

DATS6203 Final Project Proposal:

Predicting Invasive Ductal Carcinoma in Tissue Slices

Yue Li, Hsueh-Yi Lu

1 Introduction

Invasive ductal carcinoma (IDC) is one of the most common types of breast cancer. It's malicious and able to form metastases which makes it especially dangerous. Often a biopsy is done to remove small tissue samples. Then a pathologist has to decide whether a patient has IDC, another type of breast cancer or is healthy. In addition sick cells need to be located to find out how advanced the disease is and which grade should be assigned. This has to be done manually and is a time consuming process. Furthermore the decision depends on the expertise of the pathologist and his or her equipment. Therefore deep learning could be of great help to automatically detect and locate tumor tissue cells and to speed up the process. In order to exploit the full potential one could build a pipeline using massive amounts of tissue image data of various hospitals that were evaluated by different experts. This way one would be able to overcome the dependence on the pathologist which would be especially useful in regions where no experts are available.

One of the projects of this topic was done by Laura Fink, the author updated her work in the Kaggle notebook named Breast Cancer. In this project, the author implemented the Resnet18 framework to train the data set and Cyclical Learning Rates (CLR) which is a learning rate search strategy. The goal was given a patient and a patch of a tissue slice predict whether it contains IDC or not. The prediction so far is done manually by pathologists and varies from expert to expert. The goal is to assist with an automatic detection of tumors (not expert dependent).

2 Self Contribution

In this project, we seek to investigate the performance of other pretrained network and constant learning rate on the same data set, compared with the results with the previous project. First, comparison between different frameworks on their architecture and performance on this IDC dataset, we will be able to give a better understanding of deep learning frameworks and training strategies. Second, we will explore more the data augmentation methods. Besides the normal image augmentation methods using the transforms package in torchvision, we will use Generative adversarial network (GAN) to do image augmentation.