World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:18, No:04, 2024

Using Combination of Different Sets of Features of Molecules for Improved Prediction of Solubility

Authors : Muhammet Baldan, Emel Timuçin

Abstract : Generally, absorption and bioavailability increase if solubility increases; therefore, it is crucial to predict them in drug discovery applications. Molecular descriptors and Molecular properties are traditionally used for the prediction of water solubility. There are various key descriptors that are used for this purpose, namely Drogan Descriptors, Morgan Descriptors, Maccs keys, etc., and each has different prediction capabilities with differentiating successes between different data sets. Another source for the prediction of solubility is structural features; they are commonly used for the prediction of solubility. However, there are little to no studies that combine three or more properties or descriptors for prediction to produce a more powerful prediction model. Unlike available models, we used a combination of those features in a random forest machine learning model for improved solubility prediction to better predict and, therefore, contribute to drug discovery systems.

Keywords : solubility, molecular descriptors, machine learning, random forest
Conference Title : ICBBE 2024 : International Conference on Bioinformatics and Biomedical Engineering
Conference Location : London, United Kingdom
Conference Dates : April 11-12, 2024