Check for updates

# Identifying microbe-disease association based on graph convolutional attention network: Case study of liver cirrhosis and epilepsy

Kai Shi[1,2]*, Lin Li[1], Zhengfeng Wang[1], Huazhou Chen[3], Zilin Chen[4] and Shuanfeng Fang[5]*

[1]College of Information Science and Engineering, Guilin University of Technology, Guilin, China, [2]Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin University of Technology, Guilin, China, [3]College of Science, Guilin University of Technology, Guilin, China, [4]Department of Developmental and Behavioural Pediatric Department & Department of Child Primary Care, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5]Department of Children Health Care, Children's Hospital Affiliated to Zhengzhou University, Zhengzhou, China

The interactions between the microbiota and the human host can affect the physiological functions of organs (such as the brain, liver, gut, etc.). Accumulating investigations indicate that the imbalance of microbial community is closely related to the occurrence and development of diseases. Thus, the identification of potential links between microbes and diseases can provide insight into the pathogenesis of diseases. In this study, we propose a deep learning framework (MDAGCAN) based on graph convolutional attention network to identify potential microbe-disease associations. In MDAGCAN, we first construct a heterogeneous network consisting of the known microbe-disease associations and multi-similarity fusion networks of microbes and diseases. Then, the node embeddings considering the neighbor information of the heterogeneous network are learned by applying graph convolutional layers and graph attention layers. Finally, a bilinear decoder using node embedding representations reconstructs the unknown microbe-disease association. Experiments show that our method achieves reliable performance with average AUCs of 0.9778 and 0.9454 ± 0.0038 in the frameworks of Leave-one-out cross validation (LOOCV) and 5-fold cross validation (5-fold CV), respectively. Furthermore, we apply MDAGCAN to predict latent microbes for two high-risk human diseases, i.e., liver cirrhosis and epilepsy, and results illustrate that 16 and 17 out of the top 20 predicted microbes are verified by published literatures, respectively. In conclusion, our method displays effective and reliable prediction performance and can be expected to predict unknown microbe-disease associations facilitating disease diagnosis and prevention.

KEYWORDS

gut-liver-brain axis, microbe-disease associations, similarity network, graph convolutional network, graph attention network, liver cirrhosis, epilepsy

# 1. Introduction

Microbes are mainly categorized as bacteria, fungi, archaea and viruses, which inhabit all parts of the human body, but the greatest number of microbes are found in the gut (Blum, 2017; Kitamoto et al., 2020; de Vos et al., 2022). Gut microbiota plays an important role in regulating host physiological processes (e.g., immunity and metabolism), and its ecological disorders are closely related to the brain, liver and other organs (Gonzalez-Ochoa et al., 2017; Tooley, 2020; Gabanyi et al., 2022). Recently, increasing medical studies reported that the gut-liver-brain axis plays a fundamental role in the pathogenesis of various diseases (Won et al., 2022), which is the bidirectional relationship between the gut and its microbiota, the liver, and the brain. Besides, gut microbiota exert their actions at different levels of the gut-liver-brain axis, impacting disease progression via changing gut-liver-brain axis communication (Fuenzalida et al., 2021). For example, liver cirrhosis is a common chronic progressive liver disease with high mortality caused by one or more factors, such as alcohol, metabolic disorders, drugs and so on (Gan et al., 2022). Researchers found out that the gut microbiota is a key factor in the progression of chronic liver disease, while the gut microbiota (e.g., *Enterococcus* and *Escherichia coli*) in patients with liver cirrhosis has significant changes compared to healthy individuals (Hussain et al., 2020; Ren et al., 2021). Moreover, *Escherichia coli* can produce an active amino acid GABA through the metabolic pathway (Altaib et al., 2022), which can activate glucose metabolism in the brain, improve brain function and impact epileptic seizures via the genetic pathway (Feng et al., 2022). Epilepsy is another of the third most common chronic neurological disorder worldwide, which usually suffers from depression, anxiety, obsessive-compulsive disorder, migraine and other disorders (Löscher et al., 2020). Many underlying disease mechanisms can lead to epilepsy, and the cause of the disease remains unknown. Research results have revealed that intestinal microbial imbalance can impact the occurrence of epilepsy due to the close relationship between the central nervous system and the gastrointestinal tract (Al-Beltagi and Saeed, 2022). For instance, serotonin produced by Enterococcus is a neurotransmitter in the central and peripheral nervous systems and has a certain inhibitory effect on the seizure of epilepsy (Deidda et al., 2021). Hence, studying disease-associated microbes not only advances the understanding of their pathogenesis, but also provides many new medical strategies for diseases. However, traditional biological experiments are difficult to meet the requirements of biomedical research owing to complex processes and expensive cost. Therefore, it is essential to develop efficient new prediction algorithms for microbe-disease association prediction.

Current computational methods for microbe-disease association prediction can be primarily classified as path-based methods, network-based methods and feature learning methods. Path-based methods usually calculate the microbe-disease association probability based on the number and weighted scores of various types of paths between two nodes. Chen et al. (2017) proposed the first computational method for microbe-disease association prediction based on the katz measure, which identified the microbe-disease correlation by calculating all paths of different lengths between microbes and diseases. Long and Luo (2019) calculated the probability score of microbe-disease pairs based on a weighted meta-graph search algorithm on a heterogeneous network to find possible microbe-disease associations. Network-based methods

infer prospective microbe-disease associations through information propagation in a heterogeneous network. Yin et al. (2020) employed the structural similarity information of biological entities of diseases and microbes, combining spatial projection and label propagation to predict unknown microbe-disease associations. Yang et al. (2021) designed a novel identification method based on multi-similarities bilinear matrix factorization to find possible microbe-disease associations on a heterogeneous network. Yin et al. (2022) used the multiple kernel learning method to fuse similarities of microbe and disease, and then used the label propagation method to make predictions for disease-related potential microbes. Feature learning methods automatically extract features or representations from data through the model, and then reconstruct new microbe-disease associations by the features. Li et al. (2020) raised a neural network approach based on the backpropagation of a modified hyperbolic tangent activation function to predict disease-related microbes. Wang et al. (2021) applied random walk and graph embedding algorithm LINE to preserve graph structure through first-order and second-order proximity and to learn the latent feature representations of microbes and diseases, afterward obtained new microbe-disease associations by refactoring the representation. Long et al. (2021) developed an embedding representation method based on inductive matrix completion and graph attention network to infer the possible associations between microbes and diseases. Although the previous methods have achieved prominent results, more effective methods still need to be developed to screen latent microbe-disease associations.

In this study, we propose a deep learning framework to predict microbe-disease association, which combines the graph convolutional network and the graph attention network. First, we construct an informative heterogeneous network composed of the known microbe-disease association network and integrated multi-similarity networks, which fuse the Gaussian kernel similarity network and functional similarity network of microbe and disease, respectively. Then, MDAGCAN learns the feature representation of each node with the information of its neighbors and itself in the heterogeneous network by multi-layer graph convolution. Subsequently, the node representations serve as the input of graph attention layers. In graph attention layers, the node representations learned from graph convolutional layers further are enhanced by aggregating the weighted sum of neighbors' information. Ultimately, the unknown microbe-disease associations are reconstructed by a bilinear decoder. In addition, our method compares with state-of-the-art methods on the datasets HMDAD and MASI and is applied to the prediction of associated microbes in liver cirrhosis and epilepsy. The results confirm that our model is effective and reliable for inferring potential microbe-disease associations.

# 2. Materials

## 2.1. Human microbe-disease associations

In this work, we download two public databases of known microbe-disease association HMDAD[1] (Ma et al., 2017) and MASI[2]

---

(Zeng et al., 2021). HMDAD is the most frequently utilized human microbe-disease association database containing 450 non-redundant associations between 292 microbes and 39 diseases, and MASI covers microbial composition changes in different types of diseases with 629 associations involving 123 microbes and 56 diseases. The detailed statistics of the two microbe-disease association datasets above are exhibited in Table 1.

The microbe-disease association is represented as a binary adjacent matrix $\mathbf{A} \in \mathbb{R}^{nd \times nm}$, where $\mathbf{A_{ij}} = 1$ if there is an interaction between disease $d_i$ and microbe $m_i$, otherwise $\mathbf{A_{ij}} = 0$.

# 3. Methods

As shown in the flowchart of MDAGCAN (Figure 1), we introduce a graph convolutional attention network model to identify latent microbe-disease associations, which combines the graph convolutional network and graph attention network. MDAGCAN works in three stages to make predictions. Firstly, we construct a heterogeneous network consisting of a known microbe-disease association network, an integrated disease similarity network, and an integrated microbe similarity network. Secondly, latent representations of microbes and diseases are encoded and learned by graph convolutional layers and graph attention layers. Finally, MDAGCAN leverages a bilinear decoder to obtain the final association scores of microbe-disease pairs.

## 3.1. Similarity computation

### 3.1.1. Gaussian interaction profile kernel similarity for microbe and disease

We calculate the Gaussian interaction profile kernel similarity of microbes according to the assumption that microbes with similar functions are more likely trend to connect similar diseases (Long et al., 2021). First, we present $GIP(m_i)$ as the interaction profile of the specific microbe $m_i$, where it indicates the $i$th column of adjacent matrix $\mathbf{A}$. Then, the Gaussian interaction profile kernel similarity $KM(m_i, m_j)$ between microbe $m_i$ and $m_j$ can be defined as follows:

$$KM(m_i, m_j) = exp\left(-\lambda_m \|GIP(m_i) - GIP(m_j)\|^2\right) \quad (1)$$

where $\lambda_m$ indicates the normalized kernel bandwidth, the computation formula is below:

$$\lambda_m = \frac{\lambda'_m}{\frac{1}{nm}\sum_{t=1}^{nm}\|GIP(m_t)\|^2} \quad (2)$$

where $\lambda'_m$ is the original bandwidth and is usually set to 1.

Similarly, we derive the Gaussian interaction profile kernel similarity between disease pairs, and construct the disease Gaussian interaction profile kernel similarity matrix $KD \in \mathbb{R}^{nd \times nd}(0 \le KD(d_i, d_j) \le 1)$.

### 3.1.2. Microbe functional similarity

Microbe functional similarity is calculated using a similar approach to Kamneva (2017), capturing the interactions between proteins encoded in the genomes of two microbes. The protein-protein functional interaction network is retrieved from the STRING v11 database[3] to characterize the functional similarity of microbes by the similarity of microbial genomic proteins, and microbes with more common genes are more similar to each other. We use $FM(m_i, m_j)$ to denote the functional similarity between microbe $m_i$ and microbe $m_j$, where $FM \in \mathbb{R}^{nm \times nm}$.

### 3.1.3. Disease functional similarity

In this work, we calculate disease functional similarity based on functional associations between disease-related genes with the assumption that similar diseases tend to interact with similar genes (Wei and Liu, 2020). We utilize the HumanNet v2.0 database (Hwang et al., 2019) to access gene interactions, where each interaction has a log-likelihood score (LLS) assessing the probability of a functional association between genes. For disease $d_i$ and disease $d_j$, their functional similarity formula can be defined as follows:

$$FD(d_i, d_j) = \frac{\sum_{1 \le x \le m} FS_D^{G^b}(g_x^a) + \sum_{1 \le y \le n} FS_D^{G^a}(g_y^b)}{m+n} \quad (3)$$

where $FS_D^{G_b}(g_x^a) = max_{1 \le y \le n}\left(LLS\left(g_x^a, g_y^b\right)\right)$ indicates the maximum functional correlation score between a gene $g_x^a$ and a gene set $G^b = \left\{g_1^b, g_2^b, , g_n^b\right\}$, and similarly $FS_D^{G_a}(g_y^b) = max_{1 \le x \le m}\left(LLS\left(g_x^a, g_y^b\right)\right)$ expresses the maximum functional correlation score between a gene $g_y^b$ and a gene set $G^a = \left\{g_1^a, g_2^a, ..., g_m^a\right\}$. $LLS\left(g_x^a, g_y^b\right)$ is the normalization of the log-likelihood score. $G^a$ and $G^b$ are the gene sets associated with the disease $d_i$ and $d_j$, separately.

## 3.2. Different similarities integration

It is not easy to achieve functional similarities between all diseases and microbes due to incomplete biology information (i.e., disease-related genes and microbial genomic proteins). To further improve similarities for diseases and microbes, we design a new strategy to integrate Gaussian kernel similarity and functional similarity. Specifically, if there is no functional similarity $FM$ between microbe $m_i$ and $m_j$, the integrated similarity between $m_i$ and $m_j$ is defined as $GM$, otherwise, it is equal to the linear combination of microbe Gaussian interaction profile kernel similarity $GM$ and microbe functional similarity $FM$. Similarly, the integrated similarity of diseases can be calculated as follows:

$$MS(m_i, m_j) = \begin{cases} GM(m_i, m_j), & if\, FM(m_i, m_j) = 0 \\ \mu GM(m_i, m_j) + (1-\mu)FM(m_i, m_j), & otherwise \end{cases} \quad (4)$$

$$DS(d_i, d_j) = \begin{cases} GD(d_i, d_j), & if\, FD(d_i, d_j) = 0 \\ \mu GD(d_i, d_j) + (1-\mu)FD(d_i, d_j) & otherwise \end{cases} \quad (5)$$

TABLE 1 The overall statistics for the microbe-disease association dataset.

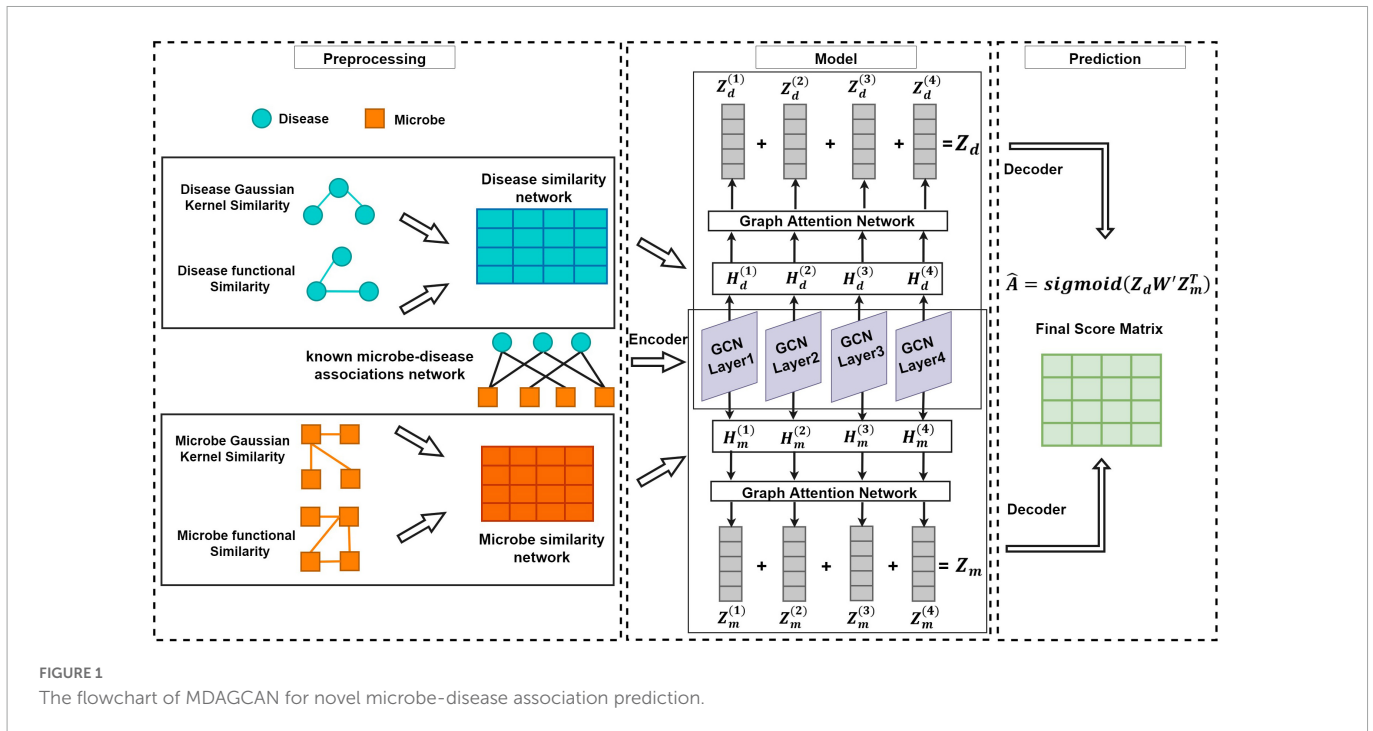| Dataset | Microbe | Disease | Associations |
|---------|---------|---------|--------------|
| HMDAD | 292 | 39 | 450 |
| MASI | 123 | 56 | 629 |

---

3   https://string-db.org

**FIGURE 1**
The flowchart of MDAGCAN for novel microbe-disease association prediction.

where μ is a control parameter for Gaussian similarity and functional similarity ranging from 0 to 1.

## 3.3. Graph convolutional network

In recent years, graph convolutional network as effective graph neural network model is widely applied in various fields with different tasks, such as node/graph classification, graph clustering and link prediction. The underlying idea of GCN is to learn node low-dimensional representations by aggregating node information from neighbors in a convolutional fashion while preserving graph structural information (Kipf and Welling, 2017; Zhang S. et al., 2019; Yue et al., 2020). Specifically, given a heterogeneous graph, the message propagation rule of GCN is expressed as:

$$H^{(l+1)} = f\left(H^{(l)}, G_{HN}\right) = tanh\left(D^{-\frac{1}{2}} G_{HN} D^{-\frac{1}{2}} H^{(l)} W_{GCN}^{(l)}\right) \quad (6)$$

where $H^{(l)}$ represents the node embedding at the $l$th layer, $W_{GCN}^{(l)}$ is the trainable weight matrix for the $l$th graph convolutional layer. $tanh$ is a nonlinear activation function. $D$ is the degree matrix of $G_{HN}$. $G_{HN} \in \mathbb{R}^{(nd+nm) \times (nd+nm)}$ is consisted of adjacent matrix $A$ and two similarity matrices ($G_{HN} = \begin{bmatrix} \beta DS^* & A \\ A^T & \beta MS^* \end{bmatrix}$). $DS^*$ and $MS^*$ are normalizations of $DS$ and $MS$, $\beta$ is a penalty factor used to control the contribution value of the similarity matrix in $G_{HN}$. The initialized embedding of the graph is denoted as $H^{(0)} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$.

## 3.4. Graph attention network

The graph attention network is another hot network architecture with the assumption that the node representation contributed from node neighbors is diverse (Veličković et al., 2018; Yu et al., 2021).

After performing graph convolutional operation, the node representations can be learned from the network structure. Thereafter, we introduce the graph attention layers to improve the node representations based on GAT, focusing on the contributions of import node neighbors for node representation learning. Specifically, there are two steps: achieving the attention distribution and averaging representations with the corresponding distribution. More definitions are described as follows:

$$e_{ij}^{(l)} = relu\left(\vec{a}^T \left[W_{GAT}^{(l)} h_i^{(l)} \| W_{GAT}^{(l)} h_j^{(l)}\right]\right) \quad (7)$$

$$Z_i^{(l)} = \sum_{j \in N_i} att_{ij}^{(l)} h_j^{(l)} \quad (8)$$

where $e_{ij}^{(l)}$ indicates the importance of node $j$ to node $i$ in the $l$th layer, $h_i^{(l)}$ is the node representations derived from the $l$th graph convolutional layer. $\|$ is the concatenation operation, $\vec{a}^T$ is a weight vector, $W_{GAT}^{(l)}$ is a shared weight matrix, $relu$ is a nonlinear activation function. $Z_i^{(l)}$ represents the representation of node $i$ by averaging representations of its neighbor nodes with normalized attention distribution. $e_{ij}^{(l)}$ is normalized as $att_{ij}^{(l)} = \frac{exp\left(e_{ij}^{(l)}\right)}{\sum_{c \in N_i} exp\left(e_{ic}^{(l)}\right)}$, $N_i$ is the neighborhood of node $i$ in the graph.

## 3.5. Decoder for microbe-disease association

We attain the learned feature representations $Z_m$ for microbes and $\mathbf{Z}_d$ for diseases from the output of GAT. Inspired by the work of Du et al. (2022), we reconstruct an association score matrix for microbe-disease associations (Equation 9) and define the local loss function which can dynamically reduce the weight of easily distinguished samples and make the distribution of loss function

balanced (Lin et al., 2020) (Equation 10).

$$\widehat{A}=sigmoid\left(Z_d W' Z_m^T\right) \tag{9}$$

$$\ell_{fl} = \sum_{(i,j)\in\Omega^+ \bigcup \Omega^-} \psi\left(\widehat{A}_{ij}, A_{ij}\right) \tag{10}$$

where $W'$ is a trainable matrix, *sigmoid* is a nonlinear activation function. $\Omega^+$ and $\Omega^-$ denote the positive and negative sample sets, respectively. Moreover, we adopt the focal loss function $\psi$ to solve the class imbalance. Focal loss (Lin et al., 2020) is based on binary cross-entropy and is a dynamically scaled cross-entropy loss.

$$\psi=\begin{cases} -\alpha\left(1-\widehat{A}_{ij}\right)^{\gamma} log\left(\widehat{A}_{ij}\right), if\, A_{ij}=1 \\ -\alpha\left(\widehat{A}_{ij}\right)^{\gamma} log\left(1-\widehat{A}_{ij}\right), otherwise \end{cases} \tag{11}$$

where $\alpha$ is a weight parameter that controls the class imbalance between positive and negative samples, and $\gamma$ is another weight parameter that controls the difficulty of sample classification. The Adam optimizer is used to minimize the loss (Kingma and Ba, 2015).

## 3.6. Parameter selection

There are several hyperparameters in MDAGCAN, such as the balance factor $\mu$, the penalty factor $\beta$, the embedding dimension $k$, the initial learning rate $lr$, two weight parameters $\alpha$ and $\gamma$ in focal loss, two dropout rates (node dropout $dp_n$ and regular dropout $dp_r$) and the iterations $epo$. These parameters consider different combinations from the ranges $\mu \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $\beta \in \{2, 4, 6, 8, 10\}$, $k \in \{32, 64, 128, 256\}$, $lr \in \{0.05, 0.005, 0.0005, 0.00005, 0.000005\}$, $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, $\gamma \in \{1, 2, 3, 4, 5\}$, $dp_n \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $dp_r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and $epo \in \{100, 200, 300, 400, 500, 600\}$. After adjusting, we set the optimal parameters $\mu = 0.5$, $\beta = 8$, $k = 64$, $lr = 0.00005$, $\alpha = 0.1$, $\gamma = 2$, $dp_n = 0.5$, $dp_r = 0.7$, and $epo = 500$ for MDAGCAN in the following experiments.

## 4. Results

## 4.1. Performance evaluation

Until now, many methods have been proposed to predict microbe-disease association. However, there are no consistent results and poor performance attributed to the single dataset usage and improper model adoption. In this paper, we conduct different experiments on two datasets to fairly compare our method with the existing methods. First, under the evaluation framework of LOOCV and 5-fold CV, we compare our method (MDAGCAN) with 10 baseline methods on HMDAD dataset, such as the katz measure-based model KATZHMDA (Chen et al., 2017), the random walk models BiRWMP (Shen et al., 2018), NTSHMDA (Luo and Long, 2020) and BRWMDA (Yan et al., 2020), the conventional machine learning model LRLSHMDA (Wang et al., 2017), the matrix decomposition model MDLPHMDA (Qu et al., 2019), the network-based models NBLPIHMDA (Wang et al., 2019) and NCPLP (Yin et al., 2020), the neural network model BPNNHMDA (Li et al., 2020) and GATMDA (Long et al., 2021). Under the evaluation framework of LOOCV and 5-fold CV, MDAGCAN obtains the highest AUC values

of 0.9778 and 0.9454, and has 4.25, 2.73% higher than the graph attention network method GATMDA, and 5.12, 1.29% better than the network consistency projection method NCPLP, respectively. All results are shown in **Figure 2**.

Besides, we perform the disease horizontal test, in which four-fifths of disease rows of the association matrix are randomly selected as the train set and the rest as the test set. Similarly, the microbe vertical test is also carried out in the columns of the association matrix. In the end, our method obtains AUC values of $0.8674 \pm 0.0175$ and $0.9290 \pm 0.0143$ on two tests, respectively. At the same time, we also compare MDAGCAN to other methods with different assessment metrics, such as F1 Score, Accuracy, Sensitivity and Specificity. More results are shown in **Tables 2, 3**. Obviously, the predictive effect of the microbe vertical test is better than the disease horizontal test due to the large degree difference of the disease node. When a disease with a large degree is used as the test set, the training set will contain less information, which will affect the prediction performance. The horizontal/vertical test suggests that our method achieves excellent performance, and is more suitable to predict new diseases and microbes.

In order to validate the robustness of methods, we perform contrast experiments on dataset MASI. The experimental results show that our method also reaches the best average AUC $(0.8730 \pm 0.0036)$, accuracy $(0.7996 \pm 0.0157)$ and specificity $(0.7691 \pm 0.0142)$ compared with the state-of-the-art methods (**Table 4**).

## 4.2. Predicting associated microbes for liver cirrhosis and epilepsy

Furthermore, we validate the prediction performance of MDAGCAN on two datasets HMDAD and MASI for two common diseases, i.e., liver cirrhosis and epilepsy. In this study, to identify the potential microbe-disease pairs, we remove all known microbe-disease associations, and select the top 20 microbes based on the ranking scores as the highly associated entities with the queried disease. Results show that 16 and 17 out of the top 20 predicted microbes for liver cirrhosis and epilepsy are verified by published literatures, respectively. Top-20 predicted candidate liver cirrhosis-related and epilepsy-related microbes also are listed in **Tables 5, 6**.

Liver cirrhosis is a common degenerative disease of the liver, caused by one or more factors such as genetics, viruses and drugs, and has a high mortality rate. In our prediction result, *Clostridium difficile* is the most associated with liver cirrhosis which is the top of the ranking list. *Clostridium difficile* infection is one of the factors leading to liver cirrhosis and is widely used to perform fecal microbial transplantation for treating liver cirrhosis (Olmedo et al., 2019). Meanwhile, Clostridiales ranked twelfth is generally considered to be beneficial bacteria, while Staphylococcus ranked fifth is the genus of pathogenic bacteria Staphylococcaceae (Bhat et al., 2016). Except for the microbes confirmed by literatures, we find four microbes, including *Clostridium coccoides*, Burkholderia, Betaproteobacteria, Bacteroides ovatus, which are not directly reported the association with liver cirrhosis. There is a report that *Clostridium coccoides* appears increased abundance in patients with nonalcoholic steatohepatitis (NASH), which leads to liver fibrosis and develops into liver cirrhosis. In other words, they may be the new biomarkers for liver cirrhosis (Mouzaki et al., 2013).
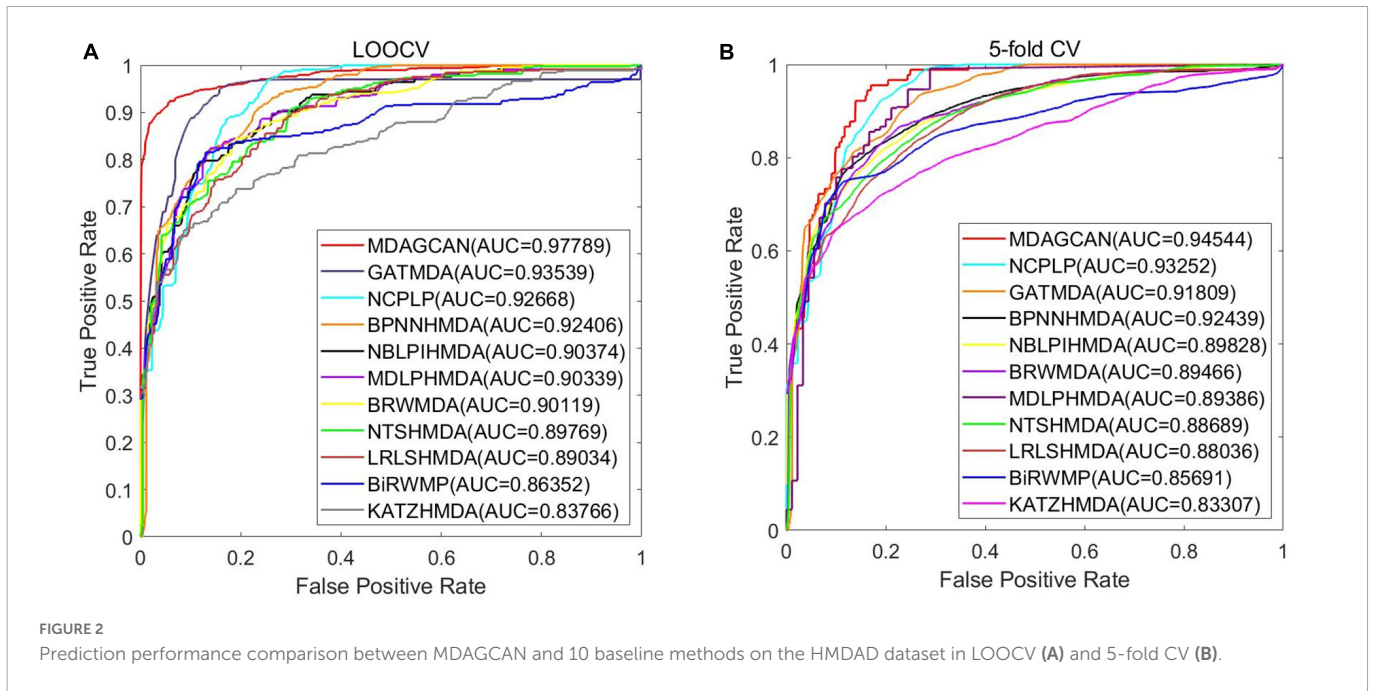
**FIGURE 2**
Prediction performance comparison between MDAGCAN and 10 baseline methods on the HMDAD dataset in LOOCV **(A)** and 5-fold CV **(B)**.

**TABLE 2**  Performance comparison between 10 baseline methods and MDAGCAN under horizontal test for diseases in 5-fold CV on HMDAD dataset.

| Methods | AUC | F1 Score | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KATZHMDA | $0.2625 \pm 0.0777$ | $0.5234 \pm 0.1151$ | $0.1649 \pm 0.0371$ | $0.3630 \pm 0.1117$ | $0.1636 \pm 0.0377$ |
| BiRWMP | $0.7345 \pm 0.0418$ | $\mathbf{0.8161 \pm 0.0789}$ | $\underline{0.7637 \pm 0.1030}$ | $0.6966 \pm 0.1100$ | $0.6936 \pm 0.1049$ |
| LRLSHMDA | $0.3794 \pm 0.1462$ | $0.5629 \pm 0.1338$ | $0.4029 \pm 0.3159$ | $0.4032 \pm 0.1266$ | $0.4022 \pm 0.3171$ |
| NTSHMDA | $0.4396 \pm 0.1082$ | $0.5032 \pm 0.1151$ | $0.4147 \pm 0.2086$ | $0.3434 \pm 0.0966$ | $0.4152 \pm 0.2090$ |
| BRWMDA | $0.3829 \pm 0.0825$ | $0.5769 \pm 0.3827$ | $0.3318 \pm 0.1231$ | $0.5114 \pm 0.4092$ | $0.3292 \pm 0.1256$ |
| MDLPHMDA | $0.4498 \pm 0.1240$ | $0.6403 \pm 0.1234$ | $0.3734 \pm 0.3990$ | $0.4833 \pm 0.1399$ | $0.3713 \pm 0.4017$ |
| NBLPIHMDA | $0.3846 \pm 0.1316$ | $0.5978 \pm 0.1496$ | $0.2481 \pm 0.1841$ | $0.4430 \pm 0.1602$ | $0.2468 \pm 0.1849$ |
| BPNNHMDA | $0.6166 \pm 0.1743$ | $0.7129 \pm 0.1619$ | $0.4321 \pm 0.1506$ | $0.6732 \pm 0.2292$ | $0.4289 \pm 0.1522$ |
| NCPLP | $\underline{0.8230 \pm 0.0372}$ | $\underline{0.7883 \pm 0.0088}$ | $0.7261 \pm 0.0552$ | $\mathbf{0.8771 \pm 0.0173}$ | $0.7252 \pm 0.0548$ |
| GATMDA | $0.4586 \pm 0.0195$ | $0.4647 \pm 0.0548$ | $0.7591 \pm 0.0509$ | $0.5050 \pm 0.0520$ | $\underline{0.7573 \pm 0.0523}$ |
| MDAGCAN | $\mathbf{0.8674 \pm 0.0175}$ | $0.7367 \pm 0.0865$ | $\mathbf{0.7826 \pm 0.0545}$ | $\underline{0.8539 \pm 0.0261}$ | $\mathbf{0.7810 \pm 0.0540}$ |

The best results are marked in bold and the second-best results are underlined.

**TABLE 3**  Performance comparison between 10 baseline methods and MDAGCAN under vertical test for microbes in 5-fold CV on HMDAD dataset.

| Methods | AUC | F1 Score | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KATZHMDA | $0.8756 \pm 0.0484$ | $0.8456 \pm 0.0263$ | $0.8641 \pm 0.0418$ | $0.7828 \pm 0.0423$ | $0.8645 \pm 0.0420$ |
| BiRWMP | $0.8993 \pm 0.0071$ | $0.8549 \pm 0.0579$ | $0.8177 \pm 0.1040$ | $0.8190 \pm 0.0987$ | $0.8159 \pm 0.1057$ |
| LRLSHMDA | $0.8465 \pm 0.0258$ | $0.8267 \pm 0.0499$ | $0.8964 \pm 0.0701$ | $0.7064 \pm 0.0561$ | $0.8979 \pm 0.0710$ |
| NTSHMDA | $0.8465 \pm 0.0258$ | $0.8430 \pm 0.0499$ | $0.8857 \pm 0.0742$ | $0.7318 \pm 0.0758$ | $0.8869 \pm 0.0754$ |
| BRWMDA | $0.8657 \pm 0.0309$ | $0.7985 \pm 0.0493$ | $\underline{0.9061 \pm 0.0049}$ | $0.6673 \pm 0.0670$ | $\mathbf{0.9438 \pm 0.0053}$ |
| MDLPHMDA | $0.8019 \pm 0.0288$ | $0.8061 \pm 0.0238$ | $0.8470 \pm 0.0473$ | $0.6759 \pm 0.0332$ | $0.8484 \pm 0.0478$ |
| NBLPIHMDA | $0.8384 \pm 0.0417$ | $0.7968 \pm 0.0496$ | $\mathbf{0.9280 \pm 0.0034}$ | $0.6651 \pm 0.0705$ | $\underline{0.9302 \pm 0.0039}$ |
| BPNNHMDA | $0.9057 \pm 0.0112$ | $0.8653 \pm 0.0485$ | $0.8739 \pm 0.0452$ | $0.8307 \pm 0.0830$ | $0.8744 \pm 0.0462$ |
| NCPLP | $\underline{0.9184 \pm 0.0093}$ | $\underline{0.9058 \pm 0.0174}$ | $0.8204 \pm 0.0440$ | $0.8533 \pm 0.0327$ | $0.8194 \pm 0.0445$ |
| GATMDA | $0.9063 \pm 0.0111$ | $0.6917 \pm 0.0263$ | $0.8644 \pm 0.0235$ | $\underline{0.9091 \pm 0.0214}$ | $0.8636 \pm 0.0238$ |
| MDAGCAN | $\mathbf{0.9290 \pm 0.0143}$ | $\mathbf{0.9062 \pm 0.0401}$ | $0.8559 \pm 0.0410$ | $\mathbf{0.9232 \pm 0.0159}$ | $0.8549 \pm 0.0418$ |

The best results are marked in bold and the second-best results are underlined.

TABLE 4 Performance comparison between 10 baseline methods and MDAGCAN in 5-fold CV on MASI dataset.

| Methods | AUC | F1 Score | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KATZHMDA | 0.6869 ± 0.0160 | 0.7371 ± 0.0382 | 0.6048 ± 0.0556 | 0.7133 ± 0.0685 | 0.6026 ± 0.0580 |
| BiRWMP | 0.7370 ± 0.0228 | 0.7285 ± 0.0366 | 0.7616 ± 0.0260 | 0.7062 ± 0.0466 | 0.7627 ± 0.0272 |
| LRLSHMDA | 0.7724 ± 0.0115 | 0.8169 ± 0.0272 | 0.6453 ± 0.0366 | 0.8378 ± 0.0487 | 0.6413 ± 0.0383 |
| NTSHMDA | 0.7861 ± 0.0152 | 0.8490 ± 0.0273 | 0.7262 ± 0.0538 | 0.7523 ± 0.0520 | 0.7257 ± 0.0557 |
| BRWMDA | 0.8128 ± 0.0174 | 0.8681 ± 0.0422 | 0.7488 ± 0.0506 | 0.7658 ± 0.0378 | 0.7485 ± 0.0523 |
| MDLPHMDA | <u>0.8324 ± 0.0156</u> | <u>0.8755 ± 0.0293</u> | 0.7638 ± 0.0404 | 0.8099 ± 0.0547 | <u>0.7629 ± 0.0421</u> |
| NBLPIHMDA | 0.8209 ± 0.0140 | **0.8818 ± 0.0187** | 0.7311 ± 0.0569 | 0.7997 ± 0.0520 | 0.7297 ± 0.0590 |
| BPNNHMDA | 0.8049 ± 0.0133 | 0.8065 ± 0.0424 | 0.6774 ± 0.0552 | 0.8246 ± 0.0596 | 0.6744 ± 0.0574 |
| NCPLP | 0.7824 ± 0.0131 | 0.8128 ± 0.0217 | 0.6596 ± 0.0286 | <u>0.8528 ± 0.0461</u> | 0.6556 ± 0.0300 |
| GATMDA | 0.8206 ± 0.0173 | 0.7534 ± 0.0243 | <u>0.7642 ± 0.0448</u> | **0.8794 ± 0.0400** | 0.7619 ± 0.0463 |
| MDAGCAN | **0.8730 ± 0.0036** | 0.7840 ± 0.0135 | **0.7996 ± 0.0157** | 0.8411 ± 0.0206 | **0.7691 ± 0.0142** |

The best results are marked in bold and the second-best results are underlined.

TABLE 5 Prediction results of top-20 liver cirrhosis-related microbes.

| Rank | Microbe | Evidence | Rank | Microbe | Evidence |
|---|---|---|---|---|---|
| 1 | *Clostridium difficile* | PMID: 26440041 | 11 | *Clostridium leptum* | PMID: 24564202 |
| 2 | *Helicobacter pylori* | PMID: 9365129 | 12 | *Clostridiales* | PMID: 31726747 |
| 3 | *Staphylococcus aureus* | PMID: 30253652 | 13 | *Bifidobacterium* | PMID: 29806520 |
| 4 | *Clostridium coccoides* | Unconfirmed | 14 | *Escherichia coli* | PMID: 36207946 |
| 5 | *Staphylococcus* | PMID: 25518533 | 15 | *Bacteroides vulgatus* | PMID: 23333527 |
| 6 | *Actinobacteria* | PMID: 32265857 | 16 | *Enterococcus* | PMID: 36035413 |
| 7 | *Clostridia* | PMID: 30661942 | 17 | *Bacteroides ovatus* | Unconfirmed |
| 8 | *Stenotrophomonas maltophilia* | PMID: 35755768 | 18 | *Bacteroides uniformis* | PMID: 33348106 |
| 9 | *Burkholderia* | Unconfirmed | 19 | *Prevotella* | PMID: 32414035 |
| 10 | *Betaproteobacteria* | Unconfirmed | 20 | *Klebsiella* | PMID: 36147601 |

TABLE 6 Prediction results of top-20 epilepsy-related microbes.

| Rank | Microbe | Evidence | Rank | Microbe | Evidence |
|---|---|---|---|---|---|
| 1 | Prevotellaceae | PMID: 35250450 | 11 | Faecalibacterium | PMID: 35069460 |
| 2 | Firmicutes | PMID: 35250450 | 12 | Coprococcus | PMID: 6699268 |
| 3 | Clostridiales | PMID: 30007242 | 13 | Erysipelotrichaceae | PMID: 33415132 |
| 4 | Enterobacteriaceae | PMID: 35069460 | 14 | Clostridium | PMID: 6699268 |
| 5 | Ruminococcaceae | PMID: 30007242 | 15 | Rikenellaceae | PMID: 30007242 |
| 6 | Clostridia | Unconfirmed | 16 | Bacteroidetes | PMID: 30007242 |
| 7 | Bacteroidaceae | Unconfirmed | 17 | Ruminococcus | PMID: 6699268 |
| 8 | Porphyromonadaceae | Unconfirmed | 18 | Streptococcus | PMID: 35250450 |
| 9 | Roseburia | PMID: 31646147 | 19 | Actinobacteria | PMID: 35250450 |
| 10 | Lachnospiraceae | PMID: 30007242 | 20 | Klebsiella | PMID: 34234109 |

Epilepsy is another common chronic neurological disorder around the world. Recent researches demonstrate that epilepsy patients tend to have dysbiosis or imbalance of gut microbial composition (Dong et al., 2022). Prevotellaceae, Actinobacteria and Streptococcus appear higher abundance compared to the healthy control group, and Firmicutes appears in the inverse pattern, where they are all ranked in our predicted top 20 score list.

In addition, Clostridia, ranked sixth in the score list, is less reported about epilepsy, but Clostridium spp appears increased relative abundance in autism spectrum disorder (ASD) (Borghi and Vignoli, 2019), where ASD and epilepsy maybe have the same heredity and physiopathologic mechanism (Mei et al., 2017). The two rarely reported microbes for epilepsy are Bacteroidaceae and Porphyromonadaceae. But there is evidence that Bacteroidaceae is

depleted after traumatic brain injury (Rogers et al., 2022) and the decrease of Porphyromonadaceae is closely linked to schizophrenia (Juckel et al., 2021). In the future, their important role in epilepsy will be further verified by wet experiments. In conclusion, results demonstrate that our method can effectively predict potential microbes for given diseases, which facilitates disease diagnosis and prevention.

# 5. Discussion

Over the last decade, increasing researchers pay more attention to the gut-liver-brain axis. The gut-liver-brain axis refers to the bidirectional relationship between the gut and its microbiota, the liver, and the brain, resulting from integrating signals generated by dietary, genetic, and environmental factors (Rocco et al., 2021). Growing evidences have emerged to consider the microbiota-gut-liver-brain axis as a comprehensive approach for better understanding diseases pathophysiology (Fuenzalida et al., 2021).

Figuring out the interactions between microbes and diseases provides a new way to diagnose and treat diseases. However, experimental identification of microbe-disease associations is time-consuming, laborious and expensive. The development of high-throughput sequencing technology has made it possible to explore the association between microbes and diseases on a large scale. In this paper, we present a deep learning framework based on the graph convolutional attention network. We integrate microbe similarity network, disease similarity network and known microbe-disease associations into a heterogeneous network. Then, we encode and learn the node feature information from its neighbors and itself via multiple graph convolutional layers and graph attention layers. Finally, MDAGCAN reconstructs the unobserved microbe-disease associations through a bilinear decoder. Comprehensive experiments demonstrate that our method MDAGCAN is promising and reliable to identify disease-related potential target microbes.

In addition, we further apply the microbe-disease association prediction model to predict liver cirrhosis and epilepsy-associated microbes and to find out the top 20 microbial candidates associated with them. Meanwhile, the indirect validation indicates that the remaining microbes are also associated with liver cirrhosis and epilepsy, respectively. They may be novel prospective biomarkers that require further experimental validation. Accumulating studies have revealed that epilepsy is associated with increased mortality in liver cirrhosis, but the underlying mechanism is still not known. Our analysis results display that there are four common microbes in the top 20 ranking score lists from liver cirrhosis and epilepsy, i.e., Actinobacteria, Clostridia, Clostridiales and Klebsiella. It is reported that the relative abundances of Actinobacteria and Klebsiella both increase in patients with liver cirrhosis and epilepsy compared with healthy controls (Chen et al., 2020; Lin et al., 2021; Dong et al., 2022; Zhou et al., 2022). Clostridiales with decreased abundance is strongly associated with the severity of liver cirrhosis and the seizure of epilepsy (Zhang et al., 2018; Fukui, 2019). Also, Clostridia appears inverse abundance pattern in liver cirrhosis and epilepsy patients (Zhang L. et al., 2019). Moreover, Actinobacteria produces SCFAs through metabolic pathways. SCFAs are vital components in the microbiota-gut-brain axis affecting the immune and endocrine systems through involvement in gut-brain signal pathways (Gong et al., 2021; Phillips-Farfan et al., 2021).

Klebsiella and Clostridiales produce an extracellular toxic complex via metabolic pathways whose main component is lipopolysaccharide (LPS). LPS release mainly affects the inflammatory response in the whole organism and the gut-liver-brain communication (Ahluwalia et al., 2016; Boeri et al., 2019). In conclusion, the gut microbe is possible as a bridge to understand the pathogenesis of liver cirrhosis and epilepsy.

Although several experiments show that our method performs well in predicting new associations, there are still some limitations. On the one hand, the known microbe-disease associations are insufficient to attain better prediction performance due to data imbalance and sparsity. On the other hand, MDAGCAN lacks a wealth of prior biological knowledge like microbial phylogeny, microbial gene sequencing and disease semantic information to improve predictive performance. In the future, we will make further research and efforts to address these shortcomings.

# Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

# Author contributions

LL implemented the experiments. ZLC analyzed the result. KS, LL, and SFF wrote the manuscript. KS and SFF designed the experiments and conducted the project. ZFW, HZC, and ZLC acquired the data and conceived the critical appraisal of the method. All authors read and approved the final manuscript.

# Funding

# Acknowledgments

# 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Ahluwalia, V., Betrapally, N. S., Hylemon, P. B., White, M. B., Gillevet, P. M., Unser, A. B., et al. (2016). Impaired gut-liver-brain axis in patients with cirrhosis. *Sci. Rep.* 6:26800. doi: 10.1038/srep26800

Al-Beltagi, M., and Saeed, N. K. (2022). Epilepsy and the gut: Perpetrator or victim? *World J. Gastrointest. Pathophysiol.* 13, 143–156. doi: 10.4291/wjgp.v13.i5.143

Altaib, H., Kozakai, T., Badr, Y., Nakao, H., El-Nouby, M. A., Yanase, E., et al. (2022). Cell factory for gamma-aminobutyric acid (GABA) production using *Bifidobacterium adolescentis*. *Microb. Cell. Fact.* 21:33. doi: 10.1186/s12934-021-01729-6

Bhat, M., Arendt, B. M., Bhat, V., Renner, E. L., Humar, A., and Allard, J. P. (2016). Implication of the intestinal microbiome in complications of cirrhosis. *World J. Hepatol.* 8, 1128–1136. doi: 10.4254/wjh.v8.i27.1128

Blum, H. E. (2017). The human microbiome. *Adv. Med. Sci.* 62, 414–420. doi: 10.1016/j.advms.2017.04.005

Boeri, L., Izzo, L., Sardelli, L., Tunesi, M., Albani, D., and Giordano, C. (2019). Advanced Organ-on-a-Chip devices to investigate liver multi-organ communication: Focus on gut, microbiota and brain. *Bioengineering (Basel)* 6:91. doi: 10.3390/bioengineering6040091

Borghi, E., and Vignoli, A. (2019). Rett syndrome and other neurodevelopmental disorders share common changes in gut microbial community: A descriptive review. *Int J. Mol. Sci.* 20:4160. doi: 10.3390/ijms20174160

Chen, X., Huang, Y., You, Z., Yan, G., and Wang, X. (2017). A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* 33, 733–739. doi: 10.1093/bioinformatics/btw715

Chen, Z., Xie, Y., Zhou, F., Zhang, B., Wu, J., Yang, L., et al. (2020). Featured gut microbiomes associated with the progression of chronic hepatitis B disease. *Front. Microbiol.* 11:383. doi: 10.3389/fmicb.2020.00383

Deidda, G., Crunelli, V., and Giovanni, G. D. (2021). 5-HT/GABA interaction in epilepsy. *Prog. Brain. Res.* 259, 265–286. doi: 10.1016/bs.pbr.2021.01.008

Dong, L., Zheng, Q., Cheng, Y., Zhou, M., Wang, M., Xu, J., et al. (2022). Gut microbial characteristics of adult patients with epilepsy. *Front. Neurosci.* 16:803538. doi: 10.3389/fnins.2022.803538

Du, Z., Wu, Y., Huang, Y., Chen, J., Pan, G., Hu, L., et al. (2022). GraphTGI: An attention-based graph embedding model for predicting TF-target gene interactions. *Brief. Bioinform.* 23:bbac148. doi: 10.1093/bib/bbac148

Feng, Y., Wei, Z., Liu, C., Li, G., Qiao, X., Gan, Y., et al. (2022). Genetic variations in GABA metabolism and epilepsy. *Seizure* 101, 22–29. doi: 10.1016/j.seizure.2022.07.007

Fuenzalida, C., Dufeu, M. S., Poniachik, J., Roblero, J. P., Valenzuela-Pérez, L., and Beltrán, C. J. (2021). Probiotics-based treatment as an integral approach for alcohol use disorder in alcoholic liver disease. *Front. Pharmacol.* 12:729950. doi: 10.3389/fphar.2021.729950

Fukui, H. (2019). Role of gut dysbiosis in liver diseases: What have we learned so far? *Diseases* 7:58. doi: 10.3390/diseases7040058

Gabanyi, I., Lepousez, G., Wheeler, R., Vieites-Prado, A., Nissant, A., Wagner, S., et al. (2022). Bacterial sensing via neuronal Nod2 regulates appetite and body temperature. *Science* 376:eabj3986. doi: 10.1126/science.abj3986

Gan, P., Huang, S., Pan, X., Xia, H., Zeng, X., Ren, W., et al. (2022). Global research trends in the field of liver cirrhosis from 2011 to 2020: A visualised and bibliometric study. *World J. Gastroenterol.* 28, 4909–4919. doi: 10.3748/wjg.v28.i33.4909

Gong, X., Cai, Q., Liu, X., An, D., Zhou, D., Luo, R., et al. (2021). Gut flora and metabolism are altered in epilepsy and partially restored after ketogenic diets. *Microb. Pathog.* 155:104899. doi: 10.1016/j.micpath.2021.104899

Gonzalez-Ochoa, G., Flores-Mendoza, L. K., Icedo-Garcia, R., Gomez-Flores, R., and Tamez-Guerra, P. (2017). Modulation of rotavirus severe gastroenteritis by the combination of probiotics and prebiotics. *Arch. Microbiol.* 199, 953–961. doi: 10.1007/s00203-017-1400-3

Hussain, S. K., Dong, T. S., Agopian, V., Pisegna, J. R., Durazo, F. A., Enayati, P., et al. (2020). Dietary protein, fiber and coffee are associated with small intestine microbiome composition and diversity in patients with liver cirrhosis. *Nutrients* 12:1395. doi: 10.3390/nu12051395

Hwang, S., Kim, C., Yang, S., Kim, E., Hart, T., Marcotte, E., et al. (2019). HumanNet v2: Human gene networks for disease research. *Nucleic Acids Res.* 47, D573–D580. doi: 10.1093/nar/gky1126

Juckel, G., Manitz, M., Freund, N., and Gatermann, S. (2021). Impact of Poly I: C induced maternal immune activation on offspring's gut microbiome diversity - Implications for schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 110:110306. doi: 10.1016/j.pnpbp.2021.110306

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

Kingma, D. P., and Ba, J. (2015). "Adam: A method for stochastic optimization," in *Proceedings of the international conference on learning representations (ICLR)*, San Diego, CA.

Kipf, T. N., and Welling, M. (2017). "Semi-supervised classification with graph convolutional networks," in *Proceedings of the international conference on learning representations (ICLR)*, Toulon.

Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T. M., and Kamada, N. (2020). The bacterial connection between the oral cavity and the Gut diseases. *J. Dent. Res.* 99, 1021–1029. doi: 10.1177/0022034520924633

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. *IEEE ACM Trans. Comput. Biol. Bioinform.* 18, 2502–2513. doi: 10.1109/TCBB.2020.2986459

Lin, P., Lin, A., Tao, K., Yang, M., Ye, Q., Chen, H., et al. (2021). Intestinal *Klebsiella pneumoniae* infection enhances susceptibility to epileptic seizure which can be reduced by microglia activation. *Cell Death Discov.* 7:175. doi: 10.1038/s41420-021-00559-0

Lin, T., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 318–327. doi: 10.1109/TPAMI.2018.2858826

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 Bioinform.* 20:541. 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. Bioinform.* 22:bbaa146. doi: 10.1093/bib/bbaa146

Löscher, W., Potschka, H., Sisodiya, S. M., and Vezzani, A. (2020). Drug resistance in epilepsy: Clinical impact, potential mechanisms, and new innovative treatment options. *Pharmacol. Rev.* 72, 606–638. doi: 10.1124/pr.120.019539

Luo, J., and Long, Y. (2020). NTSHMDA: Prediction of human microbe-disease association based on random walk by integrating network topological similarity. *IEEE ACM Trans. Comput. Biol. Bioinform.* 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. Bioinform.* 18, 85–97. doi: 10.1093/bib/bbw005

Mei, S., Zhang, Z., Liu, X., Gao, T., and Peng, X. (2017). [Association between autism spectrum disorder and epilepsy in children]. *Zhongguo Dang Dai Er Ke Za Zhi* 19, 549–554. doi: 10.7499/j.issn.1008-8830.2017.05.014

Mouzaki, M., Comelli, E. M., Arendt, B. M., Bonengel, J., Fung, S. K., Fischer, S. E., et al. (2013). Intestinal microbiota in patients with nonalcoholic fatty liver disease. *Hepatology* 58, 120–127. doi: 10.1002/hep.26319

Olmedo, M., Reigadas, E., Valerio, M., Cuesta, S. V., Pajares, J. A., et al. (2019). Is it reasonable to perform fecal microbiota Transplantation for recurrent *Clostridium* difficile Infection in patients with liver cirrhosis? *Rev. Esp. Quimioter.* 32, 205–207.

Phillips-Farfan, B., Gómez-Chávez, F., Medina-Torres, E., Vargas-Villavicencio, J., Carvajal-Aguilera, K., Camacho, L., et al. (2021). Microbiota signals during the neonatal period forge life-long immune responses. *Int. J. Mol. Sci.* 22:8162. doi: 10.3390/ijms22158162

Qu, J., Zhao, Y., and Yin, J. (2019). Identification and analysis of human microbe-disease associations by matrix decomposition and label propagation. *Front. Microbiol.* 10:291. doi: 10.3389/fmicb.2019.00291

Ren, X., Hao, S., Yang, C., Yuan, L., Zhou, X., Zhao, H., et al. (2021). Alterations of intestinal microbiota in liver cirrhosis with muscle wasting. *Nutrition* 83:111081. doi: 10.1016/j.nut.2020.111081

Rocco, A., Sgamato, C., Compare, D., Coccoli, P., Nardone, O. M., and Nardone, G. (2021). Gut microbes and hepatic encephalopathy: From the old concepts to new perspectives. *Front. Cell. Dev. Biol.* 9:748253. doi: 10.3389/fcell.2021.748253

Rogers, M. B., Simon, D., Firek, B., Silfies, L., Fabio, A., Bell, M. J., et al. (2022). Temporal and spatial changes in the microbiome following pediatric severe traumatic brain injury. *Pediatr Crit Care Med.* 23, 425–434. doi: 10.1097/PCC.0000000000002929

Shen, X., Zhu, H., Jiang, X., Hu, X., Yang, J., et al. (2018). "A novel approach based on bi-random walk to predict microbe-disease associations," in *intelligent computing methodologies*, eds D.-S. Huang, M. M. Gromiha, K. Han, and A. Hussain (Cham: Springer International Publishing).

Tooley, K. L. (2020). Effects of the human gut microbiota on cognitive performance, brain structure and function: A narrative review. *Nutrients* 12:3009. doi: 10.3390/nu12103009

Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., Bengio, Y., et al. (2018). "Graph attention networks," in *Proceedings of the international conference on learning representations (ICLR)*, Canada.

de Vos, W. M., Tilg, H., Hul, M. V., and Cani, P. D. (2022). Gut microbiome and health: Mechanistic insights. *Gut* 71, 1020–1032. doi: 10.1136/gutjnl-2021-326789

Wang, F., Huang, Z., 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, 1–11. doi: 10.1038/s41598-017-08127-2

Wang, L., Wang, Y., Li, H., Feng, X., Yuan, D., and Yang, J. (2019). A bidirectional label propagation based computational model for potential microbe-disease association prediction. *Front. Microbiol.* 10:684. doi: 10.3389/fmicb.2019.00684

Wang, Y., Lei, X., Lu, C., and Pan, Y. (2021). Predicting microbe-disease association based on multiple similarities and LINE algorithm. *IEEE ACM Trans. Comput. Biol. Bioinform.* 19, 2399–2408. doi: 10.1109/TCBB.2021.3082183

Wei, H., and Liu, B. (2020). iCircDA-MF: Identification of circRNA-disease associations based on matrix factorization. *Brief. Bioinform.* 21, 1356–1367. doi: 10.1093/bib/bbz057

Won, S., Oh, K. K., Gupta, H., Ganesan, R., Sharma, S. P., Jeong, J., et al. (2022). The link between gut microbiota and hepatic encephalopathy. *Int. J. Mol. Sci.* 23:8999. doi: 10.3390/ijms23168999

Yan, C., Duan, G., Wu, F., Pan, Y., and Wang, J. (2020). BRWMDA:Predicting microbe-disease associations based on similarities and bi-random walk on disease and microbe networks. *IEEE ACM Trans. Comput. Biol. Bioinform.* 17, 1595–1604. doi: 10.1109/tcbb.2019.2907626

Yang, X., Kuang, L., Chen, Z., and Wang, L. (2021). Multi-similarities bilinear matrix factorization-based method for predicting human microbe-disease associations. *Front. Genet.* 12:1941. doi: 10.3389/fgene.2021.754425

Yin, M., Liu, J., Gao, Y., Kong, X., and Zheng, C. (2020). NCPLP: A novel approach for predicting microbe-associated diseases with network consistency projection and label propagation. *IEEE Trans. Cybern.* 52, 5079–5087. doi: 10.1109/TCYB.2020.3026652

Yin, M.-M., Gao, Y. L., Shang, J., Zheng, C. H., and J-X, Liu (2022). Multi-similarity fusion-based label propagation for predicting microbes potentially associated with diseases. *Futur. Gener. Comp. Syst.* 134, 247–255. doi: 10.1016/j.future.2022.04.012

Yu, Z., Huang, F., Zhao, X., Xiao, W., and Zhang, W. (2021). Predicting drug-disease associations through layer attention graph convolutional network. *Brief. Bioinform.* 22:bbaa243. doi: 10.1093/bib/bbaa243

Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., et al. (2020). Graph embedding on biomedical networks: Methods, applications and evaluations. *Bioinformatics* 36, 1241–1251. doi: 10.1093/bioinformatics/btz718

Zeng, X., Yang, X., Fan, J., Tan, Y., Ju, L., Shen, W., et al. (2021). MASI: Microbiota-active substance interactions database. *Nucleic Acids Res.* 15:209. doi: 10.1093/nar/gkaa924

Zhang, S., Tong, H., Xu, J., and Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Comput. Soc. Netw.* 6:11. doi: 10.1186/s40649-019-0069-y

Zhang, L., Wu, Y., Chen, T., Ren, C., Li, X., and Liu, G. (2019). Relationship between intestinal microbial dysbiosis and primary liver cancer. *Hepatobiliary Pancreat. Dis. Int.* 18, 149–157. doi: 10.1016/j.hbpd.2019.01.002

Zhang, Y., Zhou, S., Zhou, Y., Yu, L., Zhang, L., and Wang, Y. (2018). Altered gut microbiome composition in children with refractory epilepsy after ketogenic diet. *Epilepsy. Res.* 145, 163–168. doi: 10.1016/j.eplepsyres.2018.06.015

Zhou, Z., Lv, H., Lv, J., Shi, Y., Huang, H., Chen, L., et al. (2022). Alterations of gut microbiota in cirrhotic patients with spontaneous bacterial peritonitis: A distinctive diagnostic feature. *Front. Cell. Infect. Microbiol.* 12:999418. doi: 10.3389/fcimb.2022.999418