Project No.	Project Title	
2022-01-107	Data-driven personalized treatment of inflammatory bowel disease	
Academic Advisor		**Co-Advisor**
Boaz Lerner		
Team Members		
Maya Tzemach		
Mayatzem@post.bgu.ac.il		

Abstract

Inflammatory bowel disease (IBD) is a chronic, relapse-remission disease, causing inflammation and stimulation in certain parts of the digestive system. The most popular treatments to maintain remission are immunomodulators and biological medications. Matching a treatment to a patient is a difficult task; some patient experience side effects and lack of effectiveness requiring them to change their treatment, sometimes more than once. By predicting the use of a drug, or its success in treatment, we can improve patients' quality of life and contribute to their clinical care.

In this project, I created a pipeline based on a random forest (RF) classifier using patients' medical records – lab tests, medication usage, and diagnoses – from the 5 years prior to the critical date of prediction, collected by the patients' health maintenance organization (HMO). The critical date is the diagnosis day for drug use prediction, and the first prescription of a drug for its success prediction. To address the temporality of the data, statistical moments for the features were extracted and used in a flattened model. After finding the best classifier and classification threshold (the percentage of trees in the forest classifying as true), splitting information of the RF was distilled to a single decision tree classifier to enhance explainability.

Prediction of the biological medication use had an f1 score of 0.42 and AUC score of 0.67, and that of the biological medication success had an f1 score of 0.66 and AUC score of 0.58. The immunomodulator use prediction had an f1 score of 0.55 and AUC score of 0.7 and that of the immunomodulator success prediction had an f1 score of 0.71 and AUC score of 0.63.

In conclusion, the project demonstrates that medical records, collected as part of patients' interactions with their HMO, can predict the use or success of using certain drugs for IBD, making it applicable for supporting clinicians in their decisions. Also, it points to specific medical concepts that are likely to have an impact on the prediction, whether it is an impact previously unknown to clinicians (e.g., cholesterol levels), or one previously known to clinicians (e.g., inflammation measurements) that has been validated by the findings.

Keywords: Inflammatory bowel disease, Area under curve, Machine learning, Medical concepts