[BIOINFO 2023]

Predicting adverse drug reaction signal from medical data via systematic data preprocessing and machine learning

Junhyeok Jeon[1], Eujin Hong[1], and Hyun Uk Kim[1,*]

[1]*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea*

**Corresponding author: ehukim@kaist.ac.kr*

Prescribing multiple medications in different combinations is a common practice to enhance the effectiveness of disease treatment. While this strategy aims to improve therapeutic outcomes, it can result in unintended adverse drug reactions (ADRs) due to interactions between the prescribed drugs. More than 30% of reported ADRs can be attributed to these drug-drug interactions (DDIs). Given the importance of both DDIs and ADRs, several computational models have been created to predict potential DDIs and ADRs. However, these computational models primarily rely on the molecular characteristics of drugs and often neglect clinical variables. Therefore, there is an opportunity to refine these computational models to further enhance the prediction of ADRs driven by multiple drug intake. To address this issue, we introduce a machine learning framework that predicts ADR signals potentially resulting from DDIs by using medical data. We first systematically preprocessed the MIMIC-IV dataset, which includes medical factors such as age, ethnicity, gender, patient disease history (ICD-10 codes), and drug information (including molecular structure and drug dosage). With this preprocessed dataset, we explored various machine learning methods to predict ADR signals, which are determined by classifying abnormalities in specific laboratory measurements (e.g., hematocrit, creatinine, and hemoglobin) The most successful model achieved an average AUROC of over 0.8. The machine learning model developed in this study is potentially valuable for assessing potential risks, such as ADRs, associated with the concurrent use of multiple drugs.