Abstract

This research aims to implement the Firefly-Ensemble algorithm in predicting the toxicity with the case study of the Androgen Receptor Ligand Binding Domain (AR-LBD). Toxicity is the main factor causing drug failure in clinical trials, and accurate toxicity prediction is crucial for optimizing the drug development process. This study uses the Firefly Algorithm for feature selection and Ensemble methods, which include Random Forest, AdaBoost, and XGBoost, for model parameter optimization. The research results show that the use of the Firefly Algorithm for feature selection can improve the model's accuracy on several models, especially XGBoost, with an accuracy of 99.30%, an F1-score of 99.31%, a recall of 99.88%, and a precision of 98.75%. Additionally, the application of parameter optimization using the Ensemble method also contributes to the improvement of model performance, with XGBoost showing the best performance, followed by Random Forest and AdaBoost. These results indicate that the combination of the Firefly-Ensemble algorithm improves the accuracy of toxicity prediction, thus it can be used to support the development of in silico methods in toxicology and drug discovery.

Keywords: Machine Learning, Firefly Algorithm, Ensemble, AR-LBD, Toxicity