**Early detection of pancreatic cancer through spectroscopic techniques.**

Kiman Park

**Problem:** In the United States, 1 in 2 men and 1 in 3 women will be diagnosed with cancer in their lifetime. The pancreatic cancer is the third-leading among all cancers in the United States. Also, survival rates largely depend on the stage of the cancer, which implies that early detection is enormously important. The detection of cancer usually involves harmful X-rays, expensive MRI/CAT scans, or painful, slow, and invasive biopsies to identify and confirm cancer. However, early detection is difficult with those techniques, since the cancer has not grown large enough to be detected. One of the potential solutions is to use Raman spectroscopy to detect exosomes, which is a marker for cancer. Raman spectroscopy is a powerful technique with the ability to detect small amounts of chemicals by using gold nanoparticles that amplify the electromagnetic field, which also amplifies the signal. It is a non-destructive and label-free technique. Also, Raman spectroscopy offers inexpensive and fast measurements and diagnosis. However, the blood has many components that vary by individual, which can obscure the proper identification of cancer. The publication of interest resulted in 90% accuracy with 97% specificity and 91% sensitivity for three markers and a control. These numbers could be better in order to serve as a primary detection method.

**Client:** Hospitals and general patients would be interested since the problem is expensive, slow, and/or requires a lot of labor.

**Data:** The data is from a published source in a journal, which can be acquired in Google Dataset. The format is in .txt file with intensities and frequencies as columns.

**Solution:** The solution involves various methods to improve accuracy, specificity, and sensitivity. We will begin with reproducing the results to benchmark the statistics and further use different models to improve those numbers. The models will include a random forest classifier, linear discriminant analysis, principal component analysis, and a customized neural network.

**Deliverables:** The results will be a Jupyter Notebook published on the GitHub page. This will include the code, graphs, and explanations.