International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Predicting miRNA-Disese Associations based on miRNA and Disease similarity

Aparna A Nair
MASTER of TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY,
KOCHI, ERNAKULAM, 682 022.

https://doi.org/10.55041/ijsmt.v2i6.051

Cite this Article: Nair, A. A. (2026). Predicting miRNA-Disese Associations based on miRNA and Disease similarity. International Journal of Science, Strategic Management and Technology, *02*(6). https://doi.org/10.55041/ijsmt.v2i6.051

Abstract

Many different approaches have been used to find miRNA and disease associations.But none of the state-of-art methods use global information instead of using local infor- mation. As of now, this work has been done considering maximum 2 similarities of miRNA or Disease. In the proposed method, we tried finding MiRNA disease asso- ciation using five similarities. The work also calculates the accuracy of the resulting miRNA -Disease prediction using a SOM approach.

Chapter 1 Introduction

This chapter begins with the section 1.1 which describes the background of the problem. Then the motivation for the problem. Followed by the problem statement and objectives. Subsequently the organization of the thesis and proposed works are given briefly.

1.1 Background

The discovery of the first microRNA was about 20 years ago and it involved in variety of physiological and pathological processes.Since the discovery of the powerful effect miRNAs are through biological processes,mutations affecting miRNA may play an important role in the pathogenesis of human diseases. According to latest recent years microRNAs plays an important role in various human diseases.Now a days many studies aim to apply miRNAs for diagnostic and therapeutic applications in human diseases. Over the past several years microRNAs have been found to play a major role in various human diseases. In addition, many studies aim to apply miRNAs for diagnostic and therapeutic applications in

human diseases. [5]

The miRNA's plays an important roles in many biological processes such as cell development, proliferation, differentiation, apoptosis etc. The miRNA exhibits their function by regulating expression of disease genes.As a result, the abnormality of miRNAs, the imbalance of miRNAs and dysfunction of miRNA biogenesis may result in many diseases, including cancers, inherited diseases, nervous system diseases, and so on. The miRNA exhibits their function by reg- ulating expression of disease genes. As such, the abnormality of miRNAs, the dysregulation of miRNAs and dysfunction of miRNA biogenesis may result in many diseases, including cancers, inherited diseases, nervous system diseases, and so on [6]. For example, miR-21 can control the expression of gene MAP2K3 expression, which is a tumor repressor gene and has association with hepatocel- lular carcinoma cell expansion. Therefore, identifying disease-related miRNAs can be helpful for exploring disease parthenogenesis and designing appropriate and effective treatments [6].

It has been reported that genes with similar functions are often implicated in similar diseases and vice versa [7].The homologous miRNAs are collected into the same miRNA family The seed regions (normally 2–8th nucleotide from the 59 end of miRNA) of miRNA sequences of the same family are almost identical. It has been reported that miRNAs are often found in genomic clusters.The clus- tered miRNAs are usually transcribed together and more likely associated with the similar diseases. [7]The clustered miRNAs are usually transcribed together and that are associated with the similar diseases.Based on this information we can calculate disease-disease DAG similarity, disease-disease semantic similarity and disease-disease functional similarity. Identifying the relationship between miRNAs and diseases is by using the experimental methods. And the cost of identifing relatioship is greatly increased by the probe design

. Therefore, development of computational methods that estimate the reliable disease-related miRNA candidates is a worth complement to experi- mental studies. So far, little work is obtained in predicting disease miRNAs. By combining machine learning techniques a lot more information can be obtained by integrating different similrity database and eventually a better prediction of miRNA and diseases is possible. There are quite a few papers which integrating data from different sources and using different techniques to the prediction of miRNA and diseases. These approaches have been discussed in the Literature survey section.

1.1.1 Basics of miRNA

1.1.2 Chromosome

Chromosomes are thread-like structures located inside the nucleus of living be- ings [1] . Each chromosome is made of protein and a single molecule of deoxyri- bonucleic acid (DNA) [1]and RNA. DNA contains genetic information that is transfer from one generation to another. A detailed diagram showing chromo- some structure and DNA is provided in Figure1.1. [1].

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

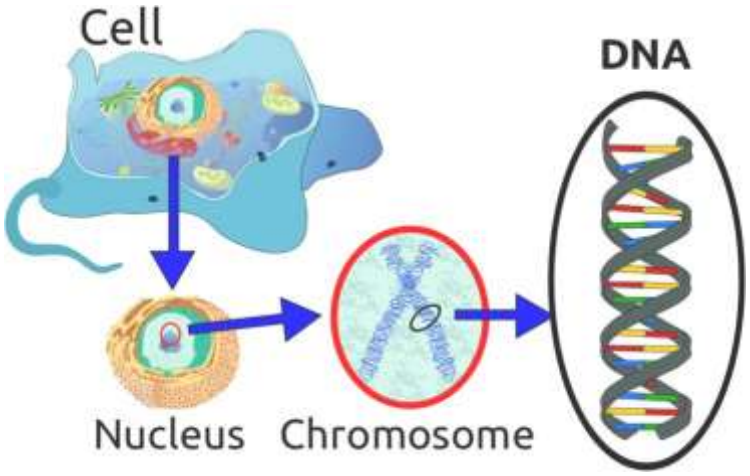

Figure 1.1: Chromosome structure [1]

.

1.1.3

1.1.4 Central Dogma of molecular biology

The central dogma of molecular biology deals with the detailed residue-by- residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid [8]. The basic three process of Replication, Transcription and Translation is depicted in the Figure 1.2.

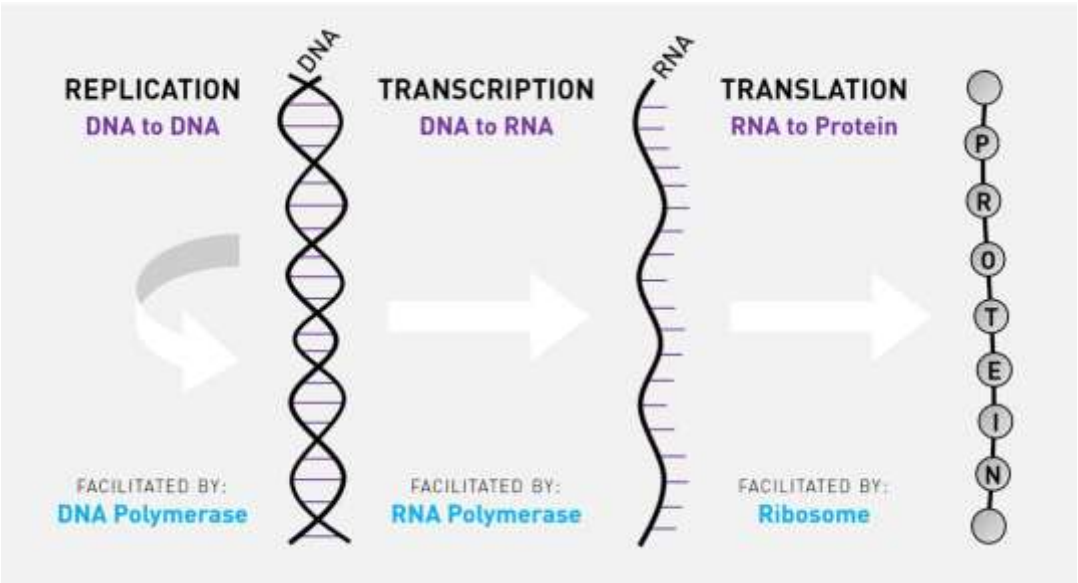

Figure 1.2: Central Dogma of Molecular Biology [2]

1.1.5 miRNA

Mature microRNAs (miRNAs) are a class of small non-coding RNA molecules, about 21–25 nucleotide in length.Due to the up-regulation and down-regulation of miRNAs causing different dangerous

diseases . miRNA plays an impor- tant roles including translational repression, mRNA cleavage, and deadenyla- tion, microRNAs are relatively peer to one or better messenger RNA (mRNA) molecules. [3].

1.2 Motivation

In the field of Genetics, integrating data from various sources is the greatest challenge due to the complexity of heterogeneous data available out there. Inte- grating such large sources of data can be tedious for humans but can be carried out easily by a computer science student. Research works carried out in the filed of Bioinformatics using Machine Learning approaches proved to be a boon to humanity by finding MiRNA disease associations which were earlier unknown to biologists. These associations in turn paved way for exploring disease pathogens and designing effective treatments.

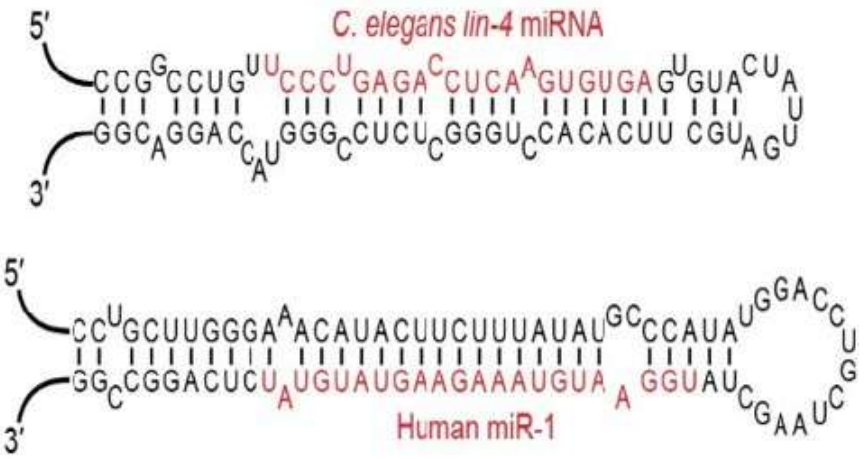

Figure 1.3: miRNAstructure [3]

.

1.3 Problem Statement and Objectives

The major objectives identified are the following:

1. Prediction of suitable disease corresponding to the miRNA based on miRNA and disease similarity.

2. Integrating different data source related to miRNA's similarities and dis- eases similarities.

1.4 Major contributions of the thesis

1. To integrate multiple side information sources by coupling matrix factor- ization.

2. A global method based on multiple kernel learning is used to predict the miRNA and disese association

1.5 Challenges in this Thesis

1. Dataset availability.
2. Integration of different datasets.
3. Understanding molecular mechanism of disease.
4. The available dataset has poor quality and missing values.
5. Understanding molecular mechanism of diseases.
6. Integration from different machine platforms raises issues.

1.6 Thesis Outline

This thesis report is organized into five chapters, including Chapter 1. The contents of chapters 2 to 5 are briefly outlined as follows:

Chapter 2: deals with the survey of the literature

Chapter 3: presents the Methodology used

Chapter 4: deals with the Result analysis and Discussion

Chapter 5: presents the Conclusion and Future Scope

Chapter 2 Literature Survey

In order to gain a better understanding on the topic a lot of references including Published papers and Web Databases were referred. Alongside that some software's out there was also referenced and used. The details on these topics are explained below.

2.1 Literature papers

The identification of human disease-related microRNAs (disease miRNAs) is important for analysing their involvement in the focalization of diseases. It is identified that miRNAs with similar functions are often associated with simi- lar diseases. Xuan.P,Han et al. [7] describes new prediction method, HDMP, based on weighted k most similar neighbors is presented for predicting dis- ease miRNAs. The paper uses the members of miRNA family or cluster are assigned higher weight since they are more probably associated with similar diseases.Experiments validated that HDMP achieved significantly higher pre- diction performance than existing FCS methods([9]) .HDMP results of 5-fold cross validation and those of the updated dataset validation demonstrated that HDMP has significantly higher accuracy in recovering the known disease miR- NAs. . The results had an average correct rate of 78.90% and highest correct rate of 82.20% .

Chen.X et al. [10] developed two machine learning methods, Regular- ized Least Squares for MiRNA-Disease Association (RLSMDA) and Restricted

Boltzmann machine for multiple types of miRNA-disease association predic- tion (RBMM MDA), to discover the certify relation-ships between diseases and miRNAs .The AUC of RWRMDA is 0.8617, which has significantly increase the performance of previous computational method based on the hyper geometric distribution .For RLSMDA, AUC in local and global LOOCV is 0.8450 It is much likely that the performance of RLSMDA would be further improved after introducing the information of miRNA family and cluster into its model. Ex- cellent performance certify RLSMDA can recover known experimentally verified miRNA–disease associations and hence has the potential to predict potential associations.

L.Cheng, et al. [11] gives idea about number of databases have been ac- quire to bring together disease-related molecular, phenotypic and environmen- tal features (DR-MPEs), such as genes, microRNAs, genetic variations, drugs, phenotypes and environmental factors. However, each of current databases fo-cused on only one or two database of Disease related molecular ,phenotypic and enviormental factors. If a integrated all existing database it is very helpful in cross checking which can establish significant associations among disease- related databases and link them to provide a global view of human disease at the biological level.To establish an combined disease-collaborated database, dis- ease vocabularies used in different databases are mapped to Disease Ontology (DO) through semantically. In Medical Subject Headings (MeSH) and Online Mendelian Inheritance there are 4,274 and 4,176 disease terms in o(OMIM) On- line Mendelian Inheritance in Man are mapped to DO. Then, the relationships between DR-MPEs and diseases are extracted and combine from different source databases for reducing the data repetition. After that a network visualization tool using Cytoscape Web plugin has been

discovered in SIDD. It increases the SIDD usage when viewing the relationships between diseases and DR-MPEs.

Juan Wang et al. [12] It is known that genes with similar functions are of- ten associated with similar diseases, and the relationship of different diseases can be represented by a structure of directed acyclic graph (DAG). This is also true for miRNA genes. Therefore, it is possible to find miRNA functional similarity by calculating the similarity of their associated disease DAG. Based on there observations and rapidly accumulated human miRNA-disease association data, we presented a method to infer the pairwise functional similarity and functional network for human miRNAs based on the structures of their disease relation- ships. There are lot of comparisons in state of arts showed that the calculated functional similarity of miRNA is well associated with previous knowledge of miRNA functional relationship. It is importantly used to predict novel miRNA biomarkers and it also infer miRNA novel potential functions and associated diseases for miRNAs.

Subsequently, this method can be easily extended to other species when sufficient miRNA-associated disease datas are available for specific species.

Zou et al. [13]proposed two methods based on social network analysis, KATZ and CATAPULT, to infer potential miRNA-disease associations.It is a supervised machine learning method. KATZ succeeds in processing social heterogeneous network links to achieve prediction.KATZ uses the simple mea- sure on the social network to predict the potential microRNA-disease relation- ship. CATAPULT is a supervised learning method which help for the predic- tion miRNA and disease Zou et al. [13]proposed two methods based on social network analysis, KATZ and CATAPULT, to infer potential miRNA-disease associations.It is a supervised machine learning method. KATZ succeeds in processing social heterogeneous network links to achieve prediction.KATZ uses the simple measure on the social network to predict the potential microRNA- disease relationship. CATAPULT is a supervised learning method that helps to find association of diseases and miRNA ,it also uses a biased SVM . Biological experiment prove than it KATS CATAPULT gives the true associations.

Biological experiment prove than it KATS CATAPULT gives the true associations.

2.2 Evaluation measures

The Metrics used to evaluate includes AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. It is one of the most important evaluation metrics for checking any classification model's performance.

2.3 Review of Different methods

Table 2.1: Review of Different methods

Authors	Data Set's	Techniques	Performance
Xuan.P,Han, et. al. [7]	• HMDD • MiRBase.	HDMP METHOD.	• Average Accuracy : 78.90% • Highest Accuracy : 82.20% • It assigned higher weights to mem- bers in the same miRNA cluster or family when miRNA func- tional similarity is calculated. • Improved the func- tional similarity es- timation method.
Chen.X et al. [10]	• SIDD Database • HMDD database	RLSMDA METHOD.	• It reconstructs the missing association for all diseases. • RLSMDA don't need negative samples selection

Continued on next page

Table 2.1 – continued from previous page

Authors	**Data Set's**	**Techniques employed**	**Performance**
L.Cheng, et al [14]	• Disease ontology Database • Mesh Database.	Integrated DBs are mapped to DO .	• integrates 18 disease-DBs,network visu- alization is used ot browse multiple types of DR-MPEs in a view
Juan Wang, et al. [7]	• HMDD • MiRbASE	MISIM METHOD .	• Region based Gene variation • Gender based Gene variation • 2 most important miRNA's

2.4 Observations

Though there are various approaches,to find the prediction of mirna and disease association.HDMP method have better accuracy but its totally dependent on the weights of mirna and genomic cluster. SIDD method give a new idea to integrte the different disease-related molecular, phenotypic and environmental features (DR-MPEs), such as genes, non-coding RNAs, genetic variations, drugs, phenotypes and environmental factors. sequence and functional similarities of miRNA's . RLSMDA have the advantage they dont take any negative samples for the prediction of miRNA and disease.Likewise diseases semantic similar- ity,functional similarity and DAG similarities. But the different approaches do gives further insight on how to attain better knowledge in prediction of diseases.

2.5 Conclusions

The main methods considered for this study was done by Juan Wang et al [7],Xuann P,Han et al. [14] and L.Cheng et al. [10]. After reading their work it became clear that there are a lot of factors that are yet to connected with the existing knowledge. This project aims to improve that insight to gain a better understanding of how predicting miRNA -Disease using existing all similarities.

Chapter 3 Methodology

In this chapter the detailed step wise methodology applied in this project is described. Starting of with the Introduction, Inputs Used, and Methods Used

3.1 Introduction

This project focuses on predicting disease and miRNA based the disease Sim- ilarities and miRNA similarities. There are different types of similarites used here that are miRNA-miRNA sequence similarity, miRNA-miRNA functional similarity,Disease-Disease functional similarity, Disease-Disease semantic simi- larity, Disease- Disease DAG similarity and that are combined into different manner. The SOM method gives the accurate disease causing miRNA based on the similarites of diseases and miRNA.

Softwares and tools used :

- Python [15]
- Matlab [16]
- multi-Harmony [17]
- Cytoscape [18]

Inputs Used :

miRNA- Disease Dataset's were used. The datasets are as follows HMDD

[19], [20], MiRbASE [21],Disese ontology [22],Mesh Database [23], miRNA- miRNA similarities [24], disease-disease similarites [25]. The dataset, MiRNA and Disease similarities used for the creation of the dataset and dimensions are given in the table below.

Table 3.1: Data Set Details

Dataset Name	**Descriptions**	**Dataset Name**	**Descriptions**
HMDD [19]	• Human miRNA-Disease Associ- ation Data • manually collected 32281 miRNA-disease associa- tion • miRNA name, disease name, the refer- ence PubMed ID, and the evidence supporting the miRNA-association	SIDD [20]	• It integrates 18 disease- associated databases • The process of mapping from MeSH to DO has three major steps: MFR, MFS,MFI • mapping results are man- ually checked to avoid the mapping errors as much as possible
Continued on next page			

Table 3.1 – continued from previous page

Dataset Name	**Descriptions**	**Dataset Name**	**Descriptions**
MIRBASE [21]	• It is a search-able database of published miRNA se-quences and annotation.	Disease Ontology [22]	• It is a Biomedical community with consistent, reusable and sustainable de-scriptions of human disease terms, pheno- type characteristics and related medical vocabulary disease concepts.
HumanNET [26]	• Identifying novel disease genes based on the observation that similar mutational phenotypes arise from functionally related genes. • Get the functional similaries be- tween diseases	miRNA-miRNA functional similari	

Table 3.1 – contin

Dataset Name	**Descriptions**

		ty [24]	mi RN A-mi RN A seq uen ce simi larit y [24]	• It is collected from miRbASE and HMDD
			Dise ase-Dise ase Fun ctio nal Sim ilari ty [25]	• Datas are col- lected from the SIDD and HumanNET
				• miRNA miRNA func- tional similarities are collected from HMDD and misim omix

Continued on next page

3.2 Methodology

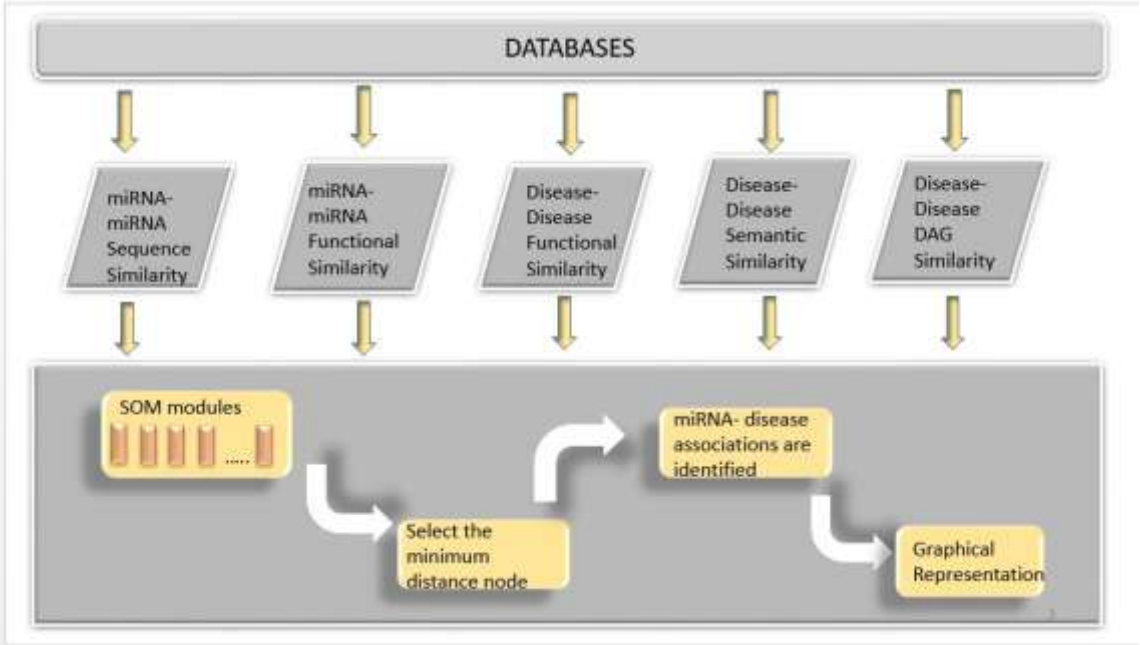

Figure 3.1: Block diagram of methodology

3.2.1 INPUTS

1. miRNA-miRNA functional similarities can be calculated based on the misim method

2. miRNA-miRNA sequence similarities can be calculated based on the needle- man wunch algorithm

3. Disease-Disease functional similarity is to measure the similarity between two diseases , that can be obtained from gene similarity of the correspond- ing diseases. Gene similarities are available in HumanNet

4. Disease-Disease DAG similarities represent the corresponding node,ancestor node and edges of the corresponding link

5. Disease-Disease semantic similarities can be taken from the mesh database and humannet

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

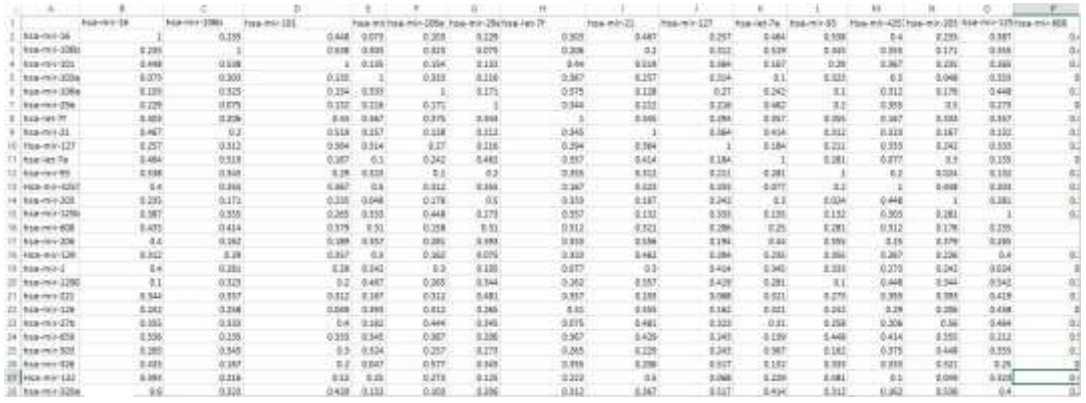

Figure 3.2: miRNA-miRNA Sequence Similarity

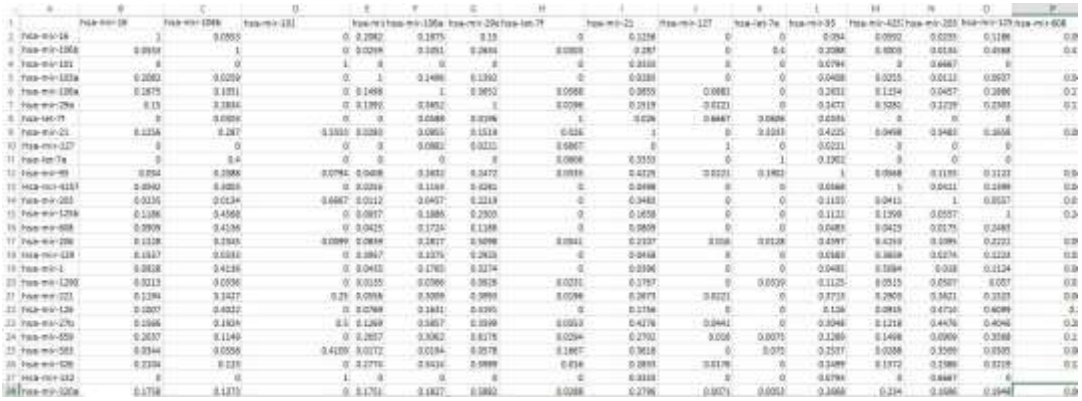

Figure 3.3: miRNA-miRNA Functional Similarity

	A	B	C	D	E	F	G
1		Pregnancy, Ectopic	Vascular Diseases	Nerve Sheath Neoplasms	Vitiligo	Ependymoma	Period
2	Pregnancy, Ectopic	1	0	0	0	0	
3	Vascular Diseases	0	1	0.5118	0.5912	0.453	
4	Nerve Sheath Neoplasms	0	0.5118	1	0.6267	0	
5	Vitiligo	0	0.5912	0.6267	1	0	
6	Ependymoma	0	0.453	0	0	1	
7	Periodontal Diseases	0	0	0	0	0	
8	Huntington Disease	0.4902	0	0	0	0	
9	Hyperlipidemias	0.4902	0	0	0	0	
10	Endomyocardial Fibrosis	0	0.5118	0.5118	0.5118	0	
11	Diabetes Mellitus, Type 2	0	0	0	0	0.5069	
12	Meningioma	0.4902	0	0	0	0	
13	Osteoarthritis	0	0	0	0	0	
14	Tuberculosis, Pulmonary	0	0	0	0	0	
15	Giant Cell Tumors	0	0.4848	0.3567	0.4848	0.453	
16	Glomerulonephritis	0	0.4848	0.3567	0.4848	0	
17	Gastric Neoplasms	0.4902	0.4029	0.4029	0.4029	0	
18	Osteoporosis	0	0.5118	0.6267	0.6267	0	
19	Irritable Bowel Syndrome	0	0.4029	0.4029	0.4029	0	
20	Panic Disorder	0	0.5118	0.6267	0.6267	0	
21	Neoplasms, Squamous Cell	0	0.4848	0.3567	0.4848	0	
22	Heart Defects, Congenital	0	0.5118	0.6267	0.6267	0	
23	Central Nervous System Diseases	0	0.5118	0.5976	0.5976	0	

	A	B	C	D	E	F	G	H	I	J	K	
1		Pregnancy, Ectopic	Vascular Diseases	Nerve Sheath Neoplasms	Vitiligo	Ependymoma	Periodontal	Huntington Disease	Hyperlipidemias	Diabetes Mellitu	Meningioma	Oste
2	Pregnancy, Ectopic	1	0.007889223	0.304660555	0.0211	0.371681757	0.01703149	0.309100548	0.261146773	0.191157124	0.21001451	0.32
3	Vascular Diseases	0.007889223	1	0.035471601	0.0278	0.02394243	0.07583026	0.019983269	0.033142578	0.037224688	0.02214091	0.02
4	Nerve Sheath Neoplasms	0.304660555	0.035471601	1	0.0928	0.465162265	0.03148574	0.498595157	0.480311692	0.435338938	0.34526794	0.46
5	Vitiligo	0.021134113	0.027837657	0.092777114	1	0.044164975	0.04131431	0.059425598	0.047372871	0.121365208	0.04352513	0.03
6	Ependymoma	0.371681757	0.02394243	0.465162265	0.0442	1	0.03060636	0.447497125	0.424243272	0.312726529	0.325338	0.49
7	Periodontal Diseases	0.017031487	0.075830256	0.031485744	0.0413	0.030606365	1	0.01757209	0.022620849	0.0266559	0.02609663	0.0
8	Huntington Disease	0.309100548	0.019983269	0.498595157	0.0594	0.447497125	0.01757209	1	0.536353157	0.397441468	0.31433546	0.51
9	Hyperlipidemias	0.261146773	0.033142578	0.480311692	0.0474	0.424243272	0.02262085	0.536353157	1	0.372879256	0.30155832	0.49
10	Endomyocardial Fibrosis	0.191157124	0.037224688	0.435338938	0.1214	0.312726529	0.0266559	0.397441468	0.372879256	1	0.29003745	0.28
11	Diabetes Mellitus, Type 2	0.210014508	0.022140914	0.345267944	0.0435	0.325338003	0.02609663	0.314335463	0.301556319	0.290037647	1	0.29
12	Meningioma	0.321766277	0.021310617	0.464411884	0.0373	0.493154183	0.0175265	0.512585794	0.490781351	0.287595357	0.29850772	
13	Osteoarthritis	0.088508129	0.00843693	0.222623943	0.1097	0.139705229	0.01855477	0.158551706	0.13820068	0.223231752	0.10820894	0.12
14	Tuberculosis, Pulmonary	0.096191863	0.045786855	0.279016828	0.1837	0.152316909	0.01182364	0.194469395	0.19273664	0.342891148	0.15216643	0.13
15	Giant Cell Tumors	0.202634288	0.031594129	0.46254741	0.1299	0.285788366	0.02488589	0.34481598	0.344456147	0.583000113	0.25862648	0.27
16	Glomerulonephritis	0.350242099	0.032054393	0.62629078	0.077	0.524201892	0.0253098	0.507504352	0.46672568	0.403133701	0.35599166	0.57
17	Gastric Neoplasms	0.051561488	0.015657785	0.163400036	0.1732	0.082742096	0.00531318	0.114112385	0.11767568	0.217052321	0.07952942	0.08
18	Osteoporosis	0.030511562	0.018955698	0.122700332	0.1896	0.057669403	0.00025666	0.088090241	0.073956651	0.152640671	0.09195906	0.05
19	Irritable Bowel Syndrome	0.001364748	0.00025666	0.004393716	0.0933	0.00025666	0.00025666	0.001771635	0.001326941	0.00249912	0.00025666	0.00
20	Panic Disorder	0.349429894	0.039317054	0.580723024	0.0926	0.453736675	0.01569095	0.498580009	0.466444635	0.470538481	0.32379088	0.43
21	Neoplasms, Squamous Cell	0.310776972	0.032375471	0.619978249	0.1189	0.412068176	0.01655012	0.480814255	0.447464285	0.511538387	0.35807413	0.37
22	Heart Defects, Congenital	0.252151035	0.036996578	0.507277227	0.0899	0.33648304	0.01698102	0.399491885	0.381874256	0.464053324	0.28941023	0.31
23	Central Nervous System Diseases	0.317196122	0.045355257	0.656827885	0.0941	0.436855148	0.02912896	0.494852309	0.472649911	0.465641936	0.33417444	0.43

Figure 3.5: Disease-Disease Functional Similarity

	A	B	C	D	E	F	G	H	I	J	K	L
1		Pregnancy, Ectopic	Vascular Diseases	Nerve Sheath Neoplasms	Vitiligo	Ependymoma	Periodontal	Huntington Disease	Hyperlipidemias	Diabetes Mellitu	Meningioma	Osteoarthri
2	Pregnancy, Ectopic	1	0.057926601	0.471699896	0.1294	0.345681422	0.03008821	0.405570764	0.386837062	0.504154367	0.28465414	0.32360087
3	Vascular Diseases	0.039020804	1	0.13961341	0.1716	0.07218395	0.00827899	0.104077161	0.101678327	0.236010881	0.06540521	0.06808567
4	Nerve Sheath Neoplasms	0.053992894	0.028493142	1	0.0644	0.121919807	0.01144058	0.100015357	0.115347033	0.158231492	0.17864003	0.08947228
5	Vitiligo	0.324238829	0.034154988	0.640729344	1	0.464390543	0.02406073	0.488548145	0.462015134	0.462401799	0.34222058	0.44823778
6	Ependymoma	0.256627517	0.024260321	0.530870108	0.1156	1	0.02614769	0.483857508	0.441416335	0.489584718	0.26901894	0.33666143
7	Periodontal Diseases	0.149967394	0.012509795	0.347357754	0.1086	0.192805564	1	0.328746706	0.319241119	0.372167753	0.15627858	0.22760905
8	Huntington Disease	0.11915531	0.02394056	0.309271402	0.2515	0.18495041	0.04448991	1	0.205226978	0.386775892	0.18281611	0.14684267
9	Hyperlipidemias	0.351215634	0.022183499	0.553353462	0.0641	0.514244844	0.02129535	0.488093221	1	0.365615256	0.32249141	0.51038323
10	Endomyocardial Fibrosis	0.329246209	0.023408498	0.448555127	0.0578	0.555894566	0.03619545	0.473989366	0.452179676	1	0.31075297	0.60190978
11	Diabetes Mellitus, Type 2	0.229731092	0.02376206	0.380978336	0.0562	0.367846684	0.02506473	0.385269875	0.385748478	0.355095328	1	0.31392074
12	Meningioma	0.020535773	0.00025666	0.031669586	0.0044	0.021703774	0.00025666	0.056865892	0.072209838	0.047097183	0.02358794	1
13	Osteoarthritis	0.06239704	0.134129724	0.190620275	0.2713	0.105165339	0.03595171	0.127988153	0.117859327	0.269985381	0.10064894	0.05810833
14	Tuberculosis, Pulmonary	0.00025666	0.00025666	0.007894028	0.0003	0.006728732	0.00025666	0.008957707	0.009060182	0.018092354	0.0079231	0.00663722
15	Giant Cell Tumors	0.060131329	0.060014295	0.165996845	0.139	0.093618641	0.00908361	0.112371858	0.107226131	0.243128643	0.07060131	0.08444266
16	Glomerulonephritis	0.098957583	0.148528532	0.344998028	0.1891	0.075281756	0.09695758	0.18133262	0.130528677	0.552626646	0.20957972	0.34688232
17	Gastric Neoplasms	0.342940495	0.020388265	0.503343969	0.0676	0.503416227	0.0509709	0.496501557	0.511430524	0.400820921	0.33856007	0.45403609
18	Osteoporosis	0.00814647	0.00025666	0.040115972	0.0042	0.021083985	0.00025666	0.035125367	0.031076061	0.072130523	0.01065671	0.01670579
19	Irritable Bowel Syndrome	0.029115352	0.02131814	0.054138134	0.0808	0.039775786	0.01651638	0.047803964	0.042506186	0.063349113	0.04395561	0.03271579
20	Panic Disorder	0.154002521	0.021042386	0.36253294	0.1525	0.212544995	0.0049226	0.333824141	0.32290262	0.434582734	0.18972739	0.21728743
21	Neoplasms, Squamous Cell	0.313585426	0.020364664	0.452816861	0.042	0.307401607	0.02033718	0.403572238	0.375285633	0.263977353	0.26705034	0.39919288
22	Heart Defects, Congenital	0.207762138	0.066133193	0.412048092	0.1207	0.269300084	0.0125706	0.313886119	0.288574354	0.435819823	0.24996269	0.26124486
23	Central Nervous System Diseases	0.311706318	0.017975653	0.388020254	0.0315	0.422682448	0.02569636	0.376136488	0.3580249	0.232434601	0.28190786	0.35789445

Figure 3.6: Disease-Disease Semantic Similarity

3.2.2 Integrating the data sets

We collect the five data sets of miRNA-miRNA sequence similarity, miRNA- miRNA functional similarity,Disease-Disease Functional Similarity, Disease-Disease Semantic similarity, Disease-Disease DAG Similarity.The main task was inte- grating these five data sets which was later fed as input to SOM module. For each disease the values in the other data sets are found and converted to an ar- ray. Similarly miRNA 's are calculated to a array. For each array we obtain five values, out of which three values are for diseases rest for miRNA. The first value
in the array consist miRNA-miRNA sequence similarity.In the data sets miRNA are represented row wise. Therefore we can't obtain miRNA-miRNA Sequence Similarity directly. For this we need to find out the miRNA with which that particular diseases has 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.

3.2.3 Input of the SOM modules

After integrating the data sets,each row of input is added even as input to the SOM. Each row consists of miRNA and it is represented as five similarities.

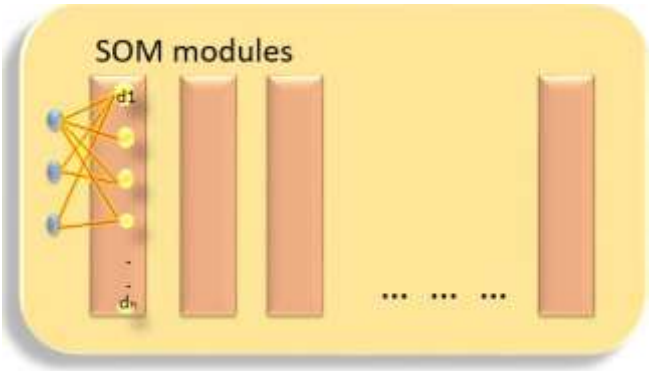

Figure 3.7: SOM MODULE

3.2.4 Working of som modules

SOM Maintains the topology of the data set . When we input a miRNA 's they initialize weight randomly .For each input, find the "winning" neuron called Best matching unit. If we input a miRNA , five similarity values are considered to rep- resented that miRNA. Then find the euclidean distance between the miRNA and all disease in the data sets . One disease is represented by mirna-mirna sequence similarity, mirna-mirna functional similarity,disease-disease functional similar- ity, disease-disease Semantic similarity, disease-disease dag similarity. But we cannot obtain mirna-mirna both similarities, so we need to find which miRNA is causing disease from the HMDD database. Similarly miRNA is also represented. fig 3.8 Shows the distance calculated between miRNA and disease.

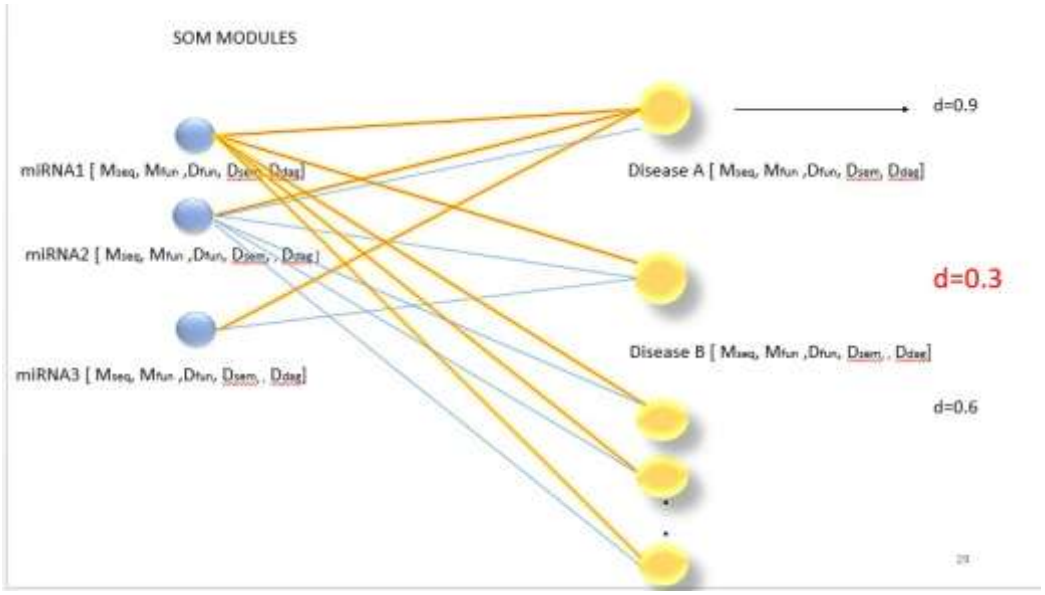

Figure 3.8: DISTANCE

3.2.5 Finding BMU

According to the euclidean distance, each som modules have a winning neuron called best matching unit. Then we integrate all the best matching unit and found the lowest distance. Because Lowest distance have high similarity. So, the high similarity of disease will be the output.

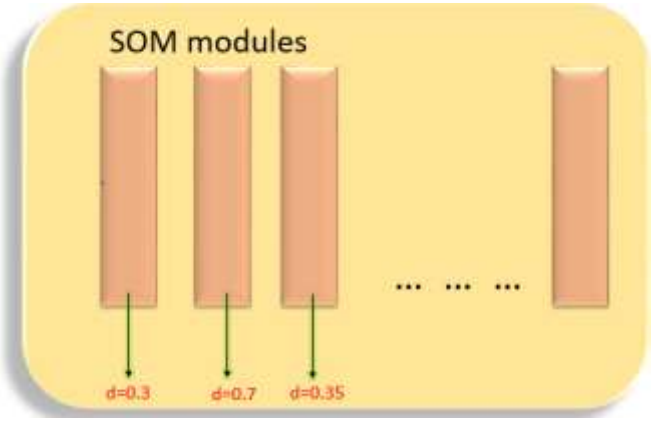

Figure 3.9: Lower Distance of som module

3.2.6 Ploting Networks from BMU

For plotting the Best match unit we use a cytoscape to visualizing the miRNA and Disease. In cytoscape we set which is the source nodes and target nodes. The network gives a great deal of idea about the key regulatory miRNA based on disease.

3.3 Conclusion

This section covered in detail the Inputs taken, methodologies used, and results are plotted using cytoscape. The results obtained at of this methodology is given in the next chapter, Result analysis and Discussion.

Chapter 4

Result Analysis and Discussion

4.1 Results Analysis

4.1.1 SOM results

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

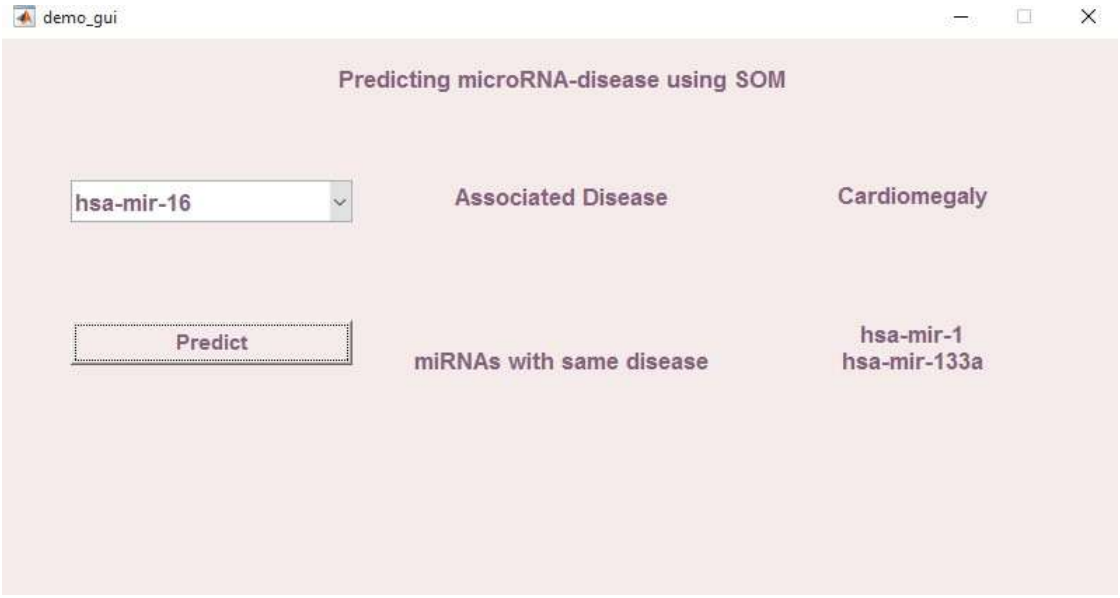

Figure 4.1: Predicting miRNA using SOM.

Fig shows the miRNA-Disease prediction using SOM . When we input a miRNA ,it takes the corresponding five values to represented them. After fed to the SOM, it calculates the best matching unit(disease) and that disease has an association with other miRNA is be calculated. So if we give a miRNA the corresponding disease is predicted. Fig 4.2 shows the miRNA-Disease associ-

result
137x271 double

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-4.3401	-3.2739	-3.0108	-2.7394	-3.3834	-3.0049	-2.2567	-2.4914	-3.2181	-3.1148	-1.9113	-2.8862	-3.2991
2	-4.1719	-3.9490	-3.2839	-2.0802	-3.4717	-1.7124	-2.8767	-3.0450	-3.6962	-3.4142	-2.6219	-3.6071	-5.4403
3	-3.1741	-2.0801	-3.5029	-1.7936	-2.7264	-3.2456	-3.0979	-2.8200	-3.6784	-1.9994	-3.5225	-3.8112	-5.5102
4	-3.1912	-2.8399	-2.2627	-3.0479	-3.2179	-2.8215	-2.2328	-1.8567	-3.4128	-2.1238	-2.8635	-4.8158	-2.9996
5	-4.6437	-4.3516	-3.1454	-3.1308	-3.5681	-2.6709	-2.9428	-2.1469	-4.0929	-3.6341	-2.8900	-4.3289	-3.8830
6	-4.1643	-4.7888	-3.1719	-3.9002	-4.0055	-3.8371	-3.0346	-3.7313	-4.1578	-3.3498	-2.1482	-3.3165	-4.2473
7	-4.6833	-4.2520	-3.7617	-3.5182	-3.5744	-4.0849	-2.5562	-2.2500	-3.7279	-3.1218	-2.0769	-3.7493	-4.9563
8	-3.3531	-2.9348	-2.6651	-2.3338	-3.4964	-3.1933	-2.3392	-2.2420	-3.0968	-2.4668	-2.8886	-2.9026	-3.3309
9	-3.0888	-3.5530	-3.5804	-3.1460	-3.8015	-2.7277	-3.3570	-3.8092	-4.1964	-3.3624	-2.8571	-2.4166	-5.0015
10	-2.5658	0.3156	-2.6029	-3.0909	-1.4777	-2.5510	-2.7903	-1.7089	-3.5332	-2.4835	-1.3216	-2.3224	-3.7243
11	-3.6446	-4.1359	-3.0115	-2.5400	-3.7723	-3.6657	-2.5870	-2.4581	-3.3744	-3.0098	-2.8084	-3.1564	-3.9342
12	-4.4936	-4.6584	-2.7924	-2.6219	-3.7479	-3.6364	-3.4701	-3.0263	-3.6264	-3.5849	-4.9672	-2.5328	-3.2384
13	-3.5534	-3.8973	-3.4994	-2.8965	-4.0004	-2.3765	-3.0150	-4.0735	-4.0597	-3.3005	-3.4931	-2.9548	-4.6132
14	-2.7561	-3.3863	-3.5413	-2.9588	-3.5990	-2.9127	-3.4599	-3.8258	-4.1782	-3.0569	-2.7914	-2.3810	-4.4632
15	-3.5907	-2.9391	-2.8542	-3.0965	1.1459	2.6728	-3.1823	-2.7208	-3.9388	-2.1249	0.9561	-2.3160	-4.6780
16	-4.3050	-4.2656	-4.0319	-3.7700	-4.0528	-4.4440	-3.0699	-3.3453	-4.3443	-3.4398	-2.7048	-4.0121	-5.1279
17	-3.7053	-3.5194	-3.5296	-3.2762	-4.1837	-3.5809	-3.3195	-3.7549	-4.2323	-3.3862	-3.2047	-3.2816	-4.9338
18	-3.9288	-3.9521	-3.1090	-3.2450	-4.0028	-4.3846	-2.5771	-2.6340	-3.6377	-2.7693	-3.2065	-2.6818	-3.6052

Figure 4.2: miRNA-Disease Association Data.

ation data using the KBMF method. The negative value shows that the worst similarity and positive value shows that high similarity. Each row represent by a miRNA and column represent by diseases. One cell value represent the association of miRNA and disease.

4.1.2 Network Results

The networks created based on the results of prediction of miRNA and disease using SOM are shown in fig 4.3. Using cytoscape we can plot the som results as network.Cytoscape Networks are easy to understand which miRNAs are related to diseases. Each edge showing the relationship between disease and miRNA (fig 4.3).

4.2 Performance

The area under the ROC curve(AUC) is utilized to measure the performance. We compare the SOM with work done by. [9]. Fig 4.4 shows that area of curve

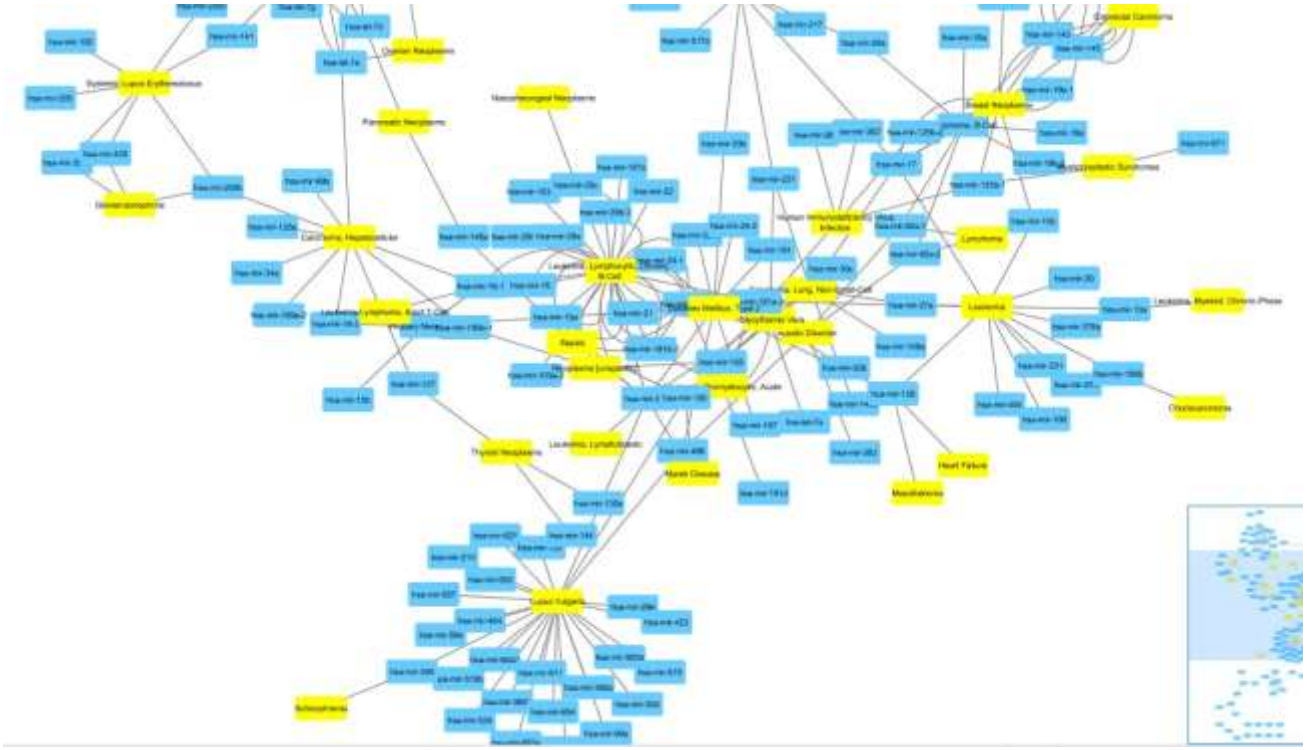

Figure 4.3: miRNA-Disease Association using SOM Module.

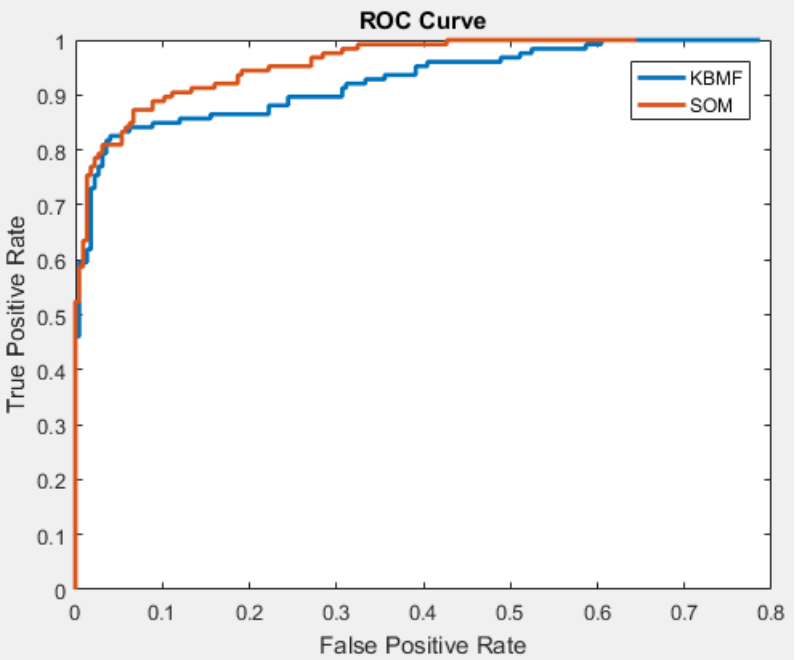

Figure 4.4: ROC CURVE

obtained from SOM is greater that KBMF method which implies better accuracy of SOM.

4.3 Discussions

The ROC curve of our SOM shows that it matches the state-of-art methods.The methods used in the previous works uses the assignment of weight based on miRNA families and miRNA Clusters on the basics of seed regions of miRNA sequence of same families are identical. Some of the state-of-art methods mea- sures only similarities of miRNA and diseases only based on disease data or miRNA data.

4.4 Conclusion

From the result analysis we can come to the following main conclusions

- The miRNA plays important role in diseases (fig:4.3).
- The uses of SOM improved the accuracy compared to the one state of art methods. Hence like mentioned in Zou et.al [13], Juan wang et al. [7], Cheng et al. [10] it seems that identifying miRNA SIMILARITIES do help in recognizing that has higher probability of causing a disease.
- Apart from the above main results showcases the miRNA regulation which plays an important role in many diseases.

Chapter 5 Conclusion

This chapter presents the summary and conclusions of the proposals and future work

5.1 Conclusion

So in this thesis accurate use of similarities of miRNA and disease are used, from which the global data sets HMDD,SIDD and HumanNET. The integrating the miRNA 's and diseases similarities helped in to find best matching diseases. In the previous state of art methods uses only either miRNA or Disease simi- larities. But in our method we uses the five similarities to predict the miRNA and diseases. So it is very helpful to discover uncover the dangerous disease pathogens behind the miRNA.

5.2 Future work

The future work includes the use add more similarities based on miRNA or diseases. There are lot of exiting methods which uses only the information of disease or miRNA which result in poor performance. So if we modify it get better result. If we add chromosomal coordinates of human miRNA from miRBase data sets can provide more clearer and accurate results.

Bibliography

[1] N. H. G. R. Institute, "Chromosomes," 2015. [Online; accessed 25-October- 2018].

[2] K. Academy, "Dna the central dogma." [Online; accessed 25-October- 2018].

[3] Wikipedia, "Dna in eukaryote cell," 2012. [Online; accessed 25-march-2019].

[4] C. Greene, A. Krishnan, A. Wong, E. Ricciotti, R. Zelaya, D. Himmelstein,
R. Zhang, B. Hartmann, E. Zaslavsky, S. Sealfon, D. Chasman, G. FitzGer- ald, K. Dolinski, T. Grosser, and O. Troyanskaya, "Understanding multi- cellular function and disease with human tissue-specific networks," *Nature Genetics*, vol. 47, pp. 569—-576.

[5] K. Tüfekci, M. Oner, R. Meuwissen, and S. Genç, "The role of micrornas in human diseases," *NCBI*, vol. 28, no. 3, 2014.

[6] M. L. J. L. F.-X. W. Wei Lan, Jianxin Wang and Y. Pan, "Predicting microrna-disease associations based on improved microrna and disease sim- ilarities," *NCBI*, vol. 15, no. 5, 2016.

[7] X. Ping, H. Ke, G. Maozu, G. Yahong, L. Jinbao, D. Jian, L. Yong,
D. Qiguo, L. Jin, T. Zhixia, and H. Yufei, "Prediction of micrornas associ- ated with human diseases

based on weighted k most similar neighbors,"

[8] C. Francis, "Central dogma of molecular biology," *Nature*, vol. 227, pp. 561– 563, 1970.

[9] M. G. Y. G. J. L. J. D. Y. L. Q. D. J. L. Z. T. Ping Xuan, Ke Han,
"Prioritizing human cancermicrornas based on genes' functional consistency between microrna andcancer," *PLOS ONE*, vol. 39, p. 1–10.

[10] Y. G. Chen X, "Semi-supervised learning for potential human microrna- disease associations inference," *Nature Research*, vol. 4, pp. 5501–5510.

[11] L. Chen, G. Wang, J. Li, T. Zhang, P. Xu, and Y. Wang, "'inferring the human microrna functional similarity and functional network based on microrna- associated diseases '," *BIOINFORMTICS*, vol. 26, no. 4767, 2013.

[12] W. G. health estimates, "The top 10 causes of death," 2018. [Online; accessed 27-May-2019].

[13] Yu, H. ZHANG, C. LIANG, and X. DONG, "Katzmda: Prediction of mirna-disease associations based on katz model," *IEEE ACESS*, vol. 16, no. 2, pp. 181–189, 2017.

[14] A. R, "'measuring disease similarity and predicting disease-related ncrnas by a novel method'," *BMC Genomics*, vol. 19, no. 668, pp. 15–25, 2016.

[15] P. Team, *python: Integrated Development Environment for R*. python, Inc., Defense Advanced Research Projects Agency (DARPA)Boston, MA, 2014.

[16] The Mathworks, Inc., Natick, Massachusetts, *MATLAB(R2018b)*, 2018.

[17] Y. K. and F. K.A, "multi-harmony: multi-group sequence harmony multi- relief," 2010. [Online; accessed 03-AUGUST-2018].

[18] S. Paul, M. Andrew, O. Owen, B. Nitin, S., W. Jonathan, T., R. Daniel,
A. Nada, S. Benno, and I. and, Trey, "Cytoscape: A software environ- ment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, pp. 2498–2504.

[19] H. Z, Shi, G. Y, C. C1, Z. S, L. J, Z. Y, and C. Q, "Hmdd v3.0: a database for experimentally supported human microrna-disease association," 2017. [Online; accessed 01-AUGUST-2018].

[20] L. Cheng, G. Wang, J. Li, T. Zhang, P. Xu, and Y. Wang, "Sidd: A semantically integrated database towards a global view of human disease," *PLOS ONE*, vol. 27, pp. 135–137, 2013.

[21] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones., "mirbase: from mi- crorna sequences to function," *PLOS ONE*, vol. 47, 2019.

[22] S. M. Bello, M. Shimoyama, E. Mitraka, S. J. F. Laulederkind, C. L. Smith,
J. T. Eppig, and L. M. Schriml, "Disease ontology: improving and unify- ing disease annotations across species," *disease models and mechanisums*, vol. 101, no. 16, pp. 2173–2178, 2018.

[23] R. F B, ""medical subject headings"," 2007. [Online; accessed 04-october- 2018].

[24] Sam and Griffiths-Jones, "mirbase: microrna sequences and annotation," 2009. [Online; accessed 01-october-2018].

[25] C. Liang, L. Jie, J. Peng, P. Jiajie, and Y. Wang, "Semfunsim: A new method for measuring disease similarity by integrating semantic and gene functional association," *PLOS ONE*, vol. 42, pp. 118–126, 2014.

[26] Insuk, Lee, B. U. Martin, Peggy, Wang, J. E. Shin, and M. M. Edward, "Prioritizing candidate disease genes by network-based boosting of genome- wide association data," *NCBI*, vol. 7, 2011.