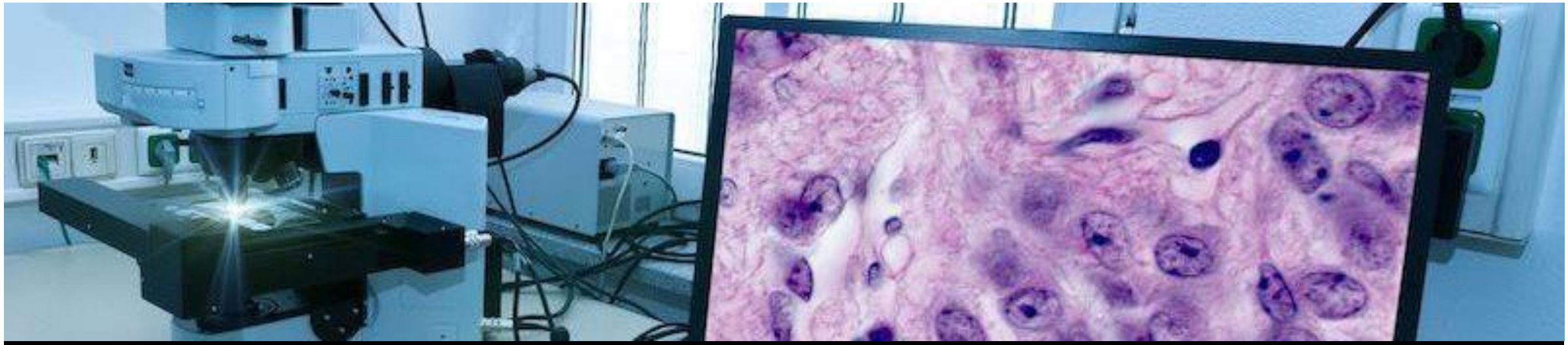
Cancer Histopathology Image Classification

DHANI POZNIAK

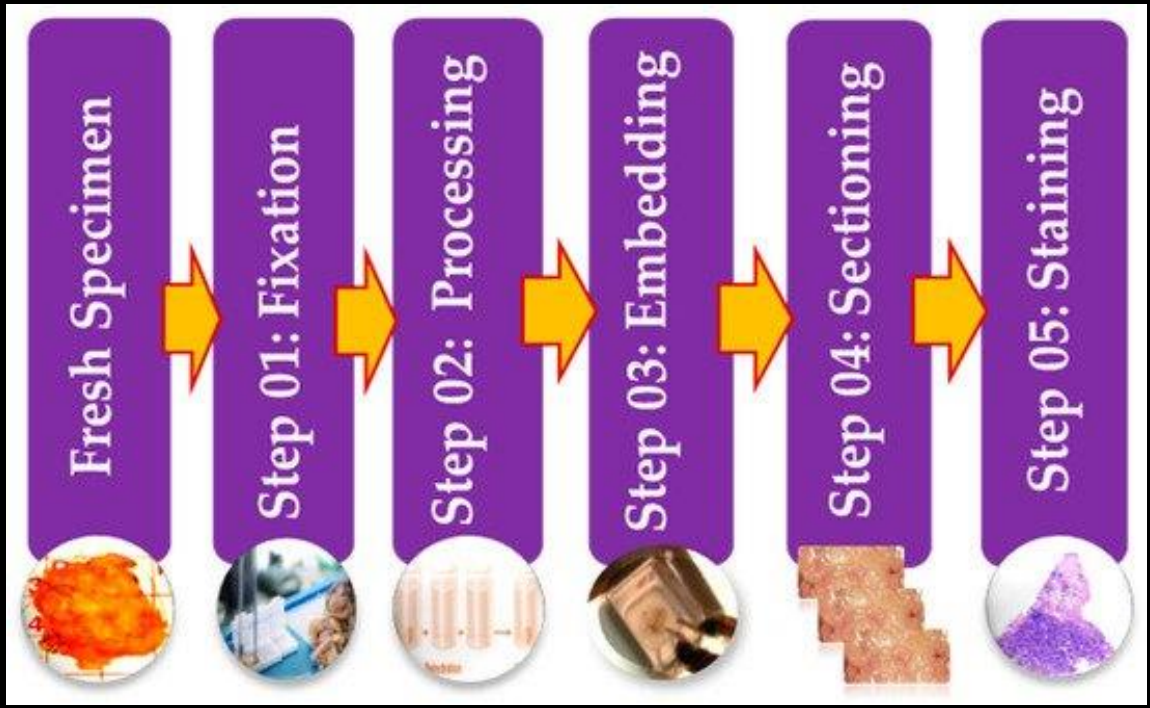


Business Understanding

- Investigation of cells and tissues at a microscopic level
- Digital histopathology has evolved significantly in modern medicine
- Visualize the status of biological structure which can be used to diagnose diseases like cancer
- Specialized cameras with a microscope

Histopathological Image Preparation

- Diagnosticians compare new processed images with those of healthy tissue samples in order to identify any differences or abnormalities



- Staining the tissue with dyes increases the contrast to the tissue sample and highlights specific features

Machine Learning in Histopathology

- ML models preserve performance while decreasing the workload for pathologists
- Essential for CAD models to precisely and confidently label classes and minimize reading interpretation run times
- The processes for a CAD model in large patient populations would include receiving a batch of raw images, pre-processing them, and then produce an array of diagnosis labels.
- Increased storage/ cloud capabilities has made it easier to store large numbers of digitized images enabling fast remote analysis and secure pathology information.
- ML models will offload some of the pressure for Doctors and diagnosticians by reducing redundancy and improving accuracy and capacity.

Data Understanding

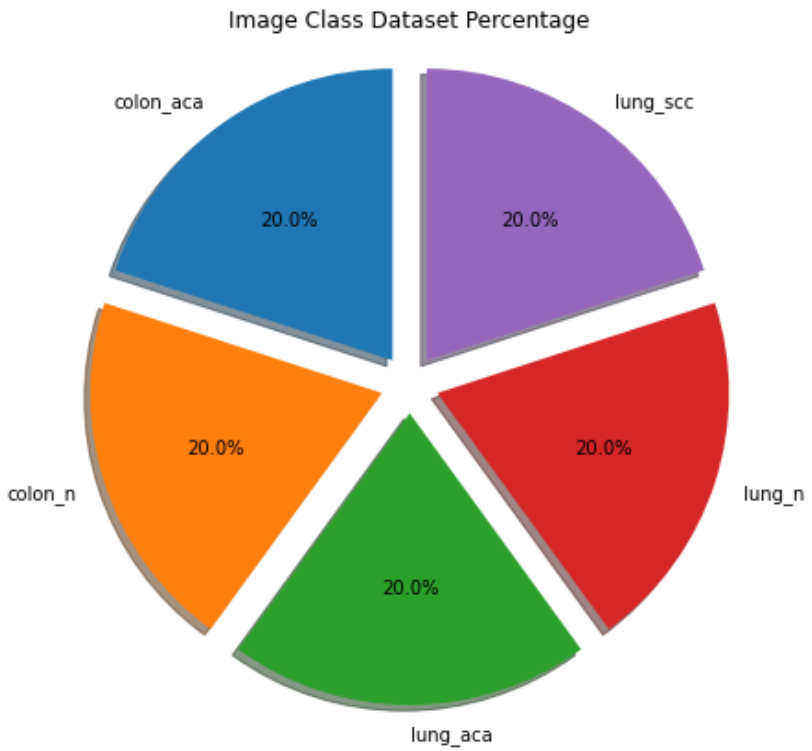
- A. Borkowski, "Lung and Colon Cancer Histopathological Images", Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images. [Accessed: 30- Jul- 2022].
- Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset (LC25000). arXiv:1912.12142v1 [eess.IV], 2019
- 25000 images, 5 Classes

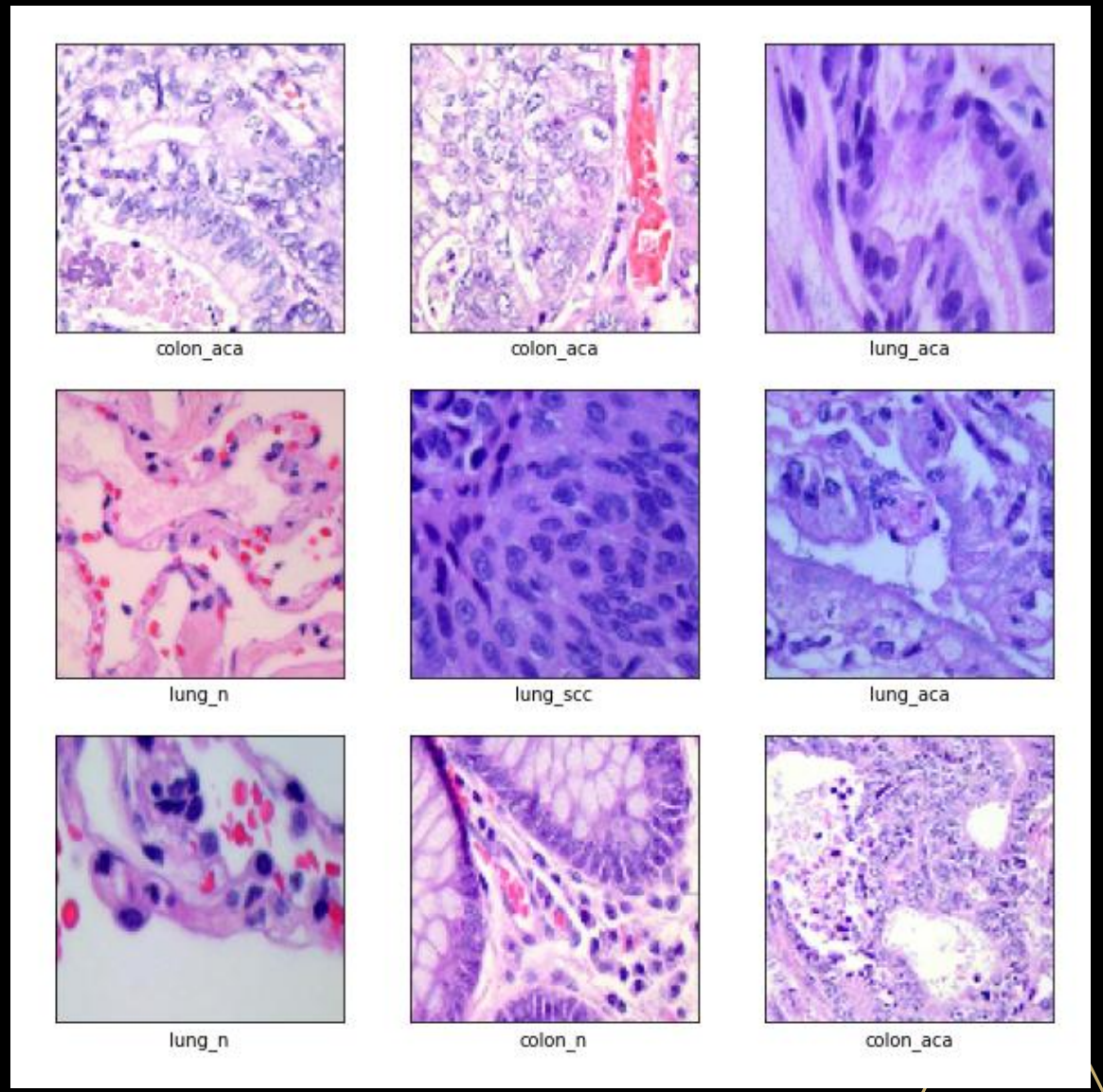
Image Labels:

- Lung benign tissue: 3
- Lung adenocarcinoma: 2
- Lung squamous cell carcinoma: 4
- Colon adenocarcinoma: 0
- Colon benign tissue: 1

Data Preparation

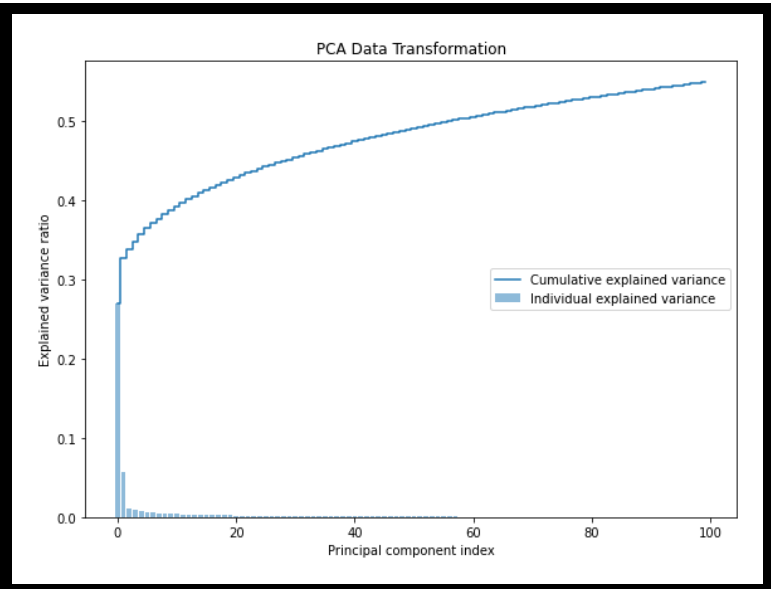
- Initial analysis shows a balanced dataset and successful loading of images
- Images resized from 768x768 to 96x96 which will be beneficial with storage capacity

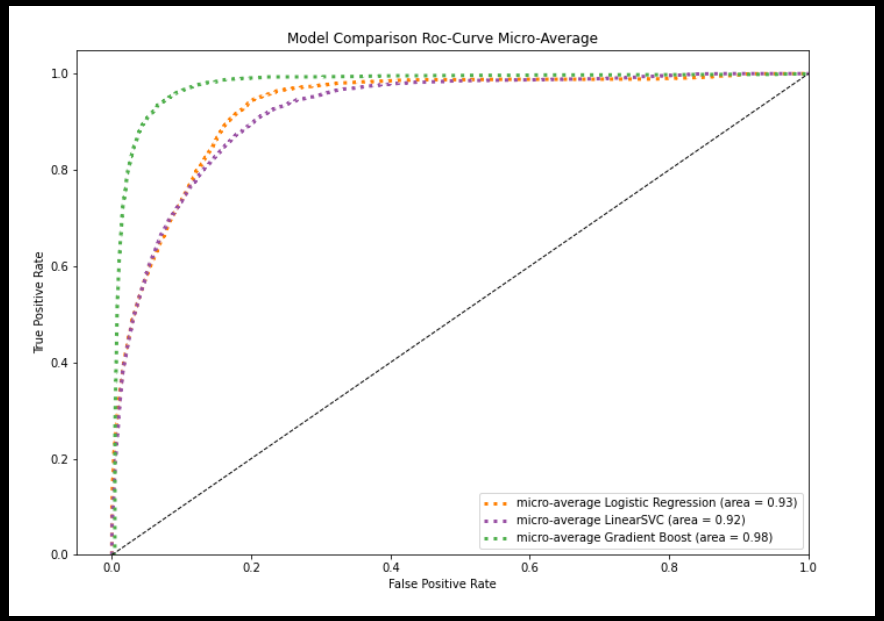


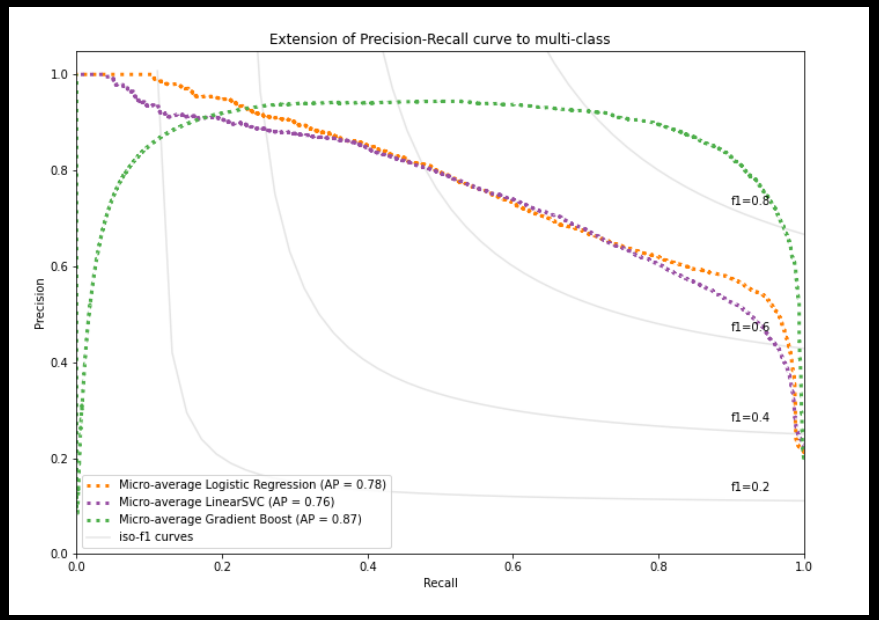


Feature Engineering

- PCA reduce features from 27649 to 100
- Represented 55% of the spread of data
- Gradient Boosting able to achieve 88% accuracy score

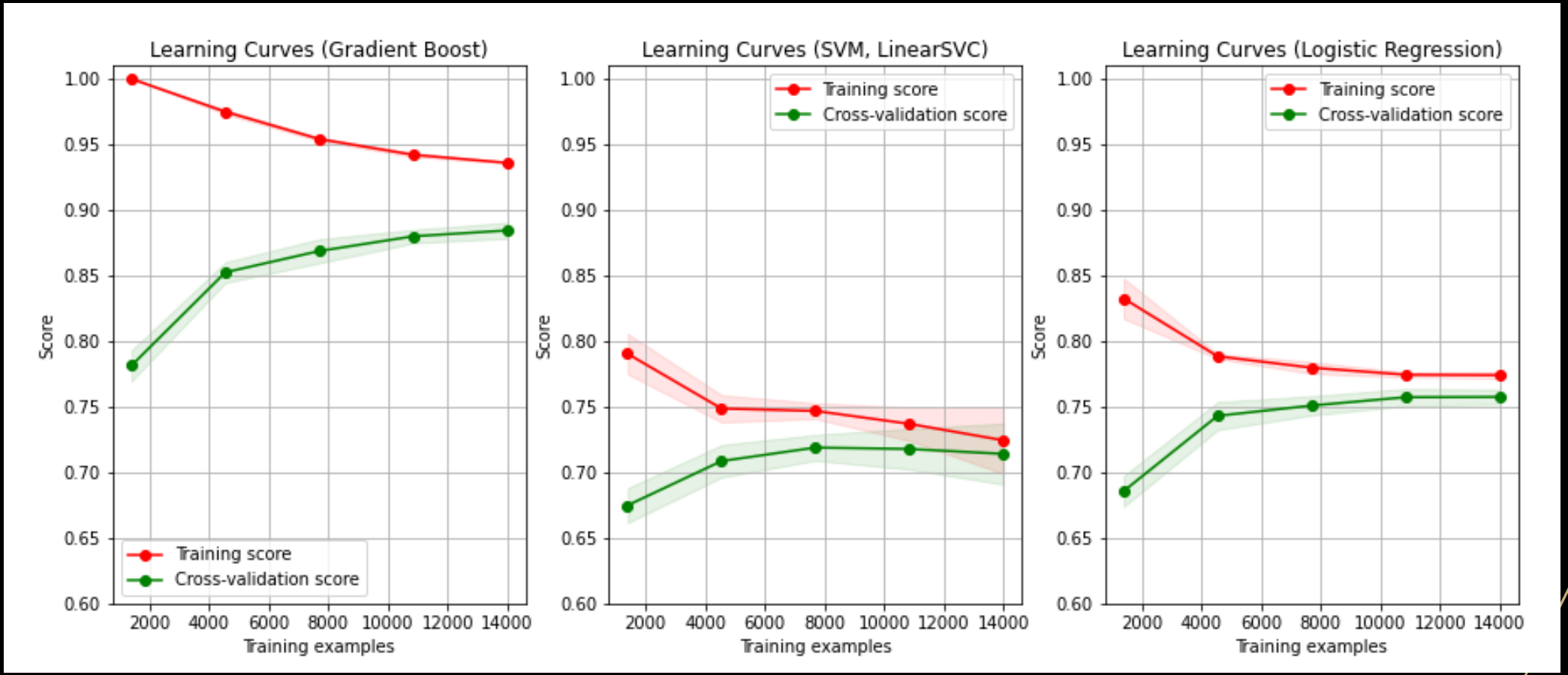






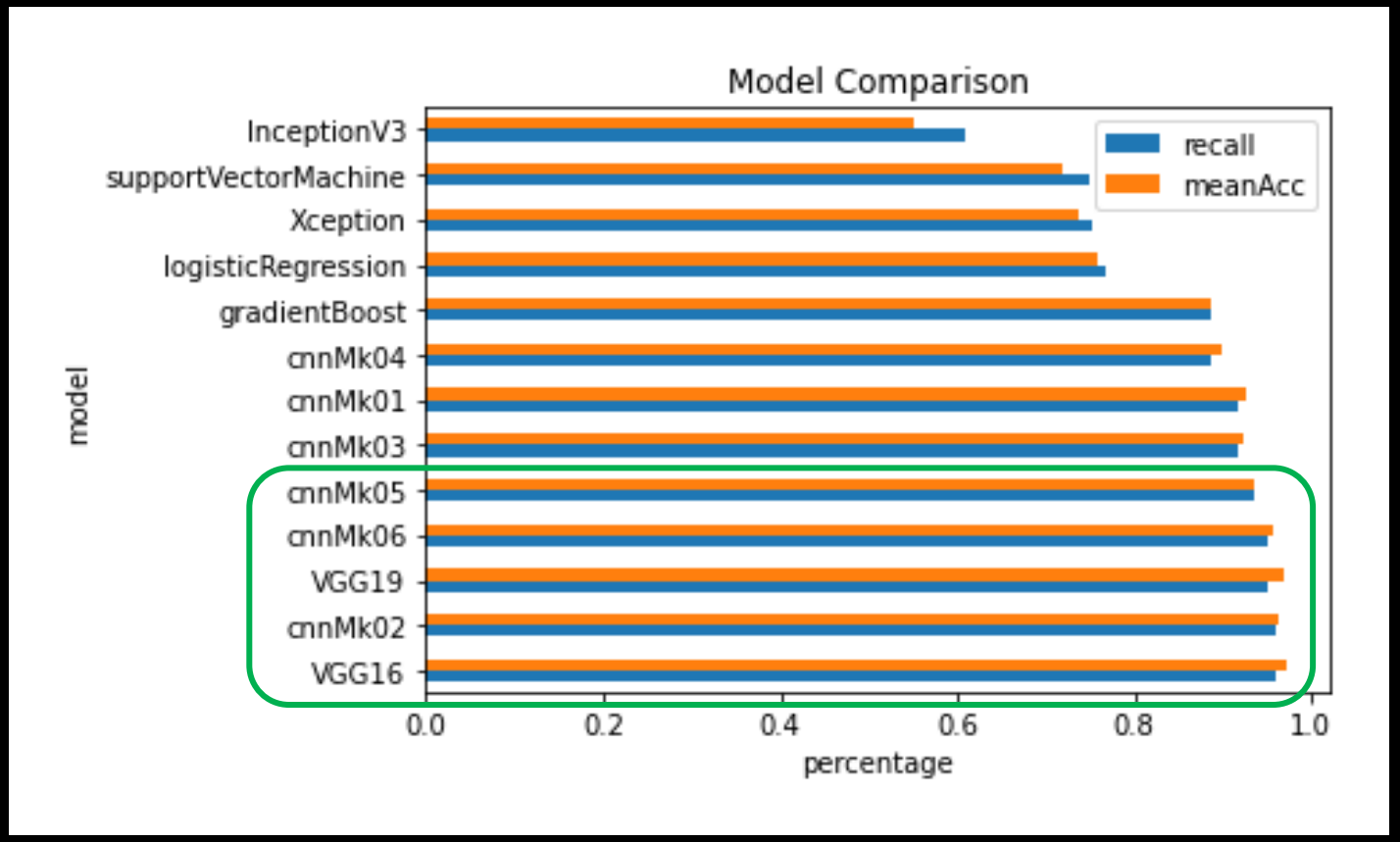
Learning Curves (GradBst, SVC, LogReg)

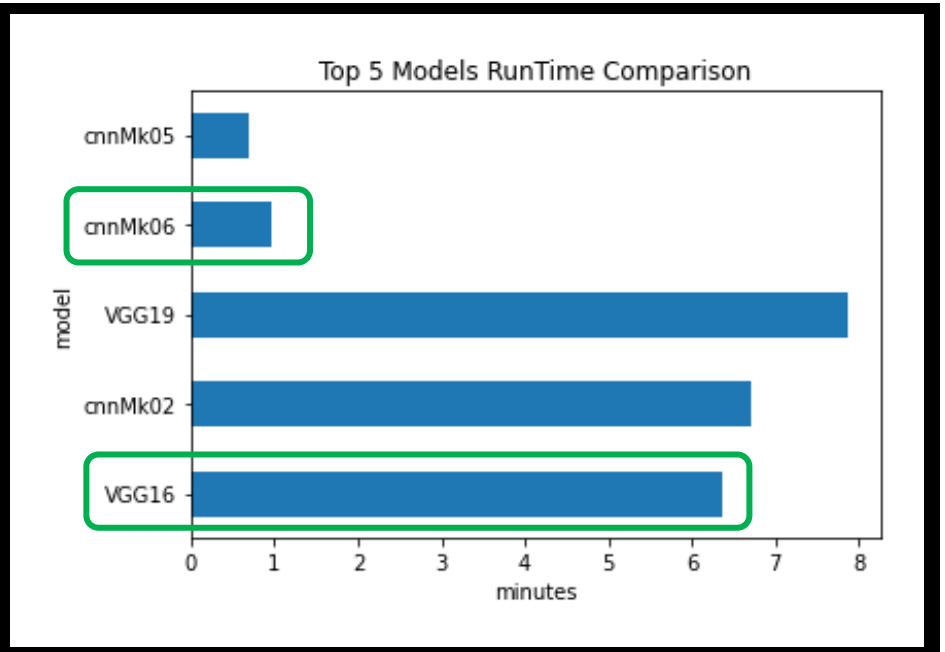
- Performed well considering half the data lost, be careful of overfitting
- Convolutional Neural Networks preferred algorithm



Model Comparison

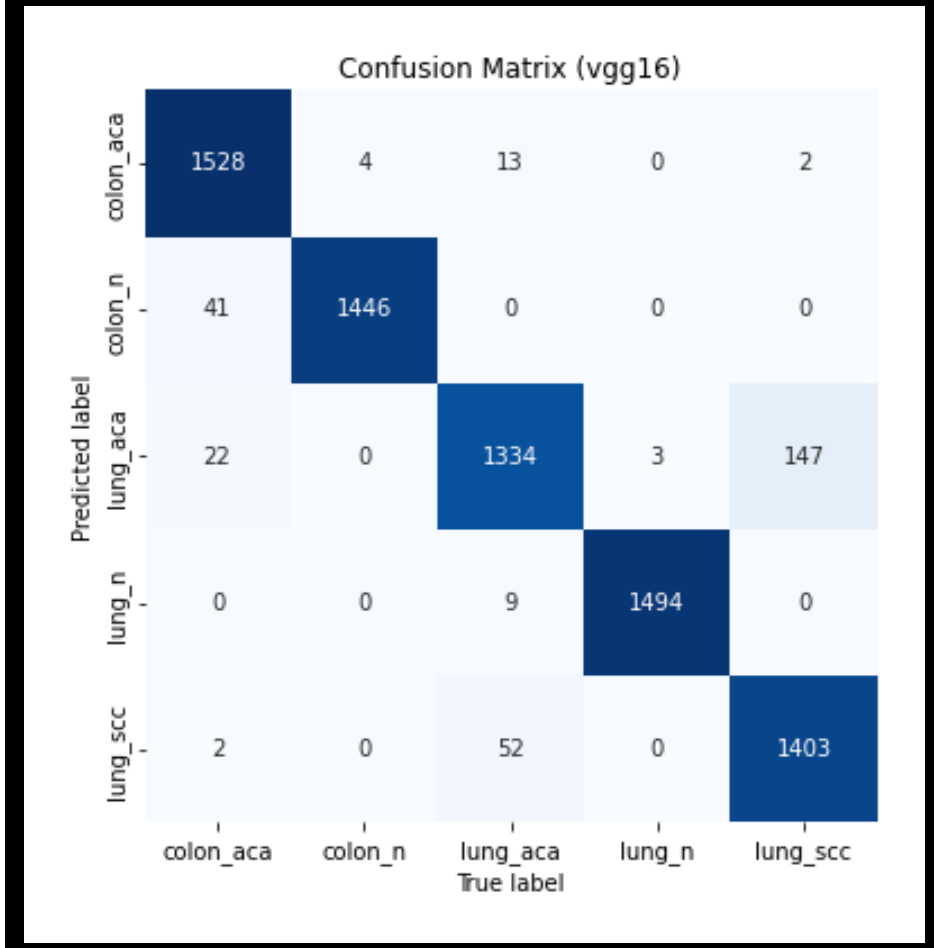
- Focus on Recall (false negatives)
- Proportion of actual positives identified correctly

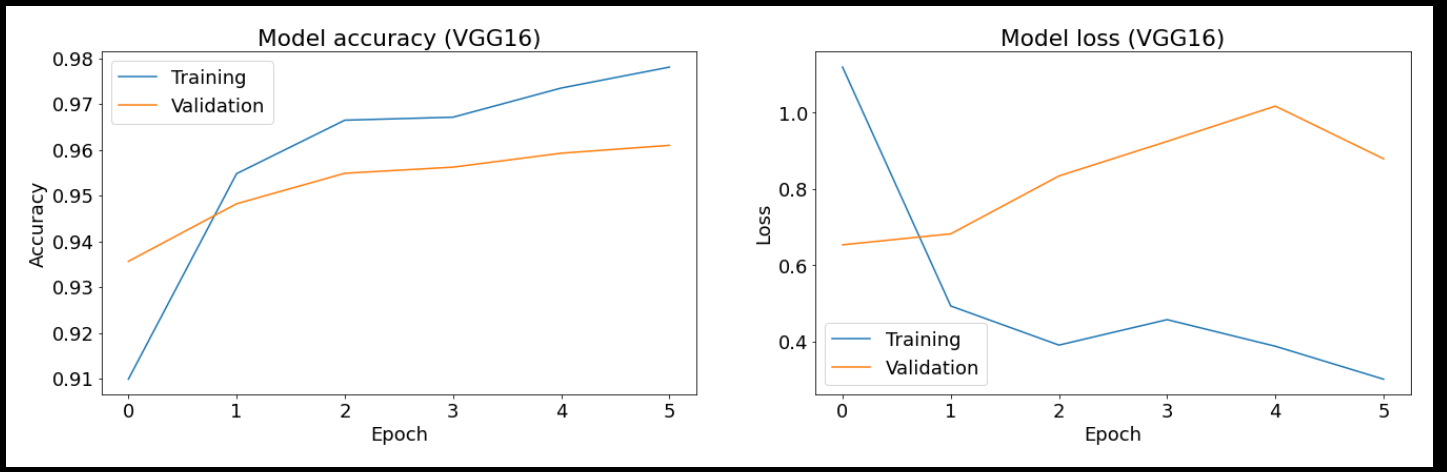




- BatchNormalization/ Dropout/ MaxPooling/ Dense Layers
- Runtime for 7500 unseen images processed and classified

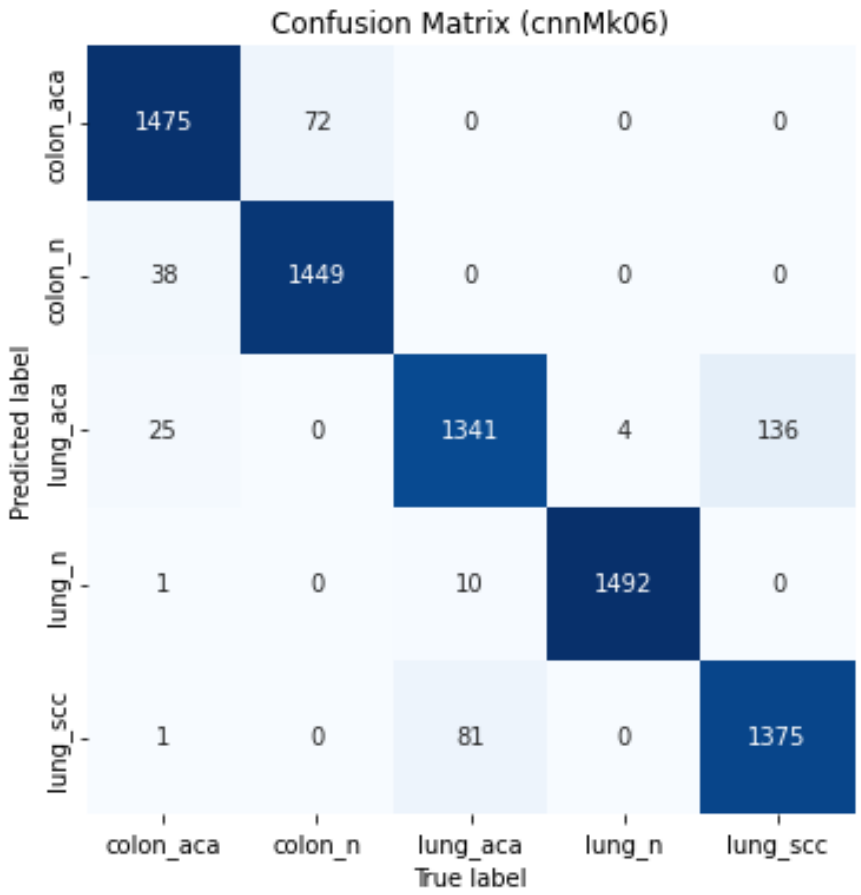
Pre-Trained CNN (VGG16)

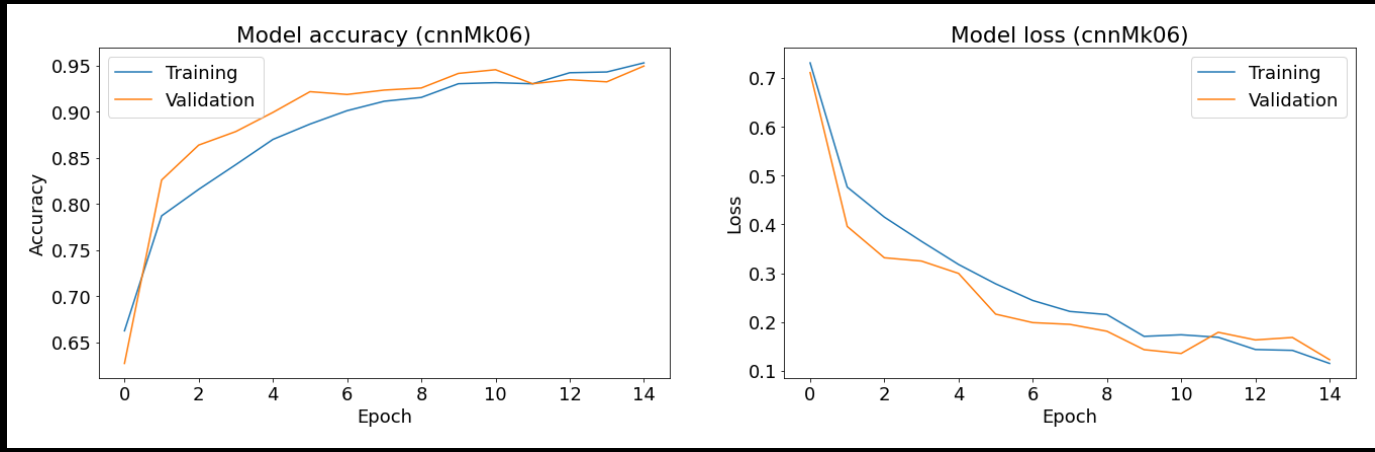




- Total Params: 14 737 733
- 21 Hidden Layers
- 10 epochs
- Classified 7500 unseen images in 6 minutes with 96% accuracy

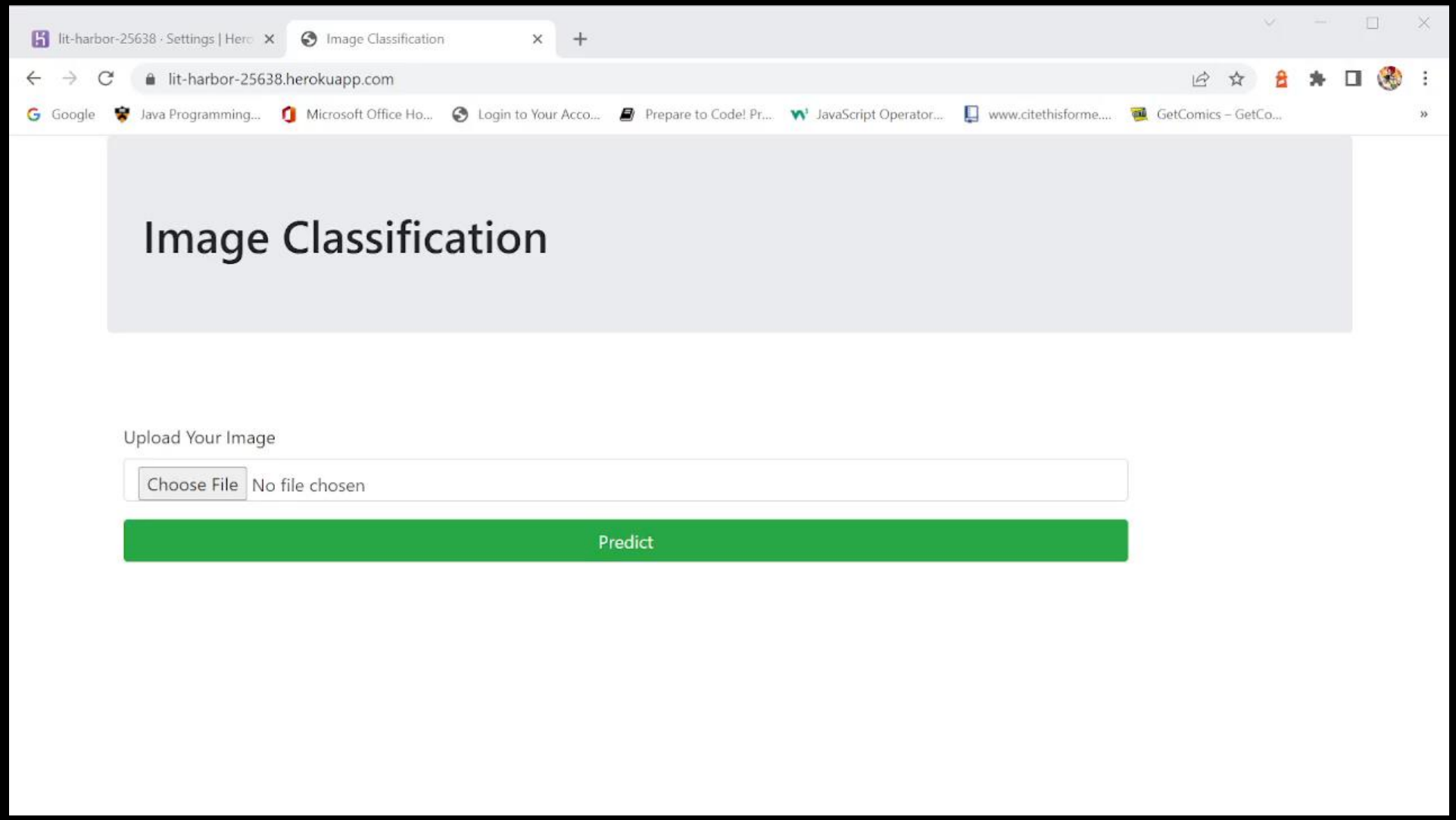
Custom CNN (cnnMk06)





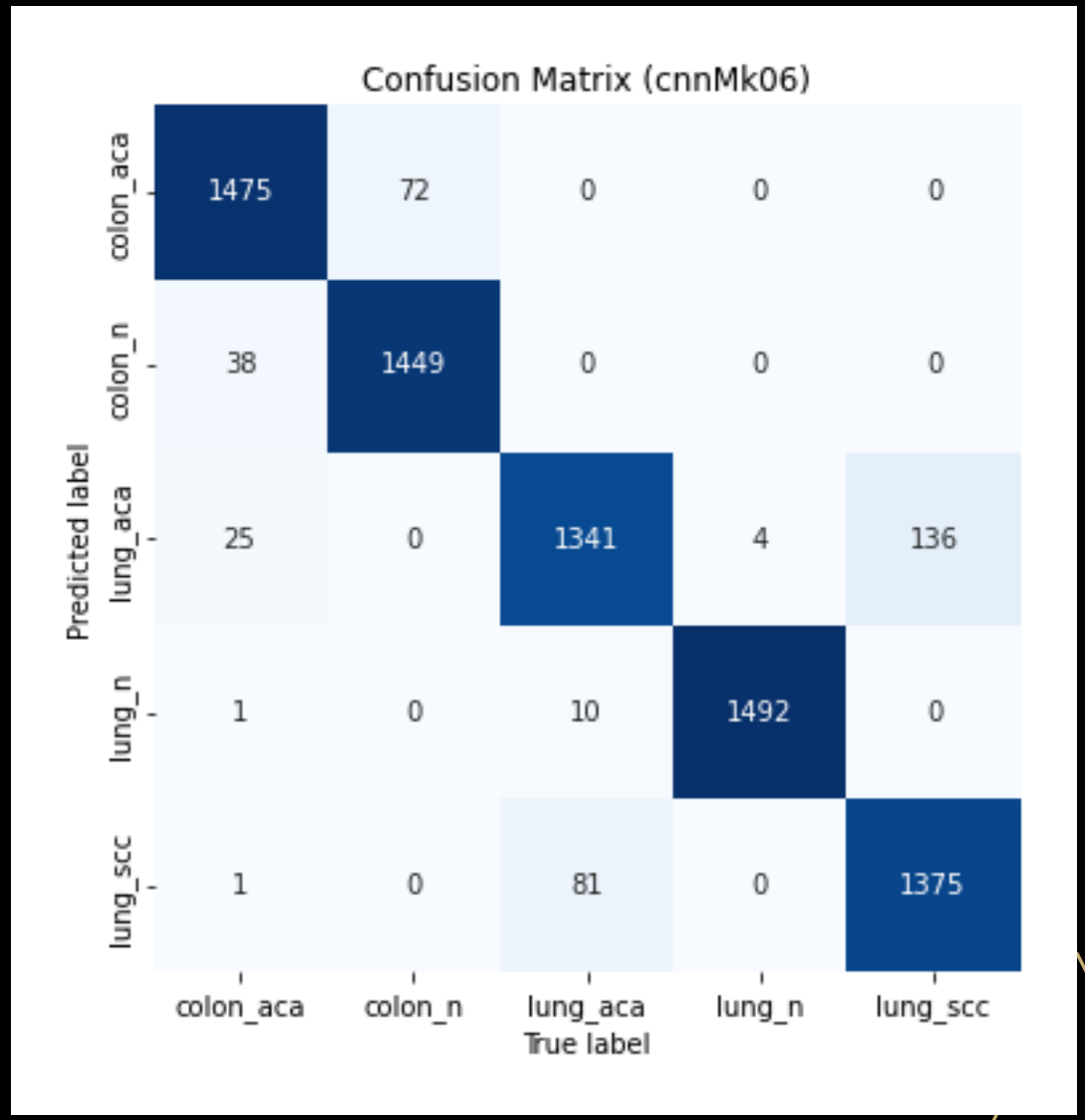
- Total Params: 66 565
- 8 Hidden Layers
- 15 epochs
- Classified 7500 unseen images in 1 minute with 95% accuracy

Deployment Demo: https://lit-harbor-25638.herokuapp.com/



Future Development

- The cnnMk06 legitimate model is able to predict with 95% confidence the class of the histopathological image
- Consider separate models for each cancer (colon/ lung)
- With better processing equipment, extend the number of epochs run
- Further finetuning of custom model
- store large numbers of digitized images enabling fast remote analysis and secure pathology information (96x96) instead of (768x768)
- preserve performance while decreasing the workload for pathologists so that they can focus on future research and development to enhance and evolve the medical industry



Contact

LinkedIn
https://www.linkedin.com/in/dhani-pozniak-34801a196/

Github
https://github.com/pozengineer/dataScienceImageClassify/

Email
pozengineer@hotmail.com

Phone
0412 966 933

Thank You

Any Questions ?