IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 29, 2019, accepted May 19, 2019, date of publication June 5, 2019, date of current version July 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920942

Predicting CircRNA-Disease Associations Through Linear Neighborhood Label Propagation Method

WEN ZHANG[1], CHENGLIN YU[2], XIAOCHAN WANG[2], AND FENG LIU[2]
[1]College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[2]School of Computer Science, Wuhan University, Wuhan 430072, China

Corresponding authors: Wen Zhang (zhangwen@mail.hzau.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772381 and Grant 61572368, in part by the Fundamental Research Funds for the Central Universities under Grant 2042017kf0219 and Grant 2042018kf0249, in part by the National Key Research and Development Program under Grant 2018YFC0407904, and in part by the Huazhong Agricultural University Scientific and Technological Self-innovation Foundation.

ABSTRACT Identification of circRNA-disease associations provides insight into the mechanism that circRNAs cause diseases. Wet experimental identification of circRNA-disease associations is time-consuming and labor-intensive, and thus developing computational methods for the circRNA-disease association prediction is an urgent task. In this paper, we propose a linear neighborhood label propagation method to predict circRNA-disease associations, named CD-LNLP. First, CD-LNLP uses association profiles based on known associations to calculate circRNA-circRNA similarities and disease-disease similarities. Next, CD-LNLP implements the label propagation based on the circRNA-circRNA similarity-based graph and the disease-disease similarity-based graph respectively to predict circRNA-disease associations. Finally, we combine the outputs from circRNA-circRNA similarity-based graph model and disease-disease similarity-based graph model to produce the results. In the experiments, CD-LNLP achieves impressive performance with the AUPR score of 0.4487 and the AUC score of 0.9007 and outperforms outstanding baseline methods (collaborative filter method, KATZ, nonnegative matrix factorization method, and resource allocation method) and the state-of-the-art method MRLDC. The case studies show that CD-LNLP identifies novel circRNA-disease associations, which are validated by up-to-date circRNA-disease databases and literature respectively. In conclusion, CD-LNLP is a promising method for predicting circRNA-disease associations.

INDEX TERMS CircRNA-disease associations, association profiles, linear neighborhood similarity, label propagation.

I. INTRODUCTION

Circular RNA (circRNA) is a novel type of endogenous non-coding RNAs (ncRNA) [1]. Different from the linear RNAs, circRNAs are generated by back splicing or lariat introns. Thus, they don't have 5' and 3' ends which reflect start and stop of the RNA polymerase on the DNA template [2]–[5].

The first circRNA was discovered in a study of RNA viruses [6] in 1976. Due to the structural specificity, unknown function and low abundance of circRNAs, circRNAs were initially assumed as artefacts or mis-splicing products. Thus circRNAs did not attract much attention [2]. In recent years, more and more circRNAs have been identified in thousands of living organisms, including archaea, plants and animals [5], [7]–[9]. At the same time, researchers have studied circRNAs to obtain their knowledge [10], [11]. To date, circRNAs have been discovered to be expressed in different tissues and implicated in many cellular processes [5], [12], including cell proliferation, invasion, and apoptosis [13], [14]. CircRNAs play critical roles in biological processes including transcription, mRNA splicing [15], [16], RNA decay and translation [17]. The misregulation of circRNAs may cause abnormal cellular functions and growth defects, and several circRNAs have been reported to be associated with human diseases. For example, Circ-FBXW7 is reduced in glioblastoma clinical samples compared with their paired tumor-adjacent tissues, and circ-FBXW7 expression is also positively associated with glioblastoma patients overall survival [18]. CircPVT1 is significantly up-regulated in the osteosarcoma tissues, and circPVT1 may be a biomarker for the diagnosis of osteosarcoma [19]. Hsa_circ_0081001 is reported correlated

The associate editor coordinating the review of this manuscript and approving it for publication was Yungang Zhu.

with poor prognosis, and its expression level may dynamically monitor the condition changes of osteosarcoma [20]. Has_circ_001569 is highly expressed in cell proliferation and invasion of colorectal cancer compared with non-cancerous samples [21]. Has_circ_0054633 is found to be upregulated in the peripheral blood of patients with type 2 diabetes, and it can be used as a candidate biomarker for the diagnosis for pre-diabetes and type 2 diabetes [22]. Clearly, exploring the circRNA-disease associations can contribute to deciphering the cellular behaviors underlying diseases.

In recent years, a lot of databases about circRNAs have also been developed to further study the function mechanism of circRNAs, such as CircBase [23], CircRNADb [24], PlantcircBase [8], PlantCircNet [25], ExoRBase [26], CircNet [27], TSTD [28], SomamiR 2.0 database [29] and CSCD [30]. In addition, there are several databases for circRNA-disease associations, including Circ2Disease [31], circRNADisease [32] and CircR2Disease [33]. Inspired by these data sources, we think it is necessary to develop computational methods for the circRNA-disease association prediction. Lei *et al.* [34] proposed a computational path weighted method PWCDA by integrating diseases' functional similarities and circRNAs' semantic similarities. Yan *et al.* [35] developed a method called DWNN-RLS to predict circRNA-disease associations based on Regularized Least Squares of Kronecker product kernel. Fan *et al.* [36] constructed a heterogeneous network by employing the circRNA and disease expression profiles, and then developed a computational model based on KATZ measure called KATZHCDA. Xiao *et al.* [37] developed a weighted low-rank approximation optimization algorithm with dual-manifold regularizations for predicting disease-associated circRNAs, named MRLDC. These methods mainly utilize either circRNA features or disease features, and known associations to predict novel circRNA-disease associations. However, circRNA features and disease features are not always available. Thus, those methods can't work when information is incomplete.

In this paper, we propose a linear neighborhood label propagation method to predict circRNA-disease associations, named CD-LNLP. First, CD-LNLP uses association profiles based on known associations to calculate circRNA-circRNA similarities and disease-disease similarities. Next, CD-LNLP implements the label propagation based on the circRNA-circRNA similarity-based graph and the disease-disease similarity-based graph respectively to predict circRNA-disease associations. Finally, we combine the outputs from circRNA-circRNA similarity-based graph model and disease-disease similarity-based graph model to produce the results. In the experiments, CD-LNLP achieves impressive performance with an AUPR of 0.4487 and an AUC of 0.9007, and outperforms outstanding baseline methods (collaborative filter method, KATZ, nonnegative matrix factorization method and resource allocation method) and the state-of-the-art method MRLDC. The case studies show that CD-LNLP identifies novel circRNA-disease associations, which are validated by up-to-date circRNA-disease databases and literature respectively. In conclusion, CD-LNLP is promising for predicting circRNA-disease associations.

TABLE 1. Details of datasets.

Datasets	No. of circRNAs	No. of diseases	No. of associations
Dataset1	312	40	331
Dataset2	603	88	650

II. MATERIALS AND METHODS

A. DATASETS

Recently, researchers collected data about circRNA-disease associations, and constructed databases, including Circ2Disease, circRNADisease and CircR2Disease.

Circ2Disease [31] is a database that manually curated experiment-supported human circRNAs and provides associations between circRNAs and human diseases. This database totally contains 273 associations between 237 circRNAs and 54 human diseases that had been recorded in the existing literature of PubMed prior to 1 November 2017. In addition, Circ2Disease also integrated experimentally verified miRNAs and miRNA targets from several databases, such as HMDD v2.0, OncomiRDB, miRTarBase, etc. CircRNADisease [32] is a manually curated database of experimentally supported circRNA and disease associations, which contains 354 high confident experimentally supported circRNA-disease associations between 330 circRNAs and 48 diseases that had been recorded in PubMed database before November 2017 from the National Center for Biotechnology Information. CircR2Disease [33] contains 661 circRNAs annotated with 725 relations to 100 diseases, which were manually curated from existing literature prior to 31 March 2018.

All these circRNA-disease databases contributed to the study on the roles of circRNAs in diseases. We use two datasets to evaluate the performance of our proposed method CD-LNLP. First, we collect a total of 354 experimentally verified associations between 330 circRNAs and 48 diseases from circRNADisease database, and then we remove duplicate associations and associations related to other species except human, and eventually retain 331 associations between 312 circRNAs and 40 diseases for humans. We name the dataset as Dataset1 and we notice the same dataset was used by Xiao *et al.* [37]. Second, we collect circRNA-disease associations from CircR2Disease, and obtain 650 associations between 603 circRNAs and 88 diseases after removing duplicated associations. We name the dataset as Dataset2. The statistics of the two datasets are shown in Table 1.

B. PROBLEM DESCRIPTION

Given a set of circRNAs $R = \{R_1, R_2, \ldots, R_m\}$ and a set of diseases $D = \{D_1, D_2, \ldots, D_n\}$, known circRNA-disease associations can be formulated as a bipartite network, which regards circRNAs, diseases as nodes and regards associations as edges. The bipartite network is represented by an

adjacency matrix A. $A(i,j) = 1$, if there is an association between circRNA R_i and disease D_j; otherwise, $A(i,j) = 0$. This work is to predict undiscovered circRNA-disease associations between these circRNAs and diseases.

Here, we introduce the association profiles of circRNAs and diseases for the sake of the following content. The association profile of a circRNA is a binary vector describing the presence or absence of association with every disease in the network. The association profile of circRNA R_i is the ith row of the circRNA-disease adjacency matrix A, i.e. $A(i,:)$. Similarly, the association profile of the disease D_j is the jth column of the circRNA-disease adjacency matrix A, i.e. $A(:,j)$.

C. LINEAR NEIGHBORHOOD SIMILARITY

After represented by association profiles, two circRNAs (diseases) can be taken as data points in the feature spaces, and there are several popular similarity measures for pairwise data points, i.e. Jaccard similarity, cosine similarity and Gaussian similarity. In our previous works, we presented the linear neighborhood similarity (LNS) measure to calculate similarity between two data points, and successfully applied it to lots of bioinformatics tasks [38]–[40], and results demonstrated that LNS can achieve superior performances compared with other similarity measures. Therefore, we adopt LNS to calculate circRNA-circRNA similarity based on their association profiles.

The calculation of association profile similarity is briefly introduced as follows. We take circRNAs as data points in the association space according to their association profiles, denoted by $X_i, i = 1, 2, \ldots, m$. Each circRNA can be represented as a vector (i.e. "association profile") in the association space, which corresponds to a row of the observed association matrix A. We calculate the Euclid distances between circRNAs' association profiles, and define K nearest neighbors of X_i as $N(X_i)$. We present the optimization problem as the following objective function:

$$\min_{w} \frac{1}{2}\left\|A-(C\odot W)A\right\|_F^2+\frac{\mu}{2}\sum_{i=1}^{n}\left\|(C\odot W)_{i\cdot}\right\|_1^2$$
$$s.t.\ (C\odot W)e=e,\quad W\geq 0 \tag{1}$$

where $\odot$ is the Hadamard product; $\|\cdot\|_F$ is the Frobenius norm of matrix; $\|\cdot\|_1$ is the 1-norm of vector. $(C\odot W)_{i\cdot}$ is the ith row of $C\odot W$, and e is a $n\times 1$ vector with all values equal to 1. $C\in\mathbb{R}^{n\times n}$ is an indicator matrix. $c_{ij}=1$ if $x_j\in N(X_i)$; otherwise, $c_{ij}=0$. By considering n rows at the same time, $\sum_{i=1}^{n}\left\|(C\odot W)_{i\cdot}\right\|_1^2$ in (1) can be rewritten as $\|(C\odot W)e\|_2^2$.

As discussed in [41], W reflects the intrinsic structure of known associations. To calculate W in (1), we introduce the Lagrange function :

$$L=\frac{1}{2}\left\|A-(C\odot W)A\right\|_F^2+\frac{\mu}{2}\left\|(C\odot W)e\right\|_2^2 -\lambda^T\left((C\odot W)e-e\right)-tr\left(\Phi W\right) \tag{2}$$

where tr is the trace of a matrix and Φ is Lagrange multiplier. Differentiating L with respect to W:

$$\nabla_W L=C\odot\left((C\odot W)AA^T+\mu(C\odot W)ee^T-AA^T-\lambda e^T\right)-\Phi^T \tag{3}$$

According to the complementary slackness condition, $\Phi_{ij}W_{ij}=0$, and we have:

$$C_{ij}\left((C\odot W)AA^T+\mu(C\odot W)ee^T-AA^T-\lambda e^T\right)_{ij}W_{ij}=0 \tag{4}$$

Thus, we obtain the update rule for W:

$$W_{ij}=\begin{cases}W_{ij}\dfrac{\left(AA^T+\lambda ee^T\right)_{ij}}{\left((C\odot W)AA^T+\mu(C\odot W)ee^T\right)_{ij}} & i\neq j\\ 0 & i=j\end{cases} \tag{5}$$

where λ is the regularization coefficient. We set λ to 1 for simplicity.

Thus, we can obtain a similarity matrix $W_{Cap}\in R^{m\times m}$ for m circRNAs. Similarly, we can obtain the association profiles of diseases, and calculate pairwise similarities between n diseases denoted by a similarity matrix.

D. LABEL PROPAGATION

After we calculate circRNA-circRNA similarity and disease-disease similarity, we use a label propagation process [41] to predict unobserved circRNA-disease associations.

First, we construct a directed graph which uses circRNAs as nodes and use $W_{Cap}(i,j)$ as the weight of edge linking circRNA R_i to circRNA R_j. The known circRNA-disease associations are considered as labels, which are propagated in the circRNA similarity graph. In each step, the labels of nodes are updated by absorbing label information from their neighborhoods with a rate of α and retaining their initial labels with a rate of $1-\alpha$. Let $Y_i^t=\{y_{1i}^t, y_{2i}^t, \ldots, y_{ni}^t\}$ denote predicted association scores of the ith disease at time t, where y_{ji}^t measures the propensities of disease D_i being associated with circRNA R_j. When $t=0$, Y_i^0 is the association profile of disease D_i. The propagation process can be written as:

$$Y_i^{t+1}=\alpha WY_i^t+(1-\alpha)Y_i^0 \tag{6}$$

It can be inferred that:

$$\lim_{t\to\infty}Y_i^t=\lim_{t\to\infty}(\alpha W)^{t-1}Y_i^0+(1-\alpha)\sum_{i=0}^{t-1}(\alpha W)^i Y_i^0 =(1-\alpha)(I-\alpha W)^{-1}Y_i^0 \tag{7}$$

Then, Y_i^t will converge to:

$$Y_i=(1-\alpha)(I-\alpha W)^{-1}Y_i^0 \tag{8}$$

where Y_i is the final circRNA-disease association scores. Taking all diseases at the same time, let $Y^t=\{Y_1^t, Y_2^t, \ldots, Y_n^t\}$, we can write above process in matrix form as:

$$Y=(1-\alpha)(I-\alpha W)^{-1}Y^0 \tag{9}$$

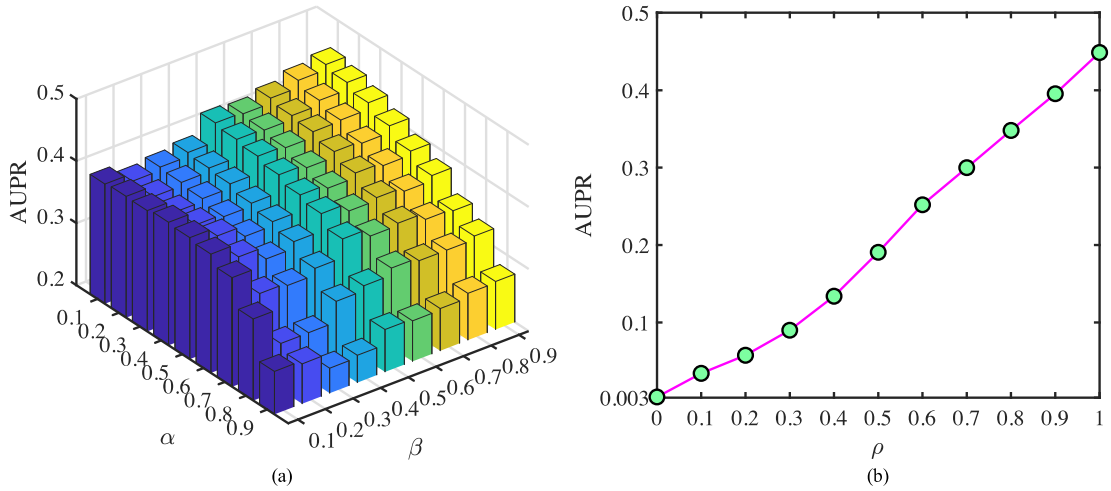


FIGURE 1. The influence of parameters on CD-LNLP (a) Effect of parameter α and β when ρ was set 1.0. (b) The influence of ρ.

The predicted matrix based on the circRNA similarity graph is denoted by $Y_{circRNA}$. In the same manner, we built the label propagation model based on the disease similarity graph, and the predicted matrix is denoted by $Y_{disease}$. We assign different weights to each prediction model and calculate the final integrated scores as follows:

$$Y_{intergration} = \rho Y_{circRNA} + (1 - \rho) Y_{disease} \tag{10}$$

where ρ is the trade-off (weight) between circRNA-based prediction and disease-based prediction.

III. EXPERIMENTS AND RESUTLS

A. EVALUATION METRICS

We adopt leave-one-out cross-validation (LOOCV) to evaluate the performance of prediction models. In LOOCV, each circRNA-disease pair is left out in turn as the testing sample, and other circRNA-disease pairs are used as the training set. In each fold, we construct prediction models based on the training set and then score the testing one. We repeat the training process and testing process until we have prediction scores for all pairs.

In addition, we adopt several evaluation metrics to evaluate performances of prediction models, i.e. the area under receiver-operating characteristic curve (AUC), the area under precision-recall curve (AUPR), sensitivity (SEN), specificity (SPEC), precision (PRE), accuracy (ACC) and F-measure (F1). The area under receiver-operating characteristic curve (AUC) is to evaluate the prediction performance of a model by considering the true positive rate and the false positive rate over different thresholds. The area under precision-recall curve (AUPR) considers the recall and precision over different thresholds. Since known circRNA-disease associations are much fewer than unknown circRNA-disease associations, we adopt AUPR as the primary metric. The experiments are conducted on Dataset1 in python 3.6 under macOS.

B. PARAMETER SETTINGS

CD-LNLP has three parameters: α, K and ρ. α is the possibility of receiving label information from neighbors, K is the number of neighbors, and ρ is the trade-off (weight) for circRNAs. In our experiments, we considered the combinations of following values for different parameters: $\{0.1, 0.2, 0.3, \ldots, 0.9\}$ for α, $\{10\%, 20\%, 30\%, \ldots, 90\%\}$ of the number of overall data points for K and $\{0.0, 0.1, 0.2, 0.3, \ldots, 0.9, 1.0\}$ for ρ. Here, we denote the proportion for K as β. Then we considered all combinations of different parameter values to build CD-LNLP models, and implemented LOOCV on Dataset1 to evaluate the influence of parameters on CD-LNLP.

In computational experiments, CD-LNLP produced the best AUPR value when $\alpha = 0.1$, $\beta = 0.9$, and $\rho = 1.0$. Then, we fixed the weight parameter ρ, and evaluated the influence of parameters α and β, and the results are shown in Fig. 1 (a). Clearly, α and β have a great impact in the model. For any β, when α is greater than 0.7, the performances fall rapidly. For any α, when β is greater than 0.5, the performances don't change too much. Further, we fixed the parameters $\alpha = 0.1$, and $\beta = 0.9$, and observed the influence of the weight ρ. The different values for ρ and AUPR values are shown in Fig. 1(b). Clearly, performances of CD-LNLP will increase as ρ increases.

C. COMPARISON WITH OTHER METHODS

In order to demonstrate the superior performance of CD-LNLP, we compared it with several baseline methods and state-of-the-art circRNA-disease association prediction methods. Here, we consider the baseline methods: collaborative filter (CF), KATZ, nonnegative matrix factorization (NMF), resource allocation (RA).

Collaborative Filter (CF) is a classic method for recommender system [42] and it has been adopted to solve a lot of bioinformatics problems. CF calculates the prediction score

between R_i and D_j as:

$$A_p(i,j) = \frac{\sum_{k=1,k\neq i}^{m} W_{ik} A(k,j)}{\sum_{k=1,k\neq i}^{m} W_{ik}} \quad (11)$$

where W_{ik} is the similarity between circRNA R_i and circRNA R_k. A is the association matrix between circRNAs and diseases. m is the number of circRNAs as mentioned before. W_{ik} is the Gaussian association profile similarity [43] between circRNAs.

KATZ model was initially used to predict associations in a social network, and it has also been successfully used to predict lncRNA-disease associations [44], disease-gene associations [45], microbe-disease associations [46] as well as miRNA-disease associations [47]. We constructed a heterogeneous network with the adjacency matrix:

$$A^* = \begin{bmatrix} W_{Cg} & A \\ A^T & W_{Dg} \end{bmatrix} \quad (12)$$

where W_{Cg} is the circRNA Gaussian association profile similarity matrix, W_{Dg} is the disease Gaussian association profile similarity matrix, A is the association matrix and A^T is the transpose of A. We apply KATZ to the heterogeneous network, and obtain a $(m+n)\times(m+n)$ matrix S^{Katz}:

$$S^{Katz} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \left(I - \eta A^*\right)^{-1} - I \quad (13)$$

where η is a free parameter (i.e. the damping factor) controlling the path weights, and we assign $\eta = 0.005$. The submatrix S_{12} is the prediction matrix for circRNAs and diseases.

Nonnegative matrix factorization (NMF) has also been successfully applied to a lot of bioinformatics problems, such as disease-miRNA prediction [48], microRNA-disease prediction [49] and so on. NMF is to factorize the circRNA-disease association matrix A into two low-rank feature matrices $X \in R^{m\times k}$ and $Y \in R^{n\times k}$, where k is the dimension of the low-rank spaces. The prediction matrix for circRNAs and diseases is calculated as $A_p = XY^T$.

The resource allocation (RA) [50] method is a graph-based inference method, which has successfully solved lots of association prediction problems [51]–[54]. The prediction score between circRNA R_i and disease D_j is

$$A_p(i,u) = \sum_{l=1}^{m} \frac{A(l,u)}{K(R_l)} \sum_{j=1}^{n} \frac{A(l,j) f_0\left(D_j\right)}{K\left(D_j\right)}, \quad j \in \{1,2,\ldots,n\}, l \in \{1,2,\ldots,m\} \quad (14)$$

where $K(R_l) = \sum_{u=1}^{n} A(l,u), u \in \{1,2,\ldots,n\}$ denotes the number of the diseases that are associated with R_l, and $K\left(D_j\right) = \sum_{k=1}^{m} A(k,j), k \in \{1,2,\ldots,m\}$ represents the number of the circRNAs that are associated with D_j.

To the best of our knowledge, several methods have been proposed to predict circRNA-disease associations, i.e. PWCDA [34], DWNN-RLS [35], KATZHCDA [36], and MRLDC [37]. The first three methods all integrate multi-data sources (e.g., circRNAs' related gene targets, genes' related annotation terms, expression profiles of circRNAs and MeSH descriptions of diseases' descriptors) to predict circRNA-disease associations. However, those features about circRNAs and diseases are not always available. The main goal of our study is to infer new circRNA-disease associations based only on the circRNA-disease association network. For this reason, we only replicate the MRLDC method, which prioritizes disease-associated circRNAs only based on the known associations and reports good performances, and adopt MRLDC as a benchmark method for comparison.

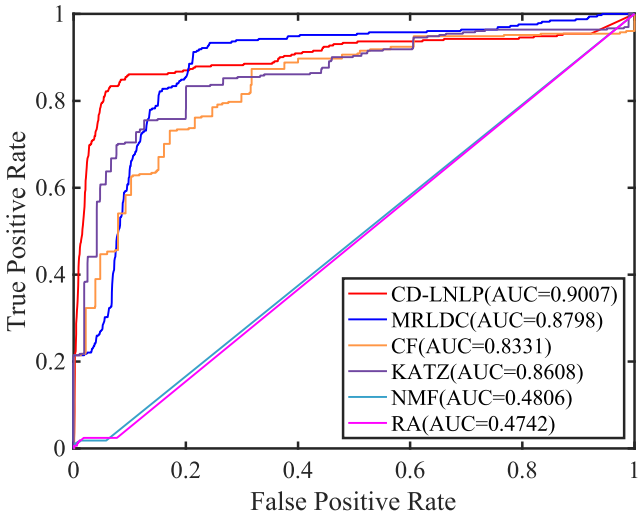


FIGURE 2. AUCs of different methods evaluated by LOOCV.

All models are evaluated based on Dataset1 by using LOOCV. As shown in Fig. 2, our method produces greater AUC scores than other methods. In addition, we also calculate AUPRs and other metrics of different methods. As shown in Table 2, our method achieves an AUPR score of 0.4487, which is higher than AUPR scores of other methods under the same condition. Furthermore, the results in Table 2 demonstrate our method has good performances in terms of different evaluation metrics.

D. INFLUENCE OF DATA RICHNESS

In our task, known circRNA-disease associations are the only information source for model construction, and the number of known circRNA-disease associations, i.e., data richness, may influence performances of our method CD-LNLP. Here, we randomly remove 5%, 10%, 15%, 20%, 25% and 30% known circRNA-disease associations from Dataset1, and implement LOOCV to evaluate CD-LNLP models by using datasets which have fewer associations.

As shown in Table 3, data richness greatly influences the performances of CD-LNLP models. Removing known associations can decrease the data richness, and the AUC score and AUPR score of CD-LNLP model decrease. In general, 5% decrease of known associations may lead to around 19.6% decrease of the AUPR score and 1.4% decrease of the AUC score. Therefore, CD-LNLP predicts unobserved associations when most associations are known.

TABLE 2. Performances of different prediction methods evaluated by LOOCV.

Methods \ Metrics	AUPR	AUC	F1	ACC	REC	SPEC	PRE
CD-LNLP	0.4487	0.9007	0.5200	0.9673	0.6677	0.9755	0.4258
MRLDC	0.3057	0.8798	0.3498	0.9788	0.2145	0.9997	0.9467
CF	0.1333	0.8331	0.3430	0.9782	0.2145	0.9990	0.8554
KATZ	0.3647	0.8608	0.3738	0.9608	0.4411	0.9750	0.3244
NMF	0.0363	0.4806	0.0517	0.0265	1.0000	0.0000	0.0265
RA	0.0316	0.4742	0.0517	0.0265	1.0000	0.0000	0.0265

TABLE 3. Performances of CD-LNLP model based on fewer associations.

Ratio \ Metrics	AUPR	AUC	F1	ACC	REC	SPEC	PRE
5%	0.3608	0.8881	0.4399	0.9712	0.4476	0.9848	0.4325
10%	0.2866	0.8753	0.3906	0.9690	0.4161	0.9825	0.3680
15%	0.2651	0.8662	0.3674	0.9658	0.4397	0.9779	0.3155
20%	0.2167	0.8622	0.3284	0.9708	0.3358	0.9846	0.3213
25%	0.1827	0.8472	0.3040	0.9695	0.3333	0.9825	0.2795
30%	0.1444	0.8450	0.2633	0.9677	0.3103	0.9802	0.2286

E. CASE STUDIES

In order to test the performance of our method in predicting new circRNA-disease associations, we use case studies to examine the performances of our method. We use all the experimentally verified associations as training samples and use those unobserved circRNA-disease associations as candidate associations. For a specific disease, we rank those candidate circRNAs based on their prediction scores.

As discussed in section II-A, associations in Dataset1 are extracted from circRNADisease database, which was established in November 2017. Dataset2 is obtained from CircR2Disease established in March 2018 and contains many newly discovered human circRNA-disease associations. Therefore, we predict novel circRNA-disease associations based on all known associations in Dataset1, and then confirm our findings in Dataset2.

First, we build the CD-LNLP model and the MRLDC model, the CF model, the KATZ model, the NMF model and the RA model by using all associations in Dataset1, and then check up on the predictions in Dataset2. We analyze the top 30 associated circRNA candidates for every disease predicted by our method (CD-LNLP) and other methods, and demonstrate the overall numbers of confirmed associations for all methods in Fig. 3. Clearly, CD-LNLP successfully identifies 5 associations among top 30 predicted associations for all diseases, and produce better results than compared methods: MRLDC (1), CF (4), KATZ (2), NMF (1) and RA (2). Further, we investigate into the correctly identified circRNA-disease associations, and list the rank of these associations in the predictions yielded by other methods in Table 4. We observe

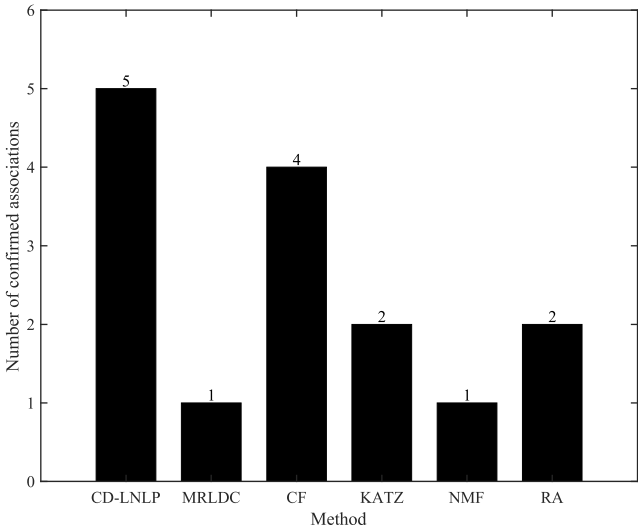


FIGURE 3. Numbers of confirmed circRNA-disease associations for different methods.

that predicted associations have low rank in the predictions of other methods, and it indicates that our method can discover novel associations ignored by other methods. We have observed from the results that NMF or RA produces the scores of 0 for many circRNA-disease pairs. As discussed in section II-A, Dataset1 contains 331 associations between 312 circRNAs and 40 diseases. The possible reason is that the circRNA-disease network is much too sparse, and NMF and RA lack capability of dealing with circRNAs or diseases without any association.

TABLE 4. Novel confirmed associations predicted by CD-LNLP for all diseases and their rank in the prediction of benchmark methods.

disease	circRNA	CD-LNLP	MRLDC	CF	KATZ	NMF	RA
acute myeloid leukemia	hsa_circ_0004277	16	60	14	81	> 8	>8
gastric cancer	circHIAT1	8	172	1	223	>75	>75
hepatocellular carcinoma	hsa_circ_0067934	1	53	1	78	>11	1
hepatocellular carcinoma	circSMARCA5	25	229	14	146	>11	>19
osteoarthritis	circRNA_100876	3	287	49	72	N.A.	N.A.
average		10.6	160.2	15.8	120	N.A.	N.A.

(1)"> n" means that only n circRNA candidates for the disease have scores and the corresponding circRNA ranks behind those circRNA candidates that have scores. (2)"N.A." (i.e., not available) means that the corresponding disease's related circRNA candidates' scores are all zeros.

TABLE 5. Top 10 circRNA-disease associations predicted by CD-LNLP.

Rank	circRNA	Disease	Evidence
1	circATP2B1/hsa_circ_000826/hsa_circ_0000426	Gastric cancer	N.A.
2	hsa_circ_0003575	Atherosclerotic vascular disease	N.A.
3	hsa_circ_0018508	Coronary artery disease	N.A.
4	CDR1as/ciRS-7/hsa_circ_0001946	Glioma	[10], [55]
5	CDR1as/ciRS-7/hsa_circ_0001946	Cholangiocarcinoma	[56]
6	hsa_circ_0001649	Glioblastoma	N.A.
7	hsa_circ_0005870	Diabetic retinopathy	N.A.
8	hsa_circ_0005870	Coronary artery disease	N.A.
9	hsa_circ_0001649	Alzheimer's disease	N.A.
10	hsa_circ_0068087	Coronary artery disease	N.A.

N.A. not available

TABLE 6. Top 10 predictions for the disease "Glioma" and "Colorectal cancer".

Disease	Rank	CircRNA	Source
Glioma	1	CDR1as/ciRS-7/hsa_circ_0001946	[62]
	2	hsa_circ_0000284/circHIPK3	[63]
	3	hsa_circ_0001313/circCCDC66	N.A.
	4	Cir-ITCH/hsa_circ_0001141/hsa_circ_001763	[64]
	5	hsa_circ_0014717	N.A.
	6	hsa_circ_0007031	N.A.
	7	hsa_circRNA_104700	N.A.
	8	hsa_circ_001988/hsa_circ_0001451	N.A.
	9	hsa_circ_0084021	N.A.
	10	hsa_circRNA_103809	N.A.
Colorectal cancer	1	hsa_circ_0067934	N.A.
	2	circular RNA100783/hsa_circ_0008887	N.A.
	3	Circ_FKBP8	N.A.
	4	circR-284	N.A.
	5	Circ-ZNF609/hsa_circ_0000615	[59]
	6	hsa-circRNA 2149	N.A.
	7	Titin circRNAs/cTtns/RBM20-dependent TTN circRNAs	N.A.
	8	Circ_HIPK3	[65]
	9	hsa_circ_0021001	N.A.
	10	Circ_CRKL	N.A.

N.A. not available

Further, we build CD-LNLP model based on Dataset2 to test its prediction ability for unseen data. All known circRNA-disease associations on Dataset2 are used to train the CD-LNLP model, and all other circRNA-disease pairs are used as the candidate circRNA-disease associations for prediction. Since all circRNA-disease associations have been used to build models, the predicted associations should be validated by the publicly available

literature. The top 10 predicted circRNA-disease associations are listed in Table 5, and we can find evidence to confirm two associations. For example, Through CDR1as [55], excessive expression of miR-671-5p markedly increases migration and proliferation rates in glioblastoma multiforme cells. The expression of CDR1as is upregulated in cholangiocarcinoma and might promote the carcinogenesis. In addition, the CDR1as expression could be considered as an independent prognostic biomarker for cholangiocarcinoma with acceptable sensitivity and specificity [56].

Moreover, we select two diseases of wide interests: Glioma and Colorectal cancer for analysis, and respectively predict circRNAs associated with them. A glioma is a type of tumor that starts in the glial cells of the brain or the spine [57]. Gliomas comprise about 30 percent of all brain tumors and central nervous system tumors, and 80 percent of all malignant brain tumors [58]. Colorectal cancer is one of the most common malignant tumors. In 2012, there were 1.4 million new cases and 693,900 death cases of colorectal cancer worldwide [59], and colorectal cancer is the third leading cause of cancer death in the United States [60], [61]. Table 6 shows the top 10 predicted circRNAs associated with the two diseases. In general, three circRNAs are proved to be associated with glioma, and two circRNAs are proved to be associated with colorectal cancer. For example, Cdr1-as is identified as downstream miR-671-5p targets in human glioblastoma multiforme, and expression of Cdr1-as significantly decreases in glioblastoma multiforme biopsies [62]. CircRNA circHIPK3 promotes glioma progression through targeting miR-654 from IGF2BP3 and circHIPK3 might be a potential target for glioma therapy [63]. Cir-ITCH is a tumor-suppressor gene in glioma and may serve as a promising prognostic biomarker for glioma patients, and restoration of Cir-ITCH expression could be a future direction to develop a novel treatment strategy [64]. For colorectal cancer, Circ-ZNF609 promotes migration of colorectal cancer by inhibiting Gli1 expression via microRNA-150 [59]. CircHIPK3 promotes colorectal cancer growth and metastasis by sponging miR-7. Co-expressing miR-7 along with a circHIPK3 inhibitor may be a promising treatment approach for patients with colorectal cancer [65]. Although there remain many predictions which have not been confirmed, they may be proved by future studies.

IV. CONCLUSION

CD-LNLP is a novel method for predicting circRNA-disease associations. The experimental results and case studies demonstrate that CD-LNLP has the high-accuracy performances and outperforms other state-of-the-art methods. The good performances of CD-LNLP are mainly attributed to the following factors. First, the application of linear neighbor similarity (LNS) guarantees the basic effectiveness of our proposed method. The LNS measure has shown good performance in many aforementioned bioinformatics problems, such as drug-disease association prediction [41], drug-disease therapeutic function prediction [39], microRNA-disease association prediction [66] and so on. Second, since not all biological information about circRNAs and diseases is available, CD-LNLP only uses the known associations, which is reliable.

However, there still exist some limitations in CD-LNLP. First, the model only utilizes the known circRNA-disease associations as prior information, and is not applicable to new circRNAs or diseases without any known association. If we have more associations between circRNAs and diseases, the model CD-LNLP would obtain better performance. Second, circRNAs can function as miRNA sponges or decays, protein sponges or decoys. In this study, we only adopt the linear neighbor similarity based on the circRNA-disease association network as circRNA similarity and disease similarity. In the future, the similarity computation of circRNAs and diseases could reasonably utilize more biological network information, such as circRNA-miRNA associations, circRNA sequence information and disease semantical information to improve predictive performance.

REFERENCES

[1] X. Fan, X. Weng, Y. Zhao, W. Chen, T. Gan, and D. Xu, "Circular RNAs in cardiovascular disease: An overview," *BioMed Res. Int.*, vol. 2017, 2017, Art. no. 5135781.

[2] J. M. Nigro, K. R. Cho, E. R. Fearon, S. E. Kern, J. M. Ruppert, J. D. Oliner, K. W. Kinzler, and B. Vogelstein, "Scrambled exons," *Cell*, vol. 64, no. 3, pp. 607–613, 1991.

[3] Y. Zhang, X.-O. Zhang, T. Chen, J.-F. Xiang, Q.-F. Yin, Y.-H. Xing, S. Zhu, L. Yang, and L.-L. Chen, "Circular intronic long noncoding RNAs," *Mol. Cell*, vol. 51, no. 6, pp. 792–806, 2013.

[4] D. Knupp and P. Miura, "CircRNA accumulation: A new hallmark of aging?" *Mech. Ageing Dev.*, vol. 173, pp. 71–79, 2018.

[5] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, and A. Loewer, "Circular RNAs are a large class of animal RNAs with regulatory potency," *Nature*, vol. 495, pp. 333–338, Feb. 2013.

[6] H. Sanger, G. Klotz, D. Riesner, H. Gross, and A. Kleinschmidt, "Viroids are single-stranded covalently closed circular rna molecules existing as highly base-paired rod-like structures," *Proc. Natl. Acad. Sci. USA*, vol. 73, no. 11, pp. 3852–3856, Nov. 1976.

[7] M. Danan, S. Schwartz, S. Edelheit, and R. Sorek, "Transcriptome-wide discovery of circular RNAs in archaea," *Nucleic Acids Res.*, vol. 40, no. 7, pp. 3131–3142, Dec. 2012.

[8] Q. Chu, X. Zhang, X. Zhu, C. Liu, L. Mao, C. Ye, Q. H. Zhu, and L. Fan, "PlantcircBase: A database for plant circular RNAs," *Mol. Plant*, vol. 10, no. 8, pp. 1126–1128, Aug. 2017.

[9] L. Chen, C. Huang, X. Wang, and G. Shan, "Circular RNAs in eukaryotic cells," *Current Genomics*, vol. 16, no. 5, pp. 312–318, Oct. 2015.

[10] L. S. Kristensen, T. B. Hansen, M. T. VenøÂÿ, and J. Kjems, "Circular RNAs in cancer: Opportunities and challenges in the field," *Oncogene*, vol. 37, no. 5, pp. 555–565, Feb. 2018.

[11] Y. Chu-Yu, C. Li, L. Chen, Z. Qian-Hao, and F. Longjiang, "Widespread noncoding circular RNA s in plants," *New Phytol.*, vol. 208, no. 1, pp. 88–95, Oct. 2015.

[12] W. R. Jeck and N. E. Sharpless, "Detecting and characterizing circular RNAs," *Nature Biotechnol.*, vol. 32, no. 5, p. 453, 2014.

[13] L. M. Holdt, A. Kohlmaier, and D. Teupser, "Molecular roles and function of circular RNAs in eukaryotic cells," *Cellular Mol. Life Sci.*, vol. 75, no. 6, pp. 1071–1098, Mar. 2018.

[14] J. Salzman, R. E. Chen, M. N. Olsen, P. L. Wang, and P. O. Brown, "Cell-type specific features of circular RNA expression," *PLoS Genet.*, vol. 9, no. 9, 2013, Art. no. 1003777.

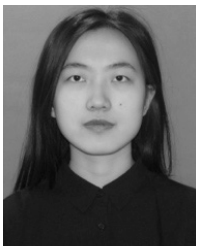
[15] T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems, "Natural RNA circles function as efficient microRNA sponges," *Nature*, vol. 495, pp. 384–388, Feb. 2013.

[16] S. Qu, X. Yang, X. Li, J. Wang, Y. Gao, R. Shang, W. Sun, K. Dou, and H. Li, "Circular RNA: A new star of noncoding RNAs," *Cancer Lett.*, vol. 365, no. 2, pp. 141–148, Sep. 2015.
[17] M. Wang, F. Yu, W. Wu, Y. Zhang, W. Chang, M. Ponnusamy, K. Wang, and P. Li, "Circular RNAs: A novel type of non-coding RNA and their potential implications in antiviral immunity," *Int. J. Biol. Sci.*, vol. 13, no. 12, pp. 1497–1506, 2017.
[18] Y. Yang, X. Gao, M. Zhang, S. Yan, C. Sun, F. Xiao, N. Huang, X. Yang, K. Zhao, H. Zhou, and S. Huang, "Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis," *J. Natl. Cancer Inst.*, vol. 110, no. 3, pp. 304–315, 2018.
[19] Z. Kun-Peng, M. Xiao-Long, and Z. Chun-Lin, "Overexpressed circPVT1, a potential new circular RNA biomarker, contributes to doxorubicin and cisplatin resistance of osteosarcoma cells by regulating ABCB1," *Int. J. Biol. Sci.*, vol. 14, no. 3, pp. 321–330, 2018.
[20] Z. Kun-Peng, Z. Chun-Lin, H. Jian-Ping, and Z. Lei, "A novel circulating hsa_circ_0081001 act as a potential biomarker for diagnosis and prognosis of osteosarcoma," *Int. J. Biol. Sci.*, vol. 14, no. 11, pp. 1513–1520, 2018.
[21] H. Xie, X. Ren, S. Xin, X. Lan, G. Lu, Y. Lin, S. Yang, Z. Zeng, W. Lia, Y. Q. Ding, and L. Liang, "Emerging roles of circRNA_001569 targeting miR-145 in the proliferation and invasion of colorectal cancer," *Oncotarget*, vol. 7, no. 18. 2016 , Art. no. 26680.
[22] Z. Zhao, X. Li, D. Jian, P. Hao, L. Rao, and M. Li, "Hsa_circ_0054633 in peripheral blood can be used as a diagnostic biomarker of pre-diabetes and type 2 diabetes mellitus," *Acta Diabetol.*, vol. 54, no. 3, pp. 237–245, Mar. 2017.
[23] P. Glažar, P. Papavasileiou, and N. Rajewsky, "circBase: A database for circular RNAs," *Rna*, vol. 20, no. 11, pp. 1666–1670, Nov. 2014.
[24] X. Chen, P. Han, T. Zhou, X. Guo, X. Song, and Y. Li, "circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations," *Sci. Rep.*, vol. 6, Oct. 2016, Art. no. 34985.
[25] *PlantCircNet: A Database for Plant circRNA-miRNA-mRNA Regulatory Networks*, document bax089, Dec. 2017.
[26] S. Li, Y. Li, B. Chen, J. Zhao, S. Yu, T. Tang, Q. Zheng, Y. Li, P. Wang, X. He, and S. Huang, "exoRBase: A database of circRNA, lncRNA and mRNA in human blood exosomes," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D106–D112, 2018.
[27] Y.-C. Liu, J.R. Li, C. H. Sun, E. Andrews, R. F. Chao, F. M. Lin, S. L. Weng, S. D. Hsu, C. C. Huang, C. Cheng, and C. C. Liu, "CircNet: A database of circular RNAs derived from transcriptome sequencing data," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D209–D215, 2016.
[28] S. Xia, J. Feng, L. Lei, J. Hu, L. Xia, J. Wang, Y. Xiang, L. Liu, S. Zhong, L. Han, and C. He, "Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes," *Brief. Bioinform.*, vol. 18, no. 6, pp. 984–992, 2017.
[29] A. Bhattacharya and Y. Cui, "SomamiR 2.0: A database of cancer somatic mutations altering microRNA-ceRNA interactions," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1005–D1010, 2016.
[30] S. Xia, J. Feng, K. Chen, Y. Ma, J. Gong, F. Cai, Y. Jin, Y. Gao, L. Xia, H. Chang, and L. Wei, "CSCD: A database for cancer-specific circular RNAs," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D925–D929, 2018.
[31] D. Yao, L. Zhang, M. Zheng, X. Sun, Y. Lu, and P. Liu, "Circ2Disease: A manually curated database of experimentally validated circRNAs in human disease," *Sci. Rep.*, vol. 8, Jul. 2018, Art. no. 11018.
[32] Z. Zhao, K. Wang, F. Wu, W. Wang, K. Zhang, H. Hu, Y. Liu, and T. Jiang, "CircRNA disease: A manually curated database of experimentally supported circRNA-disease associations," *Cell Death Dis.*, vol. 9, no. 5, pp. 4–5, Apr. 2018.
[33] C. Fan, X. Lei, Z. Fang, Q. Jiang, and F. X. Wu, "CircR2Disease: A manually curated database for experimentally supported circular RNAs associated with various diseases," *Database*, vol. 2018, pp. 1–6, Jan. 2018.
[34] X. Lei, Z. Fang, L. Chen, and F. X. Wu, "Pwcda: Path weighted method for predicting circrna-disease associations," *Int. J. Mol. Sci.*, vol. 19, no. 11, p. 3410, Oct. 2018.
[35] C. Yan, J. Wang, and F.-X. Wu, "DWNN-RLS: Regularized least squares method for predicting circRNA-disease associations," *BMC Bioinf.*, vol. 19, no. 19, p. 520, Dec. 2018.
[36] C. Fan, X. Lei, and F.-X. Wu, "Prediction of CircRNA-disease associations using KATZ model based on heterogeneous networks," *Int. J. Biol. Sci.*, vol. 14, no. 14, no. 14, pp. 1950–1959, Nov. 2018.
[37] Q. Xiao, J. Luo, and J. Dai, "Computational prediction of human disease-associated circRNAs based on manifold regularization learning framework," *IEEE J. Biomed. Heal. Informat.*, to be published.
[38] W. Zhang, Q. Qu, Y. Zhang, and W. Wang, "The linear neighborhood propagation method for predicting long non-coding RNA-protein interactions," *Neurocomputing*, vol. 273, pp. 526–534, Jan. 2018.
[39] W. Zhang, X. Yue, F. Huang, R. Liu, Y. Chen, and C. Ruan, "Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network," *Methods*, vol. 145, pp. 51–59, Jun. 2018.
[40] W. Zhang, X. Yue, G. Tang, W. Wu, F. Huang, and X. Zhang, "SFPEL-LPI: Sequence-based feature projection ensemble learning for predicting LncRNA-protein interactions," *PLOS Comput. Biol.*, vol. 14, no. 12, Dec. 2018, Art. no. e1006616.
[41] W. Zhang, X. Yue, Y. Chen, W. Lin, B. Li, F. Liu, and X. Li, "Predicting drug-disease associations based on the known association bipartite network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 503–509.
[42] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. A. Riedl, "Item-based collaborative filtering recommendation algorithms," *Www*, vol. 1, pp. 285–295, May 2001.
[43] W. Zhang, W. Yang, X. Lu, F. Huang, and F. Luo, "The bi-direction similarity integration method for predicting microbe-disease associations," *IEEE Access*, vol. 6, pp. 38052–38061, 2018.
[44] X. Chen, "KATZLDA: KATZ measure for the lncRNA-disease association prediction," *Sci. Rep.*, vol. 5, Nov. 2015, Art. no. 16840.
[45] U. M. Singh-Blom, N. Natarajan, A. Tewari, J. O. Woods, I. S. Dhillon, and E. M. Marcotte, "Prediction and validation of gene-disease associations using methods inspired by social network analyses," *PLoS One*, vol. 8, no. 5, 2013, Art. no. e58977.
[46] X. Chen, Y. A. Huang, Z. H. You, G. Y. Yan, and X. S. Wang, "A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases," *Bioinformatics*, vol. 33, no. 5, pp. 733–739, Dec. 2017.
[47] Y. Qu, H. Zhang, C. Liang, and X. Dong, "KATZMDA: Prediction of miRNA-disease associations based on KATZ model," *IEEE Access*, vol. 6, pp. 3943–3950, 2017.
[48] Y. Zhong, P. Xuan, X. Wang, T. Zhang, J. Li, Y. Liu, and W. Zhang, "A non-negative matrix factorization based method for predicting disease-associated miRNAs in miRNA-disease bilayer network," *Bioinformatics*, vol. 34, no. 2, pp. 267–277, Sep. 2017.
[49] Q. Xiao, J. Luo, C. Liang, J. Cai, and P. Ding, "A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations," *Bioinformatics*, vol. 34, no. 2, pp. 239–248, Sep. 2018.
[50] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proc. Nat. Acad. Sci.*, vol. 107, no. 10, pp. 4511–4515, 2008.
[51] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, and Y. Tang, "Prediction of drug-target interactions and drug repositioning via network-based inference," *PLoS Comput. Biol.*, vol. 8, no. 5, 2012, Art. no. 1002503.
[52] F. Cheng, W. Li, X. Wang, Y. Zhou, Z. Wu, J. Shen, and Y. Tang, "Adverse drug events: Database construction and in silico prediction," *J. Chem. Inf. Model.*, vol. 53, no. 4, pp. 744–752, Apr. 2013.
[53] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC Bioinf.*, vol. 18, no. 1, p. 18, Dec. 2017.
[54] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, "Predicting potential side effects of drugs by recommender methods and ensemble learning," *Neurocomputing*, vol. 173, pp. 979–987, 2016.
[55] X.-Y. Wang, C.-Y. You, Y.-S. Chen, H. Jiang, X. Zheng, W.-W. Tang, X.-Y. Wang, H.-Y. Xu, and F. Geng, "Microrna-671-3p promotes proliferation and migration of glioma cells via targeting cKaP4," *OncoTargets Therapy*, vol. Volume, 11, pp. 6217–6226, 2018.
[56] X.-M. Jiang, Z. L. Li, J. L. Li, Y. Xu, K. M. Leng, Y. F. Cui, and D. J. Sun, "A novel prognostic biomarker for cholangiocarcinoma: circRNA Cdr1as," *Eur. Rev. Med. Pharmacol. Sci.*, vol. 22, no. 2, pp. 365–371, 2018.
[57] A. N. Mamelak and D. B. Jacoby, "Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601)," *Expert Opinion Drug Del.*, vol. 4, no. 2, pp. 86–175, Mar. 2007.
[58] M. L. Goodenberger and R. B. Jenkins, "Genetics of adult glioma," *Cancer Genet.*, vol. 205, no. 12, pp. 21–613, Dec. 2012.
[59] L. Wu, J. Xia, J. Yang, Y. Shi, H. Xia, X. Xiang, and X. Yu, "Circ-ZNF609 promotes migration of colorectal cancer by inhibiting Gli1 expression via microRNA-150," *J. BUON.*, vol. 23, no. 5, pp. 1343–1349, 2018.

[60] S. A. Fedewa, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA Cancer J. Clinicians*, vol. 67, no. 3, pp. 177–193, 2017.
[61] A. C. Society, "Colorectal cancer facts & figures 2011–2013," *Amer. Cancer Soc. Atlanta*, 2011. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2011-2013.pdf
[62] D. Barbagallo, A. Condorelli, M. Ragusa, L. Salito, M. Sammito, B. Banelli, R. Caltabiano, G. Barbagallo, A. Zappalá, R. Battaglia, and M. Cirnigliaro, "Dysregulated miR-671-5p/CDR1-AS/CDR1/VSNL1 axis is involved in glioblastoma multiforme," *Oncotarget*, vol. 7, no. 4, p. 4746, Jan. 2016.
[63] P. Jin, Y. Huang, P. Zhu, Y. Zou, T. Shao, and O. Wang, "CircRNA circHIPK3 serves as a prognostic marker to promote glioma progression by regulating miR-654/IGF2BP3 signaling.," *Biochem. Biophys. Res. Commun.*, vol. 503, no. 3, pp. 1570–1574, 2018.
[64] F. Li, K. Ma, M. Sun, and S. Shi, "Identification of the tumor-suppressive function of circular RNA ITCH in glioma cells through sponging miR-214 and promoting linear ITCH expression," *Amer. J. Transl. Res.*, vol. 10, no. 5, pp. 1373–1386, 2018.
[65] K. Zeng, X. Chen, M. Xu, X. Liu, X. Hu, T. Xu, H. Sun, Y. Pan, B. He, and S. Wang, "CircHIPK3 promotes colorectal cancer growth and metastasis by sponging miR-7," *Cell Death Dis.*, vol. 9, no. 4, 2018, Art. no. 417.
[66] G. Li, J. Luo, Q. Xiao, C. Liang, and P. Ding, "Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity," *J. Biomed. Inform.*, vol. 82, no. February, pp. 169–177, 2018.

WEN ZHANG received the bachelor's and master's degrees in computational mathematics and the Ph.D. degree in computer science from Wuhan University, in 2003, 2006, and 2009, respectively. He is currently a Professor with the College of Informatics, Huazhong Agricultural University, China. His research interests include machine learning and bioinformatics.

CHENGLIN YU is currently pursuing the bachelor's degree with the School of Computer Science, Wuhan University, China. His research interests include machine learning and data mining.



XIAOCHAN WANG is currently pursuing the bachelor's degree with the School of Computer Science, Wuhan University, China. Her research interests include machine learning and data mining.

FENG LIU received the Ph.D. degree in computer science from Wuhan University, China, in 2007, where he is currently a Professor with the School of computer science. His research interests include machine learning and bioinformatics.