

Abstract

Abdelrahman I. Saad

Predicting Drug Target Interaction by Integrating Drug Fingerprint and Drug side effects Using Machine Learning

Drug discovery is an important step before drug development. Drug discovery is the process of identifying, testing a drug before medical use. Drugs are used to cure diseases by interacting with the target, which is the protein in the human cells. Identification of interactions between drugs and targets has a vital role in drug discovery. The very high cost of the experiments and few available public drug-target datasets make it important to develop accurate computational models, which could precisely detect the interaction between drugs and targets. Drug-target interaction prediction is an important solution to accelerate the process of drug discovery. Predicting drug-target interaction using computational approaches is an important step in drug discovery and repositioning, to predict whether there will be an interaction between a drug and a target. Cancer is one of the most influential factors causing death in the world. Humans may have cancer due to biological factors (inherited genes), exposure to radiation (X-ray radiation) and bad lifestyle habits such as smoking. Adenosine which is a molecule, found in all human cells by coupling with G protein it turns into adenosine receptor. Adenosine receptor is an important target for cancer therapy. Adenosine stops the growth of malignant tumor cells such as lymphoma, melanoma and prostate carcinoma. Target (adenosine) can be activated through interacting with drugs (compounds) to stop the tumor cells from spreading and cure the cancer disease. This research aims to predict drugs and potential drug candidates that interacts with adenosine receptor. There are two proposed frameworks in this thesis; a machine-learning model is implemented to predict drug target interaction. Drug fingerprint and side effect were used as input features to train first framework. Three different experiments were conducted using fingerprint, side effect and drug side effect integrated to drug fingerprint. Results showed improvement in prediction when integrating drug side effect to drug fingerprint. K-Nearest Neighbors scored best results in the three experiment with an average accuracy of 94.69%. The second framework goal was to predict drugs interacting with adenosine receptors using three machine-learning techniques with Synthetic Minority Oversampling Technique. We used Synthetic Minority Oversampling Technique since the dataset was imbalanced as the number of drugs interacting with adenosine molecule is relatively small compared to the non-interacting drugs. This problem affected our classification performance results. After conducting experiments before and after using Synthetic Minority Oversampling Technique, Random Forest achieved the best classification performance with accuracy 75.09%. Since drug side effect is one of the main reasons of drug design failure, we ranked the predicted drugs based on their corresponding side effects to mark its safety and eligibility for use.