Implementation of Hybrid Bat Algorithm-Ensemble on Human Oral Bioavailability Prediction of Drug Candidate

Muhamad Farhan Wirasantoso[1], Hasmawati[2], Isman Kurniawan[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]farhanwira@students.telkomuniversity.ac.id,
[2]hasmawati@telkomuniversity.ac.id,[3]ismankrn@telkomuniversity.ac.id

Abstract

One of significant parameters of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) is Human Oral Bioavailability (HOB) which is crucial for determining the total of consumed drugs inside humans body circulation. Poor HOB results in undeterminable drug effects in the human body, with approximately 50% of drug candidates failing due to low oral availability. As many as 80% of drugs in the world use the oral route of entry into the body, so HOB prediction is very important to reduce side effects and the risk of toxicity brought by drugs. Unfortunately, oral bioavailability is currently predominantly measured in vivo consequently, developing in-silico methods is considered crucial. To reckon the human oral bioavailability of medication candidates, we used the Hybrid Bat Algorithm method for feature selection and the Ensemble method, i.e. Random Forest, AdaBoost, and XGBoost for the prediction model. The result showed that XGBoost as the best model in which the value of accuracy and F1-score were 0.776, and 0.802, respectively.

Keywords: Absorption Distribution Metabolism Excretion Toxicity (ADMET), Drugs, Machine Learning, Human Oral Bioavailability (HOB), Hybrid Bat Algorithm, Ensemble.

1. Introduction

The Background

Drug development failure frequently occurs in its final stage. It caused by several factors, one of which is Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) testing [1]. ADMET is an important parameter to optimize the rate of drug discovery success and minimize risks. Human Oral Bioavailability (HOB) is one of the ADMET parameters that is important for measuring the total of drugs that enter the humans body circulation after being consumed. The HOB parameter, simply put, is checking the amount of drug that enters through the human mouth into the body circulation, which is generally the humans own blood, which will then deliver the drug to a place where the blood (which has been contaminated with drug substances) exerts a pharmacological effect through the systemic circulation. Poor HOB can lead to undeterminable drug reactions inside human body, where around 50% of drug candidates fail because of low oral availability [2], [3]. This data shows the lack of effectiveness of HOB as a main parameter tool for checking oral drugs.

As much as 80% drugs around the world are commonly taken into the body orally [4], so HOB prediction is very important in order to reduce side effects and the risk of toxicity brought by drugs. The fact that oral administration is the usual method of delivering drugs to the systemic circulation for patients, HOB represents one of the most significant pharmacokinetic parameters with ADMET features [5]. Currently, many drug candidates are introduced into the human body using the oral route. Oral bioavailability is indisputable for all drugs administered orally. The oral route is among the most effective in delivering drugs into the systemic circulation, which explains why. In fact, in vivo method is still commonly used to measure oral bioavailability; thus, the importance of developing in-silico methods is inevitable. Such as Machine Learning which can predict oral bioavailability quickly and has more effectiveness compared to in vivo measurements [6].

Studies has been carried out related to HOB prediction using diverse machine learning methods. One of them is that in 2006, Francesco Archetti, et al conducted study on empirical studies of several well-known Machine Learning techniques related to the Human Oral Bioavailability of Drugs problem, using Genetic Programming with various version, one of which is GP (Genetic Programming). Francesco Archetti also used other methods namely Artificial Neural Networks (ANN), Feature Selection, Linear and Least Square Regression, and Support Vector Machines Regression (SVMR). The results of the study show that all versions of GP outperform other machine learning methods for both correlation coefficient and Root Mean Squared Error (RMSE) [7].

In 2020, Gabriela Falcón-Cano also conducted study on ADME Prediction (Absorption, Distribution, Metabolism, and Excretion) with Konstanz Information Miner (KNIME). The result is a machine learning approach to the ensemble model, obtaining good balanced accuracy [8]. In 2021, Urban Fagerholm and his teammates also conducted study on Predicting the Oral Bioavailability of Drug Candidates in Humans with a New Machine Learning Methodology, using the Partial-Least Squares (PLS) Model and Support Vector Machine (SVM) methods. The result is efficiency regarding costs, productivity, and time [9].

We set the current study's goal to implement the Hybrid Bat Algorithm-Ensemble Method in estimating the Human Oral Bioavailability of drug candidates. Hybrid Bat Algorithm (HBA) was used as a method for selecting features, because HBA is a method that discusses bat echolocation in producing an algorithm. Therefore, HBA is a solution to be applied in this study [10]. Meanwhile, the advantage of the HBA method is the faster convergence speed. This ability is based on Bats or bats having an echolocation technique that is applied in this method. Furthermore, we used the Ensemble method for parameter optimization in this study. Initially the ensemble was developed to reduce variance, from this development the ensemble can increase the accuracy of the automatic decision making system [11].

Goals

The purpose of this study is to make the Feature selection of Hybrid Bat Algorithm as the main Feature selection, if the study results can exceed the standard of Feature selection commonly used today. The expected result of this study is that the method used can meet the requirements of the effectiveness of HOB prediction of drug candidates using the Hybrid Bat Algorithm-Ensemble method, so that this method can be a reference for the development of in silico methods to replace in vivo methods that can help increase the effectiveness of HOB prediction and also save time and costs using machine learning.

2. Related Works

In 2006, Francesco Archetti, et al. conducted research on the empirical study of several well-known Machine Learning techniques related to the Human Oral Bioavailability of Drugs problem, using various versions of Genetic Programming, one of which is GP (Genetic Programming). Franceco Archetti also used other methods such as Feature Selection, Linear and Least Square Regression, Artificial Neural Networks (ANN), and Support Vector Machines Regression (SVMR). The result of his research is that all versions of GP outperform other machine learning methods for both Root Mean Squared Error (RMSE) and correlation coefficient [7]. Marwin H. S. Segler and his colleagues in 2017 have conducted research on drug discovery with Recurrent Neural Networks. The result is that machine learning is not the only way for drug discovery, but in this study Marwin et al believe that deep neural networks can complement established approaches in drug discovery [12].

In 2019, Hao Lou and Michael J. Hageman conducted research on predicting subcutaneous human monoclonal antibody bioavailability. The database uses mAb SC by using Multi-layer perceptron (MLP), Gaussian Naïve Bayes (GaussianNB), K Nearest Neighbor (kNN) algorithm learning, and dimensionality reduction is performed using principal component analysis (PCA). The results found in this study are that the product of mAb varies from 35% to 90%, and also some of the algorithms used in this study provide acceptable prediction accuracy [13]. In 2020, Gabriela Falcón-Cano et al made a study also about Prediction (Absorption, Distribution, Metabolism and Excretion) ADME with Konstanz Information Miner (KNIME). The result is a machine learning approach to the ensemble model, obtained good balanced accuracy [8]. In 2021 Urban Fagerholm and his team also conducted research on Predicting the Oral Bioavailability of Drug Candidates in Humans with a New Machine Learning Methodology, using the SVM and PLS Model methods. As a result, efficiency regarding cost, productivity and also time [9].