A Survey of Computational Methods and Databases for lncRNA-MiRNA Interaction Prediction

Nan Sheng, Lan Huang, Ling Gao, Yangkun Cao, Xuping Xie, and Yan Wang

***Abstract*—Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are two prevalent non-coding RNAs in current research. They play critical regulatory roles in the life processes of animals and plants. Studies have shown that lncRNAs can interact with miRNAs to participate in post-transcriptional regulatory processes, mainly involved in regulating cancer development, metastatic progression, and drug resistance. Additionally, these interactions have significant effects on plant growth, development, and responses to biotic and abiotic stresses. Deciphering the potential relationships between lncRNAs and miRNAs may provide new insights into our understanding of the biological functions of lncRNAs and miRNAs, and the pathogenesis of complex diseases. In contrast, gathering information on lncRNA-miRNA interactions (LMIs) through biological experiments is expensive and time-consuming. With the accumulation of multi-omics data, computational models are extremely attractive in systematically exploring potential LMIs. To the best of our knowledge, this is the first comprehensive review of computational methods for identifying LMIs. Specifically, we first summarized the available public databases for predicting animal and plant LMIs. Second, we comprehensively reviewed the computational methods for predicting LMIs and classified them into two categories, including network-based methods and sequence-based methods. Third, we analyzed the standard evaluation methods and metrics used in LMI prediction. Finally, we pointed out some problems in the current study and discuss future research directions. Relevant databases and the latest advances in LMI prediction are summarized in a GitHub repository https://github.com/sheng-n/lncRNA-miRNA-interaction-methods, and we'll keep it updated.**

***Index Terms*—Computational methods, lncRNA-miRNA interaction prediction, network-based methods, sequence-based methods.**

I. Introduction

In THE human genome, approximately 75% is transcribed, but less than 3% encodes proteins [1], [2], [3]. In the early days, these non-coding RNAs (ncRNAs) are considered "noise" in coding regions [4]. However, a growing number of studies have found that ncRNAs play vital roles in many physiological processes, including cell proliferation and development, differentiation and apoptosis, and pathological manifestations such as disease manifestation and tumorigenesis [5], [6]. In addition, according to the transcription length, ncRNAs can be divided into small RNAs and long non-coding RNAs (lncRNAs), where the length of small RNAs is less than 200 nucleotides (nt), and common members include miRNAs, piwiRNAs (piRNAs), small interfering RNAs (siRNAs), and small nucleolar RNAs (snoRNAs) [7], [8]. It has been found that lncRNAs are participate in a variety of regulatory functions by interacting with different kinds of biomolecules, such as lncRNA-miRNA interaction and lncRNA-protein interaction [9], [10]. Here, we focus on the interactions between lncRNAs and miRNAs. The biological functional relationship between lncRNA and miRNA can be briefly summarized as miRNA sponge, competitively binding to miRNA target genes, miRNA production by lncRNA, and lncRNA degradation [11].

In recent years, many studies have been devoted to investigating the molecular mechanisms of lncRNAs and miRNAs, as well as their interactions, which will provide new insights into the mechanisms of disease development [12]. It is well known that biological experiments are essential and effective methods to accurately explore different molecules' biological functions and disease development mechanisms. There is growing evidence that LMIs lead to complex regulatory networks controlling gene expression at the transcriptional, post-transcriptional, and post-translational levels [13]. For example, Tsang et al. found that in human hepatocellular carcinoma, lncRNA HOTTIP is a novel oncogenic lncRNA that is negatively regulated by miR-125b [14]. He et al. found that in Epstein-Barr virus (EBV)-associated cancer tissues and cells, miRNA miR-BART6-3p can inhibit tumor cell migration and invasion by targeting and downregulating lncRNA LOC553103, which provides a potential new diagnostic and therapeutic marker for EBV-associated diseases [15]. Experiments have demonstrated that lncRNA HOTAIR plays an essential role in the development of many human cancers and can interact with miRNAs to influence cancer development [16], including gastric cancer [17], colon cancer [18], liver cancer [19], lung cancer [20], and pancreatic cancer [21]. In addition, Bian et al. summarized the LMIs in liver fibrosis [22].

Numerous studies have shown that interactions between miRNAs and lncRNAs also affect plant life activities. These

Manuscript received 15 November 2022; revised 20 February 2023; accepted 30 March 2023. Date of publication 4 April 2023; date of current version 9 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62072212, in part by the Development Project of Jilin Province of China under Grants 20220508125RC and 20200201290JC, in part by National Key R&D Program under Grant 2018YFC2001302, and in part by the Jilin Provincial Key Laboratory of Big Data Intelligent Cognition under Grant 20210504003GH. *(Corresponding authors: Lan Huang; Yan Wang.)*

Nan Sheng, Lan Huang, Ling Gao, Xuping Xie, and Yan Wang are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: shengnan21@mails.jlu.edu.cn; huanglan@jlu.edu.cn; gaoling22@mails.jlu.edu.cn; xiexp21@mails.jlu.edu.cn; wy6868@jlu.edu.cn).

Yangkun Cao is with the School of Artificial Intelligence, Jilin University, Changchun 130012, China (e-mail: caoyk20@mails.jlu.edu.cn).

Digital Object Identifier 10.1109/TCBB.2023.3264254

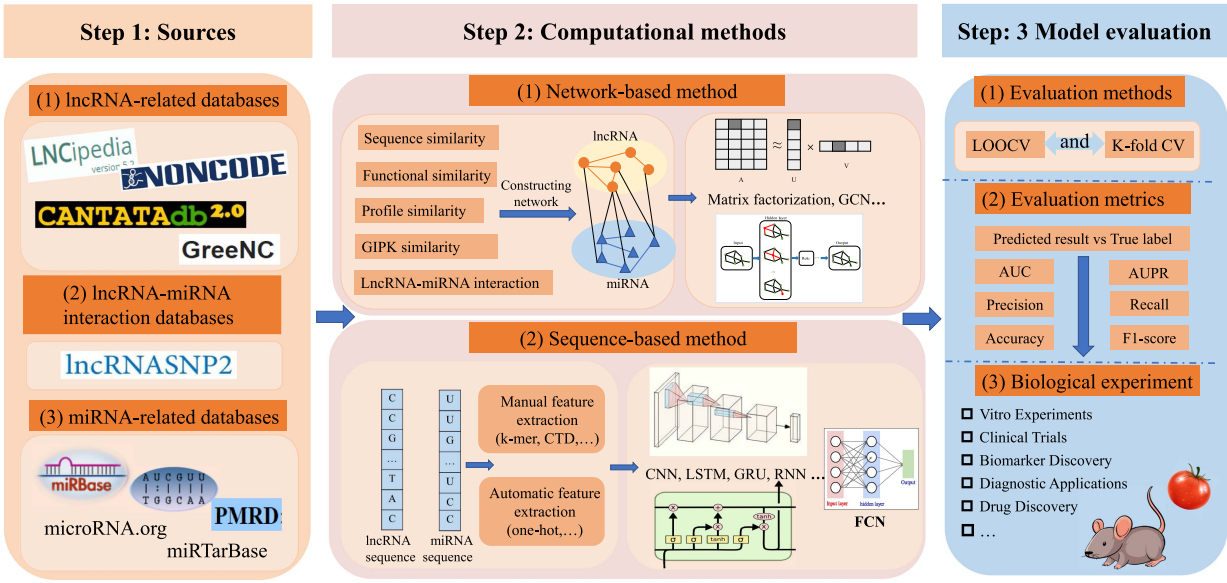

Fig. 1. A diagram illustrating the workflow of LMI prediction performing computational methods. Step 1: Collect lncRNA- and miRNA- related databases, as well as LMI databases. Step 2: Develop network-based and sequence-based computational methods, the network-based methods require the use of multiple lncRNA (miRNA) similarities to construct lncRNA-miRNA bipartite network, and the sequence-based methods only require manual or automatic feature extraction from lncRNA (miRNA) sequences. Step 3: Evaluation methods and evaluation metrics are employed to assess the performance, and validate candidate LMIs applying biological experiments.

TABLE I
THE WIDELY USED DATABASES IN LMI PREDICTION

Databases	Latest version	Description	Link
lncRNASNP [41]	lncRNASNP (v3)	It contains SNPs in lncRNAs, SNP effects on lncRNA structure, a mutation in lncRNAs, and LMIs	http://bioinfo.life.hust.edu.cn/lncRNASNP/
LNCipedia [42]	LNCipedia (v5.2)	It provides lncRNA sequence and annotation. The current release contains 127802 transcripts and 56946 genes	https://lncipedia.org/
miRBase [43]	miRbase (v22.1)	It records published miRNA sequences and annotation, involving 38589 miRNAs entries	https://mirbase.org/
NONCODE [44]	NONCODE (v6.0)	An integrated knowledge database dedicated to the ncRNAs database. It collects lncRNA expression profiles and putative functional annotations of lncRNA	http://www.noncode.org/
miRTarBase [46]	Update 2022	It presents experimentally validated miRNA-target interactions	https://miRTarBase.cuhk.edu.cn/
CANTATAdb [47]	CANTATAdb (2.0)	It records 239631 lncRNAs predicted in 36 plant species and 3 algae, and presents lncRNA sequences, genomic locations	http://yeti.amu.edu.pl/CANTATA/
PMRD [48]	Updated 2014	A plant miRNA database, including miRNA sequences and their target genes, secondary, dimension structure, expression profiling, etc.	http://bioinformatics.cau.edu.cn/PMRD/
GreeNC [49]	GreeNC (v2.0)	A plant lncRNAs database, recording lncRNA sequences, genomic coordinates, coding potential, and folding energy	http://greenc.sequentiabiotech.com/wiki2/
RNAhybrid [50]	RNAhybrid (2.1.2)	It is a tool for finding the minimum free energy hybridization of a long and a short RNA, and is primarily meant as a means for microRNA target prediction	https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid
psRNATarget [51]	psRNATarget (v2)	It is a plant small RNA target analysis server, and is used as the LMI prediction tool	https://www.zhaolab.org/psRNATarget/

interactions are crucial in plant growth, development, and response to biotic and abiotic stresses [23]. For example, in pigeonpea, lncRNA_1231 is able to isolate miR-156b during flowering, leading to increased expression of the flower-specific SPL-12 transcription factor and regulating flower development [24]. Hou et al. showed that lncRNA39026 binds miR168a and inhibits its function, thereby increasing the resistance of tomato to phytophthora infestans [25]. In addition, lncRNA42705 and lncRNA08711 are able to increase mRNA levels of MYB genes by acting as decoys for miR159, thereby enhancing resistance to phytophthora infestans [26]. Specifically, Lu et al. showed that lncRNAs enhance cold resistance in winter wheat by competitively binding miR398 [27]. Water deficit is an abiotic stress, and Chen et al. found that lncRNA TCONS_00021861 can affect drought resistance tolerance in rice by sponging miR528-3p [28]. Biswas et al. summarized the co-regulatory functions of lncRNA and miRNA in biotic and abiotic stresses in dicotyledons [29].

However, many of the laboratory experiments related to lncRNAs and miRNAs are undoubtedly expensive, complex, and time-consuming, which are often difficult to use for certain complex organisms, especially humans. Thousands of lncRNAs

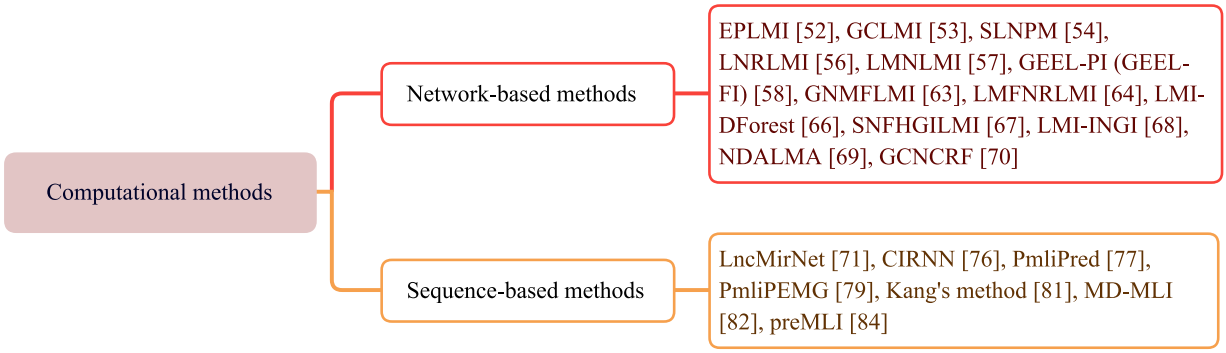

Fig. 2. Taxonomy of computational methods for LMI prediction.

and miRNAs have been identified in animals and plants, which are difficult to verify by large-scale biological experiments. There is an urgent need for computational approaches to reveal the characteristics of lncRNAs and miRNAs and guide those expensive and laborious laboratory experiments.

The rapid development and success of computational LMI prediction methods can be attributed to the following two aspects. On the one hand, with the accumulation of large amounts of high-throughput data, great opportunities have been created for the development of computational methods to mine the unobserved relationships between lncRNAs and miRNAs. On the other hand, the advancement of computer science provides an important basis for developing practical LMI prediction algorithms. Therefore, computational methods are expected to help effectively screen potential LMI candidates.

As mentioned above, computational methods are cost-effective strategies in biomedical research. In recent years, with the development of techniques like machine learning and deep learning, various computational models have been proposed for application in many studies related to different biomolecules or diseases, such as gene-disease association prediction [30], lncRNA-disease association prediction [31], [32], [33], miRNA-disease association prediction [34], [35], lncRNA-protein interaction prediction [36], [37], [38] and drug repositioning [39], [40]. In particular, the past few years have witnessed a surge in research on computational methods for LMI prediction. Many methods, databases, and applications have been proposed in the literature, and a comprehensive survey is needed to focus on this burgeoning new direction. However, to our knowledge, there is no comprehensive survey so far specifically regarding the investigation of lncRNA- and miRNA-related databases and computational methods for LMI prediction. Considering the importance of lncRNAs and miRNAs in biomedical research, a comprehensive review of studies on LMI prediction is scientifically attractive. As shown in Fig. 1, a flowchart of LMI prediction based on the computational approach is presented.

This review fills this gap by investigating computational methods for LMI prediction. We summarize the main contributions of this work as follows:

- We presented databases related to animal and plant LMI prediction, covering LMI data, and lncRNA- and miRNA-related data (such as expression profiles, sequences, and functions). These databases are widely used in computational methods and are still being updated and accessed.
- We reviewed 20 computational methods for LMI prediction that were divided into two groups, including network-based methods and sequence-based methods.
- We surveyed the commonly used evaluation methods and metrics in LMI prediction. These can help researchers effectively assess and verify the prediction ability of their developed methods in future studies.
- We further discussed the developmental challenges in LMI prediction and outlined several future research directions.

II. Databases

With the explosion of genomics, transcriptomics, and proteomics data, there is a tremendous opportunity for computational methods to discover new candidate LMIs. Researchers have established many publicly available databases or tools to store and provide data sources related to lncRNAs and miRNAs, which makes it possible for computational scientists to develop computational methods. For example, lncRNASNP v3 is a database that records a comprehensive resource of single nucleotide polymorphisms (SNPs) in lncRNA for 8 eukaryotic species, covering human, pig, mouse, etc. [41]. Specifically, this database records experimentally confirmed LMIs, which provides an accurate label for computational models. Different data types can be utilized as lncRNA and miRNA features, such as nucleotide sequence, expression profiles, target genes, and putative functional annotations. LNCipedia [42] and miRbase [43] are public databases for lncRNA and miRNA sequences and annotations, respectively. NONCODE is a comprehensive knowledge database dedicated to non-coding RNAs (excluding tRNAs and rRNAs), recording lncRNA expression profiles and putative functional annotations [44]. microRNA.org database provides a comprehensive resource for miRNA-target prediction and miRNA expression profiling [45]. miRTarBase is a comprehensive information source that stories experimental verification of miRNA-target genes interactions [46]. Unfortunately, the microRNA.org database are currently not accessible.

Some databases also specialize in recording plant lncRNA and plant miRNA information, such as CANTATAdb 2.0, the largest and most comprehensive plant lncRNA database available, containing more than 200000 lncRNA sequences

TABLE II
NETWORK-BASED METHODS FOR PREDICTING LMIS

Methods	Data types	Numbers[a]	Description	Model evaluation	Code
EPLMI [52]	LMI, lncRNA/miRNA expression profile, lncRNA/miRNA sequence, lncRNA putative functional annotations, miRNAs-target genes interaction	5348 LMIs (780 lncRNAs and 275 miRNAs)	A graph-based method based on two-way diffusion that uses expression profile similarity, sequence similarity, and functional similarity of lncRNA and miRNA	LOOCV, 5-cv (AUC)	https://github.com/TYLH/-EPLMI
GCLMI [53]	same as EPLMI	5348 LMIs (780 lncRNAs and 275 miRNAs)	An end-to-end prediction model using graph convolution autoencoder to infer LMIs	5-cv (AUC)	NA
SLNPM [54]	LMI, lncRNA/miRNA expression profile, lncRNA/miRNA sequence	Dataset1: 2272 LMIs (417 lncRNAs and 265 miRNAs), Dataset2: 2784 LMIs (642 lncRNAs and 275 miRNAs)	A sequence-derived linear neighborhood propagation method based on similarity integration to predict LMIs	5-cv (AUC, AUPR, Sen, Spe, Pre, Acc, F-measure)	NA
LNRLMI [56]	same as EPLMI	5348 LMIs (780 lncRNAs and 275 miRNAs)	An approach based on linear neighbor representation for prediction, which uses known interaction and expression profile similarity to construct a bipartite network	2-cv, 5-cv, 10-cv (AUC)	NA
LMNLMI [57]	same as EPLMI	5348 LMIs (780 lncRNAs and 275 miRNAs)	A framework based on network fusion technique and matrix completion technique for predicting interaction	5-cv (AUC, AUPR)	NA
GEEL-PI, GEEL-FI [58]	LMI, lncRNA/miRNA sequence	3784 LMIs (642 lncRNAs and 275 miRNAs)	A computational method based on graph embedding learning and ensemble learning	5-cv (AUC, AUPR, Acc, Rec, Spe, Pre, F-measure)	NA
GNMFLMI [63]	LMI, lncRNA/miRNA expression profile	8634 LMIs (468 lncRNAs and 265 miRNAs)	A graph regularized nonnegative matrix factorization for inferring interactions	5-cv (AUC, Spe, Sen, Pre, Acc, F1)	https://github.com/hai-chengyi/GNMFLMI
LMFNRLMI [64]	same as EPLMI	4966 LMIs (780 lncRNAs and 265 miRNAs)	A method based on logistic matrix factorization with neighborhood regularized for discovering interactions	LOOCV, 5-cv (AUC, AUPR, Pre, Rec, F1)	NA
LMI-DForest [66]	LMI, lncRNA/miRNA expression profile	18595 LMIs (3521 lncRNAs and 276 miRNAs)	A machine learning approach that combines deep forest and autoencoder	2-cv, 5-cv, 10-cv (AUC, Acc, Rec, Pre, F1)	NA
SNFHGILMI [67]	LMI, lncRNA/miRNA expression profile	4966 LMIs (780 lncRNAs and 275 miRNAs)	A heterogeneous graph inference method based on similarity network fusion to predict interactions	LOOCV, 5-cv (AUC, Sen, Acc, F1)	https://github.com/cj-DaSE/SNFHGILMI-master
LMI-INGI [68]	same as EPLMI	4966 LMIs (770 lncRNAs and 275 miRNAs)	A semi-supervised approach based on interactome work and graphlet interaction for prediction	5-cv (AUC, AUPR, Pre, Rec, F1)	https://github.com/Liu-Lab-Lnu/LMI-INGI
NDALMA [69]	LMI, lncRNA/miRNA sequence	Dataset1: 4966 LMIs (770 lncRNAs and 275 miRNAs), Dataset2: 15264 LMIs (1663 lncRNAs and 258 miRNAs)	A model based on network distance analysis that integrates sequence similarity and GIPK similarity of lncRNA and miRNA	5-cv (AUC, AUPR, Pre, Rec, F1)	https://github.com/Liu-Lab-Lnu/NDALMA
GCNCRF [70]	LMI, lncRNA/miRNA sequence	Dataset1: 15264 LMIs (1663 lncRNAs and 258 miRNAs), Dataset2: 4966 LMIs (770 lncRNAs and 275 miRNAs)	A method based on graph convolutional neural network and conditional random filed to infer potential interactions	5-cv (AUC, AUPR, ACC, Rec, Spe, F1)	https://github.com/zhaoqi106/GCNCRF

The "same as EPLMI" indicates that the same data type is used as in the computational method EPLMI. [a]The number of lncRNAs, miRNAs, and LMIs in methods. NA indicates that the source code is not available.

from 39 species [47]. PMRD integrates massive information of plant miRNA data, including miRNA sequences and their target genes, secondary structure, expression profiles, genome browser, etc [48]. GreeNC is the primary resource of plant lncRNAs with over 200000 annotated transcripts [49]. The database records information about the sequence, genomic coordinates, coding potential, and folding energy of all identified lncRNAs. Since there is no available public database of plant LMIs, RNAhybrid [50] and psRNATarget [51] are often employed as LMI prediction tools. Here, we briefly summarize the common databases involved in LMI prediction in Table I.

III. COMPUTATIONAL METHODS FOR LNCRNA-MIRNA INTERACTION PREDICTION

With the accumulation of various genomic data, computational approaches offer new opportunities for large-scale of screening potential LMIs. Great progress has been made in the research of lncRNAs and miRNAs, and a number of computational models have been developed to identify LMIs. As shown in Fig. 2, Tables II and III, based on the way of input, the computational methods for predicting LMIs are divided into two major groups: network-based methods and sequence-based methods.

A. Network-Based Methods

Specifically, network-based methods frequently first utilize lncRNA (miRNA) expression profile information, lncRNA (miRNA) sequence information, lncRNA (miRNA) function information, and known LMIs to calculate multiple similarities of lncRNAs and miRNAs. Then, a bipartite network is constructed by integrating the known LMIs, lncRNA similarity, and miRNA similarity. Finally, various machine learning models are

TABLE III
SEQUENCE-BASED METHODS FOR PREDICTING LMIS

Methods	Data types	Train dataset species	Test dataset species	Description	Evaluation metrics	Code
LncMirNet [71]	LMI, lncRNA/miRNA sequence	Human	Human	A method based on CNN, which uses lncRNA and miRNA sequence features, including k-mer, CTD, Doc2vec, and Role2vec features	Sen, Spe, Acc, F1, Mcc	https://github.com/-abcair/LncMirNet
CIRNN [76]	plant LMI, plant lncRNA/miRNA sequence	Zea mays	Zea mays	An ensemble deep-learning model based on CNN and IndRNN, which adopts plant lncRNA and miRNA sequence	Acc, Pre, Rec, F1	NA
PmliPred [77]	same as CIRNN	Arabidopsis thaliana, Glycine max, Medicago truncatula	Arabidopsis lyrate and Solanum lycopersicum	A framework based on deep learning model (CNN and BiGRU) and shallow machine learning model (RF) that uses sequence features and manually extracted features, involving k-mer frequency, GC content, number of base pairs, and minimum free energy. The hybridized based on fuzzy decision for prediction	AUC, Acc, F1	https://github.com/-kangzhai/PmliPred
PmliPEMG [79]	same as CIRNN	Arabidopsis thaliana, Glycine max, Oryza sativa and P. trichocarpa	Brachypodium distachyon, Medicago truncatula, Solanum tuberosum, Citrus sinensis, S. lycopersicum and T. aestivum	An ensemble deep learning model based on multi-level information enhancement (CNN and LSTM with attention mechanism) and greedy fuzzy decision for plant LMI prediction	AUC, Sen, Spe, Acc, F1	https://github.com/-kangzhai/PmliPEMG
Kangs method [81]	same as CIRNN	Arabidopsis thaliana, Brachypodium distachyon, Glycine max, Oryza sativa, Solanum tuberosum and Zea mays	Arabidopsis lyrata, Citrus sinensis, Malus domestica, Medicago truncatula, P. trichocarpa, Sorghum bicolor and T. aestivum	An adaptively prunes the base models based on dual-path parallel ensemble method to meet the challenge of cross-species prediction. it uses PmliPEMG as the base model	AUC, Sen, Spe, Acc, F1	https://github.com/-kangzhai/DPEP
MD-MLI [82]	same as CIRNN	Zea mays	Sorghum, Brachypodium distachyon, and Bryophyte	A hierarchical deep learning framework that integrates capsule network, an IndRNN with attention mechanism and Bi-LSTM	AUC, Acc, Sen, Spe, Mcc	NA
preMLI [84]	same as CIRNN	Arabidopsis thaliana,Soybean, Oryza sativa and Populus trichocarpa	Arabidopsis lyrata, Solanum lycopersicum, Medicago truncatula, Solanum tuberosum and Brachypodium distachyon	A model based on rna2vec pre-training and deep feature mining mechanism (CNN, BI-GRU, and attention layer)	AUC, AUPR, Sen, Spe, Acc, F1	https://github.com/-BioSequenceAnalysis/-preMLI

With the exception of LncMirNet, the remaining sequence-based methods have been performed to predict plant LMIs. They typically adopt different species samples in the training and test datasets to verify the generalization ability of the proposed methods for cross-species. The "same as CIRNN " indicates that the same data type is used as in the computational method CIRNN. NA indicates that the source code is not available.

performed to extract features from the bipartite network and infer LMI scores.

EPLMI

Based on the assumption that similar lncRNAs tend to have similar patterns of interaction or non-interaction with miRNAs, and vice versa. Huang et al. developed the first computational model, EPLMI, which was based on a two-way diffusion graph algorithm to predict novel LMIs [52]. The authors calculated three types of lncRNA (miRNA) similarities by integrating different information sources. The first type of similarity used the lncRNA (miRNA) expression profile data to calculate the similarity score of two lncRNAs (miRNAs) by performing Pearson correlation coefficient (PCC). The second type of RNA similarity was based on putative biological functions. Utilized the data of miRNA-target gene interactions and putative functional annotations of lncRNAs to measure the functional similarities of miRNA-miRNA pairs and lncRNA-lncRNA pairs, respectively. The third type of similarity was calculated by performing Needleman-Wunsch pairwise sequence alignment to calculate the sequence similarity of lncRNAs (miRNAs). A graph-based two-way diffusion approach was proposed to predict LMI scores. It was divided into three steps, first, two weighted LMI networks A^l and A^m were generated by integrating the known LMIs A, lncRNA (miRNA) similarity matrix $LS(MS)$, $A^l = LS \cdot A$ and $A^m = A \cdot MS$. The resource vectors of lncRNA (miRNA) R_{lncRNA} (R_{miRNA}) were calculated respectively based on the A^l and A^m. In this paper, we take the example of calculating the lncRNA resource vector, as follows:

$$R_{lncRNA_a} = \sum_{m=1}^{nm} \frac{A^w_{a,m} \cdot A_{*,m}}{\sum_{i=1}^{nl} A^w_{i,m}}, \tag{1}$$

where nl and nm are the number of lncRNA and miRNA, respectively. A^w is the A^l or A^m, and R_{lncRNA} describe the correlation scores during forward propagation. The resource vectors of lncRNA (miRNA) gained from A^l and A^m were averaged to computer the new resource vectors S_{lncRNA} (S_{miRNA}). Second, to obtain the correlation scores during back propagation, the resource vectors of lncRNA were computed based on S_{lncRNA}.

$$R'_{lncRNA_a} = \sum_{m=1}^{nm} \frac{A^w_{a,m} \cdot S_{miRNA_m}}{\sum_{i=1}^{nl} A^w_{i,m}}, \tag{2}$$

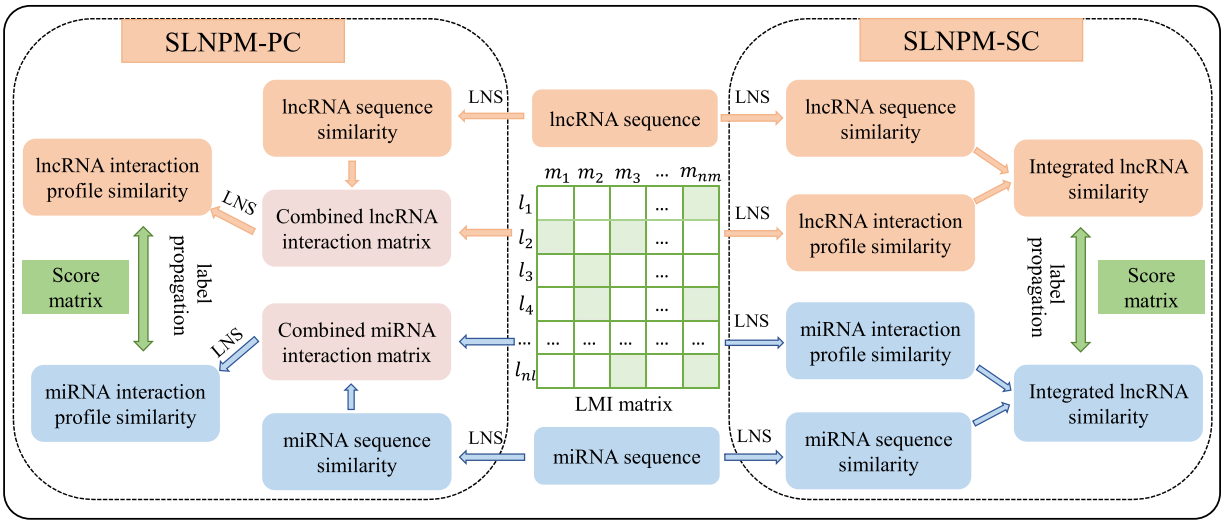

Fig. 3. The flowchart of SLNPM for predicting LMIs. SLNPM consists of two computational models: similarity-based information combination (SC) strategy (SLNPM-SC) and interaction profile-based information combination (PC) strategy (SLNPM-PC).

The two types of resource vectors were averaged to further combine into the resource vectors of the third step, namely S'_{lncRNA}. Similarly, the resource vector of miRNA S'_{miRNA} can be yielded. Third, the resource vectors for lncRNA and miRNA were combined into two $nl \times nm$ matrices, $score_{lncRNA}$ and $score_{miRNA}$, respectively.

$$score_{lncRNA} = \left[S'lncRNA_1^T, \ldots, S'lncRNA_{nl}^T\right]^T \quad (3)$$

$$score_{miRNA} = \left[S'miRNA_1, \ldots, S'miRNA_{nm}\right] \quad (4)$$

Finally, the final score matrix was calculated using the mean of $score_{miRNA}$ and $score_{miRNA}$, $score = (score_{lncRNA} + score_{miRNA})/2$.

GCLMI

Additionally, by integrating the attributes of lncRNAs (miRNAs) and interaction networks, Huang et al. proposed an end-to-end prediction model GCLMI that employs graph convolutional autoencoder [53]. GCLMI mainly consists of encoder layer and decoder layer. The feature matrixes of lncRNA and miRNA (F_l and F_m) and the known lncRNA-miRNA adjacency matrix LM were used as the input to the encoder. The lncRNA-miRNA feature matrix X was first constructed by combining F_l and F_m as follows:

$$X = \begin{bmatrix} F_l & 0 \\ 0 & F_m \end{bmatrix}. \quad (5)$$

The adjacency matrix A of the LMI network was expanded as follows:

$$A = \begin{bmatrix} 0 & LM \\ LM^T & 0 \end{bmatrix}. \quad (6)$$

In the encoder, the propagation rules between layers of the graph convolutional network are as follows:

$$H^{(l+1)} = Relu\left(D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}H^{(l)}W_e^{(l)}\right) = \begin{bmatrix} H_l \\ H_m \end{bmatrix}, \quad (7)$$

where $\tilde{A} = A + I$, and I is the identity matrix. D is the diagonal degree matrix of the matrix $\tilde{A}$, $H^{(l)}$ is the output matrix of the l-th layer, and $H^{(0)} = X$. $W_e^{(l)}$ is the trainable weight matrix of the l-th layer in the encoder. The output of the encoder layer had two components, i.e., the embedding feature matrix of lncRNA H_l; the embedding feature matrix of miRNA H_m. By introducing the trainable weight matrix W_d, the decoder layer was then constructed based on these separate matrices of the same original dimensions as follows:

$$M' = H_l W_d H_m^T. \quad (8)$$

The output matrix M' is the reconstructed LMI score matrix. Finally, the GCLMI model was trained in a semi-supervised manner using a negative sampling strategy.

SLNPM

Zhang et al. developed a new computational method SLNPM to screen potential LMIs (Fig. 3), which was based on sequence-derived linear neighborhood propagation [54]. In this method, based on lncRNA (miRNA) sequences and lncRNA (miRNA) expression profiles, the authors first adopted the linear neighborhood similarity measure method [55] to calculate lncRNA (miRNA) sequence similarity $S_{LSF}(S_{MSF})$ and lncRNA (miRNA) expression profile similarity $S_{LIP}(S_{MIP})$. Then, two similarity integration strategies were proposed, including similarity-based information combination (SC) and interaction profile-based information combination (PC). Here we employed the integration of lncRNA similarity as an example to

describe the SC strategy.

$$S_{lnc}(i,:) = \begin{cases} S_{LIP}(i,:) & \text{if lncRNA } L_i \text{ has interactions} \\ S_{LSF}(i,:) & \text{otherwise} \end{cases}. \tag{9}$$

Similarly, the integrated miRNA similarity S_m can also be computed. Unlike the SC strategy, for lncRNA L_i without interactions, their interaction profiles were complemented by sequence information before calculating the integrated lncRNA similarity in the PC strategy. The complementation process is explained as follows:

$$Y(i,:) = \frac{1}{Q_i} \sum_{i_k \in N(L_i)} S_{LSF}(i,i_k)\, Y(i_k,:), \tag{10}$$

$$Q_i = \sum_{i_k \in N(L_i)} S_{LSF}(i,i_k), \tag{11}$$

where $N(L_i)$ is the set of k lncRNAs that are most similar to that lncRNA L_i based on S_{LSF}, and each similar lncRNA has at least one interaction with miRNAs. Complementary miRNA interaction profiles can also be obtained in a similar manner. The integrated lncRNA and miRNA similarity matrices was computed based on the sequence similarity-complemented interaction profiles. Finally, the label propagation algorithm was applied on the SC and PC strategies to acquire the LMI score respectively, resulting in two versions of SLNPM: SLNPM-SC and SLNPM-PC. The propagation processes are defined as follows:

$$P_L = (1-\alpha)\left(I - \alpha S_{lnc}\right)^{-1} LM, \tag{12}$$

$$P_M = \left((1-\alpha)\left(I - \alpha S_m\right)^{-1} LM^T\right)^T, \tag{13}$$

where LM is the known LMI matrix and I is the identity matrix. The labels of nodes were updated with the labels of its neighbors with probability α, and the initial labels were kept with probability $1-\alpha$. $P_L(P_M)$ was the LMI score matrix, which was gained by applying label propagation algorithm in the lncRNA (miRNA) similarity graph. P_L and P_M were combined as the final interaction scores $P_{score} = \beta P_L + (1-\beta)P_M$, where β represented the weighting factor.

LNRLMI

Wong et al. presented a linear neighborhood representation method, LNRLMI, to identify lncRNA-related miRNAs [56]. Similarly, the model measured three types of lncRNA (miRNA) similarity utilizing lncRNA (miRNA) expression profiles, lncRNA (miRNA) functional annotations, and lncRNA (miRNA) sequences. First, the lncRNA (miRNA) network was constructed by connecting all lncRNA (miRNA) pairs whose similarities were greater than 0. Furthermore, the weights of the edges were set to their similarity. Then, according to the known LMIs, edges were added to connect lncRNA and miRNA nodes in the lncRNA and miRNA network to build the lncRNA-miRNA bipartite network M. It can be represented by the matrix as follows:

$$M = \begin{bmatrix} LS & LM \\ LM^T & MS \end{bmatrix}. \tag{14}$$

Then, the score matrix was defined as $S = MC$, where C was a weight matrix. C can be obtained by optimizing the corresponding objective function as follows:

$$\min_C \alpha \|M - MC\|_F^2 + \|C\|_F^2, \tag{15}$$

where α is set to balance the two factors, and $\|\cdot\|_F^2$ is the Frobenius norm. The final interaction score matrix can be obtained as follows:

$$S = MC^*, \tag{16}$$

where C^* is the optimized weight matrix, and the interaction probability of lncRNA and miRNA is calculated as LM' in S.

LMNLMI

Hu et al. built a novel computational approach, LMNLMI, based on fusing multiple information of lncRNAs and miRNAs to discover potential LMIs [57]. Like EPLMI, they first integrated several biological data sources, including lncRNA (miRNA) expression profiles, lncRNA putative functional annotations, miRNA-target gene interactions, and lncRNA (miRNA) sequences, to measure the three types of lncRNA and miRNA similarities. Then, considering different similar networks that are more or less interrelated and complementary to each other. A network fusion approach was utilized to obtain the lncRNA fusion network X and miRNA fusion network Y. Finally, the matrix completion method was employed to predict the probability scores of lncRNA-miRNA pairs. The interaction probability of each lncRNA-miRNA pair can be simply expressed as follows:

$$score(i,j) = x_i P y_j^T, \tag{17}$$

where P is the projection matrix to be learned. x_i and y_j are the i-th and j-th rows of matrices X and Y, respectively.

GEEL-PI and GEEL-FI

By integrating graph embedding and ensemble learning, Zhao et al. developed a model for predicting LMIs [58]. The authors first constructed the lncRNA-miRNA heterogeneous network by combining lncRNA sequence similarity SL, miRNA sequence similarity SM, and known LMI network LM, as shown in (14) above. Then, to take full advantage of the topological properties of the heterogeneous network, three categories of graph embedding methods were performed respectively, covering matrix factorization, random walk, and neural network. Among them, for the matrix factorization-based category, Laplace feature mapping (LE) [59], GraRep [60], and HOPE [61] were utilized; for the random walk-based category, DeepWalk [62] was chosen; and for the neural network-based category, the graph autoencoder was adopted. The model constructed a graph embedding ensemble learning method GEEL-PI, which was based on individual graph embedding methods to build the basic predictor, and further combine their predictions with an ensemble strategy to predict LMI. In this model, lncRNA and miRNA embedding representations, which were obtained from five embedding methods, were concatenated to construct lncRNA-miRNA pair representations. Five random forest-based classifiers were employed to predict node-pair interaction scores. Finally, Logistic

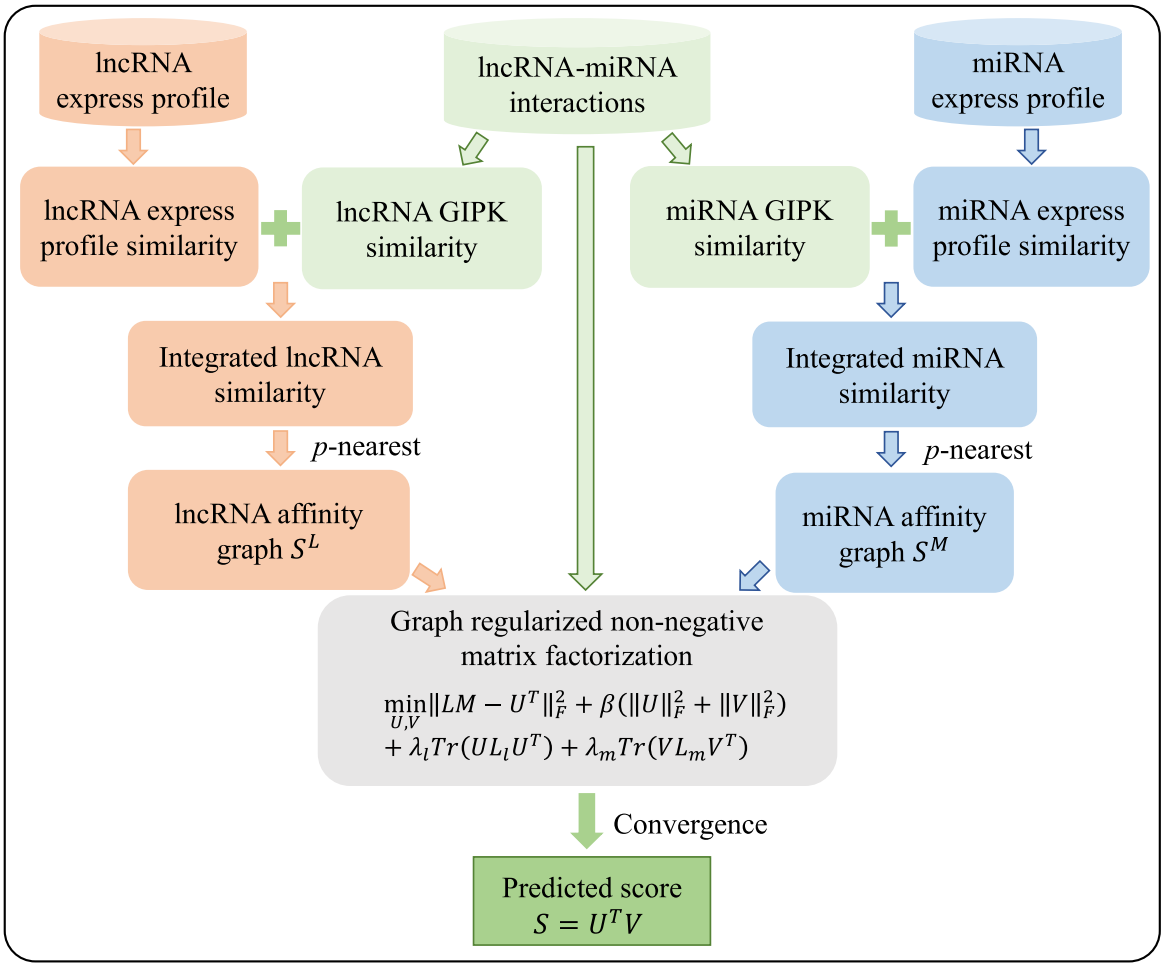

Fig. 4. The workflow of GNMFLMI based on graph regularized non-negative matrix factorization for inferring LMIs.

regression was utilized to map the outcome of multiple predictors to a label (LMI score).

In addition, this author also proposed a feature integration-based graph embedding ensemble learning model, GEEL-FI, which uses deep attention neural networks to learn lncRNA-miRNA pair representations. The deep neural network consists of two main parts: the attention layer and the deep fully connected layer (FCL). The attention layer was applied to integrate the five embedding representations of lncRNA and miRNA with the adaptively assigned weights. The FCN was employed to extract the preferred representations of node pairs and reveal potential interactions between lncRNAs and miRNAs. Binary cross entropy as the loss function:

$$\text{loss} = -\frac{1}{nl * nm}\sum_{i=1}^{nl}\sum_{j=1}^{nm}\left[p_{ij}\log\hat{p}_{ij} + (1-p_{ij})\log(1-\hat{p}_{ij})\right], \tag{18}$$

where nl and nm are the numbers of lncRNAs and miRNAs, respectively. p_{ij} denotes the true label, and $\hat{p}_{ij}$ represents the prediction label.

GNMFLMI

Wang et al. developed a new method based on graph regularized non-negative matrix factorization for LMIs prediction (Fig. 4), and they transformed the prediction task into a recommender system problem [63]. The integrated lncRNA and miRNA similarity matrix were first obtained by utilizing lncRNA Gaussian interaction profile kernel (GIPK) similarity, lncRNA expression profile similarity, miRNA GIPK similarity, and miRNA expression profile similarity. The p-nearest neighbor graph was then employed to acquire the affinity graphs S^L and S^M for lncRNA and miRNA from the similarity matrix. Finally, the authors devised a model based on non-negative matrix factorization to find two low-dimensional potential features U and V for calculating the LMI score with the following equation.

$$\min_{U,V}\left\|LM - U^TV\right\|_F^2 + \beta\left(\|U\|_F^2 + \|V\|_F^2\right) + \lambda_l Tr\left(UL_lU^T\right) + \lambda_m Tr\left(VL_mV^T\right), \tag{19}$$

where β is the sparseness constraint coefficient to adjust the sparsity of U and V. λ_l and λ_m are the graph regularization parameters. $L_l = D_l - S^L$ and $L_m = D_m - S^M$ denote the graph Laplacian matrices of S^L and S^M, respectively. D_l and

D_m are the diagonal matrixes. $\|\cdot\|_F^2$ is the Frobenius norm, and $Tr(\cdot)$ is the trace of the matrix. Updating the matrices U and V according to the above equations until convergence or reaching the upper limit of the iteration. The LMI score matrix can be calculated by $U^T V$.

LMFNRLMI

Liu et al. proposed a new computational model based on matrix factorization to reveal the potential interactions of lncRNAs with miRNAs [64]. In this model, the potential factor vectors of matrix factorization are represented as probability scores by logistic functions using known interaction information and neighbor similarity. First, the authors introduced the p-nearest neighbor to the similarity matrix and integrated it into the regularization items. Among the similarities include lncRNA (miRNA) sequence similarity, lncRNA (miRNA) expression profile similarity, and lncRNA (miRNA) function similarity. Then, the authors proposed a logistic matrix factorization based on neighborhood regularization to find the potential low-dimensional features matrixes U and V of lncRNA and miRNA. lncRNA and miRNA interaction score p_{ij} was measured by the following equation:

$$p_{ij} = \frac{\exp\left(u_i v_j^T\right)}{1 + \exp\left(u_i v_j^T\right)}, \tag{20}$$

where u_i and v_j denote the i-th and j-th rows of U and V.

LMI-DForest

Consider that DeepForest can efficiently leverage input features to generate differently grained feature vectors, and has highly competitive performance with deep neural networks in a wide range of tasks [65]. Wang et al. developed a deep learning framework to infer LMIs by integrating DeepForest and autoencoder model [66]. It consists of two main parts, first, the authors input lncRNA and miRNA expression profile features into autoencoder to extract potential low-dimensional representations of lncRNAs and miRNAs. Then, the low-dimensional representations were then fed into the random forest-based deep forest to predict potential LMIs.

SNFHGILMI

Fan et al. proposed a heterogeneous graph inference method based on similar network fusion to predict lncRNA-related miRNAs [67]. They first integrated sequence information and experimentally verified interactions to calculate the lncRNA (miRNA) sequence similarity $SL(SM)$, and lncRNA (miRNA) GIPK similarity $KL(KM)$. In this paper, two similarity fusion methods were proposed. One is a simple linear fusion as follows:

$$WL(i,j) = \alpha SL(i,j) + (1-\alpha)KL, \tag{21}$$

$$WM(i,j) = \alpha SL(i,j) + (1-\alpha)KM, \tag{22}$$

where $\alpha \in (0,1)$ is the weight parameter. The other is a nonlinear fusion method called similar network fusion (SNF). Here lncRNA is used as an example, which is expressed as follows:

$$WL = \left(SL_{\text{snf}} + KL_{\text{snf}}\right)/2, \tag{23}$$

$$WL = \left(WL + (WL)^T + I\right)/2, \tag{24}$$

where, SL_{snf} and KL_{snf} are lncRNA sequence similarity and lncRNA GIPK similarity obtained by SNF algorithm. I is the identity matrix. Similarity, integrated miRNA similarity WM can also be obtained. Combining the WL, WM and the known LMI matrix LM, the heterogeneous graph inference algorithm was adopted to predict the potential interaction scores.

$$W_{i+1} = \lambda WL \times W_i \times WM + (1-\lambda)W_0, \tag{25}$$

where $\lambda \in (0,1)$ is the decay factor. W_0 is the initial interaction adjacency matrix of lncRNA-miRNA. A new interaction matrix is generated by iteration, and when the difference between W_{i+1} and W_i is less than a certain threshold value, the matrix converges, and the LMI score matrix is obtained.

LMI-INGI

Zhang et al. developed a computational method to predict potential LMIs based on interactome network and graphlet interaction (Fig. 5) [68]. First, the authors also calculated three types of lncRNA (miRNA) similarities, including lncRNA (miRNA) sequence similarity, lncRNA (miRNA) expression profile similarity, and lncRNA (miRNA) functional similarity. Second, the authors constructed a lncRNA graph GL and a miRNA graph GM by using lncRNA and miRNA similarity values, respectively. Third, the number of graphlet interaction isomers between node i and node j in GL and GM was calculated. Here the graph GL was shown as an example.

$$N_{ij}\left(I_k\right) = \sum_{l \in V(GL)} \sum_{m \in V(GL)} b_{ij} b_{il} b_{jl} b_{im} b_{jm} b_{lm}, \tag{26}$$

$$b_{st} = \begin{cases} a_{st} & s \text{ and } t \text{ has a link in } I_k \\ 1 - a_{st} & s \text{ and } t \text{ has no link } I_k \end{cases}, \tag{27}$$

where $V(GL)$ is the set of nodes in GL, l and m are the other two nodes besides node i and node j. If there was a link from node s to node t, a_{st} was equal to the similarity value between the two nodes, otherwise $a_{st} = 0$. Similarly, the number of graphlet interaction isomers between two miRNAs in GM can be computed. Fourth, the LMI probabilities were calculated based on the graphs GL and GM, respectively. For example, for graph GL, the authors obtain the lncRNA-miRNA pair scores S_{lnc}, as follows:

$$S_{\text{lnc}}(i,j) = V_{lnc} X_{lnc}, \tag{28}$$

$$X_{lnc}(k,j) = \sum_{p \in p(i)} \frac{N_{pj}\left(I_k\right)}{\sum_{l \in L} N_{pl}\left(I_k\right)}, \tag{29}$$

where $P(i)$ is the set of lncRNAs known to be associated with miRNA, and L is the set of other lncRNAs. k is the type of graphlet interaction isomers. V_{lnc} is the weight matrix, which is measured based on training set.

$$V_{lnc} = S_{lnc} X_{lnc}^T \left(X_{lnc} X_{lnc}^T\right)^{-1} \tag{30}$$

Similarly, for the miRNA graph GM, the interaction scores S_m of the lncRNA-miRNA pairs can also be derived. Finally, the S_{lnc} and S_m were averaged as the final interaction scores of lncRNA-miRNA pairs $S = (S_{lnc} + S_m^T)/2$.

NDALMA

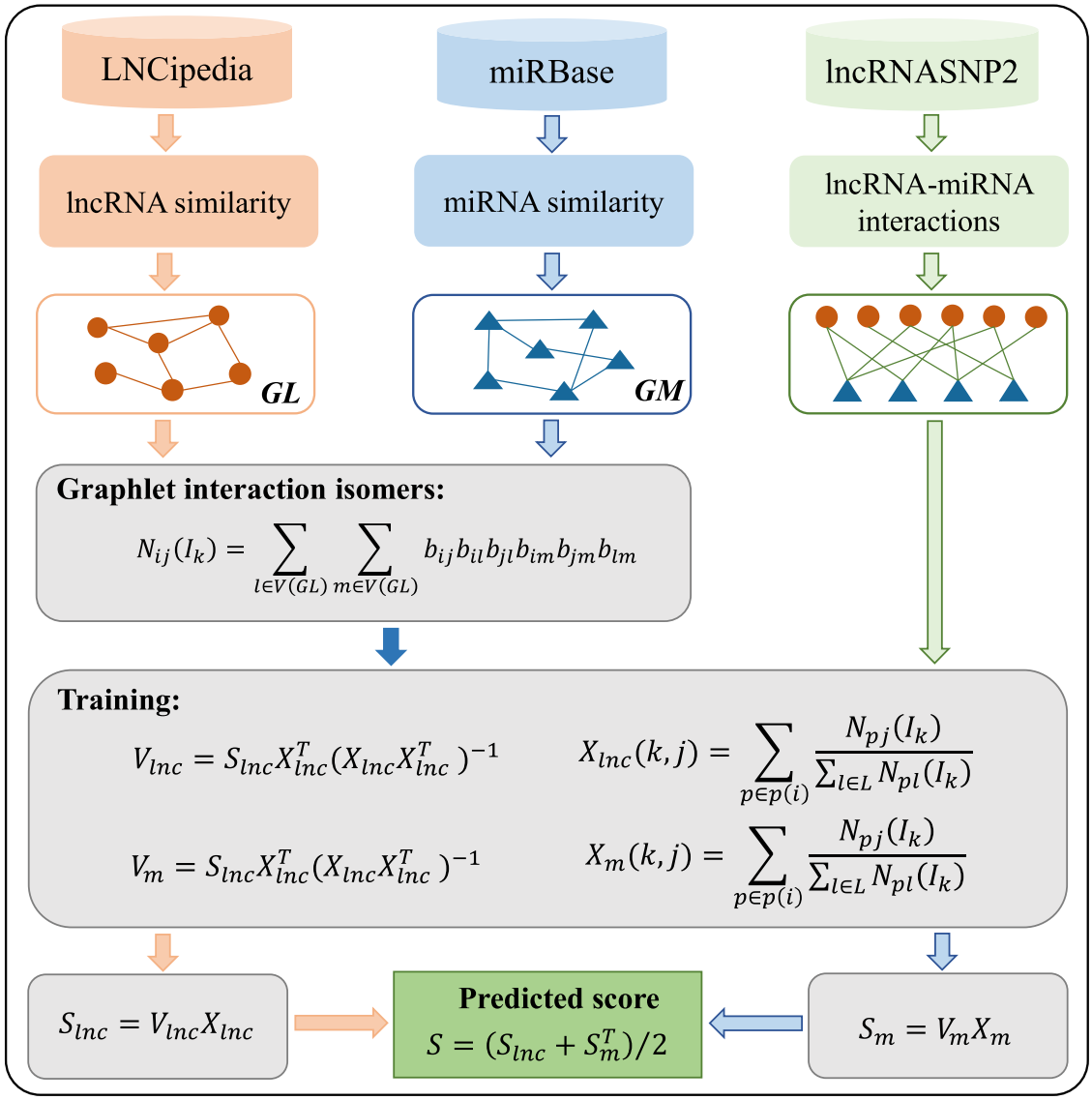

Fig. 5. The workflow of LMI-INGI based on interactome network and graphlet interaction for predicting LMIs.

Zhang et al. proposed a computational method to infer potential lncRNA-related miRNAs by applying network distance analysis [69]. In this paper, they first used lncRNA (miRNA) sequence similarity and lncRNA (miRNA) GIPK similarity to obtain the integrated lncRNA (miRNA) similarity matrix $LS(MS)$. Then, for the lncRNA similarity network, the network distance matrix of lncRNA was measured using the network distance algorithm, as follows:

$$Dl_{ij}^{adj} = \frac{Dl_{ij}}{\sqrt{\sigma_i \times \sigma_j}}, \tag{31}$$

where $Dl = 1/LS_i$. σ_i and σ_j are the average distance of the i-th and j-th lncRNA from all other lncRNAs in the original lncRNA network, respectively. Similarly, the network distance matrix of miRNA Dm_{ij}^{adj} can be derived. Finally, the confidence scores of lncRNAs and miRNAs were calculated as follows:

$$Cl_i = \frac{\sum_{j=1}^{N_l} Dl_{ij}^{adj}}{N_l} - \frac{\sum_{j=1}^{R_m} Dl_{ij}^{adj}}{R_m}, \tag{32}$$

where N_l is the number of lncRNA and R_m is the total number of lncRNAs known to interact with a given miRNA. Similarly, the confidence score Cm of miRNAs can be obtained. The score conversion was applied to get the interaction scores of lncRNAs and miRNAs. The final score is the average of the two conversion scores.

GCNCRF

Recently, Wang et al. proposed a GCN and conditional random field (CRF) based method, GCNCRF, to predict human LMIs [70]. The authors first calculated lncRNA (miRNA) sequence similarity SL (SM) of and the GIPK similarity KL (KM) of lncRNA(miRNA). Same as (21) and (22), a linear fusion approach was adopted to obtain the fused lncRNA (miRNA) similarity WL (WM). Given lncRNA similarity WL, miRNA

similarity WM and LMIs LM, the adjacency matrix of the LMI network is expressed as follows:

$$A = \begin{bmatrix} WL & LM \\ LM^T & WM \end{bmatrix}. \tag{33}$$

In addition, the authors utilized a random walk with restart in the lncRNA (miRNA) similarity matrix WL (WM) to extract the lncRNA (miRNA) features F_l (F_m). The feature matrix of the interaction network was defined as follows:

$$X = \begin{bmatrix} F_l & 0 \\ 0 & F_m \end{bmatrix}. \tag{34}$$

Given the adjacency matrix A and the feature matrix X of the LMI network, graph convolution encoder was performed to extract the topological features of the nodes. The propagation rules are defined as follows:

$$H^{(l+1)} = Relu\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W_e^{(l)}\right), \tag{35}$$

where D is the degree matrix of A, $W_e^{(l)}$ is the trainable weight matrix, and $H^{(0)} = X$. $Relu$ is the nonlinear activation function. To ensure that similar lncRNA/miRNA nodes have similar embeddings, the authors introduced a CRF layer with an attention mechanism to update the obtained preliminary embeddings. Finally, the authors acquired the lncRNA-miRNA score matrix by decoder layer.

$$A' = Q_l W_l (W_m)^T (Q_m)^T, \tag{36}$$

where W_l and W_m are trainable weight matrices, respectively. Q_l and Q_m are lncRNA and miRNA node embeddings learned in the CRF layer.

B. Sequence-Based Methods

In contrast to the network-based method, the sequence-based method only utilizes the lncRNA and miRNA sequence information. There are two ways to encode features in this type of method. The first way is to manually extract features from the lncRNA (miRNA) sequences by using k-mer, composition transition distribution (CTD), GC content, base pair number, etc. Then deep learning models like convolutional neural network (CNN) and Long Short-Term Memory (LSTM) are employed to extract deep features and predict the probability of LMIs. The second way is that lncRNA and miRNA sequences are encoded by one-hot encoding and fed into deep learning models such as CNN, LSTM, gated recurrent unit (GRU), etc., to extract deep features and infer interaction scores automatically.

LncMirNet

Considering that few computational methods are available for discovering potential interactions between lncRNA and miRNA based on sequence level, Yang et al. developed a computational method for predicting LMIs based on deep CNN (Fig. 6) [71]. The authors first extracted four types of features from lncRNA and miRNA sequences using four methods, including k-mer [72], composition transition distribution (CTD) [73], doc2vec [74], and role2vec [75]. Then, Histogram-Dd was performed to convert the four categories of feature vectors of lncRNA (miRNA) into corresponding matrices respectively, and integrates them into node feature matrix with a size of 20×20×4. Third, the feature matrix of lncRNA and miRNA were input into CNN to extract deep features, respectively. After crossing multiple CNN layers, respectively, the lncRNA tensor and miRNA tensor were combined and input into the FCLs to predict potential LMIs.

CIRNN

All of the above methods are widely employed for animal LMI prediction, and several approaches were proposed for plant LMI prediction. By integrating CNN and independently recurrent neural network (IndRNN), Zhang et al. developed a deep learning model, CIRNN, to infer plant LMIs [76]. CIRNN consists of CNN and IndRNN. First, the authors used the CNN layer to extract features from lncRNA and miRNA sequences automatically. Second, the features were sampled and processed by the pooling layer to obtain the most suitable features for classification. Third, the obtained feature was inputted into the IndRNN layer to further understand the dependencies between features. Finally, the FCLs were performed to discover plant LMIs.

PmliPred

Kang et al. proposed a model with deep learning and shallow machine learning to infer plant LMIs [77]. One of the deep learning models, CNN-BiGRU, consists of CNN and bidirectional gated recurrent unit (BiGRU). In this deep learning model, the raw sequence of lncRNA and miRNA after concatenation was encoded by one-hot encoding as input to CNN-BiGRU. CNN was employed to extract the feature map from the encoded sequences, and compress them into a one-dimensional vector by a flatten layer. The vector was fed into BiGRU, and the decision was output by the FCLs. For the machine learning model, the authors manually constructed the k-mer, GC content, number of base pairs, and minimum free energy (MFE) [78] of lncRNA and miRNA to form the feature vector as the input of random forest. Finally, inspired by fuzzy set theory, the trained CNN-BiGRU and random forest were hybridized based on fuzzy decisions to obtain the final LMI scores.

PmliPEMG and Kang's Method

Kang et al. also proposed an ensemble deep learning model by combining multi-level information enhancement and greedy fuzzy decision making to discover potential plant LMIs [79]. The authors first used k-mer and g-gap [80] to extract continuous and discontinuous sequence features from the raw sequences of lncRNAs (miRNAs). In addition, RNAfole was utilized to obtain secondary structure descriptions of lncRNAs and miRNAs. Statistical dots and brackets extracted continuous and discontinuous structural features of lncRNAs (miRNAs). Then, the base model ConvLSTM was built based on CNN and LSTM. A matrix Fcf was formed as the input of ConvLSTM by arranging vertically the values of each dimension of the complex feature vector of lncRNA and miRNA.

$$Fcf = \left[Kmer_{seq}, Ggap_{seq}, Kmer_{str}, Ggap_{str}\right], \tag{37}$$

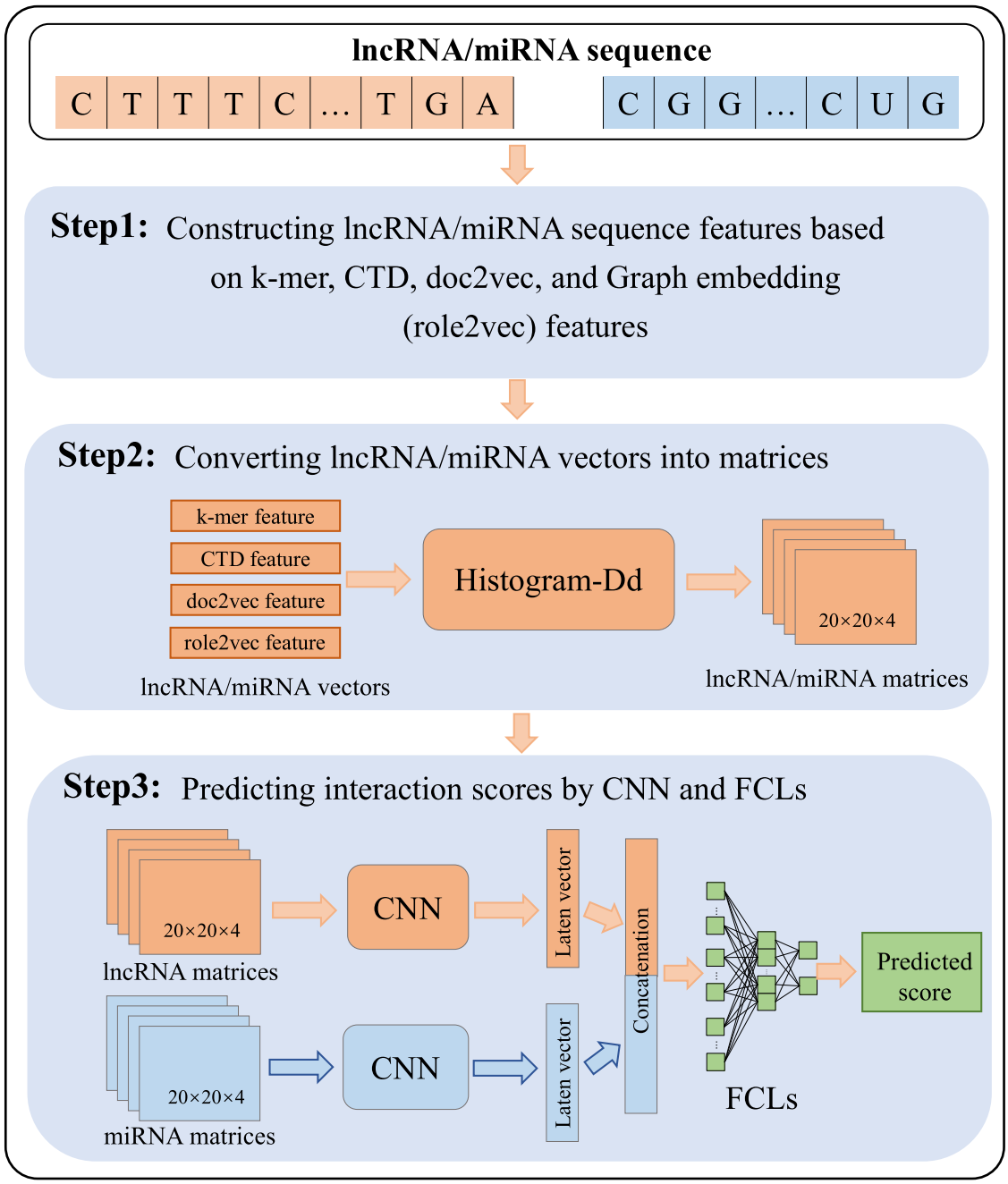

Fig. 6. The flowchart of LncMirNet based on deep convolutional neural network for LMI prediction.

where $Kmer_{seq}$ and $Ggap_{seq}$ are continuous and discontinuous sequence features of lncRNAs (miRNAs). $Kmer_{str}$ and $Ggap_{str}$ denote the continuous and discontinuous structural features of lncRNAs (miRNAs). In addition, the authors adopted an attention mechanism to assign weights to the output of the LSTM layer.

$$Fatt\left(y_d\right) = v^T \tanh\left(W y_d + B\right), \tag{38}$$

$$\omega_d = \mathrm{softmax}\left(Fatt\left(y_d\right)\right), \sum_{d=1}^{nd} \omega_d = 1, \tag{39}$$

$$A = \sum_{d=1}^{nd} \omega_d y_d, \tag{40}$$

where v denotes the parameter matrix, W and B are the weight matrix and bias vector, respectively. y_d is the d-th output of the LSTM layer, nd is the number of outputs of the LSTM layer. ω_d represents the weight of the d-th layer, A is the output after the attention mechanism enhancement. Finally, the FCLs were applied to predict the interaction probability for lncRNA-miRNA pairs. The model adopted fuzzy decision to integrate three basic

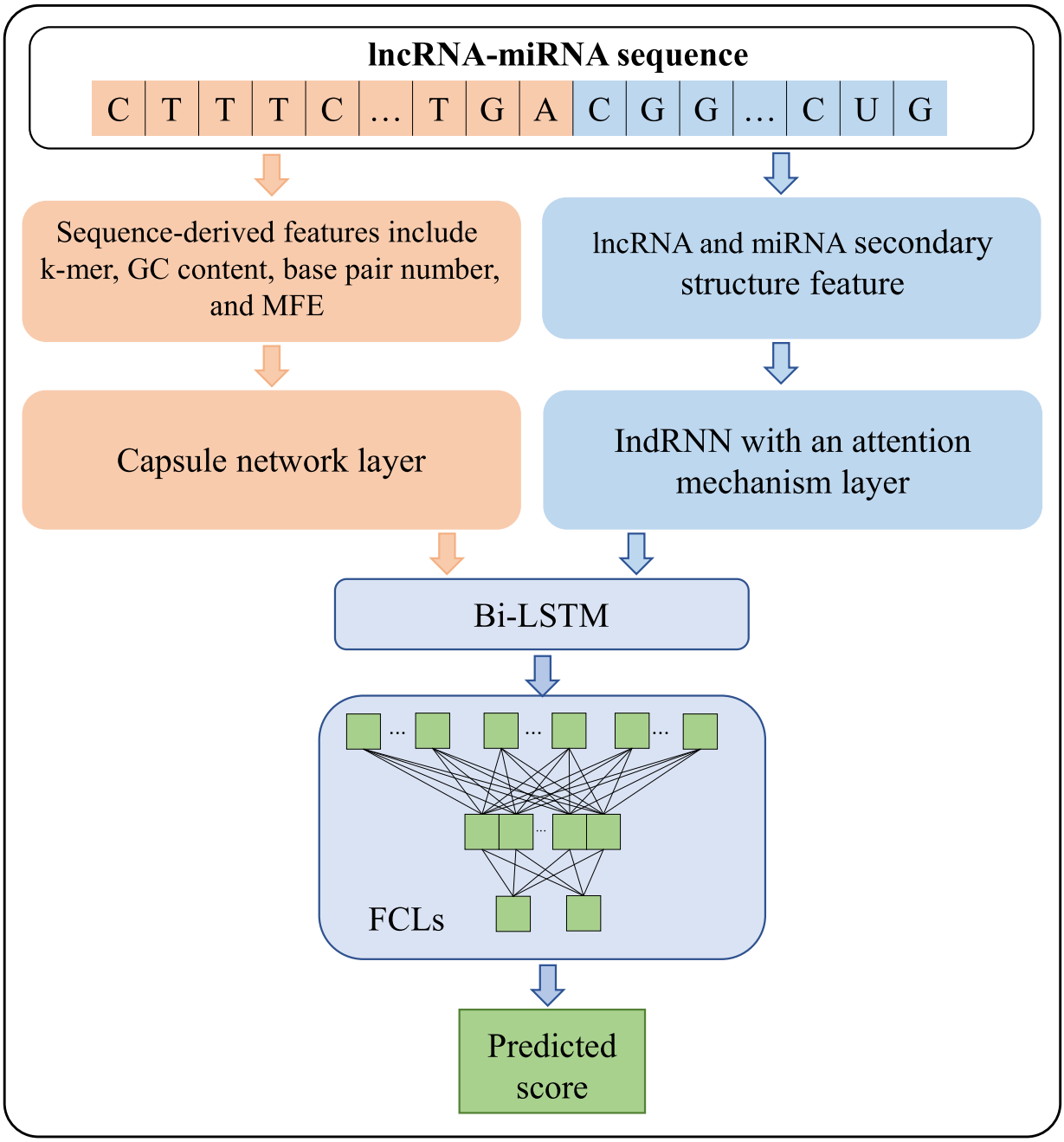

Fig. 7. The flowchart of MD-MLI based on capsule network, IndRNN with attention mechanism, and Bi-LSTM for discovering LMIs.

models and introduces greedy index to select the base model to improve efficiency and generalization ability.

Moreover, Kang et al. also proposed a dual-path parallel ensemble pruning-based approach to predict potential lncRNA-related miRNAs using PmliPEMG as the base model [81].

MD-MLI

A data-driven hierarchical deep learning framework, MD-MLI, was developed by Song et al. (Fig. 7) [82]. The framework can efficiently extract sequence-derived (intrinsic) and secondary structure features. MD-MLI consists of three components: capsule network, IndRNN with an attention mechanism, and Bi-LSTM. For the capsule network part, first, sequence-derived features of the lncRNA-miRNA pairs were folded to form sequence feature images, which were fed into 2D convolution layers to learn high-level features. The sequence-derived features of lncRNA and miRNA include k-mer frequency, GC content, base pair number, and Minimum Free Energy (MFE). Then the capsule network was then employed further to extract the high-level representation of the lncRNA-miRNA pairs. For the indRNN part, first, the universal expression of lncRNA and miRNA secondary structures were obtained using the bpRNA toolbox [83], respectively. Second, the two sequences were combined into a standard sequence, which denotes a matrix by using one-hot encoding as the input of indRNN.

The indRNN layer integrated the attention mechanism to learn the global position information of the secondary structure of the lncRNA-miRNA pair. Finally, the outputs of the two parts were fused into the Bi-LSTM network to learn the correlation between LMI sequence features. FCLs were performed to obtain classification results.

preMLI

Yu et al. proposed a deep learning model based on rna2vex pre-training and deep feature mining mechanisms for detecting potential interaction between lncRNAs and miRNAs [84]. The authors first trained a distributed representation of k-mer based on word2vec [85], which was a helpful word embedding training method consisting of a shallow two-layer neural network. Based on the pre-training tool rna2vec, low-dimensional and high-quality vectors can be acquired to represent k-mer words. Then, a hybrid structure of CNN and bi-directional gated recurrent unit (Bi-GRU) was proposed to learn the local features of lncRNA and miRNA, as well as to capture the long-term dependence of local features. In addition, an attention mechanism was added to adaptively integrate the features of different layers of Bi-GRU. Finally, lncRNA and miRNA vectors were cascaded as inputs to the FCLs to obtain interaction scores.

IV. Model Evaluation

To comprehensively evaluate the performance of computational methods, leave-one-out cross-validation (LOOCV), 5-fold cross-validation (5-cv), and 10-fold cross-validation (10-cv) are often utilized for performance evaluation. LncRNA-miRNA interaction prediction can be considered as a classification task, and some traditional machine learning metrics are often employed to measure the prediction performance of computational methods, including sensitivity (Sen), specificity (Spe), precision (Pre), recall (Rec), accuracy (Acc), F1-score (F1), the area under the receiver operating characteristic curve (AUC), the area under the precision-recall curve (AUPR), Matthews correlation coefficient (Mcc), etc., and whose formulas are defined as follows:

$$\text{Sen} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \quad \text{Spe} = \frac{\text{TN}}{\text{TN}+\text{FP}}, \tag{41}$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \quad \text{Rec} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{42}$$

$$\text{ACC} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}, \tag{43}$$

$$\text{F1} = \frac{2\ \text{Precesion} \times \text{Recall}}{\text{Precesion} + \text{Recall}}, \tag{44}$$

$$\text{Mcc} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP}+\text{FN}) \times (\text{TP}+\text{FP}) \times (\text{TN}+\text{FN}) \times (\text{TN}+\text{FP})}}, \tag{45}$$

where TP (FP) denotes true (false) positive samples and TN (FN) denotes true (false) negative samples. The receiver operating characteristic curve and the precision-recall curve are plotted based on sensitivity, specificity and precision, recall, respectively. In addition, ablation experiments and parameter sensitivity analyses are also used to further measure the capability of computational methods. Ablation experiments are conducted to assess the importance of a module in the proposed method, while parameter analyses are applied to explore the sensitivity of the parameters. Several methods also perform case studies to further assess the prediction power of the models [58], [63], [67], [69]. There are two main types of case studies, one of which utilizes the trained model to score all unlabeled lncRNA-miRNA pairs. The primary purpose of this way is to test the practical ability of the proposed method for inference of unknown LMIs. The second type is to validate the power of the model to predict relevant miRNAs for novel lncRNAs without any know interacted miRNAs. It usually requires removing all miRNAs that interact with a specific lncRNA in the training samples. Either way, the top-candidate interactions are chosen and then verified by using databases and literature. For plant LMI studies, they will select multiple test datasets of different plants to detect the cross-species prediction capabilities of the model.

V. Discussion and Future Directions

Though promising advances have been made in computational methods for predicting LMIs, each method has advantages and limitations, thus, researchers need to choose the appropriate computational method for their needs. In this section, we discuss some of the crucial problems with each type of method in the current study and outline several research directions for further investigation.

For network-based methods, the key is to construct a lncRNA-miRNA heterogeneous network, then use graph algorithms, such as matrix factorization and graph convolutional network, to extract features and predict scores. This category has the advantage of allowing for full learning of the network's structural information. Another advantage of the network-based method is that it is simple and reliable for understanding algorithm theory and computational operations. However, the models of this category usually depend on known LMIs and can suffer from possible prediction bias caused by imbalanced learning samples. For example, EPLMI requires to utilize known LMIs information for message passing. In addition, the data quality might impact the prediction performance for network-based algorithms, which frequently need to combine multiple data sources. However, sequence-based methods often outperform network-based methods by relying solely on the sequence information of lncRNAs and miRNAs. However, the success of this type of method lies in the effective extraction high-level features through different deep learning models, such as CNN, LSTM, and GRU. For example, LncMirNet performs CNN to learn non-linear features from sequence information of lncRNAs and miRNAs, and it produces accurate predictions. However, most non-linear features extracted by deep learning-based models are poorly interpreted. Moreover, these models typically have high time and space complexity. GCNCRF and preMLI are network-based and sequence-based methods, respectively, both of which have been proposed recently and have achieved the best prediction performance.

Data Quality. For network-based methods, expression profile information, sequence information, and functional information of lncRNAs and miRNAs are frequently combined to construct heterogeneous networks. While each of these resources has unique aspects and advantages, it is also critical to recognize that some of them tend to have high noise, incompleteness, and imprecision. Additionally, like PMRD database and microRNA.org database, are rarely updated or inaccessible. More importantly, there is no stored database of experimentally validated plant LMIs, which often require RNAhybrid and psRNATarge predictions to obtain labels. This approach may provide an inaccurate and unreliable label for computational models. Closer collaboration between computer scientists and biomedical scientists improves the data quality and opens data sharing is essential for future research. This is so that machine learning methods often require high-quality data to produce accurate prediction performance.

Interpretability. Sequence-based methods often perform deep learning to extract non-linear features from lncRNA and miRNA sequences. A significant challenge facing deep learning methods is interpretability. Due to the complexity of deep neural networks, it is challenging to provide biological explanations for LMI studies. However, it is essential to assess model performance and understand the underlying regulatory mechanisms through interpretability in bioinformatics. In future work, method design needs to take into account the interpretation and visualization of complex relationships, transforming the ”black box” of deep learning into a ”white box” that can be interpreted meaningfully from a biological perspective.

Negative Sample Selection. Computational methods used for LMI prediction tend to require both positive samples (known LMIs) and negative samples (unknown LMIs) to train models. In practice, negative samples are frequently rare or missing, and most studies solve this problem by randomly selecting negative samples from unlabeled data. More importantly, however, these unlabeled data may not necessarily be truly negative samples, which can affect the prediction performance of the model. Currently, this problem can be effectively improve using PU-learning, an unsupervised learning technique that learns from positive and unlabeled data. There is abundant room for further development in negative sample selection and PU-learning.

Scalability. The current applications of LMI prediction methods include two aspects, animal LMI prediction and plant LMI prediction. As mentioned by PmliPEMG, due to the differences of ncRNAs in humans and plants (Different RNA polymerases transcribe them). As a result, LncMirNet achieves good performance in animal interaction prediction, but is unsatisfactory in plants. There is still much room for development on creating a general framework to extend the two problems into one.

Multi-Task Prediction. LncRNA-disease association prediction and miRNA-disease association prediction are also two popular tasks. Studies have shown that lncRNAs and miRNAs typically interact and participate in the development of diseases. Therefore, Sheng et al. integrated lncRNA-disease associations, miRNA-disease associations, and LMIs to construct a triple heterogeneous graph and proposed an end-to-end learning framework to extract lncRNA and disease features and use them for lncRNA-disease prediction [32]. However, they ignore the learned miRNA features. It is entirely feasible to develop an unsupervised learning-based multi-task prediction model to extract lncRNA, miRNA, and disease features for predicting simultaneous LMI, lncRNA-disease association, and miRNA-disease association.

Combining Computational and Biological Experiments. Biological experiments are often neglected by computational scientists who focus more on algorithm-level performance evaluation. Most of them only employed classification metrics to assess the performance of the model, including accuracy, F1-score, AUC, and AUPR. Only Kang et al. performed PmliPred to predict plant LMIs, and then applied qRT-PCR to detect their expression levels and identify the interactions. Therefore, computational scientists and biologists should cooperate closely in future research to further understand the relationships between lncRNAs and miRNAs.

VI. Conclusion

LMIs can reveal various biological functions and mechanisms. Computational methods to screen reliable candidate LMIs are an important and promising approach. This review is the first to summarize the latest research on LMI prediction, and to survey databases, computational models, and evaluation metrics. First, we outline LMI databases, lncRNA- and miRNA-related databases, which are commonly utilized in animal and plant LMI prediction. Second, we reviewed 20 computational models for predicting LMIs and divided them into two categories, network-based methods and sequence-based methods. Third, we summarize the typical evaluation methods and metrics used in computational methods. Finally, we discuss the future challenges and directions of LMI prediction research.

References

[1] P. Kapranov et al., “RNA maps reveal new RNA classes and a possible function for pervasive transcription,” *Science*, vol. 316, no. 5830, pp. 1484–1488, 2007.
[2] R. P. Alexander et al., “Annotating non-coding regions of the genome,” *Nature Rev. Genet.*, vol. 11, no. 8, pp. 559–571, 2010.
[3] M. Guttman and J. L. Rinn, “Modular regulatory principles of large non-coding RNAs,” *Nature*, vol. 482, no. 7385, pp. 339–346, 2012.
[4] C. P. Ponting, P. L. Oliver, and W. Reik, “Evolution and functions of long noncoding RNAs,” *Cell*, vol. 136, no. 4, pp. 629–641, 2009.
[5] C. S. Sullivan, “New roles for large and small viral RNAs in evading host defences,” *Nature Rev. Genet.*, vol. 9, no. 7, pp. 503–507, 2008.
[6] L. Chen, Y. Zhou, and H. Li, “LncRNA, miRNA and lncRNA-miRNA interaction in viral infection,” *Virus Res.*, vol. 9, no. 7, pp. 503–507, 2008.
[7] M. Esteller, “Non-coding RNAs in human disease,” *Nature Rev. Genet.*, vol. 12, no. 12, pp. 861–874, 2011.
[8] K. Katsarou et al., “Infectious long non-coding RNAs,” *Biochimie*, vol. 117, pp. 37–47, 2015.
[9] J.-H. Li et al., “starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. D92–D97, 2013.
[10] Y. Zhang, C. Jia, and C. K. Kwoh, “Predicting the interaction biomolecule types for lncRNA: An ensemble deep learning approach,” *Brief. Bioinf.*, vol. 22, no. 4, 2020, Art. no. bbaa228.
[11] T. Shi et al., “The role of long non-coding RNA and microRNA networks in hepatocellular carcinoma and its tumor microenvironment,” *Int. J. Mol. Sci.*, vol. 22, no. 19, 2021, Art. no. 10630.
[12] B. Sun et al., “Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer (Review),” *Oncol. Lett.*, vol. 19, no. 1, pp. 595–605, 2020.

[13] L. Juan et al., "Potential roles of microRNAs in regulating long intergenic noncoding RNAs," *BMC Med. Genomic.*, vol. 6, no. 1, 2013, Art. no. S7.
[14] F. H. C Tsang et al., "Long non-coding RNA HOTTIP is frequently up-regulated in hepatocellular carcinoma and is targeted by tumour suppressive miR-125b," *Liver Int.*, vol. 35, no. 5, pp. 1597–1606, 2015.
[15] B. He et al., "Epstein-Barr virus-encoded miR-BART6–3p inhibits cancer cell metastasis and invasion by targeting long non-coding RNA LOC553103," *Cell Death Dis.*, vol. 7, no. 9, pp. e2353–e2353, 2016.
[16] M. Cantile et al., "Functional interaction among lncRNA HOTAIR and MicroRNAs in cancer and other human diseases," *Cancers*, vol. 13, no. 3, 2021, Art. no. 570.
[17] J. Wang et al., "Identification of the complex regulatory relationships related to gastric cancer from lncRNA-miRNA-mRNA network," *J. Cellular Biochem.*, vol. 121, no. 1, pp. 876–887, 2020.
[18] Y. Yu et al., "A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells," *Mol. Cancer*, vol. 16, no. 1, 2017, Art. no. 155.
[19] Q. Kong et al., "The lncRNA MIR4435–2HG is upregulated in hepatocellular carcinoma and promotes cancer cell proliferation by upregulating miRNA-487a," *Cellular Mol. Biol. Lett.*, vol. 24, no. 1, 2019, Art. no. 26.
[20] R. Wang et al., "MiR-326 regulates cell proliferation and migration in lung cancer by targeting Phox2a and is regulated by HOTAIR," *Amer. J. Cancer Res.*, vol. 6, no. 2, 2016, Art. no. 173.
[21] H. Cai et al., "Epigenetic inhibition of miR-663b by long non-coding RNA HOTAIR promotes pancreatic cancer cell proliferation via up-regulation of insulin-like growth factor 2," *Oncotarget*, vol. 7, no. 52, 2016, Art. no. 86857.
[22] E.-B. Bian, Z.-G. Xiong, and J. Li, "New advances of lncRNAs in liver fibrosis, with specific focus on lncRNA-miRNA interactions," *J. Cellular Physiol.*, vol. 234, no. 3, pp. 2194–2203, 2019.
[23] X. Meng et al., "Interplay between miRNAs and lncRNAs: Mode of action and biological roles in plant development and stress adaptation," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 2567–2574, 2021.
[24] A. Das et al., "Non-coding RNAs having strong positive interaction with mRNAs reveal their regulatory nature during flowering in a wild relative of pigeonpea (Cajanus scarabaeoides)," *Mol. Biol. Rep.*, vol. 47, no. 5, pp. 3305–3317, 2020.
[25] X. Hou et al., "LncRNA39026 enhances tomato resistance to phytophthora infestans by decoying miR168a and inducing PR gene expression," *Phytopathology*, vol. 110, no. 4, pp. 873–880, 2019.
[26] J. Cui et al., "Genome-wide identification of lncRNAs and analysis of ceRNA networks during tomato resistance to phytophthora infestans," *Phytopathology*, vol. 110, no. 2, pp. 456–464, 2019.
[27] Q. Lu et al., "LncRNA improves cold resistance of winter wheat by interacting with miR398," *Funct. Plant Biol.*, vol. 47, no. 6, pp. 544–557, 2020.
[28] J. Chen, Y. Zhong, and X. Qi, "LncRNA TCONS_00021861 is functionally associated with drought tolerance in rice (Oryza sativa L.) via competing endogenous RNA regulation," *BMC Plant Biol.*, vol. 21, no. 1, 2021, Art. no. 410.
[29] A. Biswas et al., "Co-regulatory functions of miRNA and lncRNA in adapting biotic and abiotic stress in economically important dicot plants," *Plant Gene*, vol. 26, 2021, Art. no. 100275.
[30] N. Natarajan and I. S. Dhillon, "Inductive matrix completion for predicting gene-disease associations," *Bioinformatics*, vol. 30, no. 12, pp. i60–i68, 2014.
[31] N. Sheng et al., "Attentional multi-level representation encoding based on convolutional and variance autoencoders for lncRNA-disease association prediction," *Brief. Bioinf.*, vol. 22, no. 3, 2021, Art. no. bbaa067.
[32] N. Sheng et al., "Multi-channel graph attention autoencoders for disease-related lncRNAs prediction," *Brief. Bioinf.*, vol. 23, no. 2, 2022, Art. no. bbab604.
[33] N. Sheng et al., "Data resources and computational methods for lncRNA-disease association prediction," *Comput. Biol. Med.*, vol. 153, 2023, Art. no. 106527.
[34] X. Tang et al., "Multi-view multichannel attention graph convolutional network for miRNA-disease association prediction," *Brief. Bioinf.*, vol. 22, no. 6, 2021, Art. no. bbab174.
[35] P. Xuan et al., "Integration of pairwise neighbor topologies and miRNA family and cluster attributes for miRNA-disease association prediction," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab428.
[36] L. Huang et al., "LGFC-CNN: Prediction of lncRNA-Protein interactions by using multiple types of features through deep learning," *Genes*, vol. 12, no. 11, 2021, Art. no. 1689.
[37] W. Zhang et al., "SFPEL-LPI: Sequence-based feature projection ensemble learning for predicting LncRNA-protein interactions," *PLoS Comput. Biol.*, vol. 14, no. 12, 2018, Art. no. e1006616.
[38] W. Zhang et al., "The linear neighborhood propagation method for predicting long non-coding RNA-protein interactions," *Neurocomputing*, vol. 273, pp. 526–534, 2018.
[39] L. Gao et al., "Prediction of drug-disease associations by integrating common topologies of heterogeneous networks and specific topologies of subnets," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab467.
[40] P. Xuan, L. Gao, N. Sheng, T. Zhang, and T. Nakaguchi, "Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1793–1804, May 2021.
[41] Y. Yang et al., "lncRNASNP v3: An updated database for functional variants in long non-coding RNAs," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D192–D198, 2022.
[42] P.-J. Volders et al., "LNCipedia 5: Towards a reference set of human long non-coding RNAs," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D135–D139, 2018.
[43] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, "miRBase: From microRNA sequences to function," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D155–D162, 2018.
[44] Y. Zhao et al., "NONCODE 2016: An informative and valuable data source of long non-coding RNAs," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D203–D208, 2015.
[45] D. Betel et al., "The microRNA.org resource: Targets and expression," *Nucleic Acids Res.*, vol. 36, no. suppl_1, pp. D149–D153, 2008.
[46] H.-Y. Huang et al., "miRTarBase update 2022: An informative resource for experimentally validated miRNA-target interactions," *Nucleic Acids Res.*, vol. 50, no. D1, pp. D222–D230, 2021.
[47] M. W. Szczesniak et al., "CANTATAdb 2.0: Expanding the collection of plant long noncoding RNAs," in *Plant Long Non-Coding RNAs: Methods and Protocols*. Berlin, Germany: Springer, 2019, pp. 415–429.
[48] Z. Zhang et al., "PMRD: Plant microRNA database," *Nucleic Acids Res.*, vol. 38, no. suppl_1, pp. D806–D813, 2009.
[49] M. Di Marsico et al., "GreeNC 2.0: A comprehensive database of plant long non-coding RNAs," *Nucleic Acids Res.*, vol. 50, no. D1, pp. D1442–D1447, 2021.
[50] J. Kruger and M. Rehmsmeier, "RNAhybrid: MicroRNA target prediction easy, fast and flexible," *Nucleic Acids Res.*, vol. 34, no. suppl_2, pp. W451–W454, 2006.
[51] X. Dai, Z. Zhuang, and P. X. Zhao, "psRNATarget: A plant small RNA target analysis server (2017 release)," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W49–W54, 2018.
[52] Y.-A. Huang, K. C. C. Chan, and Z.-H. You, "Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling," *Bioinformatics*, vol. 34, no. 5, pp. 812–819, 2017.
[53] Y.-A. Huang et al., "Predicting lncRNA-miRNA interaction via graph convolution auto-encoder," *Front. Genet.*, vol. 10, 2019, Art. no. 758.
[54] W. Zhang et al., "LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination," *BMC Genomic.*, vol. 20, no. 11, 2019, Art. no. 946.
[55] W. Zhang, W. Yang, X. Lu, F. Huang, and F. Luo, "The bi-direction similarity integration method for predicting microbe-disease associations," *IEEE Access*, vol. 6, pp. 38052–38061, 2018.
[56] L. Wong et al., "LNRLMI: Linear neighbour representation for predicting lncRNA-miRNA interactions," *J. Cellular Mol. Med.*, vol. 24, no. 1, pp. 79–87, 2020.
[57] P. Hu, Y. A. Huang, K. C. C. Chan, and Z.-H. You, "Learning multimodal networks from heterogeneous data for prediction of lncRNA-miRNA interactions," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 5, pp. 1516–1524, Sep./Oct. 2020.
[58] C. Zhao et al., "Graph embedding ensemble methods based on the heterogeneous network for lncRNA-miRNA interaction prediction," *BMC Genomic.*, vol. 21, no. 13, 2020, Art. no. 867.
[59] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 585–591.
[60] S. Cao, W. Lu, and Q. Xu, "GraRep: Learning graph representations with global structural information," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2015, pp. 891–900.

[61] M. Ou et al., "Asymmetric transitivity preserving graph embedding," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1105–1114.

[62] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.

[63] M.-N. Wang, Z.-H. You, L.-P. Li, L. Wong, Z.-H. Chen, and C.-Z. Gan, "GNMFLMI: Graph regularized nonnegative matrix factorization for predicting LncRNA-MiRNA interactions," *IEEE Access*, vol. 8, pp. 37578–37588, 2020.

[64] H. Liu et al., "Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized," *Knowl.-Based Syst.*, vol. 191, 2020, Art. no. 105261.

[65] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 3553–3559.

[66] W. Wang et al., "LMI-DForest: A deep forest model towards the prediction of lncRNA-miRNA interactions," *Comput. Biol. Chem.*, vol. 89, 2020, Art. no. 107406.

[67] Y. Fan, J. Cui, and Q. Zhu, "Heterogeneous graph inference based on similarity network fusion for predicting lncRNA-miRNA interaction," *RSC Adv.*, vol. 10, no. 20, pp. 11634–11642, 2020.

[68] L. Zhang et al., "Predicting lncRNA-miRNA interactions based on interactome network and graphlet interaction," *Genomics*, vol. 113, no. 3, pp. 874–880, 2021.

[69] L. Zhang et al., "Using network distance analysis to predict lncRNA-miRNA interactions," *Interdiscipl. Sci. Comput. Life Sci.*, vol. 13, no. 3, pp. 535–545, 2021.

[70] W. Wang et al., "Predicting the potential human lncRNA-miRNA interactions based on graph convolution network with conditional random field," *Brief. Bioinf.*, vol. 23, 2022, Art. no. bbac463.

[71] S. Yang et al., "LncMirNet: Predicting LncRNA-miRNA interaction based on deep learning of ribonucleic acid sequences," *Molecules*, vol. 25, no. 19, 2020, Art. no. 4372.

[72] B. Liu et al., "Identification of microRNA precursor with the degenerate K-tuple or Kmer strategy," *J. Theor. Biol.*, vol. 385, pp. 153–159, 2015.

[73] X. Tong and S. Liu, "CPPred: Coding potential prediction based on the global description of RNA sequence," *Nucleic Acids Res.*, vol. 47, no. 8, pp. e43–e43, 2019.

[74] J. H. Lau and T. Baldwin, "An empirical evaluation of doc2vec with practical insights into document embedding generation," in *Proc. 1st Workshop Representation Learn. NLP*, 2016, pp. 78–86.

[75] N. K. Ahmed et al., "Role-based graph embeddings," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 2401–2415, May 2022.

[76] P. Zhang et al., "Plant miRNA-lncRNA interaction prediction with the ensemble of CNN and IndRNN," *Interdiscipl. Sci. Comput. Life Sci.*, vol. 12, no. 1, pp. 82–89, 2020.

[77] Q. Kang et al., "PmliPred: A method based on hybrid model and fuzzy decision for plant miRNA-lncRNA interaction prediction," *Bioinformatics*, vol. 36, no. 10, pp. 2986–2992, 2020.

[78] T. D. C. Negri et al., "Pattern recognition analysis on long noncoding RNAs: A tool for prediction in plants," *Brief. Bioinf.*, vol. 20, no. 2, pp. 682–689, 2018.

[79] Q. Kang et al., "Ensemble deep learning based on multi-level information enhancement and greedy fuzzy decision for plant miRNA-lncRNA interaction prediction," *Interdiscipl. Sci. Comput. Life Sci.*, vol. 13, no. 4, pp. 603–614, 2021.

[80] Y. Zhang et al., "DeepCPP: A deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction," *Brief. Bioinf.*, vol. 22, no. 2, pp. 2073–2084, 2020.

[81] Q. Kang et al., "Mining plant endogenous target mimics from miRNA-lncRNA interactions based on dual-path parallel ensemble pruning method," *Brief. Bioinf.*, vol. 23, no. 1, 2021, Art. no. bbab440.

[82] J. Song et al., "MD-MLI: Prediction of miRNA-lncRNA interaction by using multiple features and hierarchical deep learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 3, pp. 1724–1733, May/Jun. 2022.

[83] P. Danaee et al., "bpRNA: Large-scale automated annotation and analysis of RNA secondary structure," *Nucleic Acids Res*, vol. 46, no. 11, pp. 5381–5394, 2018.

[84] X. Yu et al., "preMLI: A pre-trained method to uncover microRNA-lncRNA potential interactions," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab470.

[85] T. Mikolov et al., "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.

Nan Sheng is currently working toward the PhD degree with the College of Computer Science and Technology, Jilin University. His research interests include bioinformatics and data mining.

Lan Huang is a professor with the College of Computer Science and Technology, Jilin University. Her research primarily focuses on bioinformatics, data mining, and machine learning.

Ling Gao is currently working toward the PhD degree with the College of Computer Science and Technology, Jilin University. Her research interests include image processing and bioinformatics.

Yangkun Cao is currently working toward the PhD degree with the School of Artificial Intelligence, Jilin University. His research interests include computational biology and data mining.

Xuping Xie is currently working toward the PhD degree with the College of Computer Science and Technology, Jilin University. Her research interests include bioinformatics and data mining.

Yan Wang is a professor with the College of Computer Science and Technology, Jilin University. His research primarily focuses on bioinformatics, data mining, and machine learning.