IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 18, 2020, accepted January 28, 2020, date of publication February 10, 2020, date of current version February 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2972283

Prioritizing Human Microbe-Disease Associations Utilizing a Node-Information-Based Link Propagation Method

LI PENG[1], DONG ZHOU[1], WEI LIU[2], LIQIAN ZHOU[3], LEI WANG[4], BIHAI ZHAO[4], AND JIALIANG YANG[5]

[1]School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[2]College of Information Engineering, Xiangtan University, Xiangtan 411105, China
[3]School of Computer, Hunan University of Technology, Zhuzhou 412007, China
[4]School of Computer Engineering and Applied Mathematics, Changsha University, Changsha 410022, China
[5]Genesis Beijing Company, Ltd., Beijing 100025, China

Corresponding authors: Li Peng (plpeng@hnu.edu.cn) and Dong Zhou (dongzhou1979@hotmail.com)

The work of Li Peng was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61902125, in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ50187, and in part by the Scientific Research Project of Hunan Education Department under Grant 18B209. The work of Dong Zhou was supported by the NSFC under Grant 61876062. The work of Lei Wang was supported by the NSFC under Grant 61873221. The work of Bihai Zhao was supported in part by the NSFC under Grant 61772089, and in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ40325.

ABSTRACT Growing evidence shows that microbes in human body and body surface play critical roles in the development of many human diseases. Predicting the underlying associations between diseases and microbes is essential for deeply understanding the pathogenesis of diseases. However, biological experiments to find the relationship between microbes and diseases is usually laborious and time-consuming, which presents the need for effective computational tools. In this study, we propose a computational model of node-information-based Link Propagation for Human Microbe-Disease Association prediction (LPHMDA) to prioritize disease-related microbes. LPHMDA and 3 popular methods including KATZHMDA, PBHMDA, and LRLSHMDA were implemented and compared on the Human Microbe-Disease Association Database (HMDAD) based on cross-validation. As a result, LPHMDA achieved an AUC of 0.9135 in leave-one-out cross-validation (LOOCV), outperforming those of the 3 compared methods. In addition, the performances of LPHDMA on the 3-fold CV, 5-fold CV and 10-fold CV were also better than those of the other 3 canonical methods, further demonstrating its superiority. Finally, we took colorectal carcinoma, asthma and obesity as case studies. Interestingly, 9, 9 and 8 of the top 10 novel microbes predicted by LPHMDA to be associated with the 3 diseases respectively could be confirmed by literatures, providing potential disease-associated microbes for further experimental validation. In summary, LPHMDA is an effective method for prioritizing disease-associated microbes.

INDEX TERMS Microbe, disease, microbe-disease association, node-information, link propagation.

I. INTRODUCTION

Microbiota is all microbes existing on human body surfaces and cavity mucous membrane connected with the outside world [1]. Generally, microbes are divided into the following categories: bacteria, fungi, archaea, viruses and others [2]. Microbes are widespread in our bodies and body surfaces, having important effects on human metabolism, behavior, development, adaptation and even evolution [3], [4]. There are rich and diverse microbes in the intestinal tract, skin, oral cavity and genitourinary tract of the human body. It has been confirmed that the number of microbes that survive and reproduce in body and on the body surfaces is 100 billion, 10 times the number of human cells [5]. Microbiota and human body is a mutually beneficial symbiotic relationship [6]. Microbes involved in human metabolism, such as using polysaccharides and nitrogen compounds in diet [7], participating in drug metabolism [8] and affecting drug efficacy [9]; participating in the regulation of the immune system [10], endocrine system and nervous system. Scientists realized that simply focusing on the human body and the human genome does not fully grasp the key issues of human disease and health.

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

Clinical studies show that the disorder of the microbial population is related to multiple system diseases, including digestive system diseases such as irritable bowel syndrome [11], inflammatory bowel disease [12]; immune system diseases such as allergy [13], asthma [14], multiple sclerosis [15]; metabolism and endocrine system diseases such as obesity, diabetes [16]; neuropsychiatric disorders such as depression [17], autism [18], and so on.

In view of the important medical value of a disease-related microorganism, a number of sequence projects, such as the Human Microbiome project (HMP) [19] were set up to analyze the relationships between microbes and the human health. Launched by the National Institutes of Health of the United States, HMP samples 15 parts of hundreds of people to analyze the species and structure of the microorganisms on them. It isolated and cultured the microbial strains, to determine its genome, and attempted to investigate the associations between the microorganisms and the human diseases. Furthermore, some public databases have been established to collect and collate relevant information about disease-related microbes. For example, Ma *et al.* collected many experimental microbe-disease associations from the published literature, formed the human microbe-disease association database (HMDAD) [20]. These data provide a basis for our systematic analysis of the relationships between microorganisms and human diseases [21]. However, using biological experiments to find out the associations between microbes and diseases is expensive and time-consuming. Moreover, biological experiments have certain blindness and limitations, for example, some microbial strains cannot be cultured. As a result, this part of the known microbial-disease association data is still very few.

In recent years, researchers have proposed some computational methods to analyze the relationships between human microbes and diseases. These methods have become a favorable supplement to biological experimental methods. They have greatly contributed to further uncover the mysterious veil of the relationships between microorganisms and human diseases. Chen *et al*. first presented a computational method based on KATZ measure to prioritize non-infectious diseases-related microbes (KATZHMDA) [22]. This method is the first computational tool for mining microbe-disease associations. It assumes that microbes with similar functions tend to be associated with similar non-infectious diseases. As an effective computational model, it has been greatly improved prediction efficiency compared to traditional biological experiments. However, the prediction performance still needs to be improved. Wang *et al.* [23] proposed a semi-supervised method (named LRLSHMDA) for discovering Human microbe-Disease Associations, which is based on the framework of Laplacian Regularized Least Squares. However, the disadvantage of this method is that it is not suitable for the prediction of new diseases. Huang *et al.* developed an approach of Path-Based Human Microbe-Disease Association prediction (PBHMDA) [24], in which a depth-first search algorithm was introduced to traverse all paths in the heterogeneous network to discover the most likely disease-related microbes. Peng *et al.* proposed a method of Adaptive Boosting for Human Microbe-Disease Association prediction (ABHMDA) [25], in which a strong classifier has been utilized to capture the relation probability of microbe-disease pairs. In addition, Bao *et al.* proposed a method called NCPHMDA for inferring disease-related microbes by utilizing network consistency projection [26]. Wang *et al.* developed a novel model named NBLPIHMDA [27], in which the framework of bidirectional label propagation was introduced to reveal potential microbe-disease associations. However, there are some shortcomings in the above methods. For instance, some models can't work for new disease without known association information. In addition, the performance of these models needs further improvement.

In this work, we propose an approach of node-information based Link Propagation for Human Microbe-Disease Association prediction (LPHMDA) to prioritize the most possible disease-related microbes. Node similarity information that contains Gaussian profile kernel similarity and characteristics of disease symptom, has been integrated to promote strong associations between the most likely nodes through link propagation. Kronecker sum operation of the similarity matrices and the technology of Eigenvalue transformation have been adopted to simplify the solving process of the model. We applied LOOCV, k-fold cross validation and case studies to assess the prediction ability of LPHMDA. The results in these experimental situations indicate the reliable capability of LPHMDA for inferring the most possible microbe-disease associations. LPHMDA achieves a superior performance compared with previous approaches.

II. METHOD

A. DATA PREPARATION

In this study, the benchmark dataset can be downloaded from Human Microbe-Disease Association Database (HMDAD) [20], which collecting the experimental verified associations between microbes and diseases in the published literature. After data processing, 450 high-quality known associations including 292 microbes and 39 diseases have been obtained. We denoted the adjacency matrix of microbe-disease associations as P^*, whereas the element P^*_{ij} is set to 1 if there exists association between disease i and microbe j, otherwise is set to 0. We define two sets of microbe nodes and disease nodes by $M \equiv \{m_1, m_2, \cdots m_{|M|}\}$ and $D \equiv \{d_1, d_2, \cdots d_{|D|}\}$, respectively. $|M|$ and $|D|$ denote the total number of microbe nodes and disease nodes, respectively.

B. NODE SIMILARITY INFORMATION MEASUREMENT

Considering the heterogeneous network containing the nodes microbes and diseases, in order to effectively predict the association relationships among the two types of nodes, we firstly measure the similarity information of the nodes in this section. Here, we calculate Gaussian interaction profile

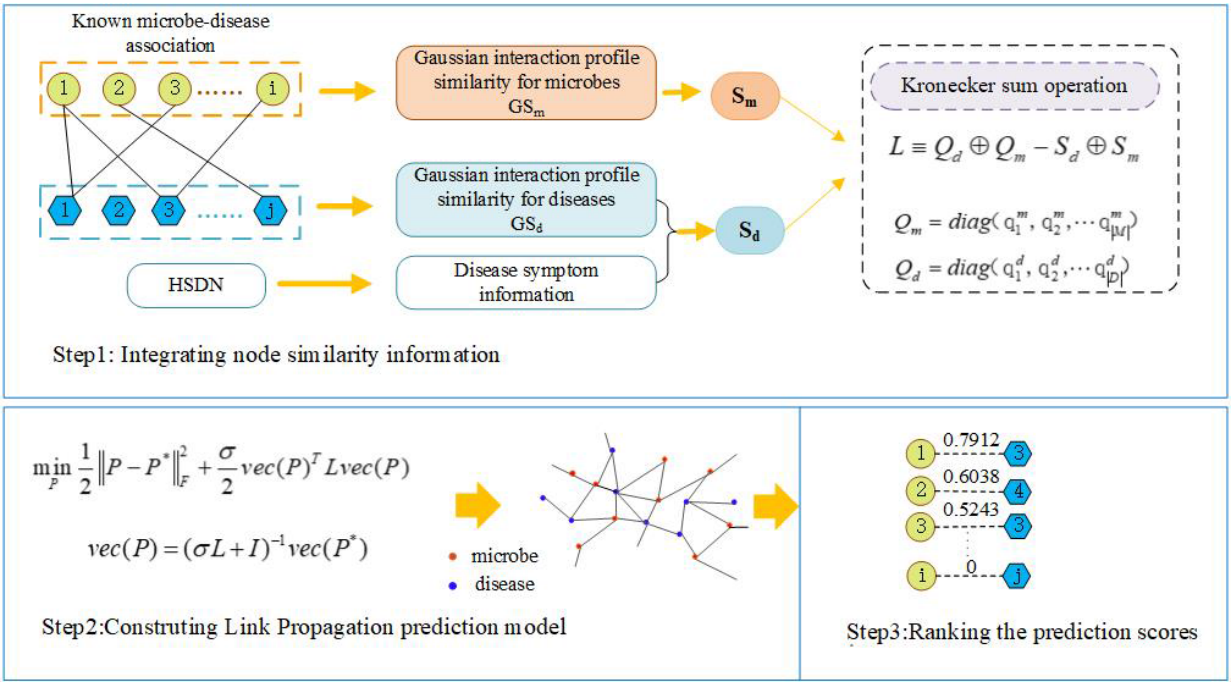

FIGURE 1. **Flowchart of LPHMDA demonstrating the basic ideas of revealing underlying microbe-disease associations by integrating node similarity information.**

kernel similarity for the two types of nodes based on known microbe-disease associations. Binary vector $Y_{(mi)}$ denotes the interaction profiles of microbe m_i, and the value (1 or 0) in $Y_{(mi)}$ records whether microbe m_i is related to each disease. The Gaussian kernel similarity for microbe m_i and m_j is defined as follows:

$$GSm(m_i, m_j) = \exp(-\gamma_m \left\| Y(m_i) - Y(m_j) \right\|^2) \quad (1)$$

Here, γ_m is introducing to adjust the kernel bandwidth and gets from normalizing another bandwidth parameter γ'_m by the mean number of associations with disease per microbe:

$$\gamma_m = \gamma'_m / \left(\frac{1}{|M|} \sum_{i=1}^{m} \| Y(m_i) \|^2 \right) \quad (2)$$

Similarly, the Gaussian kernel similarity for disease GS_d can be defined in a way similar to GS_m. Since the Gaussian kernel similarity calculated above is too dependent on known associations, it is not comprehensive enough to describe the node similarity information if there is very few known microbe-disease pairs. Hence, we integrate the information of disease and corresponding symptom, which can be obtained from the symptom-based human disease network (HSDN). Finally, by combining symptom-based disease similarity matrix SDM and Gaussian kernel similarity GS_d, the disease similarity matrix S_d is defined as follows:

$$S_d = (GS_d + SDM)/2 \quad (3)$$

1) LPHMDA

Motivated by the successful application of link propagation method in Social Networks, drug-target association prediction [28], [29], we explore the link propagation method by constructing a novel computational model for predicting potential microbe-disease associations. Link propagation is a semi-supervised method which used similar principle of label propagation to solve the problem of link prediction in a heterogeneous network. This method assumes that two pairs of similar nodes pairs have similar connection strength. We transform the problem of discovering underlying microbe-disease interactions into the task of link strength prediction between the microbe nodes and the disease nodes in a network. The flowchart of LPHMDA is shown in figure 1. Suppose P^* is an adjacency matrix based on known microbe-disease correlations supported by biological experiment report.

Our goal is to obtain a prediction matrix P, in which the element P_{ij} represents the link strength between disease i and microbe j. The higher value of link strength, the more likely disease i is associated with microbe j.

Based on the above assumptions, we can obtain the objective function of microbe-disease correlation prediction:

$$\min_P \frac{1}{2} \left\| P - P^* \right\|_F^2 + \frac{\sigma}{2} vec(P)^T L vec(P) \quad (4)$$

Here, the first term in the formula represents the constraint on prediction error, and the second term is used to promote the strong correlation of microbe-disease to have closer edge connection. Specifically, if the correlation intensity P^*_{ij} between microbe i and disease j is closer to that P^*_{lm} of microbe l and disease m, the predicted intensity of the two pairs of nodes, P^*_{ij} and P^*_{lm}, should be close to each other. The parameter σ is a regularization parameter, which is used to achieve the balance between the first term and the second term. For the purpose of the above assumption, we define the following the Laplacian

matrix based on the Kronecker sum:

$$L \equiv Q_d \oplus Q_m - S_d \oplus S_m \tag{5}$$

Here, $\oplus$ represents Kronecker sum. In this study, we use the normalized version of the above Laplacian matrix. S_m and S_d represent the microbe similarity matrix and disease similarity matrix calculated above, respectively. Then, we get the diagonal matrix $Q_m = diag(q_1^m, q_2^m, \cdots q_{|M|}^m)$ and $Q_d = diag(q_1^d, q_2^d, \cdots q_{|D|}^d)$, where q_i^{m} $(1 \le i \le |M|)$ represents the sum of similarity scores between the i^{th} microbe and all microbes, that is $q_i^m = \sum_j [S_m]_{ij}$. q_i^d $(1 \le i \le |D|)$ represents the sum of similarity scores between the j^{th} disease and all diseases, that is $q_i^d = \sum_j [S_d]_{ij}$.

In order to obtain the final predicting matrix, formula (1) can be derived from the variable p, and the analytical solution of P can be obtained as follows:

$$vec(P) = (\sigma L + I)^{-1} vec(P^*) \tag{6}$$

Note that, here we can directly calculate the value of P in formula (4). However, a large amount of memory and time overhead is required in the inverse operation of the matrix, including the Kronecker operation. (For example, $O(|M|^2 |D|^2)$ and $O(|M|^3 |D|^3)$, respectively). Because S_m and S_d is a real symmetric matrix, the eigenvalue decomposition technique is used here to improve the computational efficiency. For convenience, we first provide the following lemma. For the real symmetric matrices S_m and S_d, their eigenvalue decompositions can be represented as $\mathrm{S}_m = \mathrm{R}_m \Lambda_m R_m^T$ and $\mathrm{S}_d = \mathrm{R}_d \Lambda_d R_d^T$, respectively. Then, the Kronecker sum $\mathrm{S}_m \oplus \mathrm{S}_d$ is equal to the $\mathrm{R}\Lambda\mathrm{R}^T$, where $\mathrm{R} = \mathrm{R}_m \otimes \mathrm{R}_d$ and $\Lambda = \Lambda_m \oplus \Lambda_d$. The sign $\otimes$ represents the Kronecker product.

In detail, Let $\mathrm{S_m} = \mathrm{R_m} \bigwedge_{\mathrm{m}} \mathrm{R_m^T}$ and $\mathrm{S_d} = \mathrm{R_d} \bigwedge_{\mathrm{d}} \mathrm{R_d^T}$ be the Eigen decompositions of similarity matrices $\mathrm{S_m}$ and $\mathrm{S_d}$. Matrix $\mathrm{R_m}$ and $\mathrm{R_d}$ are composed of the eigenvectors of $\mathrm{S_m}$ and $\mathrm{S_d}$ by column. $\Lambda_m = diag(\lambda_{m_1}, \lambda_{m_2}, \cdots, \lambda_{m_{|M|}})$ and $\Lambda_d = diag(\lambda_{d_1}, \lambda_{d_2}, \cdots, \lambda_{d_{|D|}})$ represent diagonal matrices composed of the eigenvalues of symmetric matrices $\mathrm{S_m}$ and $\mathrm{S_d}$, respectively. Based on the equivalence relation vec (AXB) = $(B^T \otimes \mathrm{A})\mathrm{vec(X)}$, through the basic mathematical operation, we can further obtain the solution of the model (4) (or the equivalent solution to the solution (6) as follows:

$$\begin{aligned}
\mathrm{vec\,(P)} &= R\left((\sigma+1)I - \sigma\Lambda\right)^{-1} R^T vec\left(P^*\right) \\
&= (\mathrm{R}_m \otimes \mathrm{R}_d)\left((\sigma+1)I - \sigma\left(\Lambda_m \oplus \Lambda_d\right)\right)^{-1} \\
&\quad \times \left(R_m^T \otimes R_d^T\right) vec\left(P^*\right) \\
&= (\mathrm{R}_m \otimes \mathrm{R}_d)\left((\sigma+1)I - \sigma(\Lambda_m \oplus \Lambda_d)\right)^{-1} \\
&\quad \times vec\left(R_d^T P^* R_m\right) \\
&= (\mathrm{R}_m \otimes \mathrm{R}_d)\, vec\left(Q \odot \left(R_d^T P^* R_m\right)\right) \\
&= \mathrm{vec}(\mathrm{R}_d \left(Q \odot \left(R_d^T P^* R_m\right)\right) R_m^T)
\end{aligned} \tag{7}$$

Then, we have

$$\mathrm{P} = \mathrm{R}_d \left(Q \odot \left(R_d^T P^* R_m\right)\right) R_m^T \tag{8}$$

Here, Q is a diagonal matrix, and $[Q]_{ij} = \left(1 + \sigma\left(3 - \left(\lambda_{d_i} + \lambda_{m_j}\right)\right)\right)^{-1}$. Besides, symbol $\odot$ represents Hadamard product of matrices. At this time, the inverse operation of the matrix is converted to the reciprocal of the elements and the product of the matrix, and the calculation efficiency is accelerated.

III. RESULTS

A. PERFORMANCE EVALUATION

To comprehensively evaluate the performance of LPHMDA in discovering underlying microbe-disease interactions, we implement LOOCV based on the gold dataset. In LOOCV, to be specific, each of 450 known associations was left out in turn as test samples, and the remaining 449 were used as a training set to predict the score of associations. The score of 1 to 0 positions was compared with the predicted scores of all unknown associations. Then, the ranking values was obtained. Finally, we get the ranking list. If the prediction score is greater than the given threshold, we think the prediction is correct. A series of points can be obtained by setting different threshold values, which correspond to different transverse and longitudinal coordinates. We use ROC curve to evaluate the performance of the predicted results.

ROC curve is a comprehensive index to reflect the continuous variables of sensitivity and specificity. It uses composition method to reveal the relationship between sensitivity and specificity. By setting several different threshold values, a series of sensitivities and specificities are calculated. Then a curve is drawn with sensitivity as longitudinal coordinate and (1-specificity) as transverse coordinate. The larger the area under the curve, the higher the prediction accuracy is. On the ROC curve, the point closest to the upper left of the coordinate graph is the critical value with high sensitivity and specificity.

However, in view of the sparsity of known microbe-disease interactions, using only AUC value to estimate the predicting performance is arbitrary. Therefore, precision-recall (PR) curve and area under PR curve (AUPR) was also using as evaluation criterion to complement the performance estimation. Precision is the relative ratio of the accurately associations retrieved to all associations with score higher than given threshold; the recall is the ratio of the accurately predicted associations to all the relevant results in the database. In general, if the ROC curve and the PR curve have similar changes, meanwhile the AUC and AUPR values more close to 1, the prediction effect is better.

We compared LPHMDA with three state-of-art methods: KATZHMDA, PBHMDA, and LRLSHMDA. We executed a LOOCV for each method and the parameters of KATZHMDA, PBHMDA, and LRLSHMDA are chosen according to the description of their literatures. In our method, there are two parameters σ and α. Parameter σ is the regularization parameter and α is an eigenvalue exponent in the model of

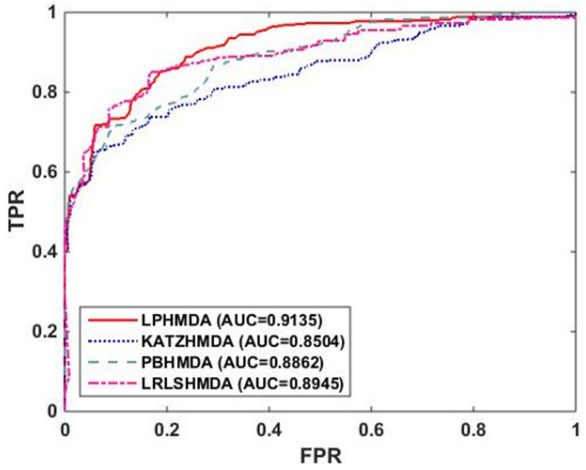

FIGURE 2. ROC curves and average AUC values of LPHMDA and other methods in LOOCV.

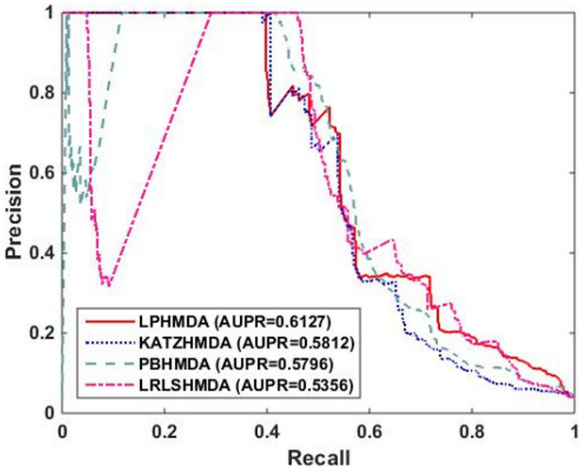

FIGURE 3. Precision-recall curves and AUPR values of LPHMDA, KATZHMDA, PBHMDA and LRLSHMDA.

LPHMDA. It is worth noting that we choose the parameters by grid search method within the algorithm. For simplicity,the eigenvalue index α was selected between 0 and 2 with a step size of 0.1. Through the grid search method, we finally set the parameters $\sigma = 0.01$ and $\alpha = 0.4$, respectively. The ROC curve of LPHMDA and other methods in LOOCV are shown in Figure 2. The average AUC values of LPHMDA, KATZHMDA, PBHMDA, and LRLSHMDA are 0.9135, 0.8504, 0.8862 and 0.8945, respectively. LPHMDA achieves the best result, and its average AUC values are 6.31%, 2.73% and 1.90% higher than other three existing computational methods. Meanwhile, the PR curve and AUPR values of LPHMDA, KATZHMDA, PBHMDA, and LRLSHMDA are also shown in Figure 3. The performance of LPHMDA in terms of PR curve is also superior to KATZHMDA, PBHMDA, and LRLSHMDA, and the average AUPR are 3.15%, 3.31% and 7.71% respectively, higher than other approaches. Evidently, these results confirmed that LPHMDA performs better than that of KATZHMDA, PBHMDA, and LRLSHMDA in LOOCV.

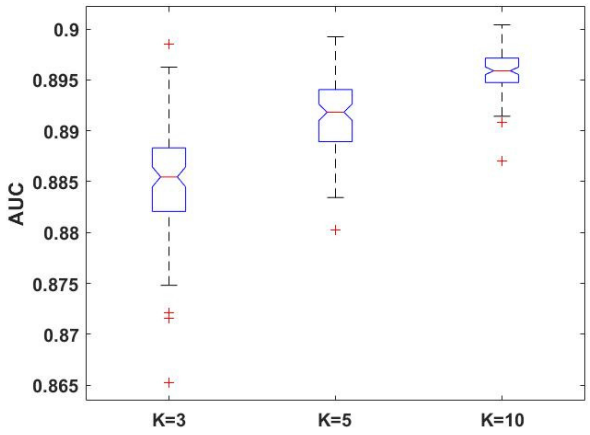

FIGURE 4. AUC values of LPHMDA for 3-fold, 5-fold, 10-fold cross-validation.

Moreover, the 3-fold, 5-fold and 10-fold cross-validation (CV) of our method on the benchmark dataset have also been implemented, respectively. As shown in Figure 4, in the framework of 3-fold CV, 5-fold CV and 10-fold CV, the average AUC value of PLHMDA is 0.8852+/−0.0110, 0.8915+/−0.0082, 0.8964+/−0.0057, respectively. This result shows that the proposed method performs well in different situations.

B. CASE STUDIES

To further demonstrate the practical effect of LPHMDA for revealing the potential relationships between microbes and human digestive system and respiratory diseases from the perspective of microorganisms, three common human diseases were selected in the case studies, including colorectal carcinoma, asthma and obesity. The prediction associations between microbes and each selected diseases are verified by previously published literatures.

Colorectal_carcinoma including colon cancer and rectal cancer, it is one of the five malignant tumors. Almost 80% of colorectal_carcinoma was late at the time of its discovery, and the mortality rate of colorectal_carcinoma is very high [30]. There are more than 1000 species of bacteria in the human intestinal tract [31]. When faced with diet, antibiotics, psychological stress and other stress, intestinal flora disorders lead to the proliferation of potentially harmful species, thus promoting the occurrence of diseases. In recent years, studies have shown that intestinal microorganisms, as important participants in the intestinal environment, may be related to the occurrence and development of colorectal cancer. The specific mechanisms include: increasing intestinal wall permeability, promoting the release of cytokines and chemokines, affecting intracellular signal transduction pathways to promote the occurrence and development of tumors [32]. In addition, a variety of enzymes with metabolic activity produced by intestinal microorganisms can convert procarcinogens into carcinogens. The top 10 potential microbes related to Colorectal_carcinoma predicted by LPHMDA are shown in table 1. As a result, 9 of top 10 have

TABLE 1. The top10 potential microbes related to Colorectal_carcinoma identified by LPHMDA.

Rank	Microbe	Evidence
1	Proteobacteria	PMID:28341746
2	Faecalibacterium_praus	PMID:28045459
3	nitzii	PMID:25182170
4	Enterobacteriaceae	PMID:21247505
5	Streptococcus	PMID:19807912
6	Clostridium_coccoides	PMID:15828052
7	Lactobacillus	PMID:24828543
8	Bacteroides_uniformis	PMID:7074582
9	Staphylococcus_aureus	PMID:22761885
10	Haemophilus	PMID:
	Clostridiales	unconfirmed

TABLE 2. The top10 potential microbes related to Asthma identified by LPHMDA.

Rank	Microbe	Evidence
1	Pseudomonas	PMID: 27076584
2	Lactobacillus	PMID: 20592920
3	Firmicutes	PMID: 23265859
4	Propionibacterium_acnes	PMID: 27433177
5	Propionibacterium	PMID: 29447223
6	Actinobacteria	PMID: 23265859
7	Burkholderia	PMID: 22919592
8	Clostridium_coccoides	PMID: 21477358
9	Enterobacter_hormaechei	PMID: unconfirmed
10	Enterobacter_aerogenes	PMID: 29318023

TABLE 3. The top10 potential microbes related to Obesity identified by LPHMDA.

Rank	Microbe	Evidence
1	Proteobacteria	PMID: 26082274
2	Clostridia	PMID: 27720396
3	Prevotella	PMID: 28017660
4	Lactobacillus	PMID: 28792488
5	Clostridium_coccoides	PMID: 30195549
6	Fusobacterium _nucleatum	PMID: unconfirmed
7	Faecalibacterium _prausnitzii	PMID: 23747247
8	Bacteroides	PMID: 27118489
9	Helicobacter_pylori	PMID: unconfirmed
10	Bacilli	PMID: 28249783

been validated by previous literature. For instance, through a microbiome-based meta-analysis for Colorectal_carcinoma, Shah et al found that Proteobacteria (1st in the rank list) increased significantly in fecal samples of patients compared to normal subjects [33]. In addition, it has been verified that, Faecalibacterium_prausnitzii (2nd in the rank list) is one of the most abundant species of bacteria in human gut. It can be used as a biomarker for the diagnosis of colorectal cancer. In different intestinal diseases, the abundance of Faecalibacterium_prausnitzii is reduced [34].

Bronchial asthma (asthma) characterized by chronic airway inflammation, has gradually become the main disease in most industrialized countries. In recent years, the incidence and mortality of asthma have gradually increased in different countries and regions, and the incidence of bronchial asthma in urban areas is significantly higher than that in rural areas [14]. Studies have shown that the lower respiratory tract is not a completely aseptic environment, in which there are a large number of microbes [35]. When chronic airway inflammatory diseases, including asthma, occur, the species and quantity of bacteria in the lower respiratory tract will change greatly [36]. To evaluate the prediction ability on asthma, we conduct a case study of asthma based on our method. As a result, 9 of top 10 have been validated by experimental evidences documented in previous literature (See table 2). For instance, infection with Pseudomonas (1st in the rank list) can increase the occurrence of asthma [37]. Lactobacillus (2nd in the rank list) has the effect of preventing and treating asthma [38]. Compared with asthmatic patients, Firmicutes (3rd in the rank list) [39] and Actinobacteria (6th in the rank list) are found more frequently in the samples of non-asthma patients [39].

Obesity is a chronic metabolic disease that is caused by multiple factors. Epidemiological data show that over 500 million overweight people in the world, even a lot of children are obese. Obesity is the main factor of various chronic diseases and has a serious effect on human health. The cause of obesity is the result of a variety of factors such as genetic, environment and so on. More and more evidences show that, intestinal microbes are involved in body weight regulation, energy metabolism and inflammation, and play an important role in the occurrence of obesity. Natural delivery, breastfeeding, and avoidance of early-life antibiotic exposure is beneficial to the maintenance of the balance of the intestinal microbes and may reduce the risk of obesity. Probiotics may alter the composition of the intestinal microbes, thereby affecting food consumption and body weight.

In order to explore the relationships between obesity and microbes, we conduct a case study of obesity based on our method. As a result, 8 of top 10 are confirmed by previous literature. For instance, some experiment results demonstrated that Proteobacteria (1st in the rank list) [40], Clostridia (2nd in the rank list) [41] and Prevotella (3rd in the rank list) [42] may be implicated in obesity. Furthermore, Lactabacillas (4th in the rank list) affects the body weight and body fat in obese people [43]. Overall, the results of case studies further verify the prediction effect on the relationships between microbes and specific diseases.

IV. DISCUSSION

A large number of microbes exist in the mouth, intestines, skin and urogenital tract of the human body. Microorganism and human body are a mutually beneficial symbiotic relationship. Microbes have important effects on human metabolism, behavior, development, adaptation and even

evolution. Increasing evidences show that microbes playing a critical role in the development of human diseases. Disorders of microbiota can lead to a variety of diseases, including digestive diseases, immune system diseases, metabolic and endocrine system diseases and so on. Studying on the potential relationships between human diseases and microbes is a very meaningful and challenging work for deepening the disease research from a microbiological point of view. In this study, we propose a novel approach LPHMDA to discover the underlying microbe-disease associations. In LOOCV, the average AUC value of LPHMDA was 0.9135, which is better than that of KATZHMDA (0.8504), PBHMDA (0.8862) and LRLSHMDA (0.8945). The results of 3-fold CV, 5-fold CV, 10-fold CV and case studies further demonstrate the strong prediction power of LPHMDA for revealing microbe-disease associations.

Link propagation is a semi-supervised method which is used in the field of social networks and achieves good performance. It assumes that two pairs of similar nodes pairs have similar connection strength. The problem of discovering underlying microbe-disease interactions can be transform into the task of link strength prediction between the microbe nodes and the disease nodes in a network. Compared with other methods, link propagation algorithm is suitable to solve this kind of association prediction problem. We have made some improvements to the traditional model to make it more efficient to predict microbe-diseases association. Some of the main factors for the reliable performance of LPHMDA can be summarized as follows. Firstly, we consider node similarity information in the framework of link propagation through combining Gaussian interaction profile kernel similarity and disease symptom information, which can promote the strong correlation of microbe-disease to have closer edge connection. Secondly, the Kronecker sum operation of the similarity matrix and the technology of Eigenvalue conversion improve the computational efficiency during the process of solving model. It reduces the large amount of memory and time overhead required in the inverse operation of the matrices. It could be further expand its application in large-scale microbe-disease association networks. Of course, LPHMDA also has some limitations. Firstly, despite the improving prediction performance of LPHMDA compared to previous approaches, it is expected that the prediction ability will be further improved if a more comprehensive similarity calculation method is taken into account. Many research teams put forward some efficient models that we can introduce them to this new field of research. For example, some excellent models for predicting disease-associated noncoding RNA [44]–[56], drug-target associations [57], [58], and some advanced and practical machine learning methods in bioinformatics [59]–[63]. Secondly, the selection of parameters can be further optimized. Third, the interactions among microbes and diseases could be formulated as a network models and network methods could be adjusted to solve this problem [64]. In addition, the evolutionary history microbes might be useful in predicting its association with diseases [65], [66]. Finally, the development of disease is a complex process involving many factors. We can consider the relationships between diseases and other molecules [67]–[72], make a comprehensive analysis of the diseases in the future.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

LP and DZ implemented the experiments, analyzed the result, and wrote the paper. WL and LW analyzed the result, and revised the paper. LQ-Z and BH-Z conceived the project, designed the experiments, analyzed the result, and revised the paper. J-LY analyzed the result and revised the paper. All authors read and approved the final manuscript.

REFERENCES

[1] The Human Microbiome Project Consortium, "A framework for human microbiome research," *Nature*, vol. 486, no. 7402, pp. 215–221, 2012.

[2] The Human Microbiome Jumpstart Reference Strains Consortium, K. E. Nelson, G. M. Weinstock, S. K. Highlander, K. C. Worley, and H. H. Creasy, "A catalog of reference genomes from the human microbiome," *Science*, vol. 328, no. 3281, pp. 994–999, 2010.

[3] The Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, pp. 207–214, 2012.

[4] V. B. Young, "The role of the microbiome in human health and disease: An introduction for clinicians," *BMJ*, vol. 356, p. j831, Mar. 2017.

[5] M. D. Harwich, M. G. Serrano, J. M. Fettweis, J. M. Alves, M. A. Reimers, and V. M. Consortium, "Genomic sequence analysis and characterization of Sneathia amnii sp. nov," *BMC Genomics*, vol. 13, no. 8, p. S4, 2012.

[6] L. Dethlefsen, M. McFall-Ngai, and D. A. Relman, "An ecological and evolutionary perspective on human–microbe mutualism and disease," *Nature*, vol. 449, no. 7164, pp. 811–818, Oct. 2007.

[7] L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, and P. J. Turnbaugh, "Diet rapidly and reproducibly alters the human gut microbiome," *Nature*, vol. 505, no. 7484, pp. 559–563, Jan. 2014.

[8] E. Holmes, J. Kinross, G. R. Gibson, R. Burcelin, W. Jia, S. Pettersson, and J. K. Nicholson, "Therapeutic modulation of microbiota-host metabolic interactions," *Sci. Transl. Med.*, vol. 4, no. 137, pp. 137rv6, Jun. 2012.

[9] M. Donia, P. Cimermancic, C. Schulze, L. W. Brown, J. Martin, M. Mitreva, J. Clardy, R. Linington, and M. Fischbach, "A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics," *Cell*, vol. 158, no. 6, pp. 1402–1414, Sep. 2014.

[10] J. M. Fettweis, M. G. Serrano, B. Huang, J. P. Brooks, A. L. Glascock, N. U. Sheth, J. F. Strauss, K. K. Jefferson, and G. A. Buck, "An emerging mycoplasma associated with trichomoniasis, vaginal infection and disease," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e110943.

[11] J. M. Davies and M. T. Abreu, "Host-microbe interactions in the small bowel," *Current Opinion Gastroenterol.*, vol. 31, pp. 118–123, 2015.

[12] J. Z. Von Martels, M. S. Sadabad, A. R. Bourgonje, T. Blokzijl, G. Dijkstra, K. N. Faber, and H. J. Harmsen, "The role of gut microbiota in health and disease: *In vitro* modeling of host-microbe interactions at the aerobe-anaerobe interphase of the human gut," *Anaerobe*, vol. 44, pp. 3–12, Apr. 2017.

[13] S. F. Bloomfield, G. A. Rook, E. A. Scott, F. Shanahan, R. Stanwell-Smith, and P. Turner, "Time to abandon the hygiene hypothesis: New perspectives on allergic disease, the human microbiome, infectious disease prevention and the role of targeted hygiene," *Perspect. Public Health*, vol. 136, no. 4, pp. 213–224, Jul. 2016.

[14] D. L. Gilstrap and M. Kraft, "Asthma and the host-microbe interaction," *J. Allergy Clin. Immunol.*, vol. 131, no. 5, pp. 1449.e3–1450.e3, May 2013.

[15] P. Rosenstiel, "Stories of love and hate: Innate immunity and host-microbe crosstalk in the intestine," *Current Opinion Gastroenterol.*, vol. 29, no. 2, pp. 125–132, 2013.
[16] B. Owens, "Gut microbe may fight obesity and diabetes," *Nature*, May 2013, doi: 10.1038/nature.2013.12975.
[17] J. Travis, "Microbe linked to Alzheimer's disease," *Sci. News*, vol. 154, no. 21, p. 325, Nov. 1998.
[18] K. Mukherjee, R. Raju, R. Fischer, and A. Vilcinskas, "Galleria Mellonella as a model host to study gut microbe homeostasis and brain infection by the human pathogen listeria monocytogenes," in *Yellow Biotechnology I* (Advances in Biochemical Engineering/Biotechnology), vol. 135. Berlin, Germany: Springer, 2013, pp. 27–39.
[19] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, "The human microbiome project," *Nature*, vol. 449, no. 7164, pp. 804–810, 2007.
[20] W. Ma, L. Zhang, P. Zeng, C. Huang, J. Li, and B. Geng, "An analysis of human microbe-disease associations," *Briefings Bioinf.*, vol. 18, no. 1, pp. 85–97, 2016.
[21] J. Chen, H. Peng, G. Han, H. Cai, and J. Cai, "HOGMMNC: A higher order graph matching with multiple network constraints model for gene–drug regulatory modules identification," *Bioinformatics*, vol. 35, no. 4, pp. 602–610, Feb. 2019.
[22] 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, 2017.
[23] F. Wang, Z. A. Huang, X. Chen, Z. Zhu, Z. Wen, and J. Zhao, "LRLSHMDA: Laplacian regularized least squares for human microbe-disease association prediction," *Sci. Rep.*, vol. 7, Aug. 2017, Art. no. 7601.
[24] Z.-A. Huang, X. Chen, Z. Zhu, H. Liu, G. Y. Yan, Z. H. You, and Z. Wen, "PBHMDA: Path-based human microbe-disease association prediction," *Frontiers Microbiol.*, vol. 8, p. 233, Feb. 2017.
[25] L.-H. Peng, J. Yin, L. Zhou, M.-X. Liu, and Y. Zhao, "Human microbe-disease association prediction based on adaptive boosting," *Frontiers Microbiol.*, vol. 9, p. 2440, May 2018.
[26] W. Bao, Z. Jiang, and D. S. Huang, "Novel human microbe-disease association prediction using network consistency projection," *BMC Bioinf.*, vol. 18, no. 16, p. 543, Dec. 2017.
[27] L. Wang, Y. Wang, H. Li, X. Feng, D. Yuan, and J. Yang, "A bidirectional label propagation based computational model for potential microbe-disease association prediction," *Frontiers Microbiol.*, vol. 10, p. 684, 2019.
[28] Q. Kuang, X. Xu, R. Li, Y. Dong, Y. Li, and Z. Huang, "An eigenvalue transformation technique for predicting drug-target interaction," *Sci. Rep.*, vol. 5, no. 1, 2015, Art. no. 13867.
[29] L. Peng, M. Peng, B. Liao, Q. Xiao, W. Liu, G. Huang, and K. Li, "A novel information fusion strategy based on a regularized framework for identifying disease-related microRNAs," *RSC Adv.*, vol. 7, no. 70, pp. 44447–44455, Sep. 2017.
[30] G. Zeller, J. Tap, A. Y. Voigt, S. Sunagawa, J. R. Kultima, and P. I. Costea, "Potential of fecal microbiota for early-stage detection of colorectal cancer," *Mol. Syst. Biol.*, vol. 10, no. 11, p. 766, 2015.
[31] J. P. Zackular, M. A. M. Rogers, M. T. Ruffin, and P. D. Schloss, "The human gut microbiome as a screening tool for colorectal cancer," *Cancer Prevention Res.*, vol. 7, no. 11, pp. 1112–1121, Nov. 2014.
[32] H. Tlaskalovahogenova, L. Vannucci, K. Klimesova, R. Stepankova, J. Krizan, and M. Kverka, "Microbiome and colorectal carcinoma: Insights from germ-free and conventional animal models," *Cancer J.*, vol. 20, no. 1, pp. 217–224, 2014.
[33] M. S. Shah, T. Z. Desantis, T. Weinmaier, P. J. Mcmurdie, J. L. Cope, A. Altrichter, J.-M. Yamal, and E. B. Hollister, "Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer," *Gut*, vol. 67, no. 5, pp. 882–891, May 2018.
[34] M. Lopez-Siles, S. H. Duncan, L. J. Garcia-Gil, and M. Martinez-Medina, "Faecalibacterium prausnitzii: From microbiology to diagnostics and prognostics," *ISME J.*, vol. 11, no. 4, pp. 841–852, Apr. 2017.
[35] A. Sullivan, E. Hunt, J. Macsharry, and D. M. Murphy, "The microbiome and the pathophysiology of asthma," *Respiratory Res.*, vol. 17, no. 1, p. 163, 2016.
[36] M. T. Shu, D. Mok, K. Pham, M. Kusel, M. Serralha, and N. Troy, "The infant airway microbiome in health and disease impacts later asthma development," *Cell Host Microbe*, vol. 17, p. 704, 2015.
[37] B. Mao, J.-W. Yang, H.-W. Lu, and J.-F. Xu, "Asthma and bronchiectasis exacerbation," *Eur. Respiratory J.*, vol. 47, no. 6, pp. 1680–1686, Jun. 2016.
[38] J. Yu, S.-O. Jang, B.-J. Kim, Y.-H. Song, J.-W. Kwon, M.-J. Kang, W.-A. Choi, H.-D. Jung, and S.-J. Hong, "The effects of Lactobacillus rhamnosus on the prevention of asthma in a murine model," *Allergy, Asthma Immunol. Res.*, vol. 2, no. 3, p. 199, 2010.
[39] P. R. Marri, D. A. Stern, A. L. Wright, D. Billheimer, and F. D. Martinez, "Asthma-associated differences in microbial composition of induced sputum," *J. Allergy Clin. Immunol.*, vol. 131, no. 2, pp. 346.e3–352.e3, Feb. 2013.
[40] J. Allard, "Obesity and the microbiome," *Can. J. Diabetes*, vol. 39, p. S7, Apr. 2015.
[41] C. Carlucci, E. O. Petrof, and E. Allen-Vercoe, "Fecal microbiota-based therapeutics for recurrent clostridium difficile infection, ulcerative colitis and obesity," *EBioMedicine*, vol. 13, pp. 37–45, Nov. 2016.
[42] J. Si, H. J. You, J. Yu, J. Sung, and G. Ko, "Prevotella as a hub for vaginal microbiota under the influence of host genetics and their association with obesity," *Cell Host Microbe*, vol. 21, no. 1, pp. 97–105, Jan. 2017.
[43] L. Crovesy, M. Ostrowski, D. M. T. P. Ferreira, E. L. Rosado, and M. Soares-Mota, "Effect of Lactobacillus on body weight and body fat in overweight subjects: A systematic review of randomized controlled clinical trials," *Int. J. Obesity*, vol. 41, no. 11, pp. 1607–1614, Nov. 2017.
[44] X. Chen, D. Xie, L. Wang, Q. Zhao, Z.-H. You, and H. Liu, "BNPMDA: Bipartite network projection for MiRNA–disease association prediction," *Bioinformatics*, vol. 34, no. 18, pp. 3178–3186, Sep. 2018.
[45] X. Chen, J. Yin, J. Qu, and L. Huang, "MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction," *PLoS Comput. Biol.*, vol. 14, no. 8, Aug. 2018, Art. no. e1006418.
[46] M. Chen, Y. Zhang, A. Li, Z. Li, W. Liu, and Z. Chen, "Bipartite heterogeneous network method based on co-neighbor for MiRNA-disease association prediction," *Frontiers Genet.*, vol. 10, p. 385, Apr. 2019.
[47] Y.-W. Niu, G.-H. Wang, G.-Y. Yan, and X. Chen, "Integrating random walk and binary regression to identify novel miRNA-disease association," *BMC Bioinf.*, vol. 20, Dec. 2019, Art. no. 59.
[48] X. Chen, L. Wang, J. Qu, N.-N. Guan, and J.-Q. Li, "Predicting miRNA-disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, 2018.
[49] W. Li, S. Wang, J. Xu, G. Mao, G. Tian, and J. Yang, "Inferring latent disease-lncRNA associations by faster matrix completion on a heterogeneous network," *Frontiers Genet.*, vol. 10, p. 769, Sep. 2019.
[50] X. Xiao, W. Zhu, B. Liao, J. Xu, C. Gu, and B. Ji, "BPLLDA: Predicting lncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network," *Frontiers Genet.*, vol. 9, p. 411, Oct. 2018.
[51] L. Peng, M. Peng, B. Liao, G. Huang, W. Li, and D. Xie, "The advances and challenges of deep learning application in biological big data processing," *Current Bioinf.*, vol. 13, no. 4, pp. 352–359, Jul. 2018.
[52] X. Liu, J. Yang, Y. Zhang, Y. Fang, F. Wang, and J. Wang, "A systematic study on drug-response associated genes using baseline gene expressions of the cancer cell line encyclopedia," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 22811.
[53] X. Chen, D. Xie, Q. Zhao, and Z.-H. You, "MicroRNAs and complex diseases: From experimental results to computational models," *Briefings Bioinf.*, vol. 20, no. 2, pp. 515–539, Mar. 2019.
[54] X. Chen, C.-C. Zhu, and J. Yin, "Ensemble of decision tree reveals potential miRNA-disease associations," *PLoS Comput. Biol.*, vol. 15, no. 7, Jul. 2019, Art. no. e1007209.
[55] X. Chen, C. C. Yan, X. Zhang, and Z.-H. You, "Long non-coding RNAs and complex diseases: From experimental results to computational models," *Briefings Bioinf.*, vol. 18, no. 4, pp. 558–576, Jul. 2017.
[56] Q. Zou, J. Li, L. Song, X. Zeng, and G. Wang, "Similarity computation strategies in the microRNA-disease network: A survey," *Briefings Funct. Genomics*, vol. 15, no. 1, pp. 55–64, 2016.
[57] X. Chen, C. C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, and Y. Zhang, "Drug-target interaction prediction: Databases, Web servers and computational models," *Briefings Bioinf.*, vol. 17, no. 4, pp. 696–712, 2016.
[58] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: Predicting synergistic drug combinations based on semi-supervised learning," *PLoS Comput. Biol.*, vol. 12, no. 7, Jul. 2016, Art. no. e1004975.
[59] Q. Zou, P. Xing, L. Wei, and B. Liu, "Gene2vec: Gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA," *RNA*, vol. 25, no. 2, pp. 205–218, Feb. 2019.

[60] X. Zeng, L. Liu, L. Lu, and Q. Zou, "Prediction of potential disease-associated microRNAs using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.
[61] Q. Zou, Q. Hu, M. Guo, and G. Wang, "HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy," *Bioinformatics*, vol. 31, no. 15, pp. 2475–2481, 2015.
[62] Q. Zou, X.-B. Li, W.-R. Jiang, Z.-Y. Lin, G.-L. Li, and K. Chen, "Survey of MapReduce frame operation in bioinformatics," *Briefings Bioinf.*, vol. 15, no. 4, pp. 637–647, Jul. 2014.
[63] Q. Zou and Q. Liu, "Advanced machine learning techniques for bioinformatics," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 4, pp. 1182–1183, Jul. 2019.
[64] J. Yang, T. Huang, W. M. Song, F. Petralia, C. V. Mobbs, and B. Zhang, "Discover the network underlying the connections between aging and age-related diseases," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 32566.
[65] J. Yang, S. Grünewald, and X.-F. Wan, "Quartet-Net: A quartet-based method to reconstruct phylogenetic networks," *Mol. Biol. Evol.*, vol. 30, no. 5, pp. 1206–1217, May 2013.
[66] J. Yang, S. Grunewald, Y. Xu, and X. F. Wan, "Quartet-based methods to reconstruct phylogenetic networks," *BMC Syst. Biol.*, vol. 8, no. 1, p. 21, 2014.
[67] X. Chen and G.-Y. Yan, "Novel human lncRNA-disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, 2013.
[68] X. Chen, Y. Z. Sun, N. Guan, J. Qu, and J. Q. Li, "Computational models for lncRNA function prediction and functional similarity calculation," *Briefings Funct. Genomics*, vol. 18, no. 1, pp. 58–82, 2019.
[69] L. Wang, Z.-H. You, X. Chen, Y.-M. Li, Y.-N. Dong, L.-P. Li, and K. Zheng, "LMTRDA: Using logistic model tree to predict MiRNA-disease associations by fusing multi-source information of sequences and similarities," *PLoS Comput. Biol.*, vol. 15, no. 3, Mar. 2019, Art. no. e1006865.
[70] Q. Zou, G. Lin, X.-P. Jiang, X.-R. Liu, and X.-X. Zeng, "Sequence clustering in bioinformatics: An empirical study," *Briefings Bioinf.*, vol. 21, no. 1, pp. 1–10, Jan. 2020, doi: 10.1093/bib/bby090.
[71] L. Peng, M. Peng, B. Liao, G. Huang, W. Liang, and K. Li, "Improved low-rank matrix recovery method for predicting miRNA-disease association," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 6007.
[72] A. Xu, J. Chen, H. Peng, G. Han, and H. Cai, "Simultaneous interrogation of cancer Omics to identify subtypes with significant clinical differences," *Frontiers Genet.*, vol. 10, p. 236, 2019.

LI PENG received the M.S. degree in computer science and technology from the Hunan University of Science and Technology, China, in 2009, and the Ph.D. degree in computer science and technology from Hunan University, China, in 2018. She is currently working as a Lecturer with the Hunan University of Science and Technology. Her current research interests include data mining, bioinformatics, and machine learning.

DONG ZHOU received the Ph.D. degree from the University of Nottingham, U.K., in 2009. He worked as a Research Fellow at the Centre for Next Generation Localization, Trinity College Dublin, Ireland, from 2008 to 2012. He is currently a Professor with the School of Computer Science and Engineering, Hunan University of Science and Technology, China. His current research interests include information retrieval, natural language processing, machine learning, and data mining.

WEI LIU received the M.S. degree in computer science and technology from Xiangtan University, China, in 2009, and the Ph.D. degree in computer science and technology from Hunan University, China, in 2017. He is currently working as a Lecturer with Xiangtan University. His current research interests include complex networks, bioinformatics, and machine learning.

LIQIAN ZHOU received the B.S. degree from the Department of Mathmatics, Xiangtang University, Xiangtang, China, in 1994, and the Ph.D. degree in applied mathematics from Xiangtang University, in 2008. His research interest is intelligent information processing. He is currently the Dean of the School of Computer, Hunan University of Technology.

LEI WANG received the Ph.D. degree in computer science from Hunan University, China, in 2005. From 2005 to 2007, he was a Postdoctoral Fellow with Tsinghua University, China. After that, he moved to USA and Canada as a Visiting Scholar with Duke University and Lakehead University successively. From 2005 to 2011, he was an Associate Professor with the College of Software, Hunan University. From 2011 to 2018, he was a Full Professor with the College of Information Engineering, Xiangtan University. He is currently a Full Professor and an Academic Leader of computer engineering with Changsha University, China. He has published more than 100 peer-reviewed articles. His main research areas include bioinformatics and the Internet of Things.

BIHAI ZHAO received the Ph.D. degree from Central South University, in 2014. He is currently an Associate Professor with the School of Computer Engineering and Applied Mathematics, Changsha University. His research interest includes bioinformatics and data mining.

JIALIANG YANG received the Ph.D. degree from the Department of Mathematics, National University of Singapore, in 2009. From 2010 to 2011, he was an Assistant Professor with the CAS-MPG Partner Institute for Computational Biology, China. After that, he moved to USA, where he became a Postdoctoral Fellow at the Department of Basic Sciences, Mississippi State University. Since 2013, he has been a Postdoctoral Fellow and a Senior Scientist with the Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai. He is currently the Vice President of Genesis Beijing Company, Ltd. He has published more than 80 peer-reviewed articles. His main research areas include bioinformatics, machine learning, oncology, aging, and evolutionary biology.