Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220612

Extreme Gradient Boosting Based Improved Classification of Blood-Brain-Barrier Drugs

Manuskandan SUBHA RAMAKRISHNAN [a], and Nagarajan GANAPATHY [b,1]
[a] *Department of Chemical Engineering, ACT Campus, Anna University, Chennai, India*
[b] *PLRI Institute of Medical Informatics of TU Braunschweig and Medical Hannover School*

Abstract. In this study, the analysis based on boosting approach namely linear and tree method are explored in extreme gradient boosting (XGBoost) to classify blood brain barrier drugs using clinical phenotype. The clinical phenotype features of BBB drugs are Public available SIDER dataset. The clinical features namely drug's side effect, drug's indication and the combination is fed to XGBoost. Results shows that the proposed approach is able to discriminate BBB drugs. The combination of XGBoost with tree boosting is found to be most accurate (F1=78.5%) in classifying BBB drugs. This method of tree boosting in XGBoost may be extended to access the drugs for precision medicine.

Keywords. Drug Discovery, Extreme gradient Boosting, Blood-Brain-Barrier-Drugs

1. Introduction

Machine learning methods is commonly used to investigate the drug discovery and drug re-purposing for improvised precision medicine. Blood brain barrier (BBB) permeability based drugs are highly important for neurological disorder prevention cure [1,2]. However, characterization and identification of BBB based drugs using clinical phenotypes is highly challenging [1]. Recently, extreme gradient boosting methods have been reported for improved classification for multi-set features.In this work, an extreme gradient boosting (XGBoost) approach in drug prediction of BBB permeability using clinical phenotype are explored.

2. Materials and Methods

2.1. Materials (Database)

For this study, the data-set reported in Doniger et al. [2], is considered the dataset contains 91 samples in total with 38 samples are characterised as BBB permeability true and the

[1]Corresponding Author: Nagarajan Ganapathy, PLRI Institute of Medical Informatics of TU Braunschweig and Hannover Medical School, Germany; E-mail: nagarajan.ganapathy@plri.de.

53 samples are characterized as BBB permeability false. Each samples have features namely side effect (SE) and Indicators (Ind). It uses SIDER datasets (SIDER) which have been proved the BBB permeability true or false in the clinic [1,2].

2.1.1. Method

Clinical phenotype features of BBB drugs namely SE and Ind are fed to XGboost. It contains deviation of the model and regular term to prevent over-fitting as objective function with optimization [3]. The result of the sample is predicted by set of regression tree with weights of the leaves and number of leaves. Regularization is considered to smooth the final learning weights with reduced over-fitting [1,3]. In this study, learning rate is set 0.01, boosting algorithm is set to linear and tree, objective function is set to hinge. Stratified k-fold cross validation technique (k= 5) is used for reliable outcome with imbalance datasets. Four performance metrics namely precision (Pr), recall (Rc), F1-measure (F1), and area under the curve (AUC) are used to evaluate the performance of XGBoost [3].

3. Results and Discussion

The performance of XGBoost in discriminating the BBB drugs for linear and tree boosting is presented in Table I. Tree boosting approach obtained the highest performance of F1 = 78.5% . Both the tree and linear boosting yielded the highest recall of 94.3% for Ind features. XGBoost with tree classifiers yields highest AUC of 81.9% for combination of SE and Ind) features. Besides Tree method, XGBoost yielded the best AUC of 79.1% for combined features. For tree boosting method, XGBoost obtained higher than 75.0% Acc for SE, Ind and its combined features.

Table I: The performance of the XGBoost for linear and tree boosting methods

Metric	Tree			Linear		
	SE	Ind	SE + Ind	SE	Ind	SE + Ind
Pr	71.8	94.3	77.2	76.6	47.1	79.2
Rc	75.4	54.6	81.1	73.9	94.6	73.6
F1	72.9	67.2	78.5	74.2	62.6	75.7
AUC	77.3	75.5	81.9	78.4	58.7	79.1

4. Conclusion

The proposed methods are found to be capable of handling the multi variant features of drug identify BBB permeability.

References

[1] Miao R, Xia LY, Chen HH, Huang HH, Liang Y. Improved classification of blood-brain-barrier drugs using deep learning. Scientific reports. 2019 Jun 19;9(1):1-1.

[2] Doniger, S., Hofmann, T. Yeh, J. Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms. Journal of Computational Biology 9, 849–864 (2002).

[3] Qiu Y, Zhou J, Khandelwal M, Yang H, Yang P, Li C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Engineering with Computers. 2021 Apr 22:1-8.