Applications of Machine Learning Models on PDAC Diagnosis

By Robert Rivera

Mentor: Jorge Pineiro

Bronx Community College

Pancreatic ductal adenocarcinoma is amongst the most common, and deadly, forms of pancreatic cancer. The standard method of diagnosing PDAC is via imaging, however by this time it has likely become metastatic due to its asymptomatic nature; with patients having a 9% survival rate past 5 years after diagnosis. In 2020, a group of 16 researchers published their work in PLOS Medicine, detailing their use of a biomarker marker panel, extracted from blood and urine, and a data model(the PancRisk score), to see the viability of the panel for early diagnosis. The results from their research was an accuracy rate of over 80%, with a dataset of less than 600 patients. Our goal is to establish a classification model that can emulate their results and, through the use of artificial neural networks, be easily scalable to larger and more complex datasets. Using fundamental principles in feature engineering, we simplified the original biomarker panel to test the results of various feature combinations, and also to remove the large amounts of missing data, which would otherwise increase the risk of false positives/negatives. Composed of a control, benign, and malignant group, we trained several versions of the dataset in an Artificial Neural Network(ANN), which can utilize large amounts of complex data and is key to the scalability of the machine model.