

Cervical Cytology Image Analysis and Classification Using Graph-Based Techniques and Deep Learning

Sudhir Sornapudi, Ph.D. student¹, Gregory T. Brown, M.D., Ph.D.², Zhiyun Xue, Ph.D.², Rodney Long, M.A.²,
Lisa Allen, B.S., CT³, Sameer Antani, Ph.D.²

¹Missouri University of Science and Technology, Rolla, MO, USA; ²Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD, USA; ³Diagnostic Systems Women's Health and Cancer, Becton Dickinson and Company, Durham, NC, USA

Sudhir Sornapudi

Website: <https://sahir.github.io>

Twitter: @Sahir23



BD

S54: Oral Presentations

Translational Bioinformatics and Big Data in Cancer

#AMIA2019



Disclosure

The authors have no commercial interests to disclose.

Those with BD are already commercial, but the work wasn't commercial in nature. Included all authors and affiliations on main slide.

Learning Objectives

After participating in this session the learner should be better able to:

- Understand the importance of image preprocessing automate the data preparation.
- Understand the concepts of image registration, superpixel, region adjacency graph, image classification.

➤ Cervical Cancer

- One of the most common cancer among women.
- 2018 Worldwide Statistics ^Δ
 - 570,000 new cases.
 - 311,000 women died from cervical cancer.
- Cervical cancer that is detected early is more likely to be treated successfully.

^Δ "Human papillomavirus (HPV) and cervical cancer," World Health Organization 2019.

➤ NLM Data

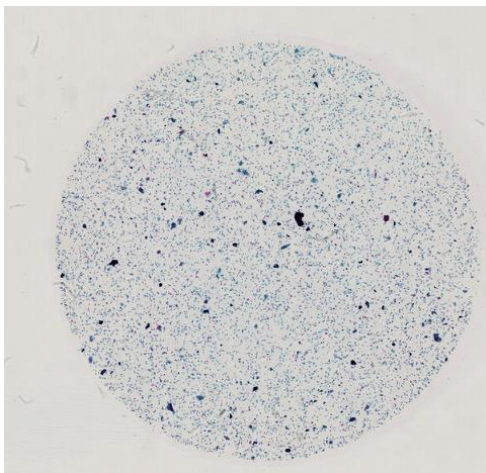
- 25 cytology slides.
- Provided by BD (Becton-Dickinson) Corporation.
- The slides are prepared through Liquid based Cytology (LBC).
 - Thin layer slide preparation technology
 - Using Sure Path technique

➤ Herlev Pap Smear Dataset

- 917 cervical cell images.
- Extracted through Conventional Pap Smear.

Dataset

Clean Slide image



Annotated Slide image

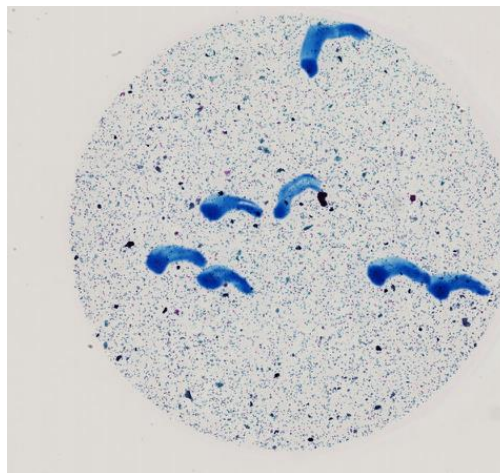


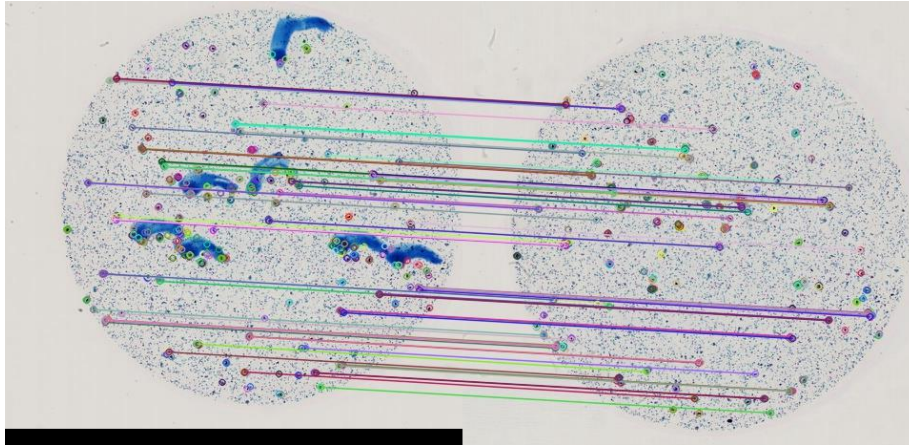
Image Properties

File Properties Macro image

Image type	Hamamatsu NDPI
File name	12XS00692.ndpi
File path	[REDACTED]
File size	342.2 MiB
Image width	61440
Image height	59136
Pixel size	0.228003 x 0.228003 um
Magnification	40X
Display color	RGB (32 bpp)
Image Centre	5.95953, 0.704433 mm
Pyramid structure	Level 0: (61440,59136) Level 1: (30720,29568) Level 2: (15360,14784) Level 3: (7680,7392) Level 4: (3840,3696) Level 5: (1920,1848) Level 6: (960,924) Level 7: (480,462) Level 8: (240,231)

Note: Displayed images are from level 7

Region of Interest (ROI) detection



Matching Key-points

➤ Image Registration

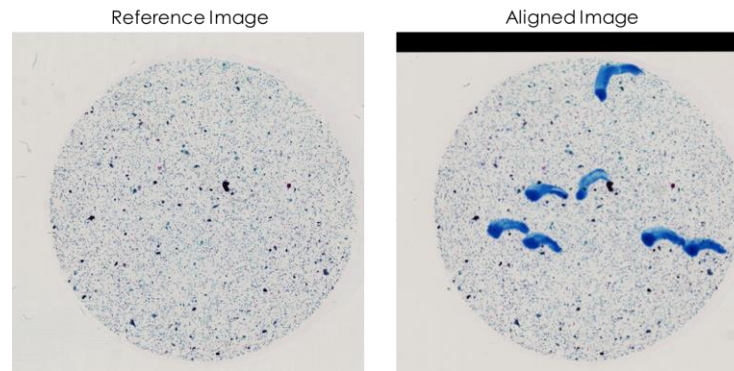
- ORB feature detector ^Δ
- Match features
- Calculate Homography
 - Uses RANSAC*
estimation technique

^Δ Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary R. Bradski: ORB: An efficient alternative to SIFT or SURF. ICCV 2011: 2564-2571

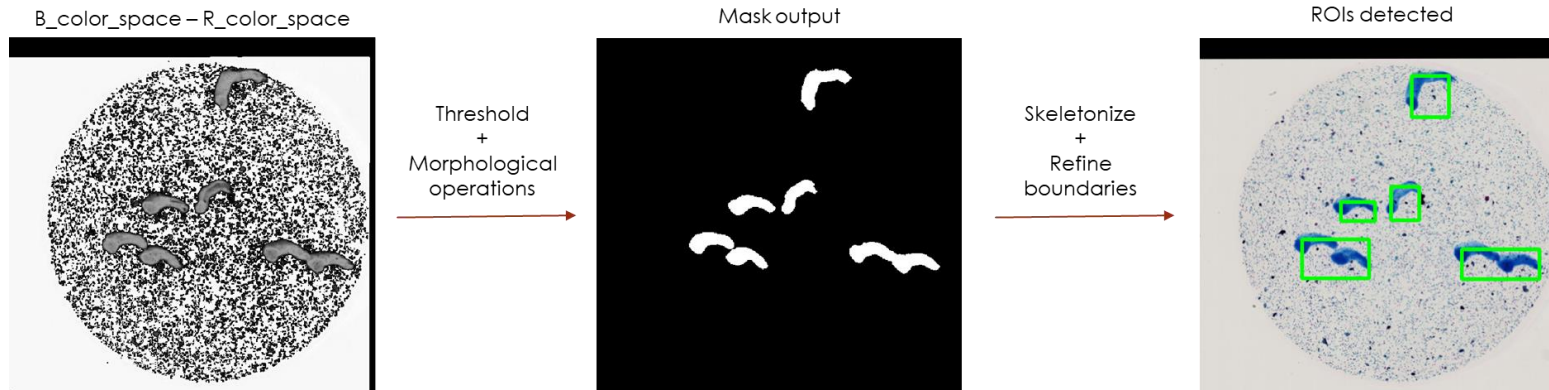
* Random sample consensus (**RANSAC**) is an iterative method to estimate parameters of a mathematical model from a set of observed data that contains outliers

Region of Interest (ROI) detection

Image Registration



Detecting marked regions



Locating Abnormal Cells

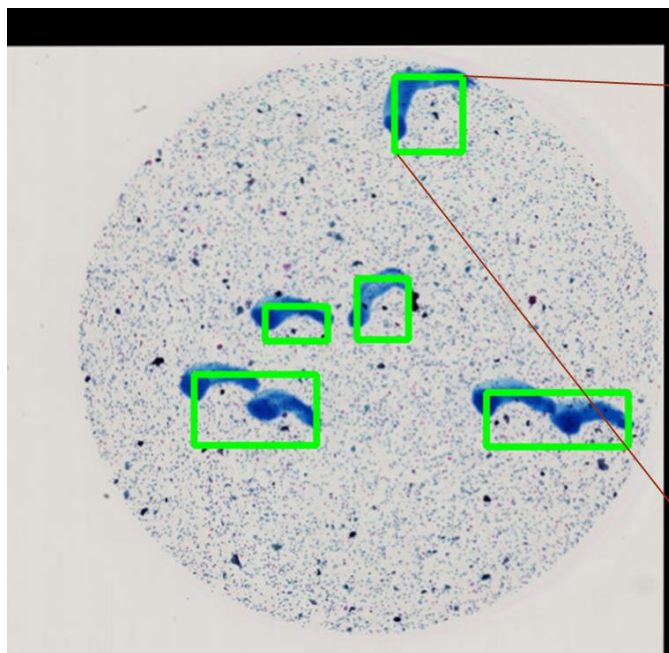


Image from level 7

Size: 3400x3079

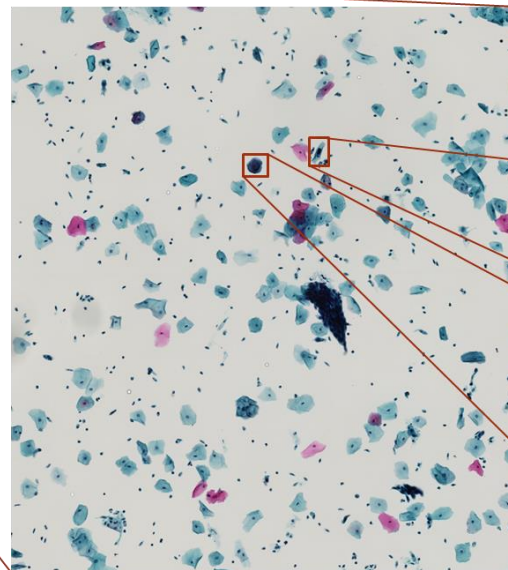
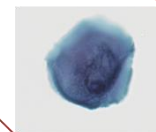
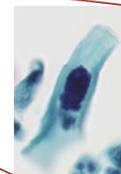


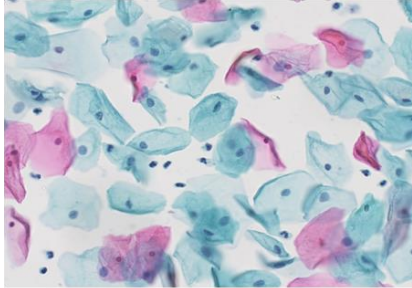
Image from level 1

Abnormal cells



Cell Classification

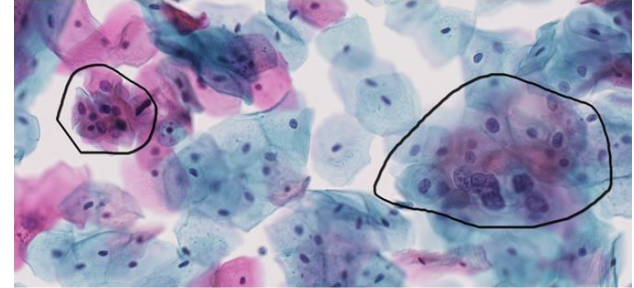
- NILM (Negative for Intraepithelial Lesion or Malignancy)



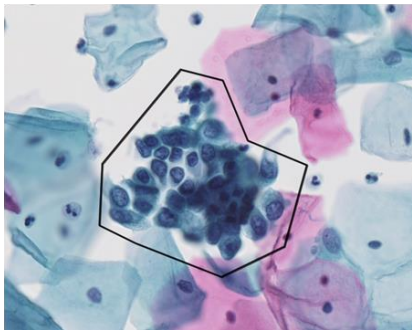
- ASCUS (Atypical squamous cells of undetermined significance)



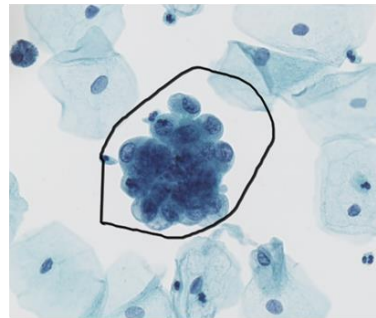
- LSIL (Lower-grade Squamous intraepithelial lesion)



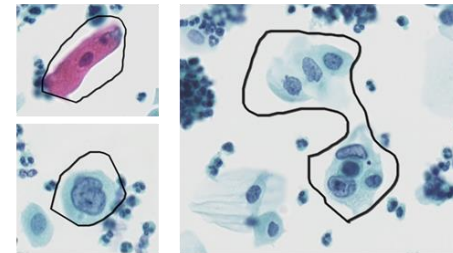
- HSIL (Higher-grade Squamous intraepithelial lesion)



- Adeno (Adenocarcinoma)



- SCC (Squamous cell carcinoma)

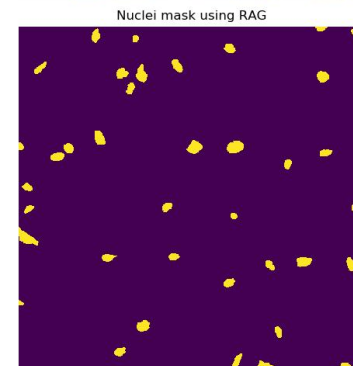
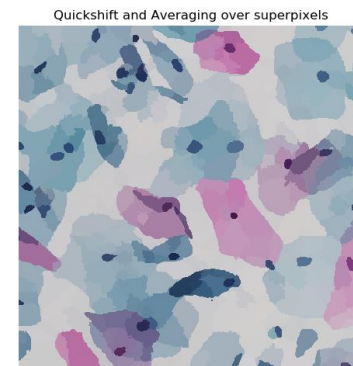
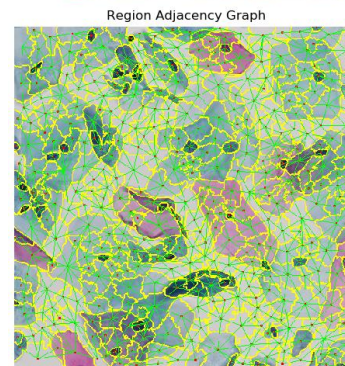
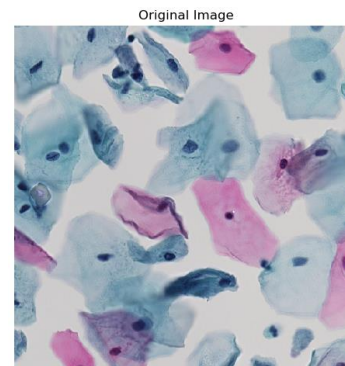


Cell Data Generation

- We employ two methods:
 - Graph based cell detection
 - Patch based data generation

Graph-based Cell Detection

- Superpixel generation using QuickShift^Δ technique
- Averaging pixel intensities over superpixel regions
- Create a Graph:
 - Node = Centroid of superpixel regions
 - Edge = Connecting adjacent nodes
 - Edge weight = $\left| \sqrt{n_1^2 - n_2^2} \right|$
- Graph Cut, with threshold = 59 (chosen empirically)
- Output: Nuclei mask

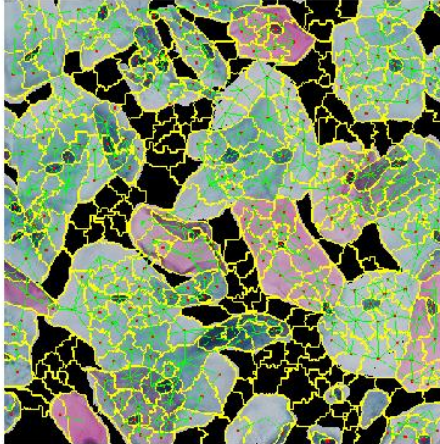


^ΔA. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In Proc. ECCV, 2008.

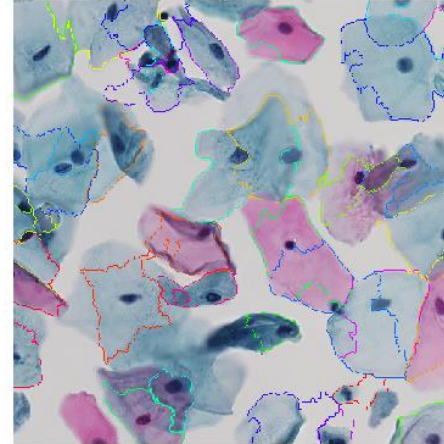
Graph-based Cell Detection

- ▶ **Cell Detection**
 - ▶ **Threshold to remove background**
 - ▶ **Join regions - nodes adjacent to nuclei node**
 - ▶ **Working on the problem to improve the cell detection**

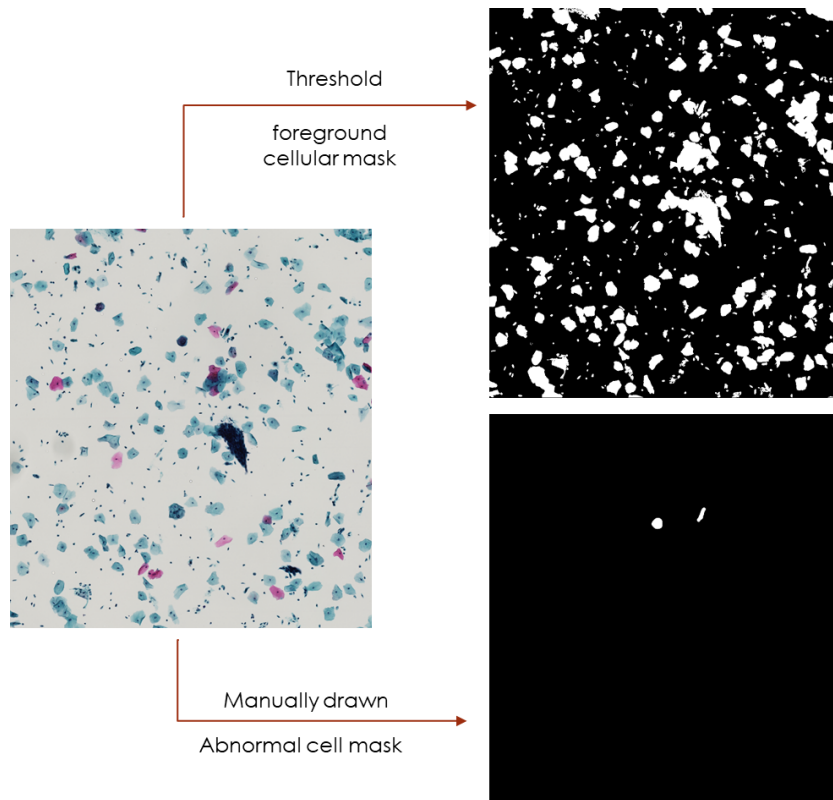
Graph structure connecting cellular regions



Cell segmentation



Patch-Based Data Generation

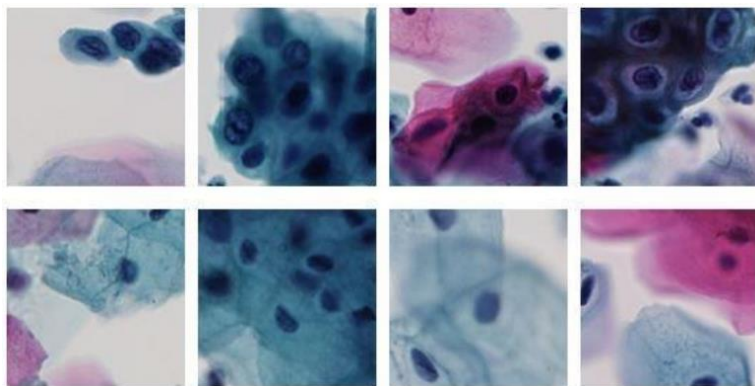


- ▶ 128x128 patch generation from sliding window technique (stride = 64)
- ▶ Patches containing more than 75% of background were discarded.
- ▶ Ground truth labels for each patch were generated w.r.t abnormal cell mask.
 - ▶ Object area > 20% Patch area → Abnormal
 - ▶ Otherwise → Normal

Cell Data for Classification

➤ Patch Based Data Generation

- 4120 multi-cell images
 - Abnormal images: 2060
 - Normal images: 2060

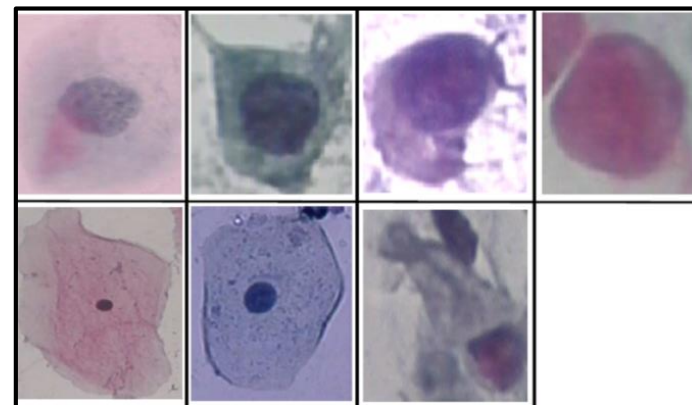


Abnormal

Normal

➤ Herlev Pap Smear Dataset

- 917 individual cell images
 - Abnormal cells: 675
 - Normal cells: 242



Classification

► CNN-based Classifier

- Fine-tuning models initialized with pre-trained ImageNet weights.
- No. of Epochs = 500
- Batch Size = 32
- Optimizer: Stochastic Gradient Descent
- Learning rate = 0.005, Momentum = 0.9
- Loss Function: Cross Entropy Loss
- Trained weights saved at epoch with best accuracy

Input for CNN Classifier		Total	Patch data	Herlev data
Training	Normal	1396	1200	196
	Abnormal	1760	1200	560
Validation	Normal	246	200	46
	Abnormal	315	200	115
Testing	Normal	660	660	-
	Abnormal	660	660	-
Total		5037	4120	917

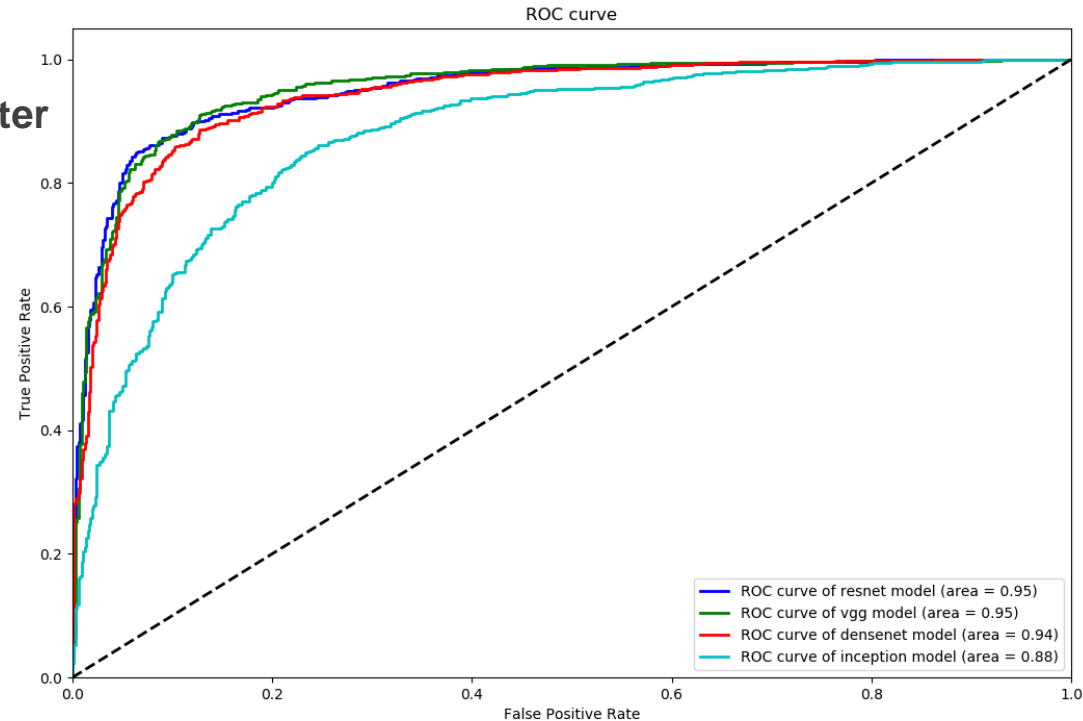
Classification Results

- PyTorch Deep Learning Platform.
- Models run on Nvidia DGX-1.
- VGG-19 was observed to give better performance

Model	Confusion matrix $\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$	Accuracy	Precision	Recall	F1-score	ROC_AUC
Resnet-50	$\begin{bmatrix} 589 & 71 \\ 78 & 582 \end{bmatrix}$	0.8871	0.8913	0.8818	0.8865	0.95
VGG-19	$\begin{bmatrix} \mathbf{581} & \mathbf{79} \\ \mathbf{68} & \mathbf{592} \end{bmatrix}$	0.8886	0.8823	0.8970	0.8896	0.95
Densenet-121	$\begin{bmatrix} 611 & 49 \\ 131 & 529 \end{bmatrix}$	0.8636	0.9152	0.8015	0.8546	0.94
Inception_v3	$\begin{bmatrix} 429 & 231 \\ 57 & 603 \end{bmatrix}$	0.7818	0.7230	0.9136	0.8072	0.88

Classification Results

- PyTorch Deep Learning Platform.
- Models run on Nvidia DGX-1.
- VGG-19 was observed to give better performance



Conclusion

- This is an opportunity to explore use of Machine Learning and Artificial Intelligence for improving efficiency in cervical cytology.
- Cell detection and classification pose great challenges due to
 - Limited availability of proper annotated data
 - Complexity due to overlapping cells
- Our method performs well as a binary classifier.
- Classification task could be extended to perform multi-class classification.
- Need to develop robust cell detection algorithm using Reinforcement learning techniques



U.S. National Library
of Medicine



Thank you!

Email me at:

ssbw5@mst.edu

