Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Computational Systems Biologoy (Shigehiko Kanaya (Professor))		
Student ID	2111417	Submission date	2023 / 7 / 25
Name	HONG RUIXUN		
Thesis title	Anti-inflammatory Activity Prediction of Plant Secondary Metabolites Based on Machine Learning Models Anti-inflammatory Activity Prediction of Plant Secondary Metabolites Based on Machine Learning Models		

Abstract

The therapeutic usage of plant secondary metabolites is currently a research focus for novel drug design, and plant secondary metabolites offer a vast array of lead active molecules to combat inflammation. In this study, we utilized machine learning-based approaches for predicting the anti-inflammatory potential of metabolites. We constructed datasets composed of 174 anti-inflammatory metabolites as positive samples, paired with an equal number of negative samples. The metabolites were represented using Extended-Connectivity Fingerprints as inputs for downstream machine learning algorithms, with a comparison made against MACCS fingerprints. The incorporated machine learning models, including DNNs, SVM, RF, and XGBoost, were compared in terms of predictive performance, and subsequently used to identify metabolites candidates with anti-inflammatory potential. Furthermore, Tanimoto similarity filtering was used to refine the dataset. As a result, the 1024-bit Extended-Connectivity Fingerprints representation with a radius of 2 demonstrated superior performance compared to other representations across most models. Among the tested models, DNNs displayed the best overall performance, while XGBoost performed less effectively. Notably, the utilization of Tanimoto similarity filtering significantly enhanced the model's predictive capability.