ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 2, Issue 11, DOI: 10.29027/IJIRASE.v2.i11.2019, 357-361 , May 2019

Drug Repositioning

Aditi Sharma , aditisharma472@gmail.com Student,CSE , IPEC
Anjali tyagi , tyagianjali260@gmail.com,
,Student,CSE, IPEC
Dr. Sweeta Bansal , sweeta.bansal@ipec.org.in,
Mentor, CSE, IPEC

Abstract

The purpose of this project is to analyze drugs known to treat a particular disease and find other drugs that can potentially treat the same disease, with some modifications, if required. Drug discovery is an expensive process. It takes a lot of time and resources to find new drugs to treat a disease. For this process, our project can be used to predict some potential drugs to treat a particular disease and those drugs can be further analyzed so as to study their properties with respect to the disease to be treated. For this, we have considered similarity between drugs chemically as well as the shared genes that are affected by these drugs.

Keywords : Drug

\

I. Introduction

Drugs generally have the tendency to act on more than one target, which is useful to check for more indications to be treated using those drugs; which is can be perceived as the main concept of drug repositioning.
Drug repositioning is the process of examining currently approved drugs and analyzing what other diseases or indications can be treated with the help of those drugs.

Drug repositioning mainly involves studying the chemical similarities and the gene sequences for target genes, that are affected by those drugs. Basically, drug repositioning depends on analyzing the currently available approved drugs and then studying those drugs for new indications.

Currently, an estimated number of 4,000 of active pharmaceutical ingredients (API) have been approved for human use in the world[4]. Approved drugs keep accumulating over the years, on average to 20 to 30 NMEs each year have been approved by US-FDA [] further expanding the space for drug repositioning[1].

Many drugs are subsequently discovered to have the potential to treat some other new Indications. Extension of the clinical use of a drug to a new indication has historically occurred through serendipitous clinical observation, e.g, sildenafil for erectile dysfunction, but recently has occurred through logical connection of a disease's pathophysiology to a drug's target, e.g, losartan for Marfan syndrome[2].

Advancements in treatment have also been made by combining new drugs with repurposed old drugs, e.g, chloroquine, historically used to treat malaria, has recently been combined with a new drug, tarceva, which kills lung cancer. It is estimated that roughly 1 in 10,000 new chemical entities that enter the pharmaceutical research and design process actually makes it to market[5]. The success rate for repurposed drugs was almost 30% in 2012[3].

In this paper, our aim is to provide a disease-specific method for drug repositioning, i.e, to provide some potential drugs that are not being currently used for that disease, but can be used to treat that disease. For our study, we have taken 2 diseases, migraine disorders

and headache disorders, to predict the same. First, we have considered the drugs that are currently being used to treat these diseases, and compared their chemical structures with currently available approved drugs. Next, we have considered the genes that are commonly affected by the drugs already being used to treat the diseases and kept a score of similarities in those genes' sequences. Then, a matrix was formed keeping a track of the commonly affected genes and whether they are directly affected by the drugs that were found to be similar to those already treating the disease.

Based on this matrix, a score was calculated for all the similar drugs and a classifier, like svm, neural network are used to predict a class, 0 or 1, for that drug. 0 indicating that the drug cannot be used to treat the disease and 1 indicating that the drug can be used to treat the disease and should be analyzed further for checking the side effects related to that particular disease.

II. Material and method

1. Dataset

i. Drug-disease association

Data was taken from DrugBank[6]. It contains 1933 drug-disease associations between 593 drugs and 313 diseases. All 593 drugs are registered in DrugBank and all 313 diseases are listed in OMIM database[7].

ii. Drug targets

This data was also obtained from DrugBank[6]. Drug targets one or more cellular molecules such as metabolites or proteins for desired effects. A list of targets corresponding to all 584 drugs were Obtained.

iii. Genomic data
Amino acid sequences of the target proteins were obtained from DrugBank[6].

2. Method
i. Drug-drug similarity

The dataset containing chemical structures of drugs, taken from DrugBank[6], was used to observe similarity between drugs, using chemical fingerprints[8].

This was done to get a list of drugs that were similar in structure, to the drugs already being used to treat the two diseases.

Similarity measure was calculated using the Tanimoto coefficient[9]. Tanimoto score gives the ratio of common elements and total elements present in the two structures. It is only applicable for a binary variable, and for binary variables, it ranges from 0 to +1 (+1 being the highest similarity).

Alternatively, the search results can be estimated using a threshold value, which was used for our research.

Drugs that were already known to treat the two diseases in consideration, migraine disorders and headache disorders,were used to calculate similarities with other approved drugs, present in the obtained database, based on chemical fingerprints and Tanimoto coefficient for similarity measure.

ii. Gene similarity

For both diseases, data was analyzed, to find the genes that were commonly affected by these diseases. This was estimated using the

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 2, Issue 11, DOI: 10.29027/IJIRASE.v2.i11.2019, 357-361 , May 2019

data for each drug known to treat the disease currently.
Using this data, pairwise gene similarity was calculated for all genes and a matrix was formed. This pairwise gene similarity was calculated using the Smith-Waterman algorithm[10].

iii. Drug-gene association

Initial association matrix was formed for both the diseases. Where row contains DrugBank ids of drugs and columns contain gene ids of the common genes, extracted in the above step.

Value 1 was assigned to a cell, if that drug affected that particular gene directly, otherwise value 0 was assigned to that cell.

3 association matrices were formed for each disease, one containing the cell values among drugs known to treat the disease and the common genes, second containing cell values among drugs, neither similar nor known to treat the disease and the third, containing cell values among drugs found to be similar and the common genes.

iv. Calculating scores

A score was calculated for each drug, i.e for each row in the initial association matrix.

When a value 1 was encountered in a row, the score for that drug was calculated by checking the similarity between all genes with the gene for which value 1 was found in the row.

By taking a weighted value for the similarity and number of drugs associated with that gene in the matrix,

A score was calculated and assigned to that particular drug. In case, there were multiple 1 values in a row, score with the highest value was assigned to that drug.
v. Assigning classes and using classifiers
Scores calculated for the first two initial matrices, one with drugs known to treat the disease (positive class) and the one with drugs neither similar nor known to treat the disease (negative class), were used as the training inputs for the two classifiers used (svm and neural network).

Positive class of drugs for a disease was assigned class 1 and the negative class of drugs was assigned class 0.

This data was used as the training data for the classifier and the scores calculated for similar drugs, was fed to the classifier so that the classifier can predict classes for those drugs.

Drugs predicted to have class 1, were provided as the result of our analysis.

III. Results

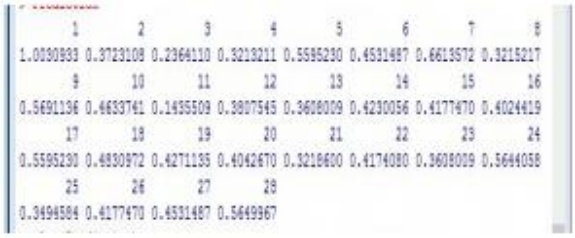
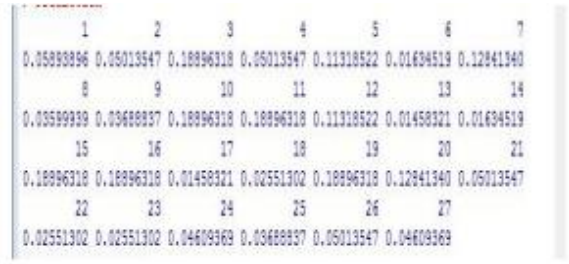
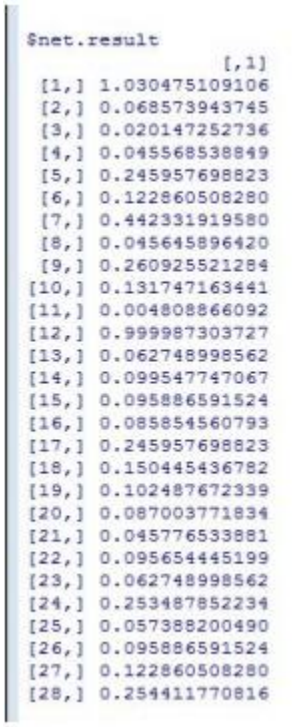


Fig 1.Predicted classes for migraine disorders using svm



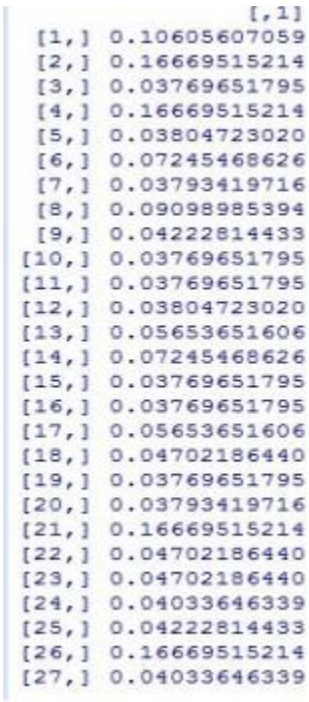
ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 2, Issue 11, DOI: 10.29027/IJIRASE.v2.i11.2019, 357-361 , May 2019

Fig 2.Predicted classes for headache disorders using svm



```
$net.result
                [,1]
 [1,] 1.030475109106
 [2,] 0.068573943745
 [3,] 0.020147252736
 [4,] 0.045568538849
 [5,] 0.245957698823
 [6,] 0.122860508280
 [7,] 0.442331919580
 [8,] 0.045645896420
 [9,] 0.260925521284
[10,] 0.131747163441
[11,] 0.004808866092
[12,] 0.999987303727
[13,] 0.062748998562
[14,] 0.099547747067
[15,] 0.095886591524
[16,] 0.085854560793
[17,] 0.245957698823
[18,] 0.150445436782
[19,] 0.102487672339
[20,] 0.087003771834
[21,] 0.045776533881
[22,] 0.095654445199
[23,] 0.062748998562
[24,] 0.253487852234
[25,] 0.057388200490
[26,] 0.095886591524
[27,] 0.122860508280
[28,] 0.254411770816
```

Fig 3. Classes predicted for migraine disorders using neural networks



```
                [,1]
 [1,] 0.10605607059
 [2,] 0.16669515214
 [3,] 0.03769651795
 [4,] 0.16669515214
 [5,] 0.03804723020
 [6,] 0.07245468626
 [7,] 0.03793419716
 [8,] 0.09098985394
 [9,] 0.04222814433
[10,] 0.03769651795
[11,] 0.03769651795
[12,] 0.03804723020
[13,] 0.05653651606
[14,] 0.07245468626
[15,] 0.03769651795
[16,] 0.03769651795
[17,] 0.05653651606
[18,] 0.04702186440
[19,] 0.03769651795
[20,] 0.03793419716
[21,] 0.16669515214
[22,] 0.04702186440
[23,] 0.04702186440
[24,] 0.04033646339
[25,] 0.04222814433
[26,] 0.16669515214
[27,] 0.04033646339
```

Fig 4. Classes predicted for headache disorders using neural networks

The drugs with highest value (near 1), or the predicted potential drugs to treat the diseases were:

For migraine disorders,

Imipramine, currently being used to treat panic disorders, pain, neuropathic pain and, Amphotericin b, currently used to treat infections, were the drugs predicted with highest scores.

For headache disorders,

Imipramine, Framycetin (currently being used to treat infection) and Trimethadione (currently being used in control of absence seizures that are refractory to treatment with other medications.

IV. Conclusion

It was observed that, for the disease where a higher number of genes were considered (migraine disorders, in our case), the results were more accurate as compared to that where a lesser number of genes were considered.

Also, for headache disorders data, multiple number of drugs in the training set had a zero score (because of having lesser number of genes in consideration), which made the results difficult to converge as compared to migraine disorders.

Our study provides some potential drugs that can be used to treat a particular disease.These drugs can be further analyzed to check whether these drugs are safe to be used for that disease, and for any other harmful side effects these drugs might have.

Case studies for migraine and headache disorders indicate that our method could be used for predicting new drugs for some indications based on the similarity between chemical structures and gene similarities.

V. REFERENCES

[1] Drugbank. [Online]. Available: https://www.drugbank.ca
[2] David Rogers and Mathew Hahn, SciTegic, Inc., High-Throughput Data Analysis. 1.Extended-Connectivity Fingerprints: High-Dimensional Descriptor for Molecular Data Analysis
[3] D. Bajusz, A. R´acz, and K. H´eberger, "Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations?" Journal of cheminformatics, vol. 7, no. 1, p. 20, 2015.
[4] T. F. Smith, M. S. Waterman, and C. Burks, "The statistical distribution of nucleic acid similarities," Nucleic Acids Research, vol. 13, no. 2, pp. 645–656, 1985.
[5] Jiao Li and Zhiyong Lu, A New Method for Computational Drug Repositioning Using Drug Pairwise Similarity.
[6] Kevin Bleakley and Yoshihiro Yamanishi. Supervised prediction of drug–target interactions using bipartite local models.
[7] Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledge base for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36(Database issue):D901–6.
[8] Sahil Manchanda, Ashish Anand. Representation learning of drug and disease terms for drug repositioning.
[9]ChemmineR documentation [online] https://www.bioconductor.org/packages/devel/bioc/vignettes/ChemmineR/inst/doc/ChemmineR.html
[10] Horacio Pérez-Sánchez*, Ginés D. Guerrero, José M. García, Jorge Peña, José M. Cecilia, Gaspar Cano, Sergio Orts-Escolano, and JoséGarcía-Rodríguez .Improving Drug Discovery Using a Neural Networks Based Parallel Scoring Function.