

# Cancer Detection using Histopathological slides

## Deep Learning Approach

Nitin Mahajan  
Data Science Career Track  
Specialization: Deep Learning

# GOAL

Classify histological slides of lymph nodes of patients to predict whether they have cancer or not using Deep Learning



# Data Source

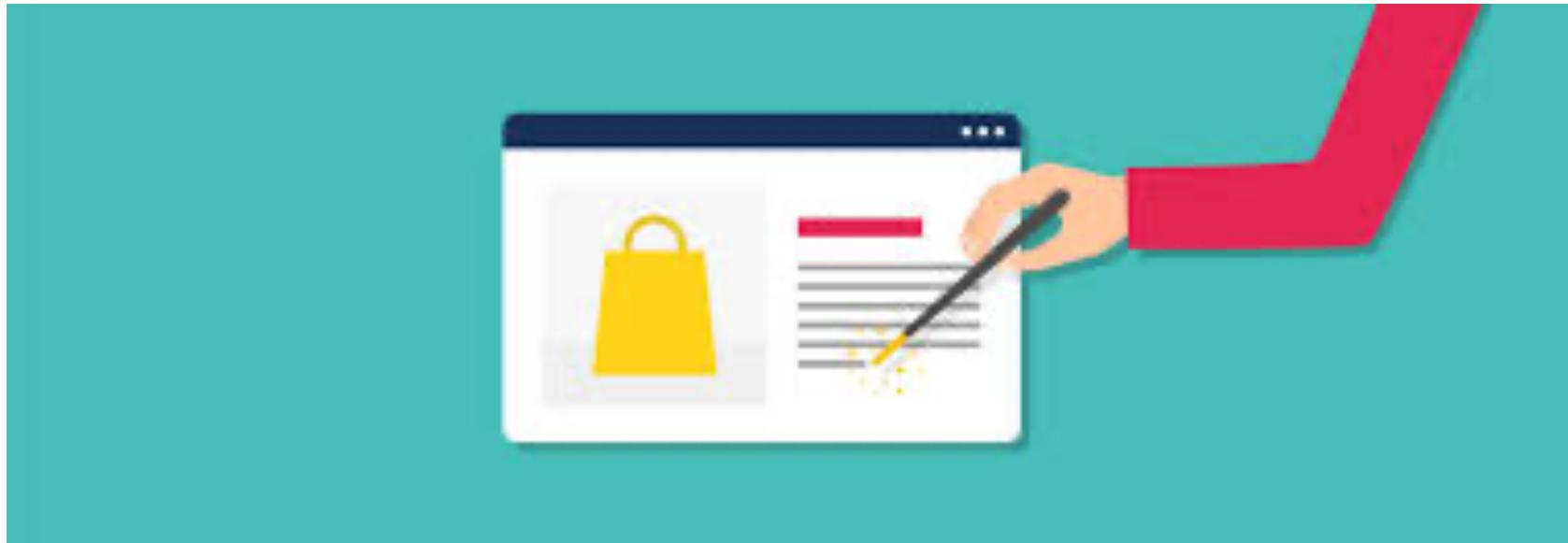


kaggle

<https://www.kaggle.com/c/histopathologic-cancer-detection/data>

# Data Description

- Train\_labels.csv files with labels 0 for normal and 1 for cancer.
- 220,025 images for training.
  - 1309081 Images with '0' label (Normal)
  - 89117 Images with '1' label (Cancer)
- 57,458 images for testing.



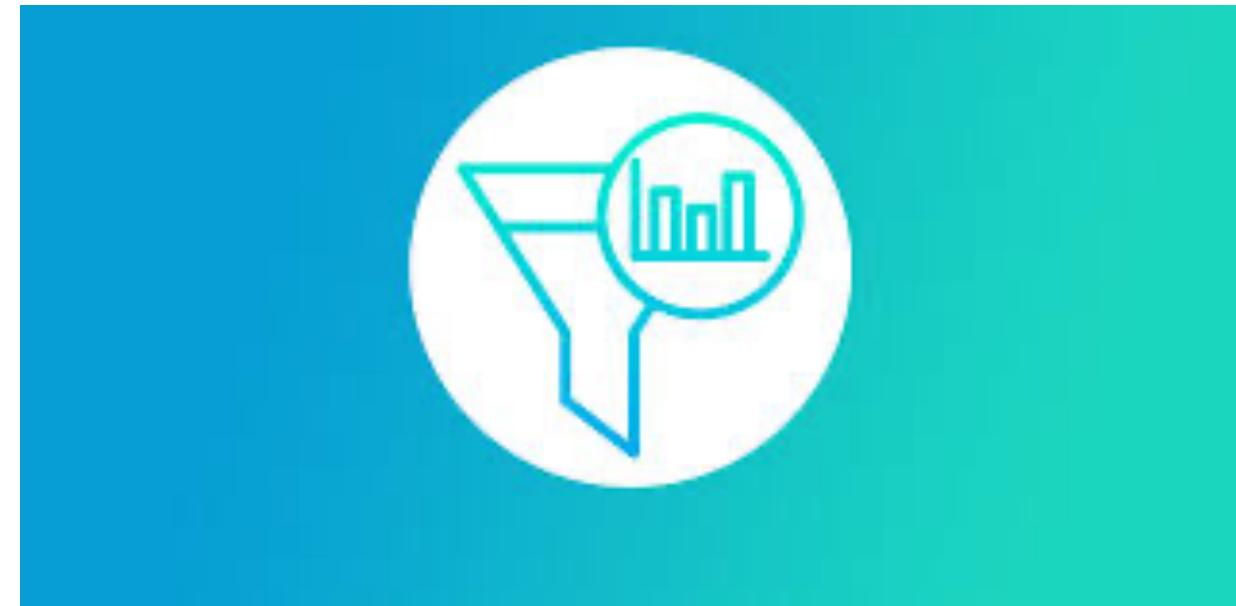
# Data Wrangling

- Data Sampling

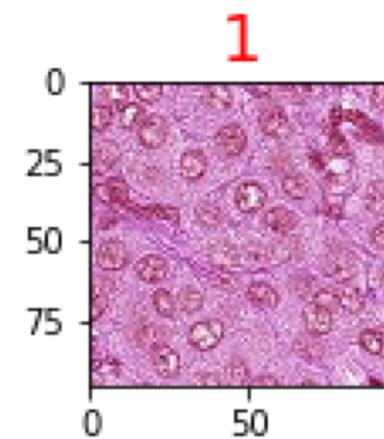
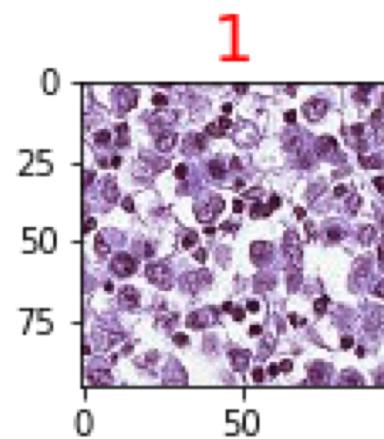
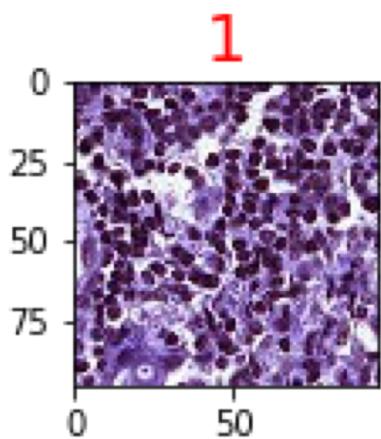
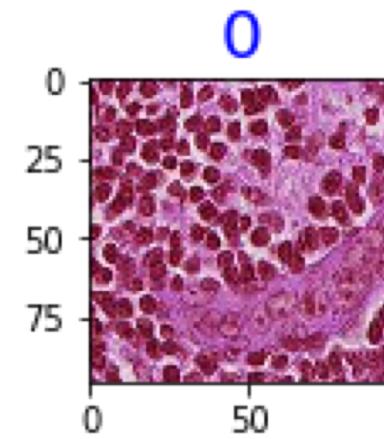
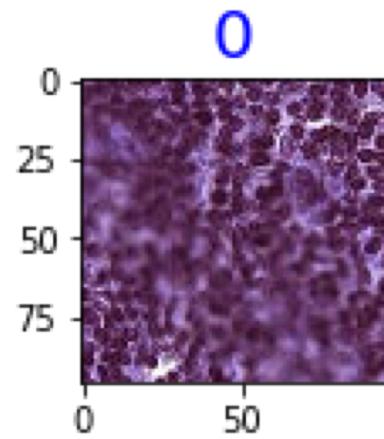
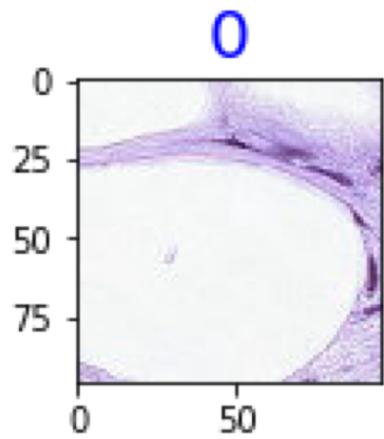
- 10,000 images for normal tissue
- 10,000 images for cancerous tissue

- Training Data splitting

- 8,000 images for training on each class (normal, cancer)
- 2,000 images for validation on each class (normal, cancer)



# Images



Normal

Cancer

# MACHINE & RESOURCES

## FOR BUILDING CONVOLUTIONAL NEURAL NETWORK MODELS

- **Machine:**

Apple MacBook Pro

Processor: 2.6GHz 6-Core Intel Core i7

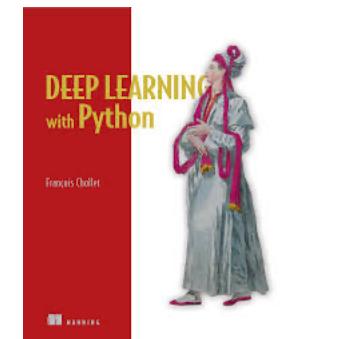
Memory: 16GB 2400 MHz DDR4



- **Modules and libraries:** Python and Keras with TensorFlow in backend



- **Reading Material:** Deep Learning with Python (Francois Chollet), Udemy Course - AZ Machine Learning and other online resources.



# What Next

- Building CNN models
- Confusion Matrix
- Apply Pretrained Models