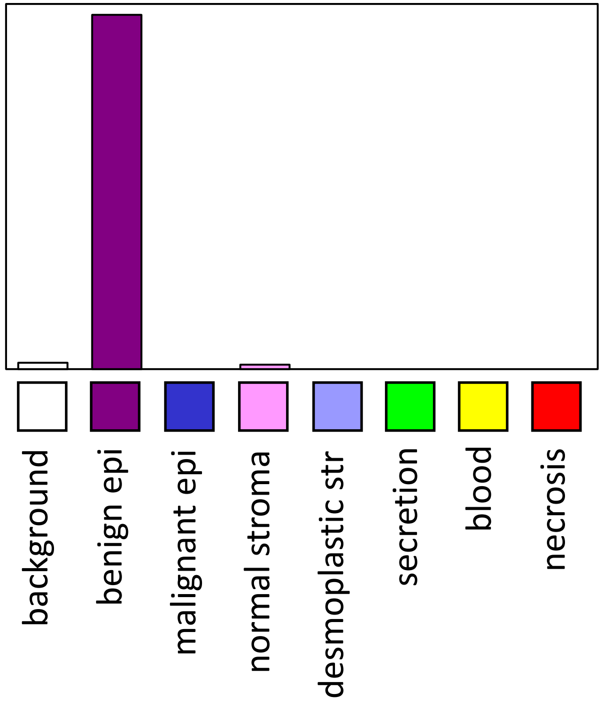


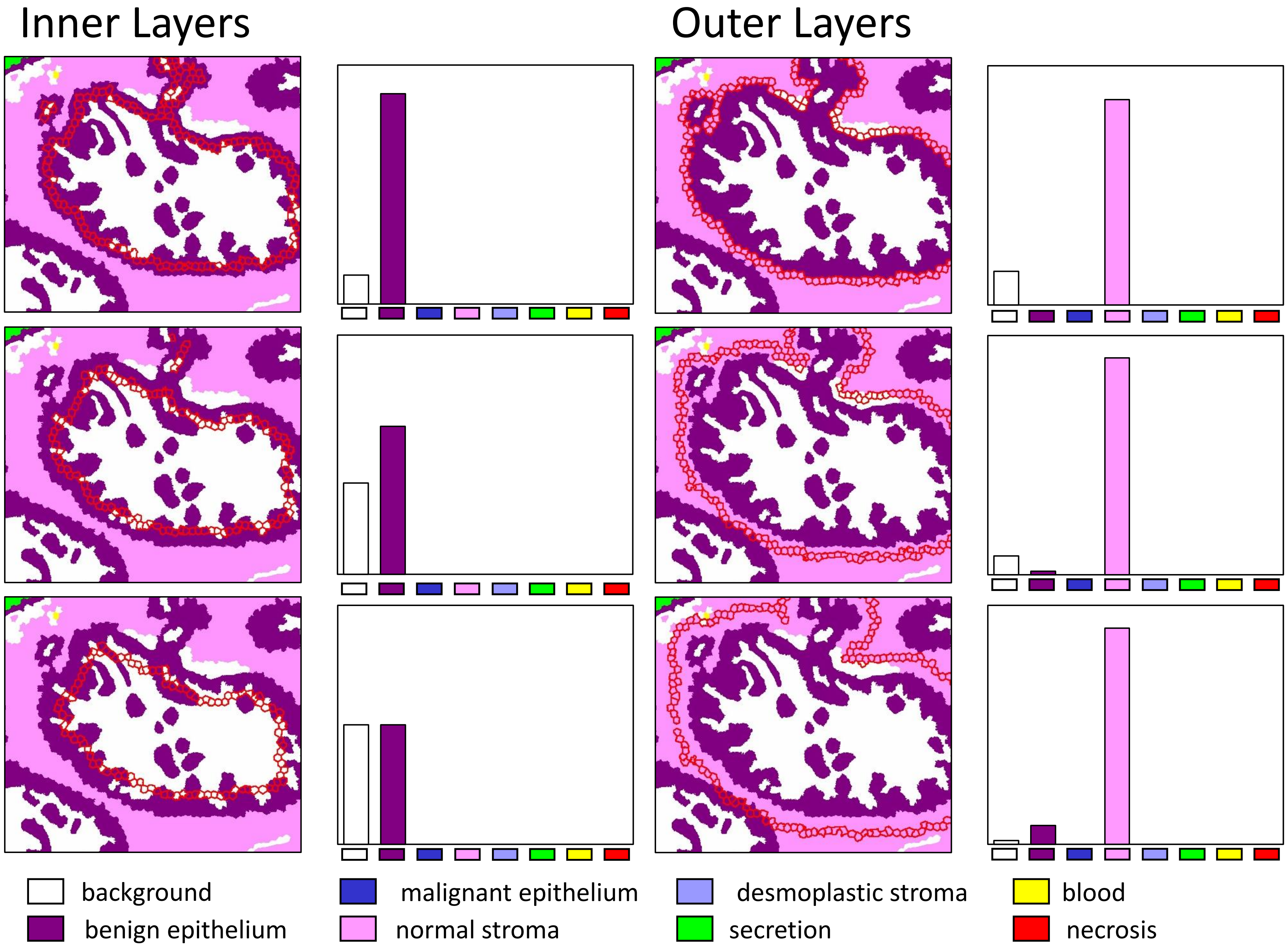
benign *atypia* *(DCIS)* *invasive*



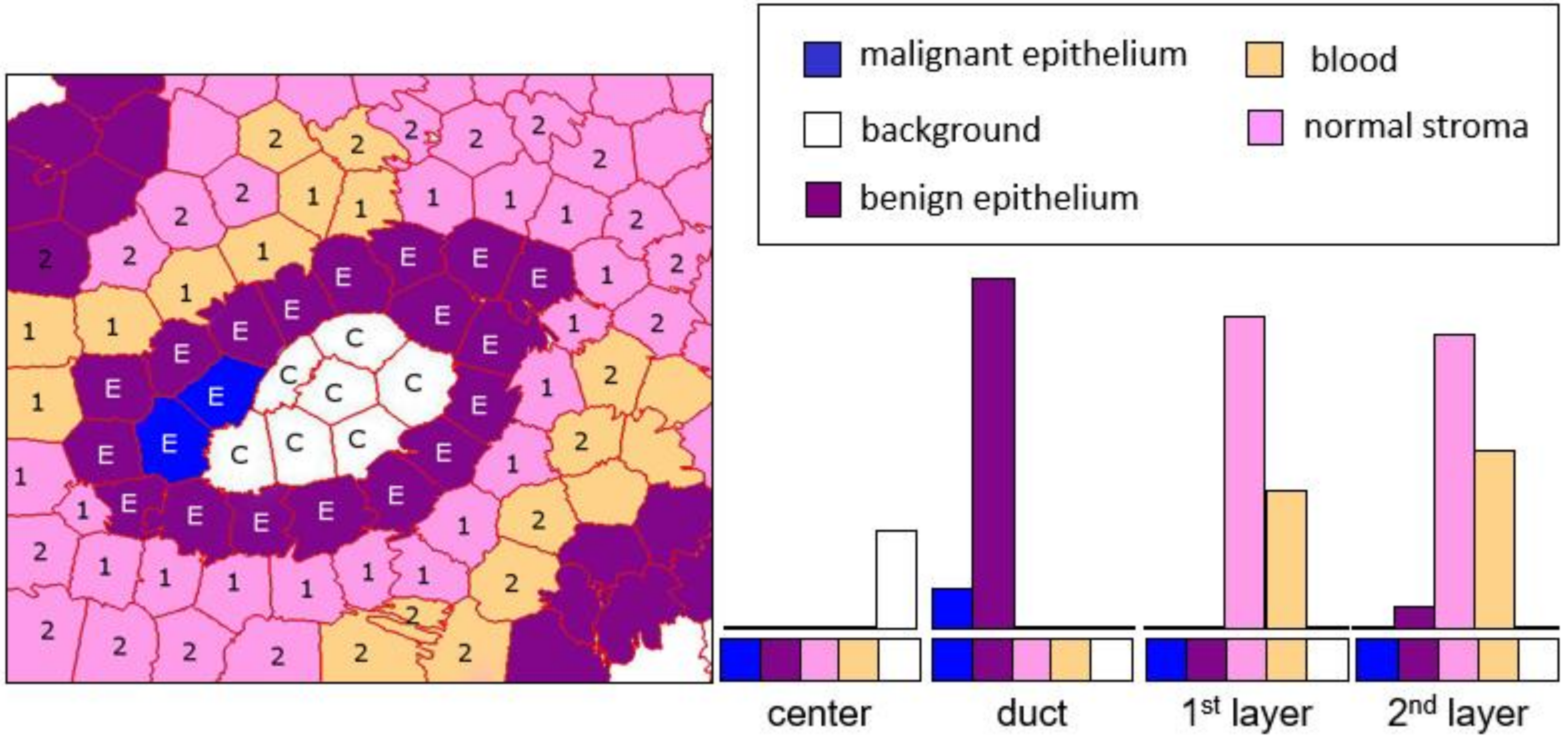


duct layer

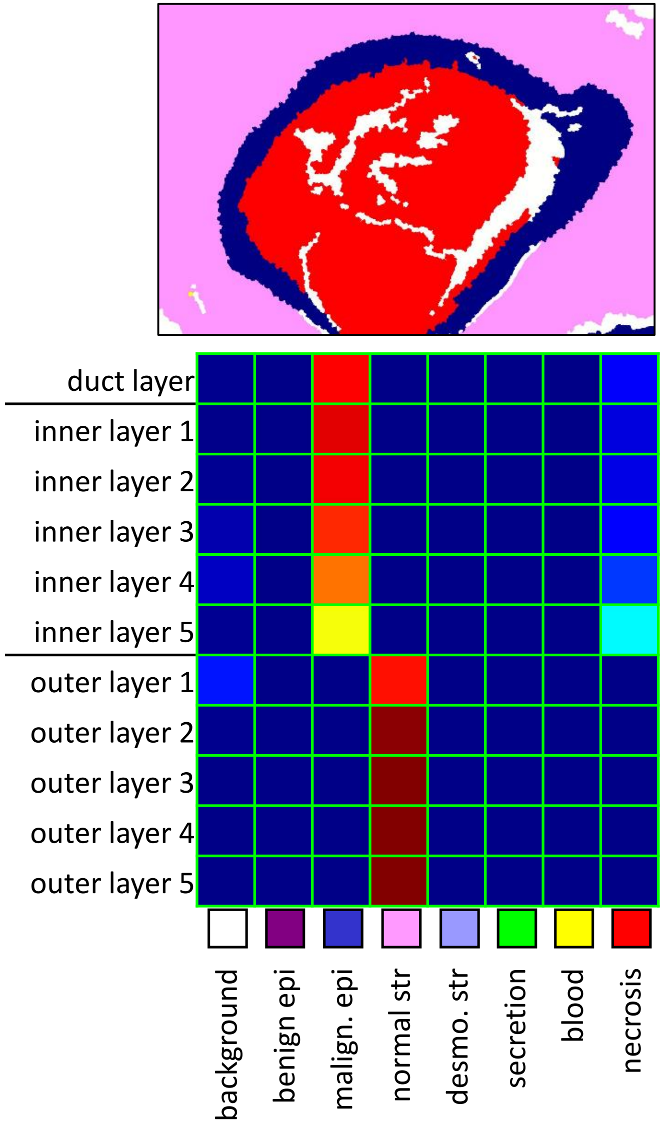


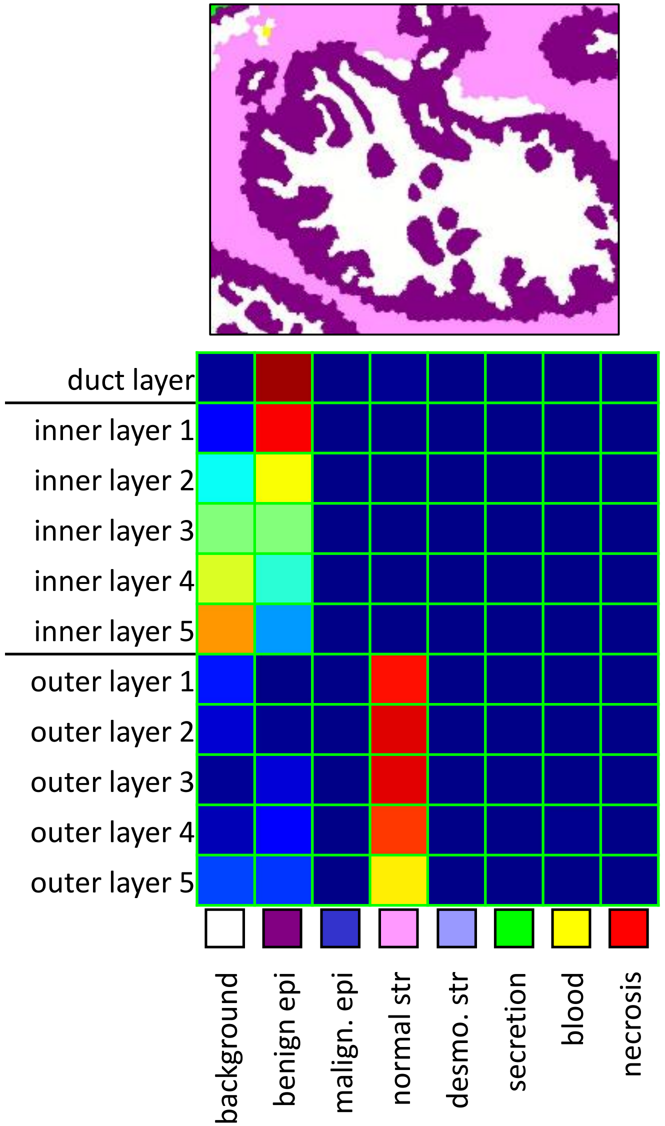

Inner Layers
Outer Layers
background
benign epithelium
malignant epithelium
normal stroma
desmoplastic stroma
secretion
blood
necrosis

Concentration of areas around and inside of ducts for breast cancer analysis



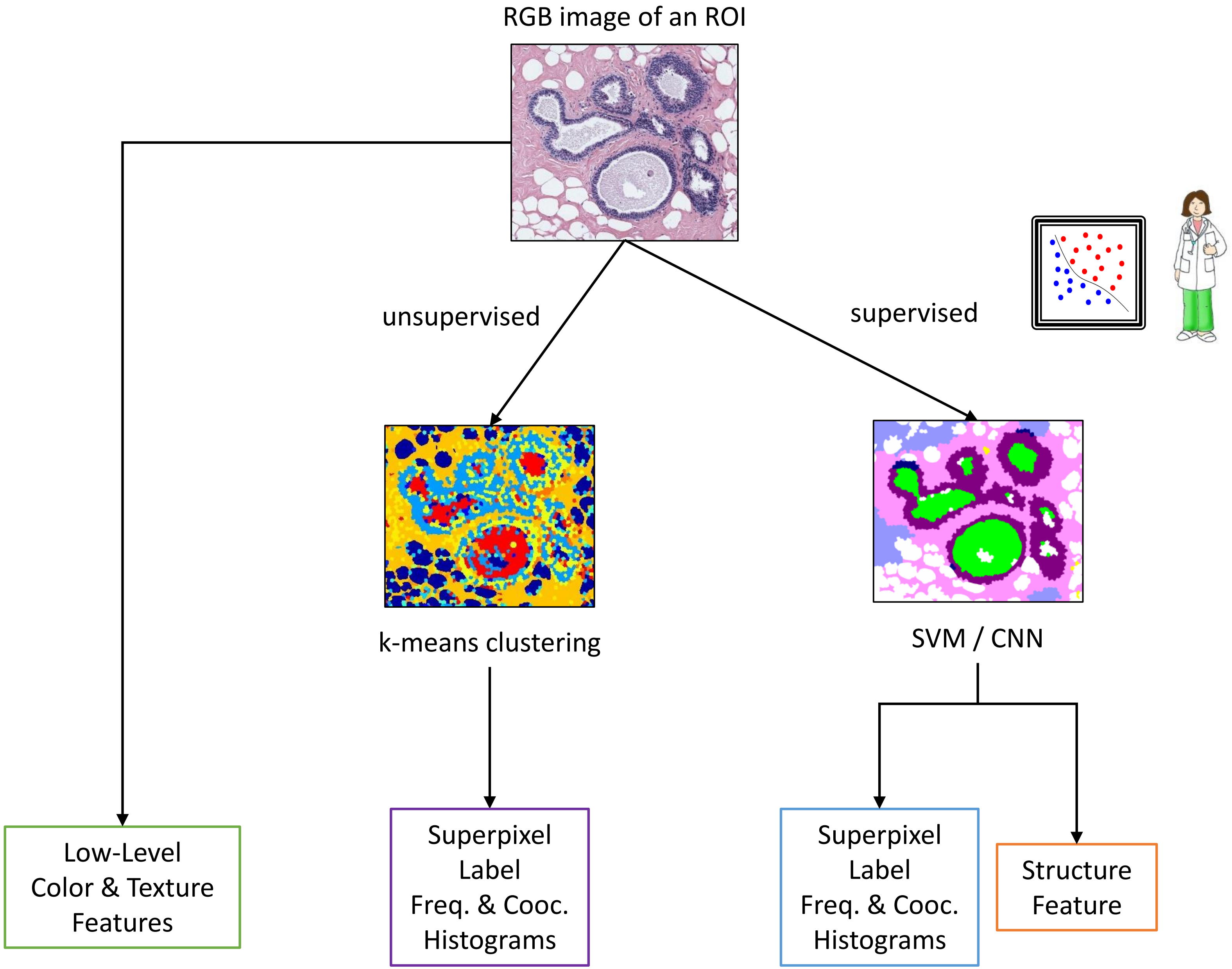
Structure Feature



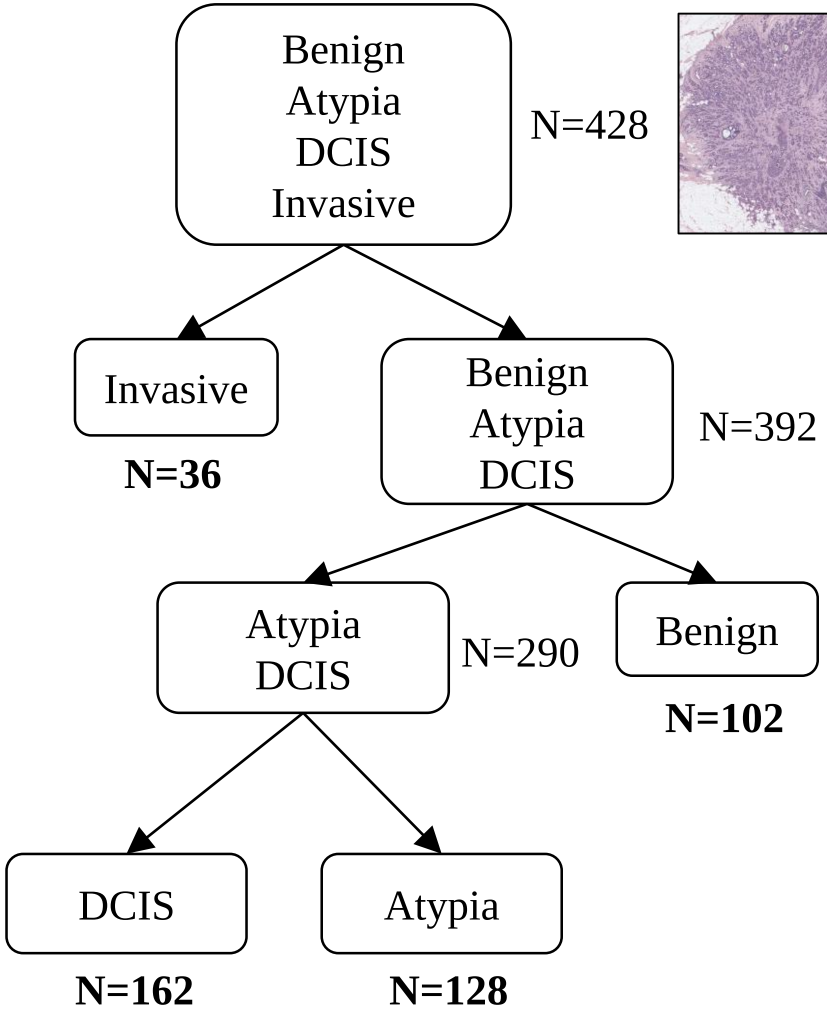


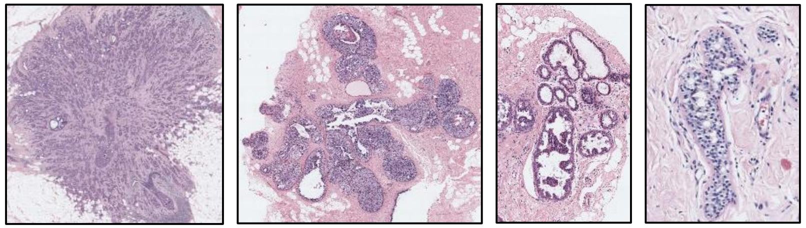
Melanoma Biopsy Analysis

- Melanoma biopsy slides are more difficult than breast biopsy slides.
- Cancerous areas can occur in multiple different areas; there is nothing equivalent to a duct to look for.
- Melanocytes may be cancerous or not depending on where they are, how many, and their size and appearance.
- The image we included in our first proposal was perhaps naïve.


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



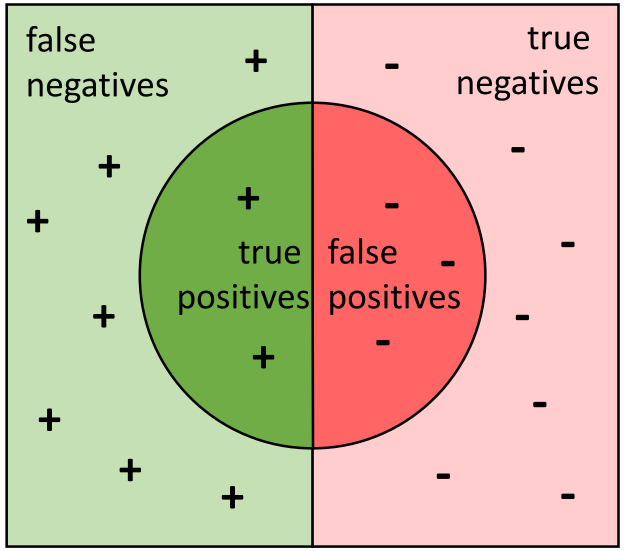


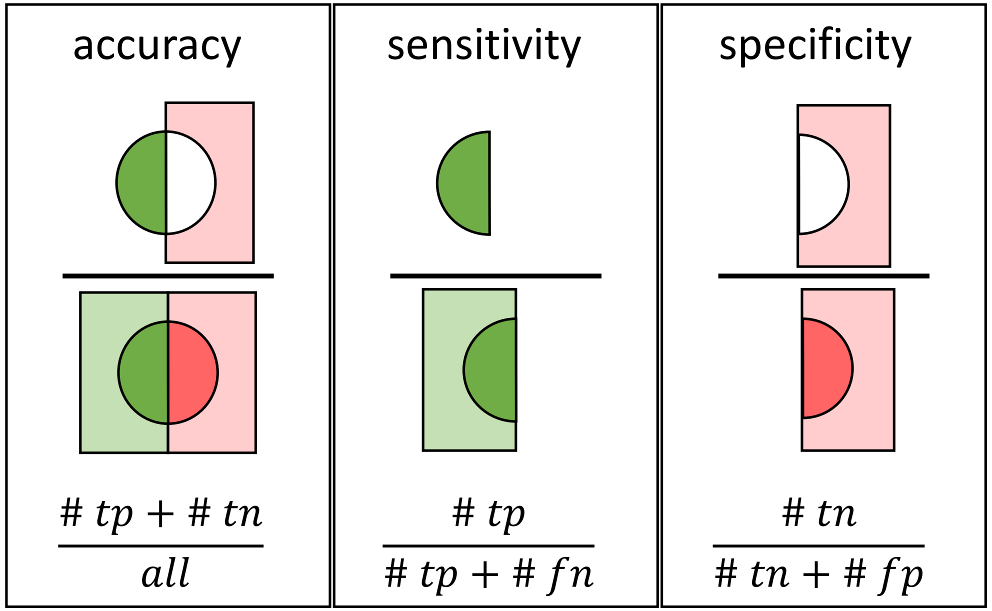
1-diagnosis-at-a-time

mimicking pathologists

Evaluation

Each ROI is a sample.

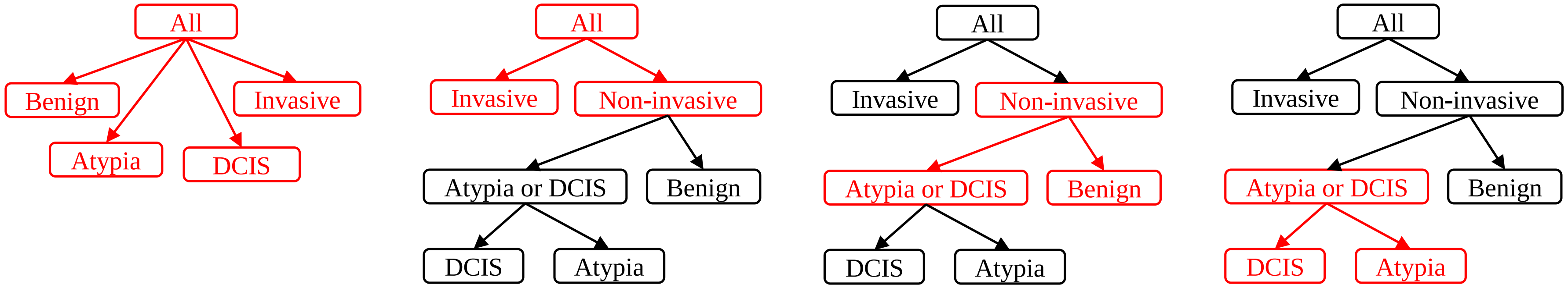




$\# tp = true\ positive$
$\# tn = true\ negative$
$\# fp = false\ positive$
$\# fn = false\ negative$

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:



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

HatNet

- We have one completed program that was developed for breast cancer, and was just published in a journal for that, but is now starting to show its first results on melanoma cancer.
- HatNet is the work of Sachin Mehta, a senior Ph.D. student in ECE, who is jointly advised by Hannenah Hajishirzi and me. He specializes in convolutional neural networks.
- HatNet is based on the concept of transformers, which comes from natural language processing.
- Sachin has extended it to medical images in a hierarchical approach.

HatNet Overview

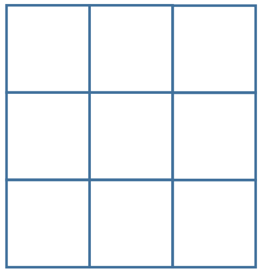
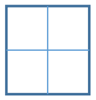


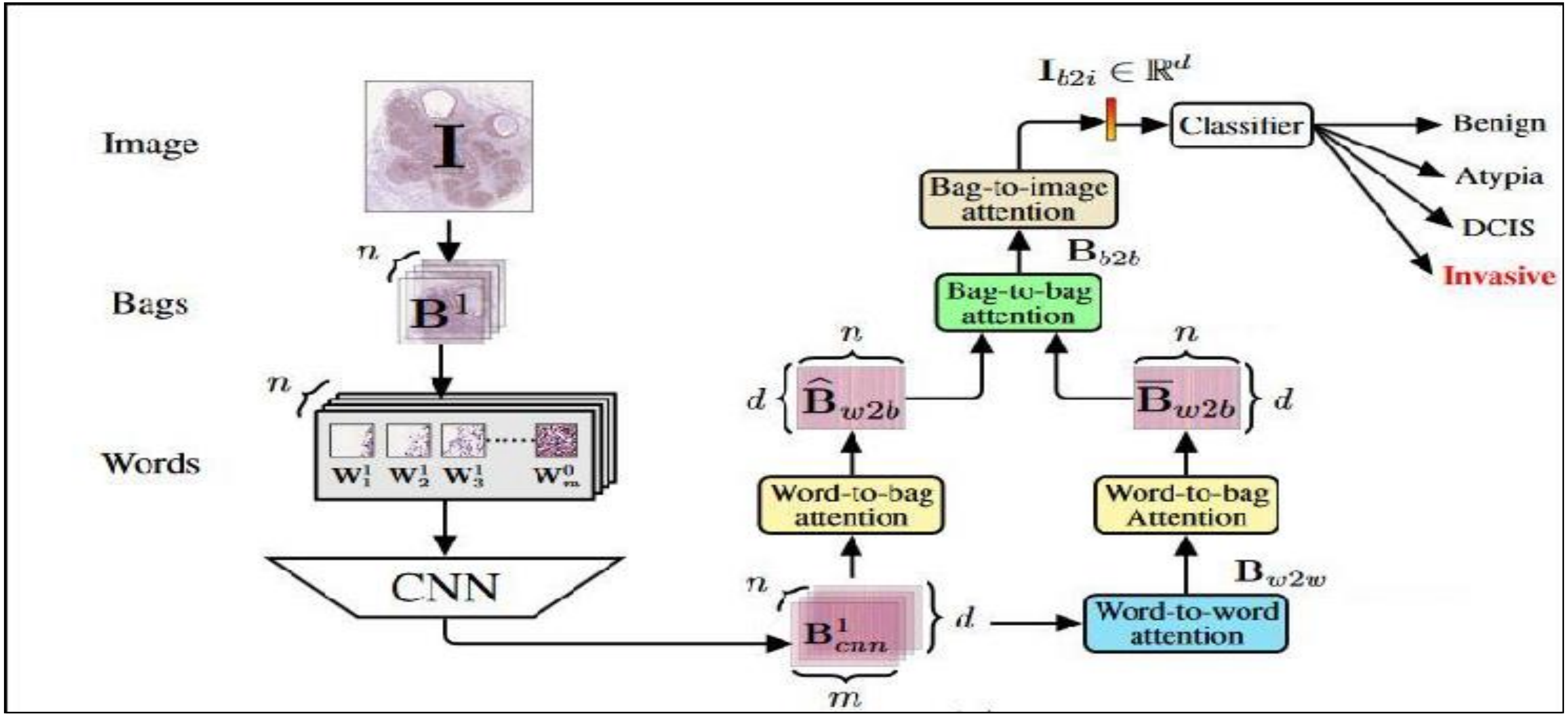
Image into bags



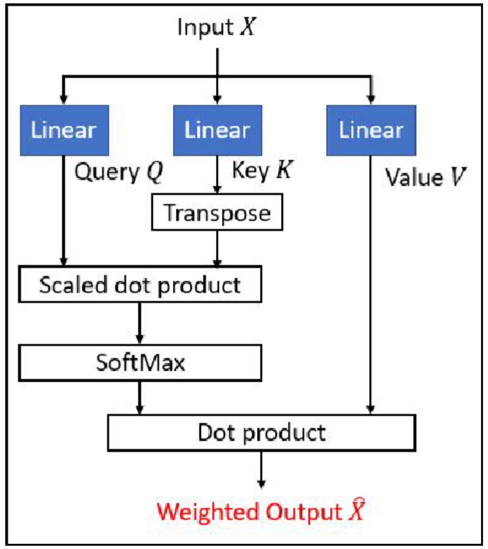
Bag into words

- The biopsy image is broken into bags and the bags into words.
- The words are input to a convolutional neural network (CNN) which converts the raw words to features.
- Next, transformers are used at multiple levels to look at word-word attention, word-bag attention, bag-bag attention, and finally bag-image attention.
- At the bag-to-image level, a simple vector of weights is produced and from it, classification is performed.

The HatNet Architecture

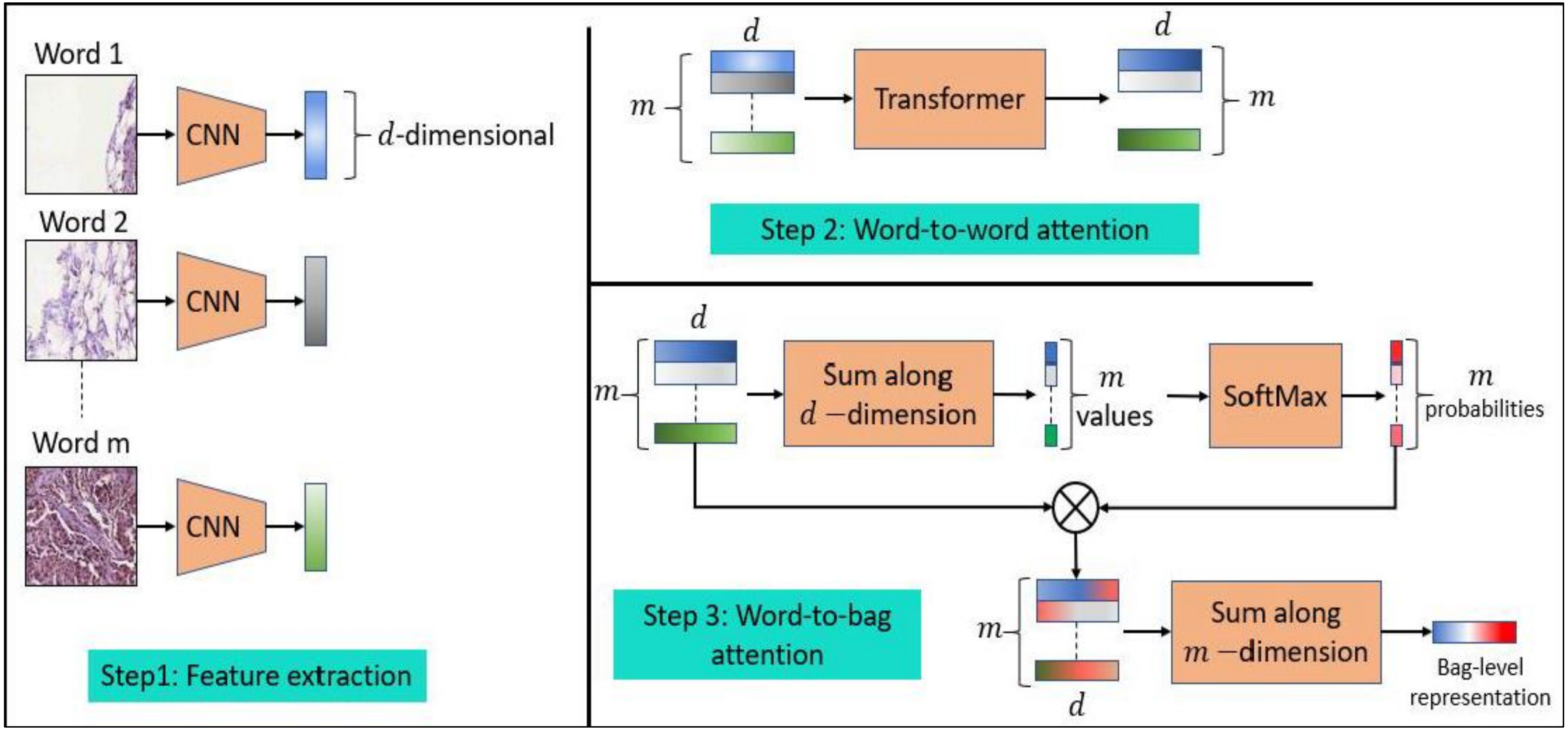


Transformer



- The transformer takes the input (words and bags, in this case) and applies three projections to obtain query (Q), key (K), and value (V) representations.
- The query and key representations are used to compute the score for each input with respect to other inputs using a dot-product (QK^T) and a softmax operation.
- The resultant scores are then combined with the value representation (V) to produce the weighted sum, which is of the same dimensionality as the input.

First three steps of HatNet



For each
bag so n
of them

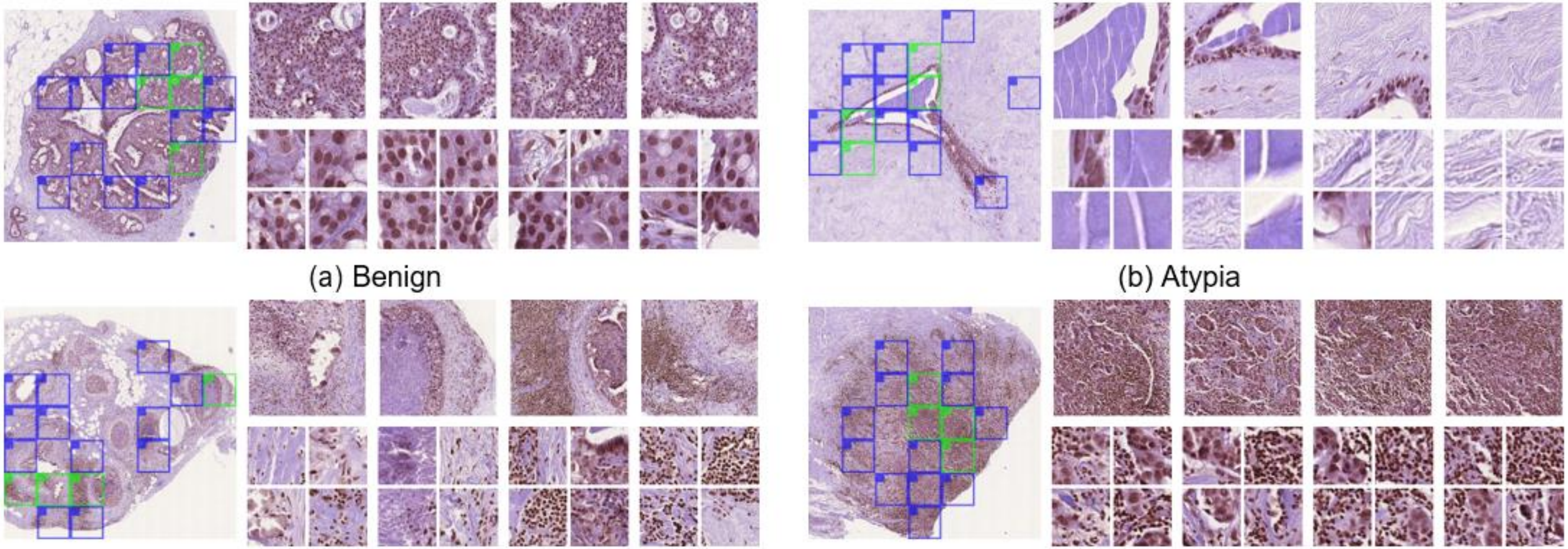
Last 3 steps

- Step 4. The Bag-Bag Attention Module, uses a transformer to assign weights to the bags.

- Step 5. The Bag-Image Attention module begins with the *n x d* attention matrix output by the Bag-Bag Module and, like the Word-Bag Module, performs projections on two different dimensions (*n* and *d*) and produces a vector of *d* weights as the image-level representation.

- Step 6. A fully-connected layer takes in this vector and outputs the diagnosis.

Examples of selected bags and words



(a) Benign (b) Atypia (c) Ductal carcinoma in-situ (DCIS) (d) Invasive

Figure 4: Example results of bags and words identified using HATNet across different diagnostic categories. HATNet aggregates information from different parts of the image and different textures. Here, each sub-figure of the breast biopsy image is shown on the left of each panel with the top-30% bags (top-4 in **green**, the rest in **blue**) identified using HATNet overlayed on image. The upper right in each panel shows the top-4 bags, and the bottom right in each panel shows the top-4 words in each bag.

Examples of HatNet identifying Stroma Tissue

Top 50% words shown in pink (stroma) and blue (other) on the right. Pathologists labeling of stroma shown at center.

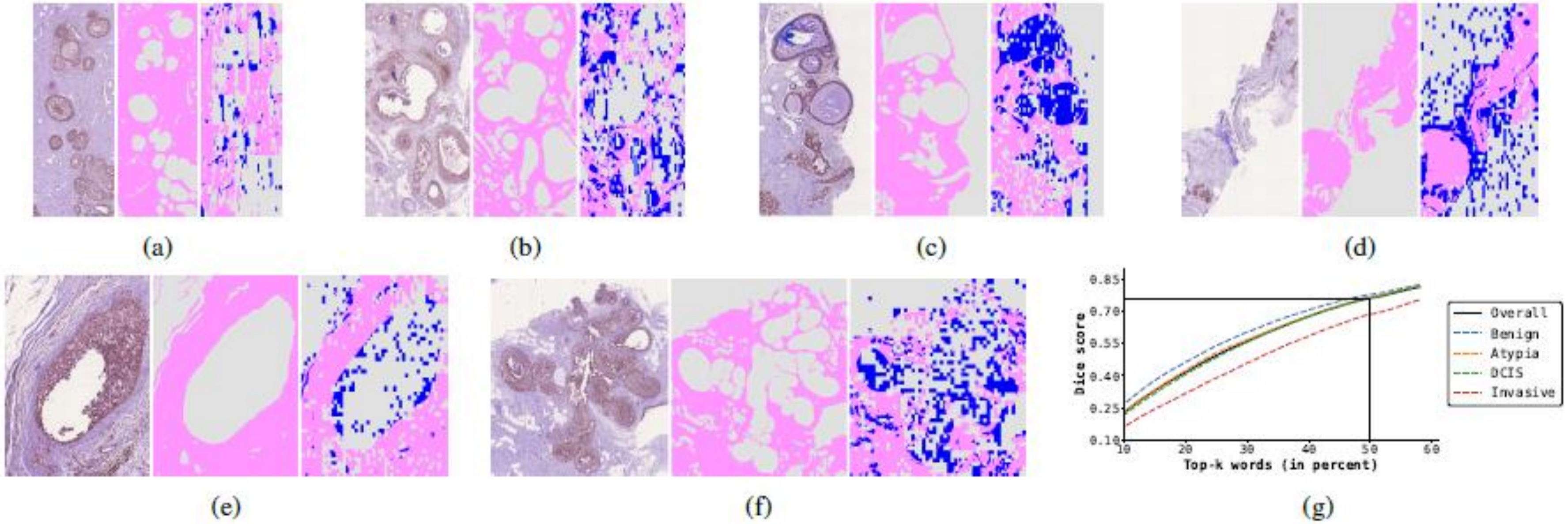


Fig. 7: HATNet identifies stroma as an important tissue. In (a-f), each sub-figure is organized from left to right as: breast biopsy image, stroma tissue labeled by pathologists, and the top-50% words (words that belong to stroma tissue are shown in pink while the remaining words are shown in blue) identified using our model. The remaining 50% words are shown in white. In (g), we plot the dice score between stromal tissue and top-k word predictions (k varies from 10% to 60%) for different diagnostic classes.

Hatnet Accuracies

- Breast Cancer—our 4 class data set, accuracy on 4-class problem

Model	Accuracy	Inference time
Y-Net	0.62	3.93 s $\pm$ 20 ms
HATNet (w/ ESPNetv2)	0.67	2.63 s $\pm$ 19 ms
HATNet (w/ MobileNetv2)	0.66	2.17 s $\pm$ 10 ms
HATNet (w/ MNASNet)	**0.70**	**2.13 s $\pm$ 12 ms**

TABLE III: **Inference time.** HATNet is fast and accurate compared to previous best model (Y-Net). Inference time is measured on a single NVIDIA GTX 1080 Ti GPU and is an average across 100 trails on the validation set.

- Melanoma Cancer – 5 classes reduced to 4 with I and II combined is up to about .53 accuracy currently.

Summary

- We have done a LOT of work on breast cancer and are just starting to get results on melanoma cancer
- Our HatNet breast cancer on the 4-class problem is at .70.
- Our HatNet melanoma on the full 4-class problem is at .53- but is only getting started.