Abstract

Drug toxicity prediction is an important aspect of bioinformatics research to reduce the risk of unwanted side effects in humans. However, the high complexity of toxicity data demands an efficient processing method for the prediction model to work optimally. This study uses a combination of Artificial Bee Colony (ABC) as a feature selection method and Support Vector Machine (SVM) as a classification algorithm to improve accuracy in toxicity prediction of Androgen Receptor-Ligand Binding Domain (AR-LBD) cases. The feature selection method using ABC aims to reduce data dimensionality without reducing important information, while SVM is used with three types of kernels (Linear, Polynomial, and RBF) to compare model performance. In addition, hyperparameter tuning is performed using Grid Search to find the best combination of parameters to improve model performance. Tests were conducted by evaluating accuracy, precision, recall, and F1-score on test data. The results showed that ABC successfully reduced the number of features without degrading model performance, with the RBF Kernel selecting 77 features and having a best fitness of 0.123, making it the most optimal in feature selection. In addition, Grid Search improved the accuracy of the model, with the Kernel RBF improving from 0.958 to 0.971 after parameter tuning. Model validation showed that the RBF Kernel performed the best, with accuracy 0.97. From the results of this study, it can be concluded that the combination of ABC and SVM with RBF Kernel is the best approach in AR-LBD toxicity prediction, because it succeeds in improving model accuracy with optimal feature selection and effective parameter tuning.

Keywords: Artificial Bee Colony, Support Vector Machine, Feature Selection, Hyperparameter Tuning, AR-LBD.