

Project Summary

Work Environment: Jupyter Notebook

Programming language: Python

Project goal: Building and trying out various CNN models for detection of cancer in colorectal histology. Using best picked to display the result in various formats.

Introduction

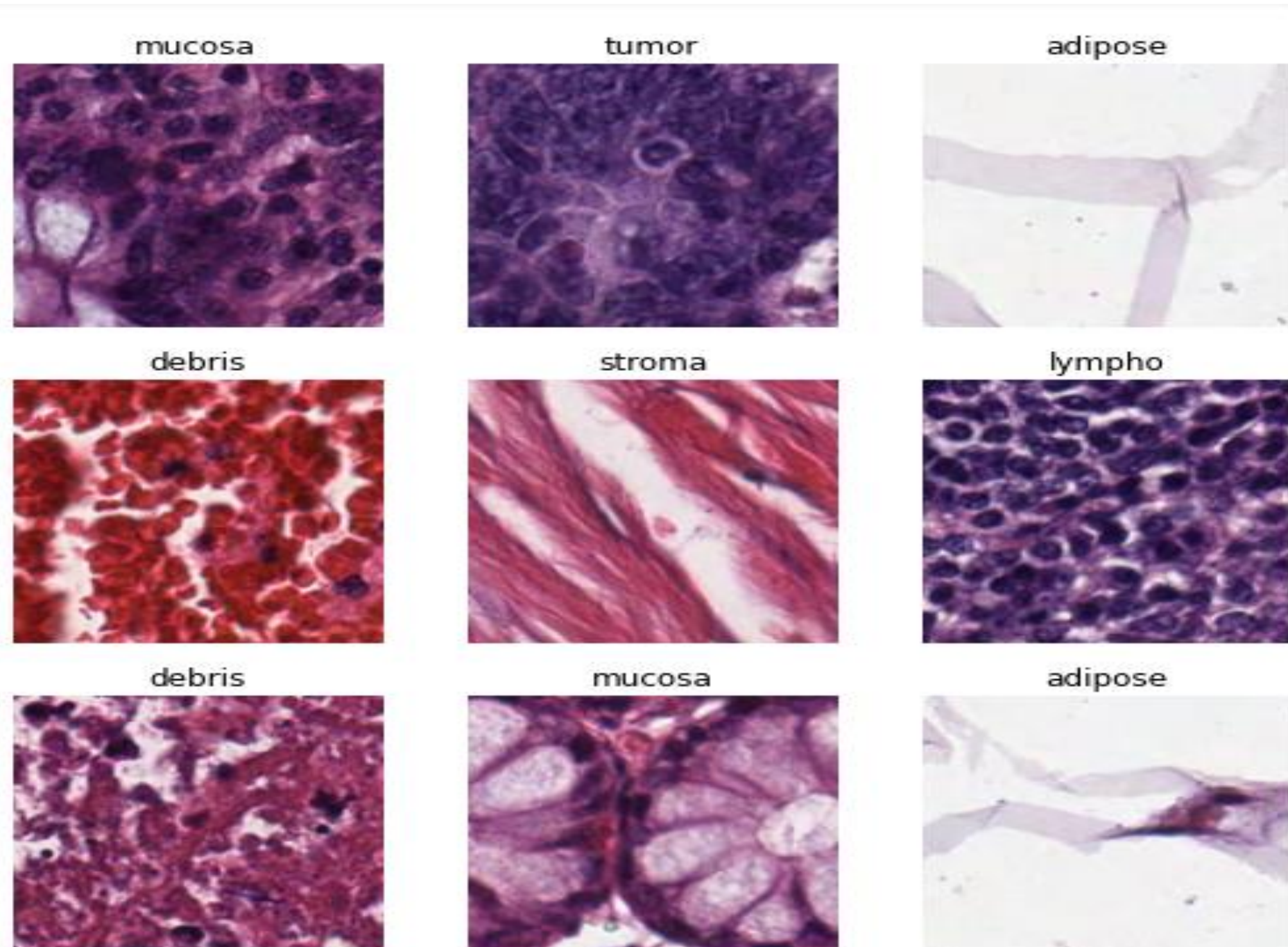
- Automated recognition of different tissue types in histological images is an essential part of digital pathology. Texture analysis is a common solution to this problem.

Our aim in this project is two-fold:

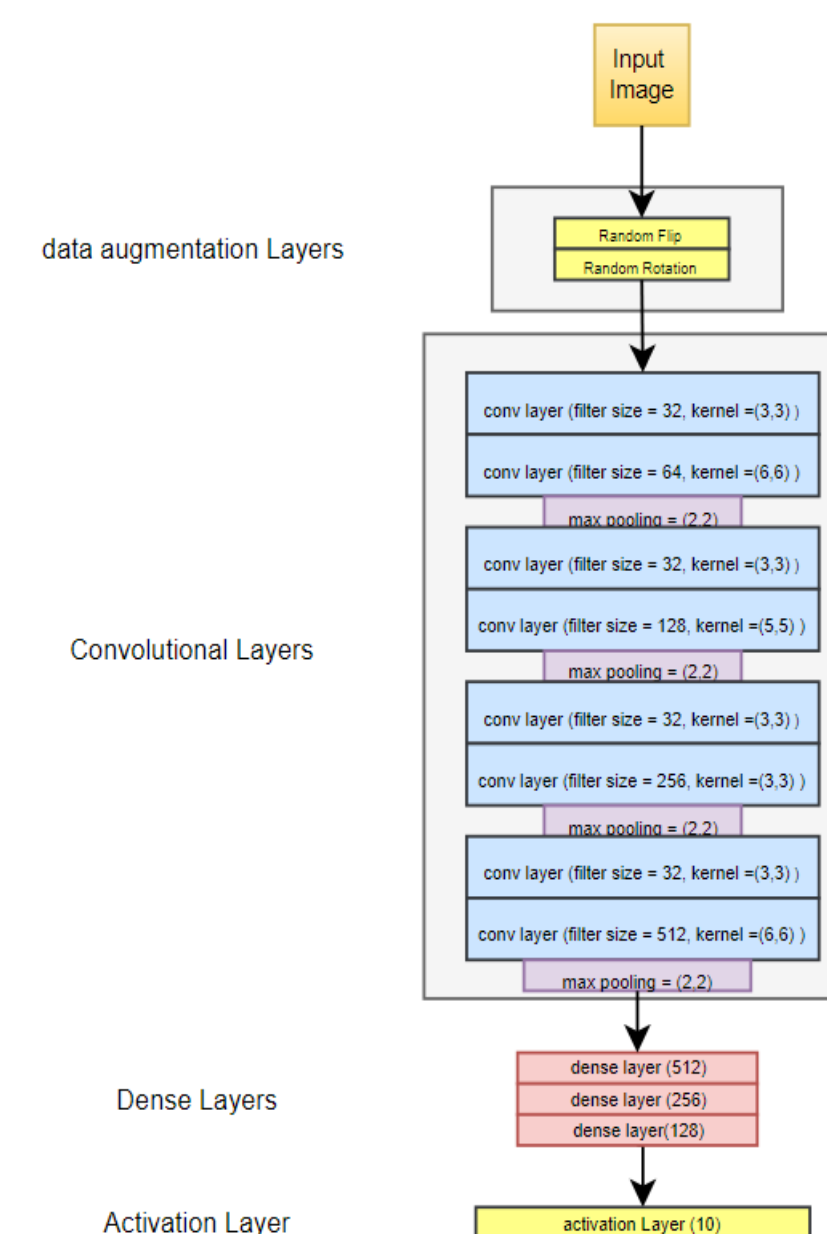
- Using a CNN model to more accurately classify tumor tissue in histological imagery.
- Building a model that can not only detect tumor tissue but also classify other types of cancerous tissue which will aid the diagnostic process.

The Training/test Dataset

- For our Training and test data we used a dataset consisting of 8 different types of texture of colorectal histology
- Each example was a 150x150x3 RGB image.

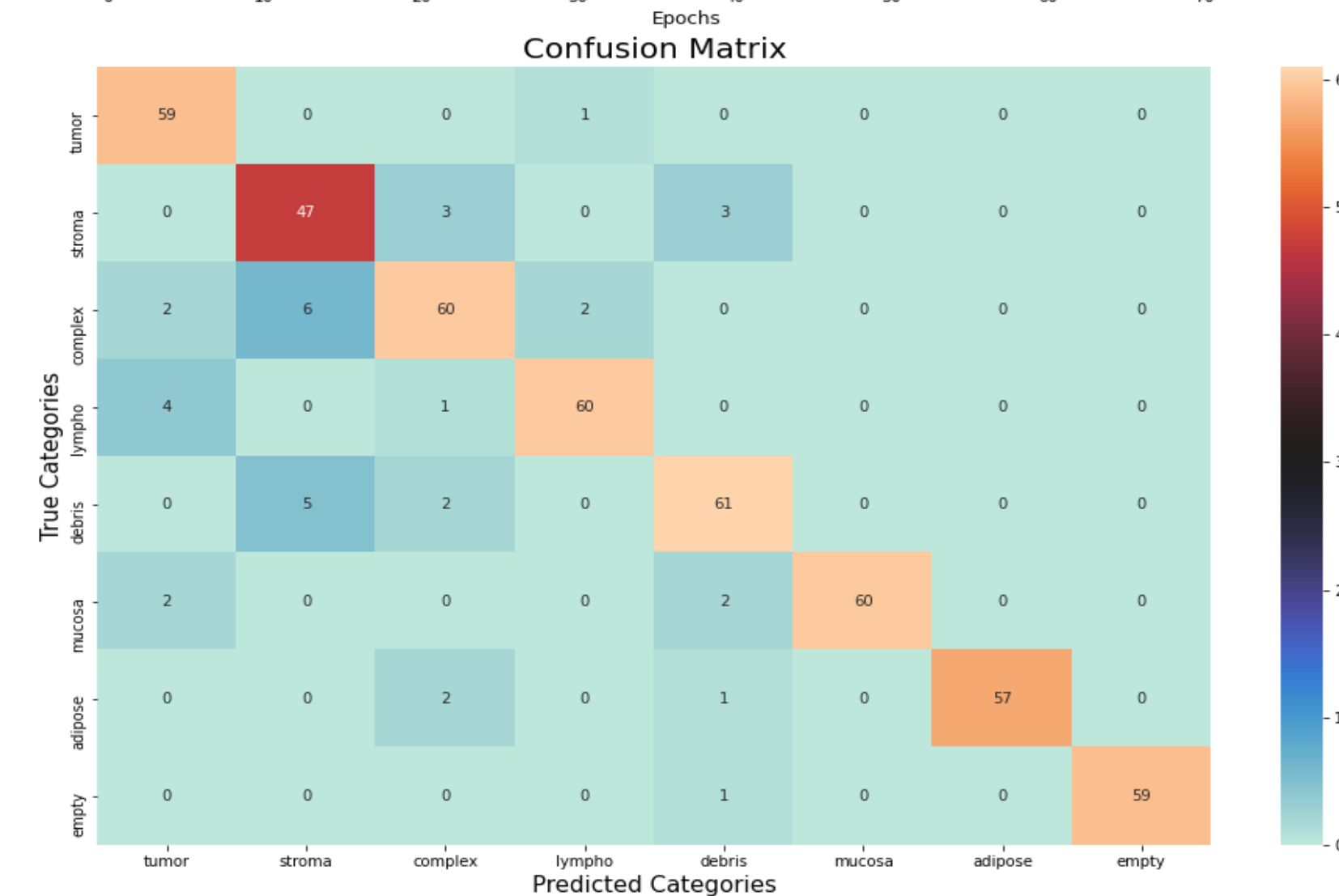
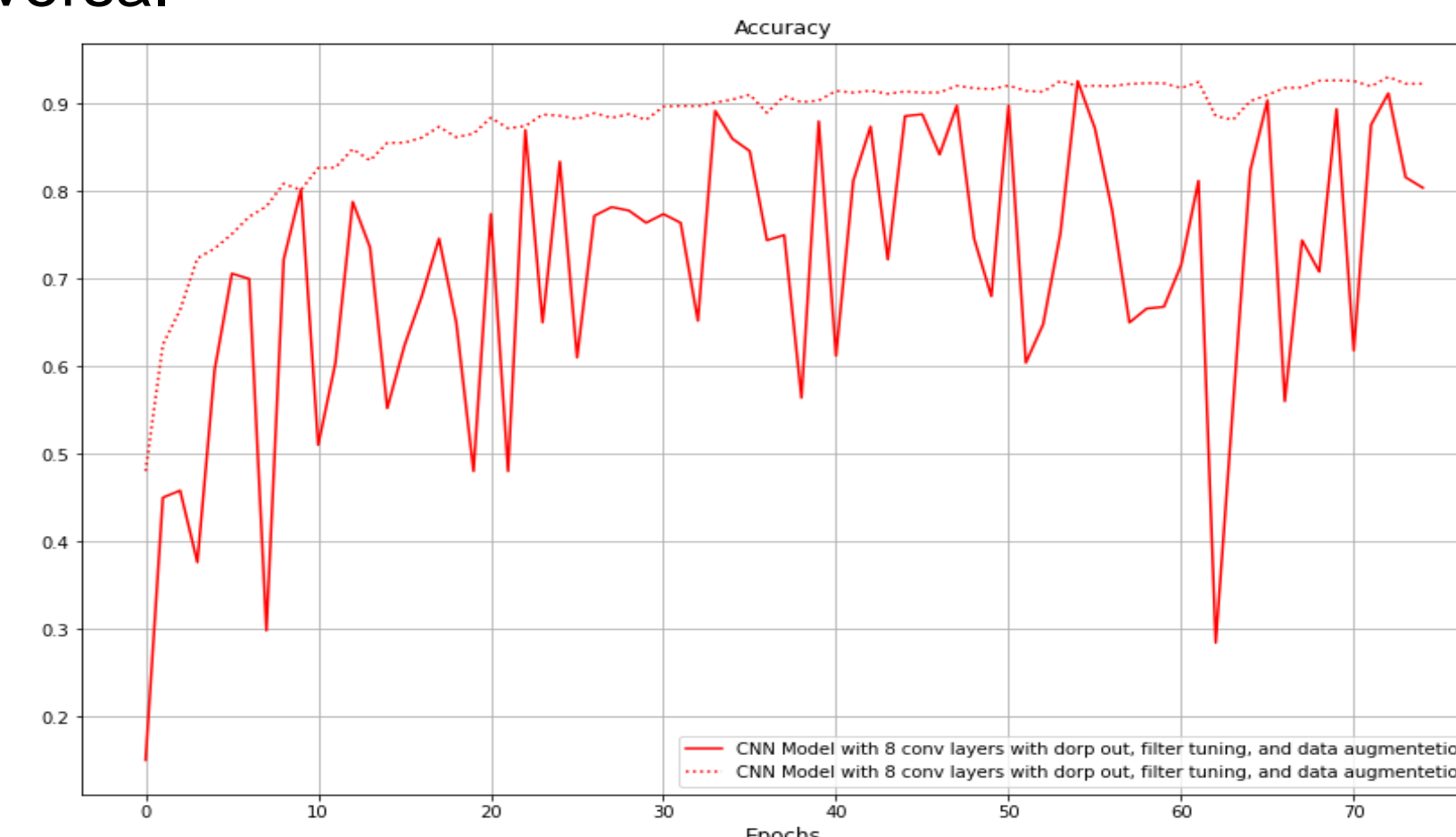


The Model We Used



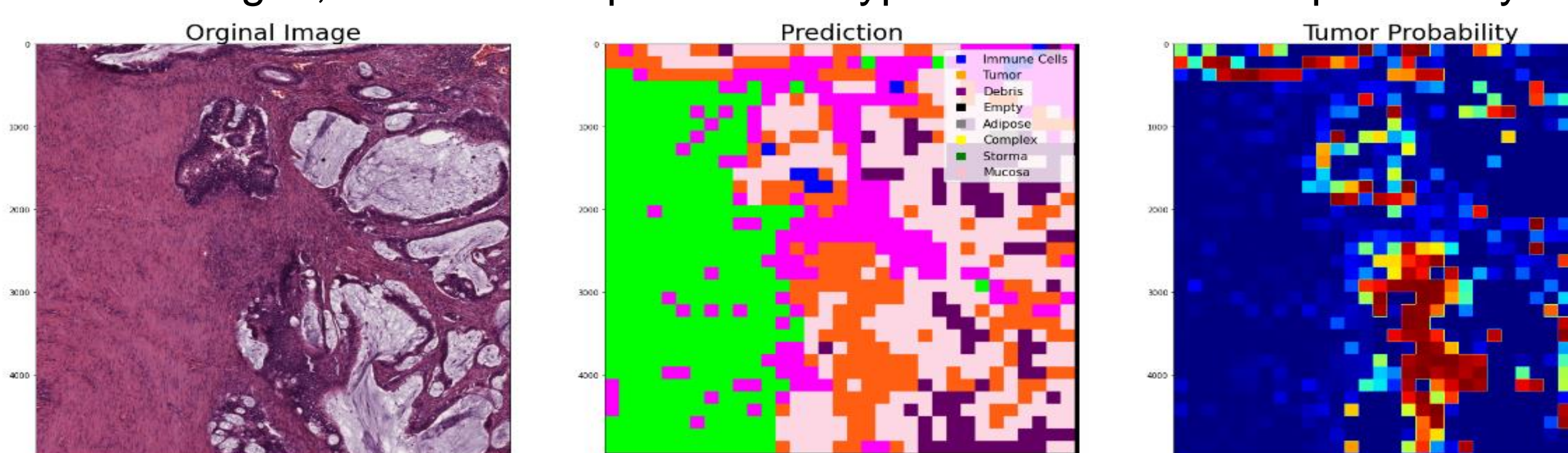
Results

- We were able to reach an accuracy of 92% on the test data
- Most of the misclassification were between one cancerous tissue and another, there were almost no healthy tissue classified as cancerous or vice versa.



Predictions on Full Images

- After we finished training our model we loaded a dataset containing full 5000x5000x3 histology images
- We separated our prediction into two images, one color map for tissue type and the second a probability heat map that a given tissue is a cancerous tumor



Discussions

- How well will this type of CNN model perform with other types of histology
- Which additional improvements we can make to our current model even more accurate than 92% percent.