PanCan Diagnosed (a miRNA Approach): Using Feature Selection, Ensemble Algorithms, and Interpretability for the Early Diagnosis and Personalized Medicine of Pancreatic Cancer

Goel, Siya (School: West Lafayette Junior/Senior High School)

The early diagnostic rate of pancreatic cancer (PC) is currently at just 9% as screening methods are unattainable, making it the fourth leading cause of cancer death. Many studies have achieved low accuracy (70-75%) as they use methods that do not take into account the 33% misdiagnosis rate of PC with other cancers. As a result, feature selection, ensemble algorithms, and interpretability techniques were used to find significant miRNAs, or biomarkers, in order to construct an early diagnostic tool (PanCan Diagnosis) for PC. In the first phase, recursive feature elimination algorithms were used to find 200 differentially expressed miRNAs in PC and no PC samples as well as early and late PC samples. In the second phase, an ensemble algorithm was constructed from K-Nearest Neighbor, Naive Bayes, Neural Network, and Logistic Regression models in order to diagnose PC and distinguish between early and late stages. In the third phase, XGBoost, SHAP, and Skater interpretability methods were used to find most significant miRNAs in model predictions. In the fourth phase, a user interface, PanCan Diagnosis, was designed to test if a person had no, early, or late stage PC and also displayed the patient's most differentially expressed miRNAs. This novel tool is the first in literature to receive a PC diagnostic accuracy of above 90%, seek miRNAs that can lead to personalized medicine of early and late stage PC samples, offers a ten-fold improvement in monetary costs, and is two times faster than current methods.

Awards Won:

Fourth Award of $500