Machine Learning for Increasing the Response Rate of Breast Cancer Patients to Neoadjuvant Therapy

Cai, Janie (School: Miami Palmetto Senior High School)

Purpose: Patients with breast cancer (BC) who undergo neoadjuvant therapy (NAT) before surgery and achieve a pathologic complete response (pCR) after NAT have a low rate of recurrence. However, the pCR rate is low, averaging about 35% in the recent I-SPY2 clinical trial. The purpose of this project is to apply machine learning methods to predict pCR using gene expression and clinical data, thereby improving the overall pCR rate. Procedure: Gene expression and clinical data of BC patients undergoing NAT with 10 different drugs in the I-SPY2 trial were downloaded. Python scripts were written to process the data and to train and test a support vector machine (SVM) model that predicts pCR for each of the 10 drugs. The permutation importance method was used to determine the top genes that influence patients' response to NAT. Results: The accuracies of SVM models for predicting pCR are greater than 0.6 for all 10 drugs and are 0.789 and 0.795 for Pembro and TDM1/P, respectively. The pCR rate based on SVM prediction is increased by 14% to 32% compared to the rate in clinical trials. The top 5 genes identified by SVM models that influence the response to Pembro and TDM1/P are all involved in tumorigenesis or the regulation of cell growth. Conclusions: The SVM models trained using gene expression and clinical data provided high accuracy for predicting BC patients' pCR to NAT, which can be used to improve the pCR rate. They may assist doctors in prescribing personalized NAT for BC patients.

Awards Won: