Poster Abstract

Title: Diagnosing Leukemia – Collecting Flow Cytometry Data of Bone Marrow and Peripheral Blood from Potential Leukemia Patients

Presenter: Jiazhen Rong, Fourth Year Bioengineering/Biotechnology Student – UC San Diego

In order to diagnose whether a patient has leukemia, hospitals collect flow cytometry data of bone marrow and peripheral blood from potential leukemia patients. While current flow cytometry technique can detect up to 10 different biomarkers, the subsequent analysis of manual gating is subject to human errors and variability. In this study, our group aims to improve leukemia diagnosis and reduce manual variability via a machine learning pipeline. The data was processed by imputing the missing data, selecting important features and tuning hyperparameters. The machine learning classifiers, including random forest, SVM, XGBoost and logistic regression, could achieve an accuracy of around 75-90%. We also found that boosting and random forest are the best classifiers based on our dataset.