ISSN No:-2456-2165

Drug – Disease Association for Drug Repositioning Using Machine Learning

Jishnu Raj R
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

Sherin K R
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

Neelima philopaul
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

Shilji Rajan
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

AmbilI M P
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

Deepthi K
Department of Computer Science
College of Engineering Vadakara
Kozhikode,India

Abstract:- Number of unknown and rare diseases are increas-ing day by day. There are only 2000 drugs and more than 35000 known diseases. Existing drugs can be used to treat new diseases.It is known as drug-repositioning.Since drug discovery is a costly and time consuming process computational methods are necessary. Machine learning methods can be applied for drug repositioning. A deep-neural network is trained with drug sim-ilarity , disease similarity and known drug-disease associations. Once trained, the neural network can be used to predict new drug-disease associations. The neural network gives new drug-disease associations with its corresponding probabilities.

Keywords:- *Machine learning, Bioinformatics, Drug reposi-tioning.*

I. INTRODUCTION

Recently a new disease COVID-19(corona virus dis-ease)causes death of more than 200,000 people all around the globe.New disease called nipah causes 17 deaths in kerala.

There are 2000 drugs and 35000+ known diseases.We trained a deep neural network model with corresponding data sets. The datasets include drug- drug similarity, disease-disease similarity and existing drug-disease associations. When we enter name of a disease and corresponding similarity data, the model gives a list of drugs corresponding to it along with their probability. This process helps to reduce the time consumption in the process of new drug discovery, and will help to act fastly on extreme conditions like COVID-19

II. RELATED WORKS

- Predicting Drug-disease Association via using Gaussian interaction profile and Kernal-based autoencoder. (Research article in Hindawi biomed international research journal)
- A deep ensemble model to predict miRNA-disease asso-ciation (Research article published in Scientific reports online journal)
- A machine Learning approach to Drug Repositioning based on Drug expression profiles : Applications to schizophrenia and depression /anxiety disorder. (A research paper on Bio Resources and molecular research of common disease)
- Drug repositioning of herbal compounds via a Machine Learning approach. (From international workshop on data and text mining in bio-medical informatics, Italy)
- Changing Trends in Computational Drug Repositioning (Pharmaceuticals Journal paper).

III. THE PROPOSED SYSTEM

The proposed system is a software which take a disease name and will give a list of drugs matches to it along with its probability. A neural network based machine learning method is used here. The algorithm is trained with all available data sets.Final output is predicted from this potential Drug-Disease association that it learned.

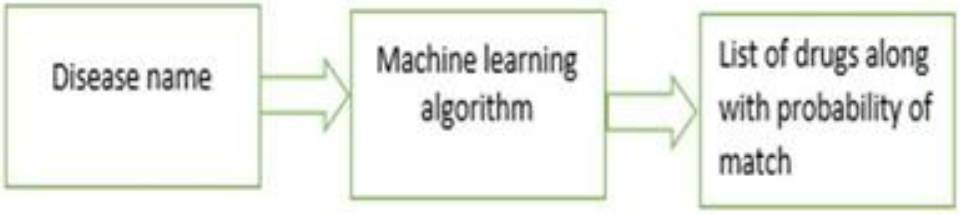

Fig 1:- User Interface

We used a 3 layer neural network for drug-disease association predictions. The learning process is based on the fundamental neural network equation.

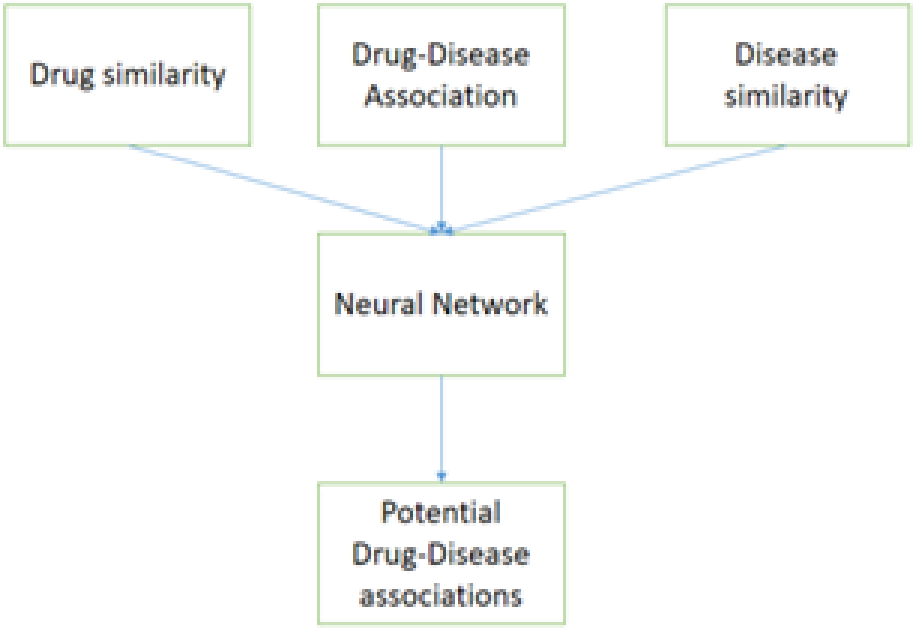

Fig 2:- system architecture of proposed system

A. Drug similarity matrices

We used two kind of data sets. First one is drug similarity matrices. Three matrices are there which represent drug sim-ilarities. All of them are represented as adjacency matrices. The rows and columns of matrices represents name of drugs. The similarity values ranges from 0 to 1. Similarity to itself will be 1.The three matrices are,

- Chemical structure similarity
- Functional similarity
- Side effect similarity

$$Di = \frac{Dc + Df + Ds}{3}$$

$$where\ Dc = Chemical\ steucture\ similarity\ matrix$$

$$Df = Functional\ similarity\ matrix$$

$$Ds = Side\ effect\ similarity\ matrix$$

The values of the three matrices are averaged together to find the drug similarity matrix.

B. Disease similarity matrix

Second kind of data set used is Disease similarity matrix. Disease similarity matrix is calculated from phenotype of each disease.

C. Neural network

The type of machine learning algorithm that we are using is neural networks.Since the most popular neural network algorithm is back propagation algorithm, we use the same here.

$$y = wx + b$$

$$Where\ y\ and\ x\ are\ parapeters$$

$$w = Weight$$

$$b = Biase$$

Here learning process is iterative where records (rows) are presented to neural network one at a time.Such input records will be feature vectors, which are combination of drug and dis-ease similarity information. Feature vectors are made by concatenating rows of drug similarity and disease similarity matrices.For example,to make the feature vector of a particular drug-disease combination, the corresponding row of drug similarity matrix is combined with the corresponding row of disease similarity matrix.Each time weight associated with the input are adjusted. This process is repeated many times. The well trained neural network will learn the similarity between each drug and disease. So it will be able to predict the most appropriate drug for a disease.

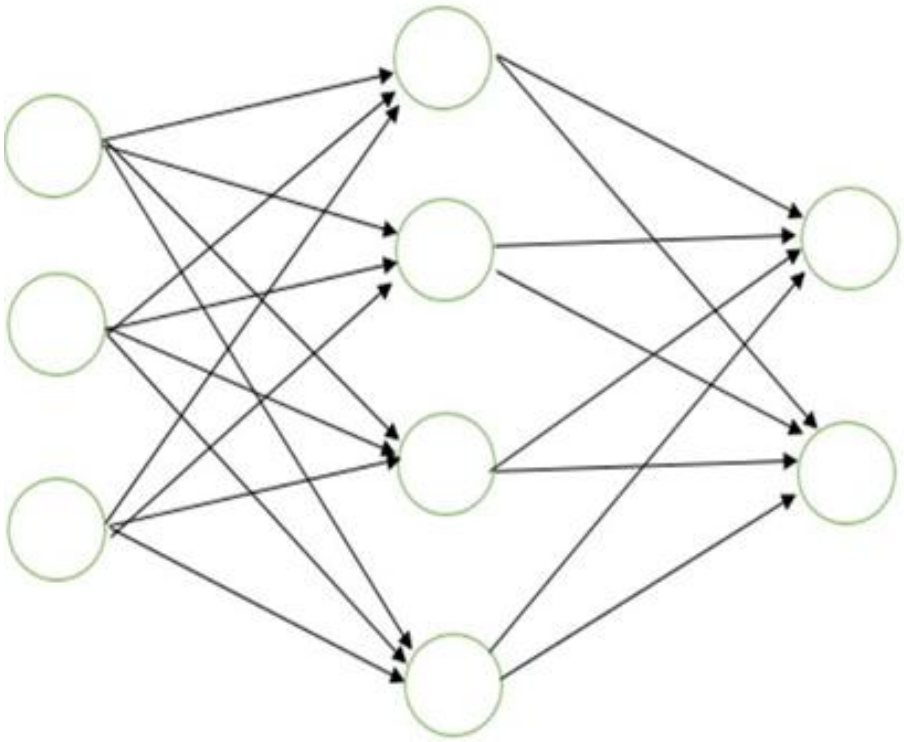

Fig 3:- Neural network structure

IV. APPLICATION

We can predict the drugs corresponding to rare diseases. Under extreme conditions like nipah and covid-19 outbreak discovery of new drug through biological process will be very difficult and time consuming. So this will help to take sudden actions against such cases.

V. CONCLUSION

When COVID-19 outbreak was happened in wuhan, china, it spread all over the world and causes death of more than 50000 people. However many patients where recovered from this using the same drug used for HIV disease. This was because COVID-19 and HIV disease has some similarity. This means drug repositioning is possible. But it is very difficult to find such similarity manually,because there are more than 35000 diseases and around 2000 drugs. So computational methods are necessary. Even though there are existing systems for the same, we can improve its efficiency through using better machine learning algorithms and better data sets. The proposed system uses a new method to do the same. This can be used in biological research centers, and they can verify the output by simple experiments (usually conducted on rats). This will help fast actions in many conditions like COVID-19 outbreak.

REFERENCES

[1]. Luo, H., Li, M., Wang, S., Liu, Q., Li, Y., Wang, J. (2018). Compu-tational drug repositioning using low-rank matrix ap- proximation and randomized algorithms. Bioinformatics, 34(11), 1904-1912.
[2]. Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., Lu, Z. (2015). A survey of current trends in computational drug repositioning. Briefings in bioinformatics, 17(1), 2-12.
[3]. Hattori, M., Okuno, Y., Goto, S., Kanehisa, M. (2003). Development of a chemical structure comparison method for in- tegrated analysis of chemical and genomic information in the metabolic pathways. Journal of the American Chemical Society, 125(39), 11853-11865.
[4]. Jiang, H. J., Huang, Y. A., You, Z. H. (2019). Predicting Drug-Disease Associations via Using Gaussian Interaction Pro- file and Kernel-Based Autoencoder. BioMed Research International, 2019.
[5]. Luo, H., Wang, J., Li, M., Luo, J., Peng, X., Wu, F. X., Pan, Y. (2016). Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. Bioinformatics, 32(17), 2664-2671.
[6]. Kuhn, M., Letunic, I., Jensen, L. J., Bork, P. (2015). The SIDER database of drugs and side effects. Nucleic acids re- search, 44(D1), D1075-D1079.
[7]. Liu, Y., Zeng, X., He, Z., Zou, Q. (2016). Inferring microRNA-disease associations by random walk on a heterogene- ous network with multiple data sources. IEEE/ACM transactions on computational biology and bioinformatics, 14(4), 905- 915.