COMS 4995 Applied Deep Learning Final Project Report

Team Members: Jingyan Xu (jx2424), Chelsea Cui (ac4788)

Background

Problem

CAMELYON16 Challenge

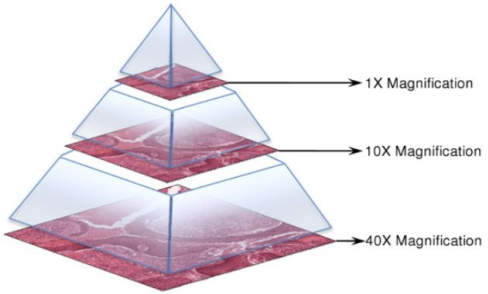
In this challenge, we will focus on the detection of micro- and macro-metastases in lymph node digitized images. This subject is highly relevant; lymph node metastases occur in most cancer types (e.g. breast, prostate, colon). Metastatic involvement of lymph nodes is one of the most important prognostic variables in breast cancer.

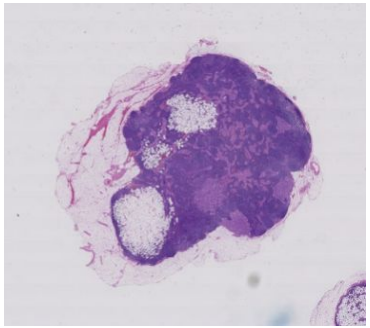
The diagnostic procedure for pathologists is, however, tedious and time-consuming and prone to misinterpretation.

Data

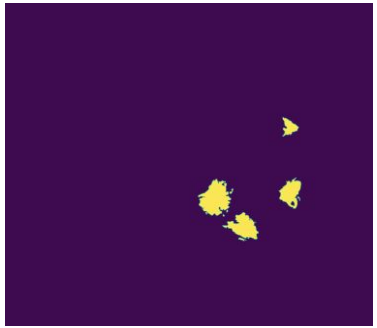
Source: Data from CAMELYON16, which was collected from Radboud UMC and UMC Utrecht, which include slides and corresponding tumor masks

Tool: OpenSlide, a library that provides a simple interface to read whole-slide images

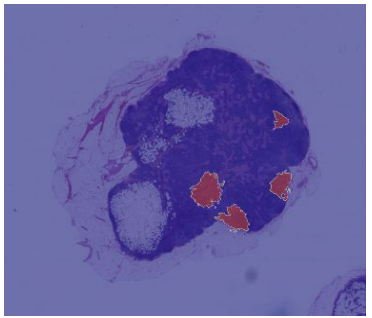




Slide



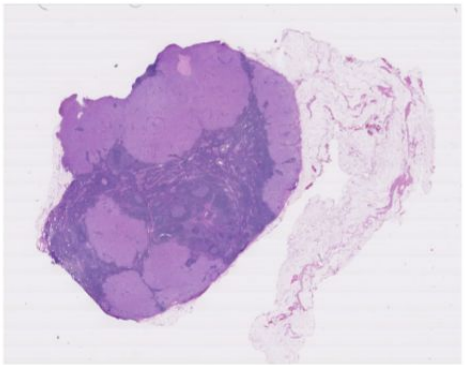
Mask



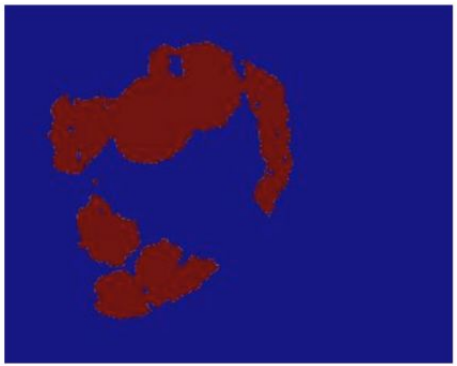
Masked Slide

Goal

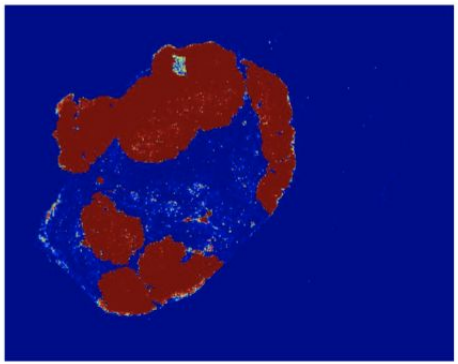
Given a collection of training data, develop a model that outputs a heatmap showing regions of a biopsy image likely to contain cancer.



Biopsy image



Ground truth (from pathologist)



Model predictions

*Image from course slide

Solution

Approach

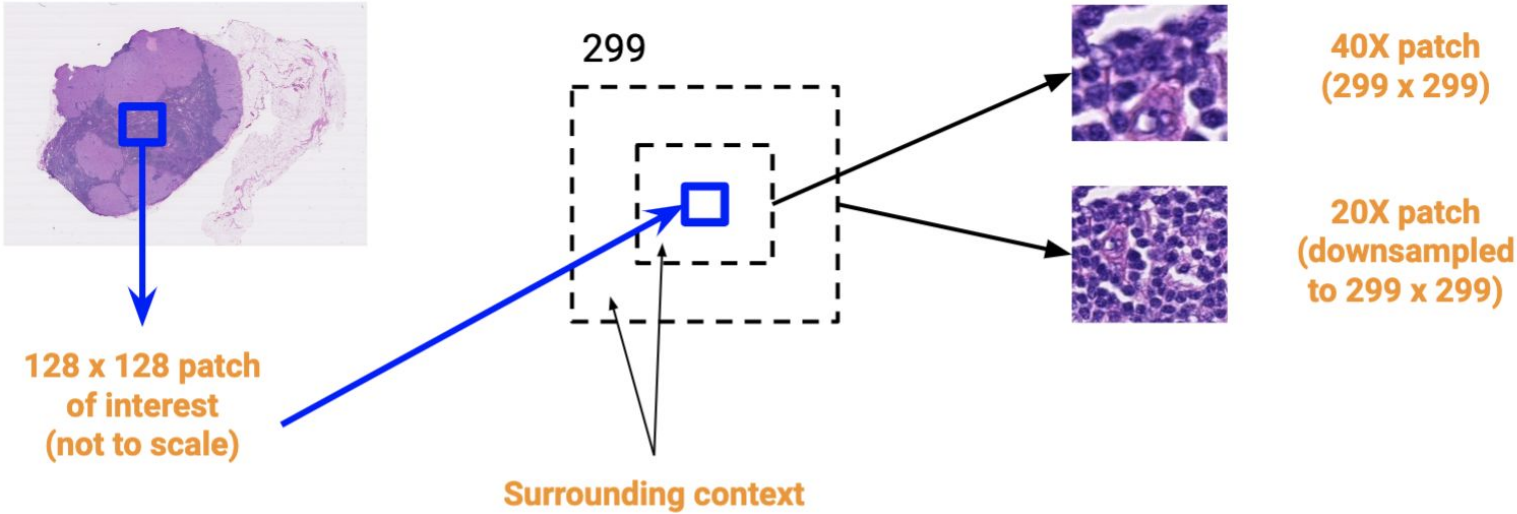
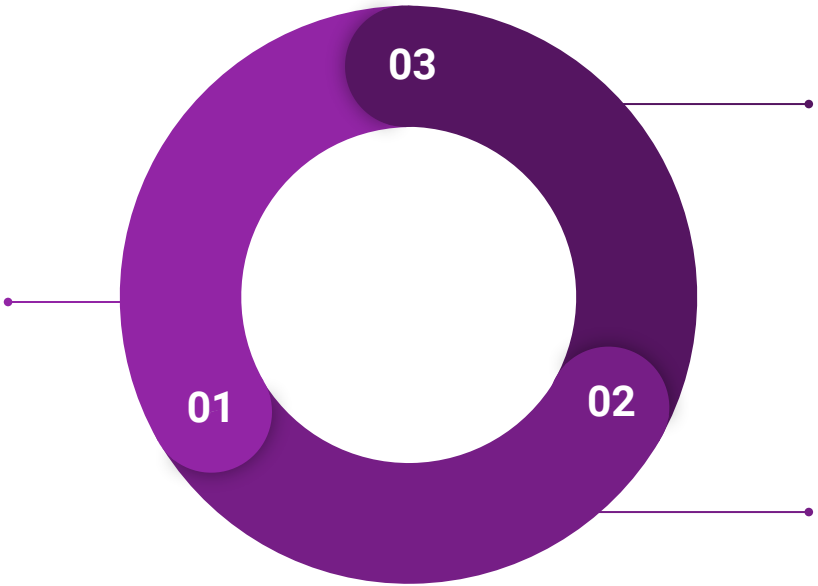


*Image from course slide

* Approach from paper Detecting Cancer Metastases on Gigapixel Pathology Images, 17

Process



Evaluation and Heatmap

Train model with batch gradient descent and visualize results

Transfer Learning and Fine-Tuning

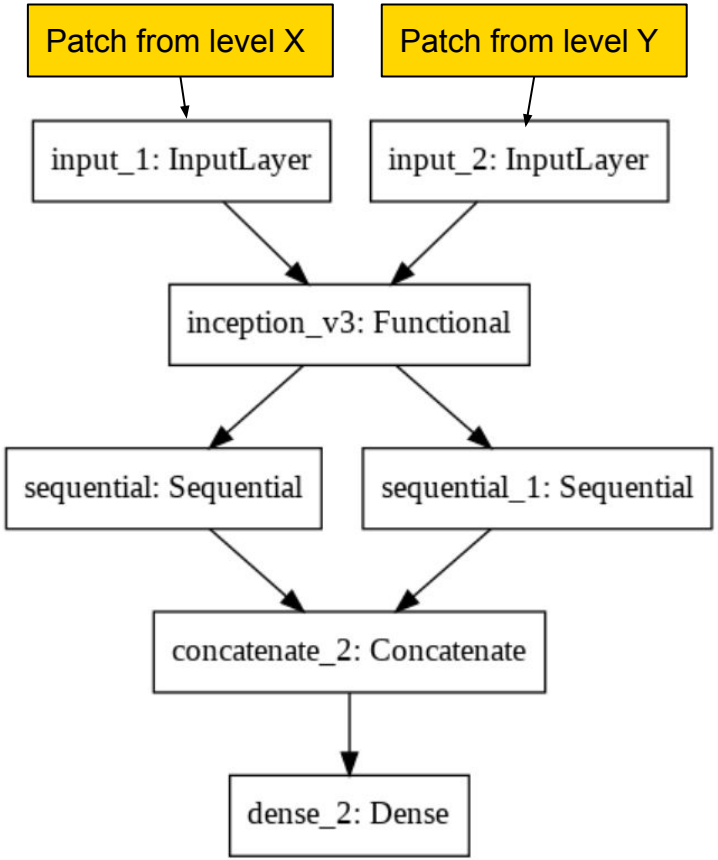
Transfer learning with image nets like VGG and Inception V3.

Data Preprocessing

Generate patches and handle imbalanced data.

Data Preprocessing

1. Use mask from the lowest magnification level to decide patch label from higher magnification levels.
2. Use size 128 * 128 as the unit patch size for prediction. This 128 * 128 patch label is the same as the 1*1 pixel from the mask of the lowest magnification level.
3. To avoid imbalanced data, at each step sample randomly for a positive or negative patch with equal probability
4. Take two magnification levels x and y for each patch
5. Apply mask on areas where gray intensity level is less than 0.8 to denoise the input



Model Structure

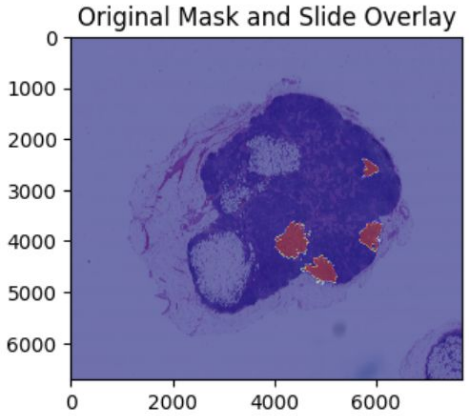
Experiments

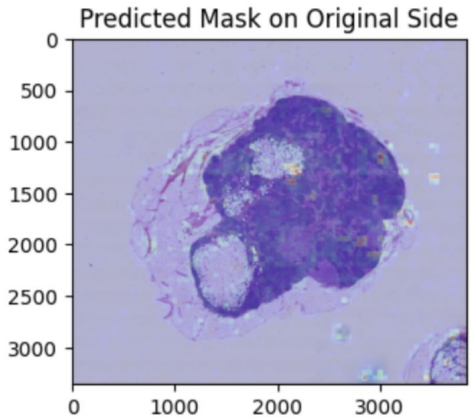
- Image Net: Inception V3, VGG16
- Magnification Levels
- Heatmap window size

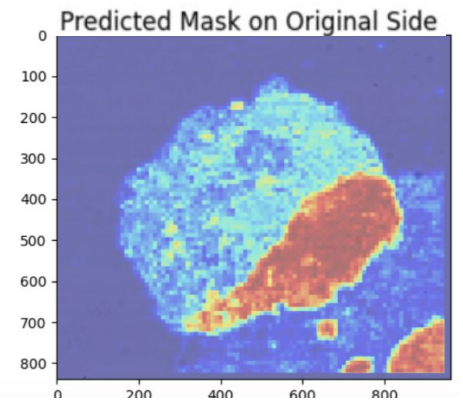
Experiments

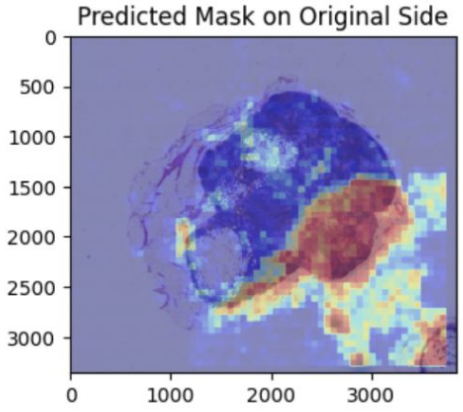
Image Net	Zoom Levels	Time/Space Cost	Performance
Inception V3	(0,1)	Very Large	Good
Inception V3	(2,3)	Large	Good
Inception V3	(2,5)	Moderate	Bad
Inception V3	(4,5)	Moderate	Ok
VGG	(4,5)	Moderate	Ok

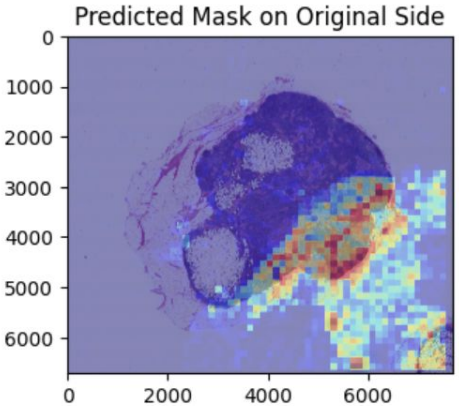
Experiment Visualizations

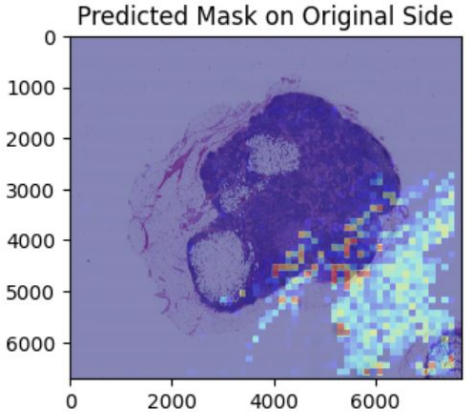








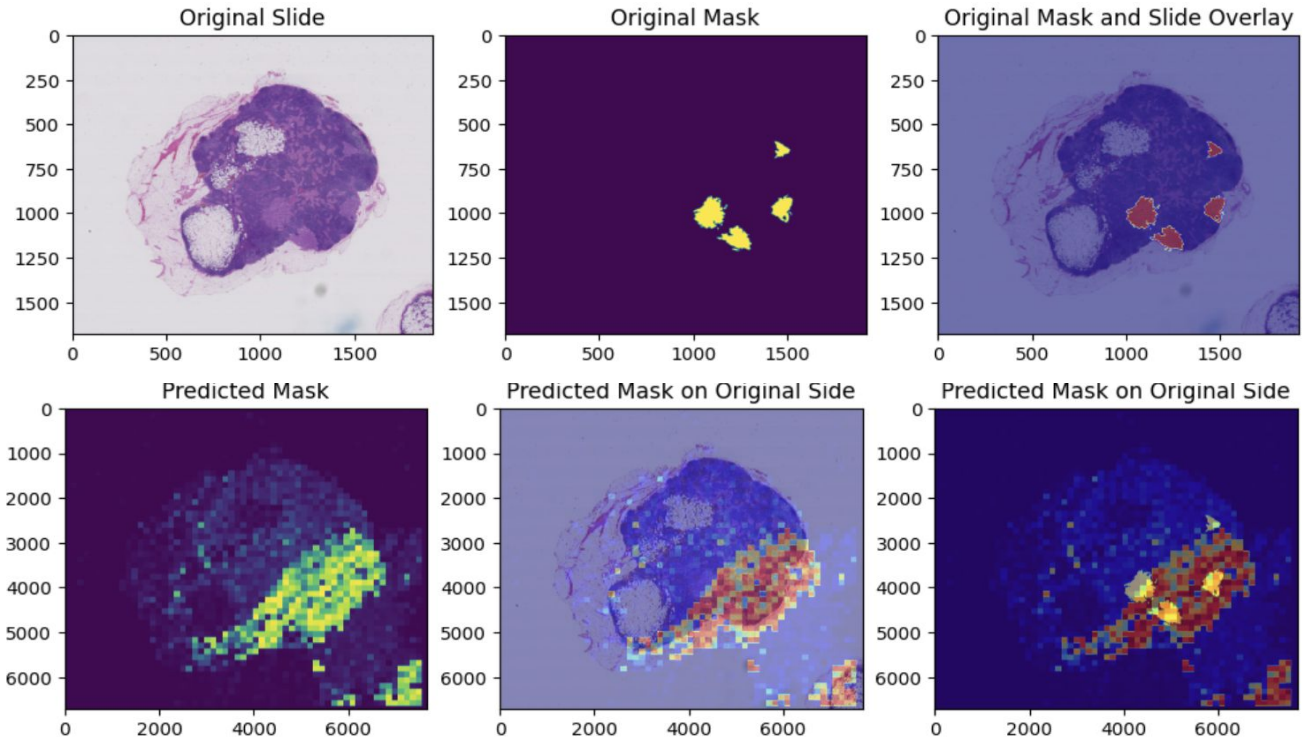




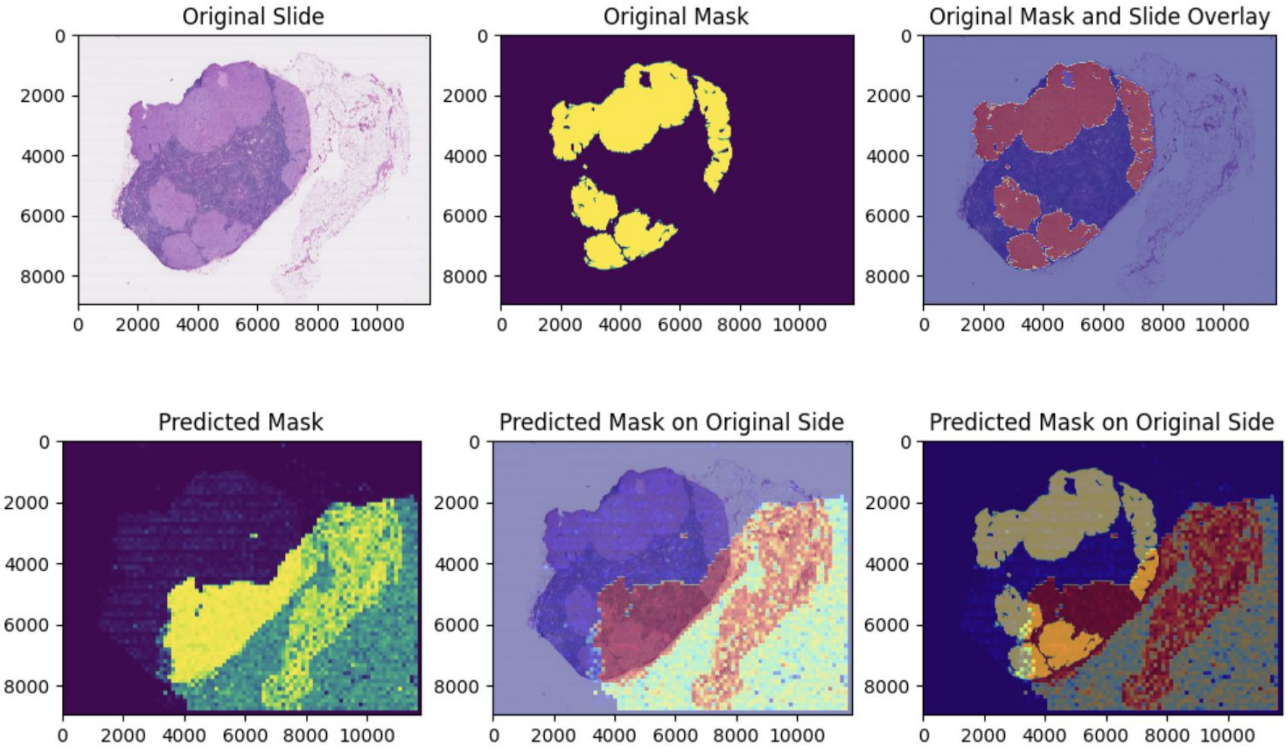
Result

Visualization

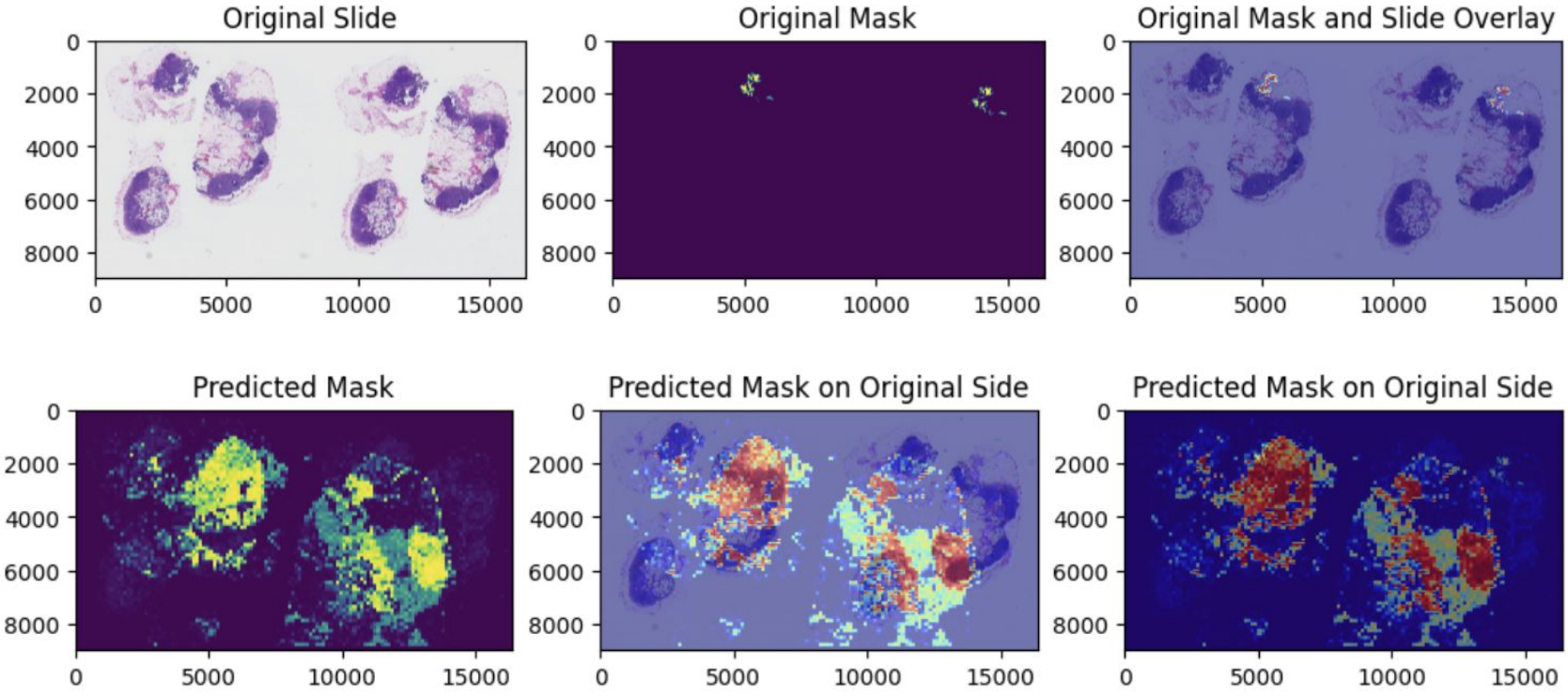
Note: Due to limitation in RAM, we couldn't print three images in our notebook at the same time, so we kept records in these slides.



Visualization



Visualization



Evaluation Metrics

Sensitivity = 0.8902

Sensitivity is defined as the probability of a positive diagnostic test in a patient with the illness or injury for which the test serves as a diagnostic tool.

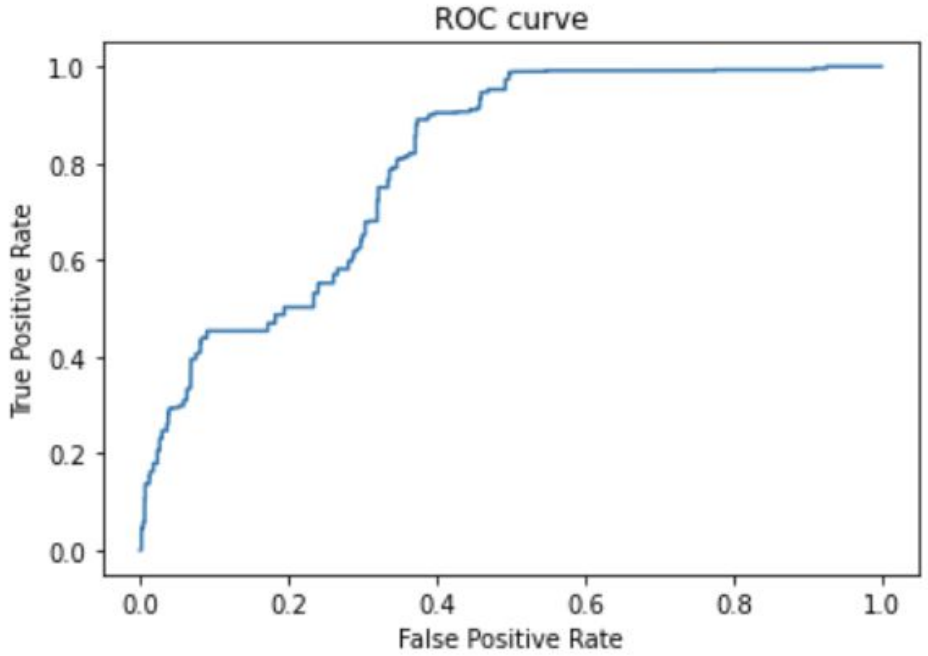
Specificity = 0.6263

Specificity is the probability of a negative diagnostic test in a patient free of the disease or injury.

Evaluation Metrics

AUC = 0.7996

ROC Curve



Next Steps

- Incorporate more scales
- Improve specificity
- Improve computation power