A Novel Multi-Task Learning Model To Predict Pan-Cancer Drug Response With Integrated Multi-Omics Data

Agrawal, Shaurya (School: Adlai E. Stevenson High School)

According to the WHO, cancer resulted in 10 million deaths in 2020, a leading cause of death, and studies show that on average any particular class of drugs is ineffective for 75% of all patients. Thus it is evident that the selection of drugs and treatment is required to be individualized, for accurate treatment. This project aims to use multi-omics data, where various omics data types are integrated accurately and with high generalizability to predict response for a large breadth of drugs and cancers. This integration of various omics data allows for complex drug response-associated biological processes to be uncovered, which in turn allows for more accurate individualized treatment. In order to integrate the data accurately, pre-processing and data wrangling was performed then feature selection methods were performed, for the first model only basic feature selection methods were utilized. The first model, which was a Multi-Task elastic net model was then trained on this data and when tested it was overfitting due to the high dimensionality of data and was quite inaccurate. The second model was also a Multi-Task Elastic Net model trained on genes that were selected by more advanced feature selection methods such as Sequential Feature Selection and were trained on different classes of drugs for more generalizability. This model was more accurate than the previous model and showed signs of less overfitting, although still failed to meet the performance criteria. Even though the model had low accuracy, it was still able to accurately uncover overarching biological processes and showed signs of improvement in terms of accuracy and generalizability. For future research, the models will be using autoencoders for better dimensionality reduction and dealing with noise.

Awards Won: