

ABSTRACT

Dengue fever is one of the endemic diseases transmitted by the dengue virus and still often spread in several countries, especially developed countries with tropical climate. There is no specific drug that can be used to treat dengue until now. Research on one of clinically investigated drugs, namely balapiravir, still not effective in inhibiting the activity of dengue virus replication. Non-structural protein 3 (NS3) can be used as a target for the development of drugs to find new compositions that can work better. The quantitative structure activity-relationship (QSAR) model can be used because this model is considered to be valid to predict and classify the biological activity of compounds that have not been tested. This study aim to build a QSAR model to predict the activity of NS3 inhibitors. Classification process was performed by using feature importance analysis for feature selection and ensemble methods with random forest, adaptive boosting and extremely randomized trees algorithm to build a prediction models. Adjustment of the hyperparameter tuning procedure is done to help increase the performance of the model. Based on the validation analysis, model 9 produces the best accuracy of 0.73 (73%) and AUC of 0.82 compared to other models. This model also not related to coincidental correlation. Its proved by 10 times of y-scrambling experiment obtained lower MCC value than actual MCC result.

Keywords: non-structural protein 3 (NS3), quantitative structure activity relationship (QSAR), ensemble methods, random forest algorithm, adaptive boosting (AdaBoost) algorithm, extremely randomized trees algorithm