Journal of Biomedical Informatics 65 (2017) 34–45

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Prediction of microRNAs involved in immune system diseases through network based features

Archana Prabahar, Jeyakumar Natarajan *

Data Mining and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore 641 046, India

ARTICLE INFO

Article history:
Received 12 July 2016
Revised 28 September 2016
Accepted 13 November 2016
Available online 15 November 2016

Keywords:
miRNA
Immune diseases
Network
Topological features
SVM
Classification

ABSTRACT

MicroRNAs are a class of small non-coding regulatory RNA molecules that modulate the expression of several genes at post-transcriptional level and play a vital role in disease pathogenesis. Recent research shows that a range of miRNAs are involved in the regulation of immunity and its deregulation results in immune mediated diseases such as cancer, inflammation and autoimmune diseases. Computational discovery of these immune miRNAs using a set of specific features is highly desirable. In the current investigation, we present a SVM based classification system which uses a set of novel network based topological and motif features in addition to the baseline sequential and structural features to predict immune specific miRNAs from other non-immune miRNAs. The classifier was trained and tested on a balanced set of equal number of positive and negative examples to show the discriminative power of our network features. Experimental results show that our approach achieves an accuracy of 90.2% and outperforms the classification accuracy of 63.2% reported using the traditional miRNA sequential and structural features. The proposed classifier was further validated with two immune disease sub-class datasets related to multiple sclerosis microarray data and psoriasis RNA-seq data with higher accuracy. These results indicate that our classifier which uses network and motif features along with sequential and structural features will lead to significant improvement in classifying immune miRNAs and hence can be applied to identify other specific classes of miRNAs as an extensible miRNA classification system.

1. Introduction

In this post genomic era, the importance of microRNAs (miRNAs) involved in the regulation of genes has become more and more apparent. miRNAs regulate their targets by translational inhibition and mRNA destabilization [1]. miRNAs are small (~22 nucleotide), noncoding RNAs that bind to their cognate mRNAs via a recognition sequence region named seed sequence which is located between second and eighth nucleotides of the miRNA [2]. With respect to the immune system, these small regulators are known to affect all facets of immune system development, from hematopoiesis to activation in response to infection during both the innate and the adaptive immune response [3]. More than 1800 human miRNAs have been discovered so far, and they are known to regulate approximately 30–60% of all protein-coding mRNA genes [4].

Generally miRNAs have multiple mRNA targets. Such miRNAs may belong to same miRNA family, with members sharing identical or highly similar seed sequences [5]. A single miRNA can regulate hundreds to thousands of target genes and play a significant role in the regulation of a large percentage of the genome including activities such as controlling the development of a normal and functional arm of the immune system [6,7]. Thus, deregulated miRNA expression can cause serious complications for the immune system. Similar to mRNA, miRNA transcription and processing involves complex regulatory mechanisms. Deregulation of these miRNAs and its target may be linked to the development and progression of several diseases including immune disorders and cancer [8]. Thus miRNAs have emerged as novel molecular regulators of numerous genes and pathways involved in normal immune responses, pathogenesis of cancer, inflammation and autoimmune diseases [9].

Autoimmunity results when the immune system fails to recognize self and directs immune responses against self-antigens, leading to cellular and tissue destruction [10]. Within the past decade, the field of immunology has increasingly intersected with the field of miRNA biology [11]. While the earliest studies of miRNA function in the immune system have demonstrated an essential role for miRNAs as a whole, recent studies have focused on the

* Corresponding author.
E-mail addresses: archana.prabahar@gmail.com (A. Prabahar), n.jeyakumar@yahoo.co.in (J. Natarajan).

http://dx.doi.org/10.1016/j.jbi.2016.11.003
1532-0464/

contribution of specific miRNAs to specific immunologic processes [12,13].

The first computational analysis which involves miRNAs and protein-protein interaction networks (PPIN) was done by Liang and Li [14]. The modeling and analysis of protein interactions as undirected graphs has been adopted for prediction of protein function [15], functional module identification [16], candidate disease gene identification [17] and prioritization of disease genes [18]. miRNA-regulated PPI networks will not only determine the functional role of miRNAs that play in PPI networks, but will also determine the potential candidates involved in the dysregulatory function of a disease [19]. Hsu et al. studied the human miRNA-regulated PPI network by utilizing Human Protein Reference Database (HPRD) data [20]. They found that miRNA often targets the hub gene of the PPI network, despite of the fact that they are not involved in characterizing the pathological events regulated by miRNA target genes [21].

A number of machine learning tools such as support vector machines, neural networks, hidden markov model and naive bayesian techniques have been widely used to predict miRNAs [22–25]. These machine learning based computational tools identify miRNAs using features such as sequence conservation, structure, and folding energy of sequences as their training data [27,28]. In addition to the above generic miRNA classification systems, Xu et al., introduced a network centric approach which uses four network topological features and one sequence fold change feature to classify miRNAs that was significantly different between prostate cancer and non-prostate cancer [26].

Motivated by the above works, we developed an immune miRNA classification system to predict immune disease associated miRNAs from non-immune miRNAs. In addition to the traditional sequence and structural features, ten network topological features and two sequence motif features were added to illustrate the performance of the classifier. The network topological features were generated using a heterogeneous network which includes miRNA-disease association, miRNA-target gene association, and target gene-gene association. In addition, network motif patterns that were over represented in the immune network were also used as network motif features. To our knowledge this is the first available approach to classify immune miRNAs from non-immune miRNAs based on network topological and motif features. The workflow of this classification approach was shown in Fig. 1.

2. Methods

2.1. Dataset

2.1.1. Positive data

Immune miRNA disease dataset was retrieved from two curated databases namely, Human miRNA associated disease database (HMDD) [29] and miR2Disease [30]. Additionally, new associations from the literature based on the MeSH disease category of immune system diseases were also included in the dataset. Finally, the positive dataset contains 245 miRNAs (Supplementary data1, Table S1)[1] involved in 92 immune system diseases (Supplementary data1, Table S2).

2.1.2. Negative data

Compiling the set of negative data is more difficult for the classification problem. Hence, we have used two different negative datasets. In the first dataset, non-immune miRNAs were retrieved using the same strategy for immune miRNA extraction from the two standard databases HMDD [29] and miR2Disease [30]. The condition applied here was that in the two databases the selected negative miRNAs should not have any association with immune system diseases such as inflammation, cancer and autoimmune diseases. In miRNA classification problem, earlier researchers demonstrated that balanced datasets with equal number of positive and negative samples provided a higher sensitivity and specificity whereas specificity is increased highly with the increase in negative samples [25]. This was also addressed by most of the current researchers where they used balanced datasets to avoid data imbalance problem [31,32]. A set of 245 non-immune miRNAs were randomly selected and used as negative dataset (Dataset 1) to avoid the imbalance problem (Supplementary data1, Table S3).

To further enhance the classification results, an additional negative gene expression dataset was chosen for the study from GEO (GSE61741). GSE61741 was specifically chosen to determine the miRNAs that are not expressed in immune system diseases. Genes that showed higher expressions in a specific disease were referred to as disease expressed genes [33,34]. Hence it is assumed that miRNAs with lowest expression in normal sample are the least involved in immune diseases. Similar to earlier balanced dataset, 245 miRNAs with lower expression values were chosen as negative dataset 2 (Supplementary data1, Table S4).

2.2. Features set

The features used for miRNA prediction includes three miRNA sequential features such as AT, GC content and miRNA length and standard structural features such as minimal free energy, self containment index (SCI) [35], stem, loop and additional 32 triplet base pair features, which were computed using RNAfold [36]. Further, we used 10 novel network based features compiled from miRNA target gene–disease associations and 2 motif features for immune specific miRNA classification. In total, we have used 51 features for the classification task (Table 1). The compilation steps of the network and motif features were explained below.

2.2.1. Network features

The compilation of network based features involves four prerequisite information extraction steps from various external databases as given below:

(i) miRNA target gene extraction
(ii) miRNA target gene – disease association extraction
(iii) miRNA – disease association extraction
(iv) miRNA target gene – protein association extraction

2.2.1.1. miRNA target gene extraction

miRNA target genes were acquired from eight miRNA target databases: miRanda [37], PicTar [38], TargetScan [39], DIANA-microT [40], RNA22 [41], RNAhybrid [42], miRDB [43] and PITA [44]. We extracted the regulatory associations between miRNAs and target genes that were found to appear in at least four of these databases in order to populate the most reliable miRNA-target associations. The miRNA-gene association data thus obtained contains 101,426 miRNA-gene associations involving 245 miRNAs and 11,130 genes.

2.2.1.2. miRNA target gene – disease association extraction

Finally, gene disease associations were obtained from curated entries of CTD database [45] for all the MeSH diseases. A total of 6537 gene-disease associations were obtained.

[1] Supplementary data S1–S7: Available for download at http://biominingbu.org/SupplementaryData.

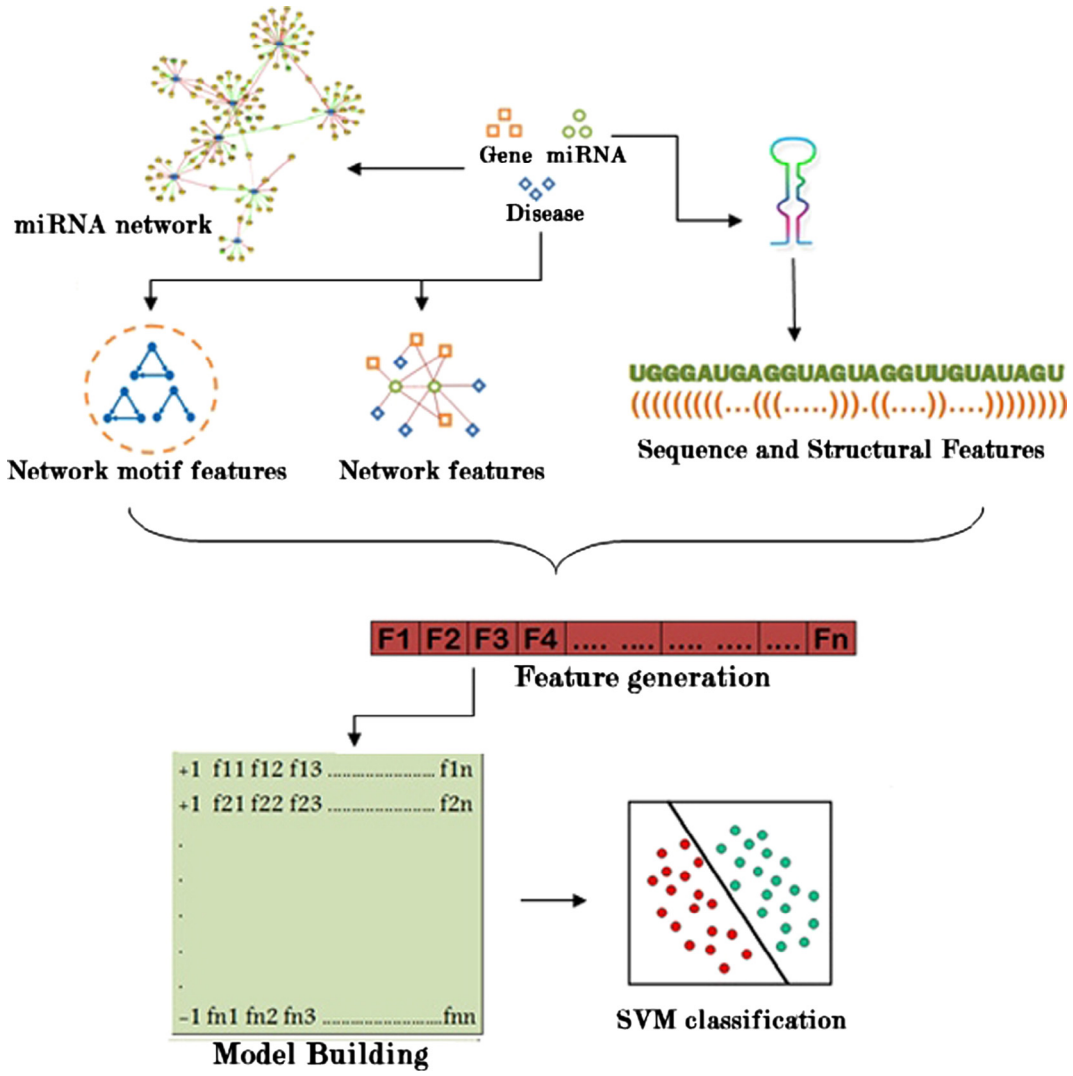


Fig. 1. Workflow of miRNA classification approach. (The complete work flow of immune miRNA classification approach depicts the various features generated and the classification schema using SVM.)

2.2.1.3. miRNA target gene – protein association extraction

miRNA target gene protein associations were extracted from the HPRD (Release 9) database [20]. The current release of HPRD contains 145,134 human PPI interactions. The miRNA target genes were mapped to the HPRD and the corresponding PPI were extracted. The extracted PPI network contains 18,753 genes and 570,870 interactions. The data retrieval of target genes, PPI and its integration workflow is shown in Fig. 2.

2.2.1.3.1. Immune gene targets. Similar to miRNA target gene protein network, the PPI were extracted for the target genes which were exclusively related to immune diseases and immune gene protein network was constructed as shown in Fig. 2. The target genes which were related to immune diseases and their PPI were found using two databases IMGT [46] and InnateDB [47]. The Immune functional gene network contains 8230 immune genes which are involved in 22,683 immune target gene interactions and 73,046 immune target gene protein interactions.

miRNA disease association data was extracted as described in Section 2.1.1. From these four associations (i.e. miRNA-disease, gene-disease, miRNA-gene and gene-gene interactions), a heterogeneous network of immune miRNAs was constructed using Cytoscape [48].

In this network, nodes represent biomedical entities (i.e. miRNA, disease, or gene), and edges between nodes represent associations between two nodes (i.e. association between miRNA and genes, miRNA and disease, gene and gene, etc.). For each miRNA in the network, we defined 10 measures (i.e. DTout, Dout, NmiRNA, RpcmiRNA, RtarpcmiRNA, miRD, miRPPI-interactome, miRTGI, miRDcoreg, miRDSW) based on network properties. Further, to assess the significance of variation in these 10 features between the immune and non-immune miRNA, the median and their p-value was computed. The discrimination between positive and negative instances is determined using p-value computed by Mann-whitney Wilcoxon's test using R-statistical package. Each feature with its description and corresponding p-value was shown in Table 1.

2.2.2. Motif and sequential features

In addition to above network features, sub networks were generated to identify smaller common patterns, or motifs. Small over-represented motifs found in the biological networks forms an essential functional unit of various biological processes [49]. The feed-forward loop (FFL) motifs are found to be over-represented in biological networks and they represent the functional units of

Table 1
SVM features for miRNA network prediction.

	Feature	Description	Median		P-value
			Positive	Negative	
Network features					
1	D_{Tout}	The total number of nodes from target miRNA	365	296	0.0137
2	Dout [Xu et al.]	Total number of target genes connected to miRNA	9	6	0.2267
3	NmiRNA [Xu et al.]	Number of coregulatory miRNAs	13	7	0.0340
4	R_{pc}-miRNA [Xu et al.]	The proportion of Immune miRNAs in coregulatory set	1.597	0.036	0.0035
5	R_{tarpc}-miRNA [Xu et al.]	Fraction of targets coregulated by itself and other immune miRNAs	0.2857	0.0739	0.0054
6	miR_{D}	Total number of diseases connected to miRNA	2	0	0.0033
7	$miR_{PPI\text{-}interactome}$	Total number of protein pairs in the interactome network	2368	2039	0.0010
8	miR_{TGI}	Total number of immune target genes connected to miRNA	32	25	0.0512
9	miR_{Dcoreg}	Number of disease coregulators of miRNA	7	3	0.0034
10	miR_{DSW}	Disease spectrum width of miRNA	2.392	0.893	0.0034
Motif features					
11	miR_{NM}	Number of motifs in the network	5	2	0.0056
12	miR_{FFL}	Number of feed-forward loops in the network	2	0	0.067
Sequential features					
13	miR_{length}	Length of the miRNA	84	80	$1.95\ e^{-06}$
14	miR_{GC}	GC content of miRNAs	4.68	5.68	0.248
15	miR_{AU}	AU content of miRNAs	3.63	3.843	0.365
Structural features					
16	miR_{MFE}	Minimal free energy of miRNA	−38.3	−24.6	0.0026
17	miR_{SCI}	Modularity of miRNA (self containment index)	0.937	0.812	0.3646
18	miR_{loop}	miRNA loops	0.222	0.15	0.0237
19	miRstem	miRNA stems	0.340	0.33	0.338

32 triplet base pair features are also computed.

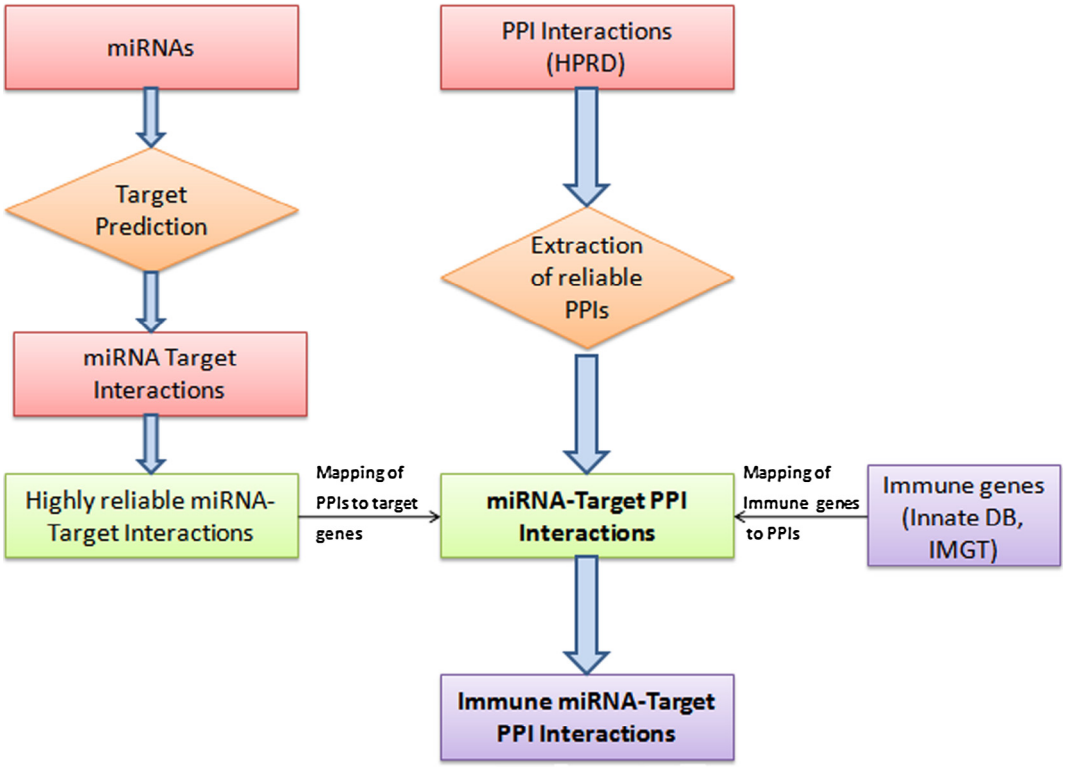


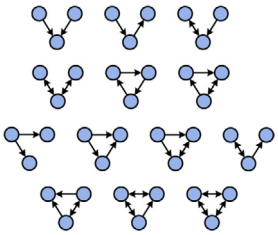
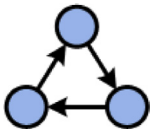
Fig. 2. Data retrieval of target genes, PPI and its integration – Workflow.

biological processes in cells [50]. For this, we defined two measures to compute motif features (i.e. miRNM, miRFFL). The motifs and its computation are shown in Table 2. For these motif features, three noded network motifs were identified from miRNA–disease–gen e-protein network. To generate statistically significant network motifs ($P < 0.05$), 1000 randomized networks were generated. Motifs that appeared at least 5 times in the network and have a Z-score greater than 2 were chosen. The Z-score and P-value were calculated by the network motif (NM) discovery procedure using FANMOD tool [51].

2.3. Functional properties of miRNA

Next, we carried out the following two experiments to demonstrate the functional properties of miRNAs used in our classification task.

Table 2
Illustration of motifs and its feature description.

Motifs	Description	Property
	Count of number of motifs in the network	miRNM
	Count of number of feed-forward loops (over represented network motif) in the network	miRFFL

(i) Human miRNA clustering
(ii) Pathway and GO enrichment analysis.

These studies help us to gain some preliminary knowledge about both immune related (positive dataset) and non immune (negative dataset) miRNAs used in this study. Both the methods were briefly discussed below:

2.3.1. Human miRNA clusters

A miRNA cluster is a set of miRNAs in which each member has at least one other member of the same set within 10 kb according to chromosomal locations. Chromosomal positions for all miRNAs in the human immune system were obtained from miRBase [52]. About 48 human miRNA clusters were obtained from miRBase [52]. Similarly miRNA family clusters were obtained for all the 245 miRNAs based on their gene family determined using miRFam as shown in Supplementary data2. The clustering results helps us to understand how miRNA clusters were formed among immune miRNAs and also to demonstrate whether the miRNAs associated with same diseases occur in same clusters.

2.3.2. Pathways and GO enrichment analysis

Functional annotation of genes involved in immune miRNA network is performed using KEGG [53] and GO [54] dataset available from miRwalk database [55]. Target genes involved in immune miRNA network were subjected to cross validation with GO annotation terms and enriched pathways from KEGG. Enriched pathways and functional annotations of various immune signaling cascades were obtained using miRwalk. The involvement of immune signaling pathways in this functional network were analyzed using Immune signaling pathway networks of KEGG. The pathway and GO enrichment analysis helps us to show whether the target genes of miRNA were associated with immune related GO categories such as t-cell activation and innate immune response and KEGG pathway categories such as B-cell and T-cell receptor signaling pathway.

2.4. Classification algorithm and validation

2.4.1. Support vector machine (SVM)

For classification, we used SVM classifier to classify immune miRNAs from other non-immune miRNAs. SVM can process classification and prediction through different kernel functions and suitable parameters. SVM classifier, using a radial bias function (RBF) available in the SVMlight package [56] was used to construct a classifier [57]. To build the good SVM models, two factors Gamma (γ) and C are especially important. C is the parameter for the soft margin cost function in the support vector which controls the influence of each component vector and has trading error penalty. Gamma is the parameter of a RBF Kernel which could cause lower data bias when the margin is small and the variance will be higher. A suitable adjustment of these parameters results in a better classification hyper-plane found by the SVM, and thereby enhances the classification accuracy. A grid search method [58] was used to tune the SVM parameters C and γ. The grid search method checks all possibilities of C and γ and determines the optimum parameters of the classifier. The SVM was trained and tested using five-fold cross-validation.

2.4.2. Principal component analysis (PCA)

In addition, we also explored a feature selection method PCA to demonstrate the impact of feature selection algorithm in miRNA classification. PCA is a powerful tool for dimension reduction and feature extraction in data analysis [59]. The main advantage of PCA is that when data compression is performed, i.e. by reducing the number of dimensions, loss of information is much limited. miRNA features generated were subjected to feature selection using PCA [60]. PCA involves the following steps (i) subtraction of mean from the feature vectors, (ii) calculation of covariance matrix, (iii) calculation of eigenvectors and eigenvalues of the covariance matrix, (iv) choosing the components from feature vector and (v) derivation of new dataset. Finally the transpose of the generated feature vector is multiplied with the original data. An in-house developed PCA feature selection program using these computations was used for miRNA feature reduction and analysis.

2.4.3. Performance evaluation

The performance of our method was evaluated using five-fold cross-validation method [26,61]. A five-fold cross-validation method was applied, where the miRNA data was divided into five equal-length datasets, four of which were used for training in each turn, with one dataset held out for testing and the assessment was performed. To measure the performance of the prediction system, values such as Accuracy (Ac), Sensitivity (Sn), Specificity (Sp) and Receiver Operating Characteristic (ROC) were calculated using the following equations:

$$\mathrm{Sn} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$
$$\mathrm{Sp} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP})$$
$$\mathrm{Acc} = (\mathrm{TN} + \mathrm{TP})/(\mathrm{TN} + \mathrm{FP} + \mathrm{TP} + \mathrm{FN})$$
$$\mathrm{ROC} = \mathrm{Sensitivity}/1 - \mathrm{specificity}$$

where:

TP = True Positives, TN = True Negatives
FP = False Positives, FN = False Negatives

3. Results and discussion

3.1. Network based classification approach

Pre-miRNA prediction methods rely on sequential and structural based features generated from miRNA sequences. These approaches aim to discriminate pseudo hairpins from known and novel miRNAs [62] or locate the mature miRNAs [63]. However, an approach to discriminate miRNAs based on their function remains unexplored. In the current investigation, we aimed to predict miRNAs that are specific to immune function mainly based on network features. This novel approach incorporating network and motif features would demonstrate a better discriminative power than traditional methods based on structural and sequential features.

For this classification approach, heterogeneous network was constructed for each of the miRNA using gene, disease and protein

interaction with the miRNA from which network and sub-network features were computed. The heterogeneous miRNA target gene network was shown in Supplementary data3.

Network features were computed based on the global properties of the miRNA. To further understand the network functionality, sub-networks were generated and provided as a feature to the classifier. More than five NMs including feed-forward NMs were determined for each of the miRNA in the network with a significant p-value of less than 0.05. We then investigated the degree distribution of different types of entities in the integrated network such as miRNA and genes and generated a scale-free network (a network whose degree distribution follows power law). The correlation value (R^2) was found to be 0.91 and 0.94 for in-degree and out-degree distribution of miRNA and target genes. The heterogeneous miRNA network and its degree distribution graph were shown in Fig. 3. The degree distribution of this miRNA network follows power law distributions and their R^2 values are both greater than 0.90, and hence the network is scale-free, which is one of the most important characteristic of real-world biological networks [64,65].

3.2. Functional properties of immune miRNAs

Functionally related miRNAs tend to be associated with phenotypically similar diseases. For example, one of the autoimmune diseases, Psoriasis is regulated by 34 miRNAs [29,30]. Based on these properties, miRNA co-regulators of diseases were chosen as one of the feature. miRNAs are pleotropic, and hence they share common mRNA targets between other miRNAs. Hence, determination of gene co-regulators would also be one of the key factors in miRNA classification. Disease spectrum width (DSW) of one miRNA could be a metric to evaluate its importance in function and human disease [29]. This is calculated as fraction of the total number of diseases associated with a particular miRNA and the total number of diseases that have been reported to be associated with the complete miRNA dataset.

miRNAs tend to be associated with same diseases, if they occur in same clusters based on chromosomal location or its family. Mir-381 located in the chromosome 14, forms one of the largest cluster with 19 miRNAs. For instance, miRNAs that belong to let-7 family are known to be involved in immune associated diseases. They form the largest cluster in the immune miRNA data with 12 known miRNAs. mir-10 family has 8 miRNA associations in its cluster. Other miRNA families such as mir-154, mir-181 and mir-515 share 5 miRNAs each in its cluster. They are also known to be associated with immune system diseases.

To further cross validate the functional properties of the target genes associated with immune miRNAs, we performed the enrichment analysis of the genes using KEGG [53] and GO [54] dataset available from miRwalk database [55]. From this analysis it is found that 2741 target genes are involved in the immune signaling functions and pathways. From the results, 790 target genes are known to be involved in KEGG immune signaling pathways and

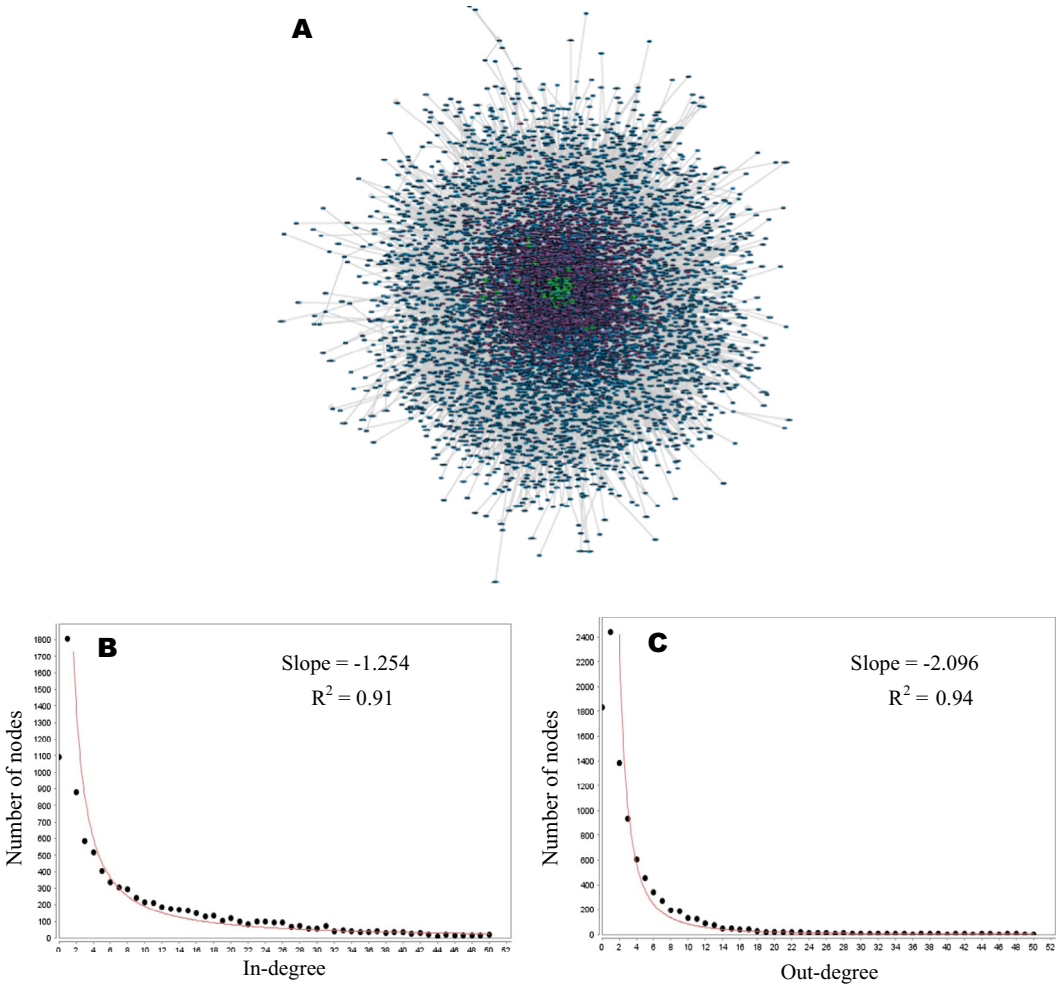


Fig. 3. Immune miRNA network. A represents the Immune miRNA network representing 245 miRNAs and its interactome. (Green nodes represent the miRNAs, Pink nodes represents target genes and the Blue nodes represent the PPIs in the network.) B and C represent the in-degree and out-degree distribution of the network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

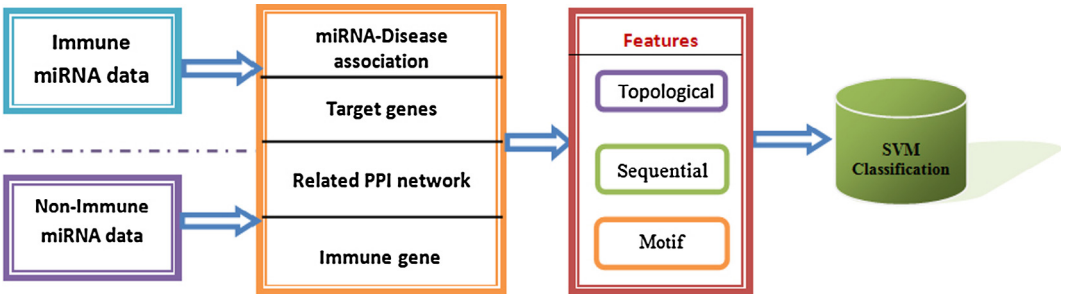


Fig. 4. Data generation process. (The flow diagram represents the steps involved in feature generation and further classification from miRNA network.)

Table 3
Five-fold cross-validation results of SVM classifier.

Features	Sensitivity		Specificity		Accuracy	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
SS	0.632 ± 0.044	0.624 ± 0.015	0.616 ± 0.019	0.640 ± 0.016	0.624 ± 0.029	0.632 ± 0.008
NM	0.869 ± 0.010	0.878 ± 0.009	0.861 ± 0.007	0.873 ± 0.007	0.865 ± 0.005	0.876 ± 0.002
NMSS	0.893 ± 0.017	0.881 ± 0.007	0.865 ± 0.004	0.898 ± 0.009	0.879 ± 0.008	0.889 ± 0.006
Combined FS	0.910 ± 0.010	0.906 ± 0.008	0.885 ± 0.010	0.897 ± 0.008	0.897 ± 0.004	0.902 ± 0.005

the remaining 1951 genes are involved in GO functions of the immune system. Target genes that are involved in the immune functions are further identified using the list of immune genes (Supplementary data4) collected from the immune gene databases such as InnateDB [47] and IMGT [46].

3.3. SVM classification and evaluation

Five-fold cross-validation evaluations were carried out with SVM classifier to show the effect of network and motif features in immune miRNA classification. Four different evaluations were carried out using two different datasets (Dataset 1: 245 positive data and 245 negative data from miRNA disease associations databases HMDD and miR2Disease, Dataset 2: 245 positive data (dataset1) and 245 negative data from microarray gene expression) using RBF Kernel in SVM classifier [58]. The data generation process for the classification study is shown in Fig. 4.

For this study, four different evaluations were performed using the following feature sets:

(i) Sequential + Structural features (SS) (3 + 36 features)
(ii) Network + Motif features (NM) (10 + 2 features)
(iii) Network + Motif + Sequential + Structural features (NMSS) (10 + 2 + 3 + 36 features)
(iv) Combined features + feature selection using principal component analysis (NMSSFs)

All the evaluation results were shown in Table 3 for dataset 1 and dataset 2. Extraction procedure for negative dataset 2 was shown in Fig. 5. Evaluation 1 includes a set of 35 structural features and 3 sequential features [61] computed from the primary sequence and its secondary structure using RNAfold program [36] along with miRNA modularity feature (SCI index) [35]. Evaluation 2 is performed using a set of 10 network topological features and 2 sub-networks features and achieves a better accuracy of 86.5% and 87.6% than evaluation 1. Evaluation 3 combines the first two features with network features and the accuracy is improved to 87.9% and 88.9% for the two different datasets. In the final evaluation, all these features are combined and a feature selection algorithm using principal component analysis is applied to the datasets in order to enhance the accuracy. This combined approach (NMSSFs) achieves the best results with an accuracy of 89.7% and 90.2% for dataset1 and dataset2 respectively. ROC plot for these four evaluations is constructed and compared for the two different datasets (Fig. 6). We then investigated the ROC curve of the classifier and found that the Area Under Curve (AUC) of the classifier was around 0.9 (0.91), suggesting that the classifier has performed a good discrimination.

3.4. Comparison of feature selection approaches

Classification performance was comparatively poor when structural features were combined with sequential features (SS) and achieved an accuracy of 62.4% and 63.2% for dataset1 and dataset2. Corresponding sensitivity and specificity were shown in Table 3. AUC metrics computed from ROC plot were shown in Table 4. These results show that topological features of network and motif (NM) could improve the prediction power of miRNAs associated with immune disease than SS features. To further illustrate the combined effect of network, motif, sequential and structural features (NMSS) on the performance of the classifier, a combination model is generated. This achieved the top scoring results with an accuracy, sensitivity and specificity of 88.9%, 88.1% and 89.8% respectively.

This combinatorial model (NMSSFs) has 51 features (10 network features, 2 sub-network features, 3 sequential features and 36 structural features). To test whether the performance accuracy improves upon feature selection, principal components were used to serve as feature selection for classification problems. Top 20 most discriminative features were further selected for the classification problem. It is obvious that applying feature selection method improves the accuracy to 89.7% and 90.2% for two different datasets. Hence, the integrated feature selection and classification algorithm is capable of identifying significant miRNAs. From the classification results, it is evident that classification model performs even better while using PCA. The best results were achieved by selecting the first 20 principal components, out of the 51 features used in the classification system. The use of PCA has not only increased the accuracy of the classifier, but also decreased the standard variation of the results. These results show that using this PCA method in combination with SVM classifier may improve the classification accuracy.

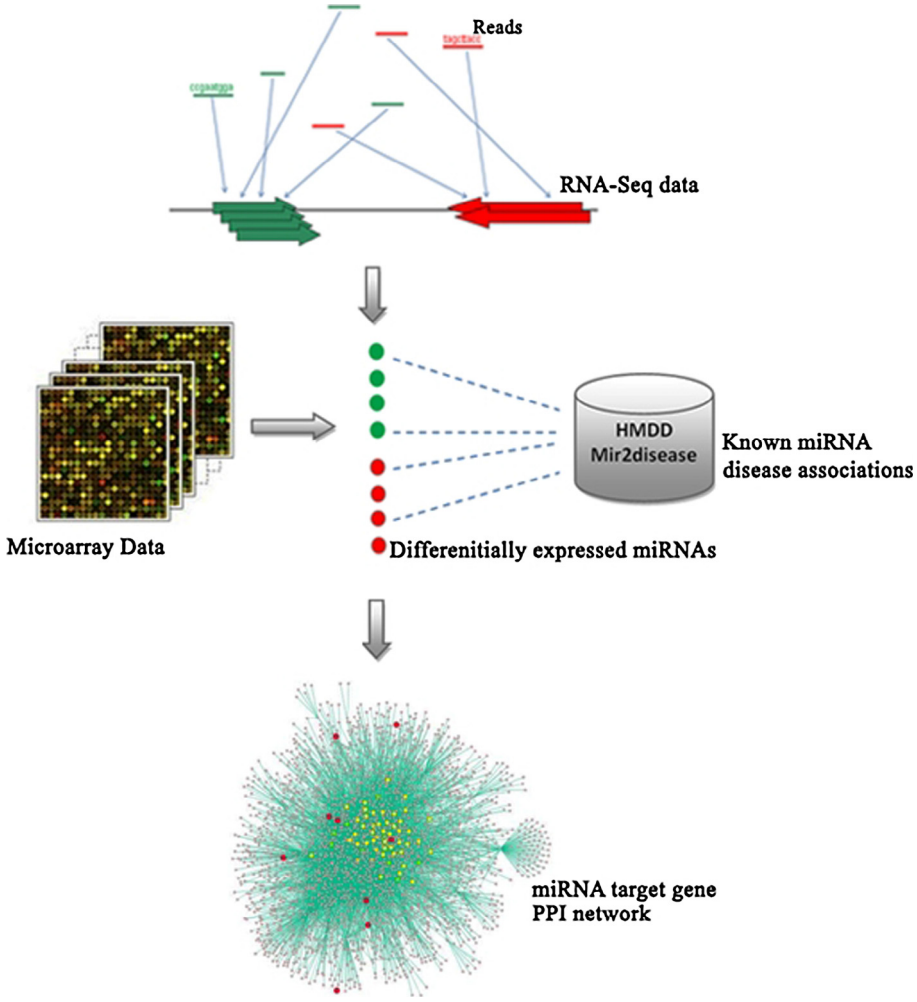


Fig. 5. Mapping of miRNA and network construction. (Differentially expressed miRNAs are mapped to known miRNA disease association from HMDD and miR2disease to be considered as classification data.)

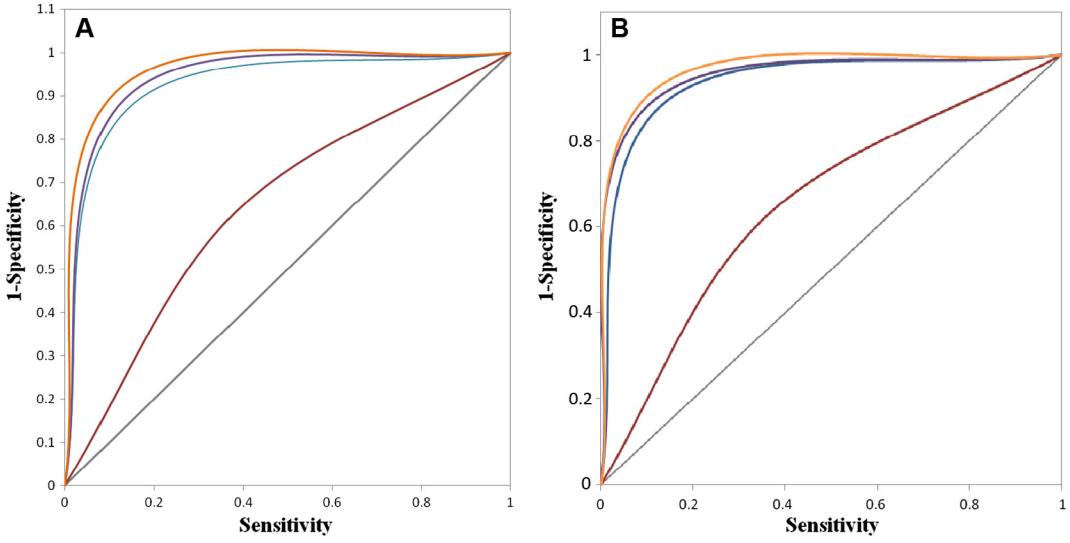


Fig. 6. ROC curve of immune miRNA classifier. This curve is plotted for two different datasets. A represents Dataset 1: positive dataset and negative dataset from miRNA disease associations. B represents Dataset 2: Positive dataset from miRNA disease association and negative dataset from microarray data. Evaluations were performed for different methods of features and plotted as SS (Red), NM (Blue), NMSS (Violet) and Combined Fs (Orange). Diagonal line represents the baseline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Area under curve metrics of ROC plot.

Features	AUC Five-fold cross-validation	
	Dataset 1	Dataset 2
SS	0.591 ± 0.031	0.632 ± 0.042
NM	0.869 ± 0.006	0.875 ± 0.007
NMSS	0.8685 ± 0.003	0.889 ± 0.006
Combined FS	0.890 ± 0.010	0.912 ± 0.008

3.5. Evaluation on unbalanced negative data

To demonstrate the size of the negative examples in classifier performance, the second evaluation was carried out by combining both negative datasets (Dataset1 and Dataset 2) and choosing a random set of 80, 160, 245, 370, 490 negative examples by retaining the positive examples as constant (245 known miRNAs).

The NMSSFs features which gives best results in balanced data classification was used as features for this evaluation. SVM classification with five-fold cross-validation was repeated for the unbalanced negative data and the evaluation results were shown in Table 5. The results shows that with equal number of 245 positive and 245 random negative data yielded 91% sensitivity, 88.5% specificity and with a higher accuracy of 89.7%. When the number of negative instances increases, there is increase in specificity by 10% but there is also decrease in the sensitivity of the classifier. The results indicate that with higher number of negative instances there is a possibility of more false predictions and agrees with the results reported by Yousef et al. [25]. Hence, balanced dataset might influence the correct predictions and appreciate the performance of the classifier with higher accuracy. Prediction performance of the classifier as a function of size of the negative class is shown in Fig. 7.

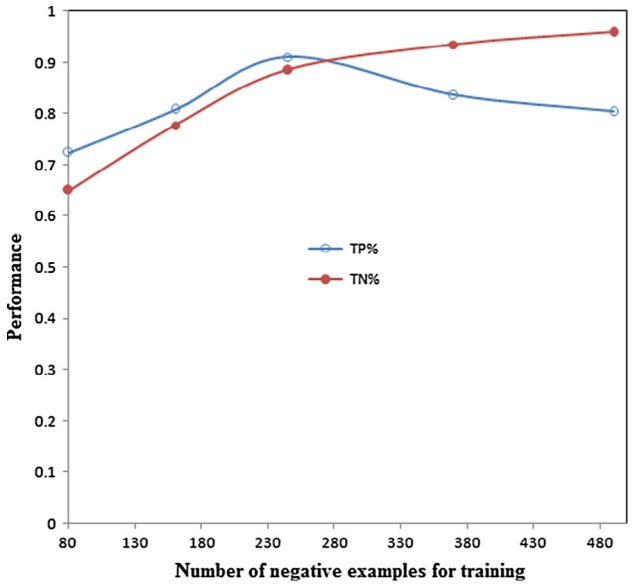


Fig. 7. Prediction performance as a function of size of the negative class for miRNAs. %TN is the true negative ratio and %TP is the true positive ratio.

Table 6
Performance of SVM classifier for Prostate cancer dataset.

Features	Sensitivity	Specificity	Accuracy
Combined features + PCA (Combined FS)	0.921 ± 0.072	0.866 ± 0.041	0.890 ± 0.003
Network features + fold change	0.864 ± 0.004	0.875 ± 0.001	0.870 ± 0.002

3.6. Comparison with other methods

To our knowledge the only other system, which utilizes network features to categorize disease specific miRNAs is the Prostate cancer classification by Xu et al. [26]. In this work, they used four network features (Dout, NmiRNA, Rpc-miRNA, Rtarpc-miRNA) and a gene expression fold change feature (Fc) in prostate cancer dataset of 37 positive samples and 44 negatives (Supplementary data1, Table S5) from a microarray experiment (GSE8126) [26]. However, in addition to the above four network features, we used 7 additional network features and 2 motif based features. We compared our classification accuracy with above datasets and results are tabulated in Table 6. Based on these results, we conclude that NMSSFs method of feature selection is the effective method for model generation and classification of heterogeneous disease network data. An ROC curve is plotted to show the performance evaluation of this approach (Fig. 8).

Chen et al. developed a novel method to classify breast cancer samples from normal sample. This method is based on network motif approach where they combine a human signaling network and high-throughput gene expression data [66]. In this study, five motif features were used such as motifs, motif genes, breast cancer (BC) genes, marker genes and bc-marker genes. Differentially expressed high-stability significant expression-correlation differential motifs (HSCDMs) were identified as classification features and a SVM classification is performed [66]. Our study is different from this approach, as they deal with gene expression dataset and HSCDMs. In our approach, we developed miRNA PPI interactive network and further defined the feature vectors using network topological and motif based features such as motifs [29] and feed forward loops (FFL) in the network.

3.7. Case studies

To further demonstrate the usefulness of our system, we conducted two case studies with our proposed network feature based classification system. The case study 1 is based on microarray gene expression data and the case study 2 is based on RNA-seq data. The result indicates that our network based classification system better discriminates the miRNA in both gene expression as well as RNA-seq data.

Table 5
Performance of five-fold cross-validation of negative dataset (Random size variation repeated 100 times).

Negative samples (size)	Sensitivity	Specificity	Accuracy	AUC
80	0.724 ± 0.046	0.65 ± 0.083	0.706 ± 0.050	0.687 ± 0.030
160	0.808 ± 0.030	0.777 ± 0.037	0.795 ± 0.037	0.792 ± 0.057
245	0.910 ± 0.023	0.885 ± 0.023	0.897 ± 0.010	0.897 ± 0.034
370	0.836 ± 0.045	0.935 ± 0.032	0.895 ± 0.020	0.885 ± 0.026
490	0.804 ± 0.055	0.959 ± 0.016	0.907 ± 0.025	0.881 ± 0.034

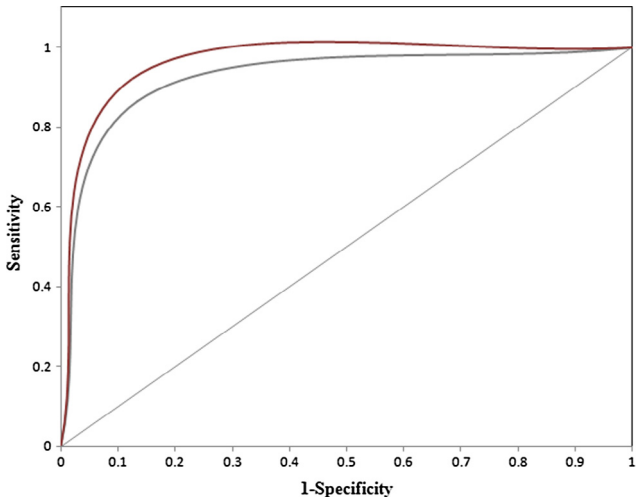


Fig. 8. ROC curve for Prostate cancer dataset (Comparison of Prostate cancer dataset generated by Xu et al. using network + Fold change features (Xu et al.) and Combined Fs features (Our approach)) red curve represents the plot of our approach and grey curve indicates the ROC curve of Xu et al. method. Diagonal line represents the baseline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

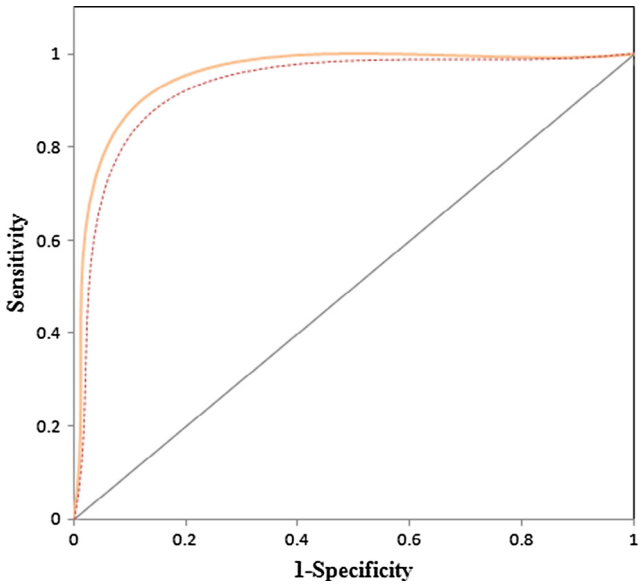


Fig. 9. ROC curve for Multiple sclerosis (MS) dataset. (Performance evaluation of MS dataset is performed using Combined Fs features (Our approach – solid line) and network + Fold change features (Xu et al. approach – dotted line.) Diagonal represents the baseline).

Table 7
Performance of SVM classifier for Multiple Sclerosis dataset.

Features	Sensitivity	Specificity	Accuracy
Combined features + PCA (Combined FS)	0.904 ± 0.046	0.908 ± 0.041	0.901 ± 0.044
4 Network features + fold change	0.858 ± 0.020	0.891 ± 0.002	87.8 ± 0.010

Table 8
Performance of SVM classifier for Psoriasis dataset.

Features	Sensitivity	Specificity	Accuracy
Combined features + PCA (Combined FS)	0.928 ± 0.018	0.904 ± 0.027	0.912 ± 0.023
4 Network features + fold change	0.906 ± 0.034	0.874 ± 0.022	0.885 ± 0.018

3.7.1. Case study 1: Network based classification of multiple sclerosis using microarray gene expression data

To investigate the efficiency of this classification approach on gene expression dataset, microarray data (GSE17846) by Keller et al. [67] related to multiple sclerosis has been used. miRNAs that has disease association with multiple sclerosis were identified using miRNA disease association databases [29,30] and mapped to miRNA expression profiles. miRNA differential expression and its fold change were computed using SAM [68]. In total 47 miRNAs that map to known multiple sclerosis miRNAs [29,30] were considered as positive dataset for the classification model. The above multiple sclerosis miRNAs were provided in Supplementary data1, Table S6. Heterogeneous miRNA network for multiple sclerosis was constructed using target genes and PPIs that interact with the miRNA. In total, there were 56,784 interactions between above 41 miRNAs, 5647 target genes and 21,384 PPI pairs.

Discrimination of negative miRNA dataset from the microarray data is quite difficult and hence miRNAs that have lower expression in the dataset has been chosen for the study. In total 47 miRNAs that were not involved with multiple sclerosis disease association were taken as negative dataset and other miRNAs were eliminated. We used the 41 positive samples and 47 negative samples, to train a new model for multiple sclerosis with five-fold cross-validation. Our model generated using NMSSFs features shows an accuracy of 90.1%, achieving a sensitivity of 90.4% and a specificity of 90.8% as shown in Table 7. The corresponding ROC plot was shown in Fig. 9. This result was compared with Network and fold change (NFc) features computed by Xu et al. [26] which achieved an accuracy of 87.8%. The results indicated that combined NMSSFs approach enhances the accuracy of the classifier than that of network and expression based features.

3.7.2. Case study 2: Psoriasis classification using network constructed from differential gene expression of small RNA-seq data

Psoriasis associated RNA-seq data from NCBI SRA database (SRP007825) was chosen as a second case study. The above RNA-seq dataset contains 67 independently sequenced libraries for psoriasis using Illumina GAIIx platform and generated 1.1 billion raw and 670 million qualified reads [69]. This resulted in a set of 512 known, 13 recently described and 47 novel mature miRNAs and miRNAs*. Psoriasis involved miRNAs and normal control datasets were compared and the differentially expressed miRNAs were determined.

A set of known Psoriasis miRNAs from HMDD V2.0 [29] and miR2disease [30] database were taken for the study. Expression values (fold change) of these psoriasis miRNAs were taken as expression based features. Identification of negative dataset for the classifier was achieved by finding the miRNAs with lowest expression based on the percentage of miRNA read counts. In total 57 Psoriasis associated miRNAs used for the classification were provided in Supplementary data, Table S7. Psoriasis miRNA disease network contains 69,213 interactions between 57 miRNAs, 7253 target genes and 27,844 PPI pairs. To construct a more reliable negative dataset, we choose a set of 63 miRNAs with lower expression and trained the classifier.

Evaluation of the constructed model was performed using five-fold cross-validation method. Our model shows an accuracy of 91.2%, achieving a sensitivity of 92.8% and a specificity of 90.4% as shown in Table 8. NFc features (Xu et al.) shows an accuracy

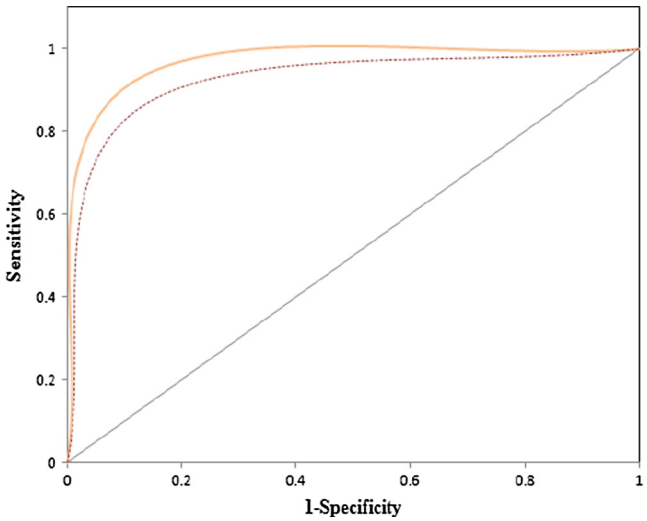


Fig. 10. ROC curve for Psoriasis dataset (Performance evaluation of Psoriais dataset is performed using Combined Fs features (Our approach – solid line) and network + Fold change features(Xu et al. approach – dotted line). Diagonal represents the baseline.

of 88.5%, achieving a sensitivity and specificity of 90.6% and 87.4% respectively. Combined NMSSFs approach enhances the accuracy of the classifier and finally ROC curve plotted for this dataset after performing the cross-validation is shown in Fig. 10 with an AUC of 0.916.

4. Conclusion

In this study, we present an immune miRNA classification system based on novel network and motif features. The average accuracy of prediction algorithm was 90.2%. Network topological features were shown to demonstrate significant differences in the discrimination of immune and non-immune miRNA-disease associations. Further, we also validated our proposed approach with datasets related to gene expression data and RNA-seq data with higher classification accuracy. The future enhancement of the system will incorporate a miRNA prediction model using an ensemble of classifiers to classify miRNAs responsible for immune diseases. This research seeks to develop a new approach to fasten the detection of miRNAs involved in the immune system, which is typically a laborious process when a novel miRNA is sequenced. Thus, future work will focus on confirming the expressions of these miRNAs using experimental methods such as real-time PCR, western blotting, and immunohistochemistry. Network topological features and sub-network patterns are known to demonstrate significant differences in the discrimination of immune and non-immune miRNA associations. The proposed method can serve as an ideal tool to discriminate immune functional miRNAs with increased classification accuracy at a reduced computational cost and time.

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been carried out at Data Mining and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, India. AP would like to thank Indian Council of Medical Research (ICMR) for Senior Research Fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2016.11.003.

References

[1] N. Bushati, S.M. Cohen, MicroRNA functions, Annu. Rev. Cell Dev. Biol. 23 (2007) 175–205.
[2] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, Cell 136 (2) (2009) 215–233.
[3] S. Jia, H. Zhai, M. Zhao, MicroRNAs regulate immune system via multiple targets, Discov. Med. 18 (100) (2014) 237–247.
[4] M. Li, C. Marin-Muller, U. Bharadwaj, K.H. Chow, Q. Yao, C. Chen, MicroRNAs: control and loss of control in human physiology and disease, World J. Surg. 33 (4) (2009) 667–684.
[5] J. Satoh, H. Tabunoki, Comprehensive analysis of human microRNA target networks, BioData Min. 4 (2011) 17.
[6] J.C. Brase, M. Johannes, T. Schlomm, M. Fälth, A. Haese, T. Steuber, T. Beissbarth, R. Kuner, H. Sültmann, Circulating miRNAs are correlated with tumor progression in prostate cancer, Int. J. Cancer 128 (3) (2011) 608–616.
[7] C.M. Groce, Causes and consequences of microRNA dysregulation in cancer, Nat. Rev. Genet. 10 (10) (2009) 704–714.
[8] J. Lu, G. Getz, E.A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B.L. Ebert, R.H. Mak, A.A. Ferrando, J.R. Downing, T. Jacks, H.R. Horvitz, T.R. Golub, MicroRNA expression profiles classify human cancers, Nature 435 (7043) (2005) 834–838.
[9] J. Davidson-Moncada, F.N. Papavasiliou, W. Tam, MicroRNAs of the immune system: roles in inflammation and cancer, Ann. N. Y. Acad. Sci. 1183 (2010) 183–194.
[10] R. Dai, S.A. Ahmed, MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases, Transl. Res. 157 (4) (2011) 163–179.
[11] R.C. Friedman, K.K. Farh, C.B. Burge, D.P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs, Genome Res. 19 (2009) 92–105.
[12] N. Lin, W. Baolin, R. Jansen, M. Gerstein, H. Zhao, Information assessment on predicting protein-protein interactions, BMC Bioinformatics 5 (2004) 154.
[13] A. Casadevall, L.A. Pirofski, Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease, Infect. Immun. 68 (12) (2000) 6511–6518.
[14] H. Liang, W.H. Li, MicroRNA regulation of human protein protein interaction network, RNA 13 (2007) 1402–1408.
[15] B. Liu, L. Liu, A. Tsykin, G.J. Goodall, J.E. Green, M. Zhu, C.H. Kim, J. Li, Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation, Bioinformatics 26 (24) (2011) 3105–3111.
[16] J. Chen, B.J. Aronow, A.G. Jegga, Disease candidate gene identification and prioritization using protein interaction networks, BMC Bioinformatics 10 (2009) 73.
[17] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[18] Z. Lubovac, J. Gamalielsson, B. Olsson, Combining functional and topological properties to identify core modules in protein interaction networks, Proteins 64 (2006) 948–959.
[19] B. Liu, J. Li, M.J. Cairns, Identifying miRNAs, targets and functions, Brief Bioinform. 15 (1) (2014) 1–19.
[20] T.S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, et al., Human protein reference database–2009 update, Nucleic Acids Res. 37 (Database issue) (2009) D767–D772.
[21] C.W. Hsu, H.F. Juan, H.C. Huang, Characterization of microRNA-regulated protein-protein interaction network, Proteomics 8 (2008) 1975–1979.
[22] Zhen Wang, Kan He, Qishan Wang, Yumei Yang, Yuchun Pan, The prediction of the porcine pre-microRNAs in genome-wide based on support vector machine (SVM) and homology searching, BMC Genomics 13 (2012) 729.
[23] M.E. Rahman, R. Islam, S. Islam, S.I. Mondal, M.R. Amin, MiRANN: a reliable approach for improved classification of precursor microRNA using Artificial Neural Network model, Genomics 99 (4) (2012) 189–194.
[24] S. Agarwal, C. Vaz, A. Bhattacharya, A. Srinivasan, Prediction of novel precursor miRNAs using a context-sensitive hidden Markov model (CSHMM), BMC Bioinformatics 11 (1) (2010) S29.
[25] M. Yousef, M. Nebozhyn, H. Shatkay, S. Kanterakis, L.C. Showe, M.K. Showe, Combining multi-species genomic data for microRNA identification using a Naïve Bayes classifier, Bioinformatics 22 (2006) 1325–1334.
[26] J. Xu, C.X. Li, J.Y. Lv, Y.S. Li, Y. Xiao, T.T. Shao, X. Huo, X. Li, Y. Zou, Q.L. Han, X. Li, L.H. Wang, H. Ren, Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer, Mol. Cancer Ther. 10 (10) (2011) 1857–1866.
[27] J.W. Nam, K.R. Shin, J. Han, Y. Lee, V.N. Kim, B.T. Zhang, Human microRNA prediction through a probabilistic co-learning model of sequence and structure, Nucleic Acids Res. 33 (2005) 3570–3581.

[28] P. Jiang, H. Wu, W. Wang, W. Ma, X. Sun, Z. Lu, MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features, Nucleic Acids Res. 35 (2007) W339–W344.
[29] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, Q. Cui, An analysis of human microRNA and disease associations, PLoS One 3 (2008) e3420.
[30] Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang, Y. Liu, miR2Disease: a manually curated database for microRNA deregulation in human disease, Nucleic Acids Res. 37 (Database issue) (2009) D98–D104.
[31] Y. Wang, X. Chen, W. Jiang, L. Li, W. Li, L. Yang, M. Liao, B. Lian, Y. Lv, S. Wang, S. Wang, X. Li, Predicting human microRNA precursors based on an optimized feature subset generated by GA-SVM, Genomics 98 (2) (2011) 73–78.
[32] B. Liu, L. Fang, J. Chen, F. Liu, X. Wang, MiRNA-dis: microRNA precursor identification based on distance structure status pairs, Mol. BioSyst. 11 (4) (2015) 1194–1204.
[33] A. Keller, P. Leidinger, B. Vogel, C. Backes, A. ElSharawy, V. Galata, MiRNAs can be generally associated with human pathologies as exemplified for miR-144*, BMC Med. 12 (2014) 224.
[34] Z. Hu, S.M. Gallo, Identification of interacting transcription factors regulating tissue gene expression in human, BMC Genomics 11 (2010) 49.
[35] T. Lee, J. Kim, Self containment, a property of modular RNA structures, distinguishes microRNAs, PLoS Comput. Biol. 4 (2008) e1000150.
[36] I.L. Hofacker, Vienna RNA secondary structure server, Nucleic Acids Res. 31 (2003) 3429–3431.
[37] D. Betel, M. Wilson, A. Gabow, D.S. Marks, C. Sander, The microRNA.org resource: targets and expression, Nucleic Acids Res. 36 (2008) D149–D153.
[38] A. Krek, D. Grün, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, et al., Combinatorial microRNA target predictions, Nat. Genet. 37 (2005) 495–500.
[39] A. Grimson, K.K. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, D.P. Bartel, MicroRNA targeting specificity in mammals: determinants beyond seed pairing, Mol. Cell 27 (2007) 91–105.
[40] M. Maragkakis, M. Reczko, V.A. Simossis, P. Alexiou, G.L. Papadopoulos, T. Dalamagas, et al., DIANAmicroT web server: elucidating microRNA functions through target prediction, Nucleic Acids Res. 37 (2009) W273–W276.
[41] K.C. Miranda, T. Huynh, Y. Tay, Y.S. Ang, W.L. Tam, A.M. Thomson, B. Lim, I. Rigoutsos, A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes, Cell 126 (2006) 1203–1217.
[42] J. Kruger, M. Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible, Nucleic Acids Res. 34 (2006) W451–W454.
[43] W. Xiaowei, MiRDB: a microRNA target prediction and functional annotation database with a wiki interface, RNA 14 (6) (2008) 1012–1017.
[44] M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, E. Segal, The role of site accessibility in microRNA target recognition, Nat. Genet. 39 (2007) 1278–1284.
[45] C.J. Mattingly, G.T. Colby, J.N. Forrest, J.L. Boyer, The comparative toxicogenomics database (CTD), Environ. Health Perspect. 111 (6) (2003) 793–795.
[46] M.P. Lefranc, V. Giudicelli, C. Ginestoux, J. Bodmer, W. Müller, R. Bontrop, et al., IMGT, the international ImMunoGeneTics database, Nucleic Acids Res. 27 (1) (1999) 209–212.
[47] K. Breuer, A.K. Foroushani, M.R. Laird, C. Chen, A. Sribnaia, R. Lo, G.L. Winsor, R. E. Hancock, F.S. Brinkman, D.J. Lynn, InnateDB: systems biology of innate immunity and beyond – recent updates and continuing curation, Nucleic Acids Res. 41 (Database issue) (2013) D1228–D1233.
[48] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[49] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.
[50] N.H. Tran, K.P. Choi, L. Zhang, Counting motifs in the human interactome, Nat. Commun. 4 (2013) 2241.
[51] S. Wernicke, F. Rasche, FANMOD: a tool for fast network motif detection, Bioinformatics 22 (9) (2006) 1152–1153.
[52] A. Kozomara, S. Griffiths-Jones, MiRBase: annotating high confidence microRNAs using deep sequencing data, Nucleic Acids Res. 42 (2014) D68–D73.
[53] S. Okuda, T. Yamada, M. Hamajima, M. Itoh, T. Katayama, P. Bork, S. Goto, M. Kanehisa, KEGG Atlas mapping for global analysis of metabolic pathways, Nucleic Acids Res. 36 (2008) W423–W426.
[54] M.A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, et al., The Gene Ontology (GO) database and informatics resource, Nucleic Acids Res. 32 (Database issue) (2004) D258–D261.
[55] H. Dweep, N. Gretz, miRWalk2.0: a comprehensive atlas of microRNA-target interactions, Nat. Methods 12 (8) (2015) 697.
[56] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and other Kernel-based Learning Methods, first ed., Cambridge University Press, 2000.
[57] T. Joachims, Making Large-Scale SVM Learning Practical. Advances in Kernel Methods – Support Vector Learning, MIT Press, 1999, pp. 169–184.
[58] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer Series in Statistics; Section 4.3, second ed., Springer, New York, NY, USA, 2008.
[59] L. Popelinsky, Combining the principal components method with different learning algorithms, in: Proceedings of the ECML/PKDD2001 IDDM Workshop, 2001.
[60] H. Ahmadi, A. Ahmadi, S. Azimzadeh-Jamalkandi, M.A. Shoorehdeli, A. Salehzadeh-Yazdi, G. Bidkhori, A. Masoudi-Nejad, HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens, Genomics 101 (2) (2013) 94–100.
[61] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng, Enhanced performance of brain tumor classification via tumor region augmentation and partition, PLoS One 10 (10) (2015) e0140381.
[62] A. Prabahar, J. Natarajan, Prediction of miRNA in human MHC that encodes different immunological functions using support vector machines, Curr. Bioinform. 9 (2014) 343–347.
[63] L. Mickael, B.D. Abdoulaye, B. Mathieu, Computational prediction of the localization of microRNAs within their pre-miRNA, Nucleic Acids Res. 41 (15) (2013) 7200–7211.
[64] A.L. Barabási, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2004) 101–113.
[65] R. Albert, Scale-free networks in cell biology, J. Cell Sci. 11 (8) (2005) 4947–4957.
[66] L. Chen, X. Qu, M. Cao, Y. Zhou, W. Li, B. Liang, W. Li, W. He, C. Feng, X. Jia, Y. He, Identification of breast cancer patients based on human signaling network motifs, Sci. Rep. 3 (2013) 3368.
[67] A. Keller, P. Leidinger, J. Lange, A. Borries, H. Schroers, M. Scheffler, H.P. Lenhof, K. Ruprecht, E. Meese, Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls, PLoS One 4 (10) (2009) e7440.
[68] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map (SAM) format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[69] C.E. Joyce, X. Zhou, J. Xia, C. Ryan, B. Thrash, A. Menter, W. Zhang, A.M. Bowcock, Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome, Hum. Mol. Genet. 20 (2011) 4025–4040.