



OPEN ACCESS

EDITED BY

Qi Zhao,
University of Science and Technology Liaoning,
China

REVIEWED BY

XianFang Tang,
Wuhan Textile University, China
Chunchun Wang,
Jiangnan University, China

*CORRESPONDENCE

Zejun Li
✉ lzjfox@hnit.edu.cn
Lian Duan
✉ duanlian301@163.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 22 June 2023

ACCEPTED 16 August 2023

PUBLISHED 18 September 2023

CITATION

Peng L, Huang L, Tian G, Wu Y, Li G, Cao J, Wang P, Li Z and Duan L (2023) Predicting potential microbe-disease associations with graph attention autoencoder, positive-unlabeled learning, and deep neural network. *Front. Microbiol.* 14:1244527. doi: 10.3389/fmicb.2023.1244527

COPYRIGHT

© 2023 Peng, Huang, Tian, Wu, Li, Cao, Wang, Li and Duan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Predicting potential microbe-disease associations with graph attention autoencoder, positive-unlabeled learning, and deep neural network

Lihong Peng^{1,2†}, Liangliang Huang^{1†}, Geng Tian³, Yan Wu³, Guang Li^{4,5,6,7}, Jianying Cao^{4,5,6,7}, Peng Wang⁸, Zejun Li^{8*} and Lian Duan^{4,5,6,7*}

¹School of Computer Science, Hunan University of Technology, Zhuzhou, China, ²College of Life Sciences and Chemistry, Hunan University of Technology, Zhuzhou, China, ³Geneis (Beijing) Co. Ltd., Beijing, China, ⁴Faculty of Pediatrics, The Chinese PLA General Hospital, Beijing, China, ⁵Department of Pediatric Surgery, The Seventh Medical Center of PLA General Hospital, Beijing, China, ⁶National Engineering Laboratory for Birth Defects Prevention and Control of Key Technology, Beijing, China, ⁷Beijing Key Laboratory of Pediatric Organ Failure, Beijing, China, ⁸School of Computer Science, Hunan Institute of Technology, Hengyang, China

Background: Microbes have dense linkages with human diseases. Balanced microorganisms protect human body against physiological disorders while unbalanced ones may cause diseases. Thus, identification of potential associations between microbes and diseases can contribute to the diagnosis and therapy of various complex diseases. Biological experiments for microbe-disease association (MDA) prediction are expensive, time-consuming, and labor-intensive.

Methods: We developed a computational MDA prediction method called GPUDMDA by combining graph attention autoencoder, positive-unlabeled learning, and deep neural network. First, GPUDMDA computes disease similarity and microbe similarity matrices by integrating their functional similarity and Gaussian association profile kernel similarity, respectively. Next, it learns the feature representation of each microbe-disease pair using graph attention autoencoder based on the obtained disease similarity and microbe similarity matrices. Third, it selects a few reliable negative MDAs based on positive-unlabeled learning. Finally, it takes the learned MDA features and the selected negative MDAs as inputs and designed a deep neural network to predict potential MDAs.

Results: GPUDMDA was compared with four state-of-the-art MDA identification models (i.e., MNNMDA, GATMDA, LRLSHMDA, and NTSHMDA) on the HMDAD and Disbiome databases under five-fold cross validations on microbes, diseases, and microbe-disease pairs. Under the three five-fold cross validations, GPUDMDA computed the best AUCs of 0.7121, 0.9454, and 0.9501 on the HMDAD database and 0.8372, 0.8908, and 0.8948 on the Disbiome database, respectively, outperforming the other four MDA prediction methods. Asthma is the most common chronic respiratory condition and affects ~339 million people worldwide. Inflammatory bowel disease is a class of globally chronic intestinal disease widely existed in the gut and gastrointestinal tract and extraintestinal organs of patients. Particularly, inflammatory bowel disease severely affects the growth and development of children. We used the proposed GPUDMDA method and found that *Enterobacter hormaechei* had potential associations with both asthma and inflammatory bowel disease and need further biological experimental validation.

Conclusion: The proposed GPUDMDA demonstrated the powerful MDA prediction ability. We anticipate that GPUDMDA helps screen the therapeutic clues for microbe-related diseases.

KEYWORDS

microbe-disease associations, graph attention autoencoder, positive-unlabeled learning, *K*-means, XGBoost, deep neural network

1. Introduction

Microorganisms or microbes exist in the form of single cell or a group of cells. Microbes mainly contain bacteria, archaea, fungi, viruses, and protozoa (Wen et al., 2021). They widely distribute on the human skin, oral cavity, respiratory tract, and gastrointestinal tract (Holmes et al., 2015). Most of human microbes are beneficial to human health. They can promote nutrient absorption, protect human body against pathogens, and strengthen metabolic capability. In addition, they have the similar metabolic ability to the liver and are even taken as “forgotten organ” of human body (Gill et al., 2006). However, their imbalance or dysbiosis could cause human diseases (Peng et al., 2022c; Tian et al., 2022), such as inflammatory bowel disease (IBD) (El Mouzan et al., 2018), diabetes (Wen et al., 2008), asthma (Demirci et al., 2019), liver diseases (Henaoui-Mejia et al., 2013), and cancer (Schwabe and Jobin, 2013). Although many evidence demonstrated that microbes have close relationships with human diseases, a comprehensive understanding about how microbes influence human healths and produce diseases remains unknown.

Microbe–disease association (MDA) identification not only help us to capture the mechanisms of complex diseases but also provide multiple possible biomarkers for their diagnosis and therapy. However, traditional wet lab remains costly, laborious, and time-consuming (Chen et al., 2019, 2020; Shen et al., 2022; Chen and Huang, 2023). With the advance of single cell sequencing (Peng et al., 2022d, 2023a,b; Wu et al., 2022; Hu et al., 2023a,b; Xu et al., 2023) and wide application of artificial intelligence (Chen et al., 2021; Lihong et al., 2022; Peng et al., 2022a; Wang et al., 2022, 2023; Zhang et al., 2022a,b; Zhang and Wu, 2023), many computational methods have been developed to discover potential MDAs. These methods mainly contain network-based algorithms and machine learning-based algorithms.

Network-based algorithms take MDA prediction as a random walk or label propagation problem. For example, to decode underlying MDAs, BRWMDA fused similarity networks and bi-random walk (Yan et al., 2019), NBLPIHMDA developed a bidirectional label propagation algorithm (Wang et al., 2019), MHEN constructed a multiplex heterogeneous network (Ma and Jiang, 2020), WMGHMDA implemented iteratively weighted meta-graph search model (Long and Luo, 2019), RWHMDA was a hypergraph-based random walk method (Niu et al., 2019), BDHNS formulated a bi-directional heterogeneous MDA network (Guan et al., 2022), and MNNMDA used low-rank matrix completion (Liu et al., 2023).

Machine learning-based algorithms take MDA prediction as a classification problem. For example, to discover potential MDAs, BPNHMDA (Li et al., 2020) adopted a neural network structure, GATMDA (Long et al., 2021) exploited a graph attention network with inductive matrix completion, DMFMDA (Liu et al., 2020) utilized a deep neural network-based deep matrix factorization model, NinimHMDA (Ma and Jiang, 2020) explored an end-to-end graph convolutional neural network structure, KGNMDA (Jiang et al., 2022) used a graph neural network model, MGATMDA (Liu et al., 2021) comprised decomposer, combiner, and predictor where the decomposer captured the latent components using node-level attention mechanism, the combiner obtained unified embedding using component-level attention mechanism, and unknown microbe–disease pairs were classified by a fully connected network. HNGFL (Wang et al., 2022) designed an embedding algorithm for feature learning and used support vector machine for MDA classification.

Although computational methods significantly improved MDA prediction and uncovered many potential MDAs, there are some limitations presented in this study. For example, network-based MDA inference methods cannot find associated entities for a new microbe or disease. Machine learning-based inference methods need reliable negative MDAs for implementing the MDA classification task. Here, we developed an MDA prediction method called GPUDMDA by combining feature extraction based on graph attention autoencoder (GATE), reliable negative MDA selection based on positive-unlabeled (PU) learning, and MDA classification based on deep neural network (DNN).

2. Materials and methods

2.1. Data preparation

We used two MDA databases to implement MDA prediction. One database is from the Human microbe–disease Association Database (HMDAD; <http://www.cuilab.cn/hmdad>) and contain 450 MDAs between 292 microbes and 39 diseases (Ma et al., 2017). The other comes from Disbiome (<https://disbiome.ugent.be/home>) (Janssens et al., 2018) and contains 4,351 MDAs between 218 diseases and 1,052 microbes. Moreover, an MDA network $Y \in \mathbb{R}^{n \times m}$ is constructed by Eq. (1) as follows:

$$y_{ij} = \begin{cases} 1, & \text{if microbe } m_i \text{ associates with disease } d_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

2.2. Pipeline for MDA prediction

In this manuscript, we developed an MDA prediction method called GPUDDMA by combining graph attention autoencoder, positive-unlabeled learning, and deep neural network. First, GPUDDMA computes disease similarity and microbe similarity matrices by integrating their functional similarity and Gaussian association profile kernel (GAPK) similarity, respectively. Next, it learns features of each microbe–disease pair using GATE. Third, it selects several reliable negative MDAs based on PU learning. Finally, it takes the extracted MDA features and the selected negative MDAs as inputs and proposes a DNN for discovering potential MDAs. Figure 1 shows the pipeline of GPUDDMA.

2.3. Similarity computation

2.3.1. Functional similarity of microbes

In the GATMDA, Long et al. (2021) computed microbe functional similarity according to their co-occurrences (Kamneva, 2017). Similarly, we use the microbe function similarity method in GATMDA and then compute a functional similarity matrix S_m^{fun} between m microbes, where $S_m^{fun}(m_i, m_j)$ denotes the similarity between two microbes m_i and m_j .

2.3.2. Functional similarity of diseases

We use the disease functional similarity assessment method proposed by Long et al. (2021) and compute functional similarity matrix S_d^{fun} between n diseases, where $S_d^{fun}(d_i, d_j)$ denotes the similarity between two diseases d_i and d_j .

2.3.3. Gaussian association profile kernel similarity

GAPK function is a symmetric function along the radial direction. It can better cluster similar examples with linearly separable form (Wang et al., 2020). Let V_{d_i} (the i th row of Y) and V_{d_j} (the j th row of Y) denote two diseases d_i and d_j , respectively, their similarity can be computed by Eq. (2) as follows:

$$G_d(d_i, d_j) = \exp\left(-\theta_d \|V_{d_i} - V_{d_j}\|^2\right) \quad (2)$$

where

$$\theta_d = \frac{1}{n} \sum_{i=1}^n \|V_{d_i}\|^2 \quad (3)$$

Similarly, microbe GAPK similarity G_m is computed.

2.3.4. Similarity integration

Functional similarity is used to measure microbe/disease similarity from the aspect of biological properties. GAPK similarity is used to evaluate microbe/disease similarity from the topological structure of MDA network. As compared with two individual similarity measurements, the combination of functional similarity and GAPK similarity can more accurately assess microbe/disease

similarity and further improve MDA identification performance. Thus, we use the two types of information for microbe/disease similarity evaluation. Moreover, the final disease similarity matrix S_d is computed by integrating their functional similarity and GAPK similarity by Eq. (4) as follows:

$$S_d(d_i, d_j) = \begin{cases} \frac{S_d^{fun}(d_i, d_j) + G_d(d_i, d_j)}{2} & \text{if } S_d^{fun}(d_i, d_j) \neq 0 \\ G_d(d_i, d_j) & \text{otherwise.} \end{cases} \quad (4)$$

Similarly, microbe similarity matrix S_m is computed by Eq. (5) as follows:

$$S_m(m_i, m_j) = \begin{cases} \frac{S_m^{fun}(m_i, m_j) + G_m(m_i, m_j)}{2} & \text{if } S_m^{fun}(m_i, m_j) \neq 0 \\ G_m(m_i, m_j) & \text{otherwise.} \end{cases} \quad (5)$$

2.4. Feature extraction

GATE can efficiently learn features from structured graph data by stacking encoders and decoders (Deng et al., 2022). In this study, we use GATE to extract features for each microbe–disease pair. The GATE structure contain multiple encoders and decoders. In the encoders, each encoder uses a self-attention mechanism to generate new representations for nodes based on their neighborhood information (Veličković et al., 2017). In the k th layer of encoder, relationship between node i and its neighbor node j is computed by Eq. (6) as follows:

$$c_{ij}^{(k)} = \text{Sigmoid}\left(V_s^{(k)T} \sigma\left(W^{(k)} h_i^{(k-1)}\right) + V_r^{(k)T} \sigma\left(W^{(k)} h_j^{(k-1)}\right)\right) \quad (6)$$

where $W^{(k)}$, $V_s^{(k)}$, and $V_r^{(k)}$ denote the trainable parameters in the k th layer of encoder with the sigmoid activation function. $h_i^{(k-1)}$ and $h_j^{(k-1)}$ denote the feature representations of nodes i and j in the $(k-1)$ th layer, respectively. For the i th node, its associations with the other nodes are taken as its initial representation, that is, $h_i^{(0)} = x_i$, and its representation in the k th layer is generated by Eq. (7) as follows:

$$h_i^{(k)} = x_i h_i^k = \sum_{j \in N_i} \alpha_{ij}^{(k)} \sigma\left(W^{(k)} h_j^{(k-1)}\right) \quad (7)$$

We use the softmax function to normalize coefficients of node i 's neighbors and solve the comparability problem by Eq. (8) as follows:

$$\alpha_{ij}^{(k)} = \frac{\exp\left(c_{ij}^{(k)}\right)}{\sum_{l \in N_i} \exp\left(c_{il}^{(k)}\right)} \quad (8)$$

where N_i represents node i and its all neighbors. Moreover, the output in the final layer of encoder is considered the node representations.

In the decoder, the initial attributes of each node are reconstructed. Its input comes from the output in the final layer of encoder. Each neighbor of the current node is assigned to different

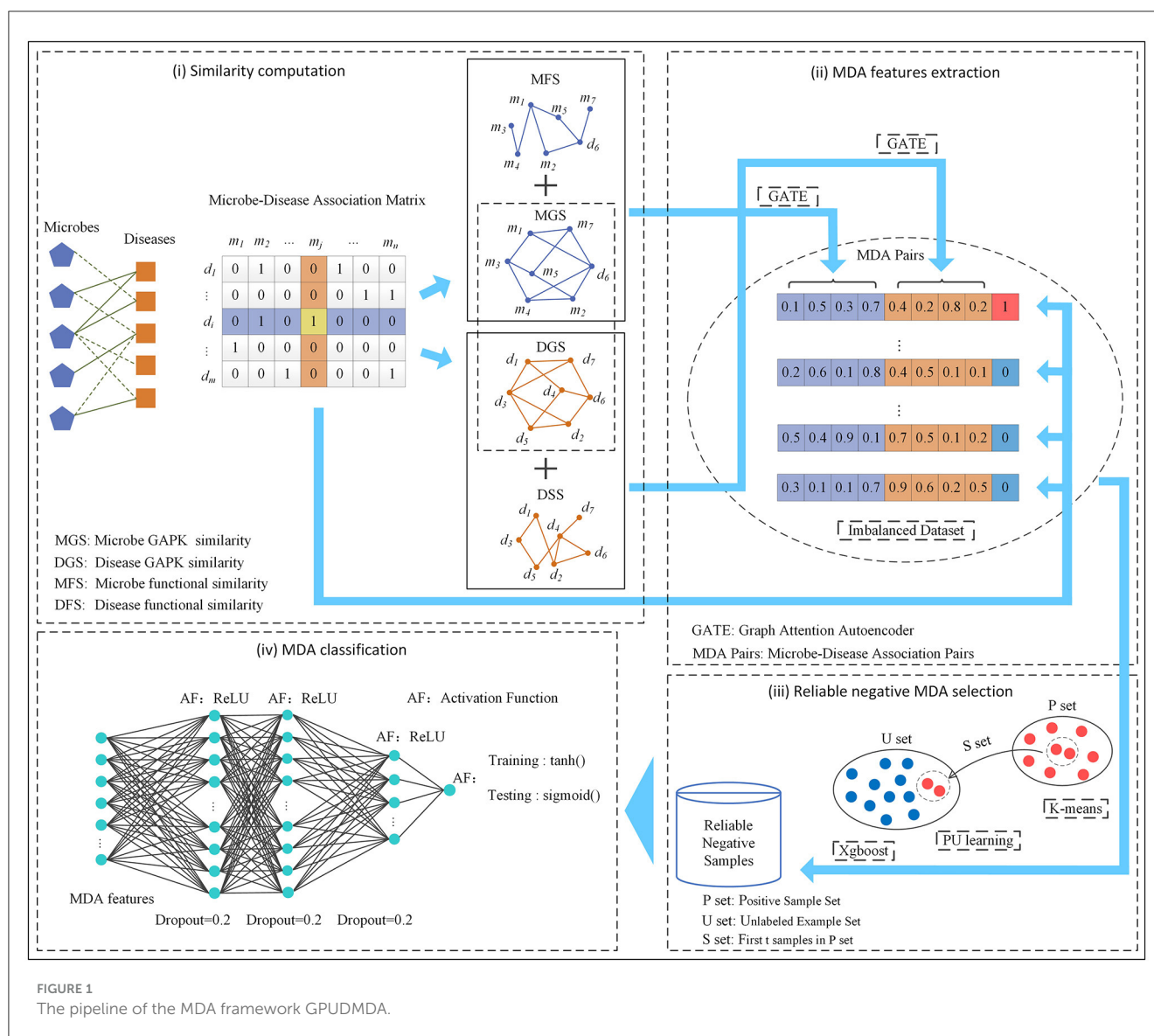


FIGURE 1
The pipeline of the MDA framework GPUDMDA.

weights by the attention mechanism. The normalized relevance between node i and its neighbor j in the k th layer of decoder is computed by Eqs (9) and (10) as follows:

$$\hat{\alpha}_{ij}^{(k)} = \frac{\exp(\hat{c}_{ij}^{(k)})}{\sum_{l \in N_i} \exp(\hat{c}_{il}^{(k)})} \quad (9)$$

$$\hat{c}_{ij}^k = \text{Sigmoid} \left(\hat{V}_s^{(k)T} \sigma \left(\hat{W}^{(k)} \hat{h}_i^{(k)} \right) + \hat{V}_r^{(k)T} \sigma \left(\hat{W}^{(k)} \hat{h}_j^{(k)} \right) \right) \quad (10)$$

where $\hat{W}^{(k)}$, $\hat{V}_s^{(k)T}$, and $\hat{V}_r^{(k)T}$ denote the trainable parameters in the k th layer of decoder. The k th layer in decoder reconstructs the node representations in the $(k-1)$ th layer by Eq. (11) as follows:

$$\hat{h}_i^{k-1} = \sum_{j \in N_i} \hat{\alpha}_{ij}^{(k)} \sigma \left(\hat{W}^{(k)} \hat{h}_j^{(k)} \right) \quad (11)$$

The loss function is defined by Eq. (12) as follows:

$$\text{Loss} = \sum_{i=1}^N \|x_i - \hat{x}_i\|_2 - \lambda \sum_{j \in N_i} \log \left(\frac{1}{1 + \exp(-h_i^T h_j)} \right) \quad (12)$$

where the first and second terms denote the reconstruction loss of node features and one of graph structure, respectively. λ is a hyperparameter used to balance the contribution of two reconstruction loss terms. x_i and \hat{x}_i represent the initial features and the reconstructed features of node i , respectively. h_j is the representation of a neighboring node j of node i .

Finally, we compute microbe feature vectors and disease feature vectors using GATE, and then, a microbe-disease pair is characterized as a a -dimensional vector by concatenating features of both the microbe and the disease.

2.5. Reliable negative MDA selection

In the area of machine learning, negative samples are equally important to final classification performance. However, there are lack of reliable negative MDAs on existing MDA databases due to the limitations of biological experiments. Thus, we design a reliable negative MDA selection method based on PU learning.

PU learning can efficiently identify high-quality negative samples from unlabeled samples and has been widely used in various practical situations (Li et al., 2022). The K -means clustering approach is one of the most popular unsupervised learning algorithms (Peng et al., 2022b). In the HMDAD and Disbiome databases, there are a few positive MDAs and multiple unknown microbe–disease pairs; that is, the two MDA databases are imbalanced. XGBoost has extremely fast parallel computation speed and demonstrates better performance in both balanced and imbalanced databases (Abdu-Aljabar and Awad, 2021).

In this manuscript, we propose a PU learning algorithm to select reliable negative MDAs by combining K -means clustering and XGBoost. Let that positive sample set P and unlabeled example set U denote known MDAs and unknown microbe–disease pairs, respectively. To select reliable negative MDAs from U , as shown in Algorithm 1, we design a PU learning algorithm.

```

1: Clustering each MDA sample with the  $K$ -means
   clustering algorithm based on the extracted MDA
   features using GATE.
2: Selecting the first  $t$  samples in  $P$  which have
   the smallest distance with cluster centroid as  $S$ 
   and adding  $S$  into  $U$ .
2: Taking  $P-S$  as positive samples, and  $U+S$ 
   as negative samples.
3: Calculating association score matrix  $A$  for all
   microbe–disease pairs based on XGBoost.
4: Ranking microbe–disease pairs in  $S$  based on
   association scores in  $A$  and obtaining the
   minimum score  $A_{min}$  in  $S$ .
5: For every sample  $x$  in  $U$ 
6:   If  $A_x$  satisfies  $A_x < A_{min}$ 
7:     then  $RN = RN \cup x$ 
8: Endfor
9: Obtaining reliable negative MDA samples  $RN$ .

```

Algorithm 1. A PU learning algorithm for selecting reliable negative MDAs.

Particularly, during PU learning, if spy samples are randomly selected from positive sample set P and placed into U , the obtained spy samples could be located at the boundary of the class cluster composed of samples in the entire P and belong to outliers. These spy samples have low spatial similarity with unknown positive examples in U . If a large number of noise or outliers are selected as spy samples, it will greatly affect the evaluation of the classifier on unlabeled samples, which could directly cause decreasing classification performance. Thus, we use K -means clustering algorithm for spy sample selection.

2.6. MDA prediction

We build a DNN to classify unknown microbe–disease pairs based on the extracted MDA features, the selected reliable negative MDAs, and known MDAs. The DNN contains an input layer, multiple hidden layers, and an output layer. In the input layer with a neurons, each MDA sample x with a -dimensional features is fed into the model by Eq. (13) as follows:

$$x = [x_1, x_2, \dots, x_a] \quad (13)$$

where x_i denotes the i th feature in x .

The j th hidden layer outputs the results by Eq. (14) as follows:

$$h_j = \sum_{i=1}^a w_i x_i + b_j \quad (14)$$

$$f(h_j) = \text{ReLU}(h_j)$$

where f denotes the ReLU activation function. Finally, the output layer with the sigmoid activation function outputs MDA classification results by Eq. (15) as follows:

$$\sigma(h) = \frac{1}{1 + e^{-h'}} \quad (15)$$

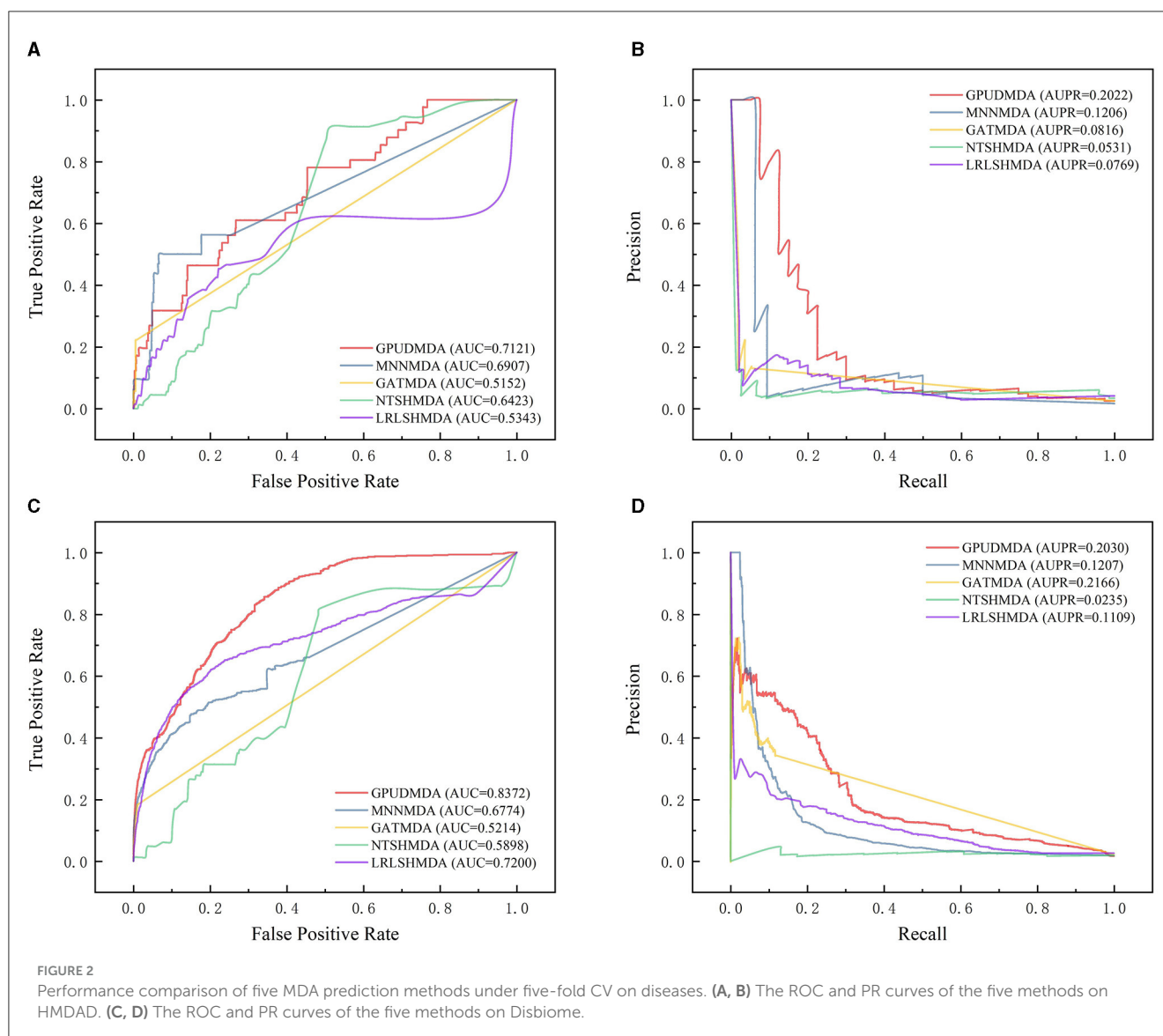
where h' denotes the output in the final hidden layer.

3. Result

3.1. Experimental settings

To evaluate the MDA prediction performance of our proposed GPUDMDA method, we compared it with other MDA identification methods (LRLSHMDA, NTSHMDA, GATMDA, and MNMDA) under five-fold cross validation (CV) on diseases, microbes, and microbe–disease pairs for 20 times. LRLSHMDA (Wang et al., 2017) is Laplacian regularized least square-based MDA identification algorithm, NTSHMDA (Luo and Long, 2018) is integrated random walk and network topology similarity, GATMDA (Long et al., 2021) combined inductive matrix completion and graph attention networks to complete missing MDAs, and MNNMDA (Liu et al., 2023) used a low-rank matrix completion model for identifying possible MDAs. During MDA prediction, it is not enough to reflect the MDA identification performance of a computational model only through cross-validation on microbe–disease pairs. Thus, in the study, we implemented cross-validations on microbes, diseases, and microbe–disease pairs to comprehensively assess the model's performance. The detailed definitions about the above three cross-validations have been proposed by Peng et al. (2020). AUC and AUPR were applied to measure the performance of MDA prediction methods.

In this study, we used GATE to extract features of microbes and diseases from their similarity networks, both of which are 64 dimensional vectors. We selected t samples from positive sample set P to place unlabeled example set U . When t was set to 15% of P on the HMDAD database and 20% of P on the



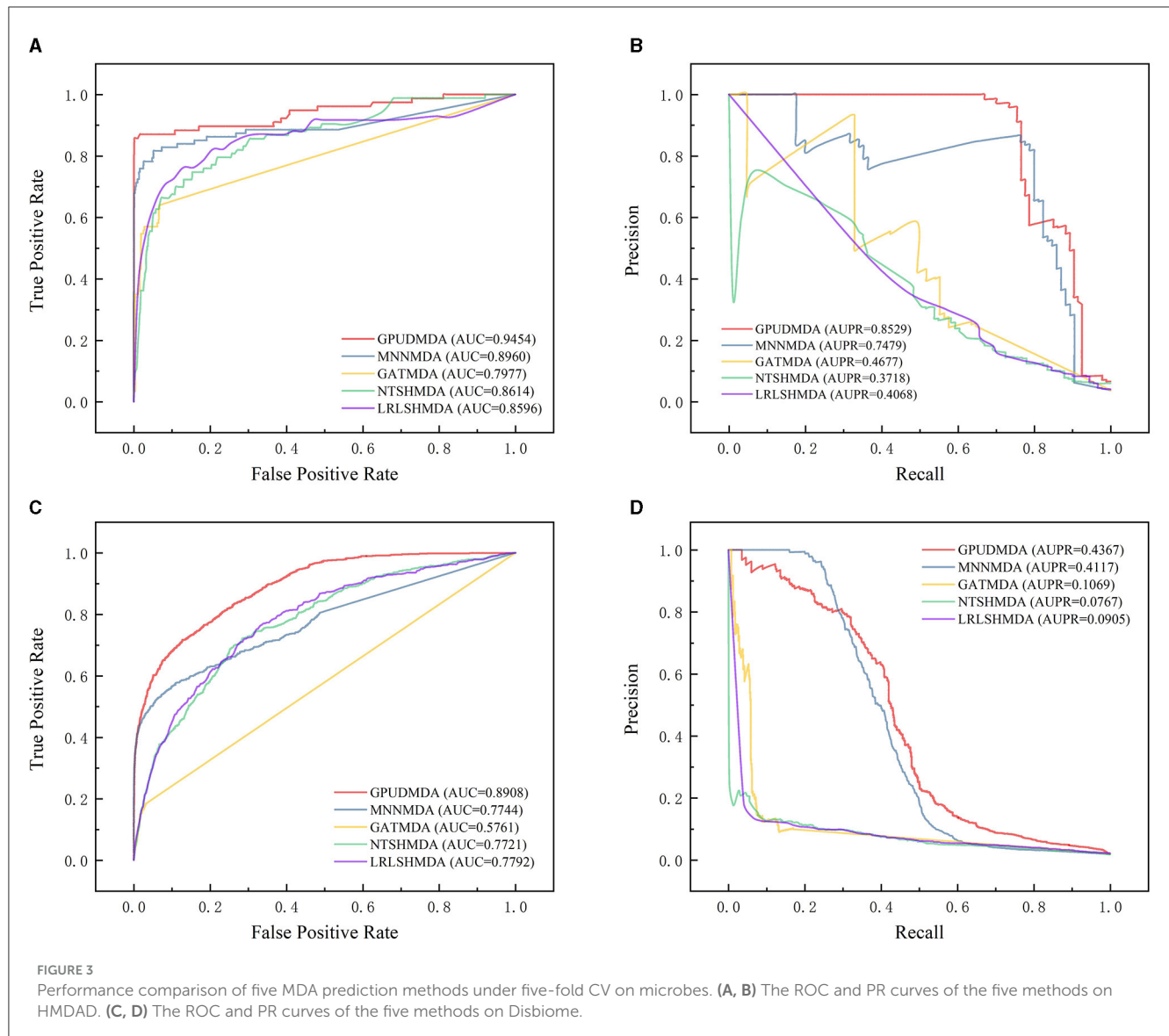
Disbiome database, GPUDDMA obtained the best performance. Thus, we set t to 15 and 20% of P on the two databases, respectively. For DNN with four layers, the input layer, the following three hidden layer, and the output layer have 128, 100, 100, 50, and one nodes, respectively. Learning rate and “dropout” were set to 0.001 and 0.2. The parameter “epoch_num,” denoting the number of training, was set to 300 and 1,500 on the two databases, respectively. Disbiome is a larger dataset, and the proposed computational model needs to be trained for enough times to obtain better classification performance; thus, the “epoch_num” value was much larger on the Disbiome database.

Additionally, the number of positive samples is the same as one of the known MDAs. The number of selected credible negative MDAs is related to the computed smallest association probability score A_{min} . Since the credible negative MDAs were selected from unknown microbe–disease pairs, unknown

microbe–disease pairs were decreased but accounted for most of all microbe–disease pairs.

3.2. Performance comparison under CV on diseases

Under CV on diseases, 80% diseases were taken as the training set and the remaining was test set. Figure 2 elucidates the receiver operating characteristic (ROC) and precision-recall (PR) curves of the five MDA prediction methods on the HMDAD and Disbiome databases under CV on diseases. Under CV on diseases, GPUDDMA obtained the best AUCs of 0.7121 and 0.8372, and the best AUPRs of 0.2022 and better AUPR of 0.2030 on the HMDAD and Disbiome databases, respectively, significantly outperforming LRLSHMDA, NTSHMDA, GATMDA, and MNNDMA.



3.3. Performance comparison under CV on microbes

Under CV on microbes, 80% microbes were taken as the training set and the remaining was test set. Figure 3 shows the ROC and PR curves of the five methods under CV on microbes. Under CV on microbes, GPUDMDA obtained better AUCs of 0.9454 and 0.8908 and AUPRs of 0.8529 and 0.4367 than LRLSHMDA, NTSHMDA, GATMDA, and MNNMDA.

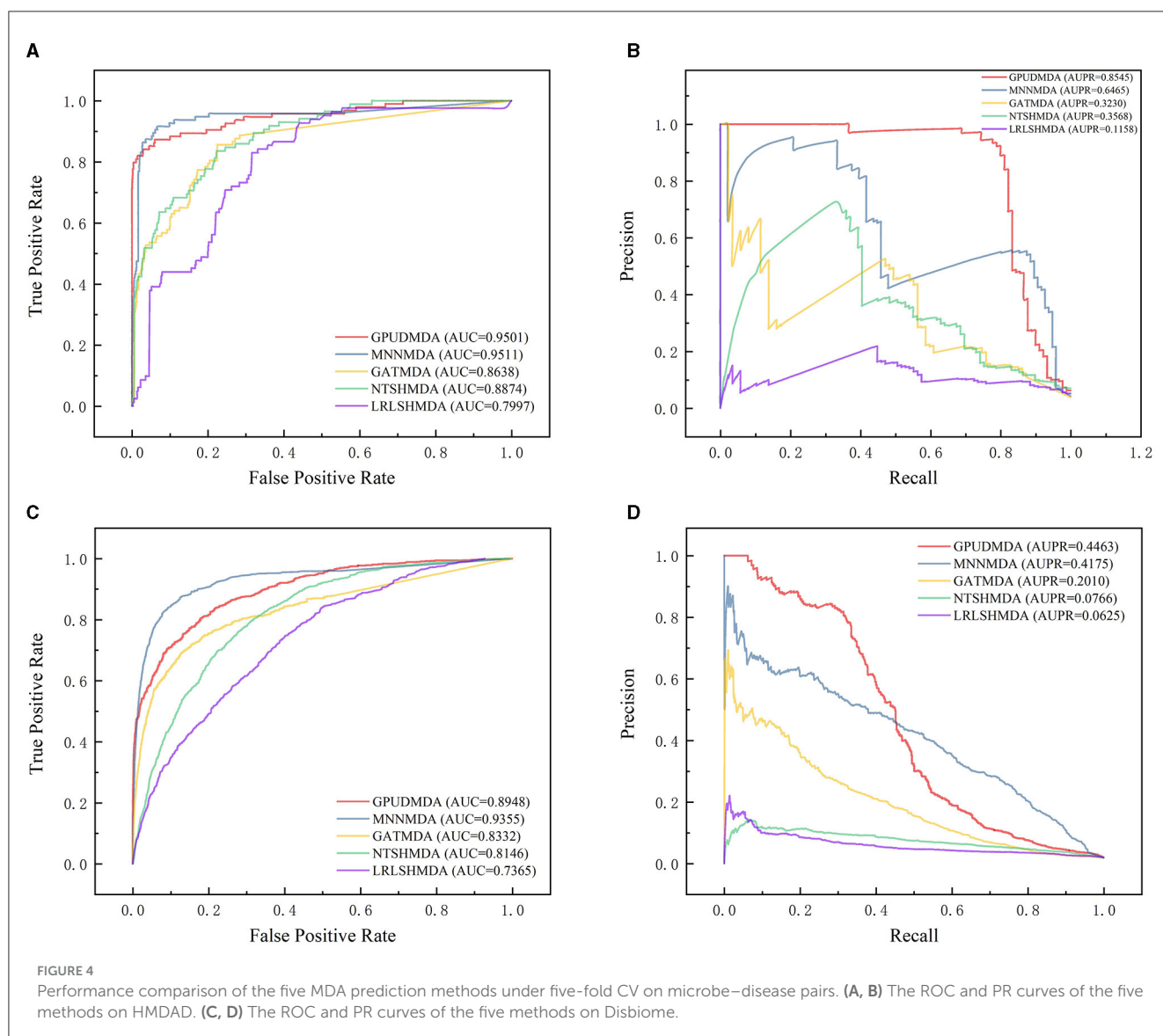
3.4. Performance comparison under CV on microbe–disease pairs

Under CV on microbe–disease pairs, 80% microbe–disease pairs were taken as the training set and the remaining was test set. Figure 4 illustrates the ROC and PR curves of the five MDA prediction methods under CV on microbe–disease pairs. Under the

CV, GPUDMDA computed better AUCs of 0.9501 and 0.8948, and the best AUPRs of 0.8545 and 0.4464 among the five methods.

3.5. The affect of PU learning on performance

Reliable negative samples can improve the classification performance of a model. To evaluate the reliability of the identified negative MDAs by GPUDMDA, we compared its performance under negative sample selection. Figure 5 demonstrates the affect of negative samples selected by PU learning on performance. The results elucidated that GPUDMDA with PU learning outperformed one without PU learning. Particularly, the performance of GPUDMDA with PU learning obtained significant improvement on Disbiome. The results suggested that reliable negative MDAs selected by PU learning can boost the MDA prediction ability.



3.6. Case study

In the above sections, we have confirmed the MDA identification accuracy of GPUDMDA. Next, we intend to find new microbes for asthma and IBD.

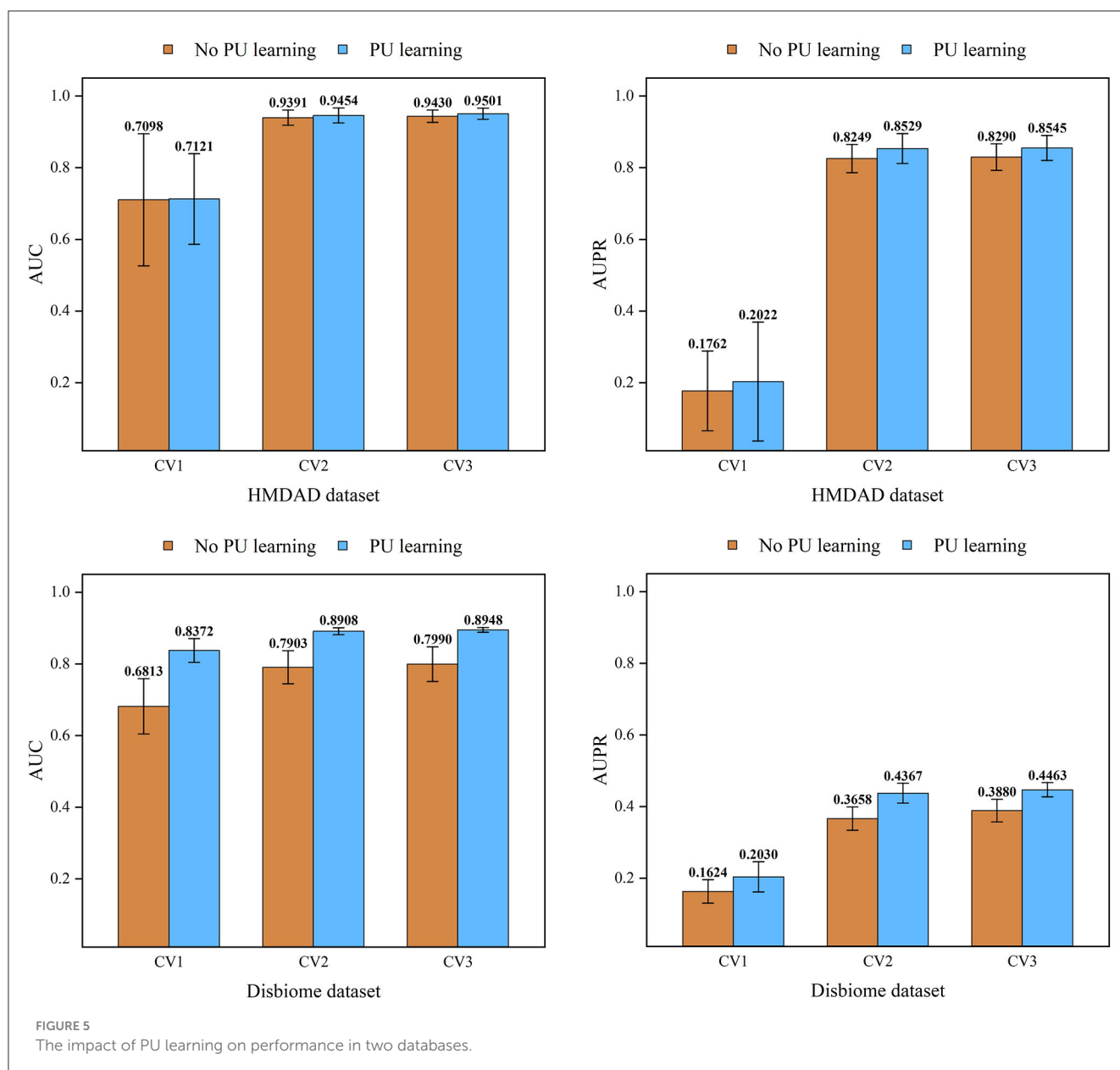
3.6.1. Identifying new microbes for asthma

Asthma is a heterogeneous disease with respect to respiratory symptoms including chest tightness, shortness of breath, wheeze, and cough. It is the most common chronic respiratory condition and affects ~339 million people worldwide. Approximately 5%–10% of these patients have severe asthma. More than 10% of adults and 2.5% of children suffered from asthma have severe asthma (Brusselle and Koppelman, 2022; Reddel et al., 2022; Rattu et al., 2023).

We used the proposed GPUDMDA method to find new microbes associated with asthma. Tables 1, 2 show the predicted top 30 microbes that may associate with asthma on the HMDAD

and Disbiome databases. The predicted 30 asthma-associated microbes included microbes with known association information with asthma and microbes without association information with asthma on the two databases. As shown in Table 1, 23 and 29 microbes can be validated by each or both of two databases or existing literatures among the identified top 30 potential asthma-associated microbes on the two databases, respectively. Furthermore, we found that *Enterobacter hormaechei* could associate with asthma with the ranking of 15 on the HMDAD database. On the Disbiome database, GPUDMDA predicted that *Enterobacter* may be a sole and unknown asthma-associated microbe among the predicted top 30 microbes associated with asthma.

Enterobacter hormaechei (Yeh et al., 2022) is a member and the most common nosocomial pathogen of the *Enterobacter cloacae* complex. It plays a key role in infectious diseases including, urinary tract infections, pneumonia, biliary tract infections, bacteremia, colitis and cellulitis. It is commonly found to be a high-pathogenicity island on its chromosome and is more virulent



compared with other *E. cloacae* complex. In this study, GPUDMDA identified that *E. hormaechei* could associate with asthma.

Figure 6 shows the association network between the predicted top 53 asthma-associated microbes and asthma, after removing the repeated associations on the two databases. In Figure 6, the gray solid lines and blue dashed lines denote known associations between microbes and asthma and the predicted associations between microbes and asthma, respectively.

3.6.2. Identifying new microbes for inflammatory bowel disease

IBD is a class of globally chronic intestinal disease (Chang, 2020; Kaplan and Windsor, 2021). It widely exists in the gut and gastrointestinal tract and extraintestinal organs in many patients (Rogler et al., 2021). Up to 2 million Europeans and

1.5 million North Americans suffer from this disease (Jairath and Feagan, 2020). It mainly comprises Crohn's disease, ulcerative colitis, and indeterminate colitis (Flynn and Eisenstein, 2019). Many studies thought that it is the result of interactions between microbial, environmental, and immune-mediated factors. In particular, microbiome has been reported to have potential roles in the development, progression, and treatment of IBD. The gut microbiome is different in the IBD patients from one in healthy bodies (Glassner et al., 2020).

In particular, IBD is very common in children. Many pediatricians and the other pediatric clinicians meet children suffered from IBD. The IBD pediatric populations demonstrate the classic features of abdominal pain, bloody diarrhea, and weight loss as well as non-classic features of anemia, isolated poor growth, or the other extraintestinal symptoms. Recently, the IBD children patients show a rising incidence (Rosen et al., 2015; Oliveira and

TABLE 1 The predicted top 30 microbes associated with Asthma on HMDAD.

Rank	Microbe	Evidence
1	<i>Proteobacteria</i>	Confirmed by HMDAD and Disbiome
2	<i>Prevotella</i>	Confirmed by HMDAD and Disbiome
3	<i>Staphylococcus</i>	Confirmed by HMDAD and Disbiome
4	<i>Bacteroidetes</i>	Confirmed by HMDAD and Disbiome
5	<i>Enterobacteriaceae</i>	Confirmed by Disbiome
6	<i>Clostridium coccoides</i>	PMID: 1477358
7	<i>Firmicutes</i>	PMID: 23265859
8	<i>Comamonadaceae</i>	Confirmed by HMDAD and Disbiome
9	<i>Oxalobacteraceae</i>	Confirmed by HMDAD and Disbiome
10	<i>Sphingomonadaceae</i>	Confirmed by HMDAD and Disbiome
11	<i>Haemophilus</i>	Confirmed by HMDAD and Disbiome
12	<i>Helicobacter pylori</i>	Confirmed by HMDAD
13	<i>Enterococcus</i>	PMID: 29788027
14	<i>Enterobacter aerogenes</i>	PMID: 23842440
15	<i>Enterobacter hormaechei</i>	Unconfirmed
16	<i>Klebsiella pneumoniae</i>	PMID: 26953325
17	<i>Shigella dysenteriae</i>	Unconfirmed
18	<i>Lactobacillus</i>	Confirmed by Disbiome
19	<i>Clostridia</i>	PMID: 21477358
20	<i>Veillonella</i>	Confirmed by Disbiome
21	<i>Klebsiella</i>	Confirmed by Disbiome
22	<i>Prevotella copri</i>	Unconfirmed
23	<i>Actinobacteria</i>	PMID: 28947029
24	<i>Shuttleworthia</i>	Unconfirmed
25	<i>Desulfovibrio</i>	PMID: 29198875
26	<i>Clostridium difficile</i>	PMID: 21872915
27	<i>Oxalobacter formigenes</i>	Unconfirmed
28	<i>Fusobacteria</i>	Unconfirmed
29	<i>Porphyromonadaceae</i>	Confirmed by Disbiome
30	<i>Verrucomicrobiaceae</i>	Unconfirmed

TABLE 2 The predicted top 30 microbes associated with Asthma on Disbiome.

Rank	Microbe	Evidence
1	<i>Actinomyces</i>	Confirmed by Disbiome
2	<i>Bacteroides stercoris</i>	Confirmed by Disbiome
3	<i>Bifidobacterium</i>	Confirmed by Disbiome
4	<i>Blautia</i>	Confirmed by Disbiome
5	<i>Clostridiaceae</i>	Confirmed by Disbiome
6	<i>Clostridium neonatale</i>	Confirmed by Disbiome
7	<i>Comamonadaceae</i>	Confirmed by HMDAD and Disbiome
8	<i>Corynebacterium</i>	Confirmed by Disbiome
9	<i>Faecalibacterium</i>	Confirmed by Disbiome
10	<i>Gallibacterium</i>	Confirmed by Disbiome
11	<i>Gammaproteobacteria</i>	Confirmed by Disbiome
12	<i>Gemella</i>	Confirmed by Disbiome
13	<i>Klebsiella</i>	Confirmed by Disbiome
14	<i>Leclercia</i>	Confirmed by Disbiome
15	<i>Moraxella</i>	Confirmed by Disbiome
16	<i>Neisseria</i>	Confirmed by Disbiome
17	<i>Nitrosomonadaceae</i>	Confirmed by Disbiome
18	<i>Oxalobacteraceae</i>	Confirmed by HMDAD and Disbiome
19	<i>Planococcaceae</i>	Confirmed by Disbiome
20	<i>Prevotella</i>	Confirmed by HMDAD and Disbiome
21	<i>Pseudomonadaceae</i>	Confirmed by Disbiome
22	<i>Sphingomonadaceae</i>	Confirmed by HMDAD and Disbiome
23	<i>Staphylococcus</i>	Confirmed by HMDAD and Disbiome
24	<i>Stenotrophomonas</i>	Confirmed by Disbiome
25	<i>Streptococcus</i>	Confirmed by Disbiome
26	<i>Sutterella wadsworthensis</i>	Confirmed by Disbiome
27	<i>Veillonella</i>	Confirmed by Disbiome
28	<i>Weeksella</i>	Confirmed by Disbiome
29	<i>Propionibacterium</i>	PMID: 13268970
30	<i>Enterobacter</i>	Unconfirmed

Monteiro, 2017). In total, 25%–30% of patients with Crohn’s disease and 20% of patients with ulcerative colitis have been diagnosed in <20 years of age. Moreover, 4% of pediatric IBD patients have been detected before 5 years (Kelsen and Baldassano, 2008). IBD severely affects normal growth and development of children. When treating children with newly diagnosed IBD, we need to consider their affects on growth and development and bone health (Rosen et al., 2015).

In this manuscript, we used the proposed GPUDMDA method to find potential microbes associated with IBD. Tables 3, 4 show the predicted top 30 IBD-associated microbes on the two MDA databases. The predicted 30 IBD-associated microbes included microbes with known association information with IBD and

microbes without association information with IBD. In total, 20 and 28 predicted IBD-associated microbes can be validated by databases or existing publications among all predicted top 30 microbes on the two databases, respectively. On HMDAD, GPUDMDA predicted that *E. hormaechei* could associate with IBD with the ranking of 7. On Disbiome, the former 28 microbes have been confirmed to associate with IBD, and GPUDMDA also identified that *E. hormaechei* could link with IBD with the ranking of 29.

Figure 7 shows the association network between the predicted top 54 IBD-associated microbes and IBD, after removing the repeated associations on the two databases. In Figure 7, the gray solid lines and blue dashed lines denote known associations



4. Discussion and conclusion

In this study, we developed a deep learning model called GPU MDMA to capture new linkages between microbes and various human complex diseases. GPU MDMA first computed disease similarity and microbe similarity matrices based on their functional similarity and GIPK similarity, respectively. Next, it extracted features for each microbe–disease pair with GATE. Third, it selected a few reliable negative MDAs based on PU learning with K -means clustering and XGBoost. Finally, it took the extracted MDA features and the selected negative MDAs as inputs and designed a DNN to predict potential MDAs.

Frontiers in Microbiology

TABLE 3 The predicted top 30 microbes associated with IBD on HMDAD.

Rank	Microbe	Evidence
1	<i>Clostridium difficile</i>	PMID: 27499718
2	<i>Helicobacter pylori</i>	PMID: 22221289
3	<i>Staphylococcus</i>	PMID: 27239107
4	<i>Clostridia</i>	PMID: 31142855
5	<i>Clostridium coccooides</i>	PMID: 19235886
6	<i>Enterobacter aerogenes</i>	PMID: 4061480
7	<i>Enterobacter hormaechei</i>	Unconfirmed
8	<i>Klebsiella pneumoniae</i>	PMID: 9930068
9	<i>Shigella dysenteriae</i>	Unconfirmed
10	<i>Prevotella copri</i>	Unconfirmed
11	<i>Enterococcus</i>	PMID: 24629344
12	<i>Klebsiella</i>	PMID: 29573336
13	<i>Actinobacteria</i>	Confirmed by HMDAD
14	<i>Bifidobacterium</i>	PMID: 24478468
15	<i>Dietzia maris</i>	Unconfirmed
16	<i>Staphylococcus epidermidis</i>	Unconfirmed
17	<i>Oxalobacter formigenes</i>	Unconfirmed
18	<i>Tropheryma whipplei</i>	Unconfirmed
19	<i>Staphylococcus aureus</i>	PMID: 11424320
20	<i>Bacteroides vulgatus</i>	PMID: 29454108
21	<i>Actinomyces</i>	PMID: 30545401
22	<i>Porphyromonas gingivalis</i>	PMID: 31652577
23	<i>Selenomonas</i>	Unconfirmed
24	<i>Treponema</i>	PMID: 31851086
25	<i>Fusobacterium nucleatum</i>	PMID: 26718210
26	<i>Bacteroides ovatus</i>	PMID: 30666959
27	<i>Verrucomicrobiaceae</i>	PMID: 22572638
28	<i>Desulfovibrio</i>	Confirmed by Disbiome
29	<i>Clostridiales</i>	Unconfirmed
30	<i>Escherichia coli</i>	PMID: 29573336

TABLE 4 The predicted top 30 microbes associated with IBD on Disbiome.

Rank	Microbe	Evidence
1	<i>Anaerostipes</i>	Confirmed by Disbiome
2	<i>Bacillus licheniformis</i>	Confirmed by Disbiome
3	<i>Blautia</i>	Confirmed by Disbiome
4	<i>Bradyrhizobiaceae</i>	Confirmed by Disbiome
5	<i>Butyrivimonas</i>	Confirmed by Disbiome
6	<i>Comamonadaceae</i>	Confirmed by Disbiome
7	<i>Christensenellaceae</i>	Confirmed by Disbiome
8	<i>Dehalobacter</i>	Confirmed by Disbiome
9	<i>Desulfovibrio</i>	Confirmed by Disbiome
10	<i>Dorea formicigenerans</i>	Confirmed by Disbiome
11	<i>Eubacterium bifforme</i>	Confirmed by Disbiome
12	<i>Gemella</i>	Confirmed by Disbiome
13	<i>Gluconobacter oxydans</i>	Confirmed by Disbiome
14	<i>Lachnobacterium</i>	Confirmed by Disbiome
15	<i>Methanobrevibacter smithii</i>	Confirmed by Disbiome
16	<i>Mogibacterium</i>	Confirmed by Disbiome
17	<i>Moraxellaceae</i>	Confirmed by Disbiome
18	<i>Pseudomonas straminea</i>	Confirmed by Disbiome
19	<i>Ruminococcus bromii</i>	Confirmed by Disbiome
20	<i>Saccharomyces cerevisiae</i>	Confirmed by Disbiome
21	<i>Streptococcus anginosus</i>	Confirmed by Disbiome
22	<i>Bacteroides ovatus</i>	PMID: 30666959
23	<i>Enterobacter aerogenes</i>	PMID: 4061480
24	<i>Fusobacterium</i>	PMID: 25307765
25	<i>Klebsiella pneumoniae</i>	PMID: 9930068
26	<i>Paraprevotella</i>	PMID: 25307765
27	<i>Propionibacterium acnes</i>	PMID: 28630242
28	<i>Staphylococcus</i>	PMID: 27239107
29	<i>Oxalobacteraceae</i>	Unconfirmed
30	<i>Enterobacter hormaechei</i>	Unconfirmed

and NTSHMDA) on the HMDAD and Disbiome databases under five-fold CVs on microbes, diseases, and microbe-disease pairs. Under the three CVs, GPUDMDA computed the best AUCs and AUPRs on the two databases, suggesting that GPUDMDA could improve MDA prediction performance. Finally, we implemented case studies for asthma and IBD. The results showed that *E. hormaechei* could densely associate with asthma and IBD and need further biological experimental validation.

In future, we will combine biological features of microbe, diseases, and MDA network to design more accurate negative MDA selection method. In addition, we will also develop

novel deep learning model to improve MDA classification performance based on the selected reliable negative MDA samples. Interestingly, we have conducted several computational models including existing classical MDA prediction methods. But the results elucidated that many models failed to compute better AUPR on the Disbiome database. It may be caused by different data structures of Disbiome. In future, we will further design a better robust computational method to improve MDA prediction on the Disbiome database. We hope that the proposed GPUDMDA method helps to identify microbes associated with related diseases and further contributes to mining the clues of treatment.



Youth Project of Hunan Provincial Education Department Grant No. 21B0802.

Acknowledgments

The authors would like to thank all the authors of the cited references.

Conflict of interest

GT and YW were employed by Geneis (Beijing) Co. Ltd.

References

- Abdu-Aljabar, R. D., and Awad, O. A. (2021). "A comparative analysis study of lung cancer detection and relapse prediction using XGBoost classifier," in *IOP Conference Series: Materials Science and Engineering*, volume 1076 (Bristol: IOP Publishing), 012048. doi: 10.1088/1757-899X/1076/1/012048
- Brusselle, G. G., and Koppelman, G. H. (2022). Biologic therapies for severe asthma. *N. Engl. J. Med.* 386, 157–171. doi: 10.1056/NEJMra2032506
- Chang, J. T. (2020). Pathophysiology of inflammatory bowel diseases. *N. Engl. J. Med.* 383, 2652–2664. doi: 10.1056/NEJMra2002697
- Chen, X., and Huang, L. (2023). Computational model for disease research. *Brief. Bioinformatics* 24, bbac615. doi: 10.1093/bib/bbac615
- Chen, X., Li, T.-H., Zhao, Y., Wang, C.-C., and Zhu, C.-C. (2021). Deep-belief network for predicting potential mirna-disease associations. *Brief. Bioinformatics* 22, bbaa186. doi: 10.1093/bib/bbaa186
- Chen, X., Liu, H., and Zhao, Q. (2020). Editorial: Bioinformatics in microbiota. *Front. Microbiol.* 11, 100. doi: 10.3389/fmicb.2020.00100
- Chen, X., Xie, D., Zhao, Q., and You, Z.-H. (2019). MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinformatics* 20, 515–539. doi: 10.1093/bib/bbx130
- Demirci, M., Tokman, H., Uysal, H., Demiryas, S., Karakullukcu, A., Saribas, S., et al. (2019). Reduced akkermansia muciniphila and faecalibacterium prausnitzii levels in the gut microbiota of children with allergic asthma. *Allergol. Immunopathol.* 47, 365–371. doi: 10.1016/j.aller.2018.12.009
- Deng, L., Liu, Z., Qian, Y., and Zhang, J. (2022). Predicting circrna-drug sensitivity associations via graph attention auto-encoder. *BMC Bioinformatics* 23, 1–15. doi: 10.1186/s12859-022-04694-y
- El Mouzan, M. I., Winter, H. S., Assiri, A. A., Korolev, K. S., Al Sarkhy, A. A., Dowd, S. E., et al. (2018). Microbiota profile in new-onset pediatric crohn's disease: data from a non-western population. *Gut Pathog.* 10, 1–10. doi: 10.1186/s13099-018-0276-3
- Flynn, S., and Eisenstein, S. (2019). Inflammatory bowel disease presentation and diagnosis. *Surg. Clin.* 99, 1051–1062. doi: 10.1016/j.suc.2019.08.001
- Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355–1359. doi: 10.1126/science.1124234
- Glassner, K. L., Abraham, B. P., and Quigley, E. M. (2020). The microbiome and inflammatory bowel disease. *J. Allergy Clin. Immunol.* 145, 16–27. doi: 10.1016/j.jaci.2019.11.003
- Guan, J., Zhang, Z. G., Liu, Y., and Wang, M. (2022). A novel bi-directional heterogeneous network selection method for disease and microbial association prediction. *BMC Bioinformatics* 23, 1–15. doi: 10.1186/s12859-022-04961-y
- Henao-Mejia, J., Elinav, E., Thaiss, C. A., Licona-Limon, P., and Flavell, R. A. (2013). Role of the intestinal microbiome in liver disease. *J. Autoimmun.* 46, 66–73. doi: 10.1016/j.jaut.2013.07.001
- Holmes, E., Wijeyesekera, A., Taylor-Robinson, S. D., and Nicholson, J. K. (2015). The promise of metabolic phenotyping in gastroenterology and hepatology. *Nat. Rev. Gastroenterol. Hepatol.* 12, 458–471. doi: 10.1038/nrgastro.2015.114
- Hu, H., Feng, Z., Lin, H., Cheng, J., Lyu, J., Zhang, Y., et al. (2023a). Gene function and cell surface protein association analysis based on single-cell multiomics data. *Comput. Biol. Med.* 157, 106733. doi: 10.1016/j.compbiomed.2023.106733
- Hu, H., Feng, Z., Lin, H., Zhao, J., Zhang, Y., Xu, F., et al. (2023b). Modeling and analyzing single-cell multimodal data with deep parametric inference. *Brief. Bioinformatics* 24, bbad005. doi: 10.1093/bib/bbad005
- Jairath, V., and Feagan, B. G. (2020). Global burden of inflammatory bowel disease. *Lancet Gastroenterol. Hepatol.* 5, 2–3. doi: 10.1016/S2468-1253(19)30358-9
- Janssens, Y., Nielandt, J., Bronselaer, A., Debonne, N., Verbeke, F., Wynendaele, E., et al. (2018). Disbiome database: linking the microbiome to disease. *BMC Microbiol.* 18, 1–6. doi: 10.1186/s12866-018-1197-5
- Jiang, C., Tang, M., Jin, S., Huang, W., and Liu, X. (2022). KGNMDA: a knowledge graph neural network method for predicting microbe-disease associations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 20, 1147–1115. doi: 10.1109/TCBB.2022.3184362
- Kamneva, O. K. (2017). Genome composition and phylogeny of microbes predict their co-occurrence in the environment. *PLoS Comput. Biol.* 13, e1005366. doi: 10.1371/journal.pcbi.1005366
- Kaplan, G. G., and Windsor, J. W. (2021). The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* 18, 56–66. doi: 10.1038/s41575-020-00360-x
- Kelsen, J., and Baldassano, R. N. (2008). Inflammatory bowel disease: the difference between children and adults. *Inflamm. Bowel Dis.* 14(suppl_2), S9–S11. doi: 10.1002/ibd.20560
- Li, F., Dong, S., Leier, A., Han, M., Guo, X., Xu, J., et al. (2022). Positive-unlabeled learning in bioinformatics and computational biology: a brief review. *Brief. Bioinformatics* 23, bbab461. doi: 10.1093/bib/bbab461
- Li, H., Wang, Y., Zhang, Z., Tan, Y., Chen, Z., Wang, X., et al. (2020). "Identifying microbe-disease association based on a novel back-propagation neural network model," in *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 18, 2502–2513. doi: 10.1109/TCBB.2020.2986459
- Lihong, P., Wang, C., Tian, X., Zhou, L., and Li, K. (2022). Finding lncrna-protein interactions based on deep learning with dual-net neural architecture. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 19, 3456–3468. doi: 10.1109/TCBB.2021.3116232
- Liu, D., Liu, J., Luo, Y., He, Q., and Deng, L. (2021). Mgatmda: predicting microbe-disease associations via multi-component graph attention network. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 19, 3578–3585. doi: 10.1109/TCBB.2021.3116318
- Liu, H., Bing, P., Zhang, M., Tian, G., Ma, J., Li, H., et al. (2023). MNNMDA: predicting human microbe-disease association via a method to minimize matrix nuclear norm. *Comput. Struct. Biotechnol. J.* 21, 1414–1423. doi: 10.1016/j.csbj.2022.12.053
- Liu, Y., Wang, S.-L., Zhang, J.-F., Zhang, W., Zhou, S., Li, W., et al. (2020). DMFMda: prediction of microbe-disease associations based on deep matrix factorization using bayesian personalized ranking. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 18, 1763–1772. doi: 10.1109/TCBB.2020.3018138
- Long, Y., and Luo, J. (2019). WMGHMDA: a novel weighted meta-graph-based model for predicting human microbe-disease association on heterogeneous information network. *BMC Bioinformatics* 20, 1–18. doi: 10.1186/s12859-019-3066-0
- Long, Y., Luo, J., Zhang, Y., and Xia, Y. (2021). Predicting human microbe-disease associations via graph attention networks with inductive matrix completion. *Brief. Bioinformatics* 22, bbaa146. doi: 10.1093/bib/bbaa146
- Luo, J., and Long, Y. (2018). Ntshmda: prediction of human microbe-disease association based on random walk by integrating network topological

- similarity. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 17, 1341–1351. doi: 10.1109/TCBB.2018.2883041
- Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2017). An analysis of human microbe-disease associations. *Brief. Bioinformatics* 18, 85–97. doi: 10.1093/bib/bbw005
- Ma, Y., and Jiang, H. (2020). Ninimhmda: neural integration of neighborhood information on a multiplex heterogeneous network for multiple types of human microbe-disease association. *Bioinformatics* 36, 5665–5671. doi: 10.1093/bioinformatics/btaa1080
- Niu, Y.-W., Qu, C.-Q., Wang, G.-H., and Yan, G.-Y. (2019). Rwhmda: random walk on hypergraph for microbe-disease association prediction. *Front. Microbiol.* 10, 1578. doi: 10.3389/fmicb.2019.01578
- Oliveira, S. B., and Monteiro, I. M. (2017). Diagnosis and management of inflammatory bowel disease in children. *BMJ* 357, j2083. doi: 10.1136/bmj.j2083
- Peng, C.-X., Zhou, X.-G., Xia, Y.-H., Liu, J., Hou, M.-H., Zhang, G.-J., et al. (2022a). Structural analogue-based protein structure domain assembly assisted by deep learning. *Bioinformatics* 38, 4513–4521. doi: 10.1093/bioinformatics/btac553
- Peng, L., Shen, L., Liao, L., Liu, G., and Zhou, L. (2020). RNMFMDA: a microbe-disease association identification method based on reliable negative sample selection and logistic matrix factorization with neighborhood regularization. *Front. Microbiol.* 11, 592430. doi: 10.3389/fmicb.2020.592430
- Peng, L., Tan, J., Xiong, W., Zhang, L., Wang, Z., Yuan, R., et al. (2023a). Deciphering ligand-receptor-mediated intercellular communication based on ensemble deep learning and the joint scoring strategy from single-cell transcriptomic data. *Comput. Biol. Med.* 163, 107137. doi: 10.1016/j.combiomed.2023.107137
- Peng, L., Wang, C., Tian, G., Liu, G., Li, G., Lu, Y., et al. (2022c). Analysis of ct scan images for COVID-19 pneumonia based on a deep ensemble framework with densenet, swin transformer, and regnet. *Front. Microbiol.* 13, 995323. doi: 10.3389/fmicb.2022.995323
- Peng, L., Wang, F., Wang, Z., Tan, J., Huang, L., Tian, X., et al. (2022d). Cell-cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: data resources and computational strategies. *Brief. Bioinformatics* 23, bbac234. doi: 10.1093/bib/bbac234
- Peng, L., Yuan, R., Han, C., Han, G., Tan, J., Wang, Z., et al. (2023b). Cellenboost: a boosting-based ligand-receptor interaction identification model for cell-to-cell communication inference. *IEEE Trans. Nanobioscience*. doi: 10.1109/TNB.2023.3278685
- Peng, L., He, X., Zhang, L., Peng, X., Lu, Y., Li, Z., et al. (2022b). “A deep learning-based unsupervised learning method for spatially resolved transcriptomic data analysis,” in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Las Vegas, NV: IEEE), 281–286. doi: 10.1109/BIBM55620.2022.9995207
- Rattu, A., Khaleva, E., Brightling, C., Dahlén, S.-E., Bossios, A., Fleming, L., et al. (2023). Identifying and appraising outcome measures for severe asthma: a systematic review. *Eur. Respir. J.* 61, 2201231. doi: 10.1183/13993003.01231-2022
- Reddel, H. K., Bacharier, L. B., Bateman, E. D., Brightling, C. E., Brusselle, G. G., Buhl, R., et al. (2022). Global initiative for asthma strategy 2021: executive summary and rationale for key changes. *Am. J. Respir. Crit. Care Med.* 205, 17–35. doi: 10.1164/rccm.202109-2205PP
- Rogler, G., Singh, A., Kavanaugh, A., and Rubin, D. T. (2021). Extraintestinal manifestations of inflammatory bowel disease: current concepts, treatment, and implications for disease management. *Gastroenterology* 161, 1118–1132. doi: 10.1053/j.gastro.2021.07.042
- Rosen, M. J., Dhawan, A., and Saeed, S. A. (2015). Inflammatory bowel disease in children and adolescents. *JAMA Pediatr.* 169, 1053–1060. doi: 10.1001/jamapediatrics.2015.1982
- Schwabe, R. F., and Jobin, C. (2013). The microbiome and cancer. *Nat. Rev. Cancer* 13, 800–812. doi: 10.1038/nrc3610
- Shen, L., Liu, F., Huang, L., Liu, G., Zhou, L., Peng, L., et al. (2022). VDA-RWLRLS: an anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and laplacian regularized least squares. *Comput. Biol. Med.* 140, 105119. doi: 10.1016/j.combiomed.2021.105119
- Tian, X., Shen, L., Gao, P., Huang, L., Liu, G., Zhou, L., et al. (2022). Discovery of potential therapeutic drugs for COVID-19 through logistic matrix factorization with Kernel diffusion. *Front. Microbiol.* 13, 740382. doi: 10.3389/fmicb.2022.740382
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2017). Graph attention networks. *arXiv*. [preprint]. doi: 10.48550/arXiv.1710.10903
- Wang, F., Huang, Z.-A., Chen, X., Zhu, Z., Wen, Z., Zhao, J., et al. (2017). LRLSHMDA: Laplacian regularized least squares for human microbe-disease association prediction. *Sci. Rep.* 7, 7601. doi: 10.1038/s41598-017-08127-2
- Wang, L., Wang, Y., Li, H., Feng, X., Yuan, D., Yang, J., et al. (2019). A bidirectional label propagation based computational model for potential microbe-disease association prediction. *Front. Microbiol.* 10, 684. doi: 10.3389/fmicb.2019.0684
- Wang, L., You, Z.-H., Huang, Y.-A., Huang, D.-S., and Chan, K. C. (2020). An efficient approach based on multi-sources information to predict circrna-disease associations using deep convolutional neural network. *Bioinformatics* 36, 4038–4046. doi: 10.1093/bioinformatics/bt z825
- Wang, T., Sun, J., and Zhao, Q. (2023). Investigating cardiotoxicity related with herg channel blockers using molecular fingerprints and graph attention mechanism. *Comput. Biol. Med.* 153, 106464. doi: 10.1016/j.combiomed.2022.106464
- Wang, W., Zhang, L., Sun, J., Zhao, Q., and Shuai, J. (2022). Predicting the potential human lncrna-mirna interactions based on graph convolution network with conditional random field. *Brief. Bioinformatics* 23, bbac463. doi: 10.1093/bib/bbac463
- Wen, L., Ley, R. E., Volchkov, P. Y., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., et al. (2008). Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* 455, 1109–1113. doi: 10.1038/nature07336
- Wen, Z., Yan, C., Duan, G., Li, S., Wu, F.-X., Wang, J., et al. (2021). A survey on predicting microbe-disease associations: biological data and computational methods. *Brief. Bioinformatics* 22, bbai157. doi: 10.1093/bib/bbaa157
- Wu, H., Wu, Y., Jiang, Y., Zhou, B., Zhou, H., Chen, Z., et al. (2022). SCHICSTACKL: a stacking ensemble learning-based method for single-cell Hi-C classification using cell embedding. *Brief. Bioinformatics* 23, bbab396. doi: 10.1093/bib/bbab396
- Xu, J., Xu, J., Meng, Y., Lu, C., Cai, L., Zeng, X., et al. (2023). Graph embedding and gaussian mixture variational autoencoder network for end-to-end analysis of single-cell RNA sequencing data. *Cell Rep. Methods* 3, 100382. doi: 10.1016/j.crmeth.2022.100382
- Yan, C., Duan, G., Wu, F.-X., Pan, Y., and Wang, J. (2019). BRWMDA: predicting microbe-disease associations based on similarities and bi-random walk on disease and microbe networks. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 17, 1595–1604. doi: 10.1109/TCBB.2019.2907626
- Yeh, T.-K., Lin, H.-J., Liu, P.-Y., Wang, J.-H., and Hsueh, P.-R. (2022). Antibiotic resistance in *Enterobacter hormaechei*. *Int. J. Antimicrob. Agents* 60, 106650. doi: 10.1016/j.ijantimicag.2022.106650
- Zhang, P., Wu, Y., Zhou, H., Zhou, B., Zhang, H., Wu, H., et al. (2022a). CLNN-loop: a deep learning model to predict CTCF-mediated chromatin loops in the different cell lines and CTCF-binding sites (CBS) pair types. *Bioinformatics* 38, 4497–4504. doi: 10.1093/bioinformatics/btac575
- Zhang, P., Zhang, H., and Wu, H. (2022b). iPro-WAEL: a comprehensive and robust framework for identifying promoters in multiple species. *Nucleic Acids Res.* 50, 10278–10289. doi: 10.1093/nar/gkac824
- Zhang, P., and Wu, H. (2023). ICHROM-deep: an attention-based deep learning model for identifying chromatin interactions. *IEEE J. Biomed. Health Inf.* doi: 10.1109/JBHI.2023.3292299