**Name(s)**
Ramya Ayyagari

**Project Number**
S1502

**Project Title**
# A Deep Learning Approach for Analyzing Tumor Histopathological Image Slides to Predict Breast Cancer

**Abstract**

**Objectives/Goals**
The goal of this project was to create a highly-accurate, automated detection platform that would be able to predict breast cancer by classifying histology slides from biopsies. The method must group images into benign and malignant subtypes, and should be integrated on a cloud server to eventually be easily accessible to any pathologist. This would greatly lower false positive/negative rates in diagnosis.

**Methods/Materials**
Using Keras and TensorFlow, I built two Convolutional Neural Networks (CNN), deep learning algorithms that were trained on the BACH (Breast Cancer Histology) dataset, to classify histology slides into binary (carcinoma, non-carcinoma) and multiple (normal, benign, in-situ, invasive) categories. The CNNs were then fine-tuned and run various times on an Amazon (AWS) GPU instance to ensure reproducibility of results. To achieve better performance, the CNNs were trained on top of deep, pre-trained neural networks, such as VGG16 and InceptionResNetV2. The CNN algorithms were then integrated into an automated detection platform that I built on the cloud.

**Results**
The fine-tuned binary classification CNN, when coupled with the pre-trained layers of the model InceptionResNetV2, achieved an accuracy of 95.0%. This is a significant improvement on the state-of-the-art accuracy for a similar dataset, 83.3%, indicating that my model surpassed a current gold standard. The multi-class classification CNN with VGG16 pre-trained layers achieved an accuracy of 83.75%, surpassing the state-of-the-art accuracy for a similar dataset, 77.8%.

**Conclusions/Discussion**
The contributions of this project include providing a clinically applicable platform to accurately diagnose breast cancer via tumor slides. This is important because the tedious work of classifying these slides manually is time-consuming, inaccurate, and costly. In the future, pathologists can use this to quickly validate their diagnoses, and patients will be able to receive results within seconds. Additionally, this project provides a deep learning framework that can be applied to other medical computer vision problems.

**Summary Statement**
I built a deep learning platform that can predict breast cancer with a 95% accuracy quickly and efficiently on a cloud server.

**Help Received**
I used the publicly available BACH dataset to train my model, and the open source platforms Keras and TensorFlow. I designed and performed all experiments myself.