frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fmicb.2022.1093615

SCCPMD: Probability matrix decomposition method subject to corrected similarity constraints for inferring long non-coding RNA–disease associations

Lieqing Lin[1], Ruibin Chen[2], Yinting Zhu[2], Weijie Xie[2], Huaiguo Jing[3]*, Langcheng Chen[1]*and Minqing Zou[4]

[1]Center of Campus Network & Modern Educational Technology, Guangdong University of Technology, Guangzhou, China, [2]School of Computer, Guangdong University of Technology, Guangzhou, China, [3]Sports Department, Guangdong University of Technology, Guangzhou, China, [4]Department of Experiment Teaching, Guangdong University of Technology, Guangzhou, China

Check for updates

OPEN ACCESS

EDITED BY
Qi Zhao,
University of Science and Technology Liaoning, China

REVIEWED BY
Chun-Chun Wang,
China University of Mining and Technology, China
Li Zhang,
Liaoning University, China

*CORRESPONDENCE
Huaiguo Jing
✉ jinghg@gdut.edu.cn
Langcheng Chen
✉ clc@gdut.edu.cn

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 09 November 2022
ACCEPTED 30 November 2022
PUBLISHED 11 January 2023

CITATION
Lin L, Chen R, Zhu Y, Xie W, Jing H, Chen L and Zou M (2023) SCCPMD: Probability matrix decomposition method subject to corrected similarity constraints for inferring long non-coding RNA–disease associations.
Front. Microbiol. 13:1093615.
doi: 10.3389/fmicb.2022.1093615

Accumulating evidence has demonstrated various associations of long non-coding RNAs (lncRNAs) with human diseases, such as abnormal expression due to microbial influences that cause disease. Gaining a deeper understanding of lncRNA–disease associations is essential for disease diagnosis, treatment, and prevention. In recent years, many matrix decomposition methods have also been used to predict potential lncRNA-disease associations. However, these methods do not consider the use of microbe-disease association information to enrich disease similarity, and also do not make more use of similarity information in the decomposition process. To address these issues, we here propose a correction-based similarity-constrained probability matrix decomposition method (SCCPMD) to predict lncRNA–disease associations. The microbe-disease associations are first used to enrich the disease semantic similarity matrix, and then the logistic function is used to correct the lncRNA and disease similarity matrix, and then these two corrected similarity matrices are added to the probability matrix decomposition as constraints to finally predict the potential lncRNA–disease associations. The experimental results show that SCCPMD outperforms the five advanced comparison algorithms. In addition, SCCPMD demonstrated excellent prediction performance in a case study for breast cancer, lung cancer, and renal cell carcinoma, with prediction accuracy reaching 80, 100, and 100%, respectively. Therefore, SCCPMD shows excellent predictive performance in identifying unknown lncRNA–disease associations.

KEYWORDS

lncRNA-long noncoding RNA, disease, similarity correction, constraint probability matrix decomposition, associations prediction

Introduction

Non-coding RNAs such as microRNAs (miRNAs), Circular RNA (circRNA) and long non-coding RNAs (lncRNAs) play crucial roles in controlling the biological processes of plants and animals (Zhang et al., 2020b; Wang et al., 2021a, 2022). Owing to their roles as genetic regulators in the development of complex disorders such as cancer, miRNAs have the potential to serve as diagnostic markers and therapeutic targets (Chen et al., 2019b; Hill and Tran, 2021; Huang et al., 2022a,b). Several algorithmic models have also been developed for the exploration of miRNA–disease and miRNA-disease associations (Chen et al., 2019c, 2021a; Zhang et al., 2021a,b). However, as medicine advances, more and more studies have also shown that lncRNAs play an important role in many different diseases (Cao et al., 2019). LncRNAs are RNA molecules with transcriptional lengths above 200 nucleotides that lack protein-coding capabilities (Xing et al., 2021). For example, *HOXA-AS2* was identified as a novel cancer-associated lncRNA, which exhibits aberrant expression in a variety of malignancies, including breast, gastric, gallbladder, hepatocellular, and pancreatic cancers (Wang et al., 2018a). With increasing recognition of the importance of lncRNAs, more in-depth research has focused on the relationship between lncRNAs and diseases. However, traditional biological validation experiments are time-consuming and costly; thus, there is an urgent need to develop accurate and effective computational methods to determine possible lncRNA–disease associations. Many computational models have recently been developed to successfully predict possible lncRNA–disease associations, which can be classified into three main categories.

The first category is characterized by machine-learning methods (Zhang et al., 2020a; Lan et al., 2022). Chen and Yan (2013) proposed the first such approach to predict lncRNA–disease associations using Laplace regularized least squares in a semi-supervised learning framework. Subsequently, by combining genomic, glomerular, and transcriptomic data, Zhao et al. (2015) devised a computational method based on a simple Bayesian classifier approach, which led to the discovery of 707 potential cancer-associated lncRNAs. Zhu et al. (2021) predicted lncRNA–disease associations by integrating several similarity matrices and combining incremental principal component analysis and random forest techniques. However, supervised learning-based models such as support vector machine and plain Bayesian classifiers rely heavily on difficult-to-obtain negative sample (Chen et al., 2017).

The second category is based on building biological networks to predict lncRNA–disease associations (Zhang et al., 2019a, 2020c). Sun et al. (2014) proposed RWRlncD, a global network computational strategy that applies restart random wandering (RWR) on lncRNA functional similarity networks to infer potential associations between human lncRNAs and disease. Zhang et al. (2019b) integrated known topological interactions of lncRNA–disease, lncRNA–miRNA, and miRNA–disease to construct a linked tripartite network, and used the topology of the obtained network to calculate the similarity of disease pairs and lncRNA pairs. Finally, rule-based inference methods were used to predict new lncRNA–disease associations. Zhou et al. (2021) employed a rotating forest classifier to train prediction models after creating a heterogeneous network by combining relationships among miRNAs, lncRNAs, proteins, drugs, and diseases. However, the heterogeneous networks constructed by these network-based approaches relying on the relationships of lncRNAs themselves, miRNAs, proteins, and drugs to lncRNAs (diseases) can result in failure in reliable predictions of new diseases and/or new lncRNAs.

The third category includes matrix decomposition methods (Chen et al., 2018a,b, 2021b; Xie et al., 2021). To effectively predict probable relationships, Fu et al. (2018) employed matrix triple decomposition to split a data matrix from heterogeneous data sources into low-rank matrices and reconstruct the lncRNA–disease association matrix. Based on probabilistic matrix decomposition, Xuan et al. (2019) deduced probable lncRNA–disease associations by assuming that low-rank matrices are positively distributed with Gaussian noise. To enhance the potential association between lncRNAs and diseases, Gao et al. (2021) optimized the lncRNA and disease space by multi-labeling and fusing these labels. Finally, co-matrix decomposition was used to predict lncRNA–disease correlations. Wang et al. (2021b) treated the discovery of disease-associated lncRNA as a recommender system problem, and predicted the relationships between lncRNA and diseases using a graph-regularized non-negative matrix decomposition approach. (Liu et al., 2021) proposed an lncRNA–disease association prediction approach based on double sparse collaborative matrix decomposition. To boost the sparsity, the L2,1-norm was introduced to the conventional co-matrix decomposition method. However, none of the algorithms presented above use similar information of lncRNA and disease as constraints to optimize the matrix decomposition algorithm. Thus, there is still some room for improvement in the prediction performance.

Traditional probabilistic matrix decomposition only uses probabilistic linear models with Gaussian noise to model the interaction of lncRNAs with diseases. Based on the assumption that similar lncRNAs/diseases are usually interrelated with the corresponding disease/lncRNA, we here propose a correction-based similarity-constrained probability matrix decomposition (SCCPMD) method for predicting lncRNA–disease associations. Considering the noise effect of the similarity matrix of lncRNAs and diseases, the noise is reduced by correcting the similarity matrix using a logistic function to highlight strong correlations within the similarity range [0,1] while diluting weak correlations. The lncRNA and disease similarity are then used as constraints in the probability matrix decomposition process, resulting in two low-rank matrices to predict the potential lncRNA–disease association. Leave-one-out cross-validation (LOOCV) and five-fold cross-validation (5-fold CV) were performed to validate the predictive performance of SCCPMD using known lncRNA–disease association datasets. The final area under the curve (AUC) values of SCCPMD reached 0.9787 and 0.9528 ± 0.0036 with LOOCV and 5-fold CV, respectively, which were both better than the prediction performances obtained with existing advanced algorithms. In addition, we confirmed the effectiveness of SCCPMD in application to three test cases of human diseases: breast cancer, lung cancer, and renal cell carcinoma (RCC).

Materials and methods

Datasets

We used the LncRNADisease database (Bao et al., 2019), which provides a dataset of lncRNA–disease associations. After removing duplicate lncRNAs and diseases as well as non-human data, 1,690 unique experimentally validated lncRNA–disease associations were obtained, including 447 unique lncRNAs and 218 unique diseases. The lncRNA–disease associations were described by building a disease–lncRNA adjacency matrix, $Y \in R^{nl \times nd}$, where nl and nd represent the number of lncRNAs and diseases, respectively. The matrix Y is defined as follows:

$$Y(i,j) = \begin{cases} 0 \text{ lncRNA } l(i) \text{ has no association with disease } d(j) \\ 1 \text{ lncRNA } l(i) \text{ is associated with disease } d(j) \end{cases} \quad (1)$$

In other words, if an lncRNA l_i is confirmed to be associated with a disease d_j, then $Y(i,j)$ is set to 1; otherwise, $Y(i,j)$ is 0.

Semantic similarity of disease

We built a directed acyclic graph (DAG) based on the descriptor data from the Medical Subject Headings (MeSH) of the National Library of Medicine[1] to determine the semantic similarity among diseases. A disease d is described by $DAG(d) = (d, V(d), E(d))$, where $V(d)$ and $E(d)$ are the vertex set and edge set of the DAG, respectively. Based on the DAG layer structure of disease d, we can calculate the semantic value (S) of disease m to disease d as follows:

$$T_d(m) = \begin{cases} 1 & , \; if \; m = d \\ \max\{0.5^* T_d(m') \mid m' \in children \; of \; m, if \; m \neq d \end{cases} \quad (2)$$

According to the DAG of a disease, the semantic value of a disease is defined as the sum of the ancestral nodes of the disease and the semantic contribution value of the disease to itself, expressed by the following equation:

$$T_d = \sum_{m \in V(d)} T_d(m) \quad (3)$$

Based on the above steps, we can construct the semantic similarity matrix SS to represent the semantic similarity between disease d_i and disease d_j :

$$SS(d_i, d_j) = \frac{\sum_{m \in V(d_i) \cap V(d_j)} \left(T_{d_i}(m) + T_{d_j}(m) \right)}{T_{d_i} + T_{d_j}} \quad (4)$$

Gaussian interaction profile kernel similarity for diseases

To address the sparsity of the semantic similarity matrix of diseases and integrate more information on disease similarity, we used microbe-disease associations to calculate Gaussian similarity of diseases. We downloaded human microbe-disease associations from the Human Microbe-Disease Association Database (HMDAD). Microbe-disease associations were described by creating a microbe-disease adjacency matrix, $A \in R^{m \times n}$, where m and d represent the number of microbes and diseases, respectively. As a measure of disease similarity, we constructed Gaussian interaction spectral kernel similarity using radial basis functions. We calculated the Gaussian interaction distribution based on the adjacency matrix A. The Gaussian interaction spectral kernel similarity between disease d_i and disease d_j can be calculated by the following equation:

$$GD(d_i, d_j) = \exp\left(-\gamma_d \left\| A(:,i) - A(:,j) \right\|^2 \right) \quad (5)$$

$$\gamma_d = \gamma / \left(\frac{1}{n} \sum_{i=1}^{n} \left\| A(:,i) \right\|^2 \right) \quad (6)$$

Integrated similarity for diseases

We combine the disease semantic similarity SS with the disease Gaussian similarity GD to construct the final disease similarity matrix SD . as follows, for disease d_i and disease d_j , $SD(d_i, d_j) = GD(d_i, d_j)$ if $SS = 0$ and $SD(d_i, d_j) = SS(d_i, d_j)$ otherwise.

[1] https://www.nlm.nih.gov/mesh/meshhome.html

$$DS(d_i,d_j)=\begin{cases}GD(d_i,d_j) & if\ \ SS(d_i,d_j)=0\\ SS(d_i,d_j) & otherwise\end{cases} \tag{7}$$

Expression similarity of LncRNAs

LncRNA expression profiles can be utilized to reflect the similarity between lncRNAs, since related lncRNAs exhibit co-expression characteristics in various tissues (Chen et al., 2019a). For this purpose, we used RNA-sequencing data retrieved from the ArrayExpress database to create lncRNA expression profiles. The Spearman correlation coefficient between the expression profiles of two lncRNAs was then used to determine the degree of similarity in their expression patterns, defined as ES , where $ES(l_i,l_j)\in[0,1]$ denotes the expression similarity of lncRNAs l_i and l_j .

SCCPMD method

Overview

SCCPMD involves the following five steps, which are schematically outlined in Figure 1: (i) constructing lncRNA–disease association networks, (ii) constructing DAGs based on MeSH information to calculate the disease semantic similarity SS and calculating disease Gaussian similarity GD based on microbe-disease associations, (iii) integration of disease semantic similarity and disease gaussian similarity to obtain disease similarity SD , (iv) calculating lncRNA expression similarity ES based on Spearman correlation coefficients, (v) performing logistic function transformation for similarity correction of disease similarity and lncRNA expression similarity to reduce the noise introduced by the similarity matrix during matrix decomposition, and (vi) using the proposed constrained probability matrix decomposition method to help predict potential lncRNA–disease associations.

Similarity correction

To reduce the noise that lncRNA and disease similarity matrices introduce during matrix decomposition, similarity correction techniques were used. The noise present in the similarity matrix is reduced by the logistic function so as to enhance the strong correlations in the similarity range [0,1] while diluting the weak correlations. This approach has previously been used in the study of disease-related genes (Vanunu et al., 2010). The logistic function is defined as follows:

$$L(x)=\frac{1}{1+e^{ax+b}} \tag{8}$$

$L(x)\approx 0$ when $x\in[0,0.3]$ and $L(x)\approx 1$ when $x\in[0.6,1]$. This means that weakly similar coefficients in the range of [0,0.3] are lost information, whereas strong similar coefficient values in the range of [0.6,1] usually exhibit significant co-expression of the relationship. Accordingly $L(0)$ needs to be close to 0; therefore, we set $L(0)=0.0001$ to obtain $b=\log(9999)$. In addition, a is a correction degree coefficient that is used for parameter adjustment of the model. The corrected lncRNA expression similarity LE and the disease similarity LD are thus obtained as follows:

$$LE(i,j)=\frac{1}{1+e^{a\times ES(i,j)+b}},\ i,j\in[1,nl] \tag{9}$$

$$LD(i,j)=\frac{1}{1+e^{a\times DS(i,j)+b}},\ i,j\in[1,nd] \tag{10}$$

Constraint probability matrix decomposition

Following the similarity correction steps outlined above, we can obtain the association matrix Y representing the relationship between lncRNA and disease from the corrected lncRNA–lncRNA expression similarity LE and the corrected disease–disease similarity LD . The values of LE and LD fall in the [0,1] interval. Let $W\in R^{k\times nl}$ and $D\in R^{k\times nd}$ be the lncRNA and disease latent feature matrices, where $k\in\min(nl,nd)$. The latent feature vectors specific to lncRNAs and diseases are represented by the column vectors W_i and D_j , respectively. The goal is then to find lncRNA and disease latent models ($W\in R^{k\times nl}$ and $D\in R^{k\times nd}$) whose product (W^TD) can reconstruct the interaction matrix Y . From a probabilistic point of view, the conditional distribution of the observed interactions $Y\in\{0,1\}$ is expressed as:

$$P(Y|W,D,\sigma^2)=\prod_{i=1}^{nl}\prod_{j=1}^{nd}\left[f\left(Y_{ij}\,|\,W_i^TD_j,\sigma^2\right)\right]^{I_{ij}} \tag{11}$$

where $f(x|,\mu|,\sigma 2)$ is the probability density function of the Gaussian normal distribution with mean μ and variance σ^2 , and I_{ij} is the indicator function that is equal to 1 if the lncRNA l_i is related with disease d_j and is 0 otherwise. A probabilistic representation of the association matrix Y is then given by $P(Y|W,D,\sigma^2)$. We use the following zero-mean spherical Gaussian priors on the lncRNA and disease eigenvectors as a generative model for the lncRNA and disease latent models:

$$P(W\,|\,\sigma_W^2)=\prod_{i=1}^{nl}f\left(W_i\,|\,0,\sigma_W^2I\right) \tag{12}$$

$$P(D\,|\,\sigma_D^2)=\prod_{i=1}^{nd}f\left(D_j\,|\,0,\sigma_D^2I\right) \tag{13}$$

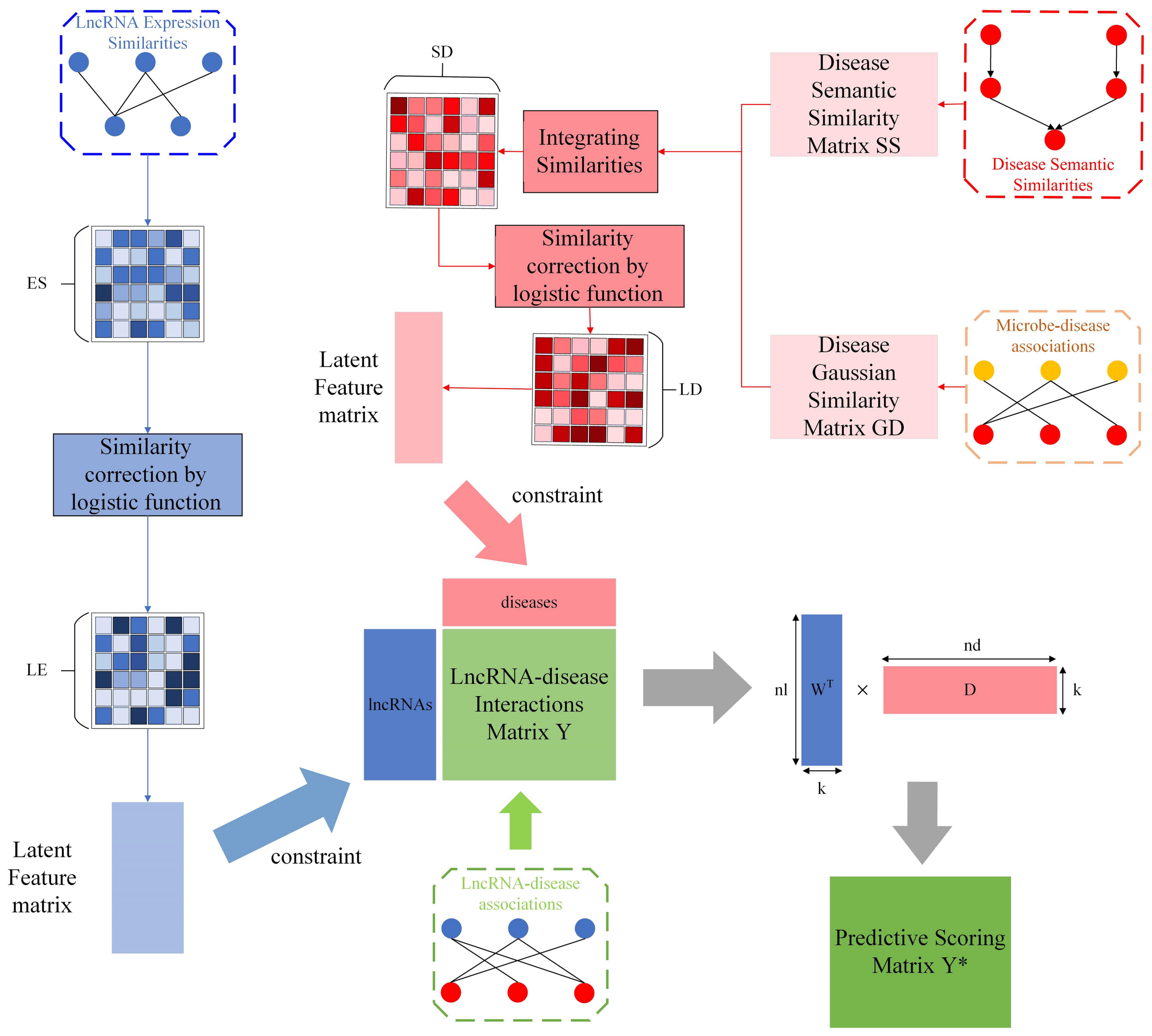
LncRNA Expression Similarities
ES
Similarity correction by logistic function
LE
Latent Feature matrix
constraint
SD
Integrating Similarities
Disease Semantic Similarity Matrix SS
Disease Semantic Similarities
Similarity correction by logistic function
LD
Latent Feature matrix
Microbe-disease associations
Disease Gaussian Similarity Matrix GD
constraint
diseases
lncRNAs
LncRNA-disease Interactions Matrix Y
LncRNA-disease associations
nl
W^T
×
nd
D
k
k
Predictive Scoring Matrix Y*

FIGURE 1
Flow chart of the SCCPMD approach.

where I is a k -dimensional identity diagonal matrix. Then, the posterior distribution of lncRNA and disease characteristics is derived as:

$$
\begin{aligned}
P\left(W, D \mid Y, \sigma^2, \sigma_W^2, \sigma_D^2\right) &= \frac{P\left(W, D, Y, \sigma^2, \sigma_W^2, \sigma_D^2\right)}{P\left(Y, \sigma^2, \sigma_W^2, \sigma_D^2\right)} \\
&= \frac{P\left(Y \mid W, D, Y, \sigma^2\right) \times P\left(W, D \mid \sigma_W^2, \sigma_D^2\right)}{P\left(Y, \sigma^2, \sigma_W^2, \sigma_D^2\right)} \\
&\sim P\left(Y \mid W, D, \sigma^2\right) \times P\left(W, D \mid \sigma_W^2, \sigma_D^2\right) \\
&= P\left(Y \mid W, D, \sigma^2\right) \times P\left(W \mid \left(\sigma_W\right)^2\right) \times P\left(D \mid \left(\sigma_D^2\right)\right) \\
&= \prod_{i=1}^{nl} \prod_{j=1}^{nd} \left[f\left(Y_{ij} \mid W_i^T D_j, \sigma^2\right)\right]^{I_{ij}} \\
&\times \prod_{i=1}^{nl} f\left(W_i \mid 0, \sigma_W^2 I\right) \times \prod_{i=1}^{nd} f\left(D_j \mid 0, \sigma_D^2 I\right)
\end{aligned} \tag{14}
$$

Taking the logarithm of equation (11), the distribution is transformed to:

$$
\begin{aligned}
\ln P\left(W, D \mid Y, \sigma^2, \sigma_W^2, \sigma_D^2\right) = & \frac{1}{2\sigma^2} \sum_{i=1}^{nl} \sum_{j=1}^{nd} I_{ij} \left(Y_{ij} - W_i^T D_j\right)^2 \\
& - \frac{1}{2\sigma^2} \sum_{i=1}^{nl} W_i^T W_i - \frac{1}{2\sigma^2} \sum_{j=1}^{nd} D_j^T D_j \\
& - \frac{1}{2} \left(\begin{array}{l} \left(\sum_{i=1}^{nl} \sum_{j=1}^{nd} I_{ij} \right) \ln \sigma^2 \\ + (nl) k \ln \sigma_W^2 + (nd) k \ln \sigma_D^2 \end{array} \right) + c
\end{aligned} \tag{15}
$$

where c is a constant. With the hyperparameters held constant, maximization of the log posterior for lncRNA and disease characteristics is identical to minimization of the sum of squared errors with a quadratic regularization term objective function:

$$\min \frac{1}{2}\sum_{i=1}^{nl}\sum_{j=1}^{nd} I_{ij}\left(Y_{ij}-W_i^T D_j\right)^2+\frac{\lambda_W}{2}\sum_{i=1}^{nl}\left\|W_i\right\|_{Fro}^2+\frac{\lambda_D}{2}\sum_{j=1}^{nd}\left\|D_j\right\|_{Fro}^2 \quad (16)$$

where $\lambda_W=\sigma^2/\sigma_W^2$, $\lambda_D=\sigma^2/\sigma_D^2$, $\|\cdot\|_{Fro}^2$ represents the Frobenius norm. However, the conventional probabilistic matrix decomposition model only uses a probabilistic linear model with Gaussian noise to depict the interaction between lncRNAs and diseases, leaving room for improvement. Based on the assumption that similar lncRNAs are usually interrelated with corresponding diseases and vice versa, CPMD takes more biological information (such as the similarity of lncRNAs and diseases) into account for the prediction. Accordingly, we suggest the following as a new objective function for CPMD:

$$\begin{aligned}&\min \frac{1}{2}\sum_{i=1}^{nl}\sum_{j=1}^{nd} I_{ij}\left(Y_{ij}-W_i^T D_j\right)^2+\frac{\lambda_W}{2}\sum_{i=1}^{nl}\left\|W_i\right\|_{Fro}^2+\frac{\lambda_D}{2}\sum_{j=1}^{nd}\left\|D_j\right\|_{Fro}^2\\&+\frac{\lambda_1}{2}\left\|W^T W-LD\right\|_{Fro}^2+\frac{\lambda_2}{2}\left\|D^T D-LE\right\|_{Fro}^2\end{aligned} \quad (17)$$

where W_i represents the k-dimensional potential feature vector of lncRNAs, $W^T W$ is the lncRNA weighted similarity matrix, and $D^T D$ is the disease weighted similarity matrix. Here, we use the gradient descent algorithm to solve the optimization problem in equation (14). First, the corresponding Lagrangian function Γ_f of equation (14) is defined as:

$$\begin{aligned}\Gamma_f=&\frac{1}{2}Tr\left(I\times\left(YY^T-YD^T W-W^T DY^T+W^T DD^T W\right)\right)\\&+\frac{\lambda_W}{2}Tr\left(WW^T\right)+\frac{\lambda_D}{2}Tr\left(DD^T\right)\\&+\frac{\lambda_1}{2}Tr\left(LD(LD)^T-LDW^T W-W^T W(LD)+W^T WW^T W\right)\\&+\frac{\lambda_2}{2}Tr\left(LE(LE)^T-LED^T D-D^T D(LE)+D^T DD^T D\right)\\&+Tr\left(\Phi W^T+Tr\left(\Psi D^T\right)\right)\end{aligned} \quad (18)$$

where $Tr(\bullet)$ denotes the trace of the matrix, and $\Phi=[\varphi_{ik}]$ and $\Psi=[\psi_{jk}]$ are the constraints $W_{ik}\geq 0$ and $D_{jk}\geq 0$ for Lagrange multipliers. The partial derivatives of W and D are:

$$\begin{aligned}\frac{\partial\Gamma_f}{\partial W}=&I\times\left(-DY^T+DD^T W\right)\\&+\lambda_W W+2\lambda_1\left(-W(LD)+WW^T W\right)+\Phi\end{aligned} \quad (19)$$

$$\begin{aligned}\frac{\partial\Gamma_f}{\partial D}=&I\times\left(-WY^T+WW^T W\right)+\lambda_D D\\&+2\lambda_2\left(-D(LE)+DD^T D\right)+\Psi\end{aligned} \quad (20)$$

Using the Karush-Kuhn-Tucker conditions $\varphi_{ik}W_{ik}=0$ and $\psi_{jk}D_{jk}=0$， the following equations for W_{ik} and D_{jk} can be obtained:

$$\begin{aligned}&\left(I\times\left(-DY^T+DD^T W\right)\right)_{ik}W_{ik}+\left(\lambda_W W\right)_{ik}W_{ik}\\&+\left(2\lambda_1\left(-W(LD)+WW^T W\right)\right)_{ik}W_{ik}=0\end{aligned} \quad (21)$$

$$\begin{aligned}&\left(I\times\left(-WY+WW^T D\right)\right)_{jk}D_{jk}+\left(\lambda_D D\right)_{jk}D_{jk}\\&+\left(2\lambda_2\left(-D(LE)+DD^T D\right)\right)_{jk}D_{jk}=0\end{aligned} \quad (22)$$

Thus, we can obtain the following update rule:

$$W_{ik}\times\frac{\left(I\times\left(DY^T\right)+2\lambda_1\left(W(LD)\right)\right)_{ik}}{\left(I\times\left(DD^T W\right)\right)_{ik}+\left(\lambda_W W\right)_{ik}+\left(2\lambda_1\left(WW^T W\right)\right)_{ik}}\rightarrow W_{ik}^{new} \quad (23)$$

$$D_{jk}\times\frac{\left(I\times(WY)+2\lambda_2\left(D(LE)\right)\right)_{jk}}{\left(I\times\left(WW^T D\right)\right)_{jk}+\left(\lambda_D D\right)_{jk}+\left(2\lambda_2\left(DD^T D\right)\right)_{jk}}\rightarrow D_{jk}^{new} \quad (24)$$

In accordance with equations (20) and (21), the matrices W and D are continuously updated until reaching the objective function's local minimum. Finally, the predicted lncRNA–disease interaction matrix is calculated using the formula $Y^*=W^T D$. In general, the jth column of Y^* indicates the interaction score between disease d_j and the lncRNA, with a higher score indicating a more significant interaction.

Results and discussion

Assessment indicators

Both LOOCV and 5-fold CV methods were utilized to assess the SCCPMD model's efficacy in predicting potential lncRNA–disease associations (Huang et al., 2022c; Sun et al., 2022). Each proven lncRNA–disease association is listed as a test sample in the LOOCV framework, whereas the other unidentified relationship pairings are listed as training samples. All confirmed lncRNA–disease associations are separated into five groups in the 5-fold CV framework, and in each experiment, one group is chosen as the test group and the other as the training group. Using this method, we ran the experiment 100 times and computed the mean of all outcomes. Since the lncRNA–disease dataset only contains a small number of known lncRNA–disease associations and the AUC is known to be insensitive to a skewed class distribution, we used the AUC of the receiver operating characteristic curve to evaluate the performance of SCCPMD (Zhao et al., 2022).

Optimal parameter selection

There are six parameters in SCCPMD: $a, k, \lambda_W, \lambda_D, \lambda_1$, and λ_2. To tease out the effect of these five parameter choices on the model, we performed 100 experiments in the 5-fold CV framework and calculated the average AUC values. First, there is a similarity correction component for parameter a. We searched for the optimal parameter in the range of −1 to −10. Figure 2 clearly shows that the highest AUC value was reached when $a = -4$.

The parameter k represents the number of lncRNA and disease latent feature matrix row vectors, which determines the size of the latent feature matrix. As shown in Figure 3, we restricted the range of k from 10 to 100. The highest AUC value was achieved for SCCPMD when $k = 20$.

Parameters $\lambda_W, \lambda_D, \lambda_1$, and λ_2 exist in the constrained probability matrix decomposition part, which controls the influence of each part in the final update rule of the lncRNA and disease characteristic matrix. As shown in Figures 4, 5, we set the range of all four parameters to be from 0.1 to 1.

Based on the above experiments, the best values of these five parameters were finally determined as $a = -4, k = 20, \lambda_W = 0.8, \lambda_D = 0.6, \lambda_1 = 0.6$, and $\lambda_2 = 0.8$.

Algorithm comparison

To evaluate the predictive performance of the SCCPMD model, SCCPMD was compared with five existing advanced methods: dual sparse collaborative matrix factorization (DSCMF; Liu et al., 2021), geometric matrix completion lncRNA–disease association (GMCLDA; Lu et al., 2020), local random walk-based prediction of human lncRNA and disease associations (Li et al., 2021), probabilistic matrix factorization method for identifying lncRNA–disease associations (PMFILDA; Xuan et al., 2019), and bi-random walks for predicting lncRNA–disease associations (BRWLDA; Yu et al., 2017). As shown in Figure 6, the AUC value of the SCCPMD curve in the LOOCV framework was 0.9787, which was larger than that obtained with the other prediction methods (DSCMF, AUC = 0.9101; GMCLDA, AUC = 0.9086; LRWHNLDA, AUC = 0.9083; PMFILDA, AUC = 0.8850; and BRWLDA, AUC = 0.8376), indicating that the

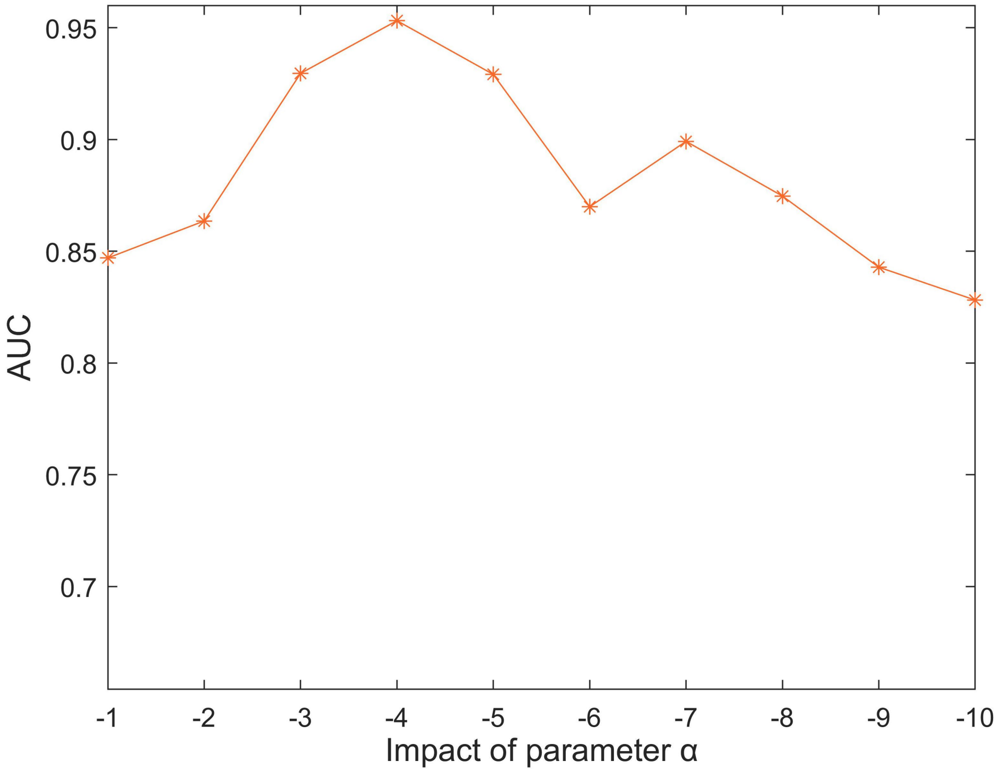

FIGURE 2
The impact of different α values under 5-fold cross-validation.

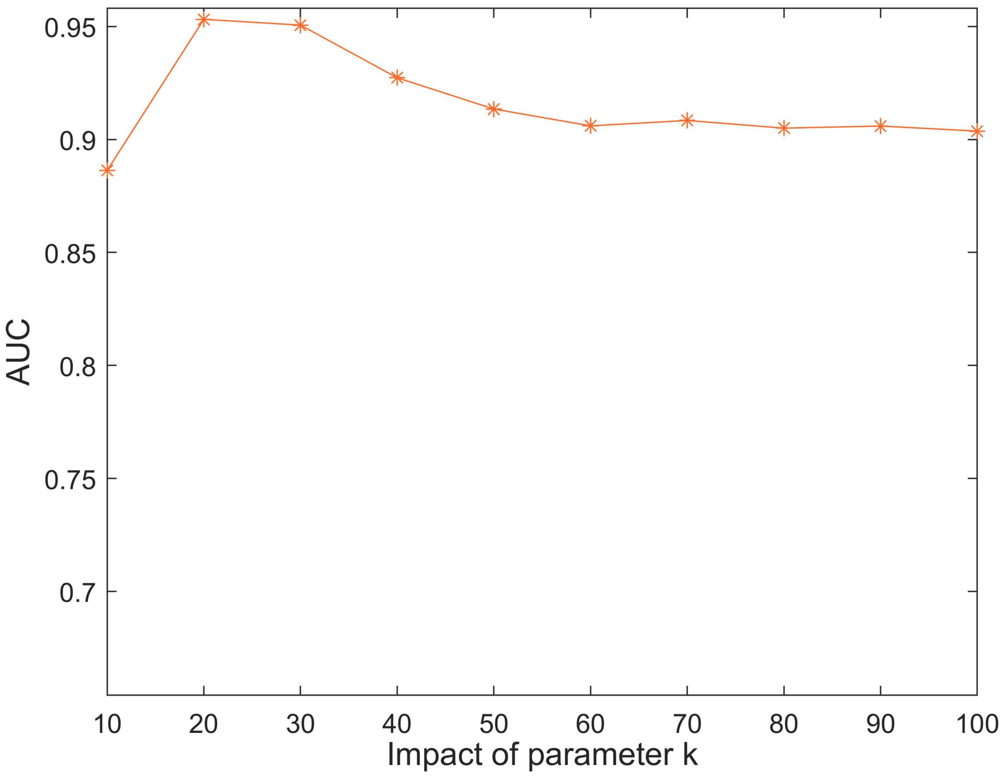

FIGURE 3
The impact of different *k* values under 5-fold cross-validation.

performance of SCCPMD is better than that of existing calculation methods. To further validate the prediction performance of SCCPMD, the 5-fold CV framework was used for validation. As shown in Figure 7, SCCPMD obtained a reliable AUC of 0.9528 ± 0.0036, which was much higher than the AUC values of 0.8946 ± 0.0038, 0.8804 ± 0.0009, 0.8844 ± 0.0014, 0.8705 ± 0.0047, and 0.8172 ± 0.0014 for the comparison methods DSCMF, GMCLDA, LRWHNLDA, PMFILDA, and BRWLDA, respectively. The computational methods we compared were only for lncRNA-disease association pairs, predicting potential associations based on the similarity between lncRNA and disease. The SCCPMD model uses microbe-disease associations to enrich disease similarities, while correcting the similarity matrix to highlight strong similarities and reduce noise in the original similarities. Therefore, SCCPMD shows better performance than these five methods and would be more favorable for the prediction of lncRNA–disease associations.

Case study

Malignancy, as a general term to refer to cancer, has a significant negative impact on human health. With a global annual mortality rate of more than 10 million, cancer remains one of the main contributors to mortality (Zaimy et al., 2017). To validate the actual predictive performance of SCCPMD for lncRNA–disease associations, three cancer types with high hazard were selected as disease case studies: breast cancer, lung cancer, and RCC. The predicted correlations were validated in three lncRNA–disease association databases: the lncRNA disease database, Lnc2cancer database, and MNDR database.

Table 1 shows the top 10 lncRNAs that were predicted to be associated with breast cancer using our model, nine of which have previously been reported to be associated with breast cancer. Breast epithelial cells can become cancerous when they proliferate uncontrollably in response to several oncogenic stimuli (Fahad, 2019). Four lncRNAs, including *LINC00667*, were identified by analysis of gene expression data from 768 breast cancer patients in The Cancer Genome Atlas database, suggesting potential predictive biomarkers for breast cancer with clinical value (Zhu et al., 2020). Among these markers, *PVT1* has been reported to affect mature adipogenic mediators by regulating p21 expression in triple-negative breast cancer cells (Wang et al., 2018b). Functional studies showed that the proliferation, migration, and invasion of breast cancer cells overexpressing *LINC01089* were significantly reduced and that epidermal growth factor reversed these effects (Yuan et al., 2019). *TSIX* is an lncRNA that has been explored as a stable non-invasive breast cancer immunological biomarker, which plays a role in X chromosome inactivation and breast cancer (Salama et al., 2020).

Table 2 shows the top 10 lncRNAs that were predicted to be associated with lung cancer using our model, all of which have been reported to play roles in lung cancer. Despite improvements

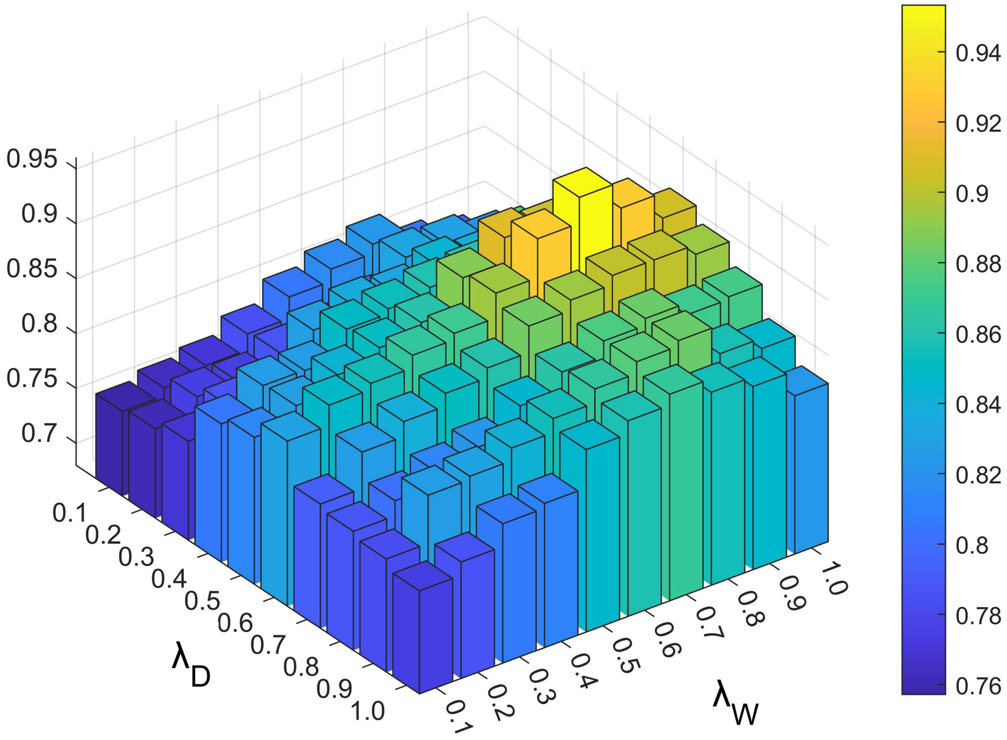

FIGURE 4
The impact of different λ_W and λ_D values under 5-fold cross-validation.

in our knowledge of lung cancer risk, progression, immunologic control, and treatment choices, lung cancer—a malignancy that starts in the bronchial mucosa or glands of the lungs—remains the most common cause of cancer-related death (Bade and Cruz, 2020). Amplification of *PVT1* in lung cancer patients was associated with a poor prognosis for survival. *PVT1* levels are increased in lung cancer cells, which promotes their growth and metastasis both *in vivo* and *in vitro* (Pan et al., 2020). The expression of *SNHG1* in non-small cell lung cancer (NSCLC) tissues and cells is high. Silencing *SNHG1* could suppress the migration and invasion of NSCLC cells, which also promoted apoptosis and decreased the cell proliferation rate (Li and Zheng, 2020). Considerable upregulation of the lncRNA *CDKN2 B-AS1* has been detected in both lung cancer tissues and cell lines (Wang et al., 2020). *In vitro* studies demonstrated that blocking *NEAT1* with short hairpin RNA prevented lung cancer cells from surviving and migrating or invading (Ma et al., 2020). Table 3 shows the top 10 lncRNAs that were predicted to be associated with RCC with our model, all of which have been associated with RCC in previous studies. RCC comprises a group of malignant tumors originating from the renal cortical epithelium, most commonly in the upper pole of the kidney (Pullen Jr, 2021). By inhibiting cell cycle progression and reversing the epithelial-to-mesenchymal transition (EMT) phenotype, *NEAT1* knockdown could reduce the rate of RCC cell proliferation and suppressed RCC migration and invasion (Liu et al., 2017). By controlling EMT-related genes, loss-of-function and gain-of-function pathways demonstrated that *CRNDE* promotes the migration and invasion of clear cell RCC cells (Ding et al., 2018). *MEG3* has been proposed to induce apoptosis in RCC cells by activating the mitochondrial pathway (Wang et al., 2015). Functional assays revealed that *MIAT* knockdown prevented kidney cancer cells from proliferating and metastasizing both *in vitro* and *in vivo* (Qu et al., 2018).

Conclusion

An increasing number of studies have shown that exploration of potential lncRNA–disease associations can be expedited and more effectively performed by developing computational models. Recent results have also showed that matrix decomposition is a reliable method for predicting lncRNA-disease associations. We here propose a novel method to predict unknown lncRNA–disease associations based on corrected similarity added as a constraint to the probability matrix decomposition (SCCPMD). We confirmed the excellent performance of SCCPMD, demonstrating superiority in prediction to existing advanced algorithms, which is attributed to the following three factors: (1) the disease Gaussian similarity obtained by fusing microbe-disease associations calculation can solve the original problem of sparse disease semantic similarity, (2) the corrected similarity performance highlights the effects of strong correlations while reducing the effects of weak correlations, thus reducing the overall noise in the matrix; and (3) introducing lncRNA and disease

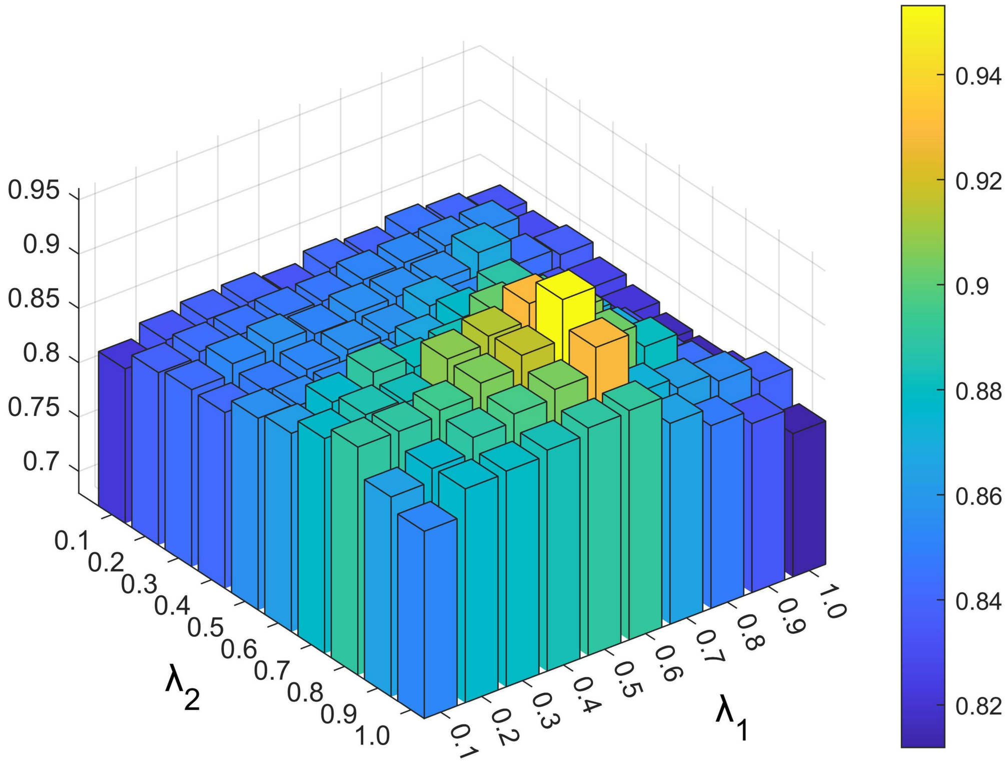

FIGURE 5
The impact of different λ_1 and λ_2 values under 5-fold cross-validation.

TABLE 1 Top 10 lncRNAs predicted by SCCPMD to be connected to breast cancer.

Rank	lncRNA name	Evidence (PubMed ID)
1	*LINC00667*	31,897,133
2	*PVT1*	30,371,726
3	*PINK1-AS*	unknown
4	*LINC01089*	31,417,284
5	*TSIX*	31,998,636
6	*MSR1*	26,967,566
7	*LINC01638*	30,002,443
8	*CDKN2B-AS1*	unknown
9	*H19*	32,124,962
10	*NEAT1*	30,957,286

TABLE 2 Top 10 lncRNAs predicted by SCCPMD to be connected to lung cancer.

Rank	lncRNA name	Evidence (PubMed ID)
1	*PVT1*	33,167,678
2	*SNHG1*	31,788,970, 28,147,312
3	*CDKN2B-AS1*	33,116,641
4	*NEAT1*	32,296,457, 31,646,570
5	*MEG8*	30,262,664
6	*KCNQ1OT1*	31,486,494
7	*MALAT1*	32,141,554
8	*H19*	31,190,899
9	*MEG3*	31,585,300
10	*PCAT6*	30,464,520

similarity constraints in the traditional probability matrix decomposition makes better use of this biological information to improve the prediction performance. The AUC values of SCCPMD in the LOOCV and 5-fold CV frameworks reached up to 0.9787 and 0.9528 ± 0.0036, respectively, which were much higher than those obtained with the comparative algorithms. Additionally, we chose three complex diseases as case studies, demonstrating that SCCPMD performs well with real-world clinical data.

Although SCCPMD enriches disease similarity using microbe-disease associations, prediction results are also affected by microbe-disease associations. In addition, relying on a single lncRNA expression similarity can also make the model limited. Integration of more similarity information is expected to make the proposed model more robust. Therefore, in future work we will try to combine more bioinformatic datasets and fuse multiple lncRNA similarities to improve the robustness and predictive performance of the model.

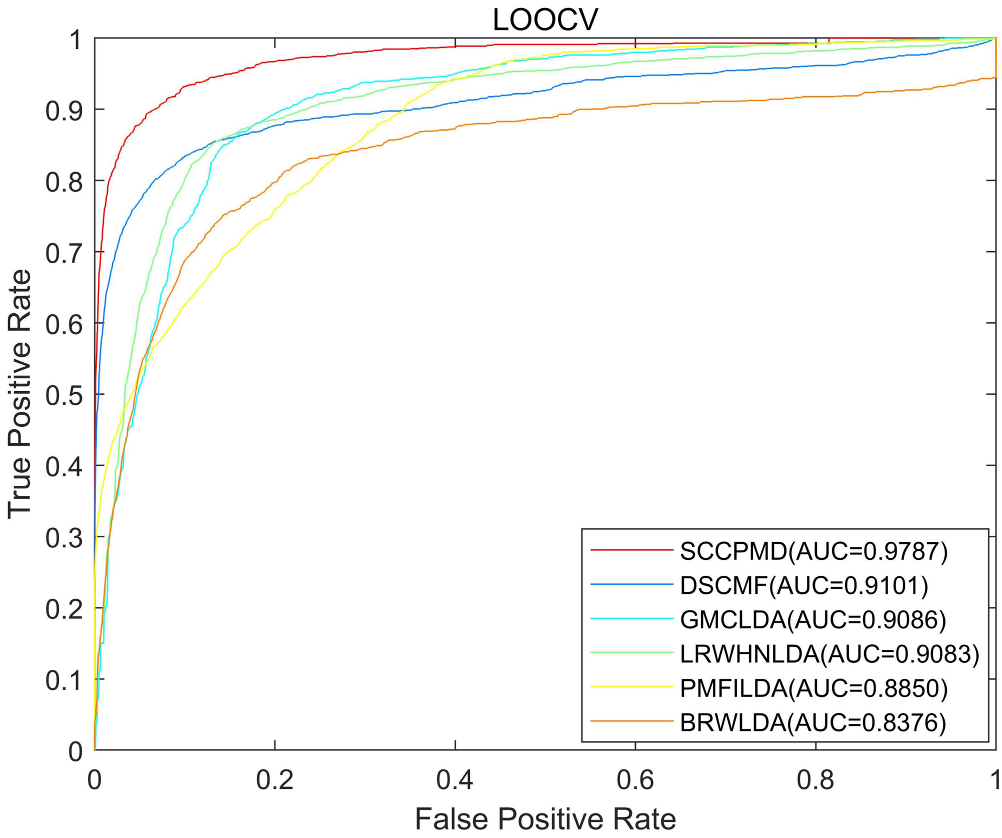

FIGURE 6
Area under the receiver operating characteristic curve (AUC) values of leave-one-out cross-validation (LOOCV) between SCCPMD and the other five comparison models.

TABLE 3 Top 10 lncRNAs predicted by SCCPMD to be connected to renal cell carcinoma.

Rank	lncRNA name	Evidence (PubMed ID)
1	*NEAT1*	28,968,960
2	*CRNDE*	30,129,055
3	*MEG3*	26,223,924
4	*MIAT*	30,041,179
5	*PVT1*	31,040,699, 29,725,470
6	*SNHG5*	32,281,285, 32,194,910
7	*HOTAIRM1*	31,862,408
8	*MEG3*	31,071,531
9	*TUG1*	31,310,753
10	*ZFAS1*	30,841,471

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical review and approval were not required for the study of human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/ participants OR patients/participants legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

LL, HJ, and LC: conceptualization. LL: data curation and resources. LL, RC, YZ, WX, HJ, LC, and MZ: formal analysis and writing—review and editing. LL, RC, and YZ: investigation. LL, RC, YZ, WX, HJ, and LC: methodology and supervision. LL and MZ: project administration. RC, YZ, and WX: validation and writing draft. RC: visualization. All authors contributed to the article and approved the submitted version.

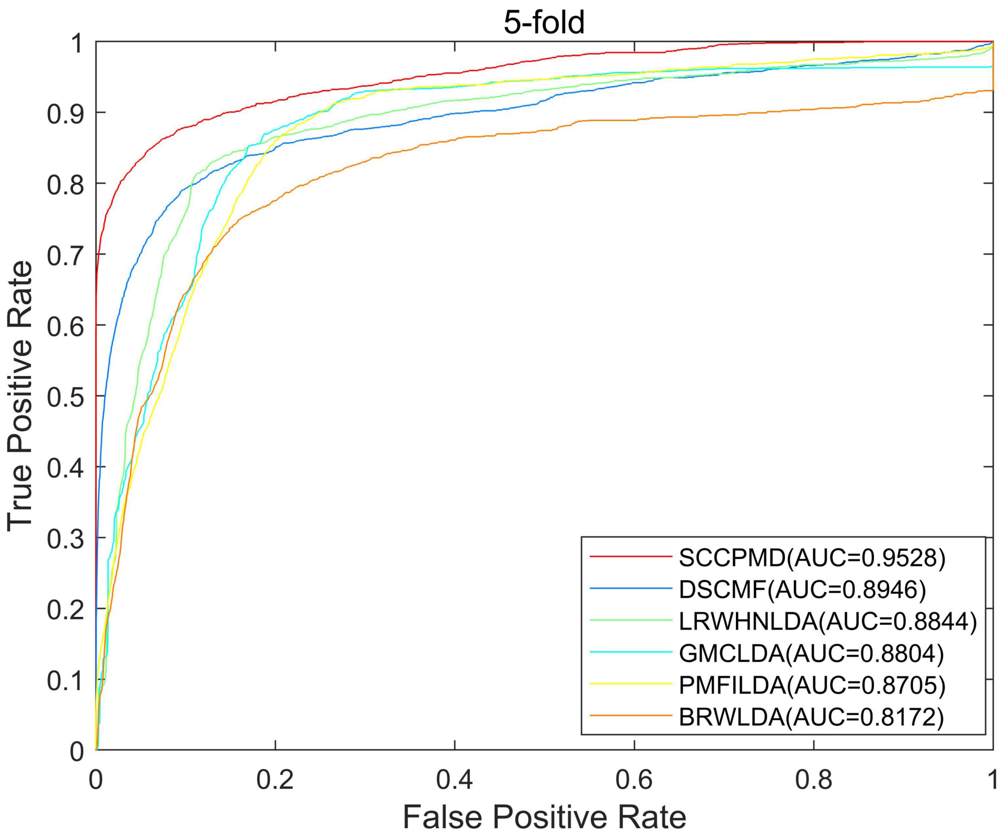

FIGURE 7
Area under the receiver operating characteristic curve (AUC) values of 5-fold cross-validation between SCCPMD and the other five comparison models.

Funding

This work was supported by the National Natural Science Foundation of China (72001202 and 62002070), the Opening Project of Guangdong Province Key Laboratory of Computational Science at Sun Yat-sen University (2021013), the Science and Technology Plan Project of Guangzhou City (202102021236), and the Philosophy and Social Science Co-Construction Project of Guangzhou City (2020GZGJ115).

Acknowledgments

We thank reviewers for valuable suggestions.

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

Bade, B. C., and Cruz, C. S. D. (2020). Lung cancer 2020: epidemiology, etiology, and prevention[J]. *Clin. Chest Med.* 41, 1–24. doi: 10.1016/j.ccm.2019.10.001

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases[J]. *Nucleic Acids Res.* 47, D1034–D1037. doi: 10.1093/nar/gky905

Cao, H. L., Liu, Z. J., Huang, P. L., Yue, Y. L., and Xi, J. N. (2019). lncRNA-RMRP promotes proliferation, migration and invasion of bladder cancer via miR-206[J]. *Eur. Rev. Med. Pharmacol. Sci.* 23, 1012–1021. doi: 10.26355/eurrev_201902_16988

Chen, X., Li, T. H., Zhao, Y., Wang, C. C., and Zhu, C. C. (2021a). Deep-belief network for predicting potential miRNA-disease associations[J]. *Brief. Bioinform.* 22:bbaa186. doi: 10.1093/bib/bbaa186

Chen, X., Sun, Y. Z., Guan, N. N., Qu, J., Huang, Z. A., Zhu, Z. X., et al. (2019a). Computational models for lncRNA function prediction and functional similarity calculation[J]. *Brief. Funct. Genomics* 18, 58–82. doi: 10.1093/bfgp/ely031

Chen, X., Sun, L. G., and Zhao, Y. (2021b). NCMCMDA: miRNA–disease association prediction through neighborhood constraint matrix completion[J]. *Brief. Bioinform.* 22, 485–496. doi: 10.1093/bib/bbz159

Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018a). Predicting miRNA–disease association based on inductive matrix completion[J]. *Bioinformatics* 34, 4256–4265. doi: 10.1093/bioinformatics/bty503

Chen, X., Xie, D., Zhao, Q., and You, Z. H. (2019b). MicroRNAs and complex diseases: from experimental results to computational models[J]. *Brief. Bioinform.* 20, 515–539. doi: 10.1093/bib/bbx130

Chen, X., and Yan, G. Y. (2013). Novel human lncRNA–disease association inference based on lncRNA expression profiles[J]. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426

Chen, X., Yan, C. C., Zhang, X., and You, Z. H. (2017). Long non-coding RNAs and complex diseases: from experimental results to computational models[J]. *Brief. Bioinform.* 18, 558–576. doi: 10.1093/bib/bbw060

Chen, X., Yin, J., Qu, J., and Huang, L. (2018b). MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction[J]. *PLoS Comput. Biol.* 14:e1006418. doi: 10.1371/journal.pcbi.1006418

Chen, X., Zhu, C. C., and Yin, J. (2019c). Ensemble of decision tree reveals potential miRNA-disease associations[J]. *PLoS Comput. Biol.* 15:e1007209. doi: 10.1371/journal.pcbi.1007209

Ding, C., Han, F., Xiang, H., Xia, X., Wang, Y., Dou, M., et al. (2018). LncRNA CRNDE is a biomarker for clinical progression and poor prognosis in clear cell renal cell carcinoma[J]. *J. Cell. Biochem.* 119, 10406–10414. doi: 10.1002/jcb.27389

Fahad, U. M. (2019). Breast cancer: current perspectives on the disease status[J]. *Breast Cancer Metastasis and Drug Resistance.* 1152, 51–64. doi: 10.1007/978-3-030-20301-6_4

Fu, G., Wang, J., Domeniconi, C., and Yu, G. (2018). Matrix factorization-based data fusion for the prediction of lncRNA–disease associations[J]. *Bioinformatics* 34, 1529–1537. doi: 10.1093/bioinformatics/btx794

Gao, M. M., Cui, Z., Gao, Y. L., Wang, J., and Liu, J. X. (2021). Multi-label fusion collaborative matrix factorization for predicting LncRNA-disease associations[J]. *IEEE J. Biomed. Health Inform.* 25, 881–890. doi: 10.1109/JBHI.2020.2988720

Hill, M., and Tran, N. (2021). miRNA interplay: mechanisms and consequences in cancer[J]. *Dis. Model. Mech.* 14:dmm047662. doi: 10.1242/dmm.047662

Huang, L., Zhang, L., and Chen, X. (2022a). Updated review of advances in microRNAs and complex diseases: experimental results, databases, webservers and data fusion[J]. *Brief. Bioinform.* 23:bbac397. doi: 10.1093/bib/bbac397

Huang, L., Zhang, L., and Chen, X. (2022b). Updated review of advances in microRNAs and complex diseases: taxonomy, trends and challenges of computational models[J]. *Brief. Bioinform.* 23:bbac358. doi: 10.1093/bib/bbac358

Huang, L., Zhang, L., and Chen, X. (2022c). Updated review of advances in microRNAs and complex diseases: towards systematic evaluation of computational models[J]. *Brief. Bioinform.* 23:bbac407. doi: 10.1093/bib/bbac407

Lan, W., Lai, D., Chen, Q., Wu, X., Chen, B., Liu, J., et al. (2022). LDICDL: LncRNA-disease association identification based on collaborative deep learning[J]. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19, 1715–1723. doi: 10.1109/TCBB.2020.3034910

Li, J., Zhao, H., Xuan, Z., Yu, J., Feng, X., Liao, B., et al. (2021). A novel approach for potential human LncRNA-disease association prediction based on local random walk[J]. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1049–1059. doi: 10.1109/TCBB.2019.2934958

Li, X., and Zheng, H. (2020). LncRNA SNHG1 influences cell proliferation, migration, invasion, and apoptosis of non-small cell lung cancer cells via the miR-361-3p/FRAT1 axis[J]. *Thoracic Cancer.* 11, 295–304. doi: 10.1111/1759-7714.13256

Liu, F., Chen, N., Gong, Y., Xiao, R., Wang, W., and Pan, Z. (2017). The long non-coding RNA NEAT1 enhances epithelial-to-mesenchymal transition and chemoresistance via the miR-34a/c-met axis in renal cell carcinoma[J]. *Oncotarget* 8, 62927–62938. doi: 10.18632/oncotarget.17757

Liu, J. X., Gao, M. M., Cui, Z., Gao, Y. L., and Li, F. (2021). DSCMF: prediction of LncRNA-disease associations based on dual sparse collaborative matrix factorization[J]. *BMC bioinformatics.* 22:241. doi: 10.1186/s12859-020-03868-w

Lu, C., Yang, M., Li, M., Li, Y., Wu, F. X., and Wang, J. (2020). Predicting human lncRNA-disease associations based on geometric matrix completion[J]. *IEEE J. Biomed. Health Inform.* 24, 2420–2429. doi: 10.1109/JBHI.2019.2958389

Ma, F., Lei, Y. Y., Ding, M. G., Luo, L. H., Xie, Y. C., and Liu, X. L. (2020). LncRNA NEAT1 interacted with DNMT1 to regulate malignant phenotype of cancer cell and cytotoxic T cell infiltration via epigenetic inhibition of p53, cGAS, and STING in lung cancer[J]. *Front. Genet.* 11:250. doi: 10.3389/fgene.2020.00250

Pan, Y., Liu, L., Cheng, Y., Yu, J., and Feng, Y. (2020). Amplified LncRNA PVT1 promotes lung cancer proliferation and metastasis by facilitating VEGFC expression[J]. *Biochem. Cell Biol.* 98, 676–682. doi: 10.1139/bcb-2019-0435

Pullen, R. L. Jr. (2021). Renal cell carcinoma, part 1[J]. *Nursing* 51, 34–40. doi: 10.1097/01.NURSE.0000753972

Qu, Y., Xiao, H., Xiao, W., Xiong, Z., Hu, W., Gao, Y., et al. (2018). Upregulation of MIAT regulates LOXL2 expression by competitively binding MiR-29c in clear cell renal cell carcinoma[J]. *Cell. Physiol. Biochem.* 48, 1075–1087. doi: 10.1159/000491974

Salama, E. A., Adbeltawab, R. E., and El Tayebi, H. M. (2020). XIST and TSIX: novel cancer immune biomarkers in PD-L1-overexpressing breast cancer patients[J]. *Front. Oncol.* 9:1459. doi: 10.3389/fonc.2019.01459

Sun, J., Shi, H., Wang, Z., Zhang, C., Liu, L., Wang, L., et al. (2014). Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network[J]. *Mol. BioSyst.* 10, 2074–2081. doi: 10.1039/c3mb70608g

Sun, F., Sun, J., and Zhao, Q. (2022). A deep learning method for predicting metabolite–disease associations via graph neural network[J]. *Brief. Bioinform.* 23:bbac266. doi: 10.1093/bib/bbac266

Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation[J]. *PLoS Comput. Biol.* 6:e1000641. doi: 10.1371/journal.pcbi.1000641

Wang, C. C., Han, C. D., Zhao, Q., and Chen, X. (2021a). Circular RNAs and complex diseases: from experimental results to computational models[J]. *Brief. Bioinform.* 22:bbab286. doi: 10.1093/bib/bbab286

Wang, M., Huang, T., Luo, G., Huang, C., Xiao, X. Y., Wang, L., et al. (2015). Long non-coding RNA MEG3 induces renal cell carcinoma cells apoptosis by activating the mitochondrial pathway[J]. *J. Huazhong Univ. Sci. Technolog. Med. Sci.* 35, 541–545. doi: 10.1007/s11596-015-1467-5

Wang, J., Su, Z., Lu, S., Fu, W., Liu, Z., Jiang, X., et al. (2018a). LncRNA HOXA-AS2 and its molecular mechanisms in human cancer[J]. *Clin. Chim. Acta* 485, 229–233. doi: 10.1016/j.cca.2018.07.004

Wang, L., Wang, R., Ye, Z., Wang, Y., Li, X., Chen, W., et al. (2018b). PVT1 affects EMT and cell proliferation and migration via regulating p21 in triple-negative breast cancer cells cultured with mature adipogenic medium[J]. *Acta Biochim. Biophys. Sin.* 50, 1211–1218. doi: 10.1093/abbs/gmy129

Wang, G., Xu, G., and Wang, W. (2020). Long noncoding RNA CDKN2B-AS1 facilitates lung cancer development through regulating miR-378b/NR2C2[J]. *Onco. Targets. Ther.* 13, 10641–10649. doi: 10.2147/OTT.S261973

Wang, M. N., You, Z. H., Wang, L., Li, L. P., and Zheng, K. (2021b). LDGRNMF: LncRNA-disease associations prediction based on graph regularized non-negative matrix factorization[J]. *Neurocomputing* 424, 236–245. doi: 10.1016/j.neucom.2020.02.062

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[J]. *Brief. Bioinform.* 23:bbac463. doi: 10.1093/bib/bbac463

Xie, G., Chen, H., Sun, Y., Gu, G., Lin, Z., Wang, W., et al. (2021). Predicting circRNA-disease associations based on deep matrix factorization with multi-source fusion[J]. *Interdisciplinary Sciences: Computational Life Sciences.* 13, 582–594. doi: 10.1007/s12539-021-00455-2

Xing, C., Sun, S., Yue, Z. Q., and Bai, F. (2021). Role of lncRNA LUCAT1 in cancer[J]. *Biomed. Pharmacother.* 134:111158. doi: 10.1016/j.biopha.2020.111158

Xuan, Z., Li, J., Yu, J., Feng, X., Zhao, B., and Wang, L. (2019). A probabilistic matrix factorization method for identifying lncRNA-disease associations[J]. *Genes.* 10:126. doi: 10.3390/genes10020126

Yu, G., Fu, G., Lu, C., Ren, Y., and Wang, J. (2017). BRWLDA: bi-random walks for predicting lncRNA-disease associations[J]. *Oncotarget* 8, 60429–60446. doi: 10.18632/oncotarget.19588

Yuan, H., Qin, Y., Zeng, B., Feng, Y., Li, Y., Xiang, T., et al. (2019). Long noncoding RNA LINC01089 predicts clinical prognosis and inhibits cell proliferation and invasion through the Wnt/β-catenin signaling pathway in breast cancer[J]. *Onco. Targets. Ther.* 12, 4883–4895. doi: 10.2147/OTT.S208830

Zaimy, M. A., Saffarzadeh, N., Mohammadi, A., Pourghadamyari, H., Izadi, P., Sarli, A., et al. (2017). New methods in the diagnosis of cancer and gene therapy of cancer based on nanoparticles[J]. *Cancer Gene Ther.* 24, 233–243. doi: 10.1038/cgt.2017.16

Zhang, Y., Chen, M., Li, A., Cheng, X., Jin, H., and Liu, Y. (2020a). LDAI-ISPS: LncRNA–disease associations inference based on integrated space projection scores[J]. *Int. J. Mol. Sci.* 21:1508. doi: 10.3390/ijms21041508

Zhang, H., Liang, Y., Han, S., Peng, C., and Li, Y. (2019a). Long noncoding RNA and protein interactions: from experimental results to computational models based on network methods[J]. *Int. J. Mol. Sci.* 20:1284. doi: 10.3390/ijms20061284

Zhang, H., Liang, Y., Peng, C., Han, S., du, W., and Li, Y. (2019b). Predicting lncRNA-disease associations using network topological similarity based on deep mining heterogeneous networks[J]. *Math. Biosci.* 315:108229. doi: 10.1016/j.mbs.2019.108229

Zhang, L., Liu, T., Chen, H., Zhao, Q., and Liu, H. (2021a). Predicting lncRNA–miRNA interactions based on interactome network and graphlet interaction[J]. *Genomics* 113, 874–880. doi: 10.1016/j.ygeno.2021.02.002

Zhang, P., Meng, J., Luan, Y., and Liu, C. (2020b). Plant miRNA–lncRNA interaction prediction with the ensemble of CNN and IndRNN[J]. *Interdisciplinary Sciences: Computational Life Sciences.* 12, 82–89. doi: 10.1007/s12539-019-00351-w

Zhang, L., Yang, P., Feng, H., Zhao, Q., and Liu, H. (2021b). Using network distance analysis to predict lncRNA–miRNA interactions[J]. *Interdisciplinary Sciences: Computational Life Sciences.* 13, 535–545. doi: 10.1007/s12539-021-00458-z

Zhang, Y., Ye, F., Xiong, D., and Gao, X. (2020c). LDNFSGB: prediction of long non-coding rna and disease association using network feature similarity and gradient boosting[J]. *BMC bioinformatics.* 21:377. doi: 10.1186/s12859-020-03721-0

Zhao, J. X., Sun, J. Q., et al. (2022). Predicting potential interactions between lncRNAs and proteins via combined graph auto-encoder methods. *Brief. Bioinform.* doi: 10.1093/bib/bbac527

Zhao, T., Xu, J., Liu, L., Bai, J., Xu, C., Xiao, Y., et al. (2015). Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features[J]. *Mol. BioSyst.* 11, 126–136. doi: 10.1039/c4mb00478g

Zhou, J. R., You, Z. H., Cheng, L., and Ji, B. Y. (2021). Prediction of lncRNA-disease associations via an embedding learning HOPE in heterogeneous information networks[J]. *Molecular Therapy-Nucleic Acids.* 23, 277–285. doi: 10.1016/j.omtn.2020.10.040

Zhu, M., Lv, Q., Huang, H., Sun, C., Pang, D., and Wu, J. (2020). Identification of a four-long non-coding RNA signature in predicting breast cancer survival[J]. *Oncol. Lett.* 19, 221–228. doi: 10.3892/ol.2019.11063

Zhu, R., Wang, Y., Liu, J. X., and Dai, L. Y. (2021). IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier[J]. *BMC bioinformatics.* 22:175. doi: 10.1186/s12859-021-04104-9