International Journal of Scientific Research and Engineering Development-– Volume 3 - Issue 5, Sep - Oct 2020

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

PREDICTION OF DRUG-TARGET INTERACTION USING BIPARTITE LOCAL MODELS AND K-NEAREST NEIGHBOR

R. Kowsalya[1]
Post Graduate Scholar
Computer Science and Engineering
Government College of Engineering
Tirunelveli
kowsalyark4697@gmail.com

Dr. E. Siva Sankari[2]
Assistant Professor
Computer Science and Engineering
Government College of Engineering
Tirunelveli
sivasankari@gcetly.ac.in

Abstract:

Computational prediction of drug-target interaction is an essential task with various applications in the pharmaceutical industry. The drug-target interaction is being evaluated using biochemical validation of hypothesized drug interaction which is considered time-consuming, laborious and expensive. To overcome the restrictions of the traditional approach there are machine learning based drug-target interactions prediction methods. In this work, the Bipartite Local Model (BLM), one of the most prominent machine learning technique is used in interaction prediction. In particular, BLM with a hubness-aware regression technique and k-nearest neighbor method is used. Similarity between drugs and target is identified using Tanimoto coefficient and Smith Waterman score method respectively. The results are then processed using the BLM to predict the interaction between the drug and the target. In this method, the interaction with multiple set of features like Chemical, genomic and interaction features are predicted. Then, hubness-aware regression is used to rectify the error that may occur while applying BLM. KNN is then applied on to the results obtained through BLM so as to identify the value and understand the nearest of the interaction predicted using BLM. Finally the accuracy of the drug-target interaction is evaluated and also the drug-target pair is identified.

Keywords — **Drugs, Targets, Drug-Target Interactions, Similarities, Error Correction**.

I. INTRODUCTION

Drug Target Interaction (DTI) defines association or reaction of drugs on target. Identifying potential interactions between drugs and targets is crucial for modern drug discovery and repurposing. Chemical compounds and protein sequences are collected from KEGG database. Similarity between drugs and target is identified using Tanimoto coefficient and Smith Waterman score method respectively. In this method, the interaction with multiple set of features like Chemical, genomic and interaction features are predicted. The interactions between drugs and targets is accepted or ignored according to its similarities. K-Nearest Neighbor is used to predict the drug-target pairs.

II. RELATED WORK

Jian-Ping Meiet et al. [1] used a supervised learning approach called Bipartite Local Model (BLM) is employed for effective prediction of drug-target interactions. However, the "local" model is

not capable of making decisions for drug-candidate compounds or target-candidate proteins that currently have no known interactions available, and therefore BLM can fail to make correct prediction when involving such new candidates.Thus an intuitive solution to the new candidate problem of BLM by integrating a NII procedure was proposed. It infers training data from neighbour's interaction profiles. Through systematic experiments with benchmark datasets, the effectiveness of BLM-NII for predicting interactions between new drug-candidate compounds and new target-candidate proteins was demonstrated allowing all neighbours to participate in training data inferring. To allow only neighbours with large similarities to contribute, a threshold was used to reduce the impact of those non-important neighbours to 0. Alternately, a Gaussian function was introduced to gradually decrease the influence of neighbors based on their distances to the new drug-target candidate in query. In the current work, only the NII procedure is applied for those completely new candidates that have no existing training data at all, and the results were good enough to show the usefulness of NII. As it is very common for drugs to activate or inhibit only a small number of targets and targets only to be activated or inhibited by very restricted drugs, the NII technique may also be applied to drugs and targets that do not have adequate training data. Four more accurate prediction models may be built by using neighbour's information to enhance the limited training examples. However, too much emphasis on neighbors tends to eliminate the local characteristics of each drug and target and could cause deterioration in the prediction performance. Moreover, the balance between local information and global information in model learning can also be explored.

. F. Cheng et al. [2] used drug-target interaction (DTI) is the basis of drug discovery and design. Visualization of network of associations between drug-target, target-disease and disease gene could provide useful information to detect new therapeutic indications or adverse effects of old drugs. If the protein is a known target of that drug, a link is placed between a drug node and a target node. The size of the drug node is the proportion of targets the drug has with established experimental evidence.

Y.Wanget al. [3] declared the silico prediction of drug-target interactions plays a major role in the detection and development of new applications of illegal or discarded drugs. Our approach uses an RBM model to effectively encode multiple sources of information about DTIs and accurately predict different types of DTIs, such as drug-target relationships or drug modes of action. Tests on two public databases showed that our algorithm can achieve excellent prediction performance with high AUPR scores.

III. PROPOSED FRAMEWORK

In this approach, the benchmark dataset containing chemical compound and protein sequence. Chemical compound structures are collected from KEGG LIGAND database. Protein sequences are collected from KEGG GENES database. The input has four components which are to processed are enzyme, ion channels, G-protein coupled receptors and nuclear receptors. The first step is preprocessing. This process involves the modification of predefined codes incases of negative values much beyond the limit. It is also used to ignore the negative, exponential terms involved. If there any negative or exponential terms are involved, normalization to zero has to be done. The second step is to find the similarities in chemical components and smith-waterman score method to find the similarities in protein target. Bipartite local models are used. It is an architecture which involves two levels of prediction.

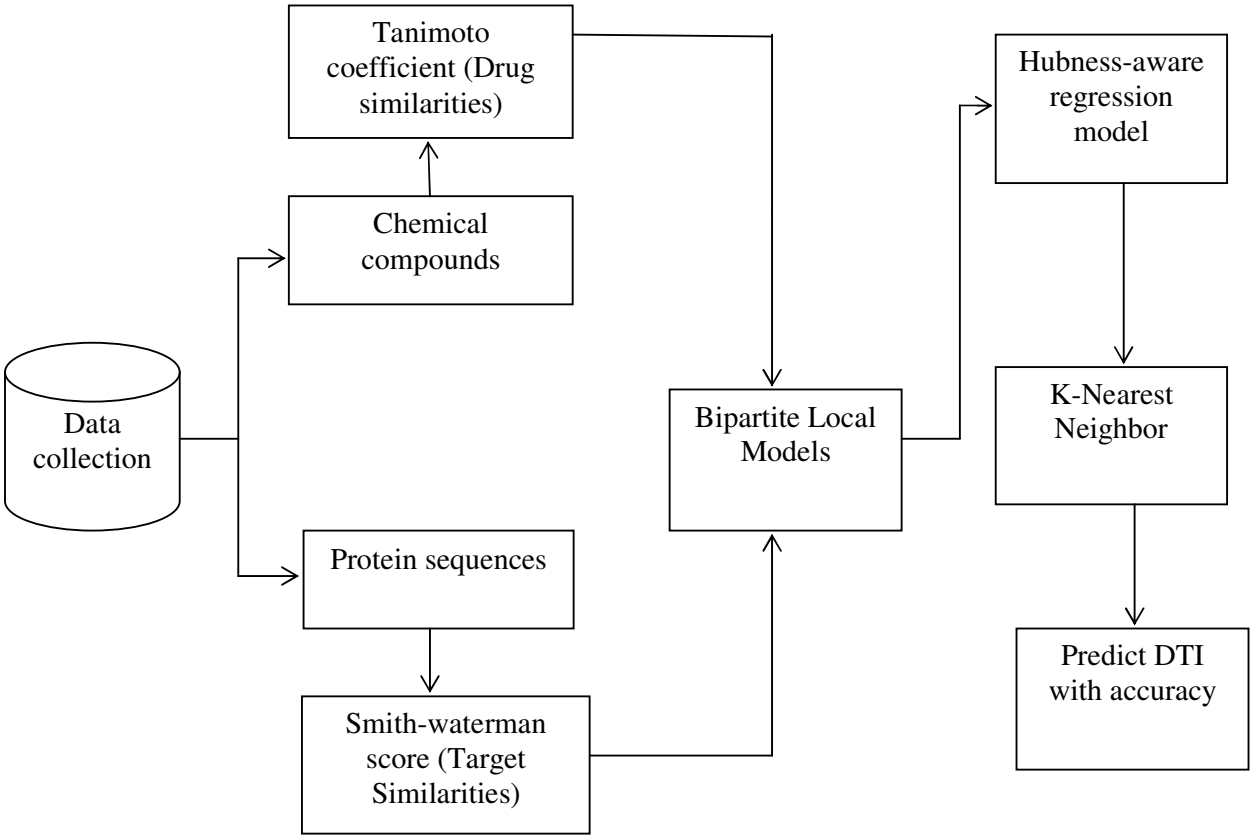

Fig 1 Overall system Architecture

The first level of prediction uses weighted method and the first level is predicted. K-nearest neighbor method is applied along with error correction. Hubness-aware regression is used as error correction method along with K-nearest neighbor method. The K-nearest neighbor method will provide the prediction of dreug-target interaction. The interactions of drugs and targets are carried out and the features available are identified. The required Drug Target Interactions (DTI) are obtained. Fig 1 represents overall work flow of the proposed system. The proposed system consists of five modules.

1. Data Preprocessing
2. Identification of tanimoto coefficient for chemical compounds and smith-waterman score for protein sequences.
3. Determining Drug-target interaction using bipartite local models.
4. Error correction by using hubness-aware regression.
5. Predicting efficient DTI using k-nearest neighbor and evaluating its accuracy.

IV. METHODOLOGY

A. Preprocessing

The first module is pre-process. There are two tasks are involved.

1) Identify and remove empty cells in the data.
2) Identify and remove negative data.

B. Similarity Measures

The entire document should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes. second module is to find the similarities. In this method, drug-drug similarity is computed using tanimoto coefficient and target-target similarity is computed using smith-waterman score method.

1) ***Tanimoto Coefficient:*** Tanimoto coefficient method is used to process the chemical compounds. The chemical structure similarity among drugs was obtained with SIMCOMP. Chemical compounds are denoted as graphs. Calculate the similarity score according to the number of the common substructures between two compounds by using tanimoto coefficient.

The tanimoto similarity occurs only for a binary variable, and the tanimoto coefficient ranges from 0 to +1 (where+1 is the highest similarity) for binary variables.

The chemical structure similarity between two compounds d_i and d_j can be calculated as,

$$Sim_{Stru\ Drug}(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|}$$

The chemical structure similarity matrix of drug compounds is described as $Sim_{StruDrug}$. It is the ratio between the modulus of the intersection value to the modulus of the union value of the respective drug compounds (d_i and d_j).

2) ***Smith-waterman Score:*** Smith-water man score method is used to process the target protein sequences. The protein sequence similarity among targets was obtained by the normalized version of the Smith waterman score. Protein sequences are denoted as structures. Calculate the similarity score according to the number of the common substructures between two targets by using Smith-waterman score method.

The protein sequence similarity between two targets t_c and t_d can be calculated as,

$$Sim_{seqTar}(t_c, t_d) = \frac{SW(t_c, t_d)}{\sqrt{SW(t_c, t_c) SW(t_d, t_d)}}$$

The protein sequence similarity matrix of target proteins is described as $Sim_{seqTar}(t_c, t_d)$. In which $SW(t_c, t_d)$ denoted as the canonical smith waterman score between the target proteins t_c and t_d..

C. Bipartite Local Models

Bipartite local models consider the drug-target interaction prediction problem as a link prediction problem in bipartite graphs. The model computes two independent predictions that are aggregated subsequently. Prediction is obtained by arranging the predictions as local models and calculated using the formula,

$$y(d_i, t_j) = \frac{y_1(d_i, t_j) + y_2(d_i, t_j)}{2}$$

D. Hubness-aware Regression

This method describes the components of the method, such as the representation of drugs and targets and the projection-based ensemble. This method is used for error correction for the output.

Algorithm for Hubness-Aware Local Models

Require : Drug-target interaction matrix I, drug-drug similarity matrix S_D,target-target similarity matrix, number of nearest neighbors k, ensemble size N, number of selected features F_D ,F_T.

Ensure : Likelihood of drug-target interactions.

1: D ← enhanced similarity-based representations of drugs.
2: T ← enhanced similarity-based representations of targets.
3: for l = 1…N do
4: D' ← random sample of D with F_D features

5: $T' \leftarrow$ random sample of T with F_T features

6: Predict interaction scores with BLM as local model.

7: end for

8: Aggregate the predictions that were made in each execution of the loop.

E. K-Nearest Neighbor Method

K-Nearest neighbor method is employed for predicting the numeric label on an instance x with k-nearest neighbor regression. The k-nearest neighbors of x (k most similar instances to x) are determined. Average of their labels is calculated as the predicted label of x.

1) ***Corrected Label:*** K-nearest neighbour is to calculate the corrected labels of training instances. Define the corrected label $y(x)$ of a training instance x as

$$y(x) = \begin{cases} \frac{1}{|R_x|} \sum_{x_i \in R_x} y(x_i) & |R_x| \geq 1 \\ y(x) & otherwise \end{cases}$$

where,

$y(x)$ → original label of instance x.

R_x → set of "reverse neighbors.

2) ***Predicted Label:*** Given a "new" (unlabeled) instance x', its predicted label $y(x')$ is calculated as follows:

$$y(x') = \frac{1}{k} \sum_{x_i \in N(x')} (x_i)$$

where the set of closest neighbours of x' denotes $N(x')$.

V. EXPERIMENTAL RESULTS

The drug-target interaction is predicted among enzyme by using these modules. The modules are drug-drug similarities, target-target similarities drug index, target protein index, drug-target interaction, binary relation of enzyme, drug-target pair and prediction of new interactions.

A. Drug-drug Similarities

The similarities among Enzymes are calculated for determining interactions of drugs. This is done using Tanimoto coefficient and is shown in Fig 2.

	D00002	D00005	D00007	D00014
D00002	1	0.515625	0.038462	0.084746
D00005	0.469697	1	0.032787	0.073529
D00007	0.038462	0.032787	1	0.428571
D00014	0.084746	0.073529	0.428571	1
D00018	0.098039	0.083333	0.1	0.066667
D00021	0.12	0.083333	0.375	0.230769
D00027	0.083333	0.109091	0	0

Fig. 2 Drug-drug similarities for enzyme

B. Target-target Similarities

The similarities among Enzyme are calculated for determining interactions of targets. This is done using Smith-waterman score and is shown in Fig 3.

	hsa10	hsa100	hsa10056
hsa10	1	0.025752149	0.021574732
hsa100	0.025752149	1	0.018325241
hsa10056	0.021574732	0.018325241	1
hsa1017	0.019324722	0.025940168	0.016284648
hsa1018	0.026672482	0.027021353	0.020770731
hsa10188	0.015293492	0.018789185	0.008853849
hsa1019	0.023485713	0.020569892	0.017779196

Fig. 3 Target-target similarities for enzyme

C. Drug Index

The drugs are collected and the numeric values are assigned to the collected drugs in ascending order. This drug index is shown in Fig 4.

1	1	D00002
2	2	D00448
3	3	D00037
4	4	D00155
5	5	D00021
6	6	D02880
7	7	D01441
8	8	D00279
9	9	D00043
10	10	D00160
11	11	D00039
12	12	D00065
13	13	D00136

Fig 4 Drug Index for enzyme

D. Target Protein Index

The target proteins are collected and the numeric values are assigned to the collected target proteins in ascending order. This target index is shown in Fig 5.

1	1	hsa:10
2	2	hsa:100
3	3	hsa:10056
4	4	hsa:1017
5	5	hsa:1018
6	6	hsa:10188
7	7	hsa:1019
8	8	hsa:1020
9	9	hsa:1021
10	10	hsa:1022
11	11	hsa:1024
12	12	hsa:1025
13	13	hsa:10269

Fig 5 Target Protein Index for enzyme

E. Drug-Target Interaction Prediction

The interaction among drugs and targets is measured using bipartite local models by considering similarity determined in previous stage. The following figure depicts the estimated drug-target interaction of the enzymes.

	D00002	D00005	D00007	D00014	D00018	D00021
hsa10	1	0	0	0	0	0
hsa100	0	0	0	0	0	0
hsa10056	0	0	0	0	0	1
hsa1017	0	0	0	0	0	0
hsa1018	0	0	0	0	0	0
hsa10188	0	0	0	0	0	0
hsa1019	0	0	0	0	0	0
hsa1020	0	0	0	0	0	0

Fig 6 Drug-Target interaction predictions for Enzyme

F. Binary Relation

The binary relation of drug-target interactions among enzyme is shown in Fig 7.

1	hsa:10	D00002	1	1
2	hsa:10	D00448	1	2
3	hsa:100	D00037	2	3
4	hsa:100	D00155	2	4
5	hsa:10056	D00021	3	5
6	hsa:1017	D02880	4	6
7	hsa:1018	D02880	5	6
8	hsa:10188	D01441	6	7
9	hsa:1019	D02880	7	6
10	hsa:1020	D02880	8	6
11	hsa:1021	D02880	9	6
12	hsa:1022	D02880	10	6
13	hsa:1024	D02880	11	6
14	hsa:1025	D02880	12	6
15	hsa:10269	D00279	13	8

Fig 7 Binary Relation of Enzyme

G. Drug-target Pair

Based on the drug-target interaction values, drug-target pair is identified. The highest value of interaction is found and corresponding enzyme drug-target pair is shown in Fig 8.

1	0.924353	hsa:1636	D00621
2	0.857192	hsa:2534	D01441
3	0.857035	hsa:5645	D00160
4	0.833499	hsa:64499	D00160
5	0.823582	hsa:5645	D00043
6	0.809633	hsa:64499	D00043
7	0.80858	hsa:1504	D00160
8	0.807243	hsa:6725	D01441
9	0.785061	hsa:3055	D01441
10	0.775562	hsa:63036	D00160
11	0.75165	hsa:23436	D00160
12	0.743193	hsa:2534	D01441
13	0.708093	hsa:7535	D01441
14	0.701674	hsa:3002	D00160
15	0.685447	hsa:2045	D01441

Fig 8 Drug-Target Pair for enzyme

H. Prediction of New Interactions

K-nearest neighbor method is used to predict possible targets for new drugs and possible drugs for new targets. Prediction of new interaction is shown in Fig 9.

hsa:1636	C11720
hsa:759	D01605
hsa:759	D01247
hsa:762	D01247
hsa:760	D01247
hsa:7153	C11230
hsa:7155	C11230
hsa:5742	C03080
hsa:7153	C12429
hsa:7155	C12429
hsa:5742	C02996
hsa:5743	C02996
hsa:7153	C01907

Fig 9 Newly Predicted DTI for enzyme

I. Drug-target Pair Identification

Drug-target interaction value is predicted and checked whether these interaction value is valid or not. If the value is 1 then the predicted DTI is valid otherwise it is not. The validated DTI is shown in Fig 10.

1	0.742999	hsa:2099	C14483	1
2	0.73479	hsa:2099	D00575	1
3	0.722097	hsa:2099	C15020	1
4	0.719486	hsa:2099	C02453	1
5	0.695891	hsa:5241	D00953	1
6	0.681868	hsa:2099	C15261	1
7	0.671919	hsa:2099	C15315	1
8	0.669423	hsa:2099	C14209	1
9	0.651191	hsa:7421	D00628	1
10	0.647253	hsa:2099	C15388	1
11	0.636598	hsa:2099	C14485	1
12	0.604187	hsa:2099	D00067	0
13	0.604067	hsa:2099	D00067	0
14	0.603016	hsa:2099	D00105	0
15	0.603016	hsa:2099	D00105	0

Fig 10 Drug-target pair identification

J. Accuracy

The accuracy of the predicted drug-target interaction among enzyme, ion channel, G-protein coupled receptor and nuclear receptor is evaluated. The drug-target prediction accuracy is shown in Table 1 and the corresponding graph representation in Fig 11.

Table 1 Accuracy of DT Prediction

	Accuracy (%)			
	Enz	GPCR	IC	NR
RPCA	95.4	94.2	94.7	93.6
Proposed Method	98.5	97.6	98.1	95.5

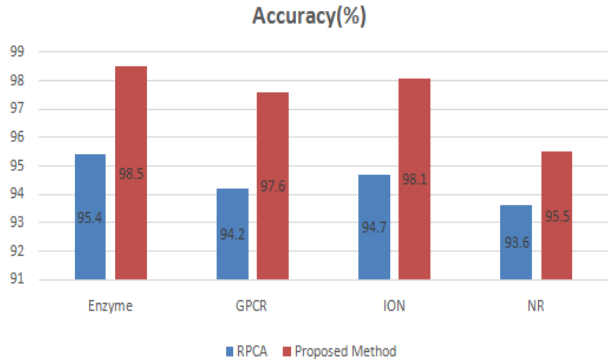

Fig 11 Accuracy of DT Prediction

VI. CONCLUSIONS

Identification potential interactions between drugs and targets are crucial for modern drug discovery and repurposing. The compound structures and protein sequences of drugs & targets are collected from KEGG databases. Initially similarity information among the drugs and targets is obtained and then computational prediction of drug-target interaction is done using the machine learning approach, bipartite local model. Bipartite local model is used to find the interactions between the chemical compounds and protein sequences by integrating the biological information. This model along with hubness-aware regression technique and k-nearest neighbor method is used to determine the drug-target pair. Error from bipartite local model is corrected by using hubness-aware regression and forwarded to k-nearest neighbor. K-nearest

neighbor is used to predict possible targets for new drugs and possible drugs for new targets. The interactions between drugs and targets is accepted or ignored according to its similarities. This model yields the drug target similarities information to predict the unknown interactions and provide improved predicted performance. The proposed method promises high accuracy above 95 % in the prediction of drug-target interactions. It is also successful in identification of drug-target pair.

In future implementation, the proposed approach can be extended to prevent the side effects of the drugs, by analyzing the side effects and designing effective treatment scheme.

REFERENCES

[1] X. Chen, M.-X. Liu, and G.-Y. Yan, "Drug–target interaction prediction by random walk on the heterogeneous network", Molecular BioSystems, vol. 8, no. 7, pp. 1970–1978, 2012.

[2] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, and Y. Tang, "Prediction of drug-target interactions and drug repositioning via network-based inference", PLoS Comput Biol, vol. 8, no. 5, p. e1002503, 2012.

[3] M. G onen, "Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization", Bioinformatics, vol. 28, no. 18, pp. 2304–2310, 2012.

[4] Jian-Ping Mei, Chee-Keong Kwoh, Peng Yang, Xiao-Li Li, Jie Zheng, "Drug-target interaction prediction by learning from local information and neighbors", Bioinformatics advance access, 2012.

[5] Hao Ding, Ichigaku Takigawa, Hiroshi Mamitsuka and Shanfeng Zhu, "Similarity-based machine learning methods for predicting drug-target interactions: a brief review", Bioinformatics advance access, 2013.

[6] Y. Wang and J. Zeng, "Predicting drug-target interactions using restricted boltzmann machines", Bioinformatics, vol. 29, no. 13, pp. i126–i134, 2013.

[7] T. van Laarhoven and E. Marchiori, "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile", PLoS one, vol. 8, no. 6, p. e66952, 2013.

[8] Weiming Yu, Yan Yan, Qing Liu, Junxiang Wang, "Predicting drug-target interaction networks of human diseases based on multiple feature information", 2013.

[9] Guillermo Palma, Maria Esther Vidal and Louiqa Raschid, "Drug-Target Interaction Prediction Using Semantic and Edge Partitioning", 2014.

[10] Xing Chen, Chenggang Clarence Yan, Xiaotian Zhang, Xu Zhang, Feng Dai, "Drug-target interaction prediction: databases,web servers and computational models", 2015.

[11] Wei Lan, Jianxin Wang, Min Li, Rqiqing Zheng,Fang-Xiang Wu, Yi Pan, "Predicting drug-target interaction based on sequence and structure information", 2015.

[12] Wei Lan, Jianxin Wang, Min Li, Jin Liu, Yaohang Li, Fang- Xiang Wu and Yi Pan, "Predicting drug-target interaction using positive-unlabeled learning", 2016.

[13] Ali Ezzat, Min Wu, Xiao-Li Li and Chee-Keong Kwoh, "Drug-Target interaction prediction via class inbalance-aware ensemble learning", 2016.

[14] Yingyi Du, Jihong Wang, Xiaodan Wang, Jiyun Chen, Huiyou Chang, "Predicting Drug-Target Interaction via Wide and Deep Learning", 2018.

[15] Kyle Yingkai Gao, Achille Fokoue, Heng Luo, Arun Iyengar, Sanjoy Dey, Ping Zhang, "Interpretable Drug Target Prediction Using Deep Neural Representation", 2018.

[16] Y. Yamanishi , M. Kotera , M. Kanehisa, and S. Goto, "Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework", Bioinformatics, vol. 26, no. 12, pp. i246– i254, 2010.

[17] M. A. Heiskanen and T. Aittokallio, "Predicting drug–target interactions through integrative analysis of chemo genetic assays in yeast", Molecular Bio Systems, vol. 9, no. 4, pp. 768–779, 2013.

[18] M. Hattori, Y. Okuno, S. Goto, and M. Kanehisa, "Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways", Journal of the American Chemical Society, vol. 125, no. 39, pp. 11853–11865, 2003.

[19] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu et al., "Kegg for linking genomes to life and the environment", Nucleic acids research, vol. 36, no. suppl 1, pp. D480–D484, 2008.

[20] Shobeir Fakhraei, Louiqa Raschid, Lise Getoor, "Drug-Target Interaction Prediction for Drug Repurposing with Probabilistic Similarity Logic", 2013.

[21] Kevin Bleakley, Yoshihiro Yamanishi, "Supervised prediction of drug-target interactions using bipartite local models", 2009.

[22] Chao Lan, Sai Nivedita Chandrasekaran, Jun Huan, "A distributed and privatized framework for drug-target interaction prediction", 2016.

[23] H. Chen and Z. Zhang, "A semi-supervised method for drug-target interaction prediction with consistency in networks", PloS one, vol. 8, no. 5, p. e62975, 2013.

[24] Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda and Minoru Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces", 2008.

[25] A. Patricia Bento, Anna Gaulton, Anna Hersey, Louisa J. Bellis, "The ChEMBL bioactivity database: an update", 2013.

[26] Nitin Bhardwaj, Morten Kallberg, Wonhwa Cho and Hui Lu, "MeTaDoR: Online Resource and Prediction server for Membrane targeting peripheral proteins", 2014.

[27] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drug–target interactions: a brief review", Briefings in Bioinformatics, vol. 15, no. 5, pp. 734–747, 2014.

[28] Vivan Law, Craig Knox, Yannick Djoumbou, "DrugBank 4.0: shedding new light on drug metabolism", 2013.

[29] Yan Qing Niu, " Supervised prediction of drug-target interactions by ensemble learning", 2014.

[30] Lihong Peng, Bo Liao, Wen Zhu, Zejun Li, Keqin Li, "Predicting Drug-Target Interactions with Multi-information Fusion", 2016.

[31] S. Fakhraei, B. Huang, L. Raschid, and L. Getoor, "Network-based drugtarget interaction prediction with probabilistic soft logic", Computational Biology and Bioinformatics, IEEE/ACM Transactions on, vol. 11, no. 5, pp. 775–787, 2014.

[32] Zheng Xia, Ling-Yun Wu, Xiaobo Zhou, Stephen T.C.Wong, "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces", 2017.

[33] Rawan S. Olayan, Haitham Ashoor and Vladimir B.Bajic, "DDR:efficient computational method to predict drug-target interactions using graph mining and machine learning approaches", 2017.

[34] S.M.Hasan Mahmud, Wenyu Chen, Hosney Jahan, Yongsheng Liu, Nasir Islam Sujan and Saeed Ahmed, "iDTi-CSsmoteB:Identification of Drug-Target interaction based on drug chemical structure and protein sequence using XGBoost with over-sampling technique SMOTE", 2019.

