



# MILNP: Plant lncRNA–miRNA Interaction Prediction Based on Improved Linear Neighborhood Similarity and Label Propagation

Lijun Cai, Mingyu Gao, Xuanbai Ren, Xiangzheng Fu\*, Junlin Xu, Peng Wang\* and Yifan Chen

College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

## OPEN ACCESS

### Edited by:

Quan Zou,  
University of Electronic Science  
and Technology of China, China

### Reviewed by:

Wen Zhang,  
Huazhong Agricultural University,  
China  
Yijie Ding,  
University of Electronic Science  
and Technology of China, China

Chu Pan,  
University Health Network (UHN),  
Canada

### \*Correspondence:

Xiangzheng Fu  
fxz326@hnu.edu.cn  
Peng Wang  
wangpenglw@126.com

### Specialty section:

This article was submitted to  
Sustainable and Intelligent  
Phytoprotection,  
a section of the journal  
Frontiers in Plant Science

**Received:** 25 January 2022

**Accepted:** 21 February 2022

**Published:** 25 March 2022

### Citation:

Cai L, Gao M, Ren X, Fu X, Xu J,  
Wang P and Chen Y (2022) MILNP:  
Plant lncRNA–miRNA Interaction  
Prediction Based on Improved Linear  
Neighborhood Similarity and Label  
Propagation.  
*Front. Plant Sci.* 13:861886.  
doi: 10.3389/fpls.2022.861886

Knowledge of the interactions between long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) is the basis of understanding various biological activities and designing new drugs. Previous computational methods for predicting lncRNA–miRNA interactions lacked for plants, and they suffer from various limitations that affect the prediction accuracy and their applicability. Research on plant lncRNA–miRNA interactions is still in its infancy. In this paper, we propose an accurate predictor, MILNP, for predicting plant lncRNA–miRNA interactions based on improved linear neighborhood similarity measurement and linear neighborhood propagation algorithm. Specifically, we propose a novel similarity measure based on linear neighborhood similarity from multiple similarity profiles of lncRNAs and miRNAs and derive more precise neighborhood ranges so as to escape the limits of the existing methods. We then simultaneously update the lncRNA–miRNA interactions predicted from both similarity matrices based on label propagation. We comprehensively evaluate MILNP on the latest plant lncRNA–miRNA interaction benchmark datasets. The results demonstrate the superior performance of MILNP than the most up-to-date methods. What's more, MILNP can be leveraged for isolated plant lncRNAs (or miRNAs). Case studies suggest that MILNP can identify novel plant lncRNA–miRNA interactions, which are confirmed by classical tools. The implementation is available on <https://github.com/HerSwain/gra/tree/MILNP>.

**Keywords:** plant lncRNA–miRNA interaction, theoretical derivation, multilevel similarity, linear reconstruction, label propagation

## INTRODUCTION

An increasing number of studies have shown that non-coding RNAs (ncRNAs), especially long non-coding RNAs (lncRNAs) and microRNAs (miRNA), act in various biological processes (Amin et al., 2019). miRNAs with a sequence length of approximately 22 nucleotides control post-transcriptional gene expression (DeVeale et al., 2021). lncRNAs, usually with a sequence length greater than 200 nucleotides, are widely engaged in essential regulatory processes (Ard et al., 2014; Chen et al., 2017; Fang et al., 2020; Statello et al., 2020; Goodall and Wickramasinghe, 2021). lncRNAs control the expression of miRNAs to influence the expression of their target genes: lncRNAs compete with mRNA for miRNAs, thereby regulating miRNA-mediated target inhibition (Geisler and Collier, 2013). For example, in the lumbar intervertebral disk degeneration (Zhu et al., 2019), lncRNAs may act as competing endogenous RNA (ceRNAs) that bind competitively to miRNAs through their miRNA response elements,

thereby regulating the expression of miRNA-targeted mRNAs. miRNAs and lncRNAs interact with each other to exert higher levels of post-transcriptional regulation.

As computer technology advances rapidly, numerous methods are employed to study miRNAs, lncRNAs, and proteins, as well as their interactions (Fu et al., 2019, 2020; Cai et al., 2020a,b, 2021; Dai et al., 2021; Li P. et al., 2021; Liu et al., 2021; Rahaman et al., 2021; Song et al., 2021; Tan et al., 2021; Zhang C. L. et al., 2021; Zhang et al., 2022). With regard to miRNAs, a miRNA that is positively selected during human evolution is identified to regulate energy expenditure, and the relevance of this positively selected locus to metabolic disorders may explain the link between this locus and metabolic diseases (Stower, 2020). With regard to lncRNAs, a lncRNA GCMA activated by SP1 acts as a competing endogenous RNA in gastric cancer *via* competition for miR-124 and miR-34a to promote tumor metastasis (Tian et al., 2020). With regard to interactions between lncRNAs and proteins, lncRNA DIGIT regulates endoderm differentiation by promoting the formation of phase-separated condensates of bromodomain and the extraterminal domain protein BRD3 (Daneshvar et al., 2020). With regard to interactions between lncRNAs and miRNAs, the targeting lnc–MGC inhibits host lnc–MGC expression while suppressing the expression of key cluster miRNAs in the kidneys and preventing early diabetic nephropathy (Allison, 2016). Studies such as these are abundant and have made important contributions.

Although predictions about lncRNA–miRNA interactions exist, most of them are not about plants (Jiang et al., 2018, 2019; Ayachit et al., 2020; Banerjee et al., 2020; Ma et al., 2020; Qazi et al., 2020; Shen et al., 2020; Aglawe et al., 2021). The confirmed plant lncRNA–miRNA interactions are very limited and have been barely covered. For instance, the NPInter4.0 (Teng et al., 2020) documents extensive functional interactions between ncRNAs and molecules of over 30 species, yet only two of them are plants. From 71 RNA–RNA interactions for the two plants, only one of them is a miRNA–lncRNA interaction. It is no secret that the mechanisms of plant miRNA–lncRNA interactions remain elusive. Also, lncRNAs are characterized by low sequence conservation, especially among distantly related species.: The lncRNA molecules of different species or the same species may vary in terms of amino acid and nucleotide fragments during biological evolution, which entails that predictions obtained from animal studies are not guaranteed to be applicable in plants (Noviello et al., 2018). As a result, conclusions about the mechanism of plant lncRNA–miRNA interactions cannot be completely copied from animals and must be explored.

Studies on lncRNA–miRNA interactions generally fall under two categories, namely, bioinformatics-based machine learning methods and similarity network-based methods (Liu et al., 2017, 2020; Peng et al., 2017; Zeng et al., 2017, 2018, 2019; Zhao et al., 2020; Chen et al., 2021; Singh et al., 2021; Wang et al., 2021; Zhang Y. et al., 2021; Zhou et al., 2021; Zhu et al., 2021). The former extracts biological features and trains models to obtain dichotomous results (i.e., the output is whether lncRNA and miRNA interact) (Intell, 2019; Li J. et al., 2021). By comparison, the latter computes single or multiple correlation similarity matrices to obtain the final predictions (Wang et al., 2014). The

works that use machine learning, even deep learning methods, do succeed. However, machine learning is flawed in terms of two aspects (Peng et al., 2018; Zhang et al., 2019b). First, it relies on data features. For certain lncRNAs or miRNAs, they may not have expression profiles or target genes. In this situation, machine-learning methods are not applicable. Moreover, for some isolated lncRNAs or miRNAs that do not have any interactions with miRNAs or lncRNAs at all, they have difficulty forecasting any unknown interaction. By contrast, similarity network-based approaches can address such imperfections. Constructing similarity networks does not necessarily depend on specific data features (Zhang et al., 2019b), and it is able to predict isolated lncRNAs and miRNAs solely on the basis of sequence information. Linear neighborhood similarity, which refers to selecting the most appropriate neighborhoods for linear reconstruction, as a new similarity measurement perspective, is currently gaining momentum in bioinformatics (Li et al., 2018; Zhang et al., 2018a, 2019a; Xie et al., 2020; Zhang W. et al., 2021; Jia and Luan, 2022; Zhu et al., 2022), such as LPLNP (Zhang et al., 2018a) and MPLPLNP (Jia and Luan, 2022), in predicting lncRNA–protein interactions, and FLNSNLI (Zhang W. et al., 2021) and LPLNS (Li et al., 2018) in predicting miRNA–disease associations. To the best of our knowledge, no similarity network-based method is available to date for predicting lncRNA–miRNA interactions in plants. Owing to the imperfections of machine-learning methods and the necessity to independently detect plant lncRNA–miRNA interactions, novel and effective methods must be constructed.

In this study, we hypothesize that lncRNA–miRNA interactions with highly similar lncRNAs will have similar interaction or non-interaction patterns with miRNAs. Under this assumption, a multi-source information-based linear neighborhood propagation method (MILNP) is proposed. The similarity is calculated through our improved linear neighborhood similarity (ILNS) algorithm, where ILNS has the advantage of obtaining a more accurate neighborhood range over the pre-improvement. First, multidimensional features are separately extracted from the sequences of lncRNAs and miRNAs to calculate sequence similarity, whereas interaction profile similarity is obtained using their interactions. These two similarities are then combined to obtain integrated similarity. Label propagation based on the integrated similarity is used to calculate individually the prediction matrix of lncRNAs and the prediction matrix of miRNAs. Finally, the two prediction matrices are summed by taking different weights to obtain the final prediction. The contribution consists of the following components.

- We proposed a novel similarity measurement, ILNS, for calculating multiple similarity profiles.
- We constructed MILNP based on ILNS to predict lncRNA–miRNA interactions and discovered new interactions in the plant.
- High-accuracy prediction results in multiple experiments and showed superiority over the existing methods and reliability for finding new interactions of MILNP.

## DATASETS AND METHODS

### Dataset Construction

The original data used herein are derived from a previous study (Kang et al., 2020) that investigated plant lncRNA-miRNA interactions. miRNA sequences are downloaded from miRBase22.1 (Kozomara et al., 2019), whereas lncRNA sequences are downloaded from GreeNC1.12 (Paytuyi-Gallart et al., 2019) and CANTATAdb2.0 (Szcześniak et al., 2019). Datasets of lncRNA-miRNA interaction from *Arabidopsis thaliana*, *Glycine max*, and *Medicago truncatula* are chosen with 2,500 positive samples from the positive dataset of each species, for a total of 7,500 positive samples. Similarly, 2,500 negative samples from the negative dataset of each species are chosen, also for a total of 7,500 negative samples. These positive and negative samples are intermixed as the training-validation set to avoid imbalance in sample distribution.

The dataset consists of five parts per sample: the symbol, the miRNA name, the lncRNA name, the sequence yielded from combining the miRNA sequence with the lncRNA sequence, and the sample label (0 for the absence of interaction, and 1 for the presence of an interaction). However, such a format is inappropriate for the method we applied herein. The processing is as follows. First, the original sequence binding files are separated by name, sequenced, and labeled to obtain the name of miRNA, name of lncRNA, binding sequences of miRNA and lncRNA, and labels. Second, all miRNA sequences are found according to the miRNA name order by checking against the

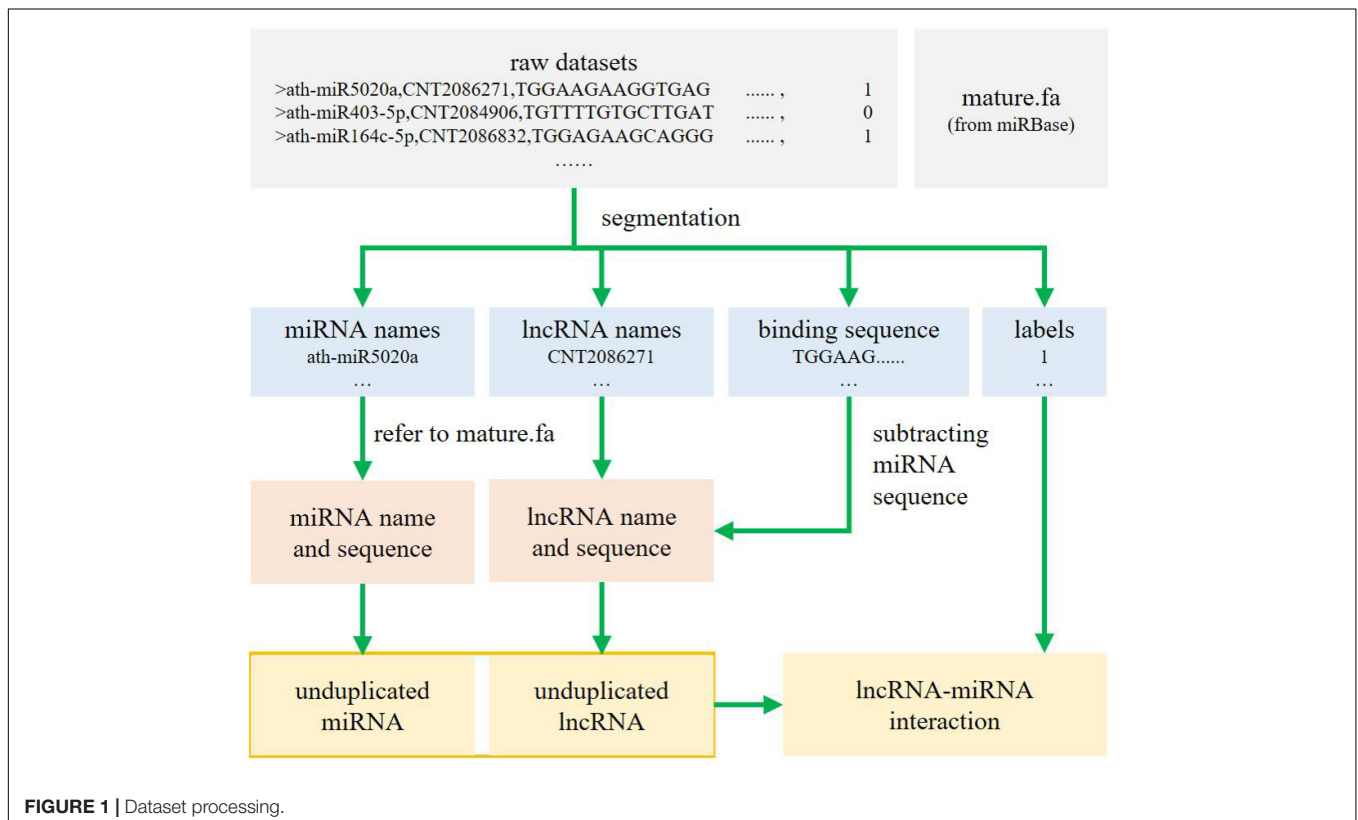
reference documents from miRBase22.1 (Kozomara et al., 2019), and the lncRNA sequence of each line is intercepted according to the binding file, which happens to follow the lncRNA name order. Third, all miRNAs and lncRNAs (originally 15,000 lines each) are de-duplicated to obtain 1,340 unduplicated miRNAs and 7,963 unduplicated lncRNAs. Finally, the serial numbers of the remaining miRNAs and lncRNAs are determined, and the miRNA-lncRNA interaction matrix is drawn in accordance with the tag file. A rough workflow is shown in **Figure 1**.

Given that multiple pieces of information are required to calculate the interactions when we adopt the linear neighborhood similarity method, we also utilize these features that represent sequence information in addition to the interaction matrix. Owing to the existence of orphaned miRNAs and lncRNAs, the sequence is more versatile than the interaction. We collate the  $k$ -mer frequency (Ahmed et al., 2020), GC content, number of base pairs, and MFE (Negri et al., 2018) of miRNAs and lncRNAs according to the de-duplication files, the original files, and the features here.

The data are summarized in **Table 1**.

### Exact Linear Neighborhood Propagation Method Based on Combined Information Improved Linear Neighborhood Similarity Measure

A matrix  $M$  is formed by  $n$ -dimensional feature vectors  $x_1, \dots, x_m$  in space, where each row  $x_i (x_{i1}, x_{i2}, \dots, x_{in})$  is regarded as a data point, and we assume that we can gather the attribute parts of other data points to get the current one. Adjacent data



points are usually viewed as possessing similar properties. Hence, the neighbors can be selected as the contributing force for the reconstruction of the point, whereas other irrelevant data points participate in the calculation but are assigned a weight of 0. For  $x_i$ , the common calculation method for selecting neighbors is Euclidean distance. Considering that various features, such as GC content and MFE, characterize the components of the data point in these dimensions and that similar data points are assumed to eventually form multidimensional vectors with similar directions, the Cosine distance is then chosen to select neighbors  $N_1(x_i)$  (a total of  $n_1$ ), and then the Euclidean distance is chosen to select nearer neighbors  $N_2(x_i)$  (a total of  $n_2$ ) on the basis of the Cosine distance to select more exact neighbors. The latter is a subset of the former, ensuring that neighbors are nearer in both direction and position. The order of the two can be swapped since the final selected neighborhood is the same. The percentages of neighbors are expressed by  $K_1$  and  $K_2$  where  $K_1 = n_1/m$  and  $K_2 = n_2/n_1$ .

To minimize the reconstruction error for  $m$  data points, we propose the objective function:

$$\begin{aligned} \min_{G,W} & \frac{1}{2} \|2M - (C_1 \odot G)M - (C_2 \odot W)M\|_F^2 \\ & + \frac{\mu}{2} \sum_{i=1}^m \|(C_1 \odot G)e + (C_2 \odot W)e\|_2^2 \end{aligned} \quad (1)$$

*s.t.*  $(C_1 \odot G)e = (C_2 \odot W)e = e, G \geq 0, W \geq 0$

where  $M$  is an  $m \times n$  matrix in the feature space, and both  $C_1$  and  $C_2$  are indicator matrices that separately indicate whether they are neighbors on the basis of the Cosine distance and nearer neighbors on the basis of the Euclidean distance.  $G$  and  $W$  are both weight matrices. Here,  $\mu$  is a weight parameter, and  $e$  is an  $m \times 1$  column vector with all elements being 1.  $\|\bullet\|_F$  is to obtain the Frobenius norm of a matrix.  $\|\bullet\|_2$  is to obtain the 2-norm of a vector. The first term of the objective function is to get the optimal weight matrix to minimize the reconstruction error of all data points, whereas the second term is to reduce overfitting during the reconstruction. For the weight matrices  $G$  and  $W$ , the elements are non-negative.

Given that the first neighbors are required to find the second ones, the objective function must be decomposed:

$$\begin{aligned} f(G) &= \frac{1}{2} \|M - (C_1 \odot G)M\|_F^2 + \frac{\mu}{2} \sum_{i=1}^m \|(C_1 \odot G)e\|_2^2 \\ \text{s.t.} & (C_1 \odot G)e = e, G \geq 0 \end{aligned} \quad (2)$$

The Lagrange multiplier method is used to solve Equation (2), which then has the following form:

$$\begin{aligned} \min_G & f(G) \\ \text{s.t.} & g(G) = e - (C_1 \odot G)e = 0 \\ & h(G) = -G \leq 0 \end{aligned}$$

Then

$$\begin{aligned} L(G, \lambda_1, \lambda_2) &= f(G) + \lambda_1^T h(G) + \lambda_2^T g(G) \\ &= \frac{1}{2} \|M - (C_1 \odot G)M\|_F^2 + \frac{\mu}{2} \sum_{i=1}^m \|(C_1 \odot G)e\|_2^2 \\ &+ \lambda_1^T (-G) + \lambda_2^T (e - (C_1 \odot G)e) \end{aligned} \quad (3)$$

where  $\lambda_1$  and  $\lambda_2$  are Lagrange factors.

According to the Karush-Kuhn-Tucker condition (Kjeldsen, 2000), the following conditions must be satisfied to determine the optimal value:

$$\begin{cases} \nabla_G L = 0 \\ g(G) = 0 \\ \lambda_1^T h(G) = 0 \end{cases}$$

The partial derivative of  $L$  is determined with respect to  $G$ :

$$\begin{aligned} L(G, \lambda_1, \lambda_2) &= \frac{1}{2} \|M - (C_1 \odot G)M\|_F^2 + \frac{\mu}{2} \sum_{i=1}^m \|(C_1 \odot G)e\|_2^2 \\ &+ \lambda_1^T (-G) + \lambda_2^T (e - (C_1 \odot G)e) \\ &= \frac{1}{2} \text{tr}((M - (C_1 \odot G)M)^T (M - (C_1 \odot G)M)) \\ &+ \frac{\mu}{2} \sum_{i=1}^m ((C_1 \odot G)e)^T ((C_1 \odot G)e) - \lambda_1^T G - \lambda_2^T ((C_1 \odot G)e - e) \\ &= \frac{1}{2} [\text{tr}(M^T M) - \text{tr}(M^T ((C_1 \odot G)M)) - \text{tr}(((C_1 \odot G)M)^T M) \\ &+ \text{tr}(((C_1 \odot G)M)^T ((C_1 \odot G)M))] \\ &+ \frac{\mu'}{2} ((C_1 \odot G)e)^T ((C_1 \odot G)e) \\ &- \lambda_1^T G - \lambda_2^T ((C_1 \odot G)e - e) \end{aligned} \quad (4)$$

$\nabla_G L$

$$\begin{aligned} &= \frac{1}{2} \left[ \frac{\partial \text{tr}(M^T M)}{\partial G} - \frac{\partial \text{tr}(M^T ((C_1 \odot G)M))}{\partial G} \right. \\ &- \frac{\partial \text{tr}(((C_1 \odot G)M)^T M)}{\partial G} \\ &+ \left. \frac{\partial \text{tr}(((C_1 \odot G)M)^T ((C_1 \odot G)M))}{\partial G} \right] \\ &+ \frac{\mu'}{2} \left( \frac{\partial \text{tr}((C_1 \odot G)e)^T ((C_1 \odot G)e)}{\partial G} \right) \\ &- \frac{\partial \text{tr}(\lambda_1^T G)}{\partial G} - \frac{\partial \text{tr}(\lambda_2^T ((C_1 \odot G)e - e))}{\partial G} \\ &= \frac{1}{2} \left[ 0 - C_1 \odot M M^T - \frac{\partial \text{tr}(M^T ((C_1 \odot G)M))}{\partial G} \right. \\ &+ \left. \frac{\partial \text{tr}(M^T (C_1 \odot G)^T (C_1 \odot G)M)}{\partial G} \right] \\ &+ \frac{\mu'}{2} \left( \frac{\partial \text{tr}(e^T (C_1 \odot G)^T (C_1 \odot G)e)}{\partial G} \right) \\ &- \lambda_1 - C_1 \odot (\lambda_2 e^T) \\ &= \frac{1}{2} \left[ -C_1 \odot M M^T - C_1 \odot M M^T \right] \end{aligned}$$



$$\begin{aligned}
& + 2C_1 \odot (C_1 \odot G)MM^T] \\
& + \frac{\mu'}{2} \times 2C_1 \odot (C_1 \odot G)ee^T \\
& - \lambda_1 - C_1 \odot (\lambda_2 e^T) = C_1 \odot (C_1 \odot G)MM^T \\
& - C_1 \odot MM^T + \mu' C_1 \odot (C_1 \odot G)ee^T \\
& - \lambda - C_1 \odot (\lambda_2 e^T) \\
& = C_1 \odot ((C_1 \odot G)MM^T + \mu'(C_1 \odot G)ee^T - MM^T \\
& - \lambda_2 e^T) - \lambda_1 \quad (5)
\end{aligned}$$

Then

$$\begin{cases} \nabla_G L = C_1 \odot ((C_1 \odot G)MM^T + \\ \mu'(C_1 \odot G)ee^T - MM^T - \lambda_2 e^T) - \lambda_1 = 0 \\ e - (C_1 \odot G)e = 0 \\ \lambda_1^T (-G) = 0 \end{cases}$$

If  $\lambda_1^T = 0$ , then there will be

$$\begin{aligned}
\nabla_G L &= C_1 \odot ((C_1 \odot G)MM^T \\
& + \mu'(C_1 \odot G)ee^T - MM^T - \lambda_2 e^T) = 0 \quad (6)
\end{aligned}$$

In that case,

$$G_{ij} = \frac{G_{ij}(MM^T + \lambda_2 e^T)_{ij}}{((C_1 \odot G)MM^T + \mu(C_1 \odot G)ee^T)_{ij}} \quad (7)$$

If  $\lambda_1^T \neq 0$ , then there will be  $G = 0$ . Thus,  $G_{ij} = 0$ .

Given the relevance of data point-based reconfiguration that  $G_{ij} \neq 0$  when  $x_j \in N_1(x_i)$ , the solution is

$$G_{ij} = \begin{cases} \frac{G_{ij}(MM^T + \lambda_2 e^T)_{ij}}{((C_1 \odot G)MM^T + \mu(C_1 \odot G)ee^T)_{ij}} & , x_j \in N_1(x_i) \\ 0 & , x_j \notin N_1(x_i) \end{cases} \quad (8)$$

Thus far, the iterative form with unknown parameters has been obtained. We inscribe its equivalent form for Equation (2) to obtain the following parameter:

$$\begin{aligned}
\min_{G_i} L^i &= \frac{1}{2} \|x_i - \sum_{j: x_j \in N_1(x_i)} G_{i,j}\|^2 \\
& + \frac{\mu}{2} (\sum_{j: x_j \in N_1(x_i)} |G_{i,j}|)^2 = \frac{1}{2} \vartheta_i^T Gr^i \vartheta_i + \frac{\mu}{2} \|\vartheta_i\|^2 \\
& \quad \text{s.t. } e^T \vartheta_i = 1, \vartheta_i \geq 0 \quad (9)
\end{aligned}$$

where  $Gr^i$  is the gram matrix. If  $x_j \in N_1(x_i)$  and  $x_k \in N_1(x_i)$ , then  $Gr_{j,k} = (x_i - x_j)^T (x_i - x_k)$ . Otherwise,  $Gr_{j,k} = 0$ . Solving Equation (9) using the Lagrange multiplier method yields

$$L^i = \frac{1}{2} \vartheta_i^T Gr^i \vartheta_i + \frac{\mu}{2} \|\vartheta_i\|^2 - \lambda_i (e^T \vartheta_i - 1) - \eta^T \vartheta_i \quad (10)$$

By taking the partial derivatives of  $g$  and  $\lambda$ ,  $\lambda_i$  can be obtained:

$$\lambda_i = (\vartheta_i^T Gr^i \vartheta_i + \mu (e^T \vartheta_i)^2) / e^T \vartheta_i \quad (11)$$

TABLE 1 | Dataset composition.

Molecule	Quantity	Features
lncRNA	7,963	$k$ -mer frequency ( $k = 1, k = 2, k = 3$ ), GC content, number of base pairs, MFE
miRNA	1,340	$k$ -mer frequency ( $k = 1, k = 2$ ), GC content, number of base pairs, MFE
interaction	7,500	-

where the reconstruction error is close to none, i.e.,  $\frac{1}{2} \vartheta_i^T Gr^i \vartheta_i \approx 0$ . According to the Lagrange multiplier method,  $e^T \vartheta_i - 1 = 0$ . Thus,  $\lambda_i = \mu$ . If  $\lambda = \mu \times e$ ,  $G$  can be represented as

$$G_{ij} = \begin{cases} \frac{G_{ij}(MM^T + \mu ee^T)_{ij}}{((C_1 \odot G)MM^T + \mu(C_1 \odot G)ee^T)_{ij}} & , x_j \in N_1(x_i) \\ 0 & , x_j \notin N_1(x_i) \end{cases} \quad (12)$$

The similarity matrix based on the Cosine distance is then acquired through iteration until convergence. Similarly,  $W$  is obtained with respect to  $G$ :

$$W_{ij} = \begin{cases} \frac{W_{ij}(MM^T + \mu ee^T)_{ij}}{((C_2 \odot G)MM^T + \mu(C_2 \odot G)ee^T)_{ij}} & , x_j \in N_2(x_i) \\ 0 & , x_j \notin N_2(x_i) \end{cases} \quad (13)$$

The final similarity matrix can be acquired by iterating until convergence or the maximum rounds.

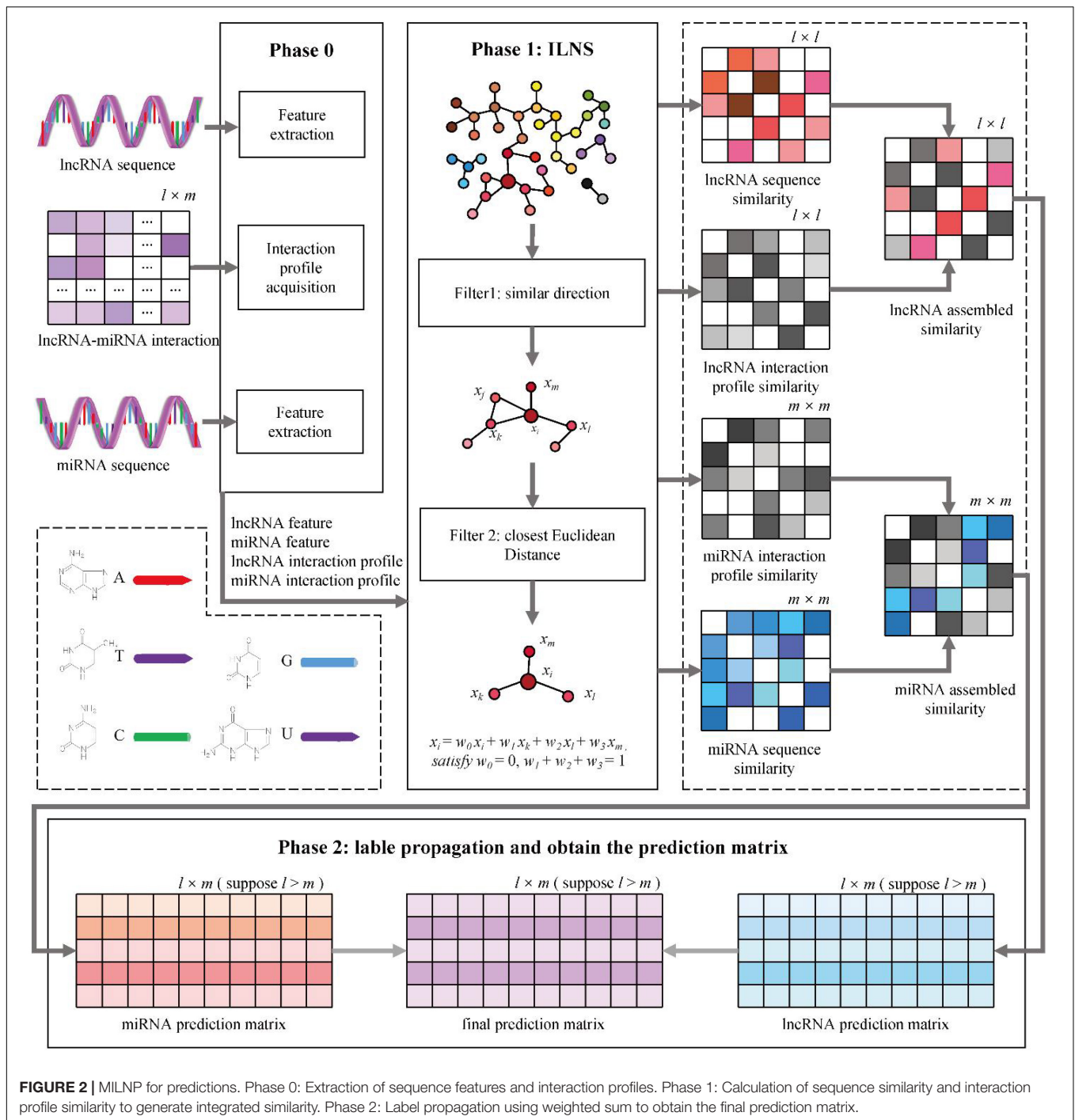
## Feature Extraction and Symbol Definition

Given a set of  $l$  lncRNAs,  $l_1, \dots, l_i, \dots, l_l$ , and a set of  $m$  miRNAs,  $m_1, \dots, m_j, \dots, m_m$ , whose interaction is represented by a matrix  $Y$  of  $l \times m$ , if there exists an interaction between lncRNA  $l_i$  and miRNA  $m_j$ , then  $Y_{ij} = 1$ ; otherwise,  $Y_{ij} = 0$ .

Four features of lncRNAs and miRNAs are extracted (110 features in total). The frequency of  $l$  lncRNAs with  $4^k$  long contiguous subsequences is calculated. Herein, we assumed  $k = 1, 2, 3$  to obtain the lncRNA-related feature vector. The sequence similarity between pairs of  $l$  lncRNAs is calculated to yield the similarity matrix of  $l \times l$ , which is denoted as  $S\_lncS$ . In the same manner, the similarity matrix of  $m$  miRNAs is denoted as  $S\_miS$  of  $m \times m$ . The interaction profiles of lncRNAs and miRNAs are derived from the interaction matrix. For lncRNA  $l_i$ , the interaction profile indicates whether it interacts with each miRNA, matching the  $i$ -th row of  $Y$ , i.e.,  $Y(i, :)$ . Similarly, for miRNA  $m_j$ , it matches the  $j$ -th row of  $Y$ , i.e.,  $Y(:, j)$ . The similarity between two interaction profiles of  $l$  lncRNAs is calculated as the matrix  $S\_lncP$  of  $l \times l$ , whereas the similarity between two interaction profiles of  $m$  miRNAs is computed as the matrix  $S\_miP$  of  $m \times m$ .

## Label Propagation Based on Improved Linear Neighborhood Similarity

Label propagation (Kato et al., 2009) assigns labels to previously unlabeled data points. During label assignment, the labels of labeled data points are propagated to unlabeled data points. The core idea of the label propagation algorithm is that similar nodes should have similar labels. It involves two stages, namely, calculating the similarity matrix and propagating the labels.



**FIGURE 2 |** MILNP for predictions. Phase 0: Extraction of sequence features and interaction profiles. Phase 1: Calculation of sequence similarity and interaction profile similarity to generate integrated similarity. Phase 2: Label propagation using weighted sum to obtain the final prediction matrix.

The edge from node  $i$  to node  $j$  represents the similarity of these nodes. All edge weights constitute a weight matrix, where the higher the similarity the larger the weight. Herein, ILNS is adopted to construct the similarity matrix and calculate the Cosine-distance neighbors and the Euclidean-distance neighbors of each node until convergence. Nearer neighbors of each data point are fixed to a certain proportion, and the weights of others are 0. The weight matrix is actually a sparse matrix. The labels are propagated through the edges between the nodes. The larger the

weight of the edge, the more similar the nodes are to each other and the easier to propagate the labels (Zang and Zhang, 2012; Zhang et al., 2016). For  $m$  data points  $x_1, \dots, x_m$ , an  $m \times m$  probability transfer matrix  $P$  is defined as

$$P_{ij} = P(i \rightarrow j) = \frac{w_{ij}}{\sum_{k=1}^m w_{ik}} \quad (14)$$

where  $P_{ij}$  represents the transferring probability and  $w_{ij}$  is the weight. The propagation involves three steps. First, a unique label

**TABLE 2** | Parameter setting.

Phase	Parameter	Range	Step length
Phase 1	$K_1$	0.1 – 0.9	0.1
Phase 1	$K_2$	0.1 – 1.0	0.1
Phase 2	$\alpha$	0.1 – 0.9	0.1
Phase 2	$\beta$	0.0 – 1.0	0.05

is allocated to each node, i.e., label one for node one and label  $i$  for node  $i$ , where the labels are different from each other. Second, for node  $j$ , all nodes are traversed to discover their neighboring nodes and obtain their labels to obtain the label with the most occurrences. If more than one label satisfies the largest number of occurrences, then one is randomly selected to replace the current label. Finally, if the label of node  $j$  no longer changes after this round of relabeling or the pre-set number of rounds is reached, then the iteration is stopped. Otherwise, step 2 is repeated.

### Prediction Model Multi-Source Information-Based Linear Neighborhood Propagation

The sequence and interaction profiles of lncRNAs and miRNAs are captured to develop our model. The workflow of MINLP is shown in **Figure 2**. The specific steps are as follows:

Step 1: the sequence feature similarity  $S_{lncS}$  and  $S_{miS}$  are calculated using the ILNS algorithm.

Step 2: the interaction profiles of all lncRNAs and all miRNAs are exported according to the interaction matrix of lncRNAs and miRNAs.

Step 3: the interaction profile similarity  $S_{lncP}$  and  $S_{miP}$  are calculated using the ILNS algorithm.

Step 4:  $S_{lncS}$  and  $S_{lncP}$  are combined to obtain lncRNA integrated similarity, and  $S_{miS}$  and  $S_{miP}$  are combined to obtain miRNA integrated similarity.

Step 5: for the two integration similarities, the linear neighborhood propagation method is used to generate the prediction of lncRNA and the prediction of miRNA.

Step 6: the weighted sum of the two prediction matrices is calculated, and the final interaction prediction matrix is determined.

## EXPERIMENTS AND RESULTS

### Evaluation Criteria

The criteria for measuring prediction models are area under curve (AUC), area under precision–recall (AUPR), REC, SPE, and ACC. AUC is the area under the receiver operating curve (ROC) coordinated by true positive rate–false positive rate, which is suitable for observing model performance in the case of a balanced positive and negative sample size. The formulae of REC, SPE, and ACC are as follows.

$$\begin{cases} REC = \frac{TP}{TP+FN} \\ SPE = \frac{TN}{FP+TN} \\ ACC = \frac{TP+TN}{TP+FN+FP+TN} \end{cases} \quad (15)$$

The performance is evaluated *via* fivefold cross-validation. To achieve more accurate outcomes, each fivefold cross-validation is repeated for 20 rounds to ensure that a sufficient number of learnings are reached.  $K$ -fold cross-validation is frequently used to upgrade model performance, where data is divided into  $K$  equal parts, one of which acts as test data and the other acts as training data. A distinct test set is selected each time, and the rest serves as a training set. Finally, the results of  $K$  experiments are averaged.

### Parameter Settings

A total of four relevant parameters are obtained in this work.

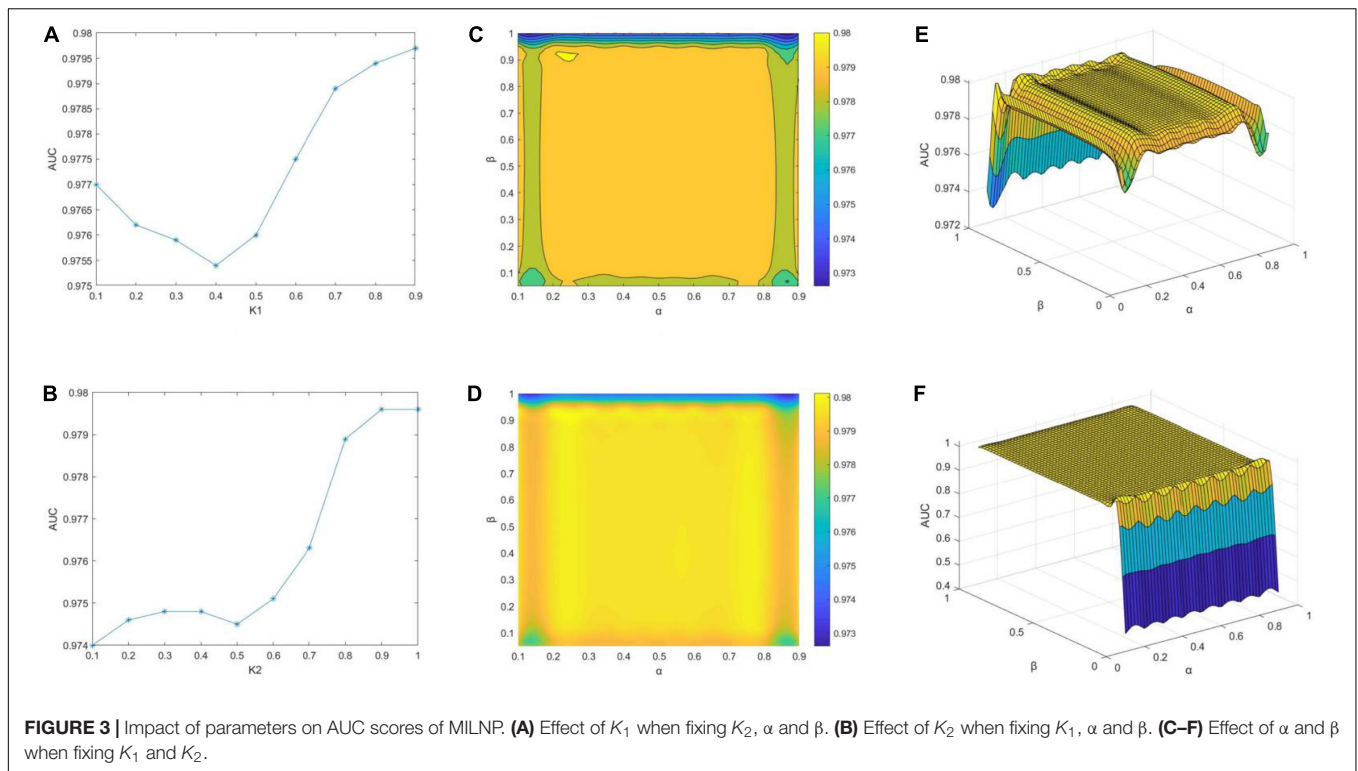
In the computation of ILNS, Cosine distance-based neighbors (With the ratio of  $K_1$ ) and Euclidean distance-based neighbors (With the ratio of  $K_2$ ) are computed.  $K_1$  is set to {0.1, 0.2, ..., 0.9}, whereas  $K_2$  is set to {0.1, 0.2, ..., 1}, and their step size is 0.1. The purpose of this arrangement is to ensure that  $K_2$  considers all neighbors generated by  $K_1$ , regardless of the size of  $n_1$ . During label propagation, the parameter  $\alpha$  is set as the probability of label absorption, i.e., for node  $x_j$  the probability of absorbing the label of its nearest neighbor node  $x_i$  is  $\alpha$ . The value of  $\alpha$  is set within the range {0.1, 0.2, ..., 0.9}, and the step size is 0.1. After the lncRNA prediction matrix  $SL$  and the miRNA prediction matrix  $SM$  are figured out,  $\beta$  is the trade-off parameter, i.e., the final prediction matrix will be measured as  $\beta \times SL + (1 - \beta) \times SM$ . The value of  $\beta$  is within {0.0, 0.05, ..., 1.0}, and the step size is 0.05.

The settings of the parameters are shown in **Table 2**. Subsequent experiments are conducted with the most optimal parameter combinations.

## Results

### Optimal Parameters

The effects of the different parameters are visualized in **Figure 3**. First,  $\alpha$  and  $\beta$  are fixed. Theoretically, the neighbors are set twice to find a more accurate batch faster. However, serious analysis reveals that a change in  $K_2$  is logical within a certain range of  $K_1$ ; otherwise, even when  $K_2$  is 100%, it will not be very helpful. Let  $K_2 = 1.0$ , as  $K_1$  changes from 0.1 to 0.9, we find that  $K_2 = 0.9$  is the optimal value, as shown in **Figure 3A**. AUC initially decreases and then increases with  $K_1$  and reaches its lowest point at 0.4. The AUC values are always above 0.975.  $K_1 = 0.9$  is then fixed, and  $K_2$  is changed. As  $K_2$  becomes larger, AUC tends to increase globally and reaches the maximum at 0.9 (**Figure 3B**). The most pronounced increase is evident from 0.7 to 0.8, clearly demonstrating that 0.8 is a cut-off point. Finally, the impact of  $\alpha$  and  $\beta$  is investigated. The contour and concentration plots in **Figures 3C,D**, respectively, demonstrate the variations in AUC with  $\alpha$  and  $\beta$ . The gradient from blue to yellow is set to indicate the increase in AUC values. Herein, the scenario with  $\beta = 0$  is dropped to detect subtle variations in the other cases owing to our prior observation that the AUC value is as low as 0.4 in the case of  $\beta = 0$ , thereby forming a cliff-like change from others. Both plots show that the best results are achieved at  $\alpha$  of 0.2–0.8 and  $\beta$  of 0.15–0.9, which are extremely close to 0.98. The yellowest areas appear locally as a result of drawing tool error, but it does not affect the fact that the results are roughly consistent. **Figures 3E,F** present the grid plots of AUC with respect to variations in  $\alpha$  and



**TABLE 3 |** Performance of models with combinations of different algorithms and information.

Model	Algorithm	Information	AUC	REC	SPE	ACC	AUPR
SLNPM-I	LNS	Sequences similarity	0.8596	0.2883	0.9962	0.9932	0.1856
SLNPM-II	LNS	IP similarity	0.8756	0.5993	0.9990	0.9973	<b>0.5981</b>
SLNPM	LNS	Sequence similarity and IP similarity	0.9768	0.9613	0.9993	0.9993	0.5132
MILNP-I	ILNS	Sequence similarity	0.8561	0.4620	0.9902	0.9916	0.1249
MILNP-II	ILNS	IP similarity	<b>0.9804</b>	0.9600	0.9994	0.9994	0.5235
MILNP	ILNS	Sequence similarity and IP similarity	0.9797	<b>0.9629</b>	<b>0.9994</b>	<b>0.9994</b>	0.5297

**TABLE 4 |** Performances of different methods.

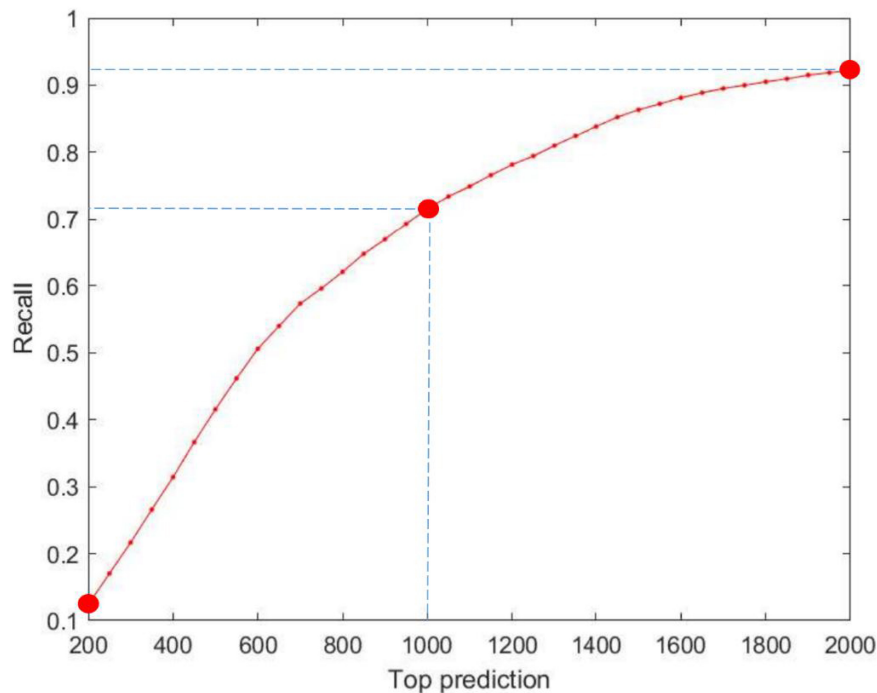
Methods	AUC	REC	SPE	ACC	AUPR
Pmlipred	0.8386	0.9493	0.9087	0.9290	0.4304
CIRNN	-	0.9413	-	0.9604	-
CNNRF1	0.8562	0.9531	0.9083	0.9307	0.4321
CNNRF2	0.8284	0.9597	0.9047	0.9322	0.4340
SLNPM	0.9768	0.9613	0.9993	0.9993	0.5132
MILNP	<b>0.9797</b>	<b>0.9629</b>	<b>0.9994</b>	<b>0.9994</b>	<b>0.5297</b>

$\beta$ , respectively. **Figure 3E** is the case with  $\beta = 0$  removed from the corresponding plots in 3(C) and 3(D). **Figure 3F** includes all cases and validates the previous interpretation that the inclusion makes distinguishing the changes from the others difficult. The models of all ranges show that the best parameters are  $K_1 = 0.9$ ,  $K_2 = 1.0$ ,  $\alpha = 0.7$ , and  $\beta = 0.35$  when AUC at this point is 0.9797. This tells that the performance of our model is attributed to Cosine distance, determining a more accurate neighborhood which is preferable to applying Euclidean distance only.

### Control Information Sources and Similarity Measurement Algorithm

To demonstrate the superiority of our model, we calculate a similarity network with a single information source and build linear neighborhood propagation models to compare with MILNP. The results are presented in **Table 3**. The optimal values obtained are set in bold typeface. We construct these models based on the optimal combination of parameters. Except for the difference in information sources, all the other processes are guaranteed to be the same. The model using only sequence similarity with ILNS as the core algorithm is named MILNP-I. The model using only interaction profile similarity with ILNS as the core algorithm is named MILNP-II. Overall, both of them are inferior to MILNP with integrated information. MILNP-II is generally very close to our model, indicating that the interaction information is a key contributor throughout the prediction. Thus, with specific optimization, MILNP-II could be credible for predicting isolated lncRNAs or miRNAs. The performance of MILNP-I is not as good as that of MILNP-II, but its AUC is barely satisfactory. We also tried another method for





**FIGURE 4 |** Performance of MILNP's top-rank predictions, where the X-axis refers to the top 200 to top 2,000 predictions and the Y-axis refers to the recall generated by MILNP.

**TABLE 5 |** Top 10 predictions for miRNA “gma-miR395a” and lncRNA “lcl| Gmax\_Glyma.18G279100.1” by MILNP.

NO	gma-miR395a	Confirmation	lcl  Gmax_Glyma.18G279100.1	Confirmation
	<b>lncRNAs</b>		<b>miRNAs</b>	
1	lcl  Gmax_Glyma.19G246900.1	NO	gma-miR319a	YES
2	lcl  Gmax_Glyma.15G199100.1	NO	gma-miR319h	YES
3	lcl  Gmax_Glyma.14G142000.1	YES	gma-miR319g	YES
4	lcl  Gmax_Glyma.08G153500.1	YES	gma-miR319i	NO
5	lcl  Mtruncatula_Medtr1g017330.1	NO	gma-miR319p	NO
6	lcl  Gmax_Glyma.16G164700.2	NO	gma-miR319c	YES
7	lcl  Gmax_Glyma.07G234600.1	YES	gma-miR319q	NO
8	lcl  Mtruncatula_Medtr8g099205.1	NO	gma-miR159a-3p	YES
9	lcl  Gmax_Glyma.12G192900.2	YES	gma-miR319f	NO
10	lcl  Gmax_Glyma.02G080100.1	NO	gma-miR5676	NO

calculating similarity, LNS (Zhang et al., 2018b), to obtain new models, named the SLNPM-series, as shown in **Table 3**. With the information source controlled, a comparison of SLNPM-II and MILNP-II reveals that ILNS is more accurate. All these results clearly validate the superiority of MILNP in terms of information integration and similarity calculation.

### Comparison With Existing Methods

As far as we know, very few studies have investigated lncRNA–miRNA interactions in plants. We select Pmlipred (Kang et al., 2020) and CIRNN (Zhang et al., 2020) as reference methods. Both methods predict plant lncRNA–miRNA interactions. Pmlipred (Kang et al., 2020) builds a prediction model by using a machine learning approach combined with a deep learning approach, and

the final prediction results are made from fuzzy decisions of the two components. We use the publicly available source code on GitHub to implement Pmlipred (Kang et al., 2020). CIRNN (Zhang et al., 2020) builds integrated deep learning models with both a CNN and an IndRNN, where the former is used to automatically extract gene sequence functional features, whereas the latter is utilized to obtain sequence feature representations and dependencies. We replicate it in the detailed description of CIRNN (Zhang et al., 2020). The results of the comparison are summarized in **Table 4**. The optimal values obtained are set in bold typeface. The performance of the two methods is clearly good, but not as good as that of our model. In particular, the AUC and ACC values have a relatively large gap. We notice that both methods stretch to deep learning.

Thus, we implement similar models on their basis to measure their effectiveness. We construct a CNN-Gate Recurrent Unit (GRU) combinatorial model. Sequential features are extracted from the original data by CNN and compressed into a one-dimensional vector in the flattened layer to input into GRU. This process is well-suited for processing sequential information. We consider GRU instead of others because GRU has fewer parameters and reduces overfitting. We first use a three-layer CNN and a single-layer GRU mixed with RF to obtain CNNRF1. We then add another layer of CNN on top of that to obtain CNNRF2. Gladly, although our MILNP is simple and built by deriving mathematical formulas *via* a top-down approach and layer by layer, the results are satisfactory. For the baseline SLNSM (Zhang et al., 2018b) that is originally created for animal prediction, we run our dataset and observe that our method is slightly better. That is because we focus on improving the computational procedure for linear neighborhood similarity by adding the spatial direction restriction. The optimal parameter combination shows that such performance is attributed to Cosine distance, determining a more accurate neighborhood, which is preferable to the approach of the baseline.

### Case Study and Discussion

Top-rank prediction is an important way of visualizing the performance of the models. We examine the top-rank predictions from 200 to 2,000 and identify the percentage of interactions that are truly correct. As shown in **Figure 4**, an average of 186 positive interactions per prediction is reached in the top 200 predictions, whereas 1,380 real interactions are determined in the top 2,000 predictions. The results demonstrate the good performance of our model.

Furthermore, we predict the interactions of isolated lncRNAs and miRNAs with MILNP. For the isolated lncRNA or miRNA, only sequence-dependent information can be used. In separate cases gma-miR395a and lcl| Gmax\_Glyma.18G279100.1 are taken as examples. We validate the prediction of the selected miRNA and lncRNA with respect to RNAhybrid2.1.2 (Rehmsmeier, 2004). All predictions are sorted in descending order of probability. For miRNA gma-miR395a, 4 of the top 10 are correctly predicted, as shown in **Table 5**, thereby confirming the predictive power of MILNP. However, the list reveals that the fifth and eighth detected lncRNAs in the prediction of miRNA “gma-miR395a” belong to *Medicago truncatula*, as evidenced by their nomenclature. The situation is worth contemplating. On the one hand, this indicates that the selected samples may affect the performance of MILNP. On the other hand, it inspires us to further explore cross-species linkages and assume that the remaining uncertified interactions are possible. Likewise, we make predictions for lncRNA lcl| Gmax\_Glyma.18G279100.1. To our surprise, five of the results happen to be identified by the tool as having interactions, a result that is very encouraging. We conjecture that the remaining ones predicted by our model can be possible. As demonstrated by the results of the comparison of the two sets of predictions, the fact that sample selection has a great influence on the prediction results should not be ignored. The association between the selected sample and other samples also affects the results. For

those that have a similarity with many samples, the prediction results may be more accurate. If a sample has little similarity with other samples, then predicting its potential interactions will be difficult.

The different results also prompt us to reconsider the results of previous experiments. We find that, although MILNP achieves good AUC and ACC, its PRE is relatively low, which may be attributed to the model itself and the distribution of the dataset. Some very similar samples may have confused the model and that induces it to arrive at a wrong judgment. Nevertheless, it could also be a new revelation that suggests these possible associations. This assumption warrants further biological laboratory validation.

## CONCLUSION

lncRNA-miRNA interactions are important because they influence various biological activity processes. Most studies on these interactions focused on animals. Although experimental results derived from studies of plants are not as easy to verify as those obtained from animals, current research is not merely a conjecture. Great improvements have been made by scientists after proposing bold assumptions and providing carefully evaluated proofs. Herein, we attempt to study plant interactions and propose a linear neighborhood propagation model based on combinatorial information. We validated it on datasets of three plants. We have obtained relatively good results. More importantly, we proposed a novel method for measuring similarity from mathematical fundamentals. We used the combined information of molecular sequences and interactions to construct a similarity network with a guarantee of being nearer in both spatial location and direction. We achieved the final prediction by label propagation. A series of experiments showed the outstanding performance of our model, demonstrating the superiority of the combinatorial information. We also attempted to predict isolated lncRNAs and miRNAs without any interaction yet and validated the predictions with existing tools. Our model possesses good generalization properties and can be used to discover new interaction relationships. Our multisource information-based linear neighborhood propagation method is a novel and unique method for predicting plant lncRNA-miRNA interactions. However, the entire study requires a large time investment of about 3 months. Hence, in a follow-up study, we will tune the parameters to make the model more efficient. We will also consider deep learning methods on this basis and combine the results that we may obtain.

## 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

MG wrote the first draft of the manuscript. XF wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This work was supported in part by the National Natural Science Foundation of China (62002111, 61872309, 61972138, and 62006074), in part by the Fundamental Research Funds

for the Central Universities (531118010355), in part by China Postdoctoral Science Foundation (2019M662770 and 2020M672487), in part by Hunan Provincial Natural Science Foundation of China (2020JJ4215), and in part by Changsha Municipal Natural Science Foundation (kq2014058).

## ACKNOWLEDGMENTS

We would like to thank Xiangxiang Zeng for kind suggestions and discussions that have helped improve the presentation of this manuscript.

## REFERENCES

- Aglawe, S. B., Verma, A. K., and Upadhyay, A. K. (2021). Bioinformatics tools and databases for genomics-assisted breeding and population genetics of plants: a review. *Curr. Bioinform.* 16, 766–773.
- Ahmed, S., Hossain, Z., Uddin, M., Taherzadeh, G., Sharma, A., Shatabda, S., et al. (2020). Accurate prediction of RNA 5-hydroxymethylcytosine modification by utilizing novel position-specific gapped k-mer descriptors. *Comput. Struct. Biotechnol. J.* 18, 3528–3538. doi: 10.1016/j.csbj.2020.10.032
- Allison, S. J. (2016). A lncRNA and miRNA megacluster in diabetic nephropathy. *Nat. Rev. Nephrol.* 12:713. doi: 10.1038/nrneph.2016.151
- Amin, N., McGrath, A., and Chen, Y.-P. P. (2019). Evaluation of deep learning in non-coding RNA classification. *Nat. Mach. Intell.* 1, 246–256. doi: 10.1038/s42256-019-0051-2
- Ard, R., Tong, P., and Allshire, R. C. (2014). Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nat. Commun.* 5:5576. doi: 10.1038/ncomms6576
- Ayachit, G., Shaikh, I., Pandya, H., and Das, J. (2020). Salient features, data and algorithms for MicroRNA screening from plants: a review on the gains and pitfalls of machine learning techniques. *Curr. Bioinform.* 15, 1091–1103.
- Banerjee, S., Yabalooru, S. R. K., and Karunakaran, D. (2020). Identification of mRNA and non-coding RNA hubs using network analysis in organ tropism regulated triple negative breast cancer metastasis. *Comp. Biol. Med.* 127:104076. doi: 10.1016/j.combiomed.2020.104076
- Cai, L., Ren, X., Fu, X., Peng, L., Gao, M., and Zeng, X. (2020a). iEnhancer-XG: Interpretable sequence-based enhancers and their strength predictor. *Bioinformatics* 2020:914. doi: 10.1093/bioinformatics/btaa914
- Cai, L., Wang, L., Fu, X., and Zeng, X. (2021). Active semisupervised model for improving the identification of anticancer peptides. *ACS Omega* 6, 23998–24008. doi: 10.1021/acsomega.1c03132
- Cai, L., Wang, L., Fu, X., Xia, C., Zeng, X., and Zou, Q. (2020b). ITP-Pred: an interpretable method for predicting therapeutic peptides with fused features low-dimension representation. *Brief. Bioinform.* 2020:367. doi: 10.1093/bib/bbaa367
- Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.* 18, 962–972. doi: 10.1038/ni.3771
- Chen, Y., Fu, X., Li, Z., Peng, L., and Zhuo, L. (2021). Prediction of lncRNA-protein interactions via the multiple information integration. *Front. Bioeng. Biotechnol.* 9:113. doi: 10.3389/fbioe.2021.647113
- Dai, Q., Chu, Y., Li, Z., Zhao, Y., Mao, X., Wang, Y., et al. (2021). MDA-CF: predicting miRNA-disease associations based on a cascade forest model by fusing multi-source information. *Comp. Biol. Med.* 136:104706. doi: 10.1016/j.combiomed.2021.104706
- Daneshvar, K., Ardehali, M. B., Klein, I. A., Hsieh, F.-K., Kratkiewicz, A. J., Mahpour, A., et al. (2020). lncRNA DIGIT and BRD3 protein form phase-separated condensates to regulate endoderm differentiation. *Nat. Cell Biol.* 22, 1211–1222. doi: 10.1038/s41556-020-0572-2
- DeVeale, B., Swindlehurst-Chan, J., and Belloch, R. (2021). The roles of microRNAs in mouse development. *Nat. Rev. Genet.* 2021:95. doi: 10.1038/s41576-020-00309-5
- Fang, H., Bonora, G., Lewandowski, J. P., Thakur, J., Filippova, G. N., Henikoff, S., et al. (2020). Trans- and cis-acting effects of Firre on epigenetic features of the inactive X chromosome. *Nat. Commun.* 11:6053. doi: 10.1038/s41467-020-19879-3
- Fu, X., Cai, L., Zeng, X., and Zou, Q. (2020). StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics* 36, 3028–3034. doi: 10.1093/bioinformatics/btaa131
- Fu, X., Zhu, W., Cai, L., Liao, B., Peng, L., Chen, Y., et al. (2019). Improved pre-miRNAs identification through mutual information of pre-miRNA sequences and structures. *Front. Genet.* 10:119. doi: 10.3389/fgene.2019.0119
- Geisler, S., and Collier, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14, 699–712. doi: 10.1038/nrm3679
- Goodall, G. J., and Wickramasinghe, V. O. (2021). RNA in cancer. *Nat. Rev. Cancer* 21, 22–36. doi: 10.1038/s41568-020-00306-0
- Intell, N. M. (2019). Remodelling machine learning: an AI that thinks like a scientist. *Nat. Mach. Intell.* 2019:263. doi: 10.1038/s42256-019-0026-3
- Jia, L., and Luan, Y. (2022). Multi-feature fusion method based on linear neighborhood propagation predict plant lncRNA-Protein Interactions. *Interdiscip. Sci. Comp. Life Sci.* 2022:17. doi: 10.1007/s12539-022-00501-7
- Jiang, J., Xing, F., Wang, C., and Zeng, X. (2018). Identification and analysis of rice yield-related candidate genes by walking on the functional network. *Front. Plant Sci.* 9:1685.
- Jiang, J., Xing, F., Wang, C., Zeng, X., and Zou, Q. (2019). Investigation and development of maize fused network analysis with multi-omics. *Plant Physiol. Biochem.* 141, 380–387.
- Kang, Q., Meng, J., Cui, J., Luan, Y., and Chen, M. (2020). PmlPred: a method based on hybrid model and fuzzy decision for plant miRNA-lncRNA interaction prediction. *Bioinformatics* 36, 2986–2992. doi: 10.1093/bioinformatics/btaa074
- Kato, T., Kashima, H., and Sugiyama, M. (2009). Robust label propagation on multiple networks. *IEEE Trans. Neural Networks* 20:35. doi: 10.1109/TNN.2008.2003354
- Kjeldsen, T. H. (2000). A contextualized historical analysis of the kuhn-tucker theorem in nonlinear programming: the impact of world war II. *Hist. Math.* 27, 331–361. doi: 10.1006/hmat.2000.2289
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi: 10.1093/nar/gky1141
- Li, G., Luo, J., Xiao, Q., Liang, C., and Ding, P. (2018). Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. *J. Biomed. Inform.* 82, 169–177. doi: 10.1016/j.jbi.2018.05.005
- Li, J., Zhang, C., Shi, Y., Li, Q., Li, N., and Mi, Y. (2021). Identification of KEY lncRNAs and mRNAs associated with oral squamous cell carcinoma progression. *Curr. Bioinform.* 16, 207–215.
- Li, P., Su, T., Zhang, D., Wang, W., Xin, X., Yu, Y., et al. (2021). Genome-wide analysis of changes in miRNA and target gene expression reveals key roles in heterosis for Chinese cabbage biomass. *Horticult. Res.* 8:39. doi: 10.1038/s41438-021-00474-6

- Liu, W., Jiang, Y., Peng, L., Sun, X., Gan, W., Zhao, Q., et al. (2021). Inferring gene regulatory networks using the improved markov blanket discovery algorithm. *Interdiscip. Sci. Comp. Life Sci.* 2021, 1–14. doi: 10.1007/s12539-021-00478-9
- Liu, X., Peng, D., Cao, Y., Zhu, Y., Yin, J., Zhang, G., et al. (2020). Upregulated lncRNA DLX6-AS1 underpins hepatocellular carcinoma progression via the miR-513c/Cul4A/ANXA10 axis. *Cancer Gene Ther.* 2020:233. doi: 10.1038/s41417-020-00233-0
- Liu, Y., Zeng, X., He, Z., and Zou, Q. (2017). Inferring MicroRNA-Disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans. Comp. Biol. Bioinform.* 14, 905–915. doi: 10.1109/tcbb.2016.2550432
- Ma, X., Thakar, S. B., Zhang, H., Yu, Z., Meng, L., and Yue, J. (2020). Bioinformatics analysis of the rhizosphere microbiota of Dangshan Su pear in different soil types. *Curr. Bioinform.* 15, 503–514.
- Negri, T., Alves, W., Bugatti, P., Saito, P., Domingues, D., and Paschoal, A. (2018). Pattern recognition analysis on long noncoding RNAs: a tool for prediction in plants. *Brief. Bioinform.* 20:34. doi: 10.1093/bib/bby034
- Noviello, T. M. R., Di Liddo, A., Ventola, G. M., Spagnuolo, A., D’Aniello, S., Ceccarelli, M., et al. (2018). Detection of long non-coding RNA homology, a comparative study on alignment and alignment-free metrics. *BMC Bioinform.* 19:407. doi: 10.1186/s12859-018-2441-6
- Paytuyi-Gallart, A., Sanseverino, W., and Aiese Cigliano, R. (2019). “A Walkthrough to the Use of GreenC: The Plant lncRNA Database,” in *Plant Long Non-Coding RNAs: Methods and Protocols*, eds J. A. Chekanova and H.-L. V. Wang (New York, NY: Springer), doi: 10.1007/978-1-4939-9045-0\_25
- Peng, L., Peng, M., Liao, B., Huang, G., Li, W., and Xie, D. (2018). The advances and challenges of deep learning application in biological big data processing. *Curr. Bioinform.* 13, 352–359.
- Peng, L., Peng, M., Liao, B., Xiao, Q., Liu, W., Huang, G., et al. (2017). A novel information fusion strategy based on a regularized framework for identifying disease-related microRNAs. *RSC Adv.* 7, 44447–44455.
- Qazi, S. R., Ahmad, S., and Shakeel, S. N. (2020). HSEAT: a tool for plant heat shock element analysis, motif identification and analysis. *Curr. Bioinform.* 15, 196–203. doi: 10.1007/s00438-011-0638-8
- Rahaman, M., Komanapalli, J., Mukherjee, M., Byram, P. K., Sahoo, S., and Chakravorty, N. (2021). Decrypting the role of predicted SARS-CoV-2 miRNAs in COVID-19 pathogenesis: A bioinformatics approach. *Comp. Biol. Med.* 136:104669. doi: 10.1016/j.compbiomed.2021.104669
- Rehmsmeier, M. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10:604. doi: 10.1261/rna.5248604
- Shen, Z., Lin, Y., and Zou, Q. (2020). Transcription factors–DNA interactions in rice: identification and verification. *Brief. Bioinform.* 21, 946–956. doi: 10.1093/bib/bbz045
- Singh, G., Swain, A. C., and Mallick, B. (2021). Delineating characteristic sequence and structural features of precursor and mature Piwi-interacting RNAs of epithelial ovarian cancer. *Curr. Bioinform.* 16, 541–552.
- Song, B., Li, Z., Lin, X., Wang, J., Wang, T., and Fu, X. (2021). Pretraining model for biological sequence data. *Brief. Funct. Genom.* 20, 181–195. doi: 10.1093/bfpg/elab025
- Stalio, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 2020:3159. doi: 10.1038/s41580-020-00315-9
- Stower, H. (2020). An miRNA linked to metabolic disease. *Nat. Med.* 26, 1677. doi: 10.1038/s41591-020-1137-4
- Szcześniak, M. W., Bryzghalov, O., Ciomborowska-Basheer, J., and Makołowska, I. (2019). “CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs,” in *Plant Long Non-Coding RNAs: Methods and Protocols*, eds J. A. Chekanova and H.-L. V. Wang (New York, NY: Springer), doi: 10.1007/978-1-4939-9045-0\_26
- Tan, X., Chen, W.-B., Lv, D.-J., Yang, T.-W., Wu, K.-H., Zou, L.-B., et al. (2021). lncRNA SNHG1 and RNA binding protein hnRNPL form a complex and coregulate CDH1 to boost the growth and metastasis of prostate cancer. *Cell Death Dis.* 12:138. doi: 10.1038/s41419-021-03413-4
- Teng, X., Chen, X., Xue, H., Tang, Y., Zhang, P., Kang, Q., et al. (2020). NPInter v4.0: an integrated database of ncRNA interactions. *Nucleic Acids Res.* 48, D160–D165. doi: 10.1093/nar/gkz969
- Tian, Y., Ma, R., Sun, Y., Liu, H., Zhang, H., Sun, Y., et al. (2020). Correction: SP1-activated long noncoding RNA lncRNA GCMA functions as a competing endogenous RNA to promote tumor metastasis by sponging miR-124 and miR-34a in gastric cancer. *Oncogene* 39, 6621–6621. doi: 10.1038/s41388-020-1377-2
- Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11, 333–337. doi: 10.1038/nmeth.2810
- Wang, Y., Zhou, M., Zou, Q., and Xu, L. (2021). Machine learning for phytopathology: from the molecular scale towards the network scale. *Brief. Bioinform.* 22:bbab037. doi: 10.1093/bib/bbab037
- Xie, G., Jiang, J., and Sun, Y. (2020). LDA-LNSUBRW: lncRNA-disease association prediction based on linear neighborhood similarity and unbalanced bi-random walk. *IEEE/ACM Trans. Comp. Biol. Bioinform.* 2020:595. doi: 10.1109/TCBB.2020.3020595
- Zang, F., and Zhang, J.-S. (2012). Label propagation through sparse neighborhood and its applications. *Neurocomputing* 97:17. doi: 10.1016/j.neucom.2012.03.017
- Zeng, X., Liao, Y., Liu, Y., and Zou, Q. (2017). Prediction and Validation of Disease Genes Using HeteSim Scores. *IEEE/ACM Trans. Comp. Biol. Bioinform.* 14, 687–695. doi: 10.1109/tcbb.2016.2520947
- Zeng, X., Liu, L., Lu, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34, 2425–2432. doi: 10.1093/bioinformatics/bty112
- Zeng, X., Zhong, Y., Lin, W., and Zou, Q. (2019). Predicting disease-associated circular RNAs using deep forests combined with positive-unlabeled learning methods. *Brief. Bioinform.* 2019:80. doi: 10.1093/bib/bbz080
- Zhang, C. L., Zhang, Y. Z., Wang, B., and Chen, P. (2021). Prediction of Drug-target Binding Affinity by An Ensemble Learning System with Network Fusion Information. *Current Bioinformatics* 16, 1223–1235.
- Zhang, C., Bi, J., Liu, C., and Chen, K. (2016). A parameter-free label propagation algorithm for person identification in stereo videos. *Neurocomputing* 218:69. doi: 10.1016/j.neucom.2016.08.069
- Zhang, P., Meng, J., Luan, Y., and Liu, C. (2020). Plant miRNA–lncRNA Interaction Prediction with the Ensemble of CNN and IndRNN. *Interdiscip. Sci. Comp. Life Sci.* 12, 82–89. doi: 10.1007/s12539-019-00351-w
- Zhang, P., Wei, Z., Che, C., and Jin, B. (2022). DeepMGT-DTI: Transformer network incorporating multilayer graph information for Drug–Target interaction prediction. *Comp. Biol. Med.* 2022:105214. doi: 10.1016/j.compbiomed.2022.105214
- Zhang, W., Jing, K., Huang, F., Chen, Y., Li, B., Li, J., et al. (2019a). SFLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inform. Sci.* 497, 189–201. doi: 10.1016/j.ins.2019.05.017
- Zhang, W., Li, Z., Guo, W., Yang, W., and Huang, F. (2021). A fast linear neighborhood similarity-based network link inference method to predict MicroRNA–Disease Associations. *IEEE/ACM Trans. Comp. Biol. Bioinform.* 18, 405–415. doi: 10.1109/TCBB.2019.2931546
- Zhang, W., Qu, Q., Zhang, Y., and Wang, W. (2018a). 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., Tang, G., Wang, S., Chen, Y., Zhou, S., and Li, X. (2018b). “Sequence-derived linear neighborhood propagation method for predicting lncRNA–miRNA interactions,” in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Piscataway. doi: 10.1109/BIBM.2018.8621184
- Zhang, W., Tang, G., Zhou, S., and Niu, Y. (2019b). lncRNA–miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination. *BMC Genomics* 20:946. doi: 10.1186/s12864-019-6284-y
- Zhang, Y., Duan, G., Yan, C., Yi, H., Wu, F.-X., and Wang, J. (2021). MDAPlatform: a component-based platform for constructing and assessing miRNA–disease association prediction methods. *Curr. Bioinform.* 16, 710–721. doi: 10.2174/1574893616999210120181506
- Zhao, Z., Zhang, C., Li, M., Yu, X., Liu, H., Chen, Q., et al. (2020). Integrative Analysis of miRNA-mediated competing endogenous RNA network reveals the lncRNAs–mRNAs interaction in glioblastoma stem cell differentiation.



- Curr. Bioinf.* 15, 1187–1196. doi: 10.2174/1574893615999200511074226
- Zhou, L., Wang, Z., Tian, X., and Peng, L. (2021). LPI-deepGBDT: a multiple-layer deep framework based on gradient boosting decision trees for lncRNA-protein interaction identification. *BMC Bioinf.* 22:1–24. doi: 10.1186/s12859-021-04399-8
- Zhu, J., Zhang, X., Gao, W., Hu, H., Wang, X., and Hao, D. (2019). lncRNA/circRNA-miRNA-mRNA ceRNA network in lumbar intervertebral disc degeneration. *Mol. Med. Rep.* 20, 3160–3174. doi: 10.3892/mmr.2019.10569
- Zhu, Q., Fan, Y., and Pan, X. (2021). Fusing multiple biological networks to effectively predict miRNA-disease Associations. *Curr. Bioinform.* 16, 371–384.
- Zhu, X., Zhu, Y., Tan, Y., Chen, Z., and Wang, L. (2022). An iterative method for predicting essential proteins based on multifeature fusion and linear neighborhood similarity. *Front. Aging Neurosci.* 2022:7995. doi: 10.3389/fnagi.2021.799500

**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.

Copyright © 2022 Cai, Gao, Ren, Fu, Xu, Wang and Chen. 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.