Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 2785436, 15 pages
https://doi.org/10.1155/2023/2785436

Research Article

SEBGLMA: Semantic Embedded Bipartite Graph Network for Predicting lncRNA-miRNA Associations

Zheng-Yang Zhao,[1,2] **Jie Lin**,[1] **Zhen Wang**,[1,2] **Jian-Xin Guo**,[1] **Xin-Ke Zhan**,[1] **Yu-An Huang**,[1] **Chuan Shi**,[1] **and Wen-Zhun Huang**[1]

[1]*School of Information Engineering, Xijing University, Xi'an 710123, China*
[2]*School of Information and Navigation, Air Force Engineering University, Xi'an 710082, China*

Correspondence should be addressed to Wen-Zhun Huang; huangwenzhun@xijing.edu.cn

Received 26 September 2022; Revised 14 November 2022; Accepted 16 November 2022; Published 24 February 2023

Academic Editor: Paolo Gastaldo

Identifying the association between long noncoding RNA (lncRNA) and micro-RNA (miRNA) is of great significance for the treatment of diseases by interfering with the combination of miRNA and messenger RNA (mRNA). Although many efforts and resources have been invested to identify lncRNA-miRNA associations (LMAs), clinical trials are still expensive and laborious. Nevertheless, the experiments also need to consult a large number of side effects. Therefore, novel computer-aided models are urgently needed to predict LMAs. This paper proposed a semantic embedded bipartite graph network for predicting lncRNA-miRNA associations (SEBGLMA), which provided a novel feature extraction method by integrating *K*-mer segmentation, word2vec, Gaussian interaction profile (GIP), and graph convolution network (GCN). Concretely, the attribute characteristics of RNA sequences are extracted by *K*-mer segmentation and word2vec modules. Afterward, the adjacent matrix is completed by GIP self-similarity. Then, the attribute characteristics and adjacent matrix are fed into GCN for embedding behavior features. Finally, the features are sent into the rotation forest (RoF) for detecting potential LMAs. The average accuracy, precision, sensitivity, specificity, Matthews correlation coefficient, and F1-Score are 87.09%, 87.66%, 87.03%, 87.84%, 74.18%, and 86.99% on the benchmark data set. For fairly validating the performance of our model, we conducted various comparisons with different classifiers. Furthermore, the case studies of hsa-miR-497-5P and NONHSAT022145.2 are also established. The results of comparisons and case studies further illustrated that our method is anticipated to become a robust and reliable tool for the identification of LMAs.

1. Introduction

As a regulatory chemical molecule, noncoding RNAs participate in many biological activities, such as epigenetic control, gene transcription, translation, chromosome organization, cell proliferation, and development programs [1]. In the biological regulatory network, each RNA plays a unique role. Long noncoding RNAs (lncRNA) with more than 200 nucleotides can take part in a variety of biological activities including cell differentiation, cell growth, disease treatment, and gene transcription [2]. The diversity and additional functions of lncRNAs have risen extensive attention. Micro RNA (miRNA) is the most common target RNA of lncRNA with 19–22 nucleotides [3, 4]. The miRNAs transcribed by viruses change the number of proteins and immune resistance by affecting the host messenger RNA (mRNA) transcription efficiency. Recently, increasing evidence emphasizes the role of long noncoding RNAs (lncRNAs) as the epigenetic factor in disease occurrence and development. Moreover, researchers are more focused on their relationships with downstream target miRNAs [5].

LncRNA has a similar structure and transcription to mRNA. Previous research studies have indicated that lncRNA and miRNA can combine with each other to control gene expression. LncRNA has the ability to regulate miRNA functions by serving as competing endogenous RNA

(ceRNA) to intervene in miRNA sponging and alter expression levels [6, 7]. Meanwhile, miRNA regulates the expression stability of lncRNA in an Argonaute2-dependent manner by reducing the host lncRNA for the virus. LncRNA provides a reliable chemical combination platform for its special segment composition. It limits the ability of miRNA to interfere with the mRNA protein encoding process by competitively binding miRNA in cells.

Up to now, only a small part of the working mechanism of lncRNA-miRNA association (LMA) has been studied. There are still a large number of lncRNA-MiRNA pairs that need to be explored. However, the existing research results have proved that tumors can be inhibited by regulating the corresponding LMA interaction. In addition to the ceRNA mechanism described above, researchers also found that LMA plays a crucial role in regulating the pathogenesis mechanisms of cancers, including ephemeral molecular transition (EMT), cancer stem cells (CSCs), drug resistance, and other molecular mechanisms. For instance, recent research studies have uncovered partial lncRNA-miRNA association networks hidden in breast, bladder, and colorectal cancer [8–10]. Therefore, identifying the associations between lncRNAs and miRNAs will benefit regulating the quantities of different RNAs for the treatment of diseases and become chemical indicators for diagnosis and prognosis. Feasible prediction models are urgently needed to find the potential LMAs for further understanding the mechanism of gene molecular networks on the molecular level. With the advancement of gene sequencing and similarity calculation. Many advanced methods are developed by researchers to detect possible LMAs.

The wet experiment is a traditional method to determine LMAs. Zhang et al. [11] find that the downregulation of miR-7 in BCSCs might be indirectly attributed to lncRNA HOTAIR by implementing ChIP-PCR and Double-Luciferase Reporter assay. Kallen et al. [12] revealed that H19 is an important regulator of the major let-7 family of microRNAs by crosslinking and real-time PCR. In recent years, the increasing clinical experimental data has provided an available platform for computer-aided prediction systems. Wang et al. [13] developed the KATZ model to construct a heterogeneous network. The network only utilized the topology information to generate the association scores of lncRNA-miRNA pairs. Yang et al. [14] proposed the multiple-index latent factor model (MILFM), which projects the predicting information onto a few local subspaces to obtain key characteristics for further fitting by regularization. Zhang et al. [15] proposed the sequence-derived linear neighborhood propagation method (SLNPM) to find the potential interacting lncRNA-miRNA pairs. Specifically, the similarity information combination and interaction profile information combination are applied in evaluating the weighted averages. Huang et al. [16] developed a novel group preference Bayesian collaborative filtering (GBCF) model for picking up a top-k probability ranking list for an individual miRNA or lncRNA. Xu et al. [17] proposed a structural perturbation method for predicting lncRNA-miRNA interactions (SPMLMI) which utilized the Pearson correlation coefficient for measuring lncRNA and miRNA similarity and form a two-layer relationship network. Recently, Liu et al. [18] developed the logistic matrix factorization with a neighborhood regularized (LMFNRLMI) model to utilize the strongest adjacent relation in the neighborhood and established an adjacency matrix through the K nearest neighbor approach to infer LMAs.

In this paper, we designed a novel model named semantic embedded bipartite graph networks for predicting lncRNA-miRNA associations (SEBGLMA). This method combines K-mer segmentation, word2vec, Gaussian interaction profile (GIP), graph convolution network (GCN), and rotation forest (RoF). The structure of this model can be separated into the following sections. (1) The sequences of RNAs are first segmented into subsequences to generate corpuses for word2vec which record different combinations of amino acids by K-mer segmentation. (2) Send the corpuses into word2vec to construct Lnc2Vec and Mi2Vec models for extracting semantic attribute characteristics. (3) Feed the known lncRNA-miRNA associations into the GIP model to obtain sample self-similarities with the same class, and utilize the self-similarities to complete the adjacency matrix. (4) Input the adjacency matrix and semantic characteristics into GCN to form a bipartite graph model. The outputs of GCN are regarded as the fusion features of nodes combining attribute and behavior information. (5) Send the features into the RoF for predicting the possible LMAs. Moreover, the case studies of hsa-miR-497-5P and NONHSAT022145.2 are also conducted. This paper makes the following contributions: (1) the RNA functional similarity of the same class is used to enrich the edges in the graph; (2) the attribute and behavior characteristics of lncRNAs and miRNAs have been effectively fused to improve sample diversity; (3) our model has great effects on screening candidates for subsequent clinical trials. The workflow of SEBGLMA is displayed as Figure 1.

The rest architecture of this study is arranged as follows: in Section 2, we introduce the benchmark data set for LMA prediction and different parts of the proposed model. In Section 3, we discuss the criteria, performance of the proposed model, and comparisons. The conclusion and feature work are illustrated in Section 4.

2. Materials and Methods

2.1. Datasets. In this paper, the benchmark dataset is established based on the clinical database lncRNASNP2 (https://bioinfo.life.hust.edu.cn/lncRNASNP#!/). LncRNASNP2 offers numerous interactions between lncRNAs and miRNAs which have been clinically validated. The specific subjects of this database include human and mouse [19]. It also provides information on single-nucleotide polymorphisms (SNPs), mutations, and diseases about lncRNA. This study excavated 780 lncRNAs and 275 miRNA-related interaction data from the database and collected 10597 interacting pairs. Furthermore, the corresponding amino acid sequences of lncRNAs were obtained from the LNCipedia database (https://lncipedia.org/) [20], and sequences of miRNAs were achieved from the MiRbase database (https://www.mirbase.org/index.shtml) [21]. Finally, there

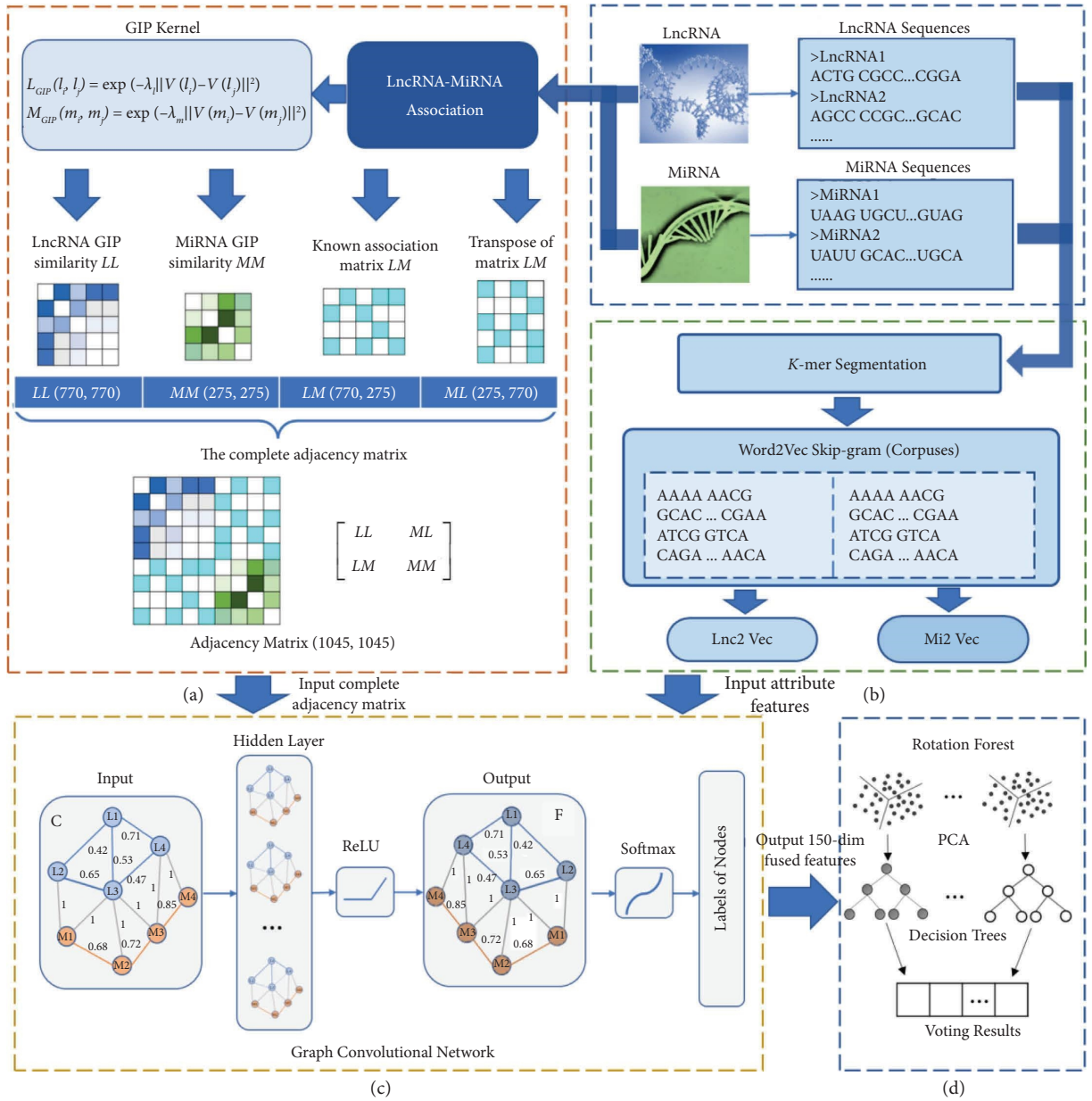

Figure 1: The flow chart of SEBGLMA. (a) Complete adjacency matrix by GIP kernel similarity. (b) Calculate semantic features by *K*-mer and word2vec. (c) Integrate semantic and behavior features by graph convolution network. (d) Feed the fusion features into the rotation forest for predicting LMAs.

are 4966 interacting pairs remaining in the benchmark dataset, involving 770 lncRNAs and 275 miRNAs after eliminating redundant information. Table 1 gives the statistics of the benchmark data set.

2.2. GIP Kernel. The similarities between the samples of the same category calculated by functional difference can explicably describe the differences between samples, compared with the similarities obtained by calculating the numerical and spatial distance between sequences. The genes with more similarities usually have a similar interoperating mechanism. To be specific, lncRNAs that combine the same target miRNAs often show more similarities.

Table 1: The statistics of the benchmark data set.

Database	Data type	Number of edges (nodes)
LncRNASNP2	lncRNA-miRNA associations	4966 (associations/ edges)
LNCipedia	lncRNA amino acid sequences	770 (lncRNA nodes)
MiRbase	miRNA amino acid sequences	275 (miRNA nodes)

Hence, the Gaussian interaction profile (GIP) kernel was utilized to separately describe the lncRNA and miRNA self-similarities for further completing the adjacency matrix

which describes the relationships between the nodes of the graph network [22–24].

The lncRNA-miRNA associations offer a topological structure for the GIP kernel to generate two types of functional self-similarities, which are denoted as L_{GIP} and M_{GIP}. The vectors $V(l_i)$ and $V(m_i)$ represent the internal associations of lncRNA and miRNA, respectively. Take miRNA as an example, $V(m_i)$ is constructed as a binary sentence by detecting whether the miRNA interacts with a series of lncRNAs. The vector element corresponding to the interaction is marked as 1, and the rest is marked as 0 [25]. Thus, the self-similarities L_{GIP} and M_{GIP} can be generated by the following equations:

$$\begin{aligned} L_{\mathrm{GIP}}\left(l_i, l_j\right) &= \exp\left(-\lambda_l \left\| V(l_i) - V\left(l_j\right)\right\|^2\right), \\ M_{\mathrm{GIP}}\left(m_i, m_j\right) &= \exp\left(-\lambda_m \left\| V(m_i) - V\left(m_j\right)\right\|^2\right), \end{aligned} \tag{1}$$

where l_i and l_j represent the ith and jth lncRNA, m_i and m_j represent the ith and jth miRNA. λ is the normalized kernel bandwidth parameter of GIP kernel similarity, λ' is the original bandwidth parameter. The definitions of λ and λ' are as follows:

$$\begin{aligned} \lambda_l &= \frac{\lambda_l'}{\left[1/n_l \sum_{i=1}^{n_l} \left\| V(l_i)\right\|^2\right]}, \\ \lambda_m &= \frac{\lambda_m'}{\left[1/n_m \sum_{i=1}^{n_m} \left\| V(m_i)\right\|^2\right]}. \end{aligned} \tag{2}$$

Finally, the self-similarities L_{GIP} and M_{GIP} composed the blocks LL and MM. The completed adjacency matrix is shown as follows:

$$A = \begin{bmatrix} \mathrm{LL} & \mathrm{LM} \\ \mathrm{ML} & \mathrm{MM} \end{bmatrix}, \tag{3}$$

where A is a 1045×1045 matrix including the topological relationships of 770 lncRNA and 275 miRNA nodes. Blocks LL and MM represent the self-similarity metrics of lncRNAs and miRNAs, respectively. As be noticed, block ML is the transpose of LM.

2.3. K-Mer Segmentation. Before numerical describing lncRNA and miRNA, the amino acid sentences have to be segmented for building corpuses. For the given sentences, K-mer method is employed to originally explore the semantic features [26, 27]. Specifically, the sentences are divided into subsequences by sliding windows, and the length of this window is K. For instance, the sentence which contains N amino acids, it will generate $N - K + 1$ subsentences. In this section, the parameter K is set as 4. The amounts of the possible lncRNA and miRNA subsentences are all 4^K, for lncRNAs containing the Alanine (A), Cysteine (C), Glycine (G), and Threonine (T) amino acids; miRNAs contain Alanine (A), Cysteine (C), Glycine (G), and Selenocysteine (U) amino acids [28]. An example of NONHSAT129051.2 converting into subsequences is shown in Figure 2.

2.4. Distribution Representation of lncRNA and miRNA Sequences. The amino acid subsequences of lncRNAs and miRNAs were utilized to construct the RNA2Vec models including Lnc2Vec and Mi2Vec models for word embedding. These models differentially characterize lncRNA and miRNA based on biological evidence. Then, the sequences were transformed into digital vectors as attribute characteristics. Considering the size of the benchmark data set, the word2vec model with skip-gram is applied to realize the word representation [29, 30]. In general, this model constructs a projection neural network to learn the distribution of words by sliding windows. The framework of the skip-gram is displayed in Figure 3.

With regard to the sentence $(w_1, w_2, \cdots, w_{N-K+1})$ where w represents the amino acid, the objective function of the model is defined as follows:

$$\max \frac{1}{N-K+1} \sum_{l=1}^{N-K+1} \sum_{-c \le m \le c, m \ne 0} \log P\left(w_{n+m} \mid w_n\right), \tag{4}$$

where c is the maximum distance between the words and the central word in the sliding window. The conditional probability $\log P(w_{n+m} \mid w_n)$ is calculated by the following equation (31):

$$\log P\left(w_a \mid w_b\right) = \log \frac{e^{v'^{T}_{w_a} v_{w_i}}}{\sum_{w=1}^{w} e^{v'^{T}_{w_a} v_{w_i}}}, \tag{5}$$

where v_w and v_w' represent the original and output formats of the word w, respectively. W stand for the established lexicon width. Similar to the nondigital text data processing, this section inputs the sequences and corpuses into the network for obtaining the numerical vectors as semantic features of genes. After optimizing, the parameters size, window, iter, and batch_words are set as 500, 5, 10, and 10, respectively. The size determines the length of the output vector; the window gives the maximum distance between a central word with contextual words; iter is the number of model iterations; batch_words represents the account of words transferred to operations. The other parameters are set as default values. Figure 4 gives an example of the semantic embedding process.

2.5. Graph Convolution Network. To date, graph representation models are divided into traditional graph algorithms and graph neural networks. The traditional graph model can effectively implement the low-dimensional representation of network nodes without the attribute characteristics of the model. However, graph neural networks can make full use of the attribute characteristics of nodes generated by the increasing semantic feature extraction methods to improve the feature difference between nodes. Moreover, the graph neural networks can better represent graph structure characteristics of the nodes.

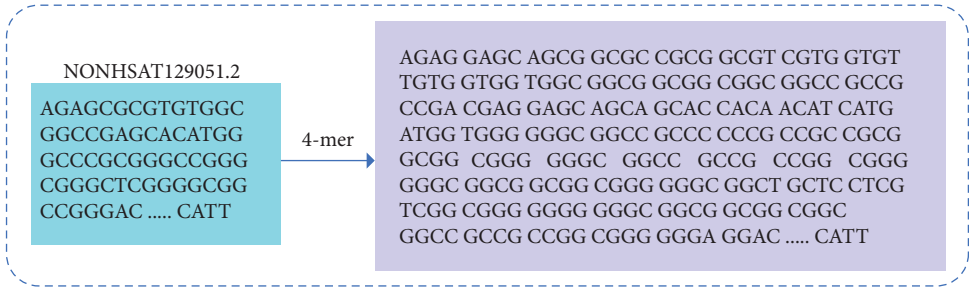

Figure 2: *K*-mer segmentation of NONHSAT129051.2.

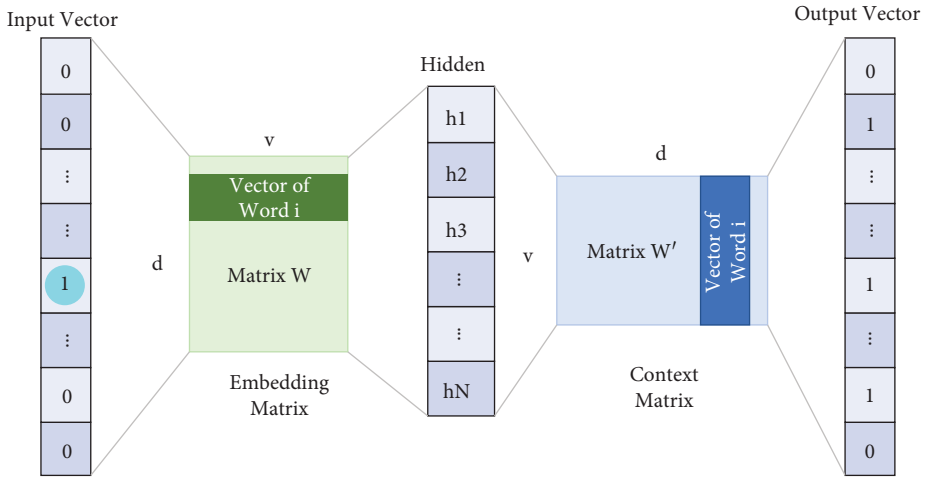

Figure 3: The structure of skip-gram.

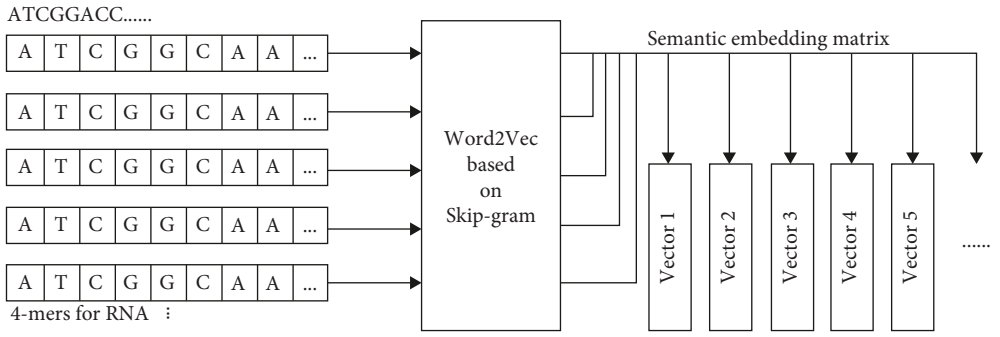

Figure 4: The example of semantic embedding process.

Kipf et al. proposed an extensible method of graph convolution network (GCN) for semisupervised learning of graph structure data, which is based on an efficient variant of a convolution neural network [32]. It is widely utilized to characterize the associations between the central node and neighbor nodes. This model linearly extends the number of edges and learns to encode the hidden layer representation of local graph structure and node features [33]. Therefore, GCN is employed to calculate the behavior features. The structure of GCN is shown in Figure 5.

In this research, the associations between lncRNAs and miRNAs are embedded in the graph $G = (V, E)$, V and E denote the nodes and edges, respectively. Meanwhile, the attribute features and adjacency matrix are fed into GCN

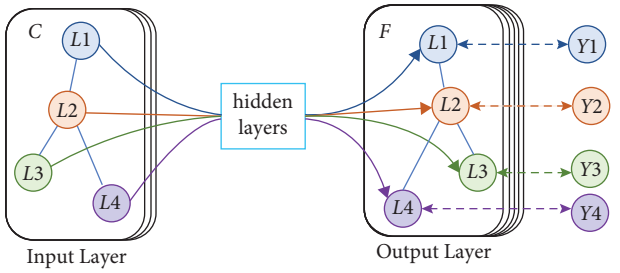

Figure 5: The structure of graph convolution network.

[34, 35]. Hence, the GCN can also integrate the attribute features with behavior features. Specifically, the Laplace regularization term is applied to define the convolution

kernel for differentially describing nodes. The Laplace regularized formula is expressed as follows:

$$L = I - D^{-1/2}AD^{-1/2}, \tag{6}$$

where I is the adaptive identity matrix, D is the degree matrix, A is the adjacency matrix. The application of the Laplace regularization term will make the convolution process smoother. The graph matrix decomposition of L and the convolution process are defined as follows:

$$\begin{aligned} L &= U\Lambda U^T, \\ g_\theta * x &= U g_\theta U^T x, \end{aligned} \tag{7}$$

where x denotes the input features of nodes. It aims to transfer nodes into Fourier space. Subsequently, the Chebyshev matrix is employed to approximate the convolution kernel for reducing the computational complexity, the new convolution kernel is as follows:

$$\begin{aligned} g_\theta(\Lambda) &= \sum_{k=0}^{K-1} \theta_k \Lambda^k \approx \sum_{k=0}^{K} \theta_k' T_k(\widetilde{\Lambda}), \\ \widetilde{\Lambda} &= \frac{2}{\lambda_{\max}}\Lambda - I_N, \theta' \in R^k, \end{aligned} \tag{8}$$

where $\lambda_{\max}$ is the maximum eigenvalue of L. In this paper, we only consider the first-order neighbors of the central node, the parameters K and $\lambda_{\max}$ are set as 1 and 2. The Chebyshev recursive formula is defined as follows:

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), T_0(x) = 1, T_1(x) = x. \tag{9}$$

The equation of graph convolution is updated to the followinf equation:

$$g_\theta * x \approx \theta_0' x + \theta_0'(L - I_N)x = \theta\left(I_N + D^{-1/2}AD^{-1/2}\right)x. \tag{10}$$

For preventing gradient explosion, the equation is further standardized as follows:

$$Z = \widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}X\Theta, \widetilde{A} = A + I_N, \widetilde{D} = \sum_j \widetilde{A}_{ij}. \tag{11}$$

Here, X is an $N \times C$ matrix containing C-dimensional semantic features of N nodes, and Θ is a trainable matrix set. According to the convolution kernel, the forward propagation formula of a two-layer model is as follows:

$$F = \widehat{A}\mathrm{ReLU}\left(\widehat{A}XW^{(0)}\right)W^{(1)}, \tag{12}$$

where $W^{(0)}$, $W^{(1)}$ represent the weight matrix of hidden layers 1 and 2. Finally, matrix F is utilized as the output feature integrating attribute and behavior characteristics.

2.6. Rotation Forest. Rodriguez et al. [36] evolved integrated forest into rotation forest (RoF) which promotes the difference by adding a rotation module. It provides a feasible pipe for dealing with the benchmark dataset. Therefore, this ensemble model is established to promote the feature difference and detect the interactions between lncRNAs and miRNAs [37]. Firstly, the whole subjects have to be stochastically disjointed into L independent subsets. Subsequently, Principal Component Analysis (PCA) algorithm attends to transfer subsets for integrating RF. Finally, these converted subsets are fed into the base classifiers for scoring subtrees. The matrix Q of $s \times S$ represents the train set with S characteristics of s samples. The corresponding labels $R = (r_1, r_2, \cdots, r_n)^T$ are also sent into the model to supervise the training process [38]. This model has K base classifiers H_i. The sequential training parts are as follows:

(i) After optimizing the parameters, the dataset Z is divided into L disjoint subsets containing $m = S/L$ features.

(ii) Regard $Z_{i,j}$ as jth subset of Z, and $Q_{i,j}$ as the feature set of $Z_{i,j}$. Then, bootstrap sampling on 75% of $Q_{i,j}$ to generate training set $Q_{i,j}'$.

(iii) Apply PCA in $Q_{i,j}'$ to obtain principal component coefficients $a_{i,j}^{(1)}, a_{i,j}^{(2)}, \cdots a_{i,j}^{(m_j)}$.

(vi) The rotation matrix P_i made up by coefficients is as follows:

$$P_i = \begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(2)}, \cdots a_{i,1}^{(M_1)} & 0 & \cdots & 0 \\ 0 & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \cdots a_{i,2}^{(M_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{i,K}^{(1)}, a_{i,K}^{(2)}, \cdots a_{i,K}^{(M_k)} \end{bmatrix}. \tag{13}$$

In the classification, the possibility that the sample x belongs to category r_i is $d_{i,j}(xP_i^a)$ calculated by base classifier H_i. Furthermore, count the confidence degree that x is assigned to each class, the method is as follows:

$$\gamma_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(xP_i^a). \tag{14}$$

The final category of sample x will be given according to the degree.

3. Results and Discussion

3.1. Evaluation Criteria. For rigorously verifying the model performance, the criteria, viz. accuracy (Acc.), precision (Prec.), sensitivity (Sen.), specificity (Spec.), Matthews correlation coefficient (MCC), and F1-Score (F1) are employed to analyze the five-fold cross-validation results. These criteria are calculated by the following formulas:

$$\begin{aligned}
\text{Acc.} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},\\
\text{Sen.} &= \frac{\text{TP}}{\text{TP} + \text{TN}},\\
\text{Pre.} &= \frac{\text{TP}}{\text{FP} + \text{TP}},\\
\text{Spec.} &= \frac{\text{TN}}{\text{TP} + \text{FP}},\\
\text{MCC} &= \frac{\text{TN} \times \text{TP} - \text{FN} \times \text{FP}}{\sqrt{(\text{TN} + \text{FN}) \times (\text{TP} \times \text{FP}) \times (\text{TN} + \text{FP}) \times (\text{TP} \times \text{FN})}},\\
F1 &= \frac{2 \times \text{Pre.} \times \textit{Sen.}}{\text{Pre.} + \textit{Sen.}},
\end{aligned} \tag{15}$$

where true positive (TP) represents the sum of associated lncRNA-miRNA pairs which are detected as positive samples; true negative (TN) records the number of nonassociated pairs which are classified as negative samples; false positive (FP) denotes the aggregate of associated lncRNA-miRNA pairs which are inferred as negative samples; false negative (FN) is the count of nonassociated pairs which are predicted as positive samples. In addition, Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves are depicted to visualize the experimental results [39, 40]. The area under the curves (AUC) and area under the PR (AUPR) values are also attached to ROC and PR curves for justifying our model and indicating the sample balance.

3.2. Parameter Discussion. In the experiment, the parameters K and L which represent the numbers of feature subsets and decision trees need to be optimized in the RoF classifier. To obtain the optimal results, the grid-search algorithm is applied to picture the accuracy surface of prediction results with different parameters. There are 2400 experiments with various groups of parameters K and L were conducted. With the increment of the L-value from 0 to 60, the surface indicates that the accuracy is continuously improved. Meanwhile, the accuracy is increased and then dramatically declined with the value rise of parameter K from 0 to 40. For high efficiency, the parameters K and L are set as 26 and 35. Figure 6 shows the prediction accuracy surface with different K-values and L-values.

3.3. Five-fold CV Results on Benchmark Data Sets. To fairly validate the prediction feasibility of SEBGLMA and prevent overfitting and underfitting. We applied the 5-fold cross-validation (CV) method on the benchmark dataset with the same optimal parameters. Specifically, the standard dataset was segmented into five independent parts of the same size. The disjointed subdatasets take turns as a test set, while the other four sets are treated as train sets. The statistics of 5-fold CV results on the benchmark dataset based on SEBGLMA are listed in Table 2. The average accuracy, precision, sensitivity, specificity, Matthews correlation coefficient, and F1-Score are 87.09%, 87.66%, 87.03%, 87.84%, 74.18, and 86.99%. The corresponding standard deviations are 0.59%, 1.02%, 0.87%, 0.65%, 0.12%, and 0.68%. Figures 7 and 8 display the performance of our model by ROC and PR curves, the mean AUC value of 0.9301 and AUPR value of 0.9323 are attached to them.

3.4. Comparison with Other Classifiers. At present, there are many supervised learning classifiers are established for identifying the association between lncRNA and miRNA based on the 5-fold cross-validation method. For further evaluating the performance of rotation forest (RoF) in SEBGLMA. We replaced the RoF with a support vector machine (SVM), deep learning with dual-net neural architecture (DLDP) [41], light gradient boosting machine (LGBM), and random forest (RF). In comparison, the parameters of RoF are set as the values which are optimized above. The inner product kernel is applied to map features into a high dimension in the SVM classifier. Furthermore, the classification process is simplified by the small sample learning mechanism. After optimization, the parameters c and g are set as 0.7 and 38, respectively with radial basis function (RBF) based on the LIBSVM tool. The DLDP classifier is a dual-net natural network combing Feature Importance Ranking (FIR), and Multiple-Layer Perceptron (MLP). We set parameter max batches = 3000 and the feature related parameters into 300, the rest parameters are set as default values. LGBM is developed from gradient boosting decision tree (GBDT) algorithm which saves computing time and cost. After optimization, the leaf number, learning rate, and training rounds were set to 60, 0.05, and 40, respectively. Within the RF classifier, the number of estimators is set as 80. Table 3 gives the results generated by different classifiers.

Figure 9 displays the results of SVM, LGBM, RF, and RoF on a benchmark data set with the same features. The comparison illustrates that the RoF-based model has a better performance than other classifiers. Compared with SVM, the average accuracy, precision, sensitivity, specificity, Matthews correlation coefficient, and F1-Score improved by 16.97%, 16.22%, 16.38%, 17.89%, 27.97%, and 15.78%, respectively. Compared with DLDP, these criteria improved by 15.62%, 11.71%, 30.09%, 15.40%, 30.96%, and 21.09%. Compared with LGBM, these criteria improved by 5.86%, 5.56%, 4.70%, 4.42%, 3.96%, and 5.86%, respectively. Compared with RF, the criteria improved by 1.51%, 1.45%, 0.72%, 2.72%, 2.95%, and 1.30%, respectively. Figures 10 and 11 give the ROC and PR curves of different classifiers. The AUC value gaps between RoF and other models have attained 0.1697, 0.1569, 0.0326, and 0.0164, respectively. The AUPR value gaps have reached 0.1544, 0.1625, 0.0385, and 0.0148, respectively. Thus, the proposed model is more efficient to predict LMAs.

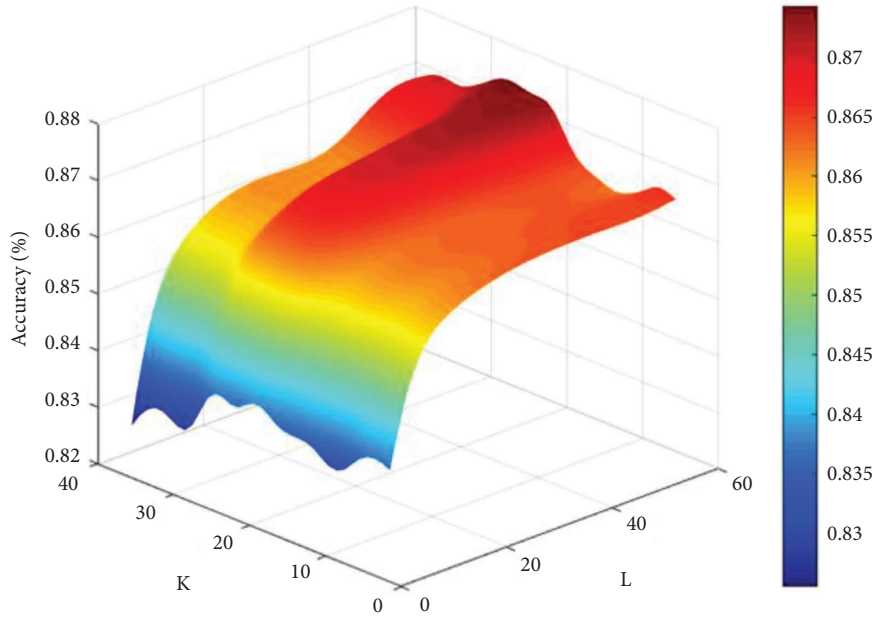

Figure 6: The accuracy surface of the optimization on K-value and L-value.

Table 2: 5-fold cross-validation results on benchmark data set obtained by SEBGLMA.

Test set	Acc. (%)	Pre. (%)	Sen. (%)	Spec. (%)	MCC	F1-score (%)
1	87.92	88.73	86.95	88.90	75.86	87.83
2	87.12	87.97	86.76	87.49	74.23	87.36
3	86.25	87.36	88.51	87.42	72.53	86.22
4	87.11	88.16	86.24	88.01	74.24	87.19
5	87.06	86.06	86.70	87.38	74.06	86.38
Average	87.09 ± 0.59	87.66 ± 1.02	87.03 ± 0.87	87.84 ± 0.65	74.18 ± 0.12	86.99 ± 0.68

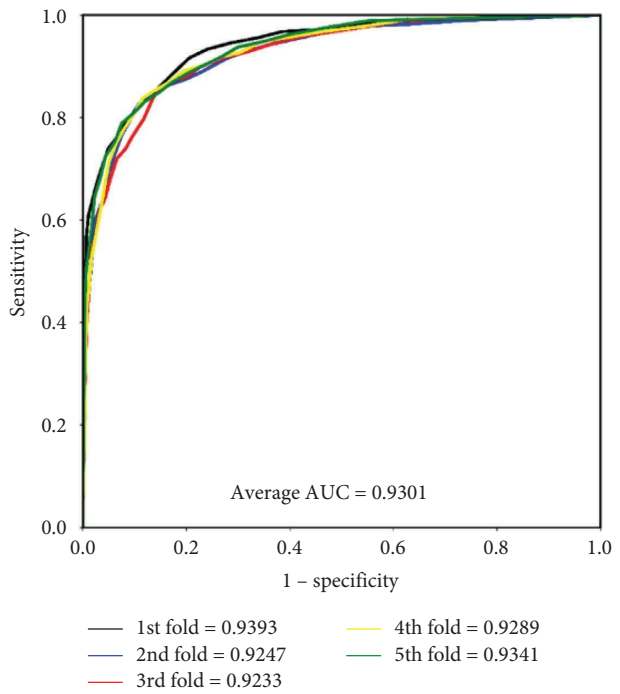

Figure 7: The ROC curves generated by SEBGLMA on benchmark data set.

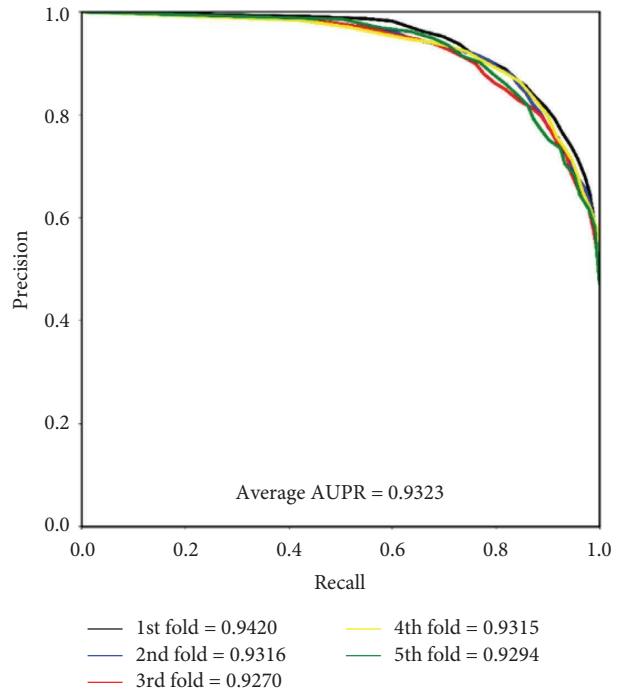

Figure 8: The PR curves generated by SEBGLMA on benchmark data set.

Table 3: 5-fold cross-validation results on benchmark dataset obtained by different classifiers.

Model	Acc. (%)	Pre. (%)	Sen. (%)	Spec. (%)	MCC (%)	F1-score (%)
SVM	70.12	71.44	70.65	69.95	46.21	71.21
DLDP	71.47	75.95	56.94	72.44	43.22	65.90
LGBM	81.23	82.10	82.33	83.42	70.22	81.13
RF	85.55	86.21	86.31	85.12	71.23	85.69
RoF	87.09	87.66	87.03	87.84	74.18	86.99

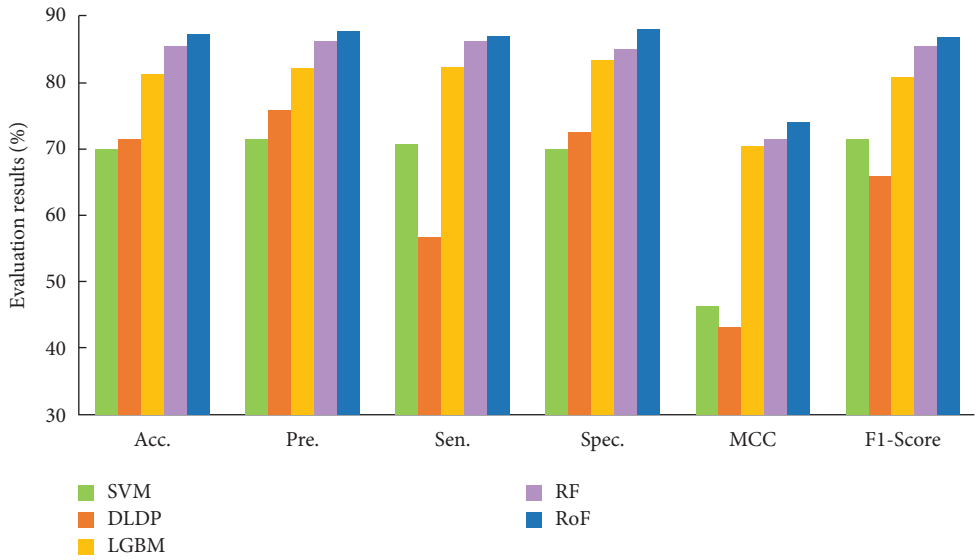

Figure 9: Comparison of various classifiers on benchmark dataset.

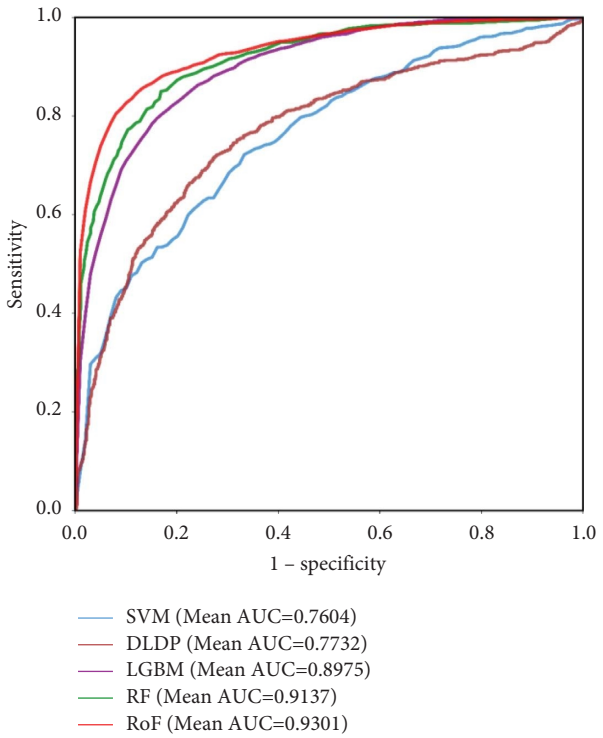

Figure 10: The ROC curves are generated by different classifiers.

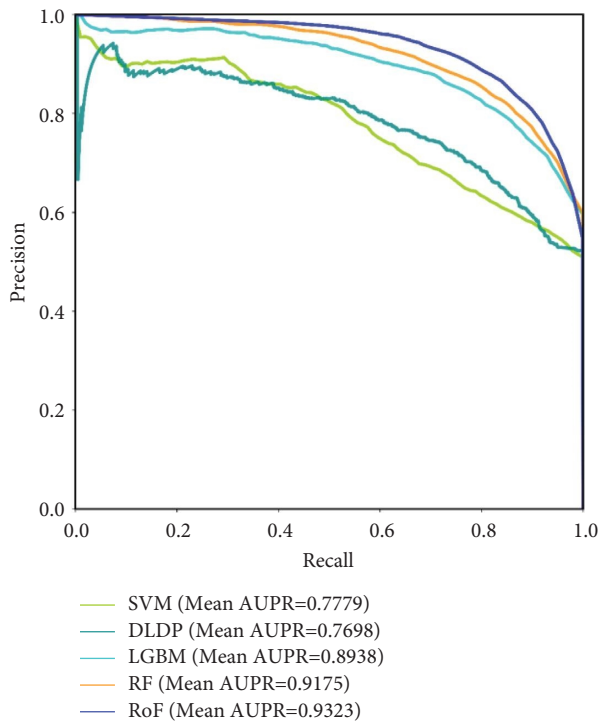

Figure 11: The PR curves are generated by different classifiers.

Table 4: 5-fold cross-validation results on benchmark data set obtained by ablation experiments.

Model	Acc. (%)	Pre. (%)	Sen. (%)	Spec. (%)	MCC (%)	F1-score (%)
RNA2Vec	67.35	67.81	66.07	68.65	34.72	66.92
RNA2Vec + GIP	81.12	82.33	80.46	82.11	46.20	82.31
RNA2Vec + GCN	84.52	85.05	83.79	85.27	69.07	84.40
SEBGLMA	87.09	87.66	87.03	87.84	74.18	86.99

3.5. Ablation Experiment. To evaluate the optimization effect of each improved module, ablation experiments were conducted on the RoF classifier. Table 4 gives the results of ablation experiments. Firstly, we employed RNA2Vec models to extract the attribute features of samples and input them into the classifier as basic experiments. Then, GIP is utilized to extract the self-similarity between the same kind of samples, and it is simply spliced with attribute features for predicting LMAs. Compared with single attribute features, the model which integrates RNA2Vec and GIP differentiates samples by splicing the semantic features and self-similarity features. The evaluation indicators of the model increased by 13.77%, 14.52%, 14.39%, 13.46%, 11.48%, and 15.39%, respectively. After that, the adjacency matrix generated by attribute features and known lncRNA-miRNA associations was fed into the GCN to generate fusing features. The criteria of the model which combines RNA2Vec and GCN were improved by 17.17%, 17.24%, 17.72%, 16.62%, 34.35%, and 17.48%. Finally, after integrating RNA2Vec, GIP, and GCN, SEBGLMA makes full use of the relationship between nodes to embedding attribute features into behavior features. The performance of the model improved by 19.74%, 19.85%, 20.96%, 19.19%, 39.46%, and 20.18% compared with the single attribute feature. Figures 12 and 13 give the ROC and PR curves of ablation experiments. The AUC value gaps between SEBGLMA and other models have attained 0.2010, 0.0213, and 0.0174, respectively. The AUPR value gaps have reached 0.2012, 0.0170, and 0.0090, respectively. Ablation experiments indicated that each module can effectively improve the performance of SEBGLMA.

3.6. Comparison with Previous Methods. By the time, various previous methods have achieved advanced performances in predicting LMAs. For more sufficiently evaluating the performance of our algorithm, we conducted the comparing experiments with SPMLMI [17], EPLMI [42], LMI-INGI [43], GNMFLMI [44], NDALMA [45], and LMFNRLMI [16]. These models are established on the same benchmark dataset with a 5-fold CV. The LMFNRLMI algorithm utilized the strongest adjacent relation to detect LMAs. The expression profile-based prediction model for lncRNA-miRNA interactions (EPLMI) method has the ability to achieve the interaction possibility of lncRNA and miRNA. Concretely, this method constructs a bipartite graph by collaborative effects and similarities of different nodes. The lncRNA-miRNA interactions based on the interactome network and graphlet interaction (LMI-INGI) model divides the complex network into subgraphs, which is called graphlet interaction isomers. The subgraphs have 9 types of structures containing 4 nodes to differently describe the nodes. The graph regularized non-negative matrix factorization for predicting lncRNA-miRNA interactions (GNMFLMI) model constructs affinity graphs by p-nearest neighbors, and scores the interacting possibility by graph regularized non-negative matrix factorization. The SPMLMI algorithm established a two-layer network for inferring the associations between lncRNAs and miRNAs. Table 5 lists the AUC values of different methods. The AUC values of the previous models are 0.9220, 0.8447, 0.8957, 0.8894, 0.8810, and 0.8767, respectively. Our method achieves a higher value of 0.9301 compared with these advanced methods. Hence, the proposed model has the potential to become a feasible and effective tool to identify LMAs.

3.7. Case Study. In this work, case studies are conducted to further verify the predictive performance of SEBGLMA. We rescreened 7263 lncRNA-miRNA interaction pairs related to 419 lncRNAs and 263 miRNAs in lncRNASNP2 to prevent accidental results. Subsequently, we took hsa-miR-497-5P and NONHSAT022145.2 as case studies. Liver cancer is a disease with a high incidence rate and is not easy to be found in the early stage. The miRNA hsa-miR-497-5P has a significant correlation with the occurrence of liver cancer by interacting with lncRNAs. Therefore, our model is established to detect the potential interacting lncRNAs for the target miRNA. In this case study, 122 case-related positive samples were subtracted from 7263 positive samples to obtain 7141 unrelated positive samples from lncRNASNP2. Subsequentially, 419 case-related samples were screened as the test set. Finally, the unrelated positive samples and the test data set are subtracted to obtain the final negative sample data set. The same number of negative samples are collected from the remaining 102637 $(419 \times 263 - 7141 - 419)$ pairs without validations. The prediction results are sorted in descending order according to the prediction score. Table 6 shows the top 30 predicted interactions. There are 24 of the top 30 candidate lncRNAs were confirmed by lncRNASNP2.

Regarding NONHSAT022145.2, it has been certificated to have an influence on the changes of renal cancer cells by regulating miRNAs. In this case study, 120 case-related positive samples were subtracted from 7263 positive samples to obtain 7143 unrelated positive samples from lncRNASNP2. Subsequentially, 263 case-related samples were screened as a test set. Finally, the unrelated positive samples and the test data set are subtracted to obtain the final negative sample data set. The same number of negative samples are obtained from the remaining 102791 $(419 \times 263 - 7143 - 263)$ pairs without validations. After sorting by prediction scores, the top 30 predicted interactions are listed in Table 7. Within the results, there are 27 of the top 30

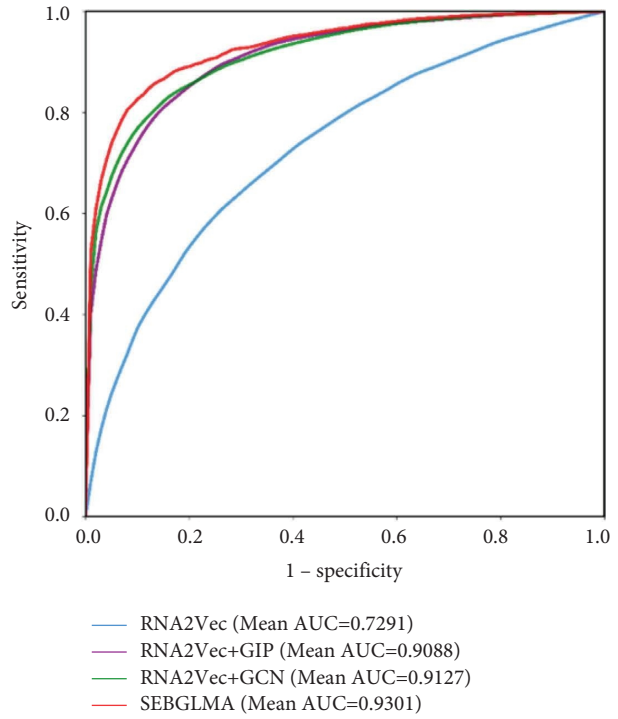

Figure 12: The ROC curves obtained by ablation experiments.

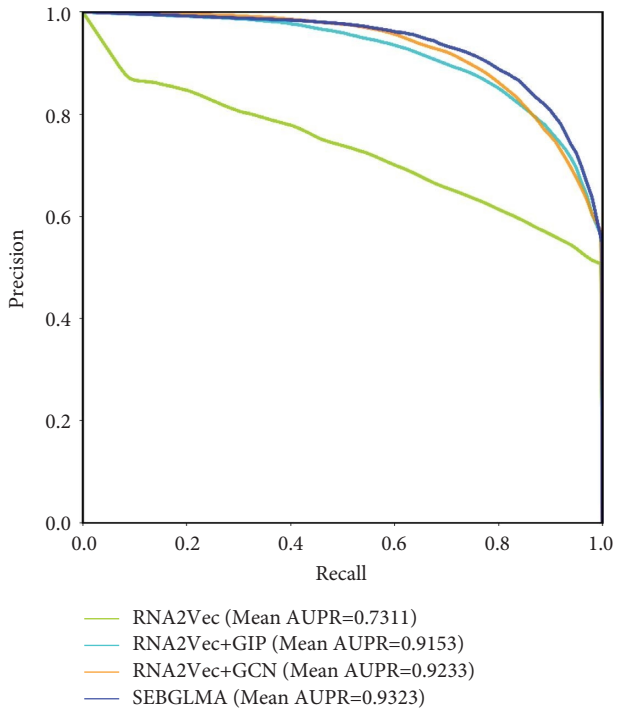

Figure 13: The PR curves obtained by ablation experiments.

Table 5: Comparison between our model with state-of-art methods in terms of benchmark data sets.

Model	AUC
SPMLMI	0.8767
EPLMI	0.8447
LMI-INGI	0.8957
GNMFLMI	0.8894
NDALMA	0.8810
LMFNRLML	0.9220
Our method	**0.9301**

The bold value represents the highest AUC value in the comparison.

Table 6: Top 30 lncRNAs associated with has-miR-497-5P predicted by SEBGLMA.

Rank	LncRNA	Evidence
1	NONHSAT137542.2	lncRNASNP2
2	NONHSAT137558.2	lncRNASNP2
3	NONHSAT137541.2	lncRNASNP2
4	NONHSAT137559.2	lncRNASNP2
5	NONHSAT137541.2	lncRNASNP2
6	NONHSAT022125.2	Unconfirmed
7	NONHSAT001468.2	lncRNASNP2
8	NONHSAT022132.2	Unconfirmed
9	NONHSAT022145.2	Unconfirmed
10	NONHSAT055703.2	lncRNASNP2
11	NONHSAT055688.2	lncRNASNP2
12	NONHSAT055676.2	lncRNASNP2
13	NONHSAT055683.2	lncRNASNP2
14	NONHSAT055670.2	lncRNASNP2
15	NONHSAT055705.2	Unconfirmed
16	NONHSAT108321.2	Unconfirmed
17	NONHSAT055684.2	lncRNASNP2
18	NONHSAT055690.2	lncRNASNP2
19	NONHSAT055692.2	lncRNASNP2
20	NONHSAT055698.2	lncRNASNP2
21	NONHSAT055694.2	lncRNASNP2
22	NONHSAT108327.2	Unconfirmed
23	NONHSAT055673.2	lncRNASNP2
24	NONHSAT055671.2	lncRNASNP2
25	NONHSAT055700.2	lncRNASNP2
26	NONHSAT021846.2	lncRNASNP2
27	NONHSAT021836.2	lncRNASNP2
28	NONHSAT056045.2	lncRNASNP2
29	NONHSAT001973.2	lncRNASNP2
30	NONHSAT001976.2	lncRNASNP2

Table 7: Top 30 miRNAs associated with NONHSAT022145.2 predicted by SEBGLMA.

Rank	miRNA	Evidence
1	hsa-miR-200c-3p	lncRNASNP2
2	hsa-miR-338-3p	lncRNASNP2
3	hsa-miR-429	lncRNASNP2
4	hsa-miR-140-5p	lncRNASNP2
5	hsa-miR-378f	lncRNASNP2
6	hsa-miR-599	Unconfirmed
7	hsa-miR-370-3p	lncRNASNP2
8	hsa-miR-181b-5p	lncRNASNP2
9	hsa-miR-378a-3p	lncRNASNP2
10	hsa-miR-378d	lncRNASNP2
11	hsa-miR-181a-5p	lncRNASNP2
12	hsa-miR-378e	lncRNASNP2

Table 7: Continued.

Rank	miRNA	Evidence
13	hsa-miR-4262	lncRNASNP2
14	hsa-miR-378b	lncRNASNP2
15	hsa-miR-378c	lncRNASNP2
16	hsa-miR-503-5p	lncRNASNP2
17	hsa-miR-320c	lncRNASNP2
18	hsa-miR-495-3p	lncRNASNP2
19	hsa-miR-181d-5p	lncRNASNP2
20	hsa-miR-194-5p	lncRNASNP2
21	hsa-miR-138-5p	Unconfirmed
22	hsa-miR-378i	lncRNASNP2
23	hsa-miR-181c-5p	lncRNASNP2
24	hsa-miR-320d	lncRNASNP2
25	hsa-miR-320a	lncRNASNP2
26	hsa-miR-320b	lncRNASNP2
27	hsa-miR-320c	lncRNASNP2
28	hsa-miR-4429	lncRNASNP2
29	hsa-miR-216a-5p	lncRNASNP2
30	hsa-miR-193a-3p	Unconfirmed

candidate miRNAs have been validated. The case studies illustrate that SEBGLMA can effectively predict LMAs.

4. Conclusion

In general, this paper provides a novel mechanism integrating *K*-mer segmentation, word2vec, Gaussian interaction profile, graph convolution network, and rotation forest to infer the LMAs. Specifically, the *K*-mer segmentation and word2vec are utilized to extract the attribute features of RNAs. Then, the adjacency matrix is completed by combining the GIP self-similarity and the known relationships between lncRNAs and miRNAs. Sequentially, the attribute features and the completed adjacency matrix are sent into GCN for embedding behavior characteristics. Finally, the fusion features are fed into RoF for identifying LMAs. The mean accuracy, precision, sensitivity, specificity, Matthews correlation coefficient, and F1-Score of the proposed model were 87.09%, 87.66%, 87.03%, 87.84%, 74.18%, and 86.99%, respectively. For ensuring the advancement of our model, we also systematically conducted comparisons. First of all, the classifier was altered by SVM, DLDP, LGBM, and RF to validate the performance of RoF. Secondly, the ablation experiments are carried on to prove the optimization efficiency of each module. Finally, many state-of-art methods are employed to evaluate the prediction performance. The comparisons indicate that our model can be a robust and efficient tool to screen reliable candidates for clinical trials.

5. Limitation and Feature Work

Besides improving the accurate prediction ability of the model, the limitations of the proposed model are also noticed. The limitations focusing on two aspects will be illustrated in this section. On the one hand, the graph convolution network only considers the mutual relationships between the target nodes and the directly connected primary nodes to obtain the local behavior feature information. It is hardly to achieve the global structure information of the target node. In future work, graph neural networks with different distances between the target and neighbor nodes will be constructed at the same time, and these networks will be cascaded to extract the global structure information of target nodes. On the other hand, the noise and feature loss in the preprocessed sequence features will reduce the robustness of the model. We will develop a sequence data denoising algorithm to improve the difference between samples. Generally, the subsequent research will emphasize excavating more robust characteristics with less noise and constructing reliable classifiers. The expansion of the high-throughput database will furnish a data foundation for building complementary identification tools.

Data Availability

https://github.com/zhaozhiya-20/SEBGLMA-semantic-embedded-bipartite-graph-network-for-predicting-lncRNA-miRNA-associations.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant no. 62072378; National Natural Science Foundation of China under Grant no. 62172338; Natural Science Foundation of Shaanxi Province under Grant no. 2022JQ-700; Natural Science Foundation of Shaanxi Province under Grant no. 2021JQ-879.

References

[1] K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Molecular Cell*, vol. 43, no. 6, pp. 904–914, 2011.

[2] M. W. Szcześniak, M. R. Kubiak, E. Wanowska, and I. Makalowska, "Comparative genomics in the search for conserved long noncoding RNAs," *Essays in Biochemistry*, vol. 65, no. 4, pp. 741–749, 2021.

[3] A. L. Jackson and A. A. Levin, "Developing microRNA therapeutics: approaching the unique complexities," *Nucleic Acid Therapeutics*, vol. 22, no. 4, pp. 213–225, 2012.

[4] Z. W. Li, T. B. Zhong, D. S. Huang, Z. H. You, and R. Nie, "Hierarchical graph attention network for miRNA-disease association prediction," *Molecular Therapy*, vol. 30, no. 4, pp. 1775–1786, 2022.

[5] G. K. Bhatti, N. Khullar, I. S. Sidhu et al., "Emerging role of non-coding RNA in health and disease," *Metabolic Brain Disease*, vol. 36, no. 6, pp. 1119–1134, 2021.

[6] K. Su, N. Wang, Q. Shao, H. Liu, B. Zhao, and S. Ma, "The role of a ceRNA regulatory network based on lncRNA MALAT1 site in cancer progression," *Biomedicine & Pharmacotherapy*, vol. 137, Article ID 111389, 2021.

[7] T. B. Zhong, Z. W. Li, Z. H. You, R. Nie, and H. Zhao, "Predicting miRNA-disease associations based on graph random propagation network and attention network," *Briefings in Bioinformatics*, vol. 23, no. 2, Article ID bbab589, 2022.

[8] J. Venkatesh, M. C. D. Wasson, J. M. Brown, W. Fernando, and P. Marcato, "LncRNA-miRNA axes in breast cancer: novel points of interaction for strategic attack," *Cancer Letters*, vol. 509, pp. 81–88, 2021.

[9] M. Huang, Y. Long, Y. Jin et al., "Comprehensive analysis of the lncRNA-miRNA-mRNA regulatory network for bladder cancer," *Translational Andrology and Urology*, vol. 10, no. 3, pp. 1286–1301, 2021.

[10] Y. Wang, J. Zhang, and S. Zheng, "The role of XBP-1-mediated unfolded protein response in colorectal cancer progression-a regulatory mechanism associated with lncRNA-miRNA-mRNA network," *Cancer Cell International*, vol. 21, no. 1, p. 488, 2021.

[11] H. Y. Zhang, K. Cai, J. Wang et al., "MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by down-regulating the STAT3 pathway," *Stem Cells*, vol. 32, no. 11, pp. 2858–2868, 2014.

[12] A. N. Kallen, X. B. Zhou, J. Xu et al., "The imprinted H19 lncRNA antagonizes let-7 microRNAs," *Molecular Cell*, vol. 52, no. 1, pp. 101–112, 2013.

[13] X. S. Wang, X. Chen, Y. A. Huang, Z. H. You, and G. Y. Yan, "A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases," *Bioinformatics*, vol. 33, no. 5, pp. 733–739, 2017.

[14] H. Yang, H. Zhu, and J. G. Ibrahim, "MILFM: multiple index latent factor model based on high-dimensional features," *Biometrics*, vol. 74, no. 3, pp. 834–844, 2018.

[15] W. Zhang, G. Tang, S. Zhou, and Y. Niu, "LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination," *BMC Genomics*, vol. 20, no. S11, Article ID 946, 2019.

[16] Z. A. Huang, Y. A. Huang, Z. H. You, Z. Zhu, and Y. Sun, "Novel link prediction for large-scale miRNA-lncRNA interaction network in a bipartite graph," *BMC Medical Genomics*, vol. 11, no. S6, pp. 113–127, 2018.

[17] M. Xu, Y. Chen, W. Lu et al., "SPMLMI: predicting lncRNA–miRNA interactions in humans using a structural perturbation method," *PeerJ*, vol. 9, Article ID e11426, 2021.

[18] H. S. Liu, G. F. Ren, H. Y. Chen, Q. Liu, Y. Yang, and Q. Zhao, "Predicting lncRNA–miRNA interactions based on logistic matrix factorization with neighborhood regularized," *Knowledge-Based Systems*, vol. 191, Article ID 105261, 2020.

[19] J. Gong, W. Liu, J. Zhang, X. Miao, and A. Y. Guo, "lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse," *Nucleic Acids Research*, vol. 43, pp. D181–D186, 2015.

[20] P. J. Volders, K. Helsens, X. W. Wang et al., "LNCipedia: a database for annotated human lncRNA transcript sequences and structures," *Nucleic Acids Research*, vol. 41, no. D1, pp. D246–D251, 2013.

[21] A. Kozomara and S. Griffiths-Jones, "miRBase: annotating high confidence microRNAs using deep sequencing data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D68–D73, 2014.

[22] H. Li, Y. Q. Wang, Z. Zhang et al., "Identifying microbe-disease association based on a novel back-propagation neural network model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, pp. 2502–2513, 2021.

[23] Z. Ma, Z. Kuang, and L. Deng, "CRPGCN: predicting circRNA-disease associations using graph convolutional network based on heterogeneous network," *BMC Bioinformatics*, vol. 22, no. 1, pp. 551–232021, 2021.

[24] L. Shen, F. Liu, L. Huang et al., "VDA-RWLRLS: an anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares," *Computers in Biology and Medicine*, vol. 140, Article ID 105119, 2022.

[25] X. Chen, C. C. Yan, X. Zhang, Z. H. You, Y. A. Huang, and G. Y. Yan, "HGIMDA: heterogeneous graph inference for miRNA-disease association prediction," *Oncotarget*, vol. 7, no. 40, pp. 65257–65269, 2016.

[26] H. C. Yi, Z. H. You, L. Cheng et al., "Learning distributed representations of RNA and protein sequences and its application for predicting lncRNA-protein interactions," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 20–26, 2020.

[27] H. C. Yi, Z. H. You, D. S. Huang, X. Li, T. H. Jiang, and L. P. Li, "A deep learning framework for robust and accurate prediction of ncRNA-protein interactions using evolutionary information," *Molecular Therapy - Nucleic Acids*, vol. 11, pp. 337–344, 2018.

[28] Y. Bussi, R. Kapon, and Z. Reich, "Large-scale k-mer-based analysis of the informational properties of genomes, comparative genomics and taxonomy," *PLoS One*, vol. 16, no. 10, Article ID e0258693, 2021.

[29] T. T. D. Nguyen, V. N. Trinh, N. Q. K. Le, and Y. Y. Ou, "Using k-mer embeddings learned from a Skip-gram based neural network for building a cross-species DNA N6-methyladenine site prediction model," *Plant Molecular Biology*, vol. 107, no. 6, pp. 533–542, 2021.

[30] S. Bankapur and N. Patil, "An enhanced protein fold recognition for low similarity datasets using convolutional and skip-gram features with deep neural network," *IEEE Transactions on NanoBioscience*, vol. 20, no. 1, pp. 42–49, 2021.

[31] L. Wang, W. Pan, Q. H. Wang et al., "A modified skip-gram algorithm for extracting drug-drug interactions from AERS reports," *Computational and Mathematical Methods in Medicine*, vol. 2020, no. 8, Article ID 1747413, 9 pages, 2020.

[32] X. Song, F. Zhou, A. F. Frangi et al., "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Medical Image Analysis*, vol. 69, no. 3, Article ID 101947, 2021.

[33] P. Xuan, S. X. Pan, T. G. Zhang, Liu, and Sun, "Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations," *Cells*, vol. 8, no. 9, pp. 1012–1028, 2019.

[34] H. T. Fu, F. Huang, X. Liu, Y. Qiu, and W. Zhang, "MVGCN: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks," *Bioinformatics*, vol. 38, no. 2, pp. 426–434, 2022.

[35] Q. M. Yuan, J. W. Chen, H. Y. Zhao, Y. Zhou, and Y. Yang, "Structure-aware protein–protein interaction site prediction using deep graph convolutional network," *Bioinformatics*, vol. 38, no. 1, pp. 125–132, 2021.

[36] J. Rodriguez, L. Kuncheva, and C. Alonso, "Rotation forest: a new classifier ensemble method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.

[37] L. Wei, R. Su, B. Wang, X. Li, Q. Zou, and X. Gao, "Integration of deep feature representations and handcrafted features to improve the prediction of N6-methyladenosine sites," *Neurocomputing*, vol. 324, no. 3–9, pp. 3–9, 2019.

[38] J. Shanmugasundaram, G. Raichal, G. Dency Flora, P. Rajasekaran, and V. Jeevanantham, "Classification of epileptic seizure using rotation forest ensemble method with 1D-LBP feature extraction," *Materials Today Proceedings*, vol. 57, no. 5, pp. 2190–2194, 2022.

[39] A. C. J. W. Janssens, F. K. Martens, and F. K. Martens, "Reflection on modern methods: revisiting the area under the ROC Curve," *International Journal of Epidemiology*, vol. 49, no. 4, pp. 1397–1403, 2020.

[40] L. Wang and C. Zhong, "gGATLDA: lncRNA-disease association prediction based on graph-level graph attention network," *BMC Bioinformatics*, vol. 23, no. 1, pp. 11–24, 2022.

[41] P. Lihong, C. Wang, X. Tian, L. Zhou, and K. Li, "Finding lncRNA-protein interactions based on deep learning with dual-net neural architecture," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2022, p. 1, Article ID 3116232, 2021.

[42] 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, 2018.

[43] L. Zhang, T. Liu, H. Y. Chen, Q. Zhao, and H. Liu, "Predicting lncRNA–miRNA interactions based on interactome network and graphlet interaction," *Genomics*, vol. 113, no. 3, pp. 874–880, 2021.

[44] 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.

[45] L. Zhang, P. Yang, H. Feng, Q. Zhao, and H. Liu, "Using network distance analysis to predict lncRNA–miRNA interactions," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 13, no. 3, pp. 535–545, 2021.