Disease-lncRNA Associations Prediction based on Fast Random Walk with Restart in Heterogeneous Networks

Jinlong Ma, and Tian Qin

***Abstract*—Long non-coding RNAs (lncRNAs) represent a fundamental category of epigenetic modulators. Recent research has revealed that lncRNAs play critical roles in gene regulatory mechanisms, substantially influencing the pathogenesis of various human diseases. In this study, a multilayer heterogeneous network was created and we introduced the fast random walk with restart (FRWR) for predicting connections between lncRNAs and diseases. By combining the similarity network of lncRNA, similarity network of disease, and association network of existing lncRNA-disease, a multilayer heterogeneous network was constructed, and the fast random walk with restart method (FRWR) was applied on this network to predict additional potential lncRNA-disease associations. The AUROC value of 0.9034, achieved through leave-one-out cross-validation, underscored the predictive precision of the FRWR technique. Furthermore, a case study of three different diseases provided further validation of the reliability of prediction results. Overall, the multilayer network FRWR method proposed in this work could effectively forecast the connections between lncRNAs and diseases, offering valuable insights into understanding the functions of lncRNAs in the context of human health and disease. The source code for the FRWR method can be accessed at: https://github.com/TianTianTian14/FRWR.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8836

***Index Terms*—lncRNA; Disease; Heterogeneous networks; Network propagation algorithm.**

I. Introduction

The role of epigenetics in regulating gene expression has increasingly attracted attention [1]. There is a growing of evidences that regulatory non-coding RNAs play an important role in epigenetic control. lncRNAs refer to a class of RNA molecules that do not undergo translation into proteins, they regulate gene expression by interacting with DNA and proteins [2]- [3]. lncRNAs exert their influence through multiple mechanisms, including binding to proteins to modulate their functions, directly regulating gene transcription, recruiting epigenetic modifiers to alter chromatin states, and affecting mRNA stability and translation [4]$^-$ [6]. These diverse functions position lncRNAs as pivotal players in cellular development, differentiation, and disease pathogenesis. With advancing understanding of their roles, lncRNAs are emerging as potential tools in disease diagnostics and therapeutics [7]$^-$ [8]. Many studies have revealed that only 2% of mammals' transcriptional level consists of ribosomal RNA and is translated into proteins. Approximately 98% of the genome consists of non-coding RNA sequences [9]$^-$ [12]. Long non-coding RNAs (lncRNAs) are defined as RNA molecules within this non-coding segment that exceed 200 nucleotides. accounting for the majority of human total genes [13]$^-$ [15]. In recent years, increasing research has found that lncRNAs play important regulatory roles in physiological and pathological processes. For example, MEG3 was a tumor suppressor gene with reduced expression. It can induce apoptosis by down-regulating the expression of BCL-2 [16]. Another common type of lncRNA is the 'pseudogenes' of small RNAs that completed gene regulatory functions by competing with the antisense binding sites of small RNAs [17]$^-$ [18]. BC200 RNA also plays a role in the occurrence and repair of stimulated neuronal injury [19]. Brain cytoplasmic RNA 1 antisense (BACE1-AS) increase Alzheimer's disease-related protein levels by promoting BACE1 demethylation [20]. In summary, lncRNAs are an important component of epigenetic regulatory mechanisms [21]. Further study of lncRNAs will help reveal their key roles in human health and disease [22].

To tackle these challenges, we have introduced a novel computational method that leverages up-to-date data from the LncRNADisease V2.0 database [28] to forecast potential links between lncRNA-diseases. Three networks were designed and integrated into a multilayer network, including the similarity network of lncRNA, similarity network of disease, and known association network of lncRNA-disease. Subsequently, we utilized the fast random walk with restart method (FRWR) to predict potential associations between lncRNAs and diseases. By incorporating disease similarity and known lncRNA-disease relevance, we updated the lncRNA-disease interaction weight matrix for newly identified diseases. FRWR achieved a robust area under the receiver operating characteristic curve (AUROC) value of 0.9034, indicating reliable predictive performance. Additionally, case studies on liver, gallbladder, and pancreatic cancers were performed, leading to the discovery of the top 15 potential lncRNA-disease relationships for each condition. The primary contribution of this study is the development of a novel computational method that integrates the similarity networks of lncRNAs and diseases along with known lncRNA-disease associations to predict potential lncRNA-disease associations through a multilayer heterogeneous network. Additionally, this paper employs the

J. Ma, and T. Qin are with the School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei, China, 050018; Jinlong Ma is also with Hebei Technology Innovation Center of Intelligent IoT, Shijiazhuang, Hebei 050018, China (e-mails: mzjinlong@163.com, and qttt144@163.com).

fast random walk with restart method (FRWR), which effectively combines disease similarity and known lncRNA-disease associations to update the interaction weight matrix for newly discovered diseases, significantly enhancing the accuracy and reliability of predictions.

II. Related Work

In recent years, the prediction of associations between lncRNAs and diseases has received considerable attention, leading to the development of various computational methods to explore potential links. Sun et al. proposed the RWRlncD method, employing a random walk strategy on an lncRNA functional network to uncover new associations [24]. However, this method relies on network construction, which may overlook some connections due to feature selection limitations. Yu et al. introduced the BRWLDA method, utilizing a bidirectional random walk to predict potential links between lncRNAs and diseases, but this method's effectiveness is constrained by the dependency on extensive existing association data [25]. Wang et al. suggested the LDGRNMF approach, which uses graph-regularized non-negative matrix factorization to maintain network structure [26], yet it depends on the quality of known models for predicting new associations. Zeng et al. offered the SDLDA method, which combines singular value decomposition with deep learning to capture complex relationships [27], though it doesn't consider network topologies. Lastly, Lu and colleagues developed the SIMCLDA method, which focuses on similarity-based inductive matrix completion that relies on existing sample quality without needing matrix decomposition [28]. Despite the capabilities of methods like RWRlncD and BRWLDA, their dependency on precise feature selection and data completeness limits their broader application. To facilitate clearer expression, we have created the following table:

TABLE I
COMPARISON OF THE METHODS FOR PREDICTING LNCRNA-DISEASE ASSOCIATIONS

Method	Key Feature	Strength	Limitation
RWRlncD	Stochastic navigation in lncRNA networks.	Mines network signals effectively.	Relies on precise feature selection.
BRWLDA	Bidirectional random walk in heterogeneous networks.	Enhances signal propagation.	Needs extensive association data.
LDGRNMF	Non-negative matrix factorization with graph regularization.	Maintains network structure.	Dependent on model quality.
SDLDA	Deep learning and SVD integration.	Captures complex relationships.	Ignores network structure.
SIMCLDA	Similarity-based inductive matrix completion.	Fits high nonlinearity.	Relies on sample quality and coverage.

III. Materials and Method

A. Dataset

By filtering out duplicate and invalid information from the most recent 2019 release of the LncRNADisease V2.0 database, we curated a standard dataset consisting of 12,865 experimentally verified lncRNA-disease associations between 3,701 lncRNAs and 486 medical conditions. The lncRNA-disease association network and the corresponding adjacency matrix (LD) were developed. In matrix LD, an entry $LD(l_u, d_v)$ is set to 1 to indicate a recorded association between lncRNA l_u and disease d_v, and to 0 when no such association exists.

B. Disease Similarity Network

1) Disease Semantic Similarity: Disease semantic similarity is calculated using Medical Subject Headings (MeSH) descriptors, which are divided into 16 categories. Category C focuses on disease relationships in a directed acyclic graph (DAG) (supplementary material available at https://github.com/TianTianTian14/FRWR), covering a broad range of classifications and more specific conditions to ensure a comprehensive analysis of disease relationships [29]$^-$ [30].

Let N denote a node of disease. $DAG(N) = (N, T(N), E(N))$ can be constructed, where $T(N)$ represents the set comprising the disease in question and its ancestral diseases, $E(N)$ refers to the set of all direct edges that are associated with the specific disease. Two respective equations are defined as follows to calculate the semantic contribution of disease d to disease node N, denoted as $D_N(d)$, which represents the measure of how much disease d contributes to the semantic information of disease node N. Similarly, the semantic contribution of disease node N itself, denoted as $D_N(N)$, quantifies the extent to which disease node N contributes to its own semantic information.

$$\begin{cases} D_N(N) = 1 \\ D_N(d) = \max\left\{\Delta * D_N\left(d'\right) \mid d' \in d\right\}, \text{ if } d \neq N \end{cases} \tag{1}$$

The semantic contribution factor Δ denotes the edge $E(N)$ connecting disease d with its offspring disease d' in the Disease Directed Acyclic Graph (DAG) associated with disease node N. Within the DAG of disease node N, the disease node N itself is considered the most specific disease. The semantic contribution of disease node N is defined as 1, indicating its highest level of relevance and specificity within the DAG. Here, we assume that ancestor nodes farther away from disease node N have lower degrees of association, which means the semantic contribution factor Δ for ancestor nodes is less than 1 but greater than 0. In this study, we set it to 0.5 [37], [39]$^-$ [40]. According to equation (1), the semantic value $DV(N)$ of disease node N is determined as:

$$DV(N) = \Sigma_{d \in T_N} D_N(d) \tag{2}$$

Disease semantic similarity is measured based on the relative positional relationships between diseases in the MeSH descriptor DAG. [31]$^-$ [33] When two diseases exhibit a greater degree of similarity in their Disease DAG, they are considered

to be more semantically similar. The semantic similarity S for a pair of diseases d_i and d_j is delineated as:

$$S\left(N_1, N_2\right) = \frac{\sum_{d \in T_{N_1} \cap T_{N_2}} \left(D_{N_1}(d) + D_{N_2}(d)\right)}{DV\left(N_1\right) + DV\left(N_2\right)} \quad (3)$$

In Equation (3), $D_{N_1}(d)$ represents the semantic value of the relationship between disease d and disease node N_1 in the DAG. $D_{N_2}(d)$ represents the semantic value of the relationship between disease d and disease node N_2. Equation (3) assesses the similarity in meaning between a pair of disease entities by taking into account their locations within the DAG and their connections to preceding diseases in terms of semantics.

2) Disease Gaussian interaction profile kernel similarity: The Gaussian interaction profile kernel operates on the premise that lncRNAs exhibiting greater functional resemblances tend to have associations with analogous diseases. To assess network topological similarity between diseases, the kernel integrates topological data obtained from the established network of lncRNA-disease associations [33]$^-$ [34]. Within this context, the binary vector $IP(d(i))$, corresponding to the i-th column in the matrix LD, serves to depict the relationship between the disease $d(i)$ and every lncRNA. Equation (4) denotes the Gaussian kernel similarity KD for diseases $d(i)$ and $d(j)$, determined through the interaction profiles of diseases $d(i)$ and $d(j)$.

$$KD\left(d_i, d_j\right) = \exp\left(-\gamma_d \left\|IP\left(d_i\right) - IP\left(d_j\right)\right\|^2\right), \quad (4)$$

where the parameter γ_d serves as the tuning parameter for the bandwidth of the kernel. The KD matrix illustrates the similarity through the Gaussian interaction profile kernel among all diseases. An updated bandwidth parameter, γ_d', calculated by dividing it by the average number of associations between diseases and lncRNAs, is required for adjusting γ_d. Informed by prior studies on lncRNA-disease association predictions, the value of γ_d' is established at 1 to regulate the bandwidth of the kernel. [33], [35]$^-$ [36] Therefore, γ_d can be expressed as:

$$\gamma_d = \gamma_d' / \left(\frac{1}{nd}\sum_{i=1}^{nd}\left\|LP\left(d_i\right)\right\|^2\right) \quad (5)$$

3) Disease Integrated Similarity: Integrating disease semantic similarity with Gaussian interaction profile kernel similarity for diseases yielded the aggregated disease similarity matrix, denoted as DD. Here, DS represents the collection of diseases associated with disease semantic similarity. Therefore, $DD(d_i, d_j)$ is defined as:

$$DD\left(d_i, d_j\right)\begin{cases} \frac{S(d_i,d_j)+KD(d_i,d_j)}{2}, i, j \in DS \\ KD\left(d_i, d_j\right), i, j \notin D_S \end{cases} \quad (6)$$

C. LncRNA Similarity Measurement

1) LncRNA Functional Similarity: The functional similarity of lncRNAs is evaluated using the LNCSIM model [37], which measures the semantic similarity among disease categories associated with two different lncRNAs. The disease categories connected to lncRNA R_u and lncRNA R_v are specified as $D(R_u)$ and $D(R_v)$, respectively. This functional similarity between lncRNAs is derived from the semantic comparison between $D(R_u)$ and $D(R_v)$. For instance, the similarity Sim for a disease d_R within the group $D(R_u)$ relative to the group $D(R_v)$ is calculated in the following manner:

$$Sim\left(d_R, D(Rv)\right) = \max_{d \in D(R_u)}\left(S\left(d_k, d\right)\right) \quad (7)$$

Rooted in the principle that lncRNAs with a greater extent of functional resemblance are likely to correlate with comparable diseases, this sentence aims to convey the same meaning by rearranging the structural elements to reduce repetitiveness. The functional similarity matrix F between R_u and R_v is calculated as follows:

$$F = \frac{\sum_{d \in D(R_w)} Sim\left(d, D\left(R_u\right)\right) + \sum_{d \in D(R_v)} Sim\left(d \cdot D\left(R_v\right)\right)}{\left|D\left(R_u\right)\right| + \left|D\left(R_v\right)\right|}, \quad (8)$$

where $|D(R_u)|$ indicates the quantity of diseases linked to lncRNA R_u, and $|D(R_v)|$ signifies the quantity of diseases connected with lncRNA R_v. The functional similarity between lncRNA R_u and R_v is denoted by $F(R_u, R_v)$.

2) LncRNA Gaussian Interaction Profile Kernel Similarity: Reflecting the previously described approach for calculating disease similarity using a Gaussian interaction profile kernel, lncRNA similarity is similarly determined via a Gaussian interaction profile kernel as follows:

$$KL\left(R_u, R_v\right) = \exp\left(-Y_R \left\|IP\left(R_u\right) - IP\left(R_v\right)\right\|^2\right), \quad (9)$$

$$y_R = y_R' / \left(\frac{1}{nR}\sum_{i=1}^{nR}\left\|IP\left(R_u\right)\right\|^2\right) \quad (10)$$

The binary vector $IP(R_u)$, the u-th row in matrix LD, marks the connection between R_u and all diseases, and is captured as the engagement of R_u.

3) LncRNA Integrated Similarity: By combining the functional similarity of lncRNAs with their Gaussian interaction profile kernel similarity, we generated the lncRNA similarity matrix, denoted as LL. The set of lncRNAs linked by functional similarity is represented by RF. Therefore, the measurement of $LL(R_u, R_v)$ is as follows:

$$LL\left(R_u, R_v\right)\begin{cases} \frac{F(R_u,R_v)+KL(R_u,R_v)}{2}, u, v \in RF \\ KL\left(R_u, R_v\right), u, v \notin RF \end{cases} \quad (11)$$

IV. Results

A. FRWR

1) Heterogeneous Network Construction: Three primary networks are designed: an lncRNA similarity network constructed based on functional and expression pattern similarities among lncRNAs; a disease similarity network established according to the common clinical manifestations and genetic expressions of diseases; and an lncRNA-disease association network directly utilizing known association data between lncRNAs and diseases. By integrating these three networks, we constructed a multilayer heterogeneous network that not only reflects the attributes of individual elements but also illustrates the interrelationships among them. Let $R = r_1, r_2, \ldots, r_p$ and $D = d_1, d_2, \ldots, d_q$ denote the sets of p lncRNA nodes and q disease nodes, respectively. The integrated similarity described in Sections 2.2.3 and 2.3.3 is used to assign the

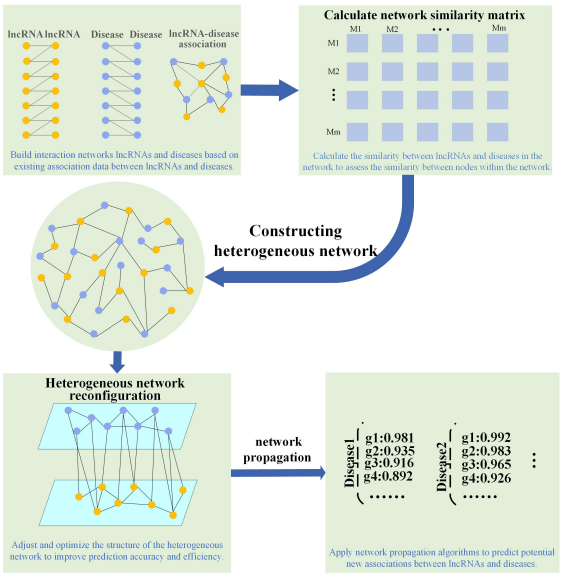

Fig. 1. Workflow diagram. The similarity network of lncRNA, similarity network of disease, and known association network of lncRNA-disease were calculated and integrated into a heterogeneous network. Then, the potential lncRNA-disease associations were predicted by applying the fast random walk with restart method on the heterogeneous network.

weights of edges between r_u and r_v, and d_i and d_j. The association network of lncRNA-disease contains p lncRNAs and q diseases, where the edge weight between lncRNA r_u and disease d_i is set to 1 if they are associated, and 0 otherwise. If the weight is 0, that is, r_u has no recorded association with a new disease d_{unexp}, a novel measure incorporating similarity of disease and association of lncRNA-disease is utilized to recalculate the matrix of lncRNA-disease. The conclusive weights are characterized by the following:

$$\begin{aligned} &LD\left(r_u,\ \text{dunexp}\ \right)=DD(\ \text{dunexp, dj}\ )*LD^{\top}, \\ &\text{if } LD\left(r_u,\ \text{dunexp}\ \right)=0 \end{aligned} \tag{12}$$

Then, a multiple heterogeneous network H is constructed and defined as:

$$H=\begin{bmatrix} L & LD \\ LD^T & D \end{bmatrix}, \tag{13}$$

where L represents the lncRNA similarity network adjacency matrix, D denotes the disease similarity network adjacency matrix, and LD, the lncRNA-disease association matrix, yields LD^T when transposed.

2) Random Walk with Restart on the Heterogeneous network: For the calculation of evaluation scores for potential lncRNAs linked to diseases, the described multi-heterogeneous network undergoes analysis through the use of a network diffusion algorithm [38]. Given the disease network DD, lncRNA network LL, and lncRNA-disease network LD, the following diagonal matrices D_D, D_L, and D_{LD} are defined, where the diagonal elements are defined as $(D_D)\,i,i = \bar{\Sigma}j(DD)j,i;(DL)\,i,i = \Sigma j(LL)j,i;(DLD)\,i,i = \Sigma j(LD)_{j,i}$. Therefore, the normalized matrices of DD, LL, and LD can be written as:

$$\begin{cases} \hat{DD}=DD\cdot D_D^{-1} \\ \hat{LL}=LLD_L^{-1} \\ \hat{LD}=LD\quad D_{LD}^{-1} \end{cases} \tag{14}$$

Based on these normalized matrices, a new matrix is constructed as:

$$\hat{H}=\begin{pmatrix} (1-\beta)\hat{LL} & \beta\hat{LD} \\ \beta\hat{LD}^{\top} & (1-\beta)\hat{DD} \end{pmatrix}, \tag{15}$$

in which the parameter β signifies the transition likelihood between the disease and lncRNA networks. When conducting a random walk, there is a probability denoted by β that the walker will jump from the disease network to the lncRNA network. Alternatively, with a probability of $1-\beta$, the walker will remain within the original network it occupied.

The random walker is only capable of performing inter-layer jumps to nodes in the adjacent network when those connecting nodes exist. Otherwise, if no inter-layer links are present, the walker can only transition to intra-layer neighboring nodes within its current network or return to the originating node. Therefore, a diagonal matrix M is defined such that the elements $(M)_{i,i}$ are equal to the sum of each row j of the normalized association matrix $\hat{H}$, which yields the final transfer matrix $M_T=\hat{H}\cdot M^{-1}$ characterizing the random walk process throughout the dual-layer heterogeneous network.

The process of network propagation via restarting random walks can be characterized as: $P_{i+1}=(1-\alpha)M_T\cdot P_i+\alpha P_0$, where p_{t+1} is the probability vector of the random walker reaching the network nodes, $p_0=(p_{0D}^T,p_{0L}^T)^T$ is the initial probability of the random walk, p_{0D}^T is the initial probability vector in the disease subnetwork, p_{0L}^T is the initial probability vector in the lncRNA subnetwork. The restarting random walk process is characterized by the parameter α, where $\alpha\in(0,1)$ represents the probability of returning to the initial node after several steps. Once the probability p_{t+1} reaches a stable state, genes can be ranked based on this steady probability to predict disease-associated genes using a stable probability approach. According to the original report, α and β were set at 0.7 and 0.8 respectively for purposes of network propagation [39]$^-$ [40].

B. Comparison with other Methods

1) Evaluation Metrics: To comprehensively assess the predictive performance of the FRWR method, this study employs three key performance metrics: the Area Under the Receiver Operating Characteristic Curve (AUROC), the Area Under the Precision-Recall Curve (AUPRC). These metrics not only quantify the model's classification ability across different threshold settings but also evaluate the model's precision in high recall areas, providing a standardized framework for comparing different methods [41].

We incorporated the Top-k Recall metric. Top-k Recall refers to the proportion of actual positives among the top k predicted positives, where k varies. This metric is widely utilized in the fields of medicine and bioinformatics to gauge model accuracy in critical predictive tasks. Specifically, different k

values such as 10, 50, 100, and 200 were set to correspond to varying levels of prediction granularity. For each k, the top k predictions with the highest scores were selected, and the proportion that were correctly identified as positives was calculated. Moreover, the results of the Top-k Recall were used to directly compare the performance of the FRWR method against existing methods. Through this comparison, we could demonstrate the advantages of FRWR in practical applications.

2) Comparison with other Methods: The FRWR method proposed in this article demonstrated exceptional predictive performance in several disease instances. Compared with the following four methods: SIMCLDA, SDLDA, RWRlncD, and BRWLDA, FRWR performed superiorly in both AUROC and AUPRC metrics. Particularly in case studies of liver, gallbladder, and pancreatic cancers, the predicted potential lncRNA-disease associations were validated by two independent databases, further proving the reliability and practicality of the FRWR method. Based on the standard dataset, the AUROC values of our FRWR method and SIMCLDA, SDLDA, RWRlncD, and BRWLDA were shown. The AUROC value of FRWR was 0.9034, SIMCLDA was 0.8245, SDLDA was 0.8692, RWRlncD was 0.8867, and BRRWLDA was 0.8963. The results demonstrated that using the standard dataset, our FRWR method achieved a higher AUROC value in predicting lncRNA-disease associations. Similarly, we also compared the AUPRC values of the five methods, which were 0.1246, 0.1201, 0.1227, 0.1188, and 0.1213 for FRWR, SIMCLDA, SDLDA, RWRlncD, and BRWLDA, respectively, as shown in the figure. As shown in Table 2, the FRWR method also had certain advantages in Top-k Recall values.

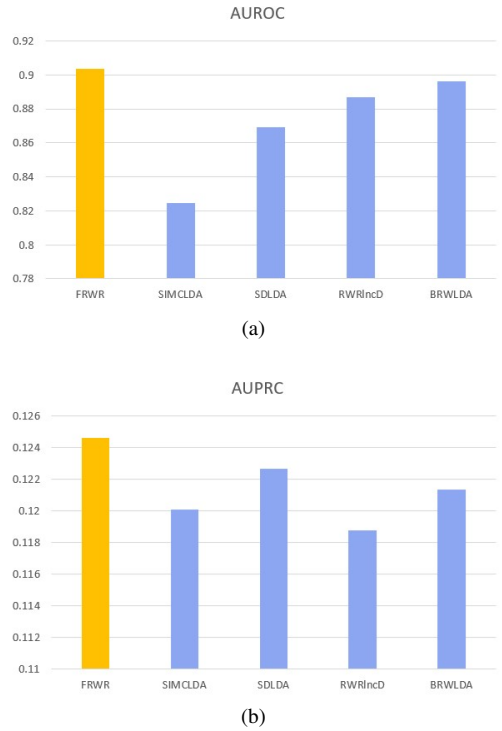

Fig. 2. Evaluating the efficacy of FRWR, focusing on (a) AUROC and (b) AUPRC metrics, and comparing these outcomes with leading-edge techniques.

TABLE II
EVALUATION OF FRWR'S EFFICACY USING TOP-K RECALL (K = 10, 50, 100, AND 200) METRICS, JUXTAPOSED WITH CUTTING-EDGE APPROACHES

Methods	Prec_10	Prec_20	Prec_50	Prec_100	Prec_200
FRWR	**0.227**	**0.287**	**0.359**	**0.418**	**0.499**
SIMCLDA	0.215	0.265	0.337	0.403	0.473
SDLDA	0.220	0.273	0.350	**0.418**	0.487
RWRlncD	0.215	0.271	0.346	**0.418**	0.496
BRWLDA	0.223	0.279	0.349	0.411	0.498

3) Case Studies: Case studies on liver, gallbladder, and pancreatic cancers were performed to demonstrate the predictive efficacy of the FRWR method. Globally, liver cancer is a highly prevalent and lethal malignancy. Its early stages are asymptomatic, while later stages may present with enlarged liver and poor liver function. Treatment methods include surgical resection, thermal therapy, chemotherapy, and targeted therapy, generally resulting in a poor prognosis [42]. Gallbladder cancer is one of the predominant malignancies within the biliary tract. Risk factors include a history of chronic cholecystitis and cholelithiasis. Its early symptoms are not obvious, while later stages may present with jaundice and indigestion [43]. The incidence of pancreatic cancer ranks fourth among causes of cancer deaths. It has an occult onset and asymptomatic early stages, with adenocarcinoma being the primary pathological type [44]. These summaries offer insights into three prevalent malignant neoplasms within the digestive system, each representing significant threats to human health and survival.

Before predicting potential lncRNAs associated with diseases, we removed all known lncRNAs linked to each disease and treated each disease as a novel entity. We then validated the top 15 predictive results using the Lnc2Cancer 3.0 [45] and RNADisease v4.0 [46] databases, which had not been utilized in our prior analyses. These databases primarily served to confirm the accuracy of the newly predicted lncRNA-disease associations. In the standard dataset, all confirmed lncRNA-disease interactions were used as training examples, with yet-to-be-studied associations treated as prospective candidate pairs. From the pool of candidate pairs, we identified the 15 most relevant lncRNAs for liver, gallbladder, and pancreatic cancers.

The majority of the predicted lncRNA-disease associations were corroborated by the Lnc2Cancer 3.0 and RNADisease v4.0 databases, while only a minority of the lncRNAs remained unconfirmed in their disease associations from the case studies. These findings underscore the accuracy of the proposed predictive method.

V. CONCLUSION

In this paper, we provide a detailed overview of various computational methods currently used to predict lncRNA-disease connections, alongside a comparison of their strengths and weaknesses. A multilayer network method named FRWR was proposed for prediction to provide an optimized solution. To elaborate, networks of lncRNA similarity, disease

TABLE III
TOP 15 RELATED GENES OF LIVER CANCER PREDICTED BY FRWR

Top 1-15 gene	Lnc2Cancer 3.0	RNADisease v4.0
MEG3	✓	✓
H19	✓	✓
ZEB1-AS1	✓	✓
ACVR2B-AS1	✓	✓
HOTAIR	✓	✓
UCA1	✓	✓
FAM215A	✓	✓
CDKN2B-AS1	✓	✓
CRNDE	✓	✓
LUCAT1	✓	✓
lnc-DILC	✓	✓
ENST00000425005	✓	✓
DLX6-AS1	✓	✓
BCAR4	✓	✓
HULC	✓	✓

TABLE IV
TOP 15 RELATED GENES OF PANCREATIC CANCER PEDICTED BY FRWR

Top 1-15 gene	Lnc2Cancer 3.0	RNADisease v4.0
HULC	✓	✓
TUG1	✓	✓
BANCR	✓	✓
SNHG16	✓	✓
XIST	✓	✓
SNHG15	✓	✓
LINC00958	✓	✓
ABHD11-AS1	✓	✓
MTA2TR	✓	✓
CCAT1	✓	✓
HOTTIP	✓	✓
MEG8	✓	✓
BCYRN1	✗	✗
MALAT1	✓	✓
BX111	✓	✗

TABLE V
TOP 15 RELATED GENES OF GALLBLADDER CANCER PREDICTED BY FRWR

Top 1-15 gene	Lnc2Cancer 3.0	RNADisease v4.0
TUG1	✓	✓
MINCR	✓	✓
ANRIL	✓	✗
MALAT1	✓	✓
SSTR5-AS1	✓	✓
LOC344887	✓	✓
GBCDRlnc1	✓	✓
SPRY4-IT1	✓	✓
HEGBC	✓	✓
UCA1	✓	✓
CRNDE	✓	✓
PAGBC	✓	✓
TP53COR1	✗	✓
AFAP1-AS1	✓	✓
PVT1	✓	✓

similarity, and established lncRNA-disease associations were developed, creating a multilayer heterogeneous network. The FRWR technique was subsequently applied to this network, leveraging principles of network propagation to deduce probable new associations. Through cross-validation, the FRWR method surpassed other methods on three evaluation metrics, demonstrating its predictive ability. Furthermore, case studies on several typical diseases validated that most prediction results were supported by real databases, showing this method's good reliability and interpretability. This method effectively harnessed data from multiple sources and introduced a structured and refined framework for forecasting lncRNA-disease links. Moving forward, we plan to delve deeper into the precise roles of lncRNAs across a broader spectrum of diseases to facilitate foundational research and its clinical applications.

REFERENCES

[1] C. Wang, L. Wang, Y. Ding et al., "LncRNA Structural Characteristics in Epigenetic Regulation", *Interbational Journal Of Molecular Sciences*, DOI:10.3390/ijms18122659, vol. 18, no. 12, pp. 26-59, 2017.

[2] X. Zhang, W. Wang, W. Zhu et al., "Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels", *Interbational Journal Of Molecular Sciences*, DOI:10.3390/ijms20225573, vol. 20, no. 22, pp. 55-73, 2019.

[3] P. Grote, B. Herrmann, "The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis", *RNA Biology*, DOI:10.4161/rna.26165, vol. 10, no. 10, pp. 1579-1585, 2013.

[4] M. Guttman, I. Amit, M. Garber et al., "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals", *Nature*, DOI:10.1038/nature07672, vol. 458, no. 7235, pp. 223-227, 2009.

[5] P. Johnsson, L. Lipovich, D. Grandér et al., "Evolutionary conservation of long non-coding RNAs; sequence, structure, function", *Biochimica et Biophysica Acta-General Subjects*, DOI:10.1016/j.bbagen.2013.10.035, vol. 1840, no. 3, pp. 1063–1071, 2014.

[6] A. N. Khachane, P. M. Harrison, "Mining mammalian transcript data for functional long non-coding RNAs", *PLoS One*, DOI:10.1371/journal.pone.0010316, vol. 5, no. 4, pp. e10316, 2010.

[7] A. Bhan, S. Mandal, "Long Noncoding RNAs: Emerging Stars in Gene Regulation, Epigenetics and Human Disease", *ChemMedChem*, DOI:10.1002/cmdc.201300534, vol. 9, no. 9, pp. 1932-1956, 2014.

[8] N. Gungaly, S. Chakrabarti, "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals", *Interbational Journal Of Molecular Sciences*, DOI:10.3892/ijmm.2021.4856, vol. 47, no. 3, pp. 23, 2021.

[9] T. R. Mercer, M. E. Dinger, S. M. Sunkin et al., "Specific expression of long noncoding RNAs in the mouse brain", *Proceedings Of The National Academy Of Sciences Of The Nited States Of America*, DOI:10.1073/pnas.0706729105, vol. 105, no. 2, pp. 716–721, 2008.

[10] L. A. Goff, A. F. Groff, M. Sauvageau et al., "Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain", *Proceedings Of The National Academy Of Sciences Of The Nited States Of America*, DOI:10.1073/pnas.1411263112, vol. 112, no. 22, pp. 6855–6862, 2015.

[11] J. Whitehead, G. K. Pandey, C. Kanduri, "Regulation of the mammalian epigenome by long noncoding RNAs", *Biochimica et Biophysica Acta-General Subjects*, DOI:10.1016/j.bbagen.2008.10.007, vol. 1790, no. 9, pp. 936–947, 2009.

[12] J. Ponjavic, P. L. Oliver, G. Lunter et al., "Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain", *PLoS Genetics*, DOI:10.1371/journal.pgen.1000617, vol. 5, no. 8, pp. e1000617, 2009.

[13] M. K. Iyer, Y. S. Niknafs, R. Malik et al., "The landscape of long noncoding RNAs in the human transcriptome", *Nature Genetics*, DOI:10.1038/ng.3192, vol. 47, no. 3, pp. 199–208, 2015.

[14] U. R. Barbara, J. Lagarde, A. Frankish et al., "Towards a complete map of the human long non-coding RNA transcriptome", *Nature Reviews Genetics*, DOI:10.1038/s41576-018-0017-y, vol. 19, no. 9, pp. 535–548, 2018.

[15] Z. Du, T. Fei, R. G. W. Verhaak et al., "Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer", *Nature Structural & Molecular Biology*, DOI:10.1038/nsmb.2591, vol. 20, no. 7, pp. 908-913, 2013.

[16] L. X. Liu, W. Deng, X. T. Zhou et al., "The mechanism of adenosine-mediated activation of lncRNA MEG3 and its antitumor effects in human hepatoma cells", *International Journal Of Oncology*, DOI:10.3892/ijo.2015.3248, vol. 48, no. 1, pp. 421–429, 2016.

[17] M. Vizoso, M. Esteller, "The activatory long non-coding RNA DBE-T reveals the epigenetic etiology of facioscapulohumeral muscular dystrophy", *Cell Research*, DOI:10.1038/cr.2012.93, vol. 22, no. 10, pp. 1413–1415, 2012.

[18] A. G. Shabgah, F. Norouzi, H. M. Mahdiyeh et al., "A comprehensive review of long non-coding RNAs in the pathogenesis and development of non-alcoholic fatty liver disease", *Nutrition & Metabolism*, DOI:10.1186/s12986-021-00552-5, vol. 18, pp. 1–15, 2021.

[19] E. Chesnokova, A. Beletskiy, P. Kolosov, "The role of transposable elements of the human genome in neuronal function and pathology", *International Journal Of Molecular Sciences*, DOI:10.3390/ijms23105847, vol. 23, no. 10, pp. 5847, 2022.

[20] M. A. Faghihi, F. Modarresi, A. M. Khalil et al., "Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase", *Nature Medicine*, DOI:10.1038/nm1784, vol. 14, no. 7, pp. 723–730, 2008.

[21] S. C. Anna, K. Yumi, Y. Yusuke et al., "Emerging roles of long noncoding RNA in cancer", *Cancer Science*, DOI:10.1111/cas.13642, vol. 109, no. 7, pp. 2093–2100, 2018.

[22] S. B. Qu, Y. Zhong, R. Z. Shang et al., "The emerging landscape of circular RNA in life processes", *RNA Biology*, DOI:10.1080/15476286.2016.1220473, vol. 14, no. 8, pp. 992-999, 2017.

[23] J. Sun, H. Shi, Z. Wang et al., "Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network", *Molecular BioSystems*, DOI:10.1039/c3mb70608g, vol. 10, no. 8, pp. 2074–2081, 2014.

[24] G. Yu, G. Fu, C. Lu et al., "BRWLDA: bi-random walks for predicting lncRNA-disease associations", *Oncotarget*, DOI:10.18632/oncotarget.19588, vol. 8, no. 36, pp. 60429, 2017.

[25] M. N. Wang, Z. H. You, L. Wang et al., "LDGRNMF: LncRNA-disease associations prediction based on graph regularized non-negative matrix factorization", *Neurocomputing*, DOI:10.1016/j.neucom.2020.02.062, vol. 424, pp. 236–245, 2021.

[26] M. Zeng, C. Lu, F. Zhang et al., "SDLDA: lncRNA-disease association prediction based on singular value decomposition and deep learning", *Methods*, DOI:10.1016/j.ymeth.2020.05.002, vol. 179, pp. 73–80, 2020.

[27] C. Lu, M. Yang, F.Luo et al., "Prediction of lncRNA–disease associations based on inductive matrix completion", *Bioinformatics*, DOI:10.1093/bioinformatics/bty327, vol. 34, no. 19, pp. 3357-3364, 2018.

[28] Z. Bao, Z. Yang, Z. Huang et al., "LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases", *Nucleic Acids Research*, DOI:10.1093/nar/gky905, vol. 47, no. D1, pp. D1034-D1037, 2019.

[29] D. Wang, J. Wang, M. Lu et al., "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases", *Bioinformatics*, DOI:10.1093/bioinformatics/btq241, vol. 26, no. 13, pp. 1644–1650, 2010.

[30] J. Luo, Q. Xiao, "A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network", *Journal Of Biomedical Informatics*, DOI:10.1016/j.jbi.2017.01.008, vol. 66, pp. 194–203, 2017.

[31] X. Chen, Z. H. You, G. Y. Yan et al., "Irwrlda: improved random walk with restart for lncrna-disease association prediction", *Oncotarget*, DOI:10.18632/oncotarget.11141, vol. 7, no. 36, pp. 57919–57931, 2016.

[32] J. Li, Y. Zhao, S. Zhou et al., "Inferring lncRNA functional similarity based on integrating heterogeneous network data", *Frontiers in Bioengineering and Biotechnology*, DOI:10.3389/fbioe.2020.00027, vol. 8, pp. 27, 2020.

[33] Z. H. You, Z. A. Huang, Z. Zhu et al., "Pbmda: a novel and effective path-based computational model for mirnadisease association prediction", *PLoS Computational Biology*, DOI:10.1371/journal.pcbi.1005455, vol. 13, no. 3, pp. e1005455, 2017.

[34] Y. A. Huang, X. Chen, Z. H. You et al., "Ilncsim: improved lncrna functional similarity calculation model", *Oncotarget*, DOI:10.18632/oncotarget.8296, vol. 7, no. 18, pp. 25902–25914, 2016.

[35] L. T. Van, S. B. Nabuurs, E. Marchiori, "Gaussian interaction profile kernels for predicting drug-target interaction", *Bioinformatics*, DOI:10.1093/bioinformatics/btr500, vol. 27, no. 21, pp. 30–36, 2011.

[36] W. Fan, J. Shang, F. Li et al., "IDSSIM: an lncRNA functional similarity calculation model based on an improved disease semantic similarity method", *BMC Bioinformatics*, DOI:10.1186/s12859-020-03699-9, vol. 21, pp. 1–14, 2020.

[37] X. Chen, C. C. Yan, C. Luo et al., "Constructing lncrna functional similarity network based on lncrna-disease associations and disease semantic similarity", *Scientific Reports*, DOI:10.1038/srep11338, vol. 5, no. 1, pp. 11338, 2015.

[38] S. Köhler, S. Bauer, D. Horn et al., "Walking the interactome for prioritization of candidate disease genes", *American Journal Of Human Genetics*, DOI:10.1016/j.ajhg.2008.02.013, vol. 82, no. 4, pp. 949–958, 2008.

[39] K. Yang, G. Liu, N. Wang et al., "Heterogeneous network propagation for herb target identification", *BMC medical informatics and decision making*, DOI:10.1186/s12911-018-0592-z, vol. 2018, no. 18, pp. 27-37, 2018.

[40] J. Xiang, N. Zhang, J. Zhang et al., "PrGeFNE: predicting disease-related genes by fast network embedding", *Methods*, DOI:10.1016/j.ymeth.2020.06.015, vol. 2021, no. 192, pp. 3-12, 2021.

[41] P. Rajarajeswari, J. Moorthy, O.A. Beg, "Simulation of diabetic retinopathy utilizing convolutional neural networks", *Journal of Mechanics in Medicine and Biology*, DOI:10.1142/S0219519422500117, vol. 22, no. 02, pp. 2250011, 2022.

[42] D. Jia, Y. He, Y. Zhang , "Long Non-coding RNAs Regulating Macrophage Polarization in Liver Cancer", *CURRENT PHARMACEUTICAL DESIGN*, DOI:10.2174/0113816128311861240523075218, vol. 30, no. 27, pp. 2120-2128, 2024.

[43] R. Quraishi, M. Dwivedi, M. Moitra et al., "Genetic Factors and MicroRNAs in the Development of Gallbladder Cancer: The Prospective Clinical Targets", *CURRENT DRUG TARGETS*, DOI:10.2174/0113894501182288240319074330, vol. 25, no. 6, pp. 375-387, 2024.

[44] N. Rodriguez, S. Syngal, "‘’’"Expanding access to genetic testing for pancreatic cancer, *FAMILIAL CANCER*, DOI:10.1007/s10689-024-00389-w, vol. 23, no. 2024, pp. 247-254, 2024.

[45] Y. Gao, S. Shang, S. Guo et al., "Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data", *Nucleic Acids Research*, DOI:10.1093/nar/gkaa1006, vol. 49, no. D1, pp. D1251–D1258, 2021.

[46] J. Chen, J. Lin, Y. Hu et al., "RNADisease v4. 0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction", *Nucleic Acids Research*, DOI:10.1093/nar/gkac814, vol. 51, no. D1, pp. D1397–D1404, 2023.

Jinlong Ma was awarded his Ph.D. in information and communication engineering in 2016 by the Harbin Institute of Technology, located in Harbin, China. Presently, he holds the position of Associate Professor at the School of Information Science and Engineering, which is part of the Hebei University of Science and Technology in Shijiazhuang, China. His research is primarily focused on the dynamics of information dissemination within complex networks, as well as analyzing data from online social networks.

Tian Qin is presently pursuing a master's degree at the Hebei University of Science and Technology in Shijiazhuang, China. His areas of research focus on big data in networks and the study of complex networks.