[BIOINFO 2024]

Development of Regression and Classification Models for Predicting CYP450 Enzyme Inhibition and Substrate Recognition

Ko Geon[1], Sung Kwang Lee[1]*

[1]Department of Chemistry, Hannam University
*Corresponding author: leesk@hnu.kr

Cytochrome P450 (CYP450) enzymes play a crucial role in drug metabolism in the human liver. When two or more drugs are administered simultaneously, the concentration of one drug in the body may increase or decrease. An increase in drug concentration can lead to unexpected side effects, while a decrease can reduce the drug's efficacy.

Previous studies used binary classification models to predict the inhibition (active/inactive) of five major CYP450 enzymes (1A2, 2C9, 2C19, 2D6, and 3A4). However, binary classification does not sufficiently reflect the degree of a compound's activity.

To evaluate the degree of activity, regression analysis was performed using XGBoost and multiple linear regression (MLR) to predict the activity score or LogIC50 for the five CYP enzymes, based on ChEMBL data and PubChem AID1851 data. Additionally, data from various sources were collected to develop an XGBoost model predicting whether a compound is a substrate or nonsubstrate for CYP enzymes, and this model was compared with existing ones.

For the LogIC50 prediction model, the cross-validation R^2 ranged from 0.206 to 0.555, while the AUC for the substrate/nonsubstrate prediction model ranged from 0.625 to 0.736.

This model is expected to be useful in identifying compounds that strongly inhibit CYP450 enzymes or recognize CYP enzyme substrates.