Project 5 Proposal
Breast cancer detection using histopathology images

JiEun Song

Overview

Breast cancer is the most common type of cancer and the second leading cause of cancer death in women. There are more than 200,000 new cases every year in US. If a patient is suspected to have breast cancer during the regular screening, small amount of cancer tissue will be taken for further examination, and pathologists will examine the each tissue samples to determine cancer. This process is critical in cancer diagnosis, thus my goal for this project is to build a model to detect breast cancer using microscope images. This model could help pathologists to expedite diagnosis process and also patients to get treatment sooner.

I found my dataset from Kaggle. The dataset consists with 198,738 cancer-negative and 78,786 cancer-positive images with labels.

Possible Tools and Models

I will use Google Colab with GPU, and train several different models such as gradient boosting in Tensorflow, XGBoost, etc. and compare the models. The final aim for this project is to build a web app that takes user's image and returns the result whether the input image is cancer or not.