MÜHENDİSLİK FAKÜLTESİ

Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2022) 14(3), s234-s243.

10.29137/umagd.1217754

Cilt/Volume:14 Sayı/Issue:3 Aralık/December 2022 Özel Sayı / Special Issue

Seçilmiş Konferans Makalesi / Selected Conference Paper

A Novel Method for miRNA-Disease Association Prediction based on Space Projection and Label Propagation (SPLPMDA)

Ahmet Toprak[1]

[1]*Department of Electricity and Energy, Selçuk University, Konya*

***Başvuru/Received**: 12/12/2022* ***Kabul / Accepted**: 23/12/2022* ***Çevrimiçi Basım / Published Online:** 31/12/2022*

***Son Versiyon/Final Version**: 31/12/2022*

Abstract

miRNAs, a subclass of non-coding small RNAs, are about 18-22 nucleotides long. It has been revealed that miRNAs are responsible many diseases such as cancer. Therefore, great efforts have been made recently by researchers to explore possible relationships between miRNAs and diseases. Experimental studies to identify new disease-associated miRNAs are very expensive and at the same time a long process. Therefore, to determine the relationships between miRNA and disease many computational methods have been developed. In this paper, a new method for the identification of miRNA-disease associations based on space projection and label propagation (SPLPMDA) is proposed. The forecast the precision of SPLPMDA was demonstrated using 5-fold cross-validation and LOOCV techniques. Values of 0.9333 in 5-fold cross validation and 0.9441 in LOOCV were obtained. Moreover, case studies on breast neoplasms and lymphoma were performed to further confirm the predictive reliability of SPLPMDA.

Key Words

"miRNA, Disease, miRNA-Disease Association, Space Projection, Label Propagation"

***Corresponding Author: ahmettoprak42@gmail.com**

1. Introduction

The first microRNA (miRNA) was defined in Caenorhabditis elegans (C. elegans) by Lee et al. at Victor Ambros' laboratory in 1993. They named this first miRNA discovered as lin-4. Reinhart et al. discovered a miRNA with a length of 22 nucleotides in C. elegans in 2000 and called let-7 (Lee, Feinbaum, & Ambros, 1993; Saydam, Değirmenci, & Güneş, 2011). Studies have revealed that miRNAs are a subclass of non-protein coding RNAs and are approximately 18-22 nucleotides long (Chandra et al., 2017). Although about 2000 human miRNAs have been identified, information about their biological functions is limited so far. Conducted research has revealed the importance miRNAs in development, proliferation, apoptosis, differentiation, signal transduction, viral infection, aging, and metabolism (Bartel, 2009; X. Chen, Zhou, & Zhao, 2018; Lan et al., 2018; Tang, Zhou, Zheng, Zhang, & Sha, 2019). The existence of important links of miRNAs with many complex human diseases has been proven by scientists with improving of the molecular biology and biotechnology (X. Chen et al., 2016; Kim, 2015). miRNAs can be oncogene or tumor suppressor in many cancer types such as breast cancer, lymphoma, lung cancer, prostate cancer, colon cancer (X. Chen, Xie, Zhao, & You, 2019; Gao, Jia, Shi, Zhou, & Cui, 2019). As an example, miRNA-21 expression level is constantly up-regulated in tissue-specific cancer types such as breast cancer, lymphoma, lung cancer, and colon cancer. In experiments, it was observed that in lung cancer and lymphoma, the let-7 miRNA family is downregulated, while miRNA-17, miRNA-18a, miRNA-19a, miRNA-19b-1, miRNA-20a and miRNA-92a-1 are overexpressed (Osada & Takahashi, 2011; Selcuklu, Donoghue, & Spillane, 2009; Tan et al., 2018). Also it is proven that both miRNA-143 and miRNA-145 are consistently downregulated in breast cancer patient by conducted research studies (Espinosa & Slack, 2006). Determination of disease-associated miRNAs is of great importance in the diagnosis, treatment, and prevention of diseases, as well as for personalized drug therapy (X. Chen et al., 2019; Yan, Zheng, Jia, Hou, & Xiao, 2019). By determining the relationships between miRNAs and diseases, many diseases such as cancer can be diagnosed at an early stage. However, experimental studies to identify new disease-associated miRNAs are very expensive and at the same time a long process. For this reason, many computational methods have been developed to determine these possible relationships between miRNAs and diseases (Pech, Lee, Hao, Po, & Zhou, 2019). Machine learning-based approaches and similarity-based approaches are generally preferred to find potential relationships between miRNAs and diseases. The assumption that similar miRNAs (diseases) likely affect the same diseases (miRNAs) is commonly used in similarity-based approaches.

Researchers have developed many computational methods to predict disease-associated miRNAs, lncRNAs, circRNAs, microbes, and environmental factor. For example, Toprak et al. (Toprak & Eryilmaz Dogan, 2021; Toprak & Eryilmaz, 2021) used two different methods for miRNA-disease associations prediction: KBMF and "weighted k-nearest known neighbors and network consistency projection". The ILDMSF method for prediction of lncRNA-disease associations was developed by Chen et al. (Q. Chen et al., 2021). Vural et al. (Vural & Kaya, 2018) used the KATZ method for prediction lncRNA-environmental factor associations. Qu et al. (Qu, Zhao, & Yin, 2019) suggested a computational technique for estimating microbe-disease associations.

Several comprehensive databases have been created to store experimentally validated results of miRNA-disease associations with advances in technology and bioinformatics: HMDD (Y. Li et al., 2014) is a database of miRNA target interactions as well as containing comprehensive information about many human diseases such as genetics and epigenetics. miRBase (Kozomara & Griffiths-Jones, 2013) database contains miRNA sequences and is also an open-source database. dbDEMC (Z. Yang et al., 2010) provides information in different cancer types about the miRNAs' expression levels. miR2Disease (Jiang et al., 2009) provides human miRNA-disease association information. deepBase (J.-H. Yang, Shao, Zhou, Chen, & Qu, 2009) comprehensively describes the role of miRNAs in biological processes of organisms. miRGen (Alexiou et al., 2009) is a database containing human miRNAs' and mouse miRNAs' genomic information.

In this paper, a new method, identification of miRNA-disease relationships based on space projection and label propagation (SPLPMDA), which integrates space projection and label propagation, has been proposed to find possible links between miRNAs and diseases. From HMDD web page, experimentally validated miRNA and disease association data, functional similarities of miRNAs, and semantic similarities of diseases were used. In addition, Gaussian interaction profile kernel similarities of the of miRNAs and diseases were calculated. Then, space projection method was implemented to miRNA space and disease space. Lastly, miRNA-disease association prediction results were obtained by applying the label propagation method.

2. Materials and Methods

2.1. Known associations between miRNA and disease

The experimentally validated miRNA-disease association data, include 495 miRNAs and 383 diseases, we used in this study were obtained from the HMDD database. A matrix of 495x383 dimensions, called the adjacency matrix, consisting of 495 rows and 383 columns, is created from the obtained data. When constructing the adjacency matrix, the experimentally proven relationship of miRNA and diseases is considered (i.e. if there is a proven relationship between miRNA $m(i)$ and disease $d(j)$, adjacency matrix $A(m(i), d(j))$ is set to 1, if there is no proven relationship between miRNA $m(i)$ and disease $d(j)$, adjacency matrix $A(m(i), d(j))$ is set to 0.

2.2. miRNAs Functional Similarity *(FS)* and Diseases Semantic Similarity *(SS)*

In 2010, a method was proposed by Wang and colleagues to calculate miRNA and disease similarity scores (D. Wang, Wang, Lu, Song, & Cui, 2010). Calculated miRNAs functional similarity scores can be obtained from the http://www.cuilab.cn/files/images/cuilab/misim.zip web page. The FS matrix with dimensions 495×495 is created from the downloaded data.

The SS matrix was calculated using the Medical Subject Headings (MeSH) definitions from the National Library of Medicine (http://www.nlm.nih.gov/) web page. Afterwards, the tree structure of each disease is defined by creating a Directed Acyclic Graph (DAG) structure of each disease. "Breast Neoplasms" and "Lymphoma" DAG structures are shown in Figure 1. For example, $DAG(A) = (A, T_A, E_A)$ structure of disease A, where T_A represents both node A itself and all sub-nodes of node A, and E_A represents the corresponding links. Here, equation 1 calculates the coefficient of disease t in DAG(A) to disease A, and equation 2 computes the disease A's semantic value (DV).

$$\begin{cases} D_A(A) = 1, \\ D_A(t) = max\{0.5 * D_A(t') | t' \in children\ of\ t\},\ if\ t \neq A, \end{cases} \quad (1)$$

$$DV(A) = \sum_{t \in T_A} D_A(t). \quad (2)$$

In order to calculate the disease A's and disease B's semantic similarity value with equation 3, first the DAG structure of each disease is created, and its semantic values are calculated. This process is performed for all diseases and a semantic similarity matrix ($SS_{383\times383}$) is created.

$$SS(A,B) = \frac{\sum_{t \in T_A \cap T_B} (D_A(t) + D_B(t))}{DV(A) + DV(B)} \quad (3)$$

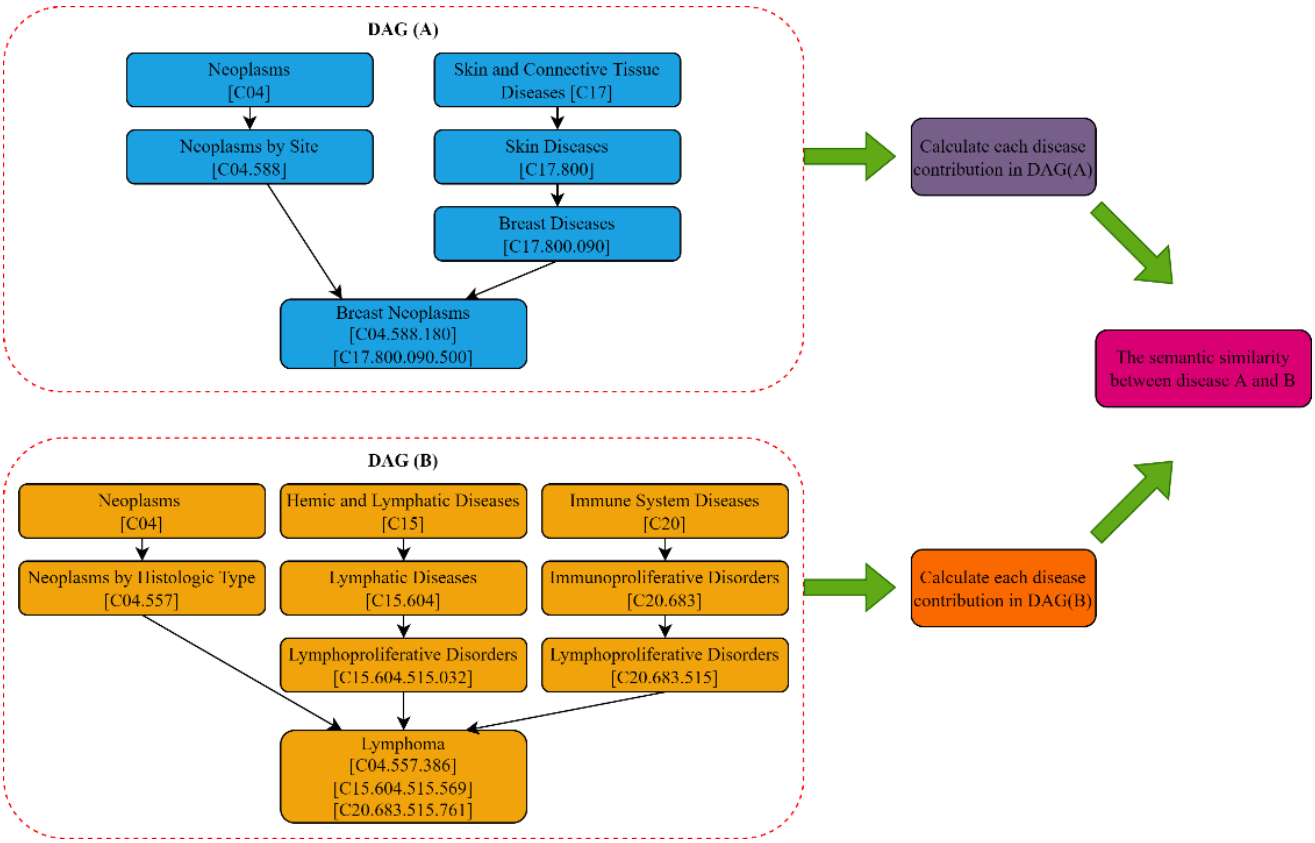

Figure 1. Breast neoplasms' and lymphoma's DAG structure

2.3. Gaussian interaction profile (GIP) kernel similarity

In 2011, Twan van Laarhoven et al. proposed the GIP kernel and used to find out drug-target relationships. "This method is based on the assumption that drugs (targets) that exhibit a similar interaction with targets (drugs) of a drug-target interaction network are likely to exhibit similar interactions" (van Laarhoven, Nabuurs, & Marchiori, 2011). The GIP method is widely used to find out disease-related miRNA, lncRNA, circRNA, and microbes.

Here, ith sequence of known human miRNA-disease matrix *A* is binary vector, represented by $IP(mi)$. The GIP kernel similarity matrix for miRNAs represented by GM is calculated by the following equation:

$$GM(m_i, m_j) = \exp\left(-\gamma_m \left\| IP(m_i) - IP(m_j) \right\|^2\right) \quad (4)$$

The GIP kernel similarity matrix for diseases represented by GD is calculated by the same way:

$$GM(d_i, d_j) = \exp\left(-\gamma_d \left\| IP(d_i) - IP(d_j) \right\|^2\right) \quad (5)$$

The γ_m and γ_d parameters seen in equations 4 and 5 that control the kernel bandwidth can be calculated as follows:

$$\gamma_m = \frac{\delta_m}{\frac{1}{n_m}\sum_{i=1}^{n_m} \|IP(m_i)\|^2} \quad (6)$$

$$\gamma_d = \frac{\delta_d}{\frac{1}{n_d}\sum_{i=1}^{n_d} \|IP(d_i)\|^2} \quad (7)$$

The new bandwidth parameters are represented by δ_m and δ_d, and all miRNA numbers and all disease numbers are represented by n_m and n_d. Twan van Laarhoven et al set the m and d parameters to 1 for simplicity.

2.4. Integration of similarities

From the known human miRNA-disease association matrix, the GIP kernel similarities of miRNAs and diseases are calculated. Then, miRNA functional similarity matrix and miRNA GIP kernel similarity matrix are integrated using equation 8. The resulting new miRNA similarity matrix is denoted by SM.

$$SM(m_i, m_j) = \begin{cases} GM(m_i, m_j), & if\ FS(m_i, m_j) = 0, \\ FS(m_i, m_j), & otherwise. \end{cases} \quad (8)$$

In the same manner, using equation 9, semantic similarities of diseases and GIP kernel similarities of diseases were integrated. The resulting new disease similarity matrix is represented by SD.

$$SD(d_i, d_j) = \begin{cases} GD(d_i, d_j), & if\ SS(d_i, d_j) = 0, \\ FS(d_i, d_j), & otherwise. \end{cases} \quad (9)$$

2.5. Network Space Projection and Label Propagation

We suggested new computational model that includes network space projection and label propagation to forecast possible associations between miRNAs and diseases. The method we have proposed consists of three parts. Firstly, integration of similarities. Secondly, network space projection for miRNA and disease. Lastly, for obtaining the prediction results, applying the label propagation on miRNA and disease space projection network. In this study, in matrix *A*, experimentally unconfirmed miRNA-disease associations were recorded as 0, but this does not mean that these experimentally unconfirmed relationships are actually unrelated. Thus, all relations that are 0 in matrix *A* are made 10^{-30}.

Network consistency projection consists of two parts, space consistency projection of miRNA and disease. Network consistency projection uses heterogeneous networks such as network of known associations between miRNA and disease, integrated similarity networks of miRNAs and diseases. The framework of SPLPMDA method is shown in Figure 2.

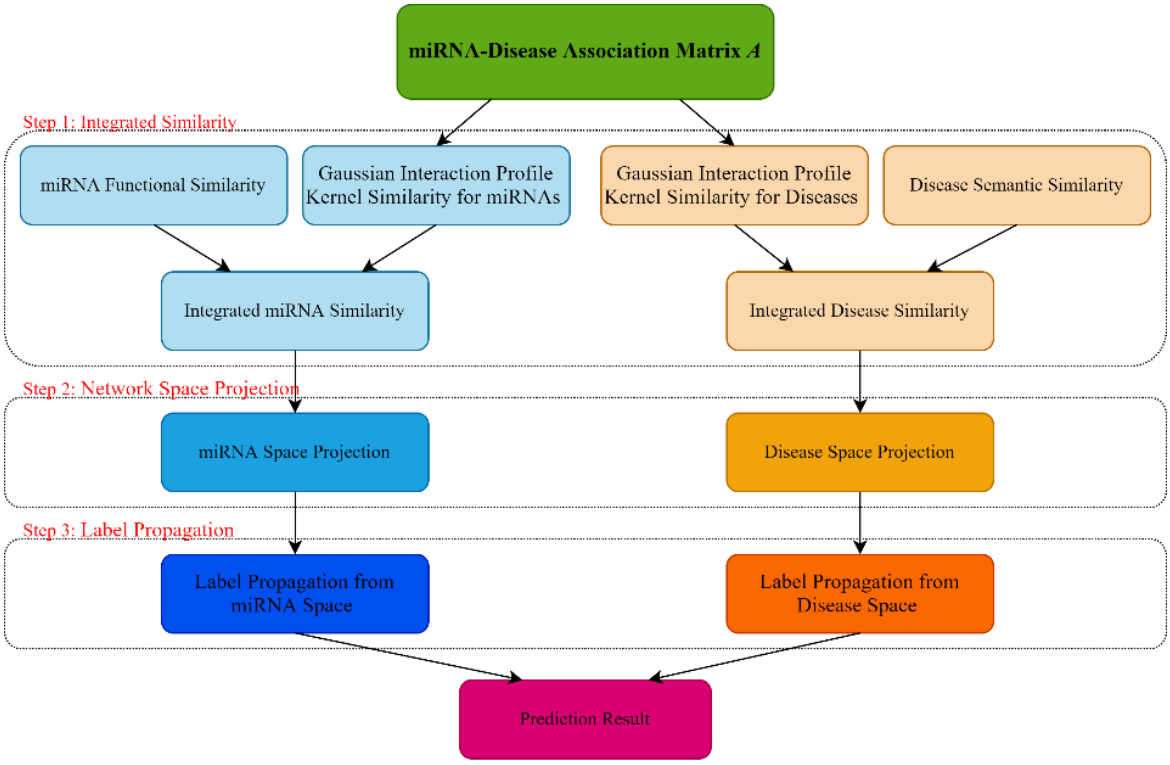

Figure 2. Flowchart

Firstly, the miRNA space projection (MSP) is determined by the following formula:

$$MSP(i,j) = \frac{SM(i,:)\times A(:,j)}{|A(:,j)|} \quad (10)$$

Then, disease space projection (DSP) is computed using equation 11:

$$DSP(i,j) = \frac{A(i,:)\times SD(:,j)}{|A(i,:)|} \quad (11)$$

Here, the label propagation technique was applied to compute the possibility of disease-associated miRNAs. "Label propagation is a semi-supervised learning method that iteratively propagates labelled information to unlabeled nodes throughout the network" (Yin, Liu, Gao, Kong, & Zheng, 2022; Yu et al., 2019). In the label propagation process, miRNAs (diseases) associated with a disease (miRNA) are regarded labeled samples, while other miRNAs (diseases) are considered unlabeled samples. The iteration formula on the miRNA space projection network can be represented by equation 12. With the same manner, the iteration equation on the disease space projection network can be written equation 13.

$$F_M(t+1) = \alpha \times SM \times F_M(t) + (1-\alpha) \times \left(\frac{MSP+A}{2}\right) \quad (12)$$

$$F_D(t+1) = \alpha \times SD \times F_D(t) + (1-\alpha) \times \left(\frac{DSP+A}{2}\right)^T \quad (13)$$

where, parameter α is between 0 and 1, and controls the rate. SM is the integrated miRNAs network and SD is the integrated diseases network. F_M and F_D are the forecast outcomes of miRNA domain and disease domain, respectively. The prediction results from both domains are combined with equation 14 and represent with F.

$$F = \beta \times F_M + (1-\beta) \times F_D^T \quad (14)$$

where β was set to 0.5.

3. Results

3.1. Performance prediction

In this chapter, the prediction performance of our model that we developed tested by five-fold (5-fold) cross validation and leave one-out cross validation (LOOCV) techniques. The known relationships between miRNA and disease are divided into five subgroups in the 5-fold cross-validation technique. While 4 randomly selected groups are used as training data, the remaining group is used as test data. In the LOOCV technique, each of the known relationships between miRNA and disease was used as test data and the rest as training data. This procedure was repeated for all known 5430 miRNA-disease associations. Also, using the true positive rate (TPR) and the false positive rate (FPR), the Receiver Operating Characteristic (ROC) curve was plotted to show the predictive accuracy of the model. Then, performance evaluation is performed with the value obtained by calculating the area under the ROC curve (AUC). The AUC value ranges from 0 to 1, and the closer the calculated result is to 1, the better the performance.

In SPLPMDA, the AUC value calculated with the 5-fold cross-validation technique was 0.9333, and the AUC value calculated with the LOOCV technique was 0.9441. ROC curves of the 5-fold cross validation and LOOCV techniques are demonstrated with Figures 3 and 4, respectively. Also, prediction performance is further proven by comparing the prediction result of SPLPMDA with 6 other methods: NDAMDA (X. Chen, Wang, & Huang, 2018), MCMDA(J.-Q. Li, Rong, Chen, Yan, & You, 2017), NSEMDA (C. C. Wang, Chen, Yin, & Qu, 2019), BNPMDA (X. Chen, D. Xie, et al., 2018), MDHGI (X. Chen, Yin, Qu, & Huang, 2018), and WBSMDA (X. Chen et al., 2016).

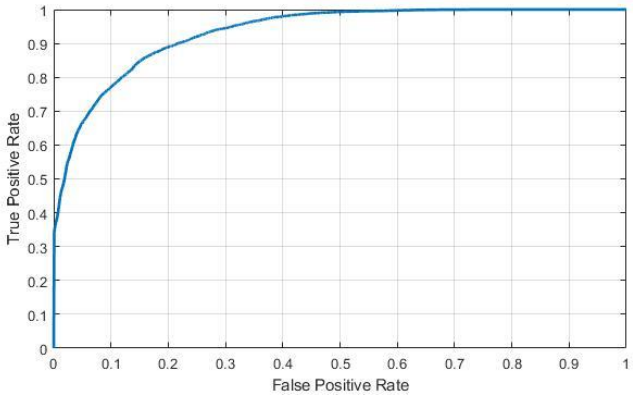

Figure 3. 5-fold CV

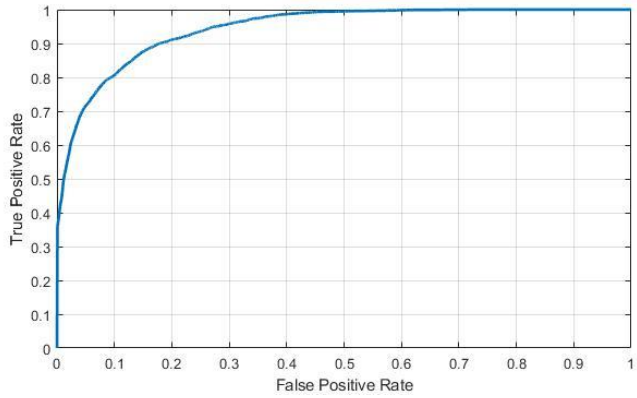

Figure 4. LOOCV

In the 5-fold cross validation technique, the AUC values obtained by NDAMDA, MCMDA, NSEMDA, BNPMDA, MDHGI, and WBSMDA were 0.8935, 0.8767, 0.8878, 0.8980, 0.8794, and 0.8185. Figure 5 shows the comparative AUC values. Moreover, in the LOOCV technique, the AUC values obtained by NDAMDA, MCMDA, NSEMDA, BNPMDA, MDHGI, and WBSMDA were 0.8920, 0.8749, 0.8899, 0.9028, 0.8945, and 0.8031. Figure 6 shows the comparative AUC values.

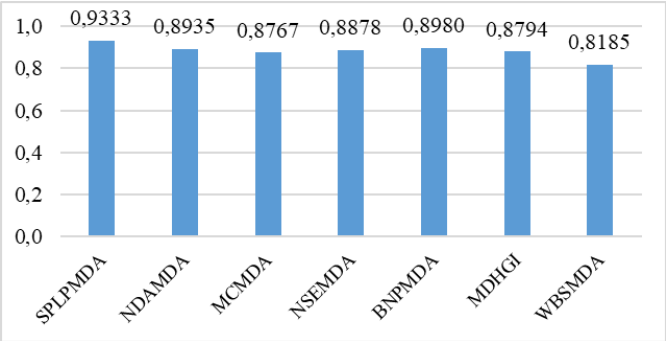

Figure 5. AUC values of SPLPMDA and others six methods in 5-fold CV

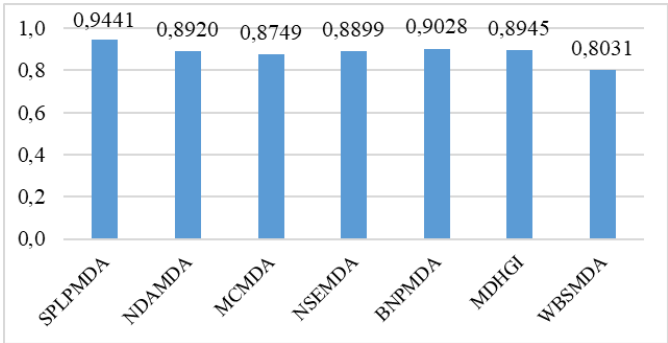

Figure 6. AUC values of SPLPMDA and others six methods in LOOCV

When the AUC values obtained in both 5-fold cross validation and LOOCV techniques are examined, it is seen that SPLPMDA gives better results compared to the other six methods.

3.2. Case studies

Here, we performed both a case study on breast cancer and a case study on lymphoma to validate the performance of SPLPMDA and predict miRNA-disease association. Two experimental databases namely HMDD v2.0 (Y. Li et al., 2014) and dbDEMC (Z. Yang et al., 2010) were selected to validate the candidate miRNAs predicted in the case studies. After applying SPLPMDA, candidate miRNAs for breast neoplasms and lymphoma were listed by scores. The top 30 candidate miRNAs listed by score for breast cancer and lymphoma were validated by the two different databases mentioned above.

Breast Neoplasms, which occurs in men as well as women, is very widespread malignant tumor in women and causes about 40000 mortalities per year in the United States (DeSantis et al., 2016). Studies have demonstrated that numerous miRNAs have an important mission in the diagnosis and development of breast cancer. For instance, in experimental studies with breast cancer patients, miR-21 and miR-155 were up-regulated, while miR-10b, miR-125b, and miR-145 were down-regulated. That's why, A case study was conducted on breast neoplasms as many miRNAs are thought to act as tumor suppressor genes or oncogenes (Feber et al., 2008; Iorio et al., 2005). In this case study, information was removed from all known miRNAs associated with breast neoplasm. Then, SPLPMDA was applied to predict potential miRNAs related to breast cancer and the top 30 candidate miRNAs proposed by SPLPMDA were listed. As can be seen from Table 1, 29 of the top 30 predicted candidate miRNAs were associated with breast neoplasms.

Lymphoma is the uncontrolled growth of lymphocytes, the body's defense cells, by becoming cancerous (Alizadeh et al., 2000), originates from the lymphatic hematopoietic system, which is the cause of many types of cancer (DeSantis, Ma, Goding Sauer, Newman, & Jemal, 2017). Recent research has shown associations between many miRNAs and lymphoma. For instance, in malignant lymphoma, tumors are suppressed by miR-150 (Watanabe et al., 2011). In addition to targeting proto-oncogenes in cutaneous T-cell lymphoma and mycosis fungoides, also miR-223 regulates cell growth (McGirt et al., 2014). Moreover, miR-200 targeting cyclin E2 is extensively suppressed in conjunctival MALT lymphoma (Cai et al., 2012). Before applying the SPLPMDA, we removed the information of all known miRNAs associated with lymphoma. When we apply SPLPMDA for potential miRNA-lymphoma association estimation, the top 30 candidate miRNAs in order of score are shown in table 2.

4. Discussion

It has been explained in many studies that non-protein-coding miRNAs cause many human diseases and miRNAs have been shown to affect many biological processes (Bartel, 2009; Xing Chen et al., 2018; Lan et al., 2018; Tang et al., 2019). For this reason, it is very important to know the relationships between miRNAs and diseases. However, finding new relationships with conventional experimental methods is time consuming and costly. In this research, we used space projection and label propagation to discover possible relationships between miRNAs and diseases. To evaluate the SPLPMDA's forecast performance 5-fold cross validation and LOOCV techniques were used, and 0.9333 AUC value and 0.9441 AUC value were gotten, respectively. When we compared SPLPMDA with NDAMDA, MCMDA, NSEMDA, BNPMDA, MDHGI, and WBSMDA, we obtained better prediction performance in both 5-fold cross-validation and LOOCV. In addition, the predictive performance of SPLPMDA was also evaluated by two case studies on breast cancer and lymphoma diseases. When the first 30 candidate miRNAs obtained as a result of the case study on breast cancer and ranked according to their scores are examined in detail, it is seen that 29 miRNAs are associated with breast cancer. As a result of the case study on lymphoma, candidate miRNAs were also ranked according to their scores. When the first 30 miRNAs were investigated in detail, it was seen that 29 miRNAs were associated with lymphoma. Candidate miRNAs from case studies on breast cancer and lymphoma were validated with HMDD and dbDEMC databases. Consequently, SPLPMDA is a powerful technique for identifying possible miRNA-disease associations without costly laboratory testing.

Table 1. Prediction of the top 30 predicted miRNAs associated with breast neoplasms

hsa-mir-1245a	HMDD
hsa-mir-1245b	HMDD
hsa-mir-1323	HMDD
hsa-mir-1469	HMDD
hsa-mir-181	*unconfirmed*
hsa-mir-2355	HMDD
hsa-mir-3130	HMDD
hsa-mir-3186	HMDD
hsa-mir-4257	HMDD
hsa-mir-4306	HMDD
hsa-mir-718	HMDD; dbDEMC
hsa-mir-320e	HMDD

Table 2. Prediction of the top 30 predicted miRNAs associated with lymphoma

hsa-mir-21	HMDD; dbDEMC
hsa-mir-155	HMDD; dbDEMC
hsa-mir-17	HMDD
hsa-mir-20a	HMDD; dbDEMC
hsa-mir-146a	HMDD; dbDEMC
hsa-mir-19b	HMDD
hsa-mir-18a	HMDD
hsa-mir-92a	HMDD
hsa-mir-19a	HMDD
hsa-mir-16	HMDD
hsa-mir-15a	HMDD
hsa-mir-126	HMDD

Table 1 ***(cont.).*** Prediction of the top 30 predicted miRNAs associated with breast neoplasms

hsa-mir-450a	HMDD
hsa-mir-450b	HMDD
hsa-mir-1915	HMDD
hsa-mir-1258	HMDD
hsa-mir-200	HMDD
hsa-mir-505	HMDD; dbDEMC
hsa-mir-632	HMDD
hsa-mir-1471	HMDD
hsa-mir-922	HMDD
hsa-mir-510	HMDD
hsa-mir-661	HMDD
hsa-mir-202	HMDD; dbDEMC
hsa-mir-298	HMDD
hsa-mir-411	HMDD
hsa-mir-516b	HMDD
hsa-mir-526a	HMDD
hsa-mir-301b	HMDD
hsa-mir-515	HMDD

Table 2 ***(cont.).*** Prediction of the top 30 predicted miRNAs associated with lymphoma

hsa-mir-125b	dbDEMC
hsa-mir-34a	dbDEMC
hsa-mir-145	dbDEMC
hsa-mir-200b	HMDD
hsa-mir-181a	HMDD; dbDEMC
hsa-mir-29c	HMDD
hsa-mir-221	HMDD; dbDEMC
hsa-mir-200a	HMDD
hsa-mir-29a	dbDEMC
hsa-mir-29b	HMDD; dbDEMC
hsa-mir-150	HMDD; dbDEMC
hsa-mir-24	*unconfirmed*
hsa-mir-200c	HMDD
hsa-mir-210	HMDD
hsa-mir-101	HMDD; dbDEMC
hsa-mir-203	HMDD
hsa-mir-125a	HMDD; dbDEMC
hsa-mir-223	dbDEMC

Acknowledgment

This study was presented at the 1st International Conference on Scientific and Academic Research ICSAR 2022.

References

Alexiou, P., Vergoulis, T., Gleditzsch, M., Prekas, G., Dalamagas, T., Megraw, M., . . . Hatzigeorgiou, A. G. (2009). miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Research, 38*(suppl_1), D137-D141. doi:10.1093/nar/gkp888

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., . . . Yu, X. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *nature, 403*(6769), 503-511.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *cell, 136*(2), 215-233. doi:10.1016/j.cell.2009.01.002

Cai, J., Liu, X., Cheng, J., Li, Y., Huang, X., Li, Y., . . . Wei, R. (2012). MicroRNA-200 is commonly repressed in conjunctival MALT lymphoma, and targets cyclin E2. *Graefe's Archive for Clinical and Experimental Ophthalmology, 250*(4), 523-531.

Chandra, S., Vimal, D., Sharma, D., Rai, V., Gupta, S. C., & Chowdhuri, D. K. (2017). Role of miRNAs in development and disease: Lessons learnt from small organisms. *Life Sci, 185*, 8-14. doi:10.1016/j.lfs.2017.07.017

Chen, Q., Lai, D., Lan, W., Wu, X., Chen, B., Liu, J., . . . Wang, J. (2021). ILDMSF: Inferring Associations Between Long Non-Coding RNA and Disease Based on Multi-Similarity Fusion. *IEEE/ACM Trans Comput Biol Bioinform, 18*(3), 1106-1112. doi:10.1109/TCBB.2019.2936476

Chen, X., Wang, L. Y., & Huang, L. (2018). NDAMDA: Network distance analysis for MiRNA-disease association prediction. *Journal of Cellular and Molecular Medicine, 22*(5), 2884-2895. doi:10.1111/jcmm.13583

Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., & Liu, H. (2018). BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics, 34*(18), 3178-3186. doi:10.1093/bioinformatics/bty333

Chen, X., Xie, D., Zhao, Q., & You, Z.-H. (2019). MicroRNAs and complex diseases: from experimental results to computational models. *Briefings in Bioinformatics, 20*(2), 515-539. doi:10.1093/bib/bbx130

Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., . . . Dai, Q. (2016). WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Scientific reports, 6*, 21106. doi:10.1038/srep21106

Chen, X., Yin, J., Qu, J., & Huang, L. (2018). MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput Biol, 14*(8), e1006418. doi:10.1371/journal.pcbi.1006418

Chen, X., Zhou, Z., & Zhao, Y. (2018). ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. *RNA biology, 15*(6), 807-818. doi:10.1080/15476286.2018.1460016

DeSantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A., & Jemal, A. (2016). Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians, 66*(1), 31-42.

DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A., & Jemal, A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians, 67*(6), 439-448.

Espinosa, C. E. S., & Slack, F. J. (2006). Cancer issue: the role of microRNAs in cancer. *The Yale journal of biology and medicine, 79*(3-4), 131-140.

Feber, A., Xi, L., Luketich, J. D., Pennathur, A., Landreneau, R. J., Wu, M., . . . Litle, V. R. (2008). MicroRNA expression profiles of esophageal cancer. *The Journal of thoracic and cardiovascular surgery, 135*(2), 255-260.

Gao, Y., Jia, K., Shi, J., Zhou, Y., & Cui, Q. (2019). A Computational Model to Predict the Causal miRNAs for Diseases. *Front Genet, 10*, 935. doi:10.3389/fgene.2019.00935

Iorio, M. V., Ferracin, M., Liu, C.-G., Veronese, A., Spizzo, R., Sabbioni, S., . . . Campiglio, M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer research, 65*(16), 7065-7070.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., . . . Liu, Y. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res, 37*(Database issue), D98-104. doi:10.1093/nar/gkn714

Kim, Y.-K. (2015). Extracellular microRNAs as biomarkers in human disease. *Chonnam medical journal, 51*(2), 51-57. doi:10.4068/cmj.2015.51.2.51

Kozomara, A., & Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research, 42*(D1), D68-D73. doi:10.1093/nar/gkt1181

Lan, W., Wang, J., Li, M., Liu, J., Wu, F. X., & Pan, Y. (2018). Predicting MicroRNA-Disease Associations Based on Improved MicroRNA and Disease Similarities. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15*(6), 1774-1782. doi:10.1109/TCBB.2016.2586190

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *cell, 75*(5), 843-854.

Li, J.-Q., Rong, Z.-H., Chen, X., Yan, G.-Y., & You, Z.-H. (2017). MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget, 8*(13), 21187. doi:10.18632/oncotarget.15061

Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., & Cui, Q. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Research, 42*(Database issue), D1070-D1074. doi:10.1093/nar/gkt1023

McGirt, L. Y., Adams, C. M., Baerenwald, D. A., Zwerner, J. P., Zic, J. A., & Eischen, C. M. (2014). miR-223 regulates cell growth and targets proto-oncogenes in mycosis fungoides/cutaneous T-cell lymphoma. *Journal of Investigative Dermatology, 134*(4), 1101-1107.

Osada, H., & Takahashi, T. (2011). let-7 and miR-17-92: Small-sized major players in lung cancer development. *Cancer science, 102*(1), 9-17.

Pech, R., Lee, Y.-L., Hao, D., Po, M., & Zhou, T. (2019). LOMDA: Linear optimization for miRNA-disease association prediction. doi:10.1101/751651

Qu, J., Zhao, Y., & Yin, J. (2019). Identification and Analysis of Human Microbe-Disease Associations by Matrix Decomposition and Label Propagation. *Front Microbiol, 10*, 291. doi:10.3389/fmicb.2019.00291

Saydam, F., Değirmenci, İ., & Güneş, H. V. (2011). MikroRNA'lar ve kanser. *Dicle Tıp Dergisi, 38*(1).
Selcuklu, S. D., Donoghue, M. T., & Spillane, C. (2009). miR-21 as a key regulator of oncogenic processes. *Biochem Soc Trans, 37*(Pt 4), 918-925. doi:10.1042/BST0370918

Tan, W., Liu, B., Qu, S., Liang, G., Luo, W., & Gong, C. (2018). MicroRNAs and cancer: Key paradigms in molecular therapy. *Oncol Lett, 15*(3), 2735-2742. doi:10.3892/ol.2017.7638

Tang, C., Zhou, H., Zheng, X., Zhang, Y., & Sha, X. (2019). Dual Laplacian regularized matrix completion for microRNA-disease associations prediction. *RNA biology, 16*(5), 601-611. doi:10.1080/15476286.2019.1570811

Toprak, A., & Eryilmaz Dogan, E. (2021). Prediction of Potential MicroRNA-Disease Association Using Kernelized Bayesian Matrix Factorization. *Interdiscip Sci, 13*(4), 595-602. doi:10.1007/s12539-021-00469-w

Toprak, A., & Eryilmaz, E. (2021). Prediction of miRNA-disease associations based on Weighted K-Nearest known neighbors and network consistency projection. *J Bioinform Comput Biol, 19*(1), 2050041. doi:10.1142/S0219720020500419

van Laarhoven, T., Nabuurs, S. B., & Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics, 27*(21), 3036-3043. doi:10.1093/bioinformatics/btr500

Vural, H., & Kaya, M. (2018). Prediction of new potential associations between LncRNAs and environmental factors based on KATZ measure. *Computers in biology and medicine, 102*, 120-125. doi:10.1016/j.compbiomed.2018.09.019

Wang, C. C., Chen, X., Yin, J., & Qu, J. (2019). An integrated framework for the identification of potential miRNA-disease association based on novel negative samples extraction strategy. *RNA biology, 16*(3), 257-269. doi:10.1080/15476286.2019.1568820

Wang, D., Wang, J., Lu, M., Song, F., & Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics, 26*(13), 1644-1650. doi:10.1093/bioinformatics/btq241

Watanabe, A., Tagawa, H., Yamashita, J., Teshima, K., Nara, M., Iwamoto, K., . . . Nakagawa, T. (2011). The role of microRNA-150 as a tumor suppressor in malignant lymphoma. *Leukemia, 25*(8), 1324-1334.

Yan, F., Zheng, Y., Jia, W., Hou, S., & Xiao, R. (2019). MAMDA: Inferring microRNA-Disease associations with manifold alignment. *Comput Biol Med, 110*, 156-163. doi:10.1016/j.compbiomed.2019.05.014

Yang, J.-H., Shao, P., Zhou, H., Chen, Y.-Q., & Qu, L.-H. (2009). deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Research, 38*(suppl_1), D123-D130. doi:10.1093/nar/gkp943

Yang, Z., Ren, F., Liu, C., He, S., Sun, G., Gao, Q., . . . Zhao, H. (2010). dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics, 11 Suppl 4*, S5. doi:10.1186/1471-2164-11-S4-S5

Yin, M. M., Liu, J. X., Gao, Y. L., Kong, X. Z., & Zheng, C. H. (2022). NCPLP: A Novel Approach for Predicting Microbe-Associated Diseases With Network Consistency Projection and Label Propagation. *IEEE Trans Cybern, 52*(6), 5079-5087. doi:10.1109/TCYB.2020.3026652

Yu, S. P., Liang, C., Xiao, Q., Li, G. H., Ding, P. J., & Luo, J. W. (2019). MCLPMDA: A novel method for miRNA-disease association prediction based on matrix completion and label propagation. *J Cell Mol Med, 23*(2), 1427-1438. doi:10.1111/jcmm.14048