Abstract

People exposed to countless chemical compounds throughout their lives, many of which have the potential to be hazardous. This chemical interaction has become an integral aspect of our daily lives. Living in a highly reactive chemical environment entails interaction with various elements, ranging from the food we consume and prescribed medications to the cosmetics we use and even the air we breathe. When developing new medications, toxicity becomes a major concern because over 30% of drug candidates have toxic effects that are not discovered during clinical trials. The toxic potential of chemicals and their combinations must be proven experimentally. Therefore, minimizing our exposure to dangerous substances in common products requires an understanding toxicity of chemical. This study used the simulated annealing algorithm and the ensemble method to predict the toxicity of a case study involving the Nuclear Receptor-Aryl Hydrocarbon Receptor toxicity type. The Tox21 Data Challenge provided the dataset used in this investigation. The simulated annealing algorithm used in the feature selection process to develop a prediction model. Ensemble method was using to build a prediction model with three methods: Random Forest, Adaptive Boosting (AdaBoost), and Extreme Gradient Boosting (XGBoost). The best model obtained from Random Forest model with accuracy value of 0.9806 and an F1-Score of 0.9808.

Keywords: Simulated Annealing Algorithm, Ensemble, NR-AHR, Toxicity Prediction, Machine Learning