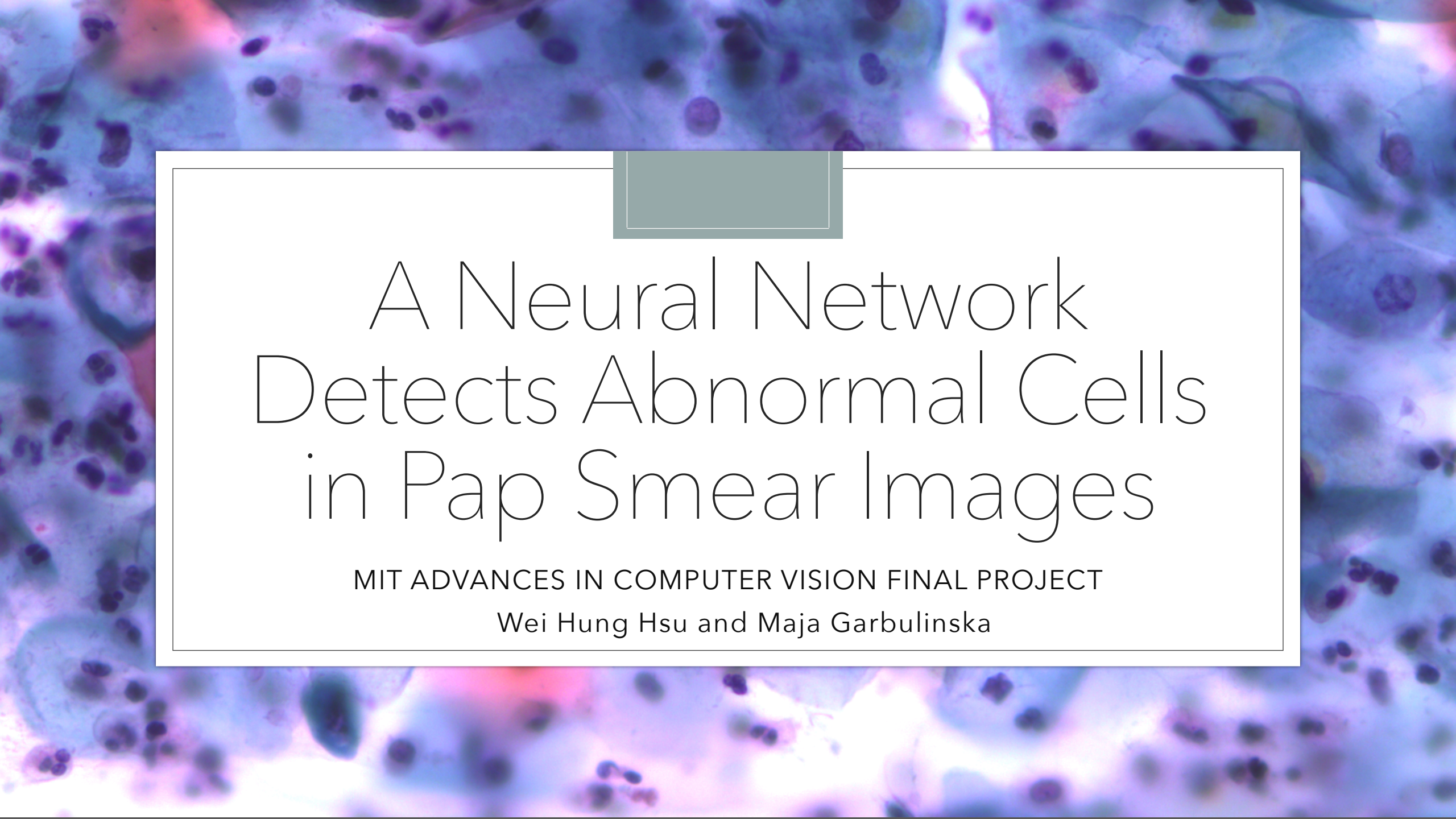
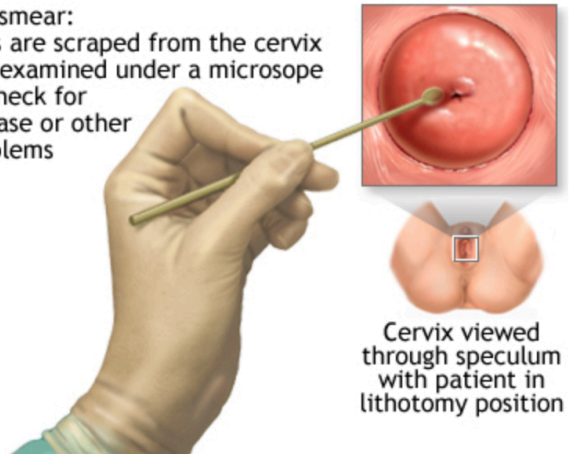
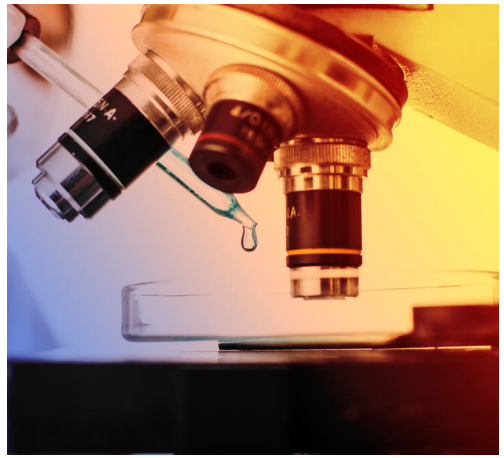


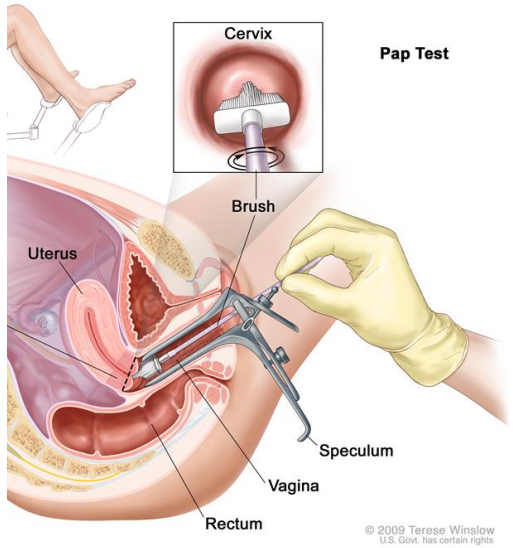
A Neural Network Detects Abnormal Cells in Pap Smear Images

MIT ADVANCES IN COMPUTER VISION FINAL PROJECT

Wei Hung Hsu and Maja Garbulinska

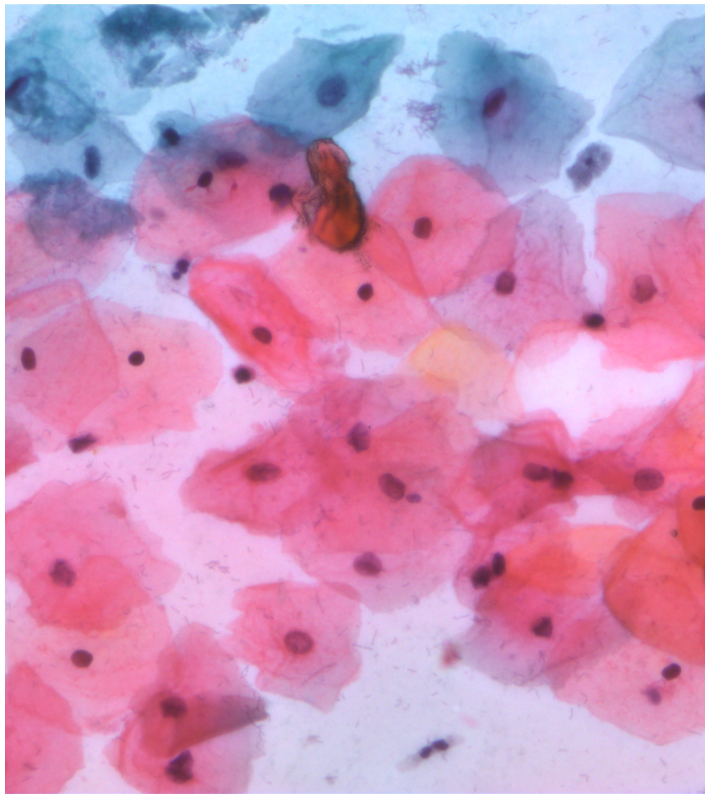


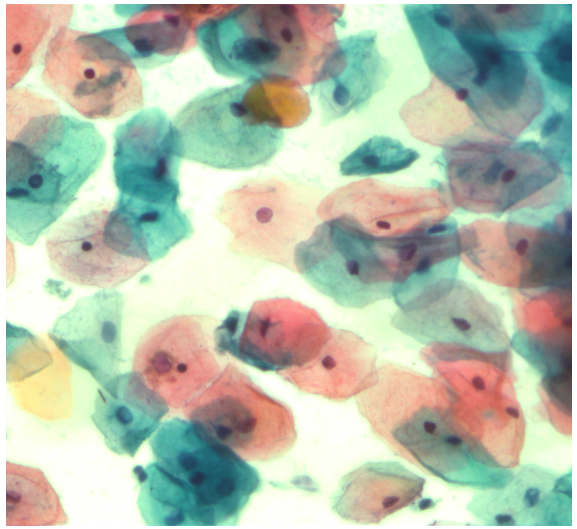


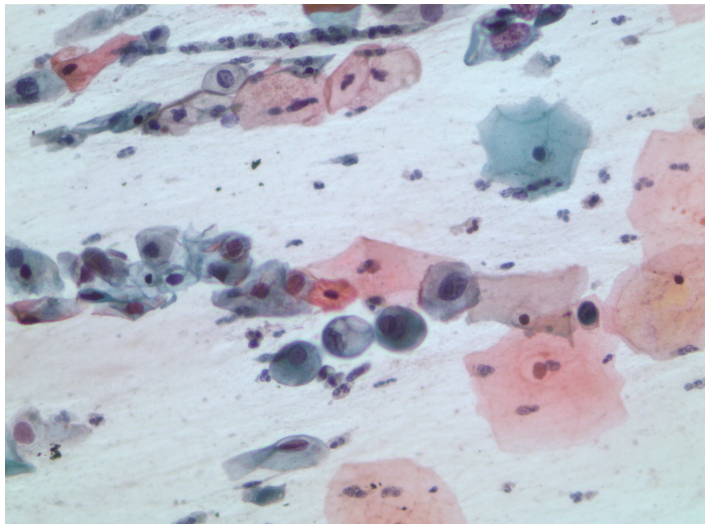


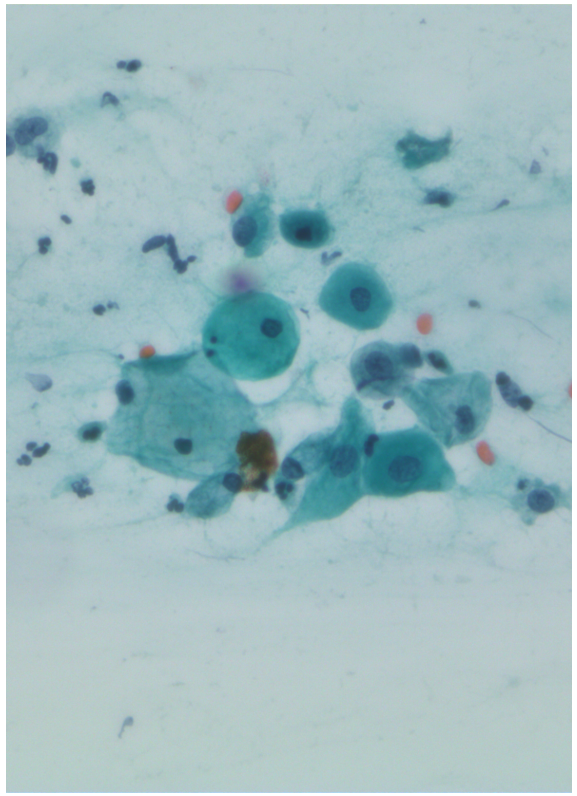
Background

- Pap smear is a test designed to detect early changes in the cervix that are either already cancer or could develop into cancer in the future
- Cervical cancer is the third most common cancer in women worldwide
- Screening and early detection leads to almost 100% survival rates
- Thousands of women die from cervical cancer worldwide, especially in developing countries, because of lack of screening
- Currently tissue samples are sent to a pathology lab and evaluated manually by a pathologist or a specialized technician
- It is a very tedious task and even the best laboratories can fail to detect between 10% and 30% of abnormal pap smear results









SIPAKMED Dataset

- A new data set consisting of 966 RBG images classified into 5 categories.

- **Normal cells**

1. Superficial- Intermediate Cells (126 images, 813 cells)
2. Parabasal Cells (108 images, 787 cells)

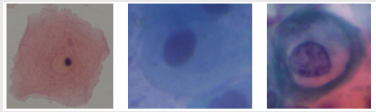
- **Abnormal cells**

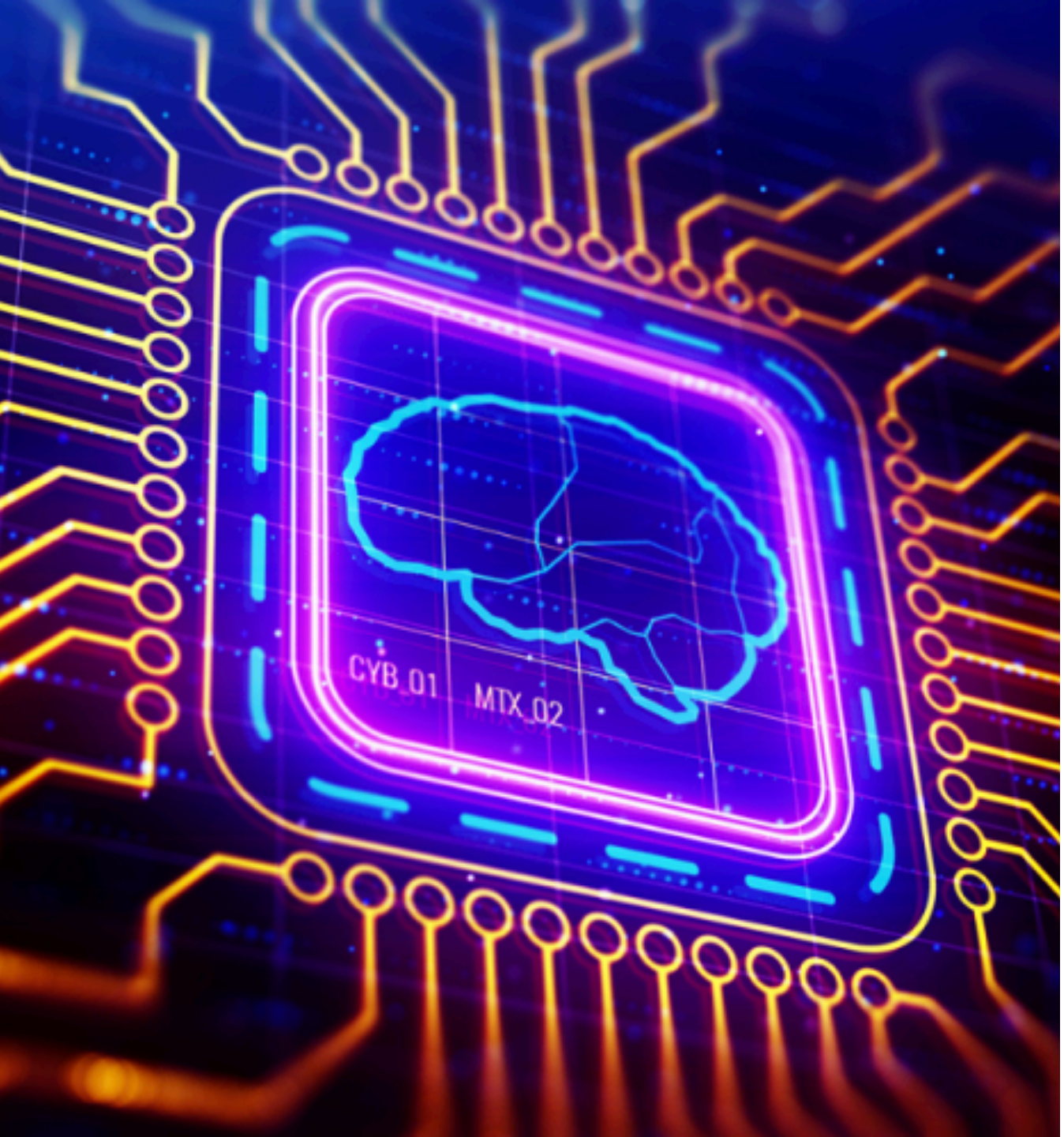
3. Koilocytotic Cells (238 images, 825 cells)
4. Dyskeratotic Cells (171 images, 793 cells)

- **Benign cells**

5. Metaplastic Cells (233 images, 813 cells)

- Total cell count: 4049. Total pap smear image: 966
- Single cells were cropped out from images and classified too.

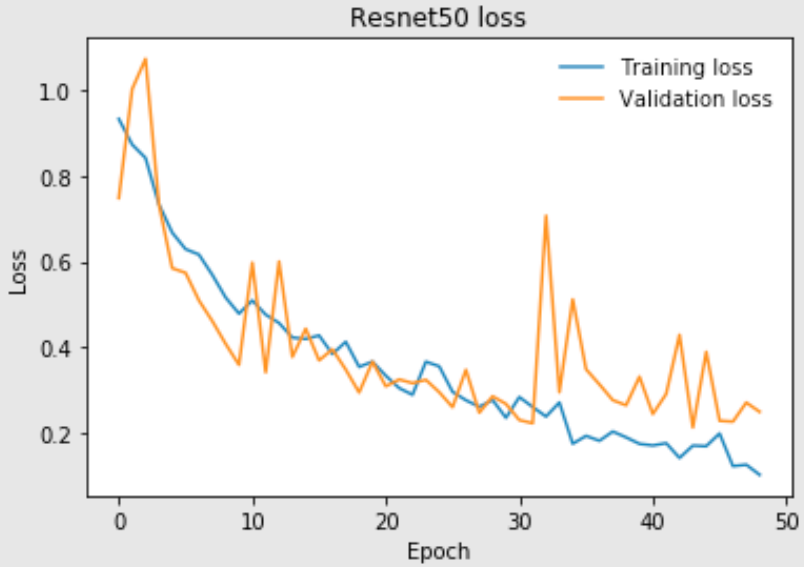


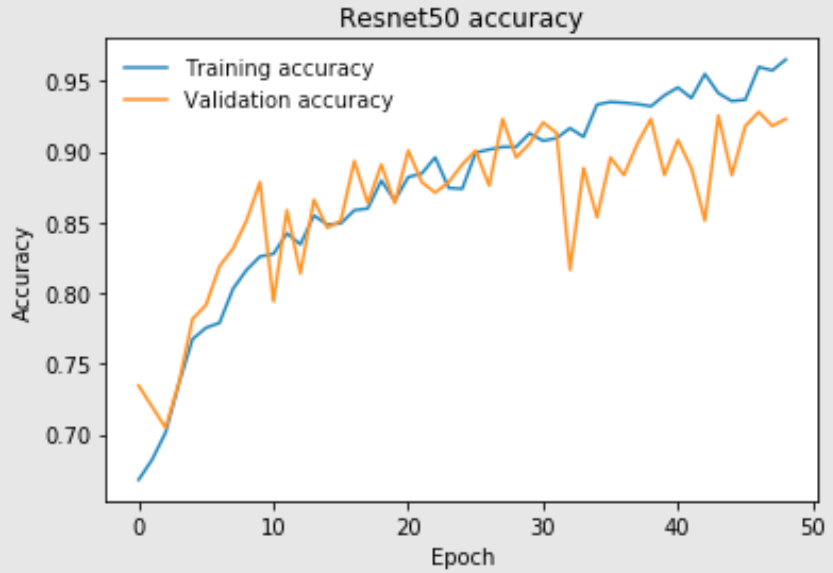


Methods

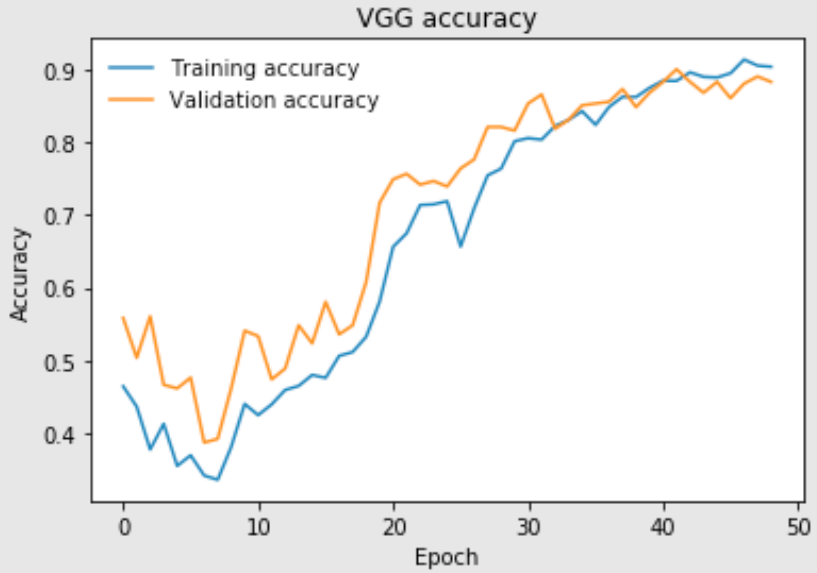
- CNNs: Resnet18, Resnet50, Resnet101, VGG
- Resnets: dropout of 50% at the final connected layer.
- Adaptive pooling layer
- Data augmentation – rotated images.
- Adjust learning rate – Annealing / Scheduling
- Adam optimizer

Results


Resnet50 loss
Training loss
Validation loss
Loss
Epoch

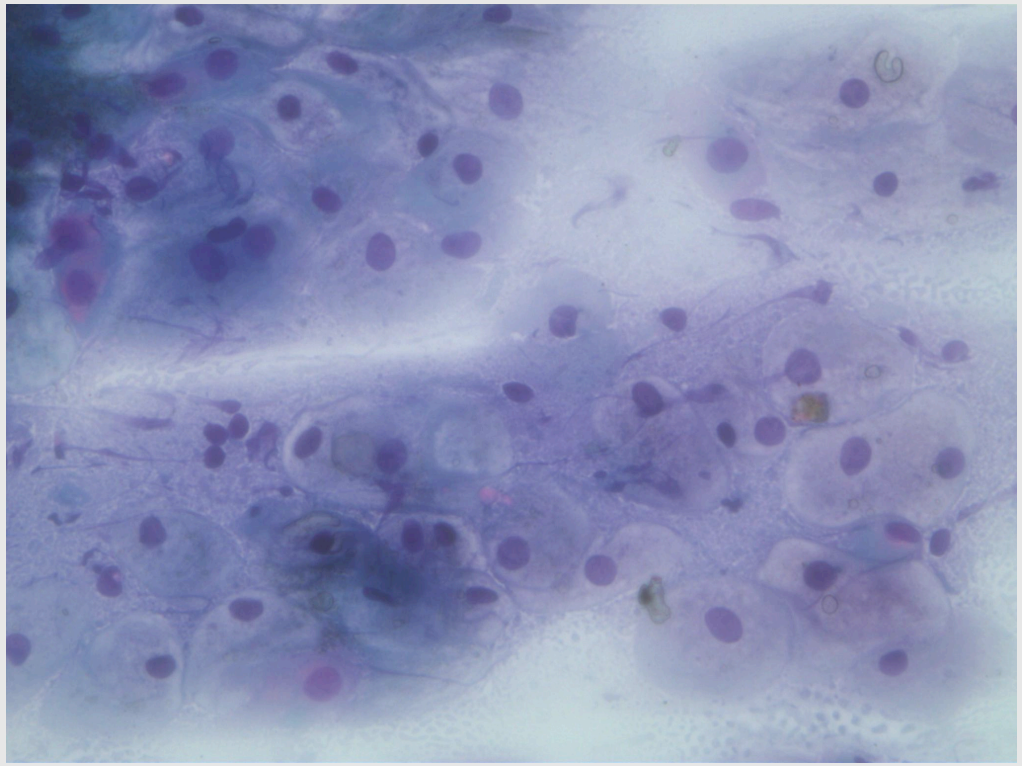

Resnet50 accuracy
Training accuracy
Validation accuracy
Accuracy
Epoch

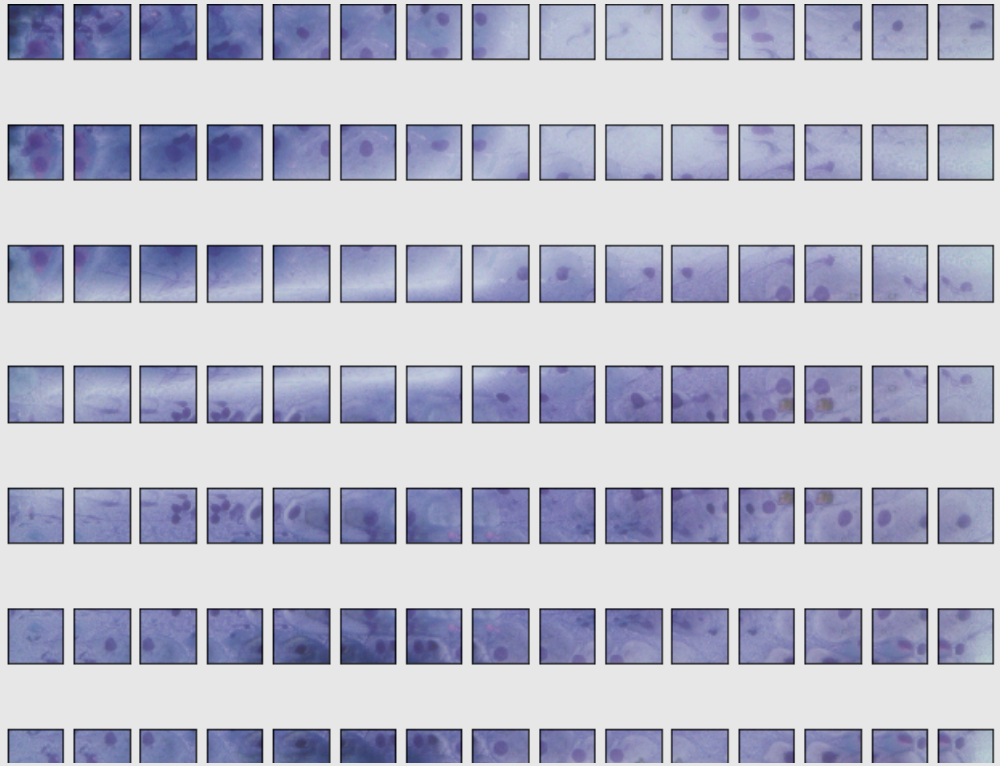

VGG loss
Training loss
Validation loss
Loss
Epoch


VGG accuracy
Training accuracy
Validation accuracy
Accuracy
Epoch

Results - Confusion Matrix for Classification

Detect Abnormal Cells on a Full Image





Preliminary Results

Imagine:

- 1000 pap smears
- Assume 10% are abnormal – 100 are abnormal and 900 normal
- Our model picks up 97 abnormal and misses 3 abnormal results
- We classify 594 normal images are “review by a specialist required”

BUT We can find 306 normal images that do not require further evaluation by a specialist – cost savings

Major Limitations

- Full images don’t have labeled cells (need a pathologist)
- By cropping the image same-size tiles, we might divide a cell in half that could lead to wrong classification
- This approach also does not account for the fact that some cells might have different sizes

THANK YOU