[BIOINFO 2024]

Machine Learning and Graph Neural Network-Based Prediction of Intestinal Permeability and Efflux Ratio : A Comparative Study Utilizing Caco-2 and MDCK Models

Na Yeong Koo[1], Sung Kwang Lee[1]*

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

The accurate prediction of intestinal permeability and efflux ratios is paramount for the assessment of drug absorption and pharmacokinetics during the early stages of drug development. This study aims to develop and compare tree-based machine learning models, including Random Forest (RF) and XGBoost, alongside graph-based deep learning models, specifically Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), for predicting intestinal permeability and efflux ratios. Data concerning Caco-2 permeability and efflux ratios (ER) data, along with MDCK ER data, were compiled from multiple sources. Molecular descriptors and graph representations of the compounds were employed for model training. Among the evaluated models, XGBoost demonstrated superior predictive performance across all evaluations of permeability and efflux ratio. In contrast, the GCN and GAT models exhibited comparable or lower diminished performance relative to the tree-based models. These finding the effectiveness of the efficacy of XGBoost in predicting drug absorption properties and its potential to minimize experimental dependencies, thereby enhancing the drug discovery process.