Predicting Interactions Between Drugs and Target Proteins From Biological Data

Kowsalya.R[1], Mrs.E.Siva Sankari[2]
[1, 2]Government College of Engineering, Tirunelveli

Abstract- *identifying potential associations between drugs and targets is critical for contemporary drug discovery and repurposing. However, predicting these associations is difficult due to the restrictions of existing computational methods. Most of the traditional models consider only the chemical structure and protein sequences. True positive interactions are analyzed using dataset. To overcome these limitations, a semi- supervised based learning framework called Norm MulInfo is used. The proposed method initially determines similarity measures and local correlations among the samples by interactions between the biological information. The similarity among chemical compounds is obtained using Tanimoto coefficient. Similarly, Smith-Waterman score provides similarity of the protein sequences. The similarity information is then integrated into Robust Principal Component Analysis method. Which accurately classifies the four classes of drug target interactions such as Enzymes, Ion channels, GPCRs and Nuclear Receptors. The data is collected from Kyoto Encyclopedia of Genes and Genomes (KEGG) and used in analysis of potential relationships between chemical compounds and protein sequences at the systematic level. This method also predicts possible targets for new drugs. Finally, the proposed method can potentially address the drug target interactions (DTI), which can be further studied to prevent side effects of drugs and design effective treatment scheme.*

Keywords- Drugs, Target proteins, Drug target interactions, Similarity measures, Kyoto Encyclopedia of Genes and Genomes database, Robust Principal Analysis, Enzymes, Ion Channels, G-Protein Coupled Receptors, Nuclear Receptors, NormMulInf, Tanimoto coefficient, Smith-Watermann score.

I. INTRODUCTION

Drug Target Interaction is important for effective drug discovery and repurposing to identify potential interactions between drugs and targets. However, it is difficult to predict these associations due to the limitations of existing computational methods. Most of the traditional models consider only the chemical structure and protein sequences. Drug is a product that is produced from living organisms or contains components of living organisms. A drug is a chemical which is given to the human body in order to treat or prevent an illness or disease. Proteins arelarge biomolecules or macromolecules made up of one or more long chains of residues of aminoalkanoic acid.Proteins perform a huge array of functions within living organisms, including catalyzing metabolic reactions, DNA replication, responding to stimuli and transporting molecules from one location to different.

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. The potential associations are analyzed and it is used to detect multi-target drugs and multi-drug target.There are four databases involved in drug target interaction networks. Enzymes are biological molecules (proteins). The chemical reactions that are taken place within cells. Serve important functions in the body, such as aiding in digestion and metabolism. Ion channels are pores.Ion channels form membrane proteins that allow ions to pass through the pores of the channel. The largest and most diverse groups of membrane receptors in eukaryotes are G-Protein-Coupled Receptors (GPCRs).These cell surface receptors act like an inbox for messages in the form of light energy, lipids, sugars, proteins. Nuclear receptors are a class of proteins found in cells that senses steroid, thyroid hormones and certain other molecules. Nuclear receptors are capable of binding directly to DNA and regulating the expression of adjacent genes; therefore, these receptors are classified as transcription factors.

The empirical data sets only include true positive interactions. To overcome these limitations, a semi-supervised based learning framework called NormMulInfo is used. The proposed method initially determines similarity measures and local correlations among the samples by interactions between the biological information. The similarity among chemical compounds is obtained using Tanimoto coefficient. Similarly, Smith-Waterman score provides similarity of the protein sequences. The similarity information is then integrated into Robust Principal Component Analysis method. Which accurately classifies the four classes of drug target interactions such as Enzymes, Ion channels, GPCRs and Nuclear Receptors. The data is collected from Kyoto Encyclopedia of Genes and Genomes (KEGG) and used in

analysis of potential relationships between chemical compounds and protein sequences at the systematic level. This method also predicts possible targets for new drugs. Finally, the proposed method can potentially address the drug target interactions (DTI), which can be further studied to prevent side effects of drugs and design effective treatment scheme.

1. Kowsalya.R is a Post Graduate Scholar in Computer Science and Engineering department at Government College of Engineering, Tirunelveli (:kowsalyark4697@gmail.com)
2. Mrs.E.Siva Sangari is an Assistant Professor in Computer Science and Engineering Department in Government College of Engineering, Tirunelveli (sivasangari@gcetly.ac.in)

II. RELATED WORK

Y. Yamanishi et al. [1] Identification of interactions between drugs and target proteins is a key area in genomic drug discovery. Develop new statistical methods for predicting unknown drug-target interaction networks simultaneously on a large scale from chemical structure and the genomic target sequence. A powerful method is docking simulation, but it requires 3D structure information for the target proteins are membrane proteins. From a technical viewpoint, the performance of our method could be improved by using more sophisticated kernel similarity functions designed for genomic sequences and chemical structures.

M. Kanehisa et al. [2] KEGG databases are highly integrated. KEGG should be seen as a computational representation of the biological system, where biological structures and their relationships are computerized as separate database entries at the genetic, cellular and organism levels. This new database also contains molecular complexes, facilitating better organizations, such as the sub unit organization of transporters or receptors, is represented by the M number that corresponds to a set of K numbers.

H. Ding et al. [3]Predicting drug target interactions is useful to select possible drug candidates for further biochemical verification. This process is also closely related to a lot of work in pharmacogenomics that attempt to understand the relationships between the chemical and genomic spaces. The assumption of similarity-based methods is that similar targets share similar drugs and vice versa. Low prediction performance on Nuclear receptor can be explained by this assumption: in Nuclear receptor, the average number of interacting targets per drug is the minimum among the four subsets.

Y.Wanget al. [4]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.

H. Chenet al. [5] Semi-supervised learning method NetCBP is presented using labeled and unlabeled interaction information to tackle this problem. Assuming cohesive interactions between drugs ranked by their relevance to a query drug, and target proteins ranked by their relevance to the query drug'shidden target proteins, we are developing a learning system that maximizes rank coherence in relation to known drug-target interactions. Once applied to four groups of significant drug-target interaction networks, our approach strengthens past methods in terms of cross validation and some highly predicted interactions are verified by the publicly accessible drug target databases, which suggest the usefulness of our process.

F. Cheng et al. [6]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 detectnew 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 drughas with established experimental evidence.

M. G onenet al. [7]In the drug discovery process for known diseases the detection of interactions between drug compounds and target proteins is of great practical significance. KBMF2K uses one kernel function for chemical similarity and another kernel function calculated on protein sequences for genomic similarity. The performance of our approach can be improved by integrating multiple kernels for both kinds of similarity.The remaining sets of results show that our novel probabilistic interpretation obtains better generalization performance than earlier optimization-based approaches.

M. A. Heiskanen et al. [8]Chemical-genomic and genetic approaches to profiling interactions are commonlyused to study drug action and resistance mechanisms. The efficiency of each of the four scoring approaches was evaluated using drug–gene interactions from the STITCH database, and the reproducibility of the rankings obtained

through the different approaches was evaluated using duplicate measurements present in the homo and heterozygous datasets.The target rankings based on the fitness defects were more accurate at recovering the STITCH links compared to the approach utilizing the conventional profile correlations.

K. Bleakleyet al. [9]High-throughput experiments analyzing the genome, transcriptome and proteome are beginning to lead to the understanding of genomic spaces populated by these classes of protein. Chemical genomics research aims to relate the chemical space with the genomic space in order to identify potentially useful compound–protein pairs. The current state-of-the-art involves integrative methods that simultaneously take into account such things as target protein sequences, drug chemical structures and the currently known drug– target network.

S. Fakhraei et al. [10]Drug-target interaction identification is an essential step of drug repurposing and drug adverse effect prediction. Drug-target associations being identified in vitro is a labor-intensive and costly process. The structure of the network and the multi relational aspects make it challenging to convert such knowledge into the (flat) data formats that are typically used with standard prediction algorithms.Predict potential interactions using drug-drug and target-target similarities and a bipartite interaction graph. Using SIMCOMP, they compute the 2D chemical drug similarities and sequence similarities for targets via the Smith-Waterman score.

III. DATA AND METHODS

A. Dataset – KEGG (Kyoto Encyclopedia of Genes and Genomes Database)

Public benchmark dataset consists of chemical compounds and protein sequences are collected from KEGG database. This proposed work uses KEGGdatabase for prediction of drug target interactions. The detailed distribution of samples in the dataset is shown in Table 4.1.

Table 3.1Distribution of samples in the dataset

Dataset	Enz	Ion	GPCRs	Nuc
Drugs (n)	445	210	223	54
Targets(m)	664	204	95	26
Interactions	2926	1476	635	90
ratio(n/m)	0.67	1.03	2.35	2.08
N_{avetar}	6.58	7.03	2.85	1.67
$N_{avedrug}$	4.41	7.24	6.68	3.46

The Determined drugs 445, 210,223,54 interact with 664, 204, 95, 26 proteins from human Enzymes, Ion channel, Guanine protein coupled receptors and Nuclear receptors respectively, with known interactions of 2926, 1476, 635 and 90. From the above mentioned table: n denotes the number of drugs, m denotes the number of targets, N_{avetar} denotes the average number of targets interacting with each drug and $N_{avedrug}$ denotes the average number of drugs interacting with each target.

B. Proposed system Architecture

In the proposed system, the benchmark dataset containing chemical compound and protein sequence is collected and used. After data collection, features are extracted from the dataset. The similarity measures are determined by utilizing extracted features from the dataset. The chemical structure similarity between two compounds can be calculated based on the Tanimoto coefficient. The target protein sequence similarity can be calculated using Smith waterman score method. The Robust Principal Component Analysis (RPCA) is used to recover the efficiency and accuracy of low rank matrix from highly corrupted measurements.The interactions between the drug and target are obtained by NormMulInf prediction. The required Drug Target Interactions (DTI) areobtained. Figure 3.1 represents overall work flow of the proposed system.

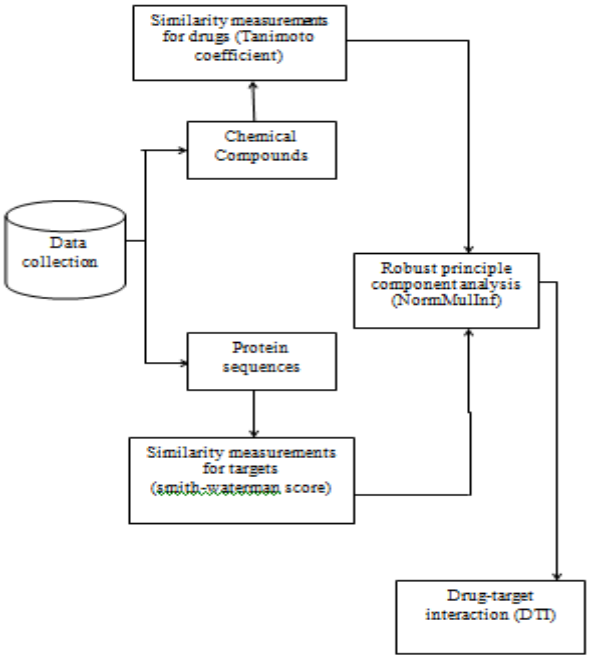


Figure 3.1 Overall block diagram of the proposed work

C. Similarity Measures

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 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 waterman 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.

3) Robust Principal Component Analysis:

Robust principal component analysis is a tool for discovering and exploiting low dimensional structures in high dimensional data.A small portion of large errors can corrupt the estimation of low rank structures of biological data. In order to recover the efficiency and accuracy of the low rank structures of biological data, the modified PCA method called RPCA is used.

Robust principal component analysis efficiently and accurately recovers the low rank matrix A from highly corrupted measurements. This can be solved within polynomial time.

$$D = A + E$$

Where,
A – Low rank matrix
E – Error matrix
D – Corrupted entries observations
This method is applied to recover the corrupted low rank matrix.

D. Prediction

1) NormDrug for DTI Prediction:

NormDrug is used to find the drug similarity matrix. Let us assign drugs as samples and each target as a label. NormDrug method is categorized into three parts: (i) The MDTIR (Masked DTI Ratio) is used to find the mask part of interactions for each sample. (ii) Laplacian matrix is used to combine the chemical structures similarities between drugs and the local correlations between the labels of samples in the DTI network. (iii) The predicted DTI matrix is achieved.In present study, drug similarity is measured by considering each drug as a vector of the frequency of interaction with the targets. There are two drug vectors.

Local correlation between the drugs can be calculated as,

$$Sim_{NetDrug}(i, j) = \frac{x_i x_j^T}{\|x_i\|\|x_j\|}$$

It denotes drug similarity matrix according to the local correlations between the labels of samples in the Drug Target Interaction (DTI) network.

Similarity between chemical compounds can be calculated as,

$$Sim_{NetDrugNorm}(i, j) = \frac{Sim_{NetDrug}(i, j)}{\sum_{k=1}^{n} Sim_{NetDrug}(i, k)}$$

If $Sim_{NetDrug}(i, j)$ is higher than $Sim_{NetDrug}(i, k)$, the i^{th} and j^{th} drugs are simultaneously associated with abundant targets.

The similarity of the drug can be obtained by combining the $Sim_{NetDrugNorm}$ and $Sim_{StruDrug}$.

$$Sim_{Drug} = Sim_{NetDrugNorm} + \alpha Sim_{StruDrug}$$

In this, α is defined as the weighted parameter which is the ratio of the $Sim_{NetDrugNorm}(i,j)$ to $Sim_{StruDrug}(i,j)$.

$$\alpha = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} Sim_{NetDrugNorm}(i,j)}{\sum_{i=1}^{n}\sum_{j=1}^{n} Sim_{StruDrug}(i,j)}$$

2) NormTarget for DTI prediction:

NormTarget is used to find the target similarity matrix. Let us assign target as samples and each drug as a label. The method is categorized into three parts: (i)The MDTIR (Masked DTI Ratio) is used to find the mask part of interactions for each sample. (ii)Laplacian matrix is used to combine the chemical structures similarities between targets and the local correlations between the labels of samples in the DTI network. (iii) The predicted DTI matrix is achieved. In present study, target similarity is measured by considering each target as a vector of the frequency of interaction with the drugs. There are two target vectors.

Local correlation between the targets can be calculated as,

$$Sim_{NetTar}(i,j) = \frac{X_i X_j^T}{\|X_i\|\|X_j\|}$$

It denotes target similarity matrix according to the local correlations between the labels of samples in the Drug Target Interaction (DTI) network.

Similarity between protein sequences can be calculated as,

$$Sim_{NetTarNorm}(i,j) = \frac{Sim_{NetTar}(i,j)}{\sum_{k=1}^{n} Sim_{NetTar}(i,k)}$$

If $Sim_{NetTar}(i,j)$ is higher than $Sim_{NetTar}(i,k)$, the i^{th} and j^{th} targets are simultaneously associated with abundant drugs.

Similarity of the drug can be obtained by combining the $Sim_{NetDrugNorm}$ and Sim_{Tar}.

$$Sim_{Tar} = Sim_{NetTarNorm} + \beta Sim_{SeqTar}$$

In this, β is defined as the weighted parameter which is the ratio of the $Sim_{NetTarNorm}(i,j)$ to $Sim_{StruTar}(i,j)$.

$$\beta = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} Sim_{NetTarNorm}(i,j)}{\sum_{i=1}^{n}\sum_{j=1}^{n} Sim_{StruTar}(i,j)}$$

3) NormMulInf for DTI Prediction:

NormMulInf for DTI Prediction is based on NormDrug and NormTarget.

$$\text{Pre} = Pre_{Drug} + \gamma Pre_{Tar}$$

The Pre_{Drug} considers drugs as samples and target as labels. It denotes DTI score matrix by NormDrug. The Pre_{Tar} considers target as samples and drugs as labels. It denotes DTI score matrix by NormTarget.

$$\gamma = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} Pre_{Drug}(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} Pre_{Tar}(i,j)}$$

γ Represents the balance between the score matrix of the Pre_{Drug} and score matrix of the Pre_{Tar}.

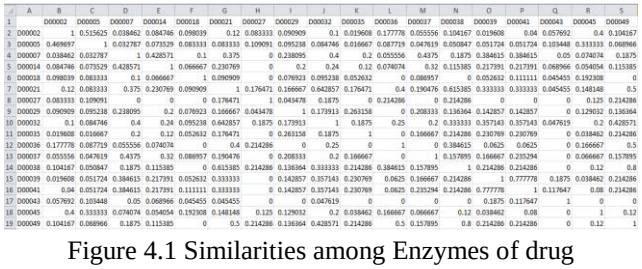
IV. EXPERIMENTAL EVALUATION AND RESULTS

A. Drug Interaction Prediction

The interaction among the drugs is measured using Tanimoto coefficient by determining similarity. The drug interaction is found among the Enzymes, Ion channels, Guanine Protein Coupled Receptors and Nuclear Receptors.

2) Similarities among Enzymes of drug:

The similarities among Enzymes are calculated for determining interactions of drugs. This is done using Tanimoto coefficient and is shown in Figure 4.1 and corresponding the corresponding graph representation in Figure 4.2.



	D00002	D00005	D00007	D00014	D00018	D00021	D00027	D00029	D00032	D00035	D00036	D00037	D00038	D00039	D00041	D00043	D00045	D00049
D00002	1	0.515625	0.038462	0.084746	0.098039	0.12	0.083333	0.090909	0.1	0.019608	0.177778	0.055556	0.104167	0.019608	0.04	0.057692	0.4	0.104167
D00005	0.469697	1	0.032787	0.073529	0.083333	0.083333	0.109091	0.095238	0.084746	0.016667	0.087719	0.047619	0.050847	0.051724	0.051724	0.103448	0.333333	0.068966
D00007	0.038462	0.032787	1	0.428571	0.1	0.375	0	0.238095	0.4	0.2	0.055556	0.4375	0.1875	0.384615	0.384615	0.05	0.074074	0.1875
D00014	0.084746	0.073529	0.428571	1	0.066667	0.230769	0	0.2	0.24	0.12	0.074074	0.32	0.115385	0.217391	0.217391	0.068966	0.054054	0.115385
D00018	0.098039	0.083333	0.1	0.066667	1	0.090909	0	0.076923	0.095238	0.052632	0	0.086957	0	0.052632	0.111111	0.045455	0.192308	0
D00021	0.12	0.083333	0.375	0.230769	0.090909	1	0.176471	0.166667	0.642857	0.176471	0.4	0.190476	0.615385	0.333333	0.333333	0.045455	0.148148	0.5
D00027	0.083333	0.109091	0	0	0	0.176471	1	0.043478	0.1875	0	0.214286	0	0.214286	0	0	0	0.125	0.214286
D00029	0.090909	0.095238	0.238095	0.2	0.076923	0.166667	0.043478	1	0.173913	0.263158	0	0.208333	0.136364	0.142857	0.142857	0	0.129032	0.136364
D00032	0.1	0.084746	0.4	0.24	0.095238	0.642857	0.1875	0.173913	1	0.1875	0.25	0.2	0.333333	0.357143	0.357143	0.047619	0.2	0.428571
D00035	0.019608	0.016667	0.2	0.12	0.052632	0.176471	0	0.263158	0.1875	1	0	0.166667	0.214286	0.230769	0.230769	0	0.038462	0.214286
D00036	0.177778	0.087719	0.055556	0.074074	0	0.4	0.214286	0	0.25	0	1	0	0.384615	0.0625	0.0625	0	0.166667	0.5
D00037	0.055556	0.047619	0.4375	0.32	0.086957	0.190476	0	0.208333	0.2	0.166667	0	1	0.157895	0.166667	0.235294	0	0.066667	0.157895
D00038	0.104167	0.050847	0.1875	0.115385	0	0.615385	0.214286	0.136364	0.333333	0.214286	0.384615	0.157895	1	0.214286	0.214286	0	0.12	0.8
D00039	0.019608	0.051724	0.384615	0.217391	0.052632	0.333333	0	0.142857	0.357143	0.230769	0.0625	0.166667	0.214286	1	0.777778	0.1875	0.038462	0.214286
D00041	0.04	0.051724	0.384615	0.217391	0.111111	0.333333	0	0.142857	0.357143	0.230769	0.0625	0.235294	0.214286	0.777778	1	0.117647	0.08	0.214286
D00043	0.057692	0.103448	0.05	0.068966	0.045455	0.045455	0	0	0.047619	0	0	0	0	0.1875	0.117647	1	0	0
D00045	0.4	0.333333	0.074074	0.054054	0.192308	0.148148	0.125	0.129032	0.2	0.038462	0.166667	0.066667	0.12	0.038462	0.08	0	1	0.12
D00049	0.104167	0.068966	0.1875	0.115385	0	0.5	0.214286	0.136364	0.428571	0.214286	0.5	0.157895	0.8	0.214286	0.214286	0	0.12	1

Figure 4.1 Similarities among Enzymes of drug

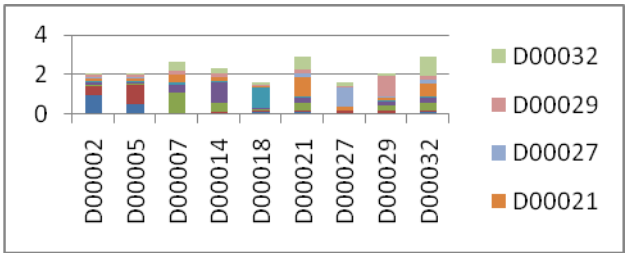


Figure 4.2Graph representations of similarities among enzymes of drug

2) Similarities among Ion Channels of drug:

The similarities among Ion channels are calculated for determining interactions of drugs. This is done using Tanimoto coefficient and is shown in Figure 4.3 and the corresponding graph representation in Figure 4.4.

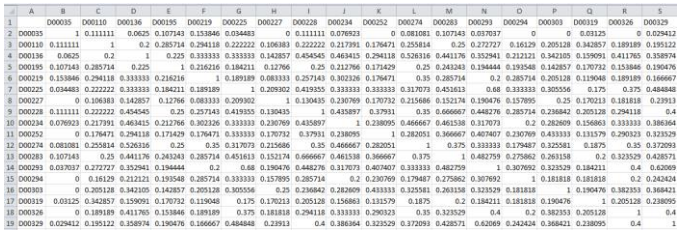


Figure 4.3 Similarities among Ion channels of drug

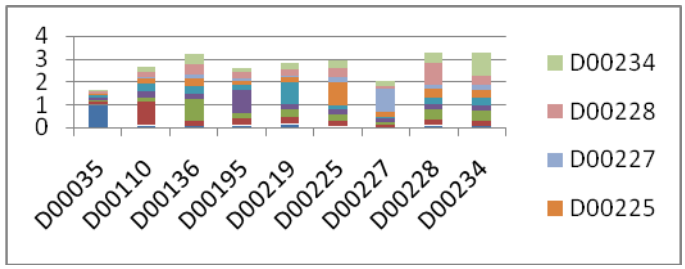


Figure 4.4Graph representations of similarities among Ion channels of drug

3) Similarities among G-Protein Coupled Receptors of drug:

The similarities among G-Protein Coupled Receptors are calculated for determining interactions of drugs. This is done using Tanimoto coefficient and is shown in Figure 4.5 and the corresponding graph representation in Figure 4.6.

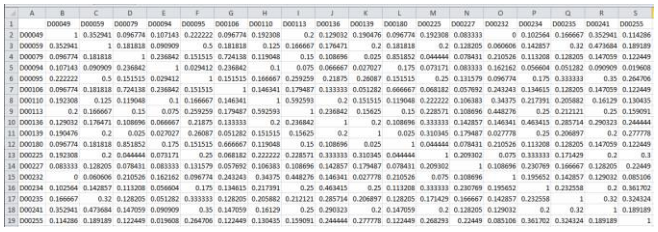


Figure 4.5 Similarities among G-Protein Coupled Receptors of drug

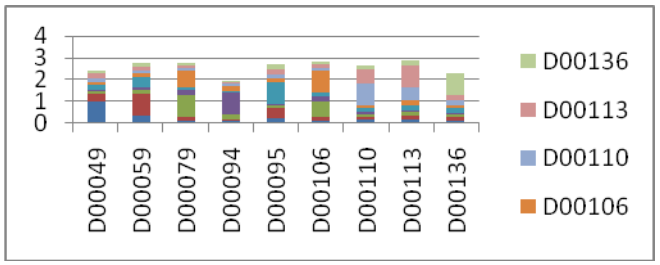


Figure 4.6Graph representations of similarities among G-Protein Coupled Receptors

4) Similarities among Nuclear Receptors of drug:

The similarities among nuclear receptors are calculated for determining interactions of drugs. This is done using Tanimoto coefficient and is shown in Figure 4.7 and the corresponding graph representation in Figure 4.8.

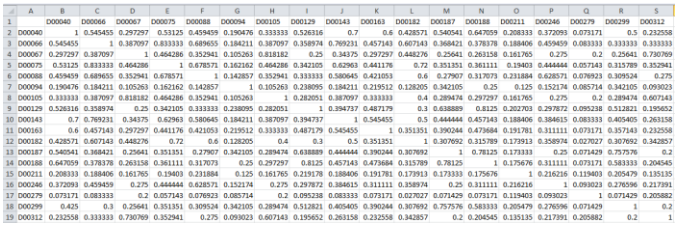


Figure 4.7 Similarities among Nuclear receptors of drug

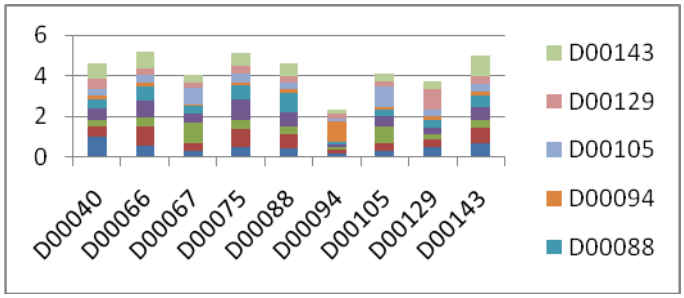


Figure 4.8Graph representations of similarities among nuclear receptors of drug

B. Target Proteins Interaction Prediction

The interaction among the target is measured using Smith-waterman score by determining similarity. The target interaction is found among the Enzymes, Ion channels, Guanine Protein Coupled Receptors and Nuclear Receptors.

1) Similarities among Enzymes of target:

The similarities among Enzymes are calculated for determining interactions of targets. This is done using Smith-waterman score and is shown in Figure 4.9 and the corresponding graph representation in Figure 4.10.

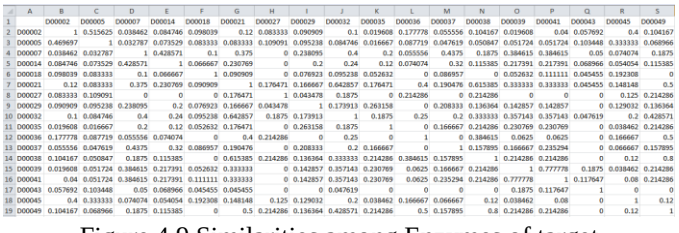


Figure 4.9 Similarities among Enzymes of target

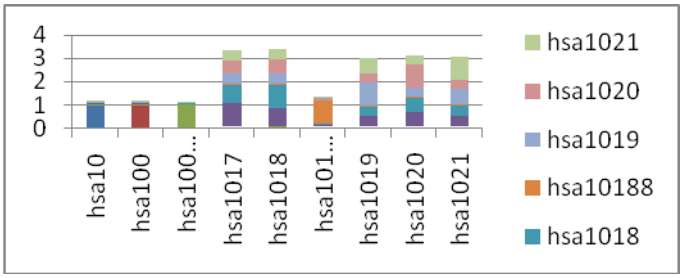


Figure 4.10Graph representations of similarities among Enzymes of target

2) Similarities among Ion Channels of target:

The similarities among Ion channels are calculated for determining interactions of targets. This is done using Smith-waterman score and is shown in Figure 5.11 and the corresponding graph representation in Figure 4.12.

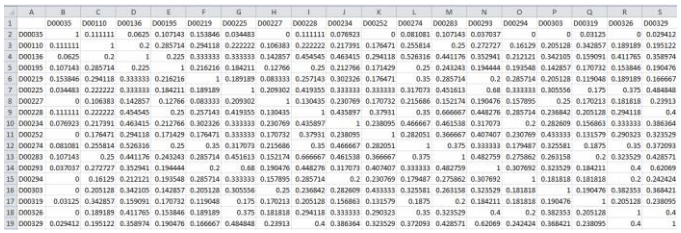


Figure 4.11 Similarities among Ion channel of target

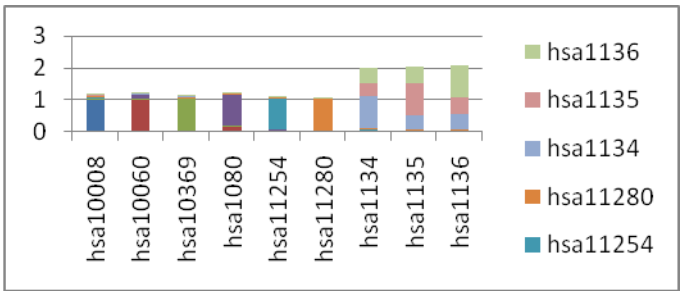


Figure 4.12Graph representations of similarities among Ion channels of target

3) Similarities among GPCRs of target:

The similarities among G-Protein Coupled Receptors are calculated for determining interactions of targets. This is done using Smith-waterman score and is shown in Figure 4.13 and the corresponding graph representation in Figure 4.14.

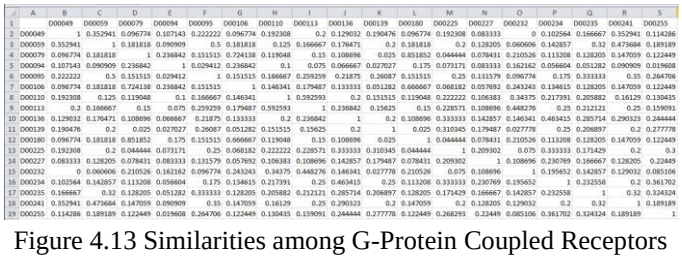


Figure 4.13 Similarities among G-Protein Coupled Receptors of target

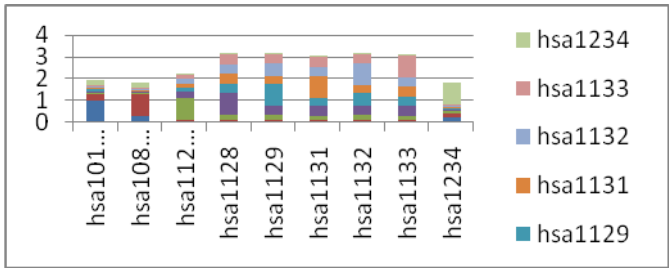


Figure 4.14Graph representations of similarities among G-Protein Coupled Receptors of target

4) Similarities among Nuclear Receptors of target:

The similarities among nuclear receptors are calculated for determining interactions of targets. This is done using Smith-waterman score and is shown in Figure 4.15 and the corresponding graph representation in Figure 4.16.

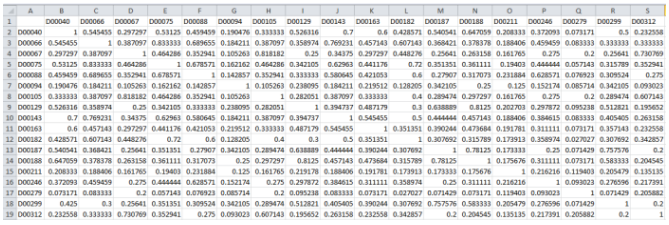


Figure 4.15 Similarities among nuclear receptors of target

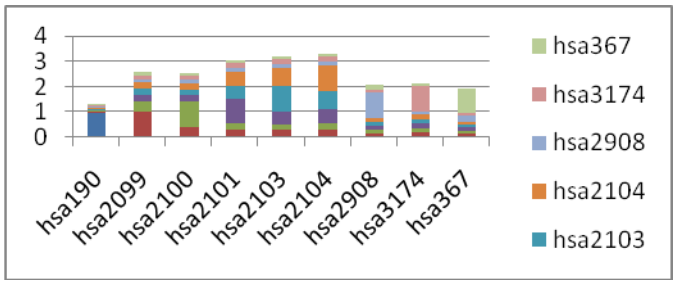


Figure 4.16Graph representations of similarities among nuclear receptors of target

C. Drug Interaction Prediction

The interaction among drugs and targets is measured using Robust Principal Analysis by considering similarity determined in previous stage. The drug and target interaction is found among the Enzymes, Ion channels, Guanine Protein Coupled Receptors and Nuclear Receptors.

1) Enzyme-Drug target Interaction Prediction:

The drug-target interactions in the enzyme were calculated using RPA. The following figure depicts the estimated drug target interaction of the enzymes.

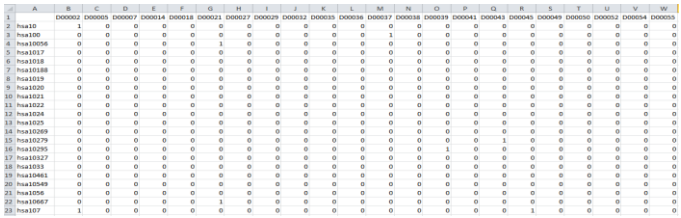


Figure 4.17 Drug Target interaction predictions for Enzymes

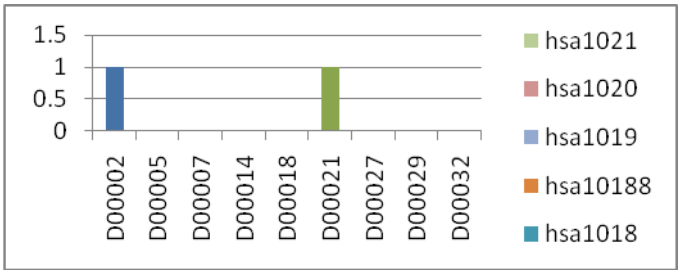


Figure 4.18 Graphical representation of drug target interaction in Enzyme

2) Ion Channel-Drug target Interaction Prediction:

The drug-target interactions in the ion channels were calculated using RPA. The following figure depicts the estimated drug target interaction of the ion channels.

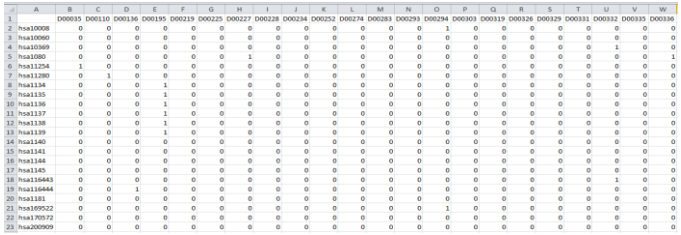


Figure 4.19 Drug Target interaction predictions for Ion channels

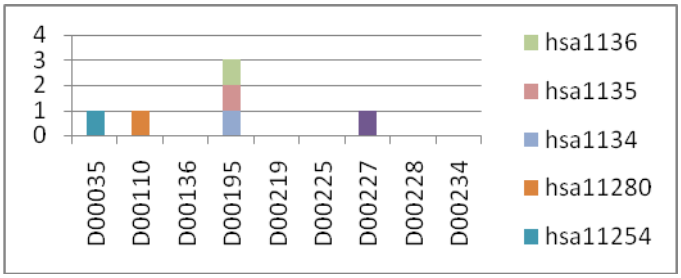


Figure 4.20 Graphical representation of drug target interaction in Ion channels

3) GPCRs-Drug target Interaction Prediction:

The drug-target interactions in the G-protein coupled receptors were calculated using RPA. The following figure depicts the estimated drug target interaction of the G-protein coupled receptors.

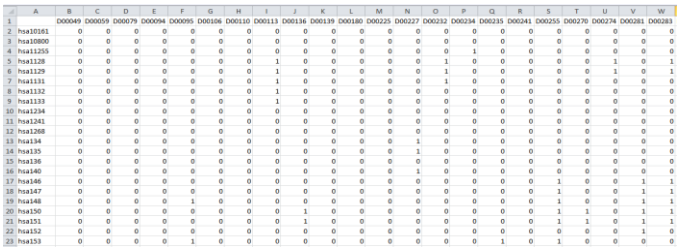


Figure 4.21 Drug Target interaction predictions for GPCRs

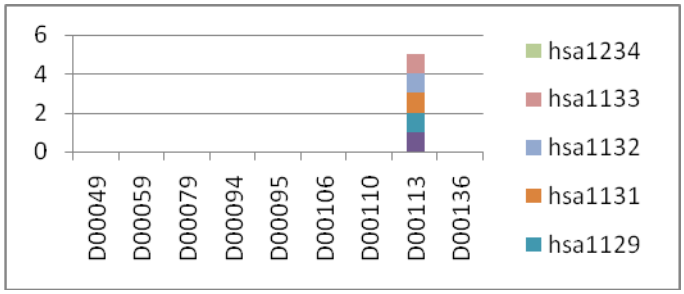


Figure 4.22 Graphical representation of drug target interaction in GPCRs

4) Similarities among Nuclear Receptors of drug:

The drug-target interactions in the G-protein coupled receptors were calculated using RPA. The following figure depicts the estimated drug target interaction of the G-protein coupled receptors.

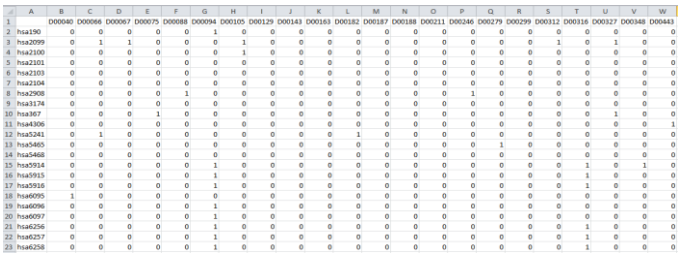


Figure 4.23 Drug Target interaction prediction for nuclear receptor

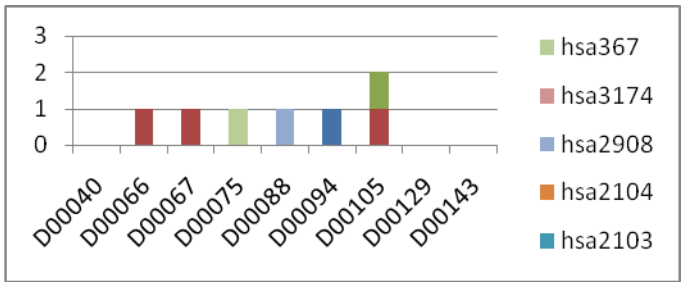


Figure 4.24 Graphical representation of drug target interaction in nuclear receptors

V. CONCLUSION

In this work, Drug-Target interaction is a crucial technique in Drug Target Interaction (DTI) prediction modern

drug discovers integrates Robust Principal Component Analysis with various similarity measures into a unified framework. The proposed approach is evaluated on KEGG, a benchmark data set. A large amount of biological information related to drugs and targets are considered and initially similarity information among the drugs and targets is obtained. Then the similarity information is fed into Robust Principal Component Analysis to build classifier model. Integrating various biological information can help identify new Drug Target Interactions. It is also used to predict possible targets for new drugs and possible drugs for new targets. The interactions between drugs and new drug target interaction 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. In future implementation, the proposed approach is used to detect multi-target drugs and multi-drug targets. The future work can also focus on prevention of side effects of the drugs, by analyzing the side effects and designing effective treatment scheme.

REFERENCES

[1] Y. Yamanishi, 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.

[2] M. Kanehisa, M. Araki, S. Goto, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu et al,. "Kegg for linking genomes to life and the environment,' Nucletic acids research,vol. 36,no. suppl 1, pp. D480-D484, 2008.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] K. Bleakley and Y. Yamanishi, "Supervised prediction of drug–target interactions using bipartite local models,"Bioinformatics,vol.25,no.18, pp. 2397–2403, 2009.

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

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

[12] 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.

[13] 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.

[14] 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", 2014.

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

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

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

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

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

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

[21] 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.
[22] Nitin Bhardwaj, Morten Kallberg, Wonhwa Cho and Hui Lu, "MeTaDoR: Online Resource and Prediction server for Membrane targeting peripheral proteins", 2014.
[23] Yan Qing Niu, "Supervised prediction of drug-target interactions by ensemble learning", 2014.
[24] Xing Chen, Chenggang Clarence Yan, Xiaotian Zhang, Xu Zhang, Feng Dai, "Drug-target interaction prediction: databases,web servers and computational models", 2015.
[25] Wei Lan, Jianxin Wang, Min Li, Rqiqing Zheng,Fang-Xiang Wu, Yi Pan, "Predicting drug-target interaction based on sequence and structure information", 2015.
[26] Lihong Peng, Bo Liao, Wen Zhu, Zejun Li, Keqin Li, "Predicting Drug-Target Interactions with Multi-information Fusion", 2016.
[27] Chao Lan, Sai Nivedita Chandrasekaran, Jun Huan, "A distributed and privatized framework for drug-target interaction prediction", 2016.
[28] Ali Ezzat, Min Wu, Xiao-Li Li and Chee-Keong Kwoh, "Drug-Target interaction prediction via class inbalance-aware ensemble learning", 2016.
[29] Zheng Xia, Ling-Yun Wu, Xiaobo Zhou, Stephen T.C.Wong, "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces", 2017.
[30] Yingyi Du, Jihong Wang, Xiaodan Wang, Jiyun Chen, Huiyou Chang, "Predicting Drug-Target Interaction via Wide and Deep Learning, 2018.
[31] Shyama M Nair, Joby George, "A survey on Drug-Target interaction prediction methods analysis of prediction mechanisms for drug target discovery", 2018.
[32] Srikanth Kuthuru, Adam T Szafran, Fabio Stossi, Michael A Mancini and Arvind Rao, "Leveraging Image-Derived phenotypic measurements for drug-target interaction predictions", 2019.
[33] 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.