Abstract

Coronavirus disease 2019 (COVID-19) is a type of disease that emerged in the last two years caused by the Sevete Acute Respiratory Syndrome Coronavirus 2 or SARS-CoV-2. COVID-19 has infected more than 215 million people and more than 4.47 million people died as of 27 August 2021. So far, no drug has been received or found that can treat and cure COVID-19 patients, despite the RT-PCR protocol. and standard DNA sequencing has been developed for diagnostic purposes. Regarding the urgency of drug design, one of the functional proteases of SARS-CoV-2 was used, namely papain-like protease (PLpro) as a drug target. Among the in silico drug discovery methods, Quantitative Structure Activity Relationship (QSAR) is one of the methods used because it has been used successfully to predict and classify the biological activity of untested compounds. This study aims to build a QSAR classification model for in-house molecules as inhibitors of the papain-like protease (PLpro) SARS-CoV-2. The classification process is carried out using the SVM (Support Vector Machine) method which is first performed with feature reduction using a genetic algorithm. The dataset used is the bioactivity label and feature descriptors in the form of PubChem and Extended fingerprints which are calculated based on the molecular structure. The number of features is reduced by eliminating features that have variances < 0.1 and < 0.2. The results showed that the best results were obtained through the implementation of the Polynomial kernel SVM on the PubChem and VarianceThreshold 0.2 descriptors with accuracy and F1-Score values of 70.370% and 63.636%, respectively.

Keywords: *Papain-like Protease* (PLpro), SARS-CoV-2, *COVID19, quantitative structure activity relationship* (QSAR), *genetic algorithm, support vector machine (SVM)*