International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 11 | Nov 2020 www.irjet.net p-ISSN: 2395-0072

Different Approaches for Finding Micro RNA and Disease Association

Jaison Mathew John[1]

[1]*Department of Computer Science and Engineering, Govt. Engineering College, RIT, Kottayam*

Abstract - *MicroRNAs are a sort of non coding RNAs with approximately 22nt nucleotides. Increasing evidences have proven that miRNAs play important roles in lots of human sicknesses. The identity of human disease related miRNAs is beneficial to discover the underlying pathogenesis of diseases. More and greater experimental proven associations among miRNAs and diseases were stated within side the latest studies, which provide beneficial records for new miRNA-disease association discovery. In this work, a computational framework, KBMFMDI, to expect the associations among miRNAs and diseases based on their similarities. The collection and characteristic records of miRNAs are used to degree similarity amongst miRNAs whilst the semantic and function records of disease are used to degree similarity amongst diseases, respectively. In addition, the kernalized Bayesian matrix factorization approach and self organizing maps is hired to infer potential miRNA disease associations via way of means of integrating those data sources.*

Key Words: ***MicroRNA, KBMFMDI, Kernel, Self organizing maps, Bayesian, machine learning***

1. INTRODUCTION

The first micro RNA was discovered 20 years ago. It includes a large range of physiologocal processes. It also includes wide ranges of compulsive or chronic processes [1]. Since It is clear fact that all the micro RNA's are find out or discovered only through biologically, miRNA's can be greatly affected by mutations, may cause a great role in human disease pathogens [2]. Recent studies have shown that micro RNA's play an important role in human life. Now days, many studies aim to apply miRNAs for diagnostic and therapeutic applications in human diseases. The miRNA exhibits their function by regulating expression of disease genes. As such, the abnormality of miRNAs, the dysregulation of miRNAs and dysfunction of miRNA bio genesis may result in many diseases, including cancers, inherited diseases, nervous system diseases, and so on. For example, miR-21 can control the expression of gene MAP2K3 expression, which is a tumor repressor gene and has association with hepatocellular carcinoma cell expansion. Therefore, identifying disease-related miRNAs can be helpful for exploring disease parthenogenesis and designing appropriate and effective treatments [3].

It has been mentioned that genes with comparable features are frequently implicated in comparable illnesses and vice versa. The homologous miRNAs are amassed into the identical miRNA own circle of relatives The seed regions (generally 2 eighth nucleotide from the fifty nine stop of miRNA) of miRNA sequences of the identical own circle of relatives are nearly identical. It has been mentioned that miRNAs are frequently determined in genomics clusters [4]. The clustered miRNAs are commonly transcribed collectively and much more likely related to the same illness. The clustered miRNAs are commonly transcribed collectively and which are related to the same illnesses. Based in this statistics we can calculate disease disease functional similarity, disease-disease semantic similarity. Identifying the connection between miRNAs and illnesses is with the aid of using the use of the experimental methods. And the cost of identifying relationship is greatly increased by the probe design.

2. LITERATURE REVIEW

M.Gohen introduces a way known as kernalized Bayesian matrix factorization [5] approach used to locate the similarity among miRNA and diseases. It integrates data sources from specific inputs. The major facts sources encompass miRNA-miRNA functional and sequence similarity and the second encompass the disease-disease functional and disease-disease semantic similarity. The identity of human associated micro rna is beneficial in coming across dangerous pathogens. It outputs a matrix that carries a rating among all of the miRNA' s and all of the diseases.

Yi Pan indicates that miRNA play vital roles in lots of biological processes. A variety of computational models were proposed to deduce miRNA-disease association [6]. MiRNA-miRNA similarity is primarily based totally on two elements taking into consideration along with the sequence and functional similarity. Kernalized Bayesian matrix factorization approach is effectively used to expect the associations among miRNAs and the diseases.

Jogile Kuklyte in his paper advise database HMDD [7]. The model of the database used is HMDDv2.0. It is a human micro-rna disease online database presenting complete development on miRNA deregulation in numerous human diseases. MiRbase acquire informations approximately nucleotide sequences. This paper consists of how successfully diseases are given. The lately released model 3.0 of Human MicroRNA Disease Database (HMDD v3.0) [8] manually collects a massive range of miRNA disease association entries from literature. Comparing to HMDD v2.zero, this new edition consists of 2-fold greater entries. Besides, the associations or the relationships had been greater as it should be classified.

L. Cheng SemFunSim [9] introduces a technique for measuring disease similarity via way of means of integrating semantic and gene functional association,. This technique makes use of powerful strategies to enhance accuracy of consequences the use of SemFunSim. It is clearly based at the semantic functional similarity. First of all, FunSim (Functional

similarity) is proposed to calculate disease similarity the use of disease-associated gene units in a weighted network of human gene function. Next, SemSim (Semantic Similarity) is devised to calculate disease similarity the use of the connection among diseases from Disease Ontology [10]. Finally, FunSim and SemSim are integrated to identify or measure the disease similarity.

Teuvo Kohonen, reduces the dimensions of data through using self-organizing neural networks [11]. In this paper, we gift an technique to cluster the different subjects of expertise from programming codes without manual labour. First, syntax trees are generated for programming codes, after which the similarities among them are computed in an effort to get a generalized mean of the syntax trees for the non-vectorial self organizing maps model. On the visualization map, the different topics of knowledge extracted from the programming codes may be gathered together.

3. METHODOLOGY

The kernalized Bayesian matrix factorization approach (KBMF) is an powerful approach to deduce a bipartite graph with the aid of using more than one statistics supply integration. MiRNAs and diseases are assigned to 2 domains. For inferring capability miRNA-disease interactions, we get a couple of kernel matrices, namely via way of means of calculating similarities of miRNAs and diseases, respectively.PM and PD denote the numbers of kernel matrices of miRNAs and diseases It can be observed that for miRNA, KBMF MDI first performs the kernel based nonlinear dimensionality reduction by using input kernel matrices AM. After projection, the kernel-specific components can be calculated. For each of the components up to PM and PD, we calculate the values of principal components. Then, the composite additives may be received with the aid of using linear aggregate of the kernel specific components. Finally based on the values of the (i, j) element of matrix F, the association of miRNA i and disease j can be predicted.

Another way is to use the self organizing maps. This is an efficient way to find the relation between various parameters. The predominant constructing block of SOM are neurons or nodes. The main task was integrating these four data sets which was later fed as input to SOM module [12]. For each disease the values in the other data sets are found and converted to an array. Similarity miRNA s are calculated to array. For each array we obtain four values, out of which three values are for diseases rest for miRNA. miRNA-miRNA functional similarities can be calculated based on the misim method [13]. miRNA-miRNA sequence similarities can be calculated based on the needleman wunch algorithm [14], Disease-Disease functional similarity [15] is to measure the similarity between two diseases, that can be obtained from gene similarity of the corresponding diseases. Gene similarities are available in HumanNet., Disease-Disease semantic similarities can be taken from the mesh database and humanNet [16].

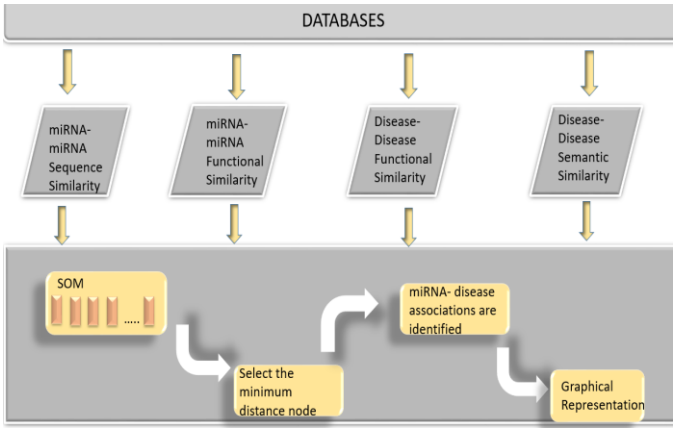

Fig -1: Methodology using SOM

We collect the four data sets of miRNA-miRNA sequence similarity, miRNA miRNA functional similarity, Disease-Disease Functional Similarity, Disease-Disease Semantic similarity. The HMDD (Human MIRNA disease dataset consists of 2163 miRNA disease association. For each Micro RNA, we have the corresponding associated disease with that micro RNA. The semantic value of diseases was found using the dag structure from the disease ontology [17]. Disease-Disease interactions or similarity were found using the idea that similar diseases tend to be associated with similar genes. The Disease Ontology(DO) dataset contains a score called log likelihood score that actually represents the value or score associated with all pairs of genes. This core is very important to find the disease disease similarity. It includes both the two types of similarity, i.e. disease-disease semantic similarity and disease-disease functional similarity.

The first value in the array consist miRNA-miRNA sequence similarity. The second value consists of miRNA-miRNA functional similarity and so on. In the data sets miRNA are represented row wise. Therefore, we cannot obtain miRNA-miRNA Sequence Similarity directly. For this we need to find out the miRNA with which that particular diseases have an association from a known database(HMDD). HMDD consist of known miRNA-disease information. If a disease has an association with more than one miRNA, then we select the miRNA which has maximum association with that disease. Similarly, miRNA-miRNA functional similarity can be found. The task of integrating all the above datasets was well done with the help of this method. The greatest task that was having the varied difficulty was eased with the help of this technique. It includes basically the steps which are dimensionality reduction [18], finding composite components, then finding the association value from the matrix that contains a score associated with a particular micro rna and a disease.

Each node is related to every different neuron, however connections are small. Each neuron related simply to 3 different neurons that we name near neighbors. There are many approaches to set up those connections, however the

maximum common one is to set up them into -dimensional grid. Each blue dot within side the parent is neuron and line among neurons manner that they're related. We name this association of neurons grid. Each node within side the grid has properties: position and connections to different nodes.

Fig -2: SOM Grid

Then we begin network training and role is the best factor that changes at some point of the training. There SOM process involves four major components. These components are very important in determining the position of the various neurons in the SOM Grid. The concept of Euclidean Distance is always taken into consideration.

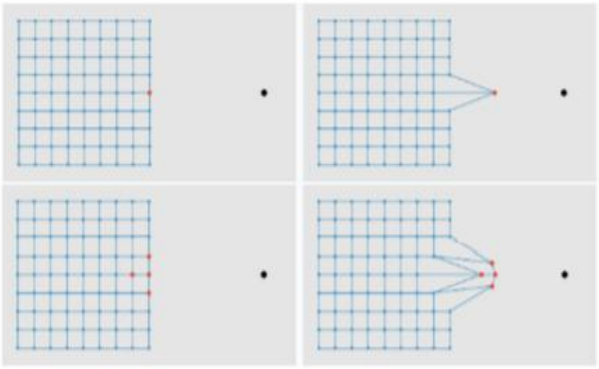

Fig -3: Working of SOM Grid

All the connections weights are initialized with randomly. Competition means all nodes compete for the ownership of the input pattern. Using Euclidean distance choose the best matching unit. Cooperation means Neuron that closest to this point we call neuron winner. But, instead of updating position of this neuron we find its neighbors. Note, that it's not the same as closest neighbors. Before training we specify special parameter known as learning radius. It defines the radius within which we consider other neuron as a neighbors. Adaptation means all the nodes values and its neighbors values are up to date. These methods are proven in determining. The input miRNA is checked with all of the nodes within side the SOM grid and the node which minimal distance is the winning node. After every mapping, the listing of diseases related to every miRNA s are up to date. But rather than updating role of the node, we discover its neighbours. Before training we specify unique parameter called learning radius [19]. It defines the radius inside which we remember different nodes as a neighbors. The weight within side the radius is up to date. After weight updating, training is performed. As end result of training, we acquire an up to date values of grid. In the testing phase, whilst a brand new miRNA is given as enter to the model, the model predicts the diseases related to that miRNA primarily based totally at the trained associations.

4. RESULTS

AN ROC (Receiver Over Characteristics) is a graph showing the overall performance of a classification model in any respect of classification thresholds. This curve plots two parameters that are True Positive Rate and False Positive Rate. An ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. AUC means Area Under the ROC Curve AUC stands for" Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1).

The area under the ROC curve(AUC) is utilized to measure the performance. We compare the SOM with work done. Figure below shows that area of curve obtained from SOM (92.76) is greater that KBMF method (86.02) which implies better accuracy of SOM.

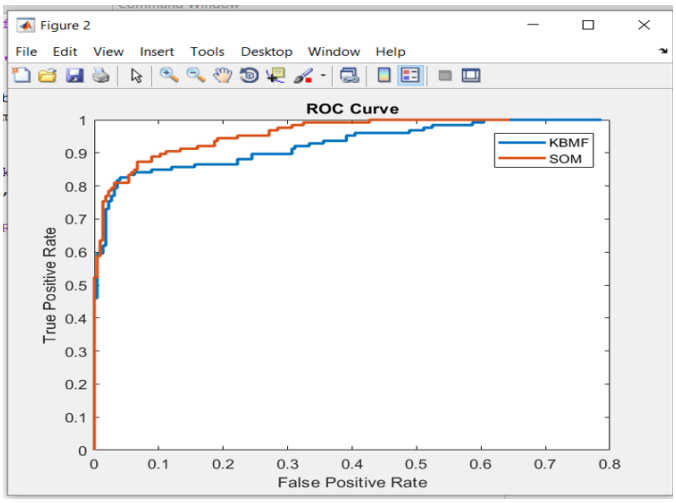

Fig -4: ROC CURVE

5. CONCLUSIONS

The integrating the miRNA's and sicknesses similarities helped in to locate first-rate matching sicknesses. In the preceding country of artwork techniques makes use of simplest both miRNA or Disease similarities. But in our approach we make use of the four similarities to expect the miRNA and sicknesses. So it's far very beneficial to find out find the damaging disorder pathogens in the back of the

miRNA. With the help of SOM, we have achieved higher accuracy than the KBMF technique.

The future work consists of the use add extra similarities primarily based totally on miRNA or diseases. There are lot of exiting strategies which makes use of simplest the information of disease or miRNA which bring about terrible performance. So if we modify it get higher result. If we add chromosomal coordinates of human miRNA sets, it can provide extra clearer and accurate results.

REFERENCES

[1] M.M. Akhtar, L. Micolucci, M.S. Islam, F. Olivieri and A.D. Pro-copio, "Bioinformatic tools for microRNA dissection," Nucleic Acids Res, vol. 4,no. 11,pp. 22-44, 2016.

[2] V. Ambros, "The functions of animal microRNAs," Nature, vol. 431, no. 7006, pp. 350-355, 2004.

[3] G.A. Calin and C.M.Croce, "MicroRNA signatures in human cancers," Nat Rev Cancer, vol. 6, no. 11,pp. 857-866, 2006.

[4] R.C. Lee,R.L. Feinbaum,V. Ambros, "The C. elegans hetero-chronic genelin-4 encodes small RNAs with antisense comple-mentarity to lin-14," Cell, vol. 75,no. 5,pp. 843-854, 1993

[5] M.Gonen and S.kaski, "Kernalized Bayesian Matrix Factorization," IEEE transactoins on pattern Analysis and Machine Intelligence, vol.36, no 10,pp. 2047-2060,2015.

[6] Q. Zou,J. Li,L. Song,X. Zeng,and G. Wang, "Similarity compu-tation strategies in the microRNA-disease network: a survey," Brief Funct Genomics, preprint, 1 Jul. 2015, pii: elv024.(PrePrint).

[7] HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Li Y, Qui C, Tu J, Geng B,Yang J, Jiang T,Cui Q.

[8] HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Huang Z, Shi J, Gao Y,Cul C, Zhang S, Li J, Zhou Y, Cui Q.

[9] L. Cheng,J.Li,P.Ju,J.Peng,Y.Wang, "SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association," PLoS One, vol. 9,no. 6,p. e99415, 2014.

[10] M. Lu,Q.Zhang,M.Deng,J.Miao,Y.Guo,W.Gaoand Q.Cui, "An analysis of human microRNA and disease associations," PLoS One, vol. 3,no. 10,p. e3420, 2008.

[11] T. Kohonen "The Self oraganizing maps" Volume: 78, Issue: 9, Sept. 1990.

[12] Taku Haraguchi, Haruna Matsushita, Yoshifumi Nishio, " Community self organizing maps and its application to data extraction.

[13] M. Ammad-ud-din,E.Georgii,M.Gönen,T.Laitinen,O.Kal-lioniemi, K. Wennerberg,A.Posoand S.Kaski, "Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization," J Chem Inf Model, vol. 54,no. 8,pp. 2347-59, 2014.

[14] S. B. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins", *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443-453, 1970.

[15] .P. Maji and S. Paul, Scalable Pattern Recognition Algorithms: Applications in Computational biology and Bioinformatics, London, U.K Springer,Apr. 2014.

[16] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov and F. Cheng, "deepDR: A network-based deep learning approach to in silico drug repositioning", *Bioinformatics*.

[17] R. Bunescu and R. Mooney, "A Shortest Path Dependency Kernel for Relation Extraction", *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pp. 724-731, 2005.

[18] H. Yin, "Nonlinear dimensionality reduction and data visualization: a review", *Int. Journal of Automation and Computing*, vol. 4, pp. 294-303, 2007.

[19] Y. E. Jian, L. D. Ge, Y. X. Wu, An application of improved RBF neural network in modulation recognition, Acta Automatic Asinica, Vol.33, No.6, 652-654, 2007.