IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 20, 2019, accepted July 3, 2019, date of publication July 22, 2019, date of current version August 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930453

mHMDA: Human Microbe-Disease Association Prediction by Matrix Completion and Multi-Source Information

CHUANYAN WU[1,3], RUI GAO[1], AND YUSEN ZHANG[2]
[1]School of Control Science and Engineering, Shandong University, Jinan 250061, China
[2]School of Mathematics and Statistics, Shandong University at Weihai, Weihai 264209, China
[3]Diabetes and Endocrinology, Lund University, 20502 Malmo, Sweden

Corresponding authors: Rui Gao (gaorui@sdu.edu.cn) and Yusen Zhang (zhangys@sdu.edu.cn)

This work was supported by the Natural Science Foundation of China under Grant U1806202, Grant 61533011, and Grant 61877064.

ABSTRACT Microbes are vital in human health. It is helpful to promote diagnostic and treatment of human disease and drug development by identifying microbe-disease associations. However, knowledge in this area still needs to be further improved. In this paper, a new computational model using matrix completion to predict human microbe-disease associations (mHMDA, Fig. 1) is developed. First, we extract the disease feature by Gaussian kernel-based similarity and symptom-based similarity. Meanwhile, the microbe feature is computed by Gaussian kernel-based similarity. As treating potential association as the missing elements of a matrix, the matrix completion is adopted to get the potential microbe-disease associations. Leave-one-out cross-validation (LOOCV) is carried out which get the AUC (The area under ROC curve) of 0.928 showing the effectiveness of mHMDA. Furthermore, 5-fold CV get the AUCs of 0.8838 ± 0.0044 (mean $\pm$ standard deviation). Moreover, through the four case studies (asthma, inflammatory bowel disease (IBD), type 2 diabetes (T2D), and type 1 diabetes (T1D)), we find that nine, ten, nine, and eight of top-ten inferred microorganisms for the four diseases are previously verified by experiments. All these results indicate the effectiveness of mHMDA. mHMDA might be helpful to infer the disease-related microorganisms.

INDEX TERMS Microbial community, microbe-disease association prediction, matrix completion.

I. INTRODUCTION

Microorganisms are very important to human health [1], [2]. Numerous lab experiments and clinic studies have found novel links between human diseases and microbes.

It will be helpful to explore the pathogenesis of diseases and prevent or treat diseases by studying the interactions of microbes and diseases (MD). To aid experiments, many computational methods were developed to exploit new relationship between microorganisms and diseases. HMDAD (human microbe-disease associations database) provides basic knowledge of the MD associations [3]. With the knowledge of this database, many models were proposed. Some prediction methods were developed by Gaussian similarity [4]. For instance, KATZHMDA adopted KATZ method to find potential MD associations [5]. PBHMDA was proposed to treat associations as links between microbes and diseases by searching depth firstly algorithm [6]. Some researchers used the matrix factorization technique to investigate the association, such as CMFHMDA utilizing factorization of the collaborative matrix to predict novel association [7]. Furthermore, Random Walk (RW) and its improved algorithms were utilized to get the probabilities of MD association. For instance, RWRHMDA [8], BiRWHMDA [9], and PRWHMDA [10] were recently developed tools utilizing RW methods on the MD network. Moreover, some researches focused on sample processing. ABHMDA utilized k-means to balance the samples to train a model to get the possible links between diseases and microbes [11]. MDPH_HMDA calculated the HeteSim measure on the heterogeneous graph fusing three levels of networks [12]. All of these efforts were to improve the identification of novel MD associations.

However, the efficiency of identifying MD associations requires to be further improved. This study is to present a computational model to infer candidate MD associations with the knowledge of the known associations. In this paper, we put forward a computational model using matrix completion to predict human microbe-disease associations (mHMDA, Fig. 1). For disease feature, we calculate the disease similarity by Gaussian kernel-based and symptom-based similarities. Then we combine the two similarities together to reflect the feature of diseases in different aspects.

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

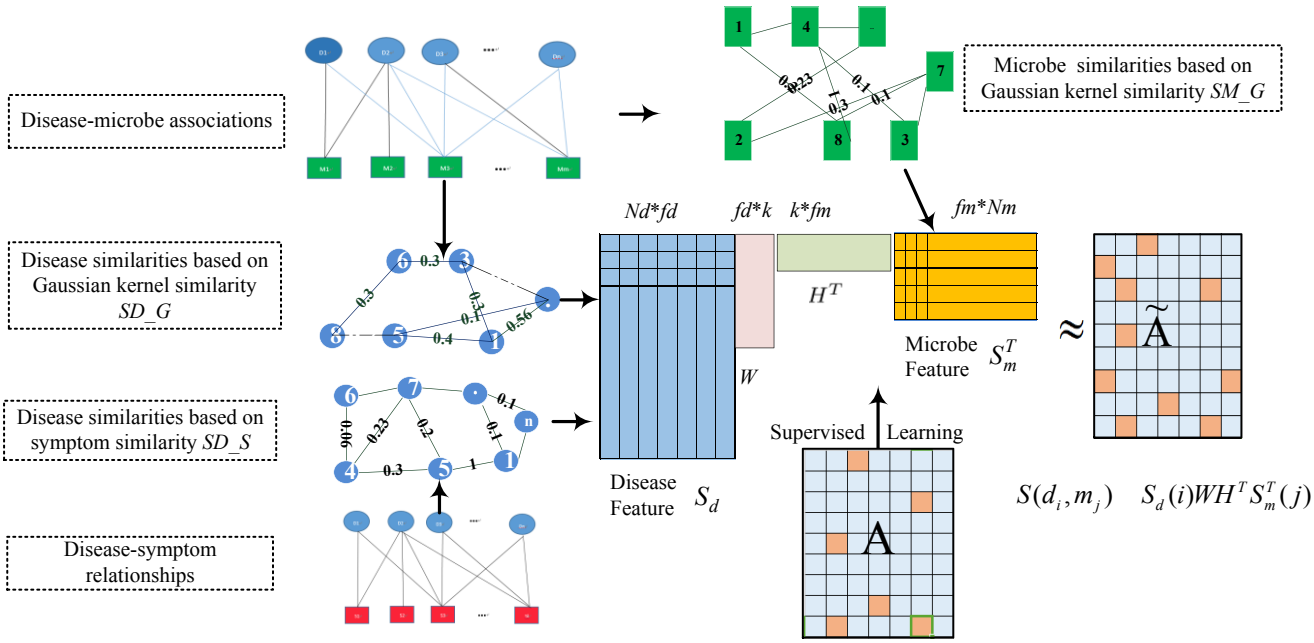


FIGURE 1. The flowchart of mHMDA.

Meanwhile, the Gaussian kernel-based similarity are utilized to calculate the microbe feature. Matrix completion is adopted to recover the missing associations between microbes and diseases. Cross-validations (CV) including LOOCV (leave-one-out cross-validation) and 5-fold CV, comparison with state-of-the-art methods, and case studies are implemented to measure the performance of mHMDA. For LOOCV, the AUC is 0.928, while it is 0.8838 $\pm$ 0.0044 (mean $\pm$ standard deviation) for 5-fold CV. The CVs show the effectiveness of mHMDA. In four case studies (asthma, IBD, T2D, and T1D), more than eight of top-ten inferred microbes have been experimentally verified for each disease. The results show that mHMDA is effective to identify disease-related microorganisms.

The structure of this paper is as follows: The dataset used in this study is depicted in the Material section. We describe MD feature calculation and the mHMDA model in the Methods section. The results of LOOCV, 5-fold CV, comparison with other existing models and the top ten inferred microbes of four diseases are given in the Results section. The results are discussed and the work is concluded in the Discussion and Conclusion part.

II. MATERIAL

HMDAD [3] was adopted in the study as basic knowledge. In HMDAD, there are $N_m = 292$ microbes, $N_d = 39$ diseases, and 483 associations between them. Removing the repetitive associations, 450 unique associations remained. Moreover, Human Symptom Disease Network (HSDN) [13] was used to calculate the symptom-based disease similarity.

III. METHODS

A. DISEASE FEATURE

Gaussian kernel-based similarity comes to one if two vectors are identical, and approaches zero as they move apart. In order to extract the feature of the disease, we calculated the Gaussian kernel-based similarity with the information from HMDAD. Furthermore, the symptom-based similarity of diseases is calculated according to the disease symptom term co-existence information in PubMed.

Firstly, let a binary vector $MS(d_i)$ denote the association information of the specific disease i with each microbe. The interactions of disease i with all the microbes denoted by $MS(d_i)$ can be calculated as follows

$$MS(d_i) = [M_1(d_i), M_2(d_i), \ldots, M_j(d_i), \ldots, M_{N_m}(d_i)], \quad (1)$$

where $M_j(d_i)$ denotes whether disease i is related to the specific microbe j. $M_j(d_i)$ can be calculated as

$$M_j(d_i) = \begin{cases} 1 & \text{microbe } j \text{ has relation with disease } i, \\ 0 & \text{otherwise.} \end{cases}$$

1) GAUSSIAN KERNEL-based DISEASE SIMILARITY

The Gaussian kernel disease-disease similarity is

$$SD\_G = \left(sd_g(d_i, d_j)\right)_{N_d \times N_d}, \quad (2)$$

where $sd_g(d_i, d_j)(i, j = 1, 2, \ldots, N_d)$ denotes the similarity between diseases i and j calculated as

$$sd_g(d_i, d_j) = \exp(-\gamma_d \left\|MS(d_i) - MS(d_j)\right\|^2), \quad (3)$$

where

$$\gamma_d = \gamma_d^{'} / (\sum_{k=1}^{N_d} ||MS(d_k)||^2 / N_d),$$

and γ_d denotes the normalized bandwidth based on initial bandwidth $\gamma_d^{'}$ which we set the value as 4.

2) SYMPTOM-BASED DISEASE-DISEASE SIMILARITY

The symptom-based human disease network (HSDN) [13] can compute disease similarity according to the symptom and disease information in PubMed. The symptom-based disease similarity (SD_S) is introduced into mHMDA model to compute the similarity of diseases.

Thus, the total disease feature S_d is computed by

$$S_d = \alpha * SD_G + \beta * SD_S, \tag{4}$$

where alpha, beta are the weights of Gaussian kernel-based similarity and symptom-based similarity, respectively.

B. MICROBE FEATURE

It is assumed that microbes sharing similar diseases tend to be functionally similar. Suppose the interactions of microbe i with all the diseases represented by $DS(m_i)$ can be calculated as follows

$$DS(m_i) = [D_1(m_i), D_2(m_i), \ldots, D_j(m_i), \ldots, D_{N_d}(m_i)], \tag{5}$$

where $D_j(m_i)$ denotes whether disease j is related to the specific microbe i. $D_j(m_i)$ can be calculated as

$$D_j(m_i) = \begin{cases} 1 & \text{disease } j \text{ has relation with microbe } i, \\ 0 & \text{otherwise.} \end{cases}$$

1) GAUSSIAN KERNEL-BASED MICROBE SIMILARITY

The Gaussian kernel-based similarity for microbes is

$$SM\_G = \left(sm_g(m_i, m_j)\right)_{N_m \times N_m}, \tag{6}$$

where $sm_g(m_i, m_j)(i, j = 1, 2, \ldots, N_m)$ denotes the Gaussian kernel-based similarity between microbes i and j calculated as

$$sm_g(m_i, m_j) = \exp(-\gamma_m ||DS(m_i) - DS(m_j)||^2), \tag{7}$$

where

$$\gamma_m = \gamma_m^{'} / \left(\sum_{k=1}^{N_m} ||DS(m_k)||^2 / N_m \right),$$

γ_m denotes the normalized bandwidth based on initial bandwidth $\gamma_m^{'}$ which we set the value as 4.

Thus, the total microbe feature S_m is computed by

$$Sm = SM_G. \tag{8}$$

C. mHMDA

Matrix completion method recovering a low-rank matrix from a partial sampling of its entries has been widely used in many fields such as collaborative filtering for recommendation [14], multi-label learning [15], [16] and clustering [17] and link prediction [18], [19]. Inspired by this, we designed to recover a matrix using the known elements of MD associations.

By feature extraction, we can get the human MD association matrix $A \in R^{N_d \times N_m}$, disease feature $S_d \in R^{N_d \times f_d}$, microbe feature $S_m \in R^{N_m \times f_m}$, where $f_d = 3N_d$, $f_m = 2N_m$. Since experimentally verified associations are very few, the matrix A is very sparse. The potential associations are treated as missing relationships. The target of our study is to complete the missing elements of A with the supervision of matrix A. The recovered matrix with missing relationships can be denoted as $\tilde{A} = S_d W H^T S_m^T$, where $\tilde{A} \in R^{N_d \times N_m}$, S_d denotes the feature of diseases, $W \in R^{f_d \times k}$, $H \in R^{f_m \times k}$, and S_m is the feature of microbes, k is the minimum rank of W and H. W and H can be obtained by optimizing

$$\min_{W,H} \varphi = \frac{1}{2} \left\| A - S_d W H^T S_m^T \right\|_F^2 + \frac{1}{2} \|W\|_F^2 + \frac{1}{2} \|H\|_F^2, \quad s.t.\ W \geq 0,\ H \geq 0, \tag{9}$$

where $\left\| A - S_d W H^T S_m^T \right\|^2 / 2$ denotes the least square cost function, $\|W\|_F^2 / 2$ and $\|H\|_F^2 / 2$ are regularizations to avoid over-fitting. The method in [20] is utilized to solve the minimum problem. W and H initialized with the random dense matrix can be updated by the iterative equations (10) and (11) until convergence. The iterative equations of W and H are

$$H_k \leftarrow H_k \frac{(S_m^T A^T S_d W)_k}{(S_m^T S_m H W^T S_d^T S_d W + H)_k}, \tag{10}$$

$$W_k \leftarrow W_k \frac{(S_d^T A S_m H)_k}{(S_d^T S_d W H^T S_m^T S_m H + W)_k}. \tag{11}$$

Finally, the probability of having associations between disease i and microbe j can be calculated by applying W and H to

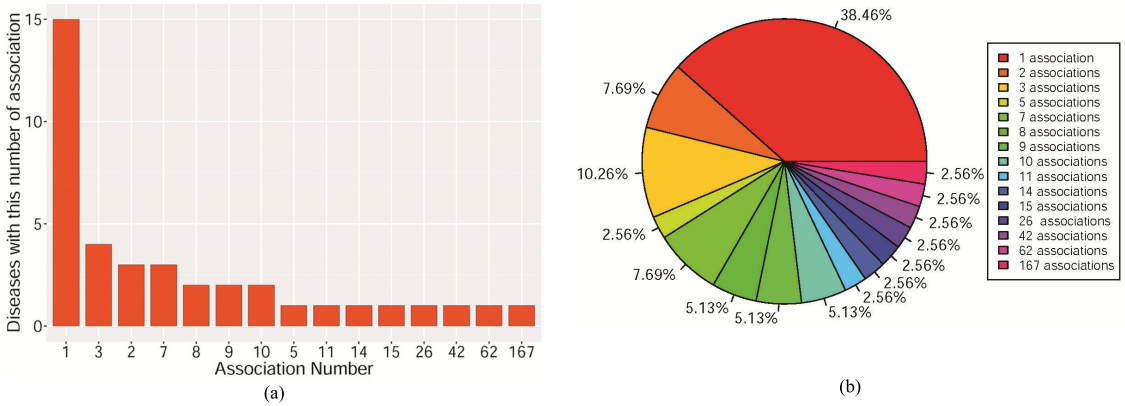
$$S(d_i, m_j) = S_d(i) W H^T S_m^T(j). \tag{12}$$

IV. RESULTS

A. RESULTS OF SIMILARITY

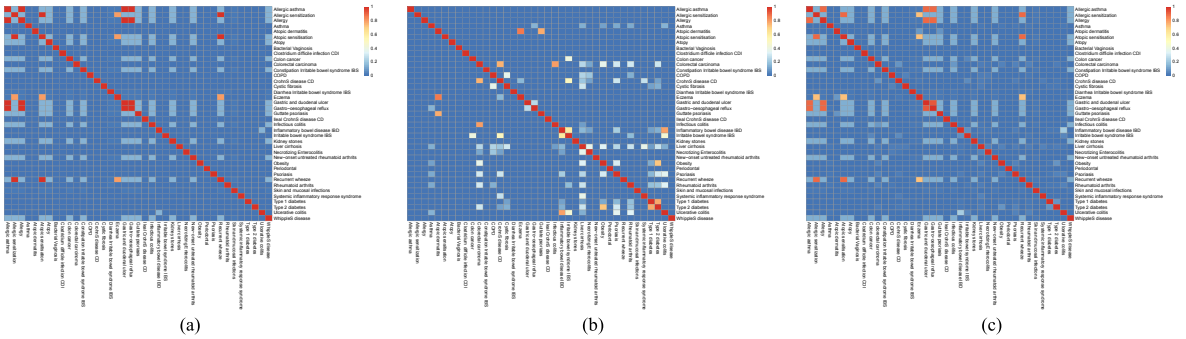
In HMDAD, the distribution of associations is shown in Fig. 2(a) and 2(b). In Fig. 2(a), X-axis denotes the association number, while the Y-axis denotes the number of disease had this number of associations. Fig. 2(a) and Fig. 2(b) indicate that there are 38.46% of diseases (15 diseases) having only one associated microbe. Therefore, some information besides HMDAD is required. Thus, we introduce the symptom-based disease feature which is calculated based on the co-existence of diseases and symptoms.

Gaussian kernel-based similarity is a measurement of similarity between two vectors, which is one if they are identical, and approaches 0 as they move further apart. Considering of this, we calculated the Gaussian kernel-based disease similarity using the interaction profile. Furthermore, symptom-based disease similarity is calculated based on the co-existence of diseases and symptoms. The two kinds of similarity reflect different features of diseases. The Gaussian



(a) (b)

FIGURE 2. The distribution of associations in HMDAD. (a) The diseases with different associations. (b) The diseases with different associations.



(a) (b) (c)

FIGURE 3. The disease-disease similarity. (a) Gaussian kernel-based similarity. (b) Symptom-based similarity. (c) Combined similarity.

kernel-based similarity of diseases is shown in Fig. 3(a). Symptom-based disease similarity is shown in Fig. 3(b). Fig. 3(a) and 3(b) show that by these two similarities, we can get more information. Combining these similarities ($\alpha = 0.9$, β=0.1), the disease feature reflects much more information (Fig. 3(c)).

B. CROSS VALIDATION

LOOCV was performed on the known MD associations from HMDAD. In each round, one known MD association is recruited to test the model trained by the other remaining associations. This is repeated until all the known associations are tested. The test sample is then ranked according to the prediction score in all unverified MD associations. After obtaining the score for each pair, the Receiver Operating Characteristic (ROC) curve can be plotted, with the x-axis representing the false positive rate (FPR) and the y-axis representing the true positive rate (TPR). TPR and FPR vary according to the threshold as a predictive criterion. For a specific threshold, the TPR is calculated as the ratio of the number of correctly predicted positive samples (i.e., having a higher score than a particular threshold) to the number of all positive samples. While FPR refers to the ratio of the number of correctly predicted negative samples to the number of all negative samples. The area under the ROC Curve (AUC) as a metric of the model is calculated to measure the performance. Finally, LOOCV AUC reached 0.928 (Fig. 4 in gray line).

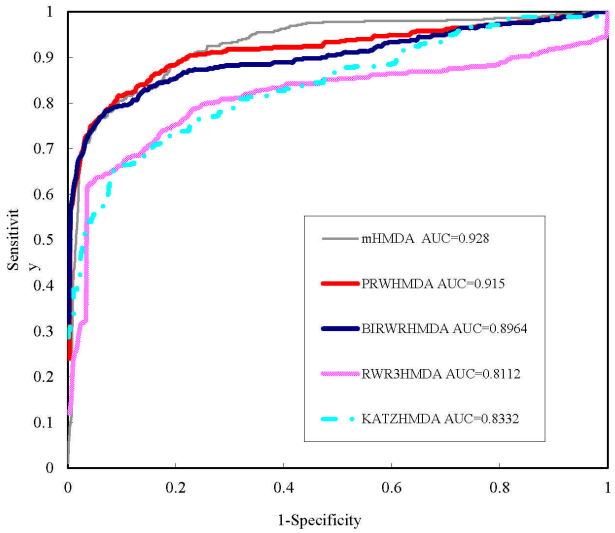


FIGURE 4. The ROC curves and AUC values of different methods.

Furthermore, 5-fold CV was implemented by randomly dividing the MD associations of HMDAD into five groups with no overlap. Of the five groups, four are used to train the model, and the remaining one is used to test the model. This process is repeated until all groups have been tested. To measure the robustness, 5-fold CV was repeated 100 times. Finally, it gets the result of 0.8838 $\pm$ 0.0044 (mean AUC $\pm$ standard deviation).

TABLE 1. The top-ten candidate microbes of Asthma.

Rank	Microbe	Evidence
1	*Lactobacillus*	PMID: 20592920
2	*Actinobacteria*	PMID: 23265859
3	*Pseudomonas*	PMID: 24451910
4	*Clostridium coccoides*	PMID: 21477358
5	*Lachnospiraceae*	PMID: 17433177
6	*Burkholderia*	PMID: 24451910
7	*Firmicutes*	PMID: 23265859
8	*Streptococcus*	PMID: 17950502
9	*Clostridium*	PMID: 27634868
10	*Bacilli*	Not confirmed

C. COMPARISON WITH OTHER METHODS

LOOCV of some state-of-the-art methods (KATZHMD [5], RW3RHMDA [8], BiRWHMDA [9], and PRWHMDA [10]) were also implemented on the same data. The ROCs are shown in Fig. 4. The AUCs of the four methods are 0.8112, 0.8332, 0.8964, and 0.915, respectively. The comparison shows that mHMDA gets a better ROC with relatively higher AUC.

D. CASE STUDIES OF NOVEL ASSOCIATIONS

In this section, we apply mHMDA to four diseases to infer related microbes as case studies.

1) ASTHMA

By mHMDA, we obtained the top-ten related microbes of asthma (Table 1). Among them, nine predicted associations have been experimentally verified.

TABLE 2. The top-ten candidate microbes of IBD.

Rank	Microbe	Evidence
1	*Clostridium coccoides*	PMID: 19235886
2	*Bacteroidetes*	PMID: 28842640, 25307765, 24013298
3	*Streptococcus*	PMID: 23679203, 28842640
4	*Prevotella*	PMID: 25307765
5	*Firmicutes*	PMID: 28842640, 25307765
6	*Lactobacillus*	PMID: 26340825
7	*Verrucomicrobiaceae*	PMID: 22572638
8	*Alistipes*	PMID: 30250234
9	*Parabacteroides*	PMID: 28683448
10	*Rikenellaceae*	PMID: 30250234

Lactobacillus, *Actinobacteria*, *Lachnospiraceae*, and *Firmicutes* (1st, 2nd, 5th, and 7th predicted association) were reported to be lower in asthmatic than non-asthmatics [21], while *Pseudomonas*, *Burkholderia*, *clostridium*(3rd, 6th, 9th predicted association) were found to be over-represented in asthmatics [22], [23]. *Clostridium coccoides* (4th predicted association) species was reported as an early predictor of developing asthma [24]. *Streptococcus* (8th predicted association) might affect the development of asthma [25].

2) INFLAMMATORY BOWEL DISEASE

By mHMDA, we obtained the top-ten related microbes of inflammatory bowel disease (IBD) (Table 2). All of the top-ten microbes in Table 2 have been verified by published work.

TABLE 3. The top-ten candidate microbes of T2D.

Rank	Microbe	Evidence
1	*Prevotella*	PMID: 23613868
2	*Actinobacteria*	PMID: 23613868
3	*Pseudomonas*	PMID: 23613868
4	*Haemophilus*	PMID: 28648853, 20140211
5	*Bacteroides*	PMID: 20140211
6	*Burkholderia*	Not confirmed
7	*Fusobacterium nucleatum*	PMID: 22762355
8	*Lachnospiraceae*	PMID: 30355671, 28701620
9	*Fusobacterium*	PMID: 22762355
10	*Staphylococcus aureus*	PMID: 16495627

TABLE 4. The top-ten candidate microbes of T1D.

Rank	Microbe	Evidence
1	*Fusobacterium nucleatum*	PMID: 25294115
2	*Clostridium coccoides*	PMID: 23433344
3	*Verrucomicrobiaceae*	Not confirmed
4	*Pseudomonas*	PMID: 22864910
5	*Clostridium leptum*	Validated by [38]
6	*Burkholderia*	Not confirmed
7	*Enterobacteriaceae*	PMID: 24475780
8	*Faecalibacterium prausnitzii*	PMID: 20613793
9	*Clostridium*	PMID: 23433344
10	*Tannerella*	PMID: 24236037

Clostridium coccoides, *Bacteroidetes*, *Alistipes*, and *Rikenellaceae*(1st, 2nd, 8th, 10th predicted association) were less enriched in IBD patients than in healthy subjects [26]–[28], whereas *Streptococcus*, *Firmicutes* (3rd, 5th predicted association) were significantly increased [28]. *Prevotella* (4th predicted association) was found to be associated with dysbacteriosis in IBD patients [29]. A recent study showed that *Lactobacillus* (6th predicted association) could help control IBD [30]. Genus of *Parabacteroides* (9th of prediction association) was the most represented in Crohn patients [31].

3) TYPE 2 DIABETES

Table 3 showed the top-ten potential related microbes of T2D, 9 of which have been experimentally validated by previous work.

There were significant differences in the enrichment of *Prevotella*, *Actinobacteria* and *Pseudomonas* (1st, 2nd, 3rd predicted association) between diabetic and non-diabetic patients [32]. The species of *Haemophilus*, order of *Bacteroidales*, the family of *Lachnospiraceae*, and*Fusobacterium* (4th, 5th, 8th, 9th predicted association) were highly enriched in the control samples [33]–[36], while *Fusobacterium nucleatum*, and*Staphylococcus aureus* (7th and 10th predicted association) were detected significantly more often in diabetic subjects than in non-diabetics [36], [37].

4) TYPE 1 DIABETES

Besides the known associated microbes of T1D, we found some novel candidate microbes associated with T1D. Eight of the top-ten associations have been experimentally verified (Table 4).

Fusobacterium nucleatum, *Clostridium*, *Pseudomonal*, *Enterobacteriaceae*, *Clostridium*, *Tannerella* (1st, genus of 2nd, 4th, 7th, 9th, 10th inferred microbes) were found with higher enrichment in T1D group [39]–[43], whereas the enrichment of *Clostridium leptum* (5th inferred microbe) was reduced [38]. It was identified that *Faecalibacterium prausnitzii*(8th inferred microbes) in the gut microbes was correlated with the development and onset of T1D [44].

The four case studies show that mHMDA can be an effective tool to exploit potential MD associations. For all the diseases in HMDAD, the top-ten inferred related microbes are available in S1 File.

V. DISCUSSION AND CONCLUSION

mHMDA was proposed as a novel computational model with the knowledge of previously experimentally validated associations between diseases and microorganisms. To get the disease feature, Gaussian kernel-based and symptom-based similarities were computed. Then we combined the two similarities. To get the microbe feature, Gaussian kernel-based similarity is calculated. Then matrix completion was employed to recover the missing MD associations (possible associations). Apart from CVs (LOOCV and 5-fold CV), comparison, and case studies were carried out to verify the effectiveness of mHMDA. Finally, mHMDA got the AUCs of 0.928 (LOOCV) and 0.8838 $\pm$ 0.0044 (5-fold CV). The case researches show that more than 8 of top-ten inferred microbes have been experimentally verified to have associations with asthma, IBD, T2D, and T1D, respectively. It is believed that mHMDA could be helpful to identify novel MD associations.

The good performance of the proposed method may benefit from the following aspects. (1) Multi-source of information was introduced to calculate the disease feature. We used not only the information of HMDAD, but also the information on symptoms of the disease. (2) Matrix completion method as a semi-supervised learning method was utilized to recover the missing associations with validated microbe-disease associations.

However, there are some limitations to mHMDA. (1) The experimentally validated MD associations used to train the mHMDA are inadequate. For some disease, we have more associations, but most of the diseases have little known associations. This issue is expected to be resolved when more microbe-disease associations are available in the future. (2) Although for disease, we have used multi-source information. For microbe, some other substantial datasets such as microbe homologous sequence similarity should be used to enhance the reliability of microbe feature. We hope that mHMDA might aid medical experiments to get potential associations.

APPENDIX

The top-ten potential inferred microbes of the researched diseases are available in S1 File.

ABBREVIATIONS

mHMDA	computational model using Matrix completion to predict Human Microbe-Disease Associations
CV	Cross-Validation
LOOCV	Leave–One-Out Cross-Validation
IBD	Inflammatory Bowel Disease
T2D	Type 2 Diabetes
T1D	Type 1 Diabetes
MD	Microbes and Diseases
HMDAD	Human Microbe-Disease Associations Database
RW	Random Walk
HSDN	Human Symptom Disease Network
ROC	Receiver Operating Characteristic
FPR	False Positive Rate
TPR	True Positive Rate
AUC	Area Under the ROC Curve

ACKNOWLEDGMENT

C. Wu would like to thank Dr. Zhi-Ping Liu for the valuable suggestions.

REFERENCES

[1] F. Guarner and J.-R. Malagelada, "Gut flora in health and disease," *Lancet*, vol. 361, pp. 512–519, Feb. 2003.
[2] J. M. Pickard, M. Y. Zeng, R. Caruso, and G. Núñez, "Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease," *Immunol. Rev.*, vol. 279, pp. 70–89, 2017.
[3] W. Ma, L. Zhang, P. Zeng, C. Huang, J. Li, B. Geng, J. Yang, W. Kong, X. Zhou, and Q. Cui, "An analysis of human microbe–disease associations," *Brief. Bioinform.*, vol. 18, no. 1, pp. 85–97, 2016.
[4] F. Wang, Z. A. Huang, X. Chen, Z. Zhu, Z. Wen, J. Zhao, and G. Y. Yan, "LRLSHMDA: Laplacian regularized least squares for human microbe–disease association prediction," *Sci. Rep.*, vol. 7, Aug. 2017, Art. no. 7601.
[5] 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, 2016.
[6] 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," *Front. Microbiol.*, vol. 8, p. 233, Feb. 2017.
[7] Z. Shen, Z. Jiang, and W. Bao, "CMFHMDA: Collaborative matrix factorization for human microbe-disease association prediction," in *Proc. Int. Conf. Intell. Comput.*, Jul. 2017, pp. 261–269.
[8] X. Shen, Y. Chen, X. Jiang, X. Hu, T. He, and J. Yang, "Predicting disease-microbe association by random walking on the heterogeneous network," in *Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, Dec. 2016, pp. 771–774.
[9] S. Zou, J. Zhang, and Z. Zhang, "A novel approach for predicting microbe-disease associations by bi-random walk on the heterogeneous network," *PLoS One*, vol. 12, Sep. 2017, Art. no. e0184394.
[10] C. Wu, R. Gao, D. Zhang, S. Han, and Y. Zhang, "PRWHMDA: Human microbe-disease association prediction by random walk on the heterogeneous network with PSO," *Int. J. Biol. Sci.*, vol. 14, no. 8, p. 849, 2018.
[11] L.-H. Peng, J. Yin, L. Zhou, M.-X. Liu, and Y. Zhao, "Human microbe-disease association prediction based on adaptive boosting," *Front. Microbiol.*, vol. 9, p. 2440, May 2018.
[12] C. Fan, X. Lei, L. Guo, and A. Zhang, "Predicting the associations between microbes and diseases by integrating multiple data sources and path-based HeteSim scores," *Neurocomputing*, vol. 323, pp. 76–85, Jan. 2019.
[13] X. Zhou, J. Menche, A.-L. Barabási, and A. Sharma, "Human symptoms–disease network," *Nat. Commun.*, vol. 5, p. 4212, Jun. 2014.
[14] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 8, pp. 30–37, Aug. 2009.
[15] R. S. Cabral, F. Torre, J. P. Costeira, and A. Bernardino, "Matrix completion for multi-label image classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 190–198.

[16] H.-F. Yu, P. Jain, P. Kar, and I. Dhillon, "Large-scale multi-label learning with missing labels," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2014, pp. 593–601.

[17] C.-J. Hsieh, K.-Y. Chiang, and I. S. Dhillon, "Low rank modeling of signed networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2012, pp. 507–515.

[18] N. Natarajan and I. S. Dhillon, "Inductive matrix completion for predicting gene–disease associations," *Bioinformatics*, vol. 30, no. 12, pp. i60–i68, 2014.

[19] 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.

[20] P. Jain and I. S. Dhillon, "Provable inductive matrix completion," 2013, *arXiv:1306.0626*. [Online]. Available: https://arxiv.org/abs/1306.0626

[21] 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, pp. 199–205, Jul. 2010.

[22] A. Beigelman, G. M. Weinstock, and L. B. Bacharier, "The relationships between environmental bacterial exposure, airway bacterial colonization and asthma," *Current Opinion Allergy Clin. Immunol.*, vol. 14, p. 137, Apr. 2014.

[23] L. T. Stiemsma, M. C. Arrieta, P. A. Dimitriu, J. Cheng, L. Thorson, D. L. Lefebvre, M. B. Azad, P. Subbarao, P. Mandhane, A. Becker, and S. R. Sears, "Shifts in Lachnospira and Clostridium sp. In the 3-month stool microbiome are associated with preschool age asthma," *Clin. Sci.*, vol. 130, pp. 2199–2207, Dec. 2016.

[24] C. Vael, L. Vanheirstraeten, K. N. Desager, and H. Goossens, "Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma," *BMC Microbiol.*, vol. 11, p. 68, Dec. 2011.

[25] J. A. Preston, A. T. Essilfie, J. C. Horvat, M. A. Wade, K. W. Beagley, P. G. Gibson, P. S. Foster, and P. M. Hansbro, "Inhibition of allergic airways disease by immunomodulatory therapy with whole killed *Streptococcus pneumoniae*," *Vaccine*, vol. 25, pp. 8154–8162, Nov. 2007.

[26] H. Sokol, P. Seksik, J. P. Furet, O. Firmesse, I. Nion-Larmurier, L. Beaugerie, J. Cosnes, G. Corthier, P. Marteau, and J. Doré, "Low counts of Faecalibacterium prausnitzii in colitis microbiota," *Inflammatory Bowel Diseases*, vol. 15, pp. 1183–1189, Aug. 2009.

[27] A. Butera, M. Di Paola, L. Pavarini, F. Strati, M. Pindo, M. Sanchez, D. Cavalieri, M. Boirivant, and C. De Filippo, "Nod2 deficiency in mice is associated with microbiota variation favouring the expansion of mucosal CD4+ LAP+ regulatory cells," *Sci. Rep.*, vol. 8, Sep. 2018, Art. no. 14241.

[28] M. L. Santoru, C. Piras, A. Murgia, V. Palmas, T. Camboni, S. Liggi, I. Ibba, M. A. Lai, S. Orrú, S. Blois, and A. L. Loizedda, "Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients," *Sci. Rep.*, vol. 7, Aug. 2017, Art. no. 9523.

[29] H. S. Said, W. Suda, S. Nakagome, H. Chinen, K. Oshima, S. Kim, R. Kimura, A. Iraha, H. Ishida, J. Fujita, and S. Mano, "Dysbiosis of salivary microbiota in inflammatory bowel disease and its association with oral immunological biomarkers," *DNA Res.*, vol. 21, pp. 15–25, Sep. 2013.

[30] M. Thomas, P. Langella, and O. Neyrolles, "Lactobacillus acidophilus, un futur outil thérapeutique dans le traitement des maladies inflammatoires chroniques de l'intestin?" *Médecine/Sci.*, vol. 31, no. 8, pp. 715–717, Aug. 2015.

[31] L. R. Lopetuso, V. Petito, C. Graziani, E. Schiavoni, F. P. Sterbini, A. Poscia, E. Gaetani, F. Franceschi, G. Cammarota, M. Sanguinetti, and L. Masucci, "Gut microbiota in health, diverticular disease, irritable bowel syndrome, and inflammatory bowel diseases: Time for microbial marker of gastrointestinal disorders," *Digestive Diseases*, vol. 36, no. 1, pp. 56–65, 2018.

[32] M. Zhou, R. Rong, D. Munro, C. Zhu, X. Gao, Q. Zhang, and Q. Dong, "Investigation of the effect of type 2 diabetes mellitus on subgingival plaque microbiota by high-throughput 16S rDNA pyrosequencing," *PLoS One*, vol. 8, Apr. 2013, Art. no. e61516.

[33] J. Qin *et al.*, "A metagenome-wide association study of gut microbiota in type 2 diabetes," *Nature*, vol. 490, pp. 55–60, Oct. 2012.

[34] M. Horie, T. Miura, S. Hirakata, A. Hosoyama, S. Sugino, A. Umeno, K. Murotomi, Y. Yoshida, and T. Koike, "Comparative analysis of the intestinal flora in type 2 diabetes and nondiabetic mice," *Exp. Anim.*, vol. 1, pp. 17–21, Jan. 2017.

[35] A. Tanca, A. Palomba, C. Fraumene, V. Manghina, M. Silverman, and S. Uzzau, "Clostridial butyrate biosynthesis enzymes are significantly depleted in the gut microbiota of nonobese diabetic mice," *mSphere*, vol. 3, no. 5, 2018, Art. no. e00492.

[36] R. Casarin, A. Barbagallo, T. Meulman, V. R. Santos, E. A. Sallum, F. H. Nociti, P. M. Duarte, M. Z. Casati, and R. B. GonÃğalves, "Subgingival biodiversity in subjects with uncontrolled type-2 diabetes and chronic periodontitis," *J. Periodontal Res.*, vol. 48, pp. 30–36, Feb. 2013.

[37] A. Tamer, O. Karabay, and H. Ekerbicer, "Staphylococcus aureus nasal carriage and associated factors in type 2 diabetic patients," *Jpn. J. Infect. Dis.*, vol. 59, no. 1, p. 10, Feb. 2006.

[38] M. Knip and O. Simell, "Environmental risk factors for type 1 diabetes," *Cold Spring Harb. Perspect. Med.*, vol. 2, Jun. 2012, Art. no. a007690.

[39] J. Sakalauskiene, R. Kubilius, A. Gleiznys, A. Vitkauskiene, E. Ivanauskiene, and V. Šaferis, "Relationship of clinical and microbiological variables in patients with type 1 diabetes mellitus and periodontitis," *Med. Sci. Monit.*, vol. 20, p. 1871, Feb. 2014.

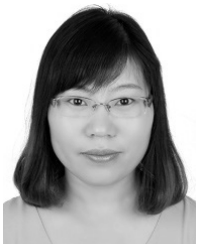
[40] M. Murri, I. Leiva, J. M. Gomez-Zumaquero, F. J. Tinahones, F. Cardona, F. Soriguer, and M. I. Queipo-Ortuño, "Gut microbiota in children with type 1 diabetes differs from that in healthy children: A case-control study," *BMC Med.*, vol. 11, p. 46, Dec. 2013.

[41] L. Peräneva, C. L. Fogarty, P. J. Pussinen, C. Forsblom, P.-H. Groop, and M. Lehto, "Systemic exposure to Pseudomonal bacteria: A potential link between type 1 diabetes and chronic inflammation," *Acta Diabetol.*, vol. 50, pp. 351–361, Jun. 2013.

[42] E. Soyucen, A. Gulcan, A. C. Aktuglu–Zeybek, H. Onal, E. Kiykim, and A. Aydin, "Differences in the gut microbiota of healthy children and those with type 1 diabetes," *Pediatr. Int.*, vol. 56, pp. 336–343, 2014.

[43] E. V. Marietta, A. M. Gomez, C. Yeoman, A. Y. Tilahun, C. R. Clark, D. H. Luckey, J. A. Murray, B. A. White, Y. C. Kudva, and G. Rajagopalan, "Low Incidence of Spontaneous Type 1 Diabetes in Non-Obese Diabetic Mice Raised on Gluten-Free Diets Is Associated with Changes in the Intestinal Microbiome," *PLoS One*, vol. 8, Nov. 2013, Art. no. e78687.

[44] A. Giongo, K. A. Gano, D. B. Crabb, N. Mukherjee, L. L. Novelo, G. Casella, J. C. Drew, J. Ilonen, M. Knip, H. Hyöty, and R. Veijola, "Toward defining the autoimmune microbiome for type 1 diabetes," *ISME J.*, vol. 5, pp. 82–91, Jan. 2011.



CHUANYAN WU received the B.S. degree in information and computing science from the Shandong University, China, in 2003, and the M.S. degree in pattern recognition and intelligent systems from the Hangzhou Dianzi University, China, in 2006. She became a Ph.D. student in Shandong University, China, in 2016. She has been with Lund University, Sweden, since 2018. Her research interests include the algorithms in computational biology, bioinformatics, pattern recognition and artificial intelligence, and big data mining.

RUI GAO received the Ph.D. degree in applied mathematics from Shandong University, China, in 2003. From 2008 to 2009, he was a Visiting Scholar with Washington University in St. Louis, USA. He is currently a Professor with the School of Control Science and Engineering, Shandong University. His current research interests include hybrid dynamical systems, optimal control theory, mathematical modeling of molecular biology, and bioinformatics.

YUSEN ZHANG received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China. From 2004 to 2006, he was a Postdoctoral Fellow in bioinformatics and computer applications with the Graduate University of Chinese Academy of Sciences. He is currently a Visiting Scholar with the Department of Computer Science and Engineering, Wahsington University in St Louis. He is also a Full Professor and an Academic Leader of computer engineering with Shandong University at Weihai, China. Prof. Zhang has published more than 50 peer-reviewed articles. His research interests include algorithms in computational biology, bioinformatics, and statistical genetics.