



## OPEN ACCESS

## EDITED BY

Qi Zhao,  
University of Science and Technology  
Liaoning, China

## REVIEWED BY

Yi Xiong,  
Shanghai Jiao Tong University, China  
Li Zhang,  
China University of Mining and  
Technology, China

## \*CORRESPONDENCE

Lieqing Lin  
tiger@gdut.edu.cn

## SPECIALTY SECTION

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

RECEIVED 09 September 2022

ACCEPTED 06 October 2022

PUBLISHED 17 November 2022

## CITATION

Chen L, Lin D, Xu H, Li J and Lin L  
(2022) WLLP: A weighted  
reconstruction-based linear label  
propagation algorithm for predicting  
potential therapeutic agents for  
COVID-19.

*Front. Microbiol.* 13:1040252.  
doi: 10.3389/fmicb.2022.1040252

## COPYRIGHT

© 2022 Chen, Lin, Xu, Li and Lin. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# WLLP: A weighted reconstruction-based linear label propagation algorithm for predicting potential therapeutic agents for COVID-19

Langcheng Chen<sup>1</sup>, Dongying Lin<sup>2</sup>, Haojie Xu<sup>2</sup>, Jianming Li<sup>2</sup>  
and Lieqing Lin<sup>1\*</sup>

<sup>1</sup>Center of Campus Network and Modern Educational Technology, Guangdong University of Technology, Guangzhou, China, <sup>2</sup>School of Computer Science, Guangdong University of Technology, Guangzhou, China

The global coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV) has led to a huge health and economic crises. However, the research required to develop new drugs and vaccines is very expensive in terms of labor, money, and time. Owing to recent advances in data science, drug-repositioning technologies have become one of the most promising strategies available for developing effective treatment options. Using the previously reported human drug virus database (HDVD), we proposed a model to predict possible drug regimens based on a weighted reconstruction-based linear label propagation algorithm (WLLP). For the drug–virus association matrix, we used the weighted  $K$ -nearest known neighbors method for preprocessing and label propagation of the network based on the linear neighborhood similarity of drugs and viruses to obtain the final prediction results. In the framework of 10 times 10-fold cross-validated area under the receiver operating characteristic (ROC) curve (AUC), WLLP exhibited excellent performance with an AUC of  $0.8828 \pm 0.0037$  and an area under the precision-recall curve of  $0.5277 \pm 0.0053$ , outperforming the other four models used for comparison. We also predicted effective drug regimens against SARS-CoV-2, and this case study showed that WLLP can be used to suggest potential drugs for the treatment of COVID-19.

## KEYWORDS

COVID-19, drug repositioning, linear neighborhood similarity, label propagation, WKNKN

## 1. Introduction

In November 2019, a novel coronavirus disease broke out in Wuhan, China, for unknown reasons, which was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO) (Zhu et al., 2020). COVID-19 is caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). To date, seven human coronaviruses (HCoV) have been identified, namely HCoV-229E, HCoV-OC43,

HCoV-NL63, HCoV-HKU1, SARS-CoV, Middle East respiratory syndrome (MERS) coronavirus (MERS-CoV), and SARS-CoV-2. Specifically, HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 are frequently found and have low pathogenicity, generally causing only common cold symptoms, whereas MERS-CoV and SARS-CoV are zoonotic viruses that are first reported in the twenty-first century (Sohrabi et al., 2020). SARS-CoV-2 is recognized as the most pathogenic human coronavirus ever discovered (Guan et al., 2020). As of September 2022, 613 million confirmed SARS-CoV-2 infections were reported around the world, with nearly 6 million deaths (Organization, 2020). Until now, there is no cure for COVID-19.

Despite substantial increases in investment by pharmaceutical companies in response to COVID-19, the successful development and approval of a new drug typically requires billions of dollars and an average of 10 years (Liu S. et al., 2020), with the disadvantages of being time consuming (Pushpakom et al., 2019), expensive, and risky. Therefore, drug repositioning (drug repurposing) has been identified as a viable solution to improve the overall process of drug development, especially following recent advances in information technology and data science. The primary goal of drug repositioning is the use of existing drugs to treat new symptoms. Compared with traditional drug development methods, drug repositioning can significantly reduce research and development time and costs while minimizing risks. In short, drug repositioning is considered a promising strategy to accelerate the development of COVID-19 therapeutics.

As Xue et al. (2018) described, current work on drug repositioning is supported by various prediction models, among which the association prediction models for computational drug repositioning applicable to COVID-19 can be broadly classified into the following three categories (Dotolo et al., 2021): (I) network-based models, (II) artificial intelligence algorithms, and (III) matrix completion.

Network-based approaches construct heterogeneous networks by integrating multiple data to predict drug–virus associations. Such approaches are mostly based on the assumption that drugs with similar functions are often associated with viruses having similar phenotypes (Chen et al., 2018b). Prediction approaches based on complex networks (Liu et al., 2022b) have important and widespread applications in drug repositioning because of their ability to integrate multiple datasets of interest (Fan et al., 2020; Zhou et al., 2020). More specifically, network nodes represent drugs, diseases, viruses, or genes, while edges represent interactions or relationships between nodes (Re and Valentini, 2013; Chen et al., 2015). The obtained predictions may contribute to the process of structure-directed drug and diagnostic research and help to identify new potential biological targets (Barlow et al., 2020). In this regard, there are two network-based approaches applicable to drug repositioning for COVID-19: the network-based clustering approach and the network-based propagation

approach. Macropoul et al. (2009) proposed the repeated random walks (RRW) method that uses RRW on the protein–protein interaction (PPI) network for local clustering of the network and then predicts some protein complexes. Although this was found to be a precise and general approach, it requires a great deal of time and memory overhead and cannot detect overlapping clusters. King et al. (2012) introduced a new model named restricted neighborhood search clustering (RNSC), which is a global network algorithm for identifying protein clusters on PPI networks. It considers both global and local information from the network and can also detect overlapping clusters, but some information may be lost if the cluster size is too small. Luo et al. (2016) proposed the bidirectional random walk (BiRW) algorithm for predicting relationships between diseases and drugs. It uses the similarity of diseases and drugs with the original correlation matrix to form a heterogeneous network and then clusters this network by a double random walk. The resulting prediction is accurate, but more biological information is needed to improve the confidence of the similarity metric. In addition to the network clustering approach, Vanunu et al. (2010) proposed an overall propagation algorithm called PRINCE, which combines weighted PPI and disease similarity networks for overall disease gene ranking and protein complex association inference. An integrated propagation method for predicting propagation strategies in different sub-networks was proposed by Martinez et al. (2015) and named DrugNet. Zhang et al. (2017b) developed the linear neighborhood similarity (LNS) method to calculate drug–drug similarities in the drug characteristic space. Peng et al. (2021), in response to COVID-19, combined the virus–drug association network topology and a random walk with restart method (VDA-RWR) to predict potential drug–virus associations using a  $2 \times 2$  similarity matrix and known associations between drugs and viruses. Zhang et al. (2021b) developed a network distance analysis model for the prediction of lncRNA–miRNA association (NDALMA). It is worth mentioning that the primary approach in recent years has been to update the network mainly by similarity and network inference (Zhang et al., 2021a; Liu et al., 2022a).

For drug repositioning, artificial intelligence-based models mainly use machine learning methods. Numerous common machine learning algorithms have been applied to predict potential therapeutic agents, such as decision trees (Chen et al., 2019b) and Laplacian regularization (Chen and Huang, 2017). The influence of deep learning models that belong to machine learning has been particularly remarkable (Chen et al., 2019a, 2021a; Keshavarzi Arshadi et al., 2020). In terms of prediction, graph convolutional neural networks (GCNNs) are the most popular tools for drug discovery applications because they can process graphs and extract features by encoding adjacency information within features to learn representations from molecules. Based on drug–target interactions in this model, Tornø and Altman (2019) made correlation predictions. In recent years, sequence-based models,

such as genomics, proteomics, and transcriptomics, have also attracted considerable attention. Vaswani et al. (2017) and Devlin et al. (2018) advanced a transformer model for extracting features from sequences through the attention mechanism and self-supervision, which are widely used in the field of natural language processing. Moreover, Shin et al. (2019) demonstrated that drug–target interactions can be predicted by using the transformer model. Pollastri et al. (2002) demonstrated that recurrent neural networks (RNNs) and long short-term memory (LSTM) networks can predict the secondary structure of molecular or protein sequences. Through an ensemble strategy of three mainstream machine learning algorithms, Hu et al. (2018) proposed a model named HLPI-Ensemble that was specifically designed for human lncRNA–protein interactions. Matrix completion mainly relies on the matrix decomposition algorithms (Chen et al., 2018a,c). Specifically, these algorithms decompose a matrix into two lower-order potential factor matrices based on known association matrices of diseases and drugs (Liu H. et al., 2020). Gönen (Gönen, 2012) put forward a prediction method by using Bayesian probabilistic matrix factorization (BPMF) based on chemical and nuclear genomes. Yang et al. (2019) developed a model based on bounded nuclear norm regularization for drug repositioning. Considering the similarity information between drugs and diseases, Meng et al. (2021) proposed a method called similarity-constrained PMF (SCPMF) to examine the potential value of existing drugs. Liu et al. (2022b) proposed a new computational method *via* deep forest ensemble learning based on an autoencoder (DFELMDA) to predict miRNA–disease associations.

The novel similarity measure of LNS proposed by Zhang et al. has been successfully applied to several bioinformatics problems (Zhang et al., 2017a, 2018a). In this method, the data points are reconstructed by linear neighborhood information and are used to measure the similarity between two points in the association network. Inspired by this, we applied this similarity measure to our model. In recent years, label propagation has been widely used for biological association prediction owing to its various advantages, such as simple logic algorithm and fast optimization. Thus, we adopted the label propagation method for network propagation of the drug–virus association matrix.

Herein, we reported on the development of a method termed label propagation through linear neighborhood similarity for the prediction of undetected drug–virus associations. More specifically, we represented drugs or viruses as feature vectors and treated them as data points in the feature space, from which we computed pairwise linear neighborhood similarities between drugs and drugs or between viruses and viruses. The computed drug and virus similarities and the known disease–virus association networks were treated as a weighted directed graph, which was then input to the label propagation algorithm. Each drug–virus interaction was scored using the label propagation method. Experiments showed that the WLLP model offered superior prediction results when compared

with other models, with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.8828 in the framework of 10 times 10-fold cross-validated.

## 2. Materials and methods

### 2.1. Experimental data

#### 2.1.1. Human drug virus database

The collection of data concerning viruses, drugs, and drug–virus associations is a crucial precursor to using bioinformatics methods to predict novel drug–virus associations. Moreover, systematic collection and management of relevant information are important for further studying the mechanism of virus action (Wang et al., 2021). Meng et al. (2021) collected a large number of experimentally validated drug–virus interaction entries from the literature by using text mining techniques and then constructed the HDVD, which is a database of human drug–virus associations. The HDVD includes 34 viruses, 219 drugs, and 455 confirmed human drug–virus interactions.

#### 2.1.2. Construction of the drug–virus interaction network

From the HDVD dataset, we constructed an association network using known drug–virus interactions, where the points represent the drugs and viruses and the edges represent drug–virus associations. Let  $G = (D, V, I)$  represents the drug–virus association network, where  $D = \{d_1, d_2, \dots, d_n\}$  represents the known drugs in the dataset,  $V = \{v_1, v_2, \dots, v_m\}$  represents the known viruses in the dataset, and  $I$  represents the interaction relationship between  $D$  and  $V$ . Let  $A_{n \times m}$  represents the adjacency matrix of graph  $G$ . If  $d_i$  and  $v_j$  are related,  $A_{ij} = 1$ ; otherwise,  $A_{ij} = 0$ . Also, let  $A^T$  represent the inversion of  $A_{n \times m}$ .

#### 2.1.3. Chemical structure similarity of drug pairs

The chemical structure similarity between two drugs can be calculated from their molecular structure information. In the current study, we downloaded the chemical structure information of drugs from the DrugBank database in the SMILES format (Öztürk et al., 2016) and then calculated their molecular access system (MACCS) fingerprints (O'Boyle et al., 2011). Finally, we used the Tanimoto index to measure the absolute similarity between two molecules (Bajusz et al., 2015). Specifically, we set two drug molecules as A and B, respectively,  $a$  is the number of bits in molecule A, and  $b$  is number of bits in molecule B.  $c$  is the number of bits that are in both molecules. The formula is as follows:

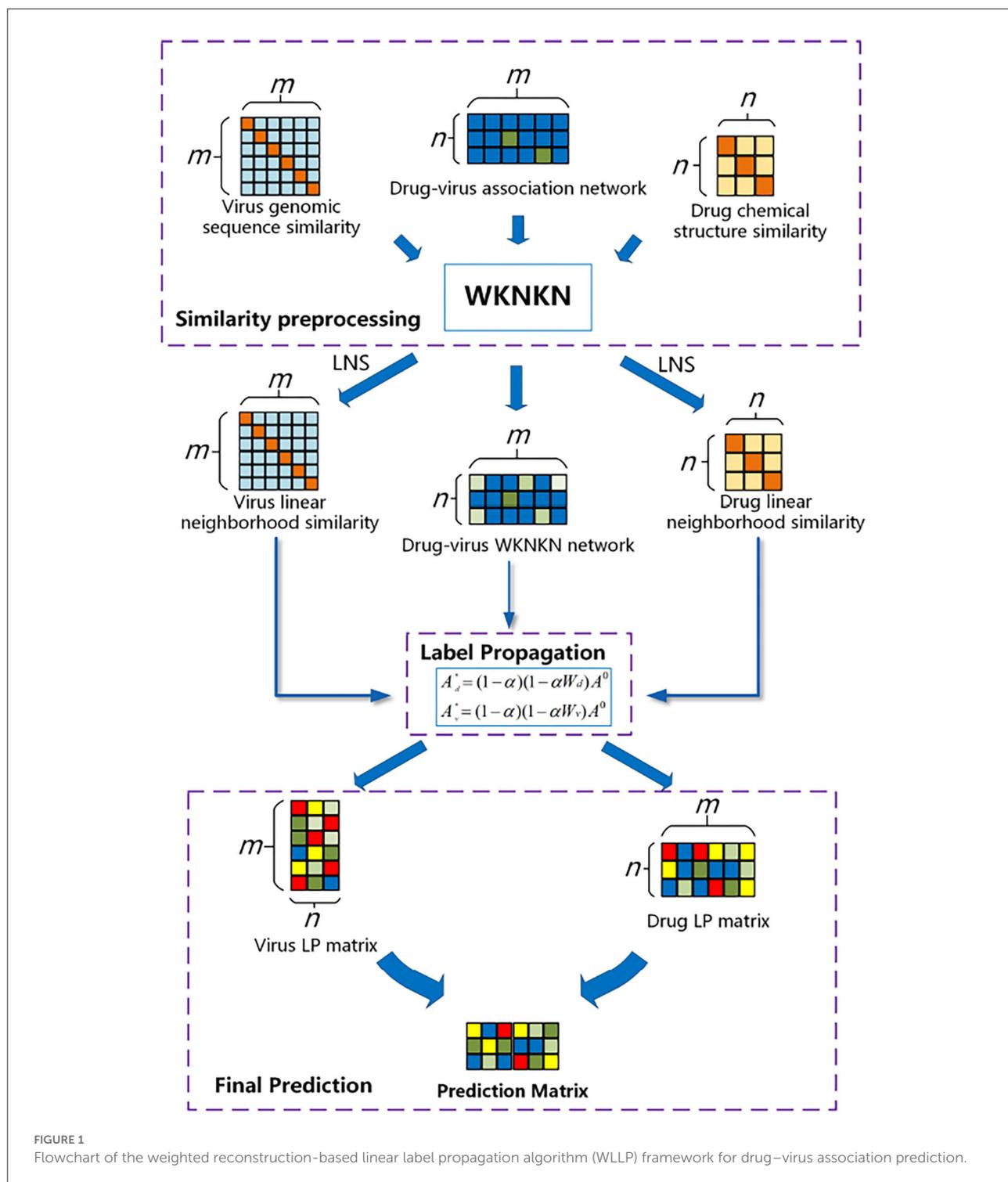
$$T = c/(a + b - c) \quad (1)$$

We used this formula to construct the drug chemical structure similarity matrix  $DD_{n \times n}$ . This is a two-dimensional

matrix whose values represent the chemical fingerprint scores between drugs. In general, the size of this score is between 0 and 1, with larger values representing greater drug–drug similarity.

#### 2.1.4. Genomic sequence similarity of virus pairs

The sequence similarity between viruses can be calculated from their genomic nucleotide sequences. We downloaded the



genomic nucleotide sequences of viruses from the National Center for Biotechnology Information (Wheeler et al., 2002). To calculate the sequence similarity, we used the multiple sequence alignment program MAFFT on account of its high performance (Katoh and Standley, 2013). Finally, the virus sequence similarity matrix  $VV_{m \times m}$  was constructed, which is a two-dimensional matrix whose values represent the sequence similarity between viruses. In general, the value of this matrix is between 0 and 1, and larger values represent greater virus–virus similarity.

## 2.2. Methods

### 2.2.1. Overview of WLLP

In this study, we developed the WLLP framework for predicting disease–virus associations based on LNS in conjunction with label propagation. As shown in Figure 1, the framework consists of three main steps: (I) Label set preprocessing: considering the sparse nature of the drug–virus interaction matrix, we introduced the weighted K-nearest known neighbors (WKNKN) algorithm to make a correction for the potential interactions between the drugs and viruses. (II) LNS information for the drugs and viruses was mined separately based on drug–virus interaction information. (III) Label propagation: a weighted directed graph consisting of known association information, drug–drug LNS, and virus–virus LNS matrices was constructed, and the drug label information was iteratively updated by the label propagation algorithm to reveal unknown potential drug–virus associations.

The flowchart of the WLLP algorithm is shown in Algorithm 1. The details of the principle and process of each WLLP module are described in the following sections.

### 2.2.2. WKNKN

Because it is hard to construct expression datasets, coming up with datasets that contain a large number of samples is generally difficult. A small number of samples complicates the knowledge discovery task (Sirin et al., 2016). The unknown nature of a large part of the information makes the drug–virus association matrix very sparse. Here, we used the WKNKN algorithm to preprocess the original association matrix (Ezzat et al., 2016). Specifically, WKNKN replaces  $A_{0ij}=0$  with the interaction likelihood in the following three steps:

**Step 1.** For each known drug, the chemical structure similarities of the closest  $K$  known drugs are calculated by the  $k$ -nearest neighbors (KNN) method and their corresponding interaction profiles are used to estimate the interaction likelihood profiles. The derived formula is

$$A_d(i, :) = \frac{\sum_{k=1}^K T^{k-1} DD(i, D_k) A(D_k, :)}{\sum_{k=1}^K DD(i, D_k)}, \quad (2)$$

**Input:** Matrices  $A_{n \times m}$ ,  $DD_{n \times n}$ , and  $VV_{m \times m}$ ; Number of neighbors  $K$  and decay factor  $r$ ; LNS size parameters  $dN$  and  $DN$ ; Probability retention factors for drugs and viruses,  $\alpha$  and  $\beta$ , respectively;

**Output:** Predictive association matrix  $A_{m \times n}^*$

```

1: Step 1: Reconstruct the association matrix
2: for  $i = 1$  to  $n$  do
3:   construct  $A_d$  using Equation (2)
4: for  $i = 1$  to  $n$  do
5:   construct  $A_v$  using Equation (3)
6:  $A_{dv} = (A_d + A_v^T) / 2$ 
7:  $A = \max(A, A_{dv})$ 
8: Step 2: Construct the drug LNS matrix  $W_d$ 
9: for  $i = 1$  to  $DN$  do
10:   construct refactoring weight  $w_i$  for each
      drug using Equation (5)
11: Step 3: Similarly, construct the virus LNS
      matrix  $W_v$ 
12: Step 4: Update the associated network by label
      propagation
13: Predict new association matrix  $A_d^*$  using
      Equation (8) and weight  $W_d$ 
14: Similarly, predict the new association matrix
       $A_v^*$ 
15:  $A^* = (A_d^* + A_v^{*T}) / 2$ 
16: return  $A^*$ 

```

Algorithm 1. WLLP.

where  $i$  denotes the drug index,  $T$  is the decay factor, and in general,  $T \leq 1$ .  $D_k$  denotes the  $k$ -th drug index that is most similar to drug  $i$ . It is worthwhile to mention that the denominator part is the normalization term.

**Step 2.** For each known virus, the sequence similarities of the closest  $K$  known viruses are calculated by the KNN method and their corresponding interaction profiles are used to estimate the interaction likelihood profiles:

$$A_v(:, j) = \frac{\sum_{k=1}^K T^{k-1} VV(j, V_k) A^T(V_k, :)}{\sum_{k=1}^K VV(j, V_k)}, \quad (3)$$

where  $j$  denotes the virus index,  $T$  is the decay factor, and in general,  $T \leq 1$ .  $V_k$  denotes the  $k$ -th virus index that is most similar to virus  $j$ . Similarly, the denominator part is the normalization term.

**Step 3.** If  $A_{ij} = 0$ , then we average the interaction likelihood values calculated by Equations (2) and (3) and replace the original values. Using WKNKN, we finally calculate a weighted nearest neighbor interaction spectrum, which we will substitute into the prediction model later.

### 2.2.3. LNS

Previous studies have demonstrated that each data point can be perfectly reconstructed with linear neighborhood information (Wang and Zhang, 2006; Chen et al., 2021b). Based on these studies, we used the known drug–virus interactions to update the degree of drug–virus similarity. Inspired by Zhang et al. (2018b), we established linear neighborhood similarity. In the following, we analyzed the drugs as an example. We take the association matrix of drugs as  $X = \{x_1, x_2, \dots, x_n\}$ , and each vector  $x_i$  is reconstructed from a linear combination of its neighboring data points. The objective function is to minimize the reconstruction loss with the following expression:

$$\min_{w_i} L_i = \left| x_i - \sum_{i_j: x_{i_j} \in N(x_i)} w_{i,i_j} x_{i_j} \right|^2 = \omega_i^T G^i \omega_i \quad (4)$$

$$s.t. \quad \sum_{i_j: x_{i_j} \in N(x_i)} \omega_{i,i_j} x_{i_j} = 1, \omega_i \geq 0, j = 1, \dots, DN,$$

where  $N(x_i)$  denotes the set of  $DN$  nearest neighbors and  $DN (0 < DN < n)$  is a conditioning parameter that indicates the number of neighbors.  $x_{i_j}$  denotes the  $j$ -th neighbor of the vector  $x_i$ .  $w_i = \{w_{i,i_1}, w_{i,i_2}, \dots, w_{i,i_{DN}}\}$  is a vector whose size is  $DN \times 1$  representing the weight size of the  $k$  nearest neighbors of  $x_i$  and also indicates the similarity between  $x_{i_j}$  and  $x_i$ .  $G_i$  denotes the gram matrix whose size is  $DN \times DN$ , where  $G_{i_p, i_q}^i = (x_i - x_{i_p})(x_i - x_{i_q})^T$ . To prevent overfitting, we incorporated the Tikhonov regularization term, which makes the minimization reconstruction loss normalized. The formula is as follows:

$$\min_{w_i} L_i = \omega_i^T G^i \omega_i + \mu \|\omega_i\|_1^2 = \omega_i^T (G^i + \mu I) \omega_i, \quad (5)$$

$$s.t. \quad \sum_{i_j: x_{i_j} \in N(x_i)} \omega_{i,i_j} x_{i_j} = 1, \omega_i \geq 0, j = 1, \dots, DN,$$

where  $\mu$  is the regularization factor. For simplicity, we set  $\mu$  to 1. Finally, we used the standard quadratic programming method to solve the objective function, and the result can be regarded as the reconstruction weight of each data point  $x_i$ . We thus obtained two weight matrices,  $W_d \in R^{n \times n}$  and  $W_v \in R^{m \times m}$ , which were the LNS matrices for the drugs and viruses, respectively.

### 2.2.4. Label propagation

From the previous calculation steps, we finally obtained three matrices: the drug–virus association matrix  $A_{n \times m}$  after WKNKN processing, the drug–drug LNS matrix  $W_d$ , and the virus–virus LNS matrix  $W_v$ . In the following, as a representative example, we considered the drug–drug LNS matrix as a directed weighted graph, with drugs as the nodes and the degree of similarity as the weights of the lines. It is worth noting that the

similarity matrix is not diagonally symmetric, i.e.,  $w_{ij} \neq w_{ji}$ . Based on this, we used a label propagation approach to circularly and iteratively propagate the label information of the drugs to reveal potential drug–virus associations. On the association network, the neighboring edge information of each drug was computed and updated at each label propagation. Meanwhile, we set a probability parameter  $\alpha$  to retain its updated state and retain its initial state with a probability of  $1 - \alpha$ . The specific updated equation is as follows:

$$A_j^{t+1} = \alpha W_d A_j^t + (1 - \alpha) A_j^0 \quad (6)$$

where, for the exact virus  $v_j$ ,  $A_j^0$  denotes all known original drug interaction relationships and  $A_j^t$  denotes the predicted label at iteration  $t$ . For all viruses, we expressed the prediction matrix as  $A^t = \{A_1^t, A_2^t, \dots, A_m^t\}$  and represented the equation further by the following matrix form:

$$A^{t+1} = \alpha W_d A^t + (1 - \alpha) A^0 \quad (7)$$

As  $t$  tends to infinity, the expression converges to the following form:

$$A^* = (1 - \alpha) (I - \alpha W_d)^{-1} A^0 \quad (8)$$

where  $I \in R^{n \times n}$  is the identity matrix and  $A^*$  is the association score matrix. For more details on the convergence analysis of label propagation, please refer to the analysis (Wang and Zhang, 2006).

## 3. Results

### 3.1. Experimental setting

In this study, we used 10 times 10-fold cross-validation to evaluate the performance of our proposed WLLP method. Specifically, 90% of the interaction data was used as the training set, and the remainder was used as the test set. For the evaluation results of the 10 prediction matrices, we averaged them. The true positive rate (TPR or recall), false positive rate (FPR), precision, AUC, and area under the precision-recall curve (AUPR) were used as evaluation metrics. The TPR and FPR indicate the ability of the model to correctly predict positive and negative labels. Precision is the ratio of correctly predicted positive labels to all predicted positive labels, and greater precision indicates better prediction performance. The formulas for the TPR, FPR, and precision are as follows:

$$TPR = \frac{TP}{TP + FN}, \quad (9)$$

$$FPR = \frac{FP}{TN + FP}, \text{ and} \quad (10)$$

$$Precision = \frac{TP}{TP + FP}, \quad (11)$$

where  $TP$  denotes the number of labels correctly predicted as positive,  $TN$  denotes the number of labels correctly predicted as negative,  $FP$  denotes the number of labels incorrectly predicted as positive, and  $FN$  denotes the number of labels incorrectly predicted as negative.

Area under the receiver operating characteristic curve and AUPR are widely used to evaluate the performance of binary classifiers. We constructed the ROC curve and the precision-recall (PR) curve by calculating the TPR, FPR, and precision. The ROC curve is a probability curve with FPR on the x-axis and TPR on the y-axis at various thresholds (Kumar and Indrayan, 2011; Pegoraro et al., 2021; Sun et al., 2022). The AUC is then the area under the ROC curve, which is primarily used to describe the global prediction performance, where larger values indicate better performance (Tang et al., 2022). An AUC of 1 indicates excellent performance and an AUC of 0.5 indicates stochastic performance (Peng et al., 2020). In addition, the PR curve is more effective than the ROC curve for representing highly unbalanced data, thus we also used the AUPR to fully evaluate the performance of the WLLP model. Similar to the AUC, a larger AUPR corresponds to better prediction performance.

### 3.2. Model comparison

In this study, we compared the WLLP model with four other models, namely SCPMF (Meng et al., 2021), NTSIM (Zhang et al., 2018c), TP-NRWRH (Liu et al., 2016), and VDA-RWR (Peng et al., 2021), for the same HDVD dataset. SCPMF is a drug-virus interaction prediction algorithm based on a novel SCPMF. NTSIM is a drug-disease association prediction method that considers only LNS and label propagation. TP-NRWRH uses the bipartite network projection to enhance similarity and propagates it over a heterogeneous network of drugs and diseases with the help of RWR. VDA-RWR applies RWR to the prediction of the newest drug-coronavirus association.

Table 1 shows a comparison of the results obtained from the five prediction models for the HDVD dataset with 10 times 10-fold cross-validation. Figure 2 shows the corresponding ROC and PR curves for the five models. The experimental results demonstrated that the ROC and PR curves of our WLLP model were higher than those of the other four models. It was also apparent that our proposed model offered the best performance in terms of the average AUC and AUPR values. More concretely, the AUC value of WLLP was 0.8828, which was higher than that of the other four approaches (SCPMF: 0.8596; NTSIM: 0.8552; TP-NRWRH: 0.8090; and VDA-RWR: 0.7999). Meanwhile, the AUPR value of WLLP was 0.5277, which was also higher than the other four methods (SCPMF: 0.4958; NTSIM: 0.4778; TP-NRWRH: 0.4929; and VDA-RWR: 0.4781). It was not difficult to find

TABLE 1 Performances of the five prediction methods on the human drug virus database (HDVD) dataset.

| Method   | 10 times<br>10-fold CV<br>AUC | 10 times 10-fold CV AUPR |
|----------|-------------------------------|--------------------------|
| WLLP     | 0.8828 ± 0.0037               | 0.5277 ± 0.0053          |
| SCPMF    | 0.8596 ± 0.0011               | 0.4958 ± 0.0010          |
| NTSIM    | 0.8552 ± 0.0051               | 0.4778 ± 0.0110          |
| TP-NRWRH | 0.8090 ± 0.0079               | 0.4929 ± 0.0175          |
| VDA-RWR  | 0.7999 ± 0.0071               | 0.4781 ± 0.0143          |

that the NTSIM model produced much better results on AUC than the TP-NRWRH and VDA-RWR models, which implied that using LNS was superior to using the original similarity alone, and indicated that using more complex and effective similarity performance provided more important information for association prediction. Due to the effect of the WKNKN pre-training method on the sparsity of the original interaction matrix, the WLLP model produced better prediction results than the NTSIM model, and it also supported the usefulness of the preprocessing procedure (WKNKN) by comparing with the SCPMF model. In summary, the WLLP model exhibited excellent performance.

## 4. Discussion

### 4.1. Ablation experiments

To investigate the plausibility of the WLLP structure, we also tested the model with ablation experiments. We again applied 10 times 10-fold cross-validation to calculate the AUC and AUPR values of the compared models, and the average results were used as the final evaluation indices. The WLLP model comprises three components: WKNKN, LNS, and label propagation (LP). As shown in Table 2, model 1 uses only LNS to set the weights between the nodes on the original label graph and uses label propagation for network diffusion, while model 2 directly applies label propagation to the association network.

The results presented in Table 2 demonstrated that the WLLP model resulted in better AUC and AUPR values for the HDVD dataset than the other two models. Specifically, for model 1, owing to the sparsity of the original drug-virus association matrix, the lack of diffusion channels without preprocessing using the WKNKN algorithm made the nodes with blank labels receive little or no resources during network diffusion, and the propagated information was concentrated on the nodes with high association probability in the global prediction. The introduction of WKNKN alleviated the sparsity of the matrix, and the association prediction of blank labels by WLLP became very simple. Therefore, the

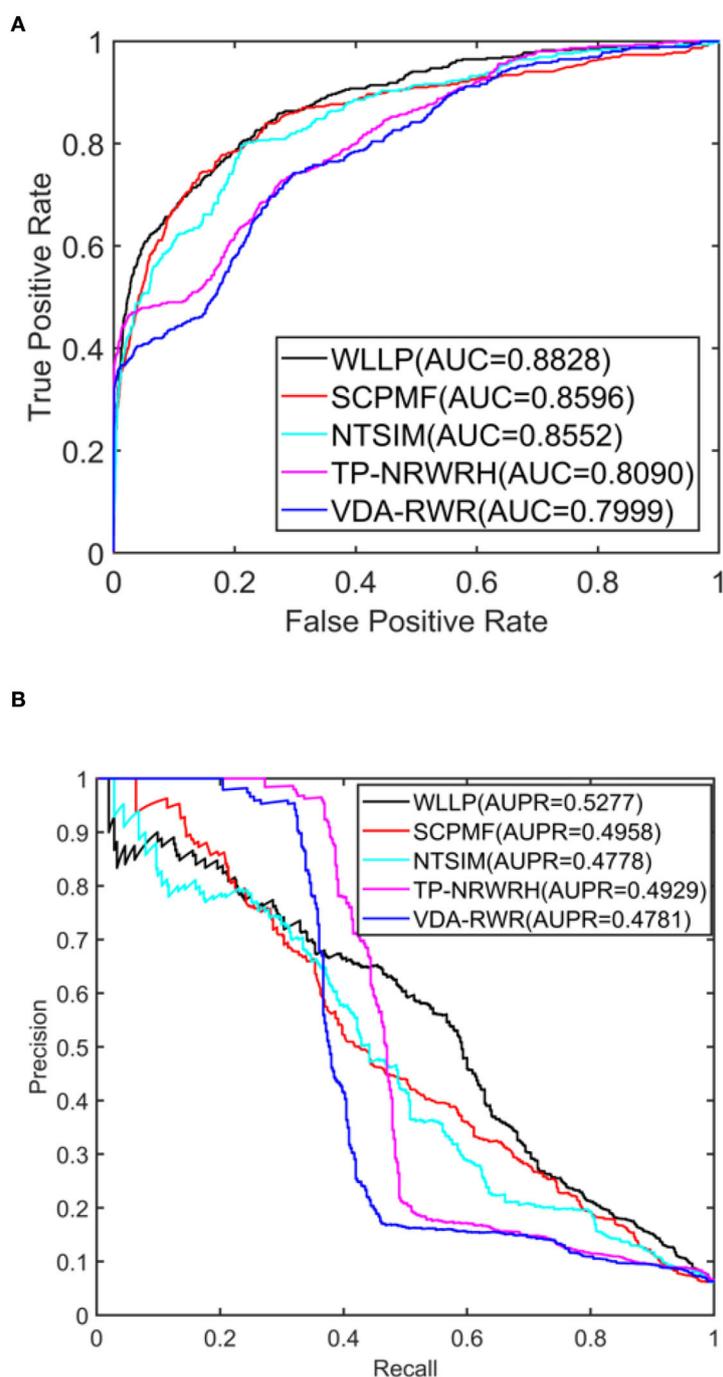


FIGURE 2

Area under the receiver operating characteristic curve (AUC) and area under the precision-recall curve (AUPR) values of the five prediction methods on the human drug virus database (HDVD) dataset. (A) AUC values of the five prediction methods. (B) AUPR values of the five prediction methods.

WKNKN algorithm can be considered an indispensable part of WLLP. Furthermore, a comparison of model 2 and model 1 clearly revealed that the label propagation algorithm in conjunction with LNS took more information into account than using the chemical structure and sequence similarity

alone. The lack of a linear relationship between nodes can make the connections less compact, which in turn leads to poor association prediction for highly unbalanced samples, which is the main reason why the AUPR of model 2 was only 0.1028.

TABLE 2 Results of ablation experiments for the weighted reconstruction-based linear label propagation algorithm (WLLP) model.

| Model | WKNKN | LNS | LP | 10 times 10-fold CV AUC | 10 times 10-fold CV AUPR |
|-------|-------|-----|----|-------------------------|--------------------------|
| WLLP  | ✓     | ✓   | ✓  | 0.8828 ± 0.0037         | 0.5277 ± 0.0053          |
| Model |       | ✓   | ✓  | 0.8552 ± 0.0051         | 0.4778 ± 0.0110          |
| Model |       |     | ✓  | 0.7886 ± 0.0045         | 0.1028 ± 0.0005          |

## 4.2. Parameter settings

We conducted experiments to analyze the effect of parameters on model WLLP. To determine the optimal combination of parameters, we used the grid search method. The WLLP model used seven parameters, namely  $K$ ,  $T$ ,  $DN$ ,  $dN$ ,  $\alpha$ ,  $\beta$ , and  $w$ , where  $K$  and  $T$  are the parameters appearing in the WKNKN algorithm.  $K$  denotes the maximum neighborhood value in the KNN function, while  $T$  denotes the decay factor. The adjustment range of parameter  $K$  is from 1 to 10, while the adjustment range of parameter  $T$  is from 1 to 0.1. We end up with  $K$  set to 8 and  $T$  set to 1 (Figure 3).  $DN$  and  $dN$  correspond to the number of elements in the set of nearest neighbors for the drugs and viruses in the LNS calculation process. The number of drug neighbors  $DN$  should be less than the number of all drugs, and the same is true for the number of virus neighbors  $dN$  based on previous experience (Chen et al., 2021b). We varied the values from 10 to 100, increasing by 10 each time. In Figure 4, for the label propagation algorithm, we used  $\alpha$  and  $\beta$  to represent the retention probability of the update status for drugs and viruses. Thus, we set the different values of  $\alpha$  and  $\beta$  from 0.1 to 1 with step 0.1 (Figure 5). Meanwhile,  $w$  is the label fusion parameter for the final matrix from 0.9 to 0.1 with step 0.1. The effect of the parameter selection of  $w$  is shown in Figure 6, where we observed that good performance is achieved at  $w = 0.4$ . The optimal parameter values for the best model performance were found to be as follows:  $K = 8$ ,  $T = 1$ ,  $DN = 100$ ,  $dN = 6$ ,  $\alpha = 0.2$ ,  $\beta = 0.5$ , and  $w = 0.4$ .

## 4.3. Case study

The overall aim of this work was to identify possible clues for the treatment of COVID-19 after confirming the performance of the WLLP model. Table 3 lists the top 15 drugs predicted from the HDVD dataset, showing the ranking, drug name, DrugBank ID, and literature evidence for each drug. It can be observed that a majority (80%) of the predicted drugs were supported by a variety of literature evidence. Ribavirin was initially recommended for clinical use in China 2019-nCoV Pneumonia Treatment Plan Version 5-Revised (Khalili et al., 2020). It is the eight predicted drug candidate for the potential treatment

of COVID-19. Remdesivir is a nucleotide analog precursor drug with a broad viral spectrum that includes filoviruses, pneumoviruses, parvoviruses, and coronaviruses (Al-Tawfiq et al., 2020; Grein et al., 2020). Remdesivir inhibits viral RNA polymerase and displays *in vitro* activity against COVID-19 (Al-Tawfiq et al., 2020; De Wit et al., 2020; Grein et al., 2020). The combination of remdesivir with emetine may provide better clinical efficacy (Touret and de Lamballerie, 2020). Chloroquine is an inexpensive, safe, and widely administered antimalarial drug that has been used for more than 70 years and is very effective in controlling COVID-19 infection *in vitro* and therefore may be used for the clinical treatment of COVID-19 (Choy et al., 2020). The combination of chloroquine and remdesivir was reported to be very effective in controlling COVID-19 infection *in vitro* (Wang et al., 2020). Based on their combined pathophysiological and pharmacological potential, camostat and nitazoxanide may be recommended for early evaluation and clinical trials against COVID-19 (Khatrri and Mago, 2020). Another study provided preliminary evidence for the use of favipiravir in the treatment of SARS-CoV-2 infection (Cai et al., 2020). Umifenovir is a broad-spectrum antiviral drug. In recent years, clinical trials of umifenovir have been initiated in China (O'Boyle et al., 2011). Sodium lauryl sulfate, an anionic surfactant with protein denaturing ability, effectively inhibits the infectivity of several enveloped viruses through denaturation of the viral envelope. Mouthwash containing sodium lauryl sulfate may be effective in preventing SARS-CoV-2 infection through the oral cavity (Sawa et al., 2021). The 18-kDa cytoplasmic protein procyclin A is an important cellular biomolecule required for RNA virus replication, and recent studies have shown that non-immunosuppressive analogs, such as alisporivir, inhibit the activity of procyclins (Almasi and Mohammadipanah, 2020). Saracatinib, sirolimus, and suramin have also been indicated as therapeutic agents for COVID-19 in recent studies (Romanelli and Mascolo, 2020; Salgado-Benvindo et al., 2020; Tatar et al., 2021).

For hexachlorophene, rifamycin, and tacrolimus, there are no studies proving their activity against COVID-19. However, hexachlorophene is a common detergent additive used for hand washing and disinfection, while rifamycin is an anti-tuberculosis agent that exhibits antiviral properties against various infectious viruses. Tacrolimus, an immunosuppressant, is commonly used in immunotherapy. Although no studies

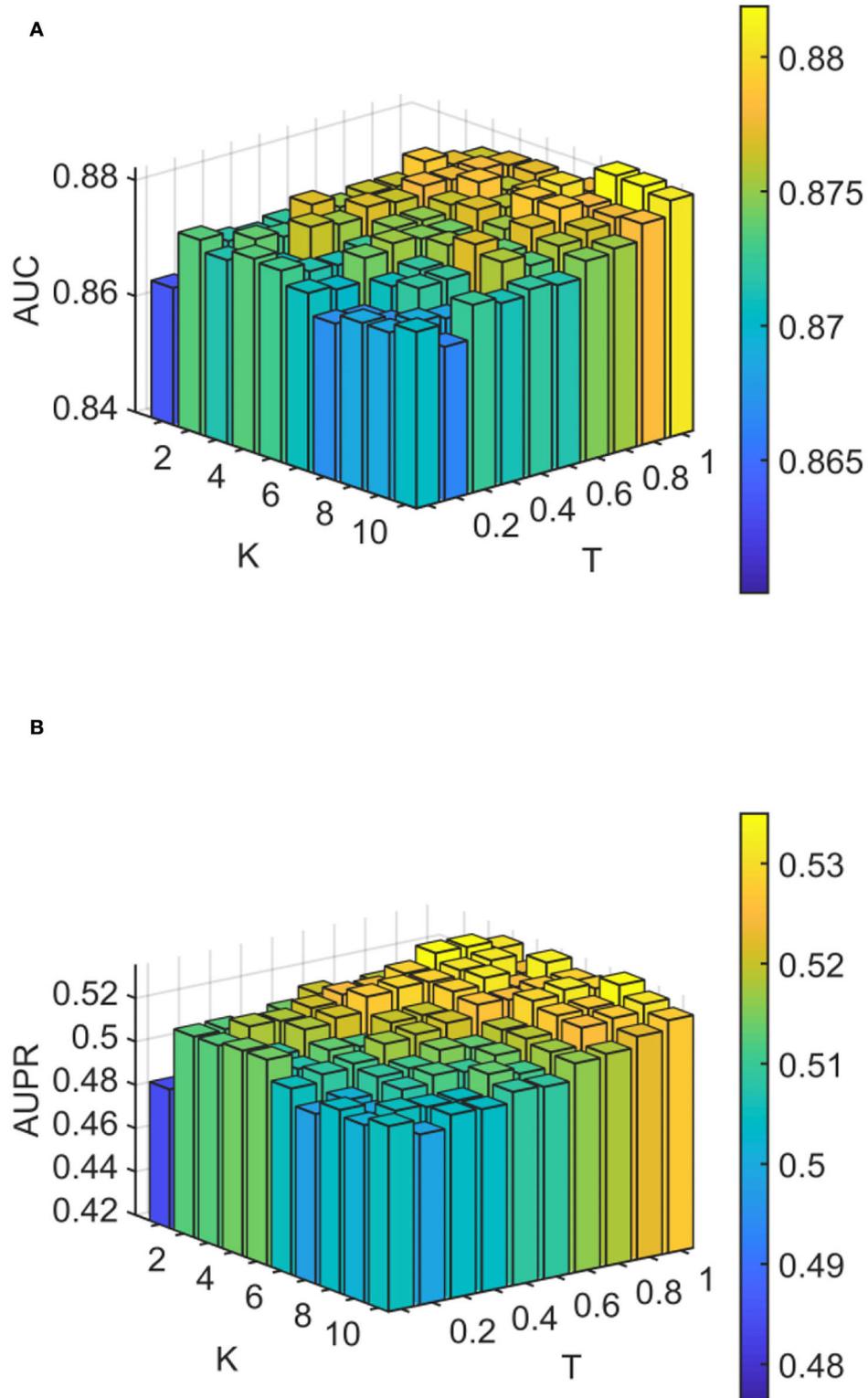
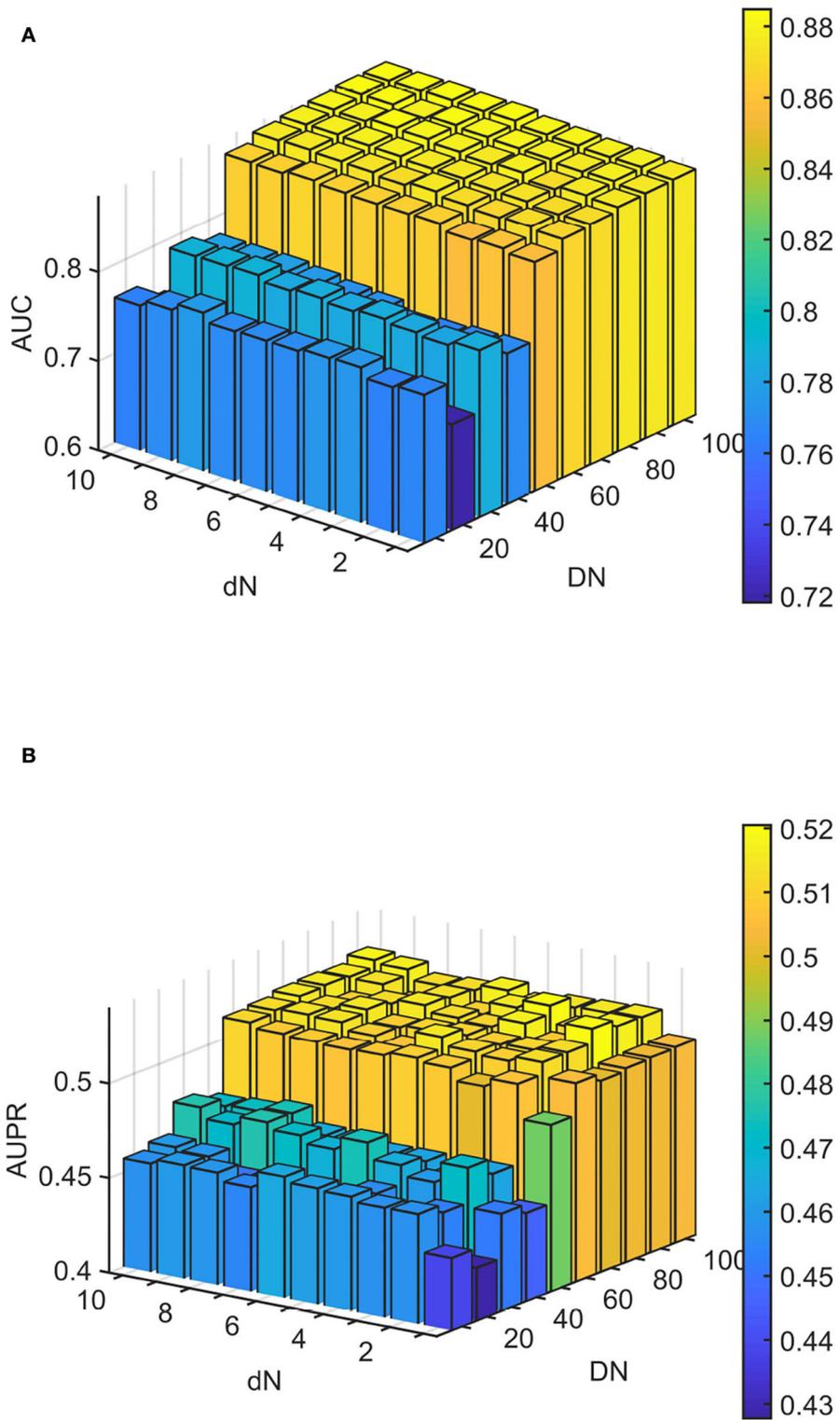


FIGURE 3

Analytical plots of AUC and AUPR for  $K$  and  $T$  in the weighted  $K$ -nearest known neighbors (WKNKN) algorithm. (A) Analytical plots of AUC for  $K$  and  $T$ . (B) Analytical plots of AUPR for  $K$  and  $T$ .



**FIGURE 4**  
 Analytical plots of AUC and AUPR for DN and dN in the linear neighborhood similarity (LNS) algorithm. **(A)** Analytical plots of AUC for *DN* and *dN*. **(B)** Analytical plots of AUPR for *DN* and *dN*.

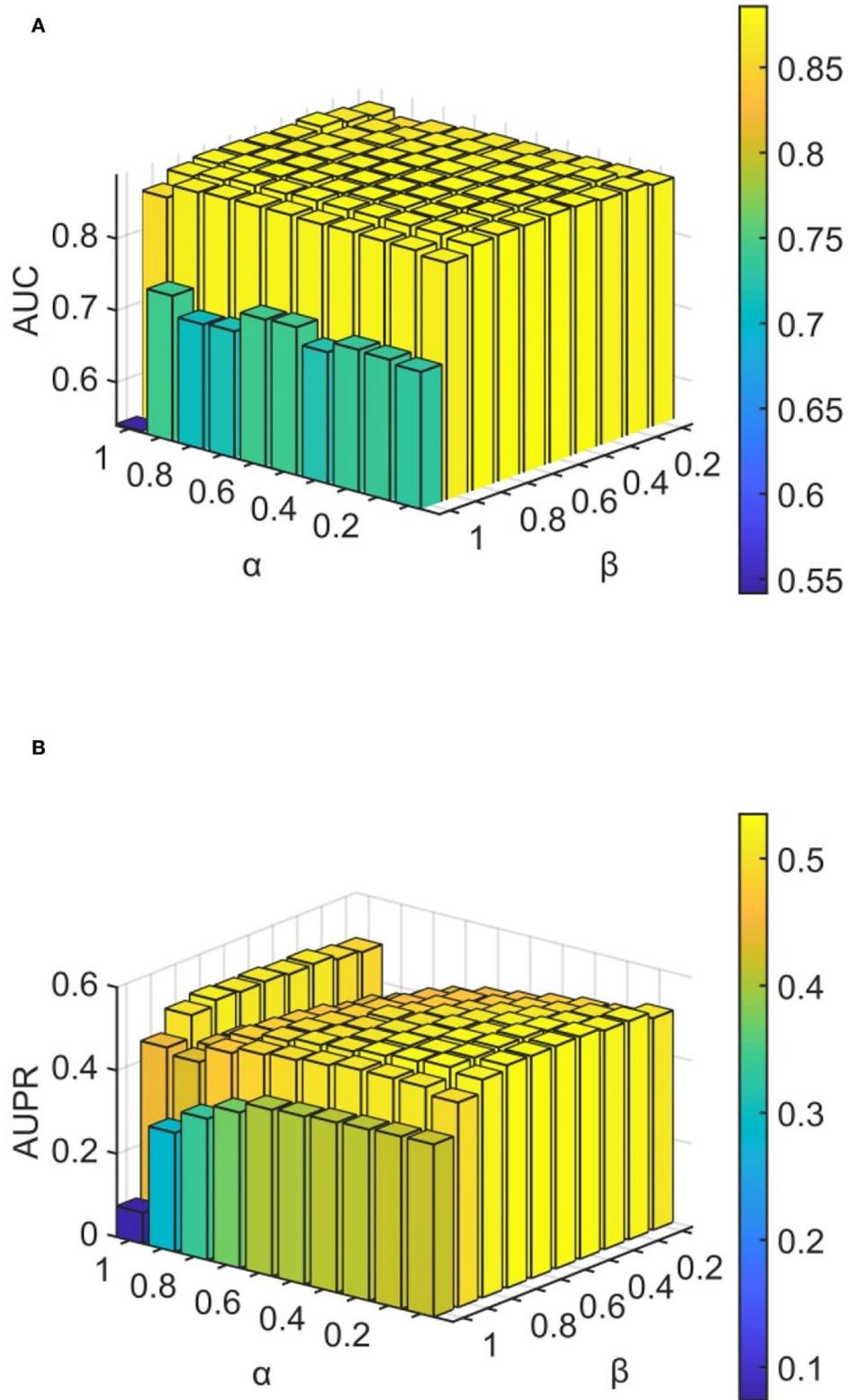
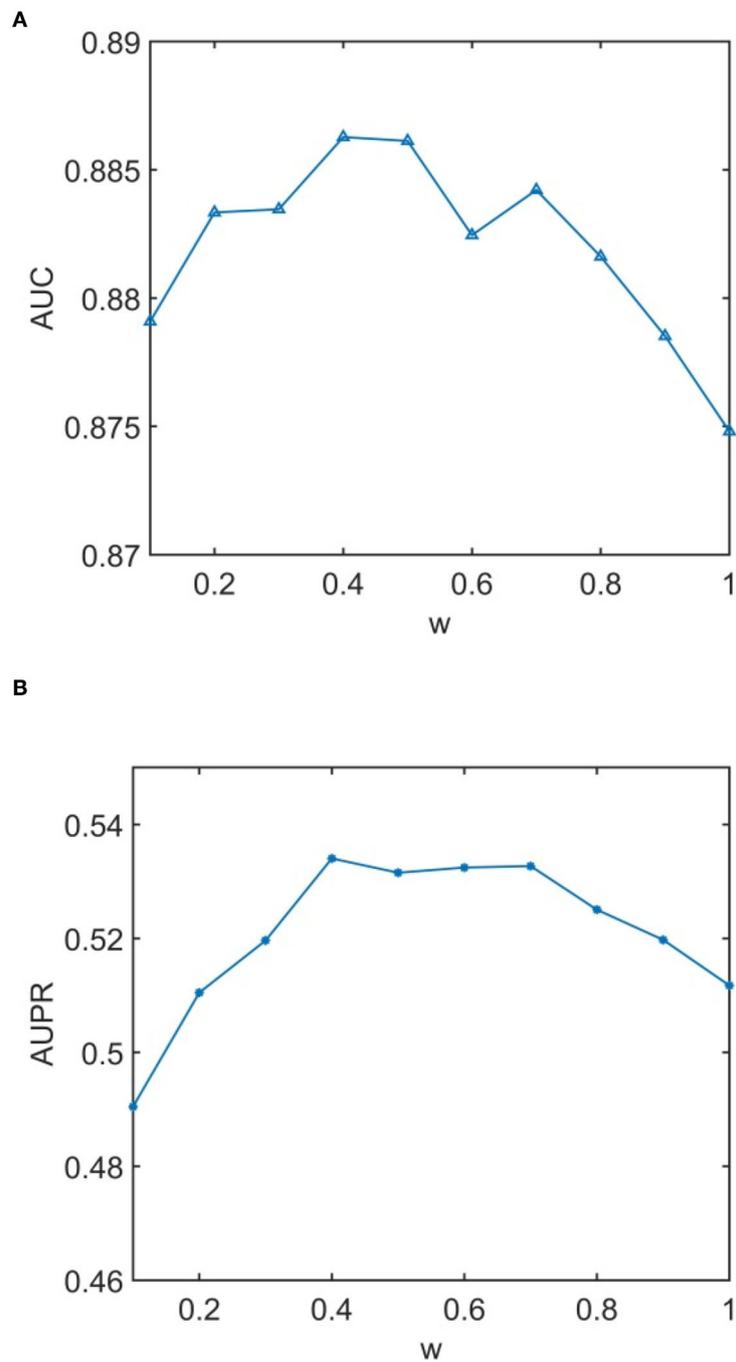


FIGURE 5 Analytical plots of AUC and AUPR for  $\alpha$  and  $\beta$  in the LP algorithm. (A) Analytical plots of AUC for  $\alpha$  and  $\beta$ . (B) Analytical plots of AUPR for  $\alpha$  and  $\beta$ .



**FIGURE 6**  
Analytical plots of AUC and AUPR for  $w$  in the label propagation (LP) algorithm. **(A)** Analytical plots of AUC for  $w$ . **(B)** Analytical plots of AUPR for  $w$ .

have been conducted to demonstrate the efficacy of these three drugs against COVID-19, they still have considerable potential, which remains to be further validated by subsequent work of drug developers.

## 5. Summary

To prevent the spread of SARS-CoV-2, it is critical to deepening our understanding of the association between the

TABLE 3 Top 15 drugs predicted from the HDVD dataset.

| Rank | Drug name (DrugBank ID)         | Evidence                                                                                                           |
|------|---------------------------------|--------------------------------------------------------------------------------------------------------------------|
| 1    | Chloroquine (DB00608)           | <a href="#">Choy et al., 2020</a>                                                                                  |
| 2    | Hexachlorophene (DB00756)       | Unknown                                                                                                            |
| 3    | Nitazoxanide (DB00507)          | <a href="#">Khatri and Mago, 2020</a>                                                                              |
| 4    | Rifamycin (DB11753)             | Unknown                                                                                                            |
| 5    | Remdesivir (DB14761)            | <a href="#">Al-Tawfiq et al., 2020</a> ;<br><a href="#">Grein et al., 2020</a> ; <a href="#">Meng et al., 2021</a> |
| 6    | Oidium lauryl sulfate (DB00815) | <a href="#">Sawa et al., 2021</a>                                                                                  |
| 7    | Camostat (DB13729)              | <a href="#">Zhou et al., 2015</a> ;<br><a href="#">Hoffmann et al., 2020</a>                                       |
| 8    | Ribavirin (DB00811)             | <a href="#">Khalili et al., 2020</a>                                                                               |
| 9    | Saracatinib (DB11805)           | <a href="#">Tatar et al., 2021</a>                                                                                 |
| 10   | Alisporivir (DB12139)           | <a href="#">Almasi and Mohammadipanah, 2020</a>                                                                    |
| 11   | Tacrolimus (DB00864)            | Unknown                                                                                                            |
| 12   | Favipiravir (DB12466)           | <a href="#">Cai et al., 2020</a>                                                                                   |
| 13   | Sirolimus (DB00877)             | <a href="#">Romanelli and Mascolo, 2020</a>                                                                        |
| 14   | Suramin (DB04786)               | <a href="#">Salgado-Benvindo et al., 2020</a>                                                                      |
| 15   | Umifenovir (DB13609)            | <a href="#">McKee et al., 2020</a>                                                                                 |

virus, target proteins, and potential drugs. In the short term, it may be unrealistic to rely on conventional laboratory techniques to develop new drugs against COVID-19, and drug repositioning may represent a more powerful approach. Drug repositioning provides an effective method for prioritizing chemical agents associated with SARS-CoV-2. In this study, a WLLP approach was used to predict the relevance of unknown associations based on drug-virus heterogeneous association networks by combining LNS with LP. The algorithm performs LP on the drug-virus association network, the drug-drug LNS network, and the virus-virus LNS network to diffuse the existing information. With 10 times 10-fold cross-validation, our model achieved an AUC of 0.8828 and an AUPR of 0.5277, both of which were higher than the other methods used for comparison. Furthermore, the information and feasibility of the first 15 drugs were determined by a case study of SARS-CoV-2. Even so, our model still has room for improvement. The predictive performance of the proposed method is limited owing to the current scarcity of data. In the future, we will attempt to tap into drug library and pharmacological resources, and with the addition and integration of more data from recent studies, the prediction results of our model should be improved.

## Data availability statement

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

## Author contributions

LC and DL: conceptualization, investigation, and project administration. LC: data curation and resources. LC, DL, HX, JL, and LL: formal analysis, funding acquisition, and writing—review and editing. LC, DL, and HX: methodology. HX, JL, and LL: supervision. HX and JL: validation and writing draft. JL: visualization. All authors contributed to the article and approved the submitted version.

## 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), and the Science and Technology Plan Project of Guangzhou City (202102021236).

## Acknowledgments

We thank reviewers for their 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

- Almasi, F., and Mohammadipanah, F. (2020). Potential targets and plausible drugs of coronavirus infection caused by 2019-ncov. *Authorea [Preprints]*. doi: 10.22541/au.158766083.33108969
- Al-Tawfiq, J. A., Al-Homoud, A. H., and Memish, Z. A. (2020). Remdesivir as a possible therapeutic option for the COVID-19. *Travel. Med. Infect. Dis.* 34, 101615. doi: 10.1016/j.tmaid.2020.101615
- Bajusz, D., Rácz, A., and Héberger, K. (2015). Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* 7, 1–13. doi: 10.1186/s13321-015-0069-3
- Barlow, A., Landolf, K. M., Barlow, B., Yeung, S. Y. A., Heavner, J. J., Claassen, C. W., et al. (2020). Review of emerging pharmacotherapy for the treatment of coronavirus disease 2019. *Pharmacotherapy* 40, 416–437. doi: 10.1002/phar.2398
- Cai, Q., Yang, M., Liu, D., Chen, J., Shu, D., Xia, J., et al. (2020). Experimental treatment with favipiravir for COVID-19: an open-label control study. *Engineering* 6, 1192–1198. doi: 10.1016/j.eng.2020.03.007
- Chen, H., Zhang, H., Zhang, Z., Cao, Y., and Tang, W. (2015). Network-based inference methods for drug repositioning. *Comput. Math. Methods Med.* 2015, 130620. doi: 10.1155/2015/130620
- Chen, X., and Huang, L. (2017). Lrsslmda: Laplacian regularized sparse subspace learning for mirna-disease association prediction. *PLoS Comput Biol.* 13, e1005912. doi: 10.1371/journal.pcbi.1005912
- Chen, X., Li, T.-H., Zhao, Y., Wang, C.-C., and Zhu, C.-C. (2021a). Deep-belief network for predicting potential mirna-disease associations. *Brief. Bioinform.* 22, bbaa186. doi: 10.1093/bib/bbaa186
- Chen, X., Sun, L.-G., and Zhao, Y. (2021b). Ncmcmda: mirna-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* 22, 485–496. doi: 10.1093/bib/bz159
- Chen, X., Wang, L., Qu, J., Guan, N.-N., and Li, J.-Q. (2018a). Predicting mirna-disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265. doi: 10.1093/bioinformatics/bty503
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z.-H., and Liu, H. (2018b). Bnpmda: bipartite network projection for mirna-disease association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333
- Chen, X., Xie, D., Zhao, Q., and You, Z.-H. (2019a). Micrnas and complex diseases: from experimental results to computational models. *Brief. Bioinform.* 20, 515–539. doi: 10.1093/bib/bbx130
- Chen, X., Yin, J., Qu, J., and Huang, L. (2018c). Mdhgi: matrix decomposition and heterogeneous graph inference for mirna-disease association prediction. *PLoS Comput. Biol.* 14, e1006418. doi: 10.1371/journal.pcbi.1006418
- Chen, X., Zhu, C.-C., and Yin, J. (2019b). Ensemble of decision tree reveals potential mirna-disease associations. *PLoS Comput. Biol.* 15, e1007209. doi: 10.1371/journal.pcbi.1007209
- Choy, K.-T., Wong, A. Y.-L., Kaewpreedee, P., Sia, S. F., Chen, D., Hui, K. P. Y., et al. (2020). Remdesivir, lopinavir, emetine, and homoharringtonine inhibit sars-cov-2 replication *in vitro*. *Antiviral Res.* 178, 104786. doi: 10.1016/j.antiviral.2020.104786
- De Wit, E., Feldmann, F., Cronin, J., Jordan, R., Okumura, A., Thomas, T., et al. (2020). Prophylactic and therapeutic remdesivir (gs-5734) treatment in the rhesus macaque model of mers-cov infection. *Proc. Natl. Acad. Sci. U.S.A.* 117, 6771–6776. doi: 10.1073/pnas.1922083117
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. doi: 10.48550/arXiv.1810.04805
- Dotolo, S., Marabotti, A., Facchiano, A., and Tagliaferri, R. (2021). A review on drug repurposing applicable to COVID-19. *Brief. Bioinform.* 22, 726–741. doi: 10.1093/bib/bbaa288
- Ezzat, A., Zhao, P., Wu, M., Li, X.-L., and Kwok, C.-K. (2016). Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 14, 646–656. doi: 10.1109/TCBB.2016.2530062
- Fan, H.-H., Wang, L.-Q., Liu, W.-L., An, X.-P., Liu, Z.-D., He, X.-Q., et al. (2020). Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coronavirus-related coronavirus model. *Chin. Med. J.* 133, 1051–1056. doi: 10.1097/CM9.0000000000000797
- Gönen, M. (2012). Predicting drug-target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics* 28, 2304–2310. doi: 10.1093/bioinformatics/bts360
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., et al. (2020). Compassionate use of remdesivir for patients with severe COVID-19. *N. Engl. J. Med.* 382, 2327–2336. doi: 10.1056/NEJMoa2007016
- Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi: 10.1016/j.cell.2020.02.052
- Hu, H., Zhang, L., Ai, H., Zhang, H., Fan, Y., Zhao, Q., et al. (2018). Hlpi-ensemble: prediction of human lncrna-protein interactions based on ensemble strategy. *RNA Biol.* 15, 797–806. doi: 10.1080/15476286.2018.1457935
- Katoh, K., and Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., et al. (2020). Artificial intelligence for COVID-19 drug discovery and vaccine development. *Front. Artif. Intell.* 3, 65. doi: 10.3389/frai.2020.00065
- Khalili, J. S., Zhu, H., Mak, N. S. A., Yan, Y., and Zhu, Y. (2020). Novel coronavirus treatment with ribavirin: groundwork for an evaluation concerning COVID-19. *J. Med. Virol.* 92, 740–746. doi: 10.1002/jmv.25798
- Khatri, M., and Mago, P. (2020). Nitazoxanide/camostat combination for COVID-19: an unexplored potential therapy. *Chem. Biol. Lett.* 7, 192–196. Available online at: <http://pubs.science.in/journal/index.php/cbl/article/view/1085>
- King, A. D., Pržulj, N., and Jurisica, I. (2012). Protein complex prediction with RNSC. *Methods Mol. Biol.* 804, 297–312. doi: 10.1007/978-1-61779-361-5\_16
- Kumar, R., and Indrayan, A. (2011). Receiver operating characteristic (roc) curve for medical researchers. *Indian Pediatr.* 48, 277–287. doi: 10.1007/s13312-011-0055-4
- Liu, H., Ren, G., Chen, H., Liu, Q., Yang, Y., and Zhao, Q. (2020). Predicting lncrna-mirna interactions based on logistic matrix factorization with neighborhood regularized. *Knowl. Based Syst.* 191, 105261. doi: 10.1016/j.knsys.2019.105261
- Liu, H., Song, Y., Guan, J., Luo, L., and Zhuang, Z. (2016). Inferring new indications for approved drugs via random walk on drug-disease heterogeneous networks. *BMC Bioinform.* 17, 269–277. doi: 10.1186/s12859-016-1336-7
- Liu, S., Zheng, Q., and Wang, Z. (2020). Potential covalent drugs targeting the main protease of the sars-cov-2 coronavirus. *Bioinformatics* 36, 3295–3298. doi: 10.1093/bioinformatics/btaa224
- Liu, W., Jiang, Y., Peng, L., Sun, X., Gan, W., Zhao, Q., et al. (2022a). Inferring gene regulatory networks using the improved markov blanket discovery algorithm. *Interdisc. Sci.* 14, 168–181. doi: 10.1007/s12539-021-00478-9
- Liu, W., Lin, H., Huang, L., Peng, L., Tang, T., Zhao, Q., et al. (2022b). Identification of mirna-disease associations via deep forest ensemble learning based on autoencoder. *Brief. Bioinform.* 23, bbac104. doi: 10.1093/bib/bbac104
- Luo, H., Wang, J., Li, M., Luo, J., Peng, X., Wu, F.-X., et al. (2016). Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics* 32, 2664–2671. doi: 10.1093/bioinformatics/btw228
- Macropol, K., Can, T., and Singh, A. K. (2009). Rrw: repeated random walks on genome-scale protein networks for local cluster discovery. *BMC Bioinformatics* 10, 1–10. doi: 10.1186/1471-2105-10-283
- Martinez, V., Navarro, C., Cano, C., Fajardo, W., and Blanco, A. (2015). Drugnet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif. Intell. Med.* 63, 41–49. doi: 10.1016/j.artmed.2014.11.003
- McKee, D. L., Sternberg, A., Stange, U., Lauffer, S., and Naujokat, C. (2020). Candidate drugs against sars-cov-2 and covid-19. *Pharmacol. Res.* 157, 104859. doi: 10.1016/j.phrs.2020.104859
- Meng, Y., Jin, M., Tang, X., and Xu, J. (2021). Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. *Appl. Soft Comput.* 103, 107135. doi: 10.1016/j.asoc.2021.107135
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel: An open chemical toolbox. *J. Cheminform.* 3, 1–14. doi: 10.1186/1758-2946-3-33
- Organization, W. H. (2020). *Global surveillance for covid-19 disease caused by human infection with novel coronavirus (covid-19):*

*interim guidance*, 27 february 2020. Technical report, World Health Organization.

Öztürk, H., Ozkirimli, E., and Özgür, A. (2016). A comparative study of smiles-based compound similarity functions for drug-target interaction prediction. *BMC Bioinform.* 17, 1–11. doi: 10.1186/s12859-016-0977-x

Pegoraro, J. A., Lavault, S., Wattiez, N., Similowski, T., Gonzalez-Bermejo, J., and Birmelé, E. (2021). Machine-learning based feature selection for a non-invasive breathing change detection. *BioData Min.* 14, 1–16. doi: 10.1186/s13040-021-00265-8

Peng, L., Shen, L., Xu, J., Tian, X., Liu, F., Wang, J., et al. (2021). Prioritizing antiviral drugs against sars-cov-2 by integrating viral complete genome sequences and drug chemical structures. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-83737-5

Peng, L.-H., Zhou, L.-Q., Chen, X., and Piao, X. (2020). A computational study of potential miRNA-disease association inference based on ensemble learning and kernel ridge regression. *Front. Bioeng. Biotechnol.* 8, 40. doi: 10.3389/fbioe.2020.00040

Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47, 228–235. doi: 10.1002/prot.10082

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58. doi: 10.1038/nrd.2018.168

Re, M., and Valentini, G. (2013). Network-based drug ranking and repositioning with respect to drugbank therapeutic categories. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10, 1359–1371. doi: 10.1109/TCBB.2013.62

Romanelli, A., and Mascolo, S. (2020). Sirolimus to treat sars-cov-2 infection: an old drug for a new disease. *Respir. Med.* 8, 420–422. doi: 10.34172/jrcm.2020.044

Salgado-Benvindo, C., Thaler, M., Tas, A., Ogando, N. S., Bredenbeek, P. J., Ninaber, D. K., et al. (2020). Suramin inhibits sars-cov-2 infection in cell culture by interfering with early steps of the replication cycle. *Antimicrob. Agents Chemother.* 64, e00900-e00920. doi: 10.1128/AAC.00900-20

Sawa, Y., Ibaragi, S., Okui, T., Yamashita, J., Ikebe, T., and Harada, H. (2021). Expression of sars-cov-2 entry factors in human oral tissue. *J. Anat.* 238, 1341–1354. doi: 10.1111/joa.13391

Shin, B., Park, S., Kang, K., and Ho, J. C. (2019). “Self-attention based molecule representation for predicting drug-target interaction,” in *Machine Learning for Healthcare Conference* (PMLR), 230–248. Available online at: <http://proceedings.mlr.press/v106/shin19a.html?ref=https://githubhelp.com>

Sirin, U., Erdogdu, U., Polat, F., Tan, M., and Alhaji, R. (2016). Effective gene expression data generation framework based on multi-model approach. *Artif. Intell. Med.* 70, 41–61. doi: 10.1016/j.artmed.2016.05.003

Sohrabi, C., Alsafi, Z., O’neill, N., Khan, M., Kerwan, A., Al-Jabir, A., et al. (2020). World health organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int. J. Surgery* 76, 71–76. doi: 10.1016/j.ijssu.2020.02.034

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

Tang, Q., Nie, F., Zhao, Q., and Chen, W. (2022). A merged molecular representation deep learning method for blood-brain barrier permeability prediction. *Brief. Bioinform.* 23, bbac357. doi: 10.1093/bib/bbac357

Tatar, G., Ozyurt, E., and Turhan, K. (2021). Computational drug repurposing study of the rna binding domain of sars-cov-2 nucleocapsid protein with antiviral agents. *Biotechnol. Prog.* 37, e3110. doi: 10.1002/btpr.3110

Torng, W., and Altman, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* 59, 4131–4149. doi: 10.1021/acs.jcim.9b00628

Touret, F., and de Lamballerie, X. (2020). Of chloroquine and COVID-19. *Antiviral Res.* 177, 104762. doi: 10.1016/j.antiviral.2020.104762

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in Neural Information Processing Systems*, Vol. 30. Available online at: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>

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

Wang, F., and Zhang, C. (2006). “Label propagation through linear neighborhoods,” in *Proceedings of the 23rd International Conference on Machine Learning*, 985–992. doi: 10.1145/1143844

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-ncov) in vitro. *Cell Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0

Wheeler, D. L., Church, D. M., Lash, A. E., Leipe, D. D., Madden, T. L., Pontius, J. U., et al. (2002). Database resources of the national center for biotechnology information: 2002 update. *Nucleic Acids Res.* 30, 13–16. doi: 10.1093/nar/30.1.13

Xue, H., Li, J., Xie, H., and Wang, Y. (2018). Review of drug repositioning approaches and resources. *Int. J. Biol. Sci.* 14, 1232. doi: 10.7150/ijbs.24612

Yang, M., Luo, H., Li, Y., and Wang, J. (2019). Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 35, i455–i463. doi: 10.1093/bioinformatics/btz331

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

Zhang, L., Yang, P., Feng, H., Zhao, Q., and Liu, H. (2021b). Using network distance analysis to predict lncrna-mirna interactions. *Interdisc. Sci.* 13, 535–545. doi: 10.1007/s12539-021-00458-z

Zhang, W., Chen, Y., and Li, D. (2017a). Drug-target interaction prediction through label propagation with linear neighborhood information. *Molecules* 22, 2056. doi: 10.3390/molecules22122056

Zhang, W., Chen, Y., Li, D., and Yue, X. (2018a). Manifold regularized matrix factorization for drug-drug interaction prediction. *J. Biomed. Inform.* 88, 90–97. doi: 10.1016/j.jbi.2018.11.005

Zhang, W., Qu, Q., Zhang, Y., and Wang, W. (2018b). The linear neighborhood propagation method for predicting long non-coding rna-protein interactions. *Neurocomputing* 273, 526–534. doi: 10.1016/j.neucom.2017.07.065

Zhang, W., Yue, X., Huang, F., Liu, R., Chen, Y., and Ruan, C. (2018c). Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods* 145, 51–59. doi: 10.1016/j.jymeth.2018.06.001

Zhang, W., Yue, X., Liu, F., Chen, Y., Tu, S., and Zhang, X. (2017b). A unified frame of predicting side effects of drugs by using linear neighborhood similarity. *BMC Syst. Biol.* 11, 23–34. doi: 10.1186/s12918-017-0477-2

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell Discov.* 6, 1–18. doi: 10.1038/s41421-020-0153-3

Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion Jr, R., Nunneley, J. W., et al. (2015). Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Res.* 116, 76–84. doi: 10.1016/j.antiviral.2015.01.011

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 82, 727–733. doi: 10.1056/NEJMoa2001017