Research Article

A Novel Method for Predicting Disease-Associated LncRNA-MiRNA Pairs Based on the Higher-Order Orthogonal Iteration

Zhanwei Xuan,[1,2] **Xiang Feng** ,[1,2] **Jingwen Yu,**[2] **Pengyao Ping,**[2] **Haochen Zhao** ,[2] **Xianyou Zhu,**[2,3] **and Lei Wang** [1,2]

[1]*College of Computer Engineering & Applied Mathematics, Changsha University, Changsha 410001, Hunan, China*
[2]*Key Laboratory of Intelligent Computing & Information Processing, Xiangtan University, 411105 Xiangtan, China*
[3]*Department of Computer Science, Hengyang Normal University, 421008 Hengyang, China*

Correspondence should be addressed to Xiang Feng; fengxiang@xtu.edu.cn and Lei Wang; wanglei@xtu.edu.cn

A lot of research studies have shown that many complex human diseases are associated not only with microRNAs (miRNAs) but also with long noncoding RNAs (lncRNAs). However, most of the current existing studies focus on the prediction of disease-related miRNAs or lncRNAs, and to our knowledge, until now, there are few literature studies reported to pay attention to the study of impact of miRNA-lncRNA pairs on diseases, although more and more studies have shown that both lncRNAs and miRNAs play important roles in cell proliferation and differentiation during the recent years. The identification of disease-related genes provides great insight into the underlying pathogenesis of diseases at a system level. In this study, a novel model called PADLMHOOI was proposed to predict potential associations between diseases and lncRNA-miRNA pairs based on the higher-order orthogonal iteration, and in order to evaluate its prediction performance, the global and local LOOCV were implemented, respectively, and simulation results demonstrated that PADLMHOOI could achieve reliable AUCs of 0.9545 and 0.8874 in global and local LOOCV separately. Moreover, case studies further demonstrated the effectiveness of PADLMHOOI to infer unknown disease-related lncRNA-miRNA pairs.

1. Introduction

Noncoding RNA, according to its size, can be divided into small and long noncoding RNAs approximately. Generally, small RNAs include tRNAs, miRNAs, piRNAs, and snoRNAs [1–4], and miRNAs are widely present in the cytoplasm of eukaryotic cells and are approximately 18–22 nucleotides in length, which can bind to $3'$-untranslated region of mRNA ($3'$-UTR) to inhibit the translation process of mRNA or to degrade mRNA, thereby affecting the expression of related genes [5–7]. miRNAs play important roles in a series of life activities such as cell differentiation of living body [8], growth and development [9], and apoptosis [10]. Compared to small-molecule ncRNA, lncRNA has a longer nucleotide chain with more than 200 nucleotides and has a specific and complex secondary space structure inside the molecule and can provide multiple sites for protein binding [11]. In addition, both lncRNAs and miRNAs are key members of noncoding RNAs and play important roles in coding and regulation of many complex human diseases [12–16].

Up to now, there have been many studies on relationships between diseases and miRNAs. For example, some important methods proposed by Xing Chen et al. [17–20] and Zou et al. [21–24]. In terms of prediction of potential associations between lncRNAs and diseases, Yu et al. [25] and Xing et al. [26] proposed two kinds of computational models called NBCLAD and LRLSLDA, respectively. Moreover, studies have also shown that there exist relationships between lncRNAs and miRNAs. For example, Gernapudi et al. demonstrated that miRNA 140

can induce the expression of lncRNA NEAT1 [27]. Dey et al. showed that the silencing of lncRNA H19 and knockout of H19 gene in myoblasts significantly decreased skeletal muscle differentiation [28]. Yilong et al. discovered that, after low XIST expression in gliomas, XIST could regulate miR-152 glioma stem cells to inhibit cell proliferation, migration, and invasion [29]. Xinyu et al. demonstrated that lncRNA MALAT1 could achieve post-transcriptional regulation of esophageal squamous cell carcinoma cells through miR-101 and miR-217 [30]. Er-bao et al. proposed that lncRNA ANRIL interacted with miR-99a/miR449a to regulate cell proliferation during gastric cancer formation [31]. You et al. found that the expression of miR-449a and the expression of lncRNA NEAT1 in lung cancer cell L9981 inhibited each other. When miR-449a was overexpressed, NEAT1 expression was decreased, cell proliferation was inhibited, and apoptosis was increased, and vice versa [32]. Emmrich et al. found that the expression of lncRNA MONC and MIR100HG was closely related to the miRNA groups of miR-99a~125b-2 and miR-100~125b-1. After silencing of lncRNA MONC and MIR100HG, acute megakaryocytes in the early stage of the disease, the tumor cells of leukemia patients, were severely inhibited [33]. Amy et al. found that lncRNA Ang362 was the host transcriptor of miR-211 and miR-222, and their interactions regulated Ang II and induced proliferation of vascular smooth muscle cells [34]. Miaojun et al. found that the interactions between lncRNA H19 and miRNA-675 play an important role in the metastasis of prostate cancer [35]. Obviously, the exploration of these relationships was conducive to the construction of gene regulatory networks and the identification of the mechanisms of complex human diseases [36–38].

From the above description, it is easy to see that more and more studies have shown that lncRNA-miRNA interactions are involved in the development of complex diseases. However, to the best of our knowledge, so far, in addition to the model of PADLMP proposed by Zhou et al. [39], few models have been proposed for large-scale prediction of potential associations between diseases and lncRNA-miRNA interactions. Hence, inspired by state-of-the-art methods [40–44], which show that the miRNA-miRNA pairs can work cooperatively to regulate a single gene or gene clusters being involved in similar processes [45], and simultaneously, based on the reasonable assumption that functionally similar lncRNA-miRNA pairs tend to be associated with similar diseases, in this paper, a new prediction model called PADLMHOOI was proposed to infer potential associations between diseases and the lncRNA-miRNA pairs. And, as illustrated in Figure 1, our newly proposed prediction model PADLMHOOI consists of the following four major steps:

Step 1 (Data Integration and Network Construction). In this step, first of all, we downloaded known disease-lncRNA associations from three different disease-lncRNA databases such as disease-lncRNA [46], MNDR [47, 48], and lnc2cancer [49], respectively, and then, based on these datasets, we constructed a bipartite

network of disease-lncRNA. Next, we downloaded known disease-miRNA associations from three different databases such as miR2Disease [50], HMDD [51], and miRCancer [52] separately, and then, based on these datasets, we constructed a bipartite network of disease-miRNA. Moreover, we downloaded the 2015 and 2017 versions of known lncRNA-miRNA associations from the starBasev2.0 database [53] (http://starbase.sysu.edu.cn/) on Feb 2, 2017, and based on these datasets, we constructed a bipartite network of lncRNA-miRNA. Finally, based on these three kinds of bipartite networks, we constructed an integrated tripartite network of disease-lncRNA-miRNA, which could be denoted as a tensor T.

Step 2 (Similarity Calculation). In this step, we would integrate the disease semantic similarity and Gaussian Interaction Profile Kernel similarity firstly to measure the similarity of diseases. Next, we would integrate the lncRNA functional similarity and miRNA functional similarity in three different ways to measure the functional similarity of lncRNA-miRNA pairs.

Step 3 (Weighted K-Nearest Neighbor Profile). Considering that there may be diseases that are unrelated to all lncRNA-miRNA pairs, which may lead to unsatisfactory prediction results while implementing PADLMHOOI to infer potential associations between diseases and lncRNA-miRNA pairs. Hence, in this step, we would introduce the weighted K-nearest neighbor profile (WKNNP) to add more interaction information between diseases, lncRNAs, and miRNAs to improve the prediction performance of PADLMHOOI.

Step 4 (Tensor Decomposition). In this step, we would perform tensor decomposition on the newly constructed disease-lncRNA-miRNA tensor T. Since the results of tensor decomposition include a core tensor and three matrices, we can define the final predicted association tensor as the modal product between the core tensor and these three matrices. Thereafter, we would sort scores of the lncRNA-miRNA pairs associated with each disease in the descending order in the final predicted association tensor, and it is obvious that the higher the ranking of the score, the bigger the possibility that there may exist potential association between the disease and the lncRNA-miRNA pair would be.

2. Materials and Methods

2.1. Construction of the Bipartite Network of Disease-lncRNA. In order to construct the bipartite network of disease-lncRNA, firstly, known associations between diseases and lncRNAs were downloaded from three different databases such as the LncRNADisease, MNDR, and Lnc2Cancer, respectively, and then, after feature processing (including feature cleaning and data imbalance processing etc.), 2048 different disease-lncRNA associations were finally obtained (Supplementary Table 1). Thereafter, based on these newly obtained 2048 known disease-lncRNA associations, we can
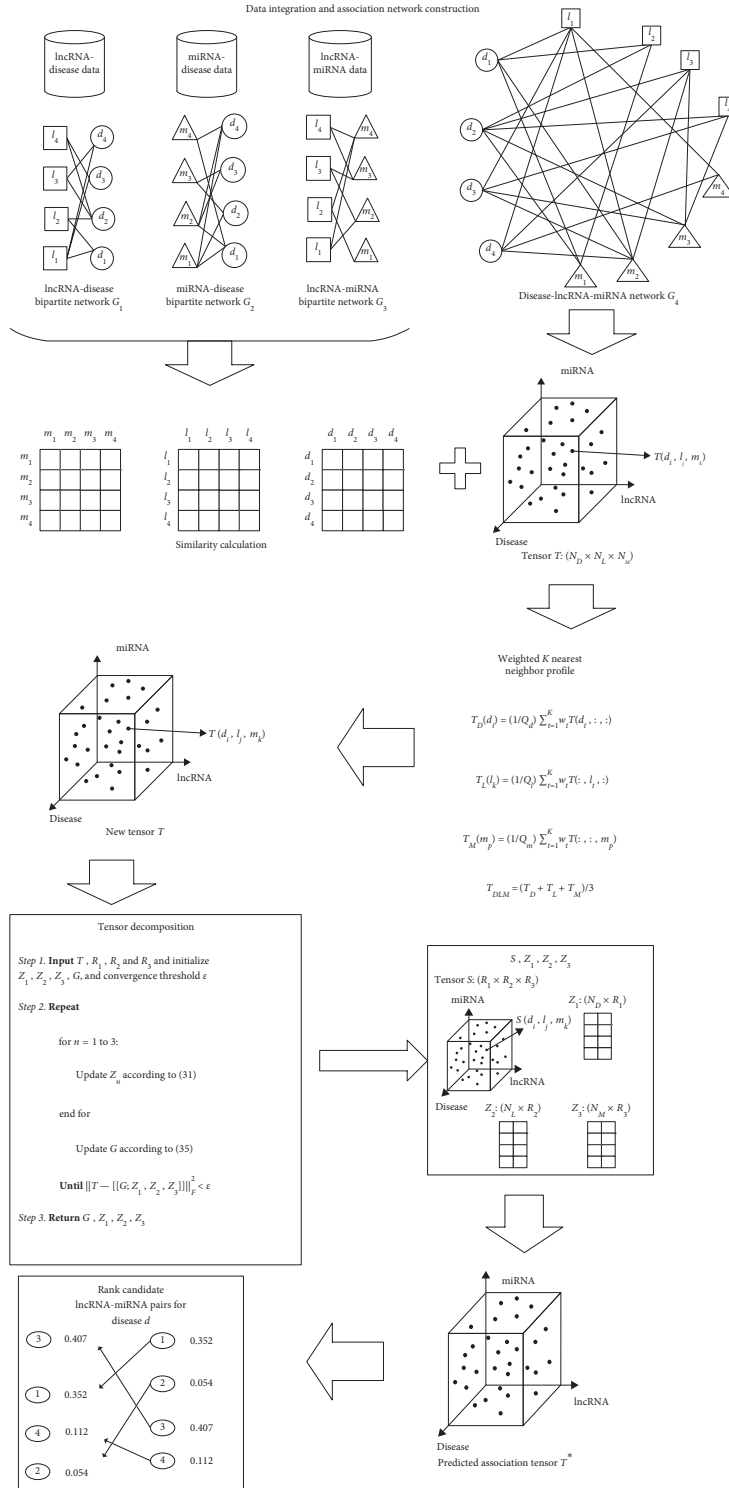
FIGURE 1: Flow chart of PADLMHOOI for predicting potential associations between diseases and lncRNA-miRNA pairs.

construct a disease-lncRNA bipartite network $G_1 = (V_1, E_1)$ according to the following steps:

Step 1. Let $V_{l_1} = \left\{ l_i | i \in [1, n_{l_1}] \right\}$ be the set of all different lncRNAs in these 2048 known disease-lncRNA associations and $V_{d_1} = \left\{ d_i | i \in [1, n_{d_1}] \right\}$ be the set of all different diseases in these 2048 known disease-lncRNA

associations, then we define $V_1 = V_{l_1} \cup V_{d_1}$ as the vertex set in G_1.

Step 2. $\forall l_i \in V_{l_1}, d_j \in V_{d_1}$, if (l_i, d_j) belongs to these 2048 downloaded known disease-lncRNA associations, then we define that there is an edge between l_i and d_j in G_1; thereafter, we can obtain the edge set E_1 in G_1.

2.2. Construction of the Bipartite Network of Disease-miRNA. In order to construct the bipartite network of disease-miRNA, at first, known disease-miRNA associations were downloaded from three different databases such as the miR2Disease, HMDD, and miRCancer separately, and then, after these newly acquired miRNAs and diseases being mapped to the database miRBase v21 [54] and disease ontology (DO) [55], respectively, 4041 different disease-miRNA associations were finally obtained (Supplementary Table 2). Hence, based on these newly obtained 4041 known disease-miRNA associations, we can construct a disease-miRNA bipartite network $G_2 = (V_2, E_2)$ according to the following steps:

Step 1. Let $V_{m_1} = \{m_i | i \in [1, n_{m_1}]\}$ be the set of all different miRNAs in these 4041 known disease-miRNA associations and $V_{d_2} = \{d_i | i \in [1, n_{d_2}]\}$ be the set of all different diseases in these 4041 known disease-miRNA associations, then we define $V_2 = V_{m_1} \cup V_{d_2}$ as the vertex set in G_2.

Step 2. $\forall m_i \in V_{m_1}, d_j \in V_{d_2}$, if (m_i, d_j) belongs to these 4041 known disease-miRNA associations, then we define that there is an edge between m_i and d_j in G_2; thereafter, we can obtain the edge set E_2 in G_2.

2.3. Construction of the Bipartite Network of lncRNA-miRNA. In order to construct the bipartite network of lncRNA-miRNA, at first, two different versions (2015 and 2017) of lncRNA-miRNA dataset were downloaded from the starBasev2.0 database separately, and then, after feature processing (including feature cleaning and data imbalance processing), 20324 different lncRNA-miRNA interactions were finally obtained (Supplementary Table 3). Thereafter, based on these newly obtained 20324 known lncRNA-miRNA associations, we can construct a lncRNA-miRNA bipartite network $G_3 = (V_3, E_3)$ according to the following steps:

Step 1. Let $V_{l_2} = \{l_i | i \in [1, n_{l_2}]\}$ denote the set of all different lncRNAs in these 20324 known lncRNA-miRNA associations and $V_{m_2} = \{m_i | i \in [1, n_{m_2}]\}$ denote the set of all different miRNAs in these 20324 known lncRNA-miRNA associations, then we define $V_3 = V_{m_2} \cup V_{l_2}$ as the vertex set in G_3.

Step 2. $\forall l_i \in V_{l_2}, m_j \in V_{m_2}$, if (l_i, m_j) belongs to these 20324 known lncRNA-miRNA associations, then we define that there is an edge between l_i and m_j in G_3; thereafter, we can obtain the edge set E_3 in G_3.

2.4. Construction of the Tripartite Network of Disease-lncRNA-miRNA. Based on the above newly obtained networks such as G_1, G_2, and G_3, we can construct a tripartite network $G_4 = (V_4, E_4)$ according to the following steps:

Step 1. Let $V_d = V_{d_1} \cup V_{d_2}, V_m = V_{m_1} \cup V_{m_2}, V_l = V_{l_1} \cup V_{l_2}, V_4 = \{\}, E_4 = \{\}$, and $V_{d'} = \{\}$.

Step 2. While V_d is not null, Repeat:

$\forall d_i \in V_d$,

If $\exists l_j \in V_l$ and $m_k \in V_m$ satisfyies the following three kinds of conditions simultaneously:

(a) $(d_i, l_j) \in E_1$
(b) $(d_i, m_k) \in E_2$
(c) $(l_j, m_k) \in E_3$

Then (d_i, l_j), (d_i, m_k), and (l_j, m_k) will be added into E_4 firstly, and then, d_i will be added into $V_{d'}$ and removed from V_d. Finally, l_j and m_k will be added into V_4 if l_j and m_k are not in V_4.

Else, d_i will be removed from V_d.

Step 3. Let $V_4 = V_4 + V_{d'}$.

According to above steps, a tripartite disease-lncRNA-miRNA association network can be obtained finally. And, it is obvious that, in the tripartite network, there are three kinds of different nodes such as disease nodes, lncRNA nodes, and miRNA nodes; moreover, the number of disease nodes, lncRNA nodes, and miRNA nodes is 68, 44, and 211, respectively, and the number of associations between diseases and lncRNA-miRNA pairs is 3,047.

2.5. Construction of the Disease-lncRNA-miRNA Tensor. Based on the newly constructed tripartite network, for any given disease node d_i, lncRNA node l_j, and miRNA node m_k in G_4, we can define a tensor T as follows:

$$T(i, j, k) = \begin{cases} 1, & [(d_i, l_j), (d_i, m_k), (l_j, m_k) \in E_4,] \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

2.6. Calculation of the Similarity of Disease Pairs

2.6.1. Calculation of the Disease Semantic Similarity (DisSemSim). In order to estimate the semantic similarity between diseases, we first downloaded the MeSH descriptor from the National Medical Library (http://www.nlm.nih.gov/) and selected the standard MeSH disease terminology. And then, for each disease d, we can construct a Directed Acyclic Graph (DAG) such as $\mathrm{DAG}(d) = (T(d), E(d))$, where $T(d)$ denotes the set of nodes containing the node d itself and its ancestors and $E(d)$ denotes the set of edges of the respective direct links from parent to child nodes [56]. Thereafter, based on the newly constructed directed acyclic graph $\mathrm{DAG}(d)$, the semantic contribution of an ancestor node d_s to the disease d can be calculated as follows:

$$D_d(d_s) = \begin{cases} 1, & \text{if } d_s = d, \\ \max\{\Delta * D_d(d') \mid d' \in \text{children of } d_s\}, & \text{otherwise,} \end{cases}$$

$$\mathrm{DV}(d) = \sum_{d_i \in T(d)} D_d(d_i), \quad (2)$$

where Δ is the semantic contribution decay factor with value between 0 and 1. And, in addition, according to the experimental results of some previous state-of-the-art methods [57, 58], the most appropriate value for Δ will be 0.5. Hence, based on the assumption that two diseases with more common ancestor nodes in their *DAG*s shall have higher semantic similarity, the semantic similarity between two diseases d_i and d_j can be defined as follows:

$$\text{DisSemSim}(i, j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} \left(D_{d_i}(t) + D_{d_j}(t) \right)}{\text{DV}(d_i) + \text{DV}(d_j)}. \tag{3}$$

2.6.2. Calculation of the Gaussian Interaction Profile Kernel Similarity for Diseases (GIPSim).

Based on the hypothesis that functionally similar genes are often associated with similar diseases, in this section, we will adopt the Gaussian Interaction Profile Kernel to calculate the similarity of diseases according to the following steps:

Firstly, based on the networks G_1 and G_2 constructed above, for any given lncRNA l_i and disease d_j, we define that

$$Y_1(i, j) = \begin{cases} 1, & \text{if } (l_i, d_j) \in G_1, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Next, for any given miRNA m_i and disease d_j, we define that

$$Y_2(i, j) = \begin{cases} 1, & \text{if } (m_i, d_j) \in G_2, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Hence, let $\text{IP}_l(d_i)$ denote the ith column of the matrix Y_1, then we can calculate the Gaussian Kernel Similarity between diseases d_i and d_j based on their interaction profiles as follows:

$$\text{GIP}_{\text{LDSIM}}(i, j) = \exp\left(-\gamma_{d1} \left\| \text{IP}_l(d_i) - \text{IP}_l(d_j) \right\|^2 \right), \tag{6}$$

$$\gamma_{d1} = \frac{1}{(1/n_{d1}) \sum_{i=1}^{n_{d1}} \left\| \text{IP}(d_l(i))^2 \right\|}, \tag{7}$$

where the parameter n_{d1} denotes the number of different diseases in G_1.

In a similar way, let $\text{IP}_m(d_i)$ denote the ith column of matrix Y_2, then we can calculate the Gaussian Kernel Similarity between diseases d_i and d_j based on their interaction profiles as follows:

$$\text{GIP}_{\text{MDSIM}}(i, j) = \exp\left(-\gamma_{d2} \left\| \text{IP}_m(d_i) - \text{IP}_m(d_j) \right\|^2 \right),$$

$$\gamma_{d2} = \frac{1}{(1/n_{d2}) \sum_{i=1}^{n_{d2}} \left\| \text{IP}_m(d_i) \right\|^2}, \tag{8}$$

Here, the parameter n_{d2} denotes the number of different diseases in G_2.

Thereafter, based on these above formulas, we can calculate the Gaussian Interaction Profile Kernel Similarity between diseases d_i and d_j as follows:

$$\text{GIPSim}(i, j) = \frac{\text{GIP}_{\text{LDSIM}}(i, j) + \text{GIP}_{\text{MDSIM}}(i, j)}{2}. \tag{9}$$

2.7. Calculation of the Similarity of lncRNA Pairs (lncSim)

2.7.1. Calculation of the lncRNA Functional Similarity (lncfunSim).

For any two given lncRNAs such as l_i and l_j, let $\text{DT}_1 = \{dt_{11}, dt_{12}, \ldots, dt_{1m}\}$ be all the diseases related to l_i in G_1 and $\text{DT}_2 = \{dt_{21}, dt_{22}, \ldots, dt_{2n}\}$ be all the diseases related to l_j in G_1, then we can define the functional similarity between l_i and l_j as follows:

$$\text{lncfunSim}(i, j) = \frac{\sum_{1 \le k \le m} \text{SemSims}(dt_{1k}, \text{DT}_2) + \sum_{1 \le k \le n} \text{SemSims}(dt_{2k}, \text{DT}_1)}{m + n}, \tag{10}$$

where

$$\text{SemSims}(dt_{1k}, \text{DT}_2) = \max_{1 \le l \le n} (\text{DisSemSim}(dt_{1k}, dt_{2l})),$$

$$\text{SemSims}(dt_{2k}, \text{DT}_1) = \max_{1 \le l \le m} (\text{DisSemSim}(dt_{2k}, dt_{1l})). \tag{11}$$

2.7.2. Calculation of the Gaussian Interaction Profile Kernel Similarity for lncRNAs (GIP_lncSim).

For any two given lncRNAs such as l_i and l_j, similar to the definition of formula (6), let $\text{IP}(l_i)$ and $\text{IP}(l_j)$ denote the ith and the jth row of the matrix Y_1, respectively, then we can calculate the Gaussian Kernel Similarity between diseases l_i and l_j based on their interaction profiles as follows:

$$\text{GIP}_{\text{lncSim}}(i, j) = \exp\left(-\gamma_l \left\| \text{IP}(l_i) - \text{IP}(l_j) \right\|^2 \right),$$

$$\gamma_l = \frac{1}{(1/n_l) \sum_{i=1}^{n_l} \left\| \text{IP}(l_i) \right\|^2}, \tag{12}$$

where n_{l1} denotes the number of different lncRNAs in G_1.

Hence, based on these formulas given above, we can finally define the similarity measurement between lncRNAs l_i and l_j as follows:

$$\text{lncSim}(i, j) = \frac{\text{lncfunSim}(i, j) + \text{GIP}_{\text{lncSim}}(i, j)}{2}. \tag{13}$$

2.8. Calculation of the Similarity between miRNAs (miRSim)

2.8.1. Calculation of the miRNA Function Similarity (miRfunSim).

For any two given miRNAs, such as m_i and m_j, let $\text{DT}_3 = \{dt_{31}, dt_{32}, \ldots, dt_{3p}\}$ be all the diseases related to m_i in G_2 and $\text{DT}_4 = \{dt_{41}, dt_{42}, \ldots, dt_{4q}\}$ be all the diseases related to m_j in G_2, then we can define the functional similarity between m_i and m_j as follows:

$$\text{miRfunSim}(i, j) = \frac{\sum_{1 \le k \le p} \text{SemSims}(\text{dt}_{1k}, \text{DT}_3) + \sum_{1 \le k \le p} \text{SemSims}(\text{dt}_{2k}, \text{DT}_4)}{p + q}. \tag{14}$$

2.8.2. Calculation of the Gaussian Interaction Profile Kernel Similarity for miRNAs (GIP_{miRSim}). For any two given miRNAs, such as m_i and m_j, in a similar way, let IP(m_i) and IP(m_j) represent the ith and jth row in matrix Y_2, respectively, then we can calculate the Gaussian Kernel Similarity between diseases m_i and m_j based on their interaction profiles as follows:

$$\text{GIP}_{\text{miRSim}}(i, j) = \exp\left(-\gamma_m \left\| \text{IP}(m_i) - \text{IP}(m_j) \right\|^2\right),$$

$$\gamma_m = \frac{1}{(1/n_m) \sum_{i=1}^{n_m} \left\| \text{IP}(m_i) \right\|^2}, \tag{15}$$

where n_{m2} denotes the number of miRNAs in G_2.

Hence, based on these formulas presented above, we can finally define the similarity measurement between miRNAs m_i and m_j as follows:

$$\text{miRSim}(i, j) = \frac{\text{miRfunSim}(i, j) + \text{GIP}_{\text{miRSim}}(i, j)}{2}. \tag{16}$$

2.9. Weighted K Nearest Neighbor Profiles for Diseases, lncRNAs, and miRNAs (WKNNP). Let $D = \{d_1, d_2, \ldots, d_m\}$, $L = \{l_1, l_2, \ldots, l_n\}$, and $M = \{m_1, m_2, \ldots, m_k\}$ denote the set of m diseases, n lncRNAs, and k miRNAs, respectively. Let $T(d_i) = T(i, :, :)$ denote the ith horizontal slice matrix in disease axis of the tensor T, hence, $T(d_i)$ also represents the interaction profile for the disease d_i. Let $T(l_j) = T(:, j, :)$ denote the jth lateral slice matrix in lncRNA axis of the tensor T, hence, $T(l_j)$ also represents the interaction profile for lncRNA l_j. Let $T(m_p) = T(:, :, p)$ denote the pth frontal slice matrix in miRNA axis of the tensor T, hence, $T(m_p)$ also denotes the interaction profile for miRNA m_p. Then, it is obvious that the values in these three kinds of interaction profiles of any novel diseases, lncRNAs, or miRNAs are all zeros, which may lead to unsatisfactory prediction performance during inferring potential associations between diseases and lncRNA-miRNA pairs. Hence, in this section, we will perform a procedure for the construction of new interaction profiles to address the problem mentioned above. And, in this procedure, for each disease d_i, its association with other K nearest known diseases (including at least one experimentally verified association) and corresponding K interaction profiles will be utilized to obtain the following interaction profile:

$$T_D(d_i) = \frac{1}{Q_d} \sum_{t=1}^{K} w_t T(d_t, :, :), \tag{17}$$

where, $\{d_1, d_2, \ldots, d_K\}$ are the diseases sorted in descending order based on their similarity to d_i, w_t is the weight coefficient, and $w_t = \alpha^{t-1} * \text{disSim}(d_t, d_i)$, which means that a higher weight will be assigned if d_t is more similar to d_i. The parameter α is a decay term with values between 0 and 1. The parameter Q_d is a normalization term, and there is $Q_d = \sum_{t=1}^{K} \text{disSim}(d_t, d_i)$.

In the same manner, the new interaction profile for each l_k can be determined as follows:

$$T_L(l_k) = \frac{1}{Q_l} \sum_{t=1}^{K} w_t T(:, l_t, :), \tag{18}$$

where $\{l_1, l_2, \ldots, l_K\}$ are the lncRNAs sorted in the descending order based on their similarity to l_k, w_t is the weight coefficient, and $w_t = \alpha^{t-1} * \text{lncSim}(l_t, l_k)$, which means that a higher weight will be assigned if l_t is more similar to l_k. The parameter Q_l is a normalization term, and there is $Q_l = \sum_{t=1}^{K} \text{lncSim}(l_t, l_k)$.

Similarly, the new interaction profile for each m_p can be determined as follows:

$$T_M(m_p) = \frac{1}{Q_m} \sum_{t=1}^{K} w_t T(:, :, m_p), \tag{19}$$

where $\{m_1, m_2, \ldots, m_K\}$ are the miRNAs sorted in the descending order based on their similarity to m_p, w_t is the weight coefficient, and $w_t = \alpha^{t-1} * \text{miRSim}(m_t, m_p)$, which means that a higher weight is assigned if m_t is more similar to m_p. The parameter Q_m is a normalization term, and there is $Q_m = \sum_{t=1}^{K} \text{miRSim}(m_t, m_p)$.

Thereafter, after combining the above three kinds of tensors T_D, T_L, and T_M obtained from different data spaces and replacing $T(i, j, k) = 0$ with an associated likelihood score, we can update the original adjacency matrix T as follows:

$$T = \max(T, T_{DLM}), \tag{20}$$

where $T_{DLM} = (a_1 T_D + a_2 T_L + a_3 T_M / \sum a_i)$, $(i = 1, 2, 3)$.

2.10. PADLMHOOI. Inspired by the successful application of tensor decomposition in the field of link prediction and the application of nonnegative matrix decomposition methods in inferring disease-miRNA associations, in this section, we proposed a novel model called PADLMHOOI to predict new associations between diseases and miRNA-lncRNA pairs. From above descriptions, it is easy to know that a tensor is a multidimensional array. Currently, the most commonly used tensor decomposition techniques include Tucker decomposition [59], HOSVD [60], and HOOI [61]. In this section, we will perform Tucker decomposition on above constructed tensor T. Assuming $T \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tucker decomposition aims at finding $Z_\alpha (\alpha \in (1, 2, 3))$ and core tensor $G \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ that can solve the following optimization problem:

$$\text{minimize} \quad \left\| T - \widehat{T} \right\|_F^2,$$

$$\text{s.t.} \quad \widehat{T}^{i,j,k} = \sum_{r_1, r_2, r_3} Z_1^{i,r_1} Z_2^{j,r_2} Z_3^{k,r_3} G^{r_1, r_2, r_3}, \quad \forall\, i, j, k. \tag{21}$$

Hence, based on formula (21), we can further transform this equation to following simple form:

$$\text{minimize} \quad \|T - \widehat{T}\|_F^2,$$
$$\text{s.t.} \qquad \widehat{T} = GX_1Z_1X_2Z_2X_3Z_3 = [\![G; Z_1, Z_2, Z_3]\!],$$
$$(22)$$

$Z_1 \in \mathbb{R}^{n_1 \times R_1}, Z_2 \in \mathbb{R}^{n_2 \times R_2}$, and $Z_3 \in \mathbb{R}^{n_3 \times R_3}$ are the factor matrices, which are usually orthogonal and can be considered as the main component of each mode. R_1, R_2, and R_3 are the number of columns $(\max(R_1, R_2, R_3) \ll \min(n_1, n_2, n_3))$ in the factor matrices Z_1, Z_2, and Z_3 respectively. The notation X_n denotes n-mode product; $[\![G; Z_1, Z_2, Z_3]\!]$ is the shorthand introduced by Kolda and Gibson [62] (Supplementary File A).

Based on equation (22), the above optimization problem can be solved according to the following steps:

Considering that the derivation forms of Z_1, Z_2, and Z_3 are similar, we will only derive the iterative formula of Z_1 as an example. Firstly, as illustrated in formula (23), the objective function given in formula (22) can be rewritten as a matrix form of T along the first dimension:

$$\left\|T_{(1)} - Z_1[\![G; Z_2, Z_3]\!]_{(1)}\right\|_F^2, \tag{23}$$

where $T_{(1)} \in \mathbb{R}^{n_1 \times (n_2 * n_3)}$ is the unfolding of T along the first dimension (Supplementary File A). Assuming that the optimal solution Z_1 satisfies all the constraints in equation (22), we have

$$T_{(1)} = Z_1[\![G; Z_2, Z_3]\!]_{(1)} = Z_1 G_{(1)}\left(Z_2 \otimes Z_3\right)^T, \tag{24}$$

where $\otimes$ denotes the Kronecker product, and moreover, we have

$$S_1 = G_{(1)}\left(Z_2 \otimes Z_3\right)^T. \tag{25}$$

Hence, formula (24) can be regarded as a nonnegative matrix factorization (NMF) form [63]. Then, we can finally obtain the solution of Z_1 by updating NMF as follows:

$$Z_1 \longleftarrow Z_1 * \frac{T_{(1)}S_{(1)}^T}{Z_1 S_{(1)}S_{(1)}^T}. \tag{26}$$

Hence, we can finally obtain the factor matrices Z_2 and Z_3 in a similar way. Thereafter, while fixing the factor matrices Z_1, Z_2, and Z_3, the objective function in formula (22) can be converted to the following form:

$$\left\|T - \widehat{T}\right\|_F^2 = \left\|\text{vec}(T) - \left(Z_3 \otimes Z_2 \otimes Z_1\right)\text{vec}(G)\right\|_F^2, \tag{27}$$

where $\text{vec}(\cdot)$ denotes the vectorization of the tensor. And moreover, based on formula (27), the following linear equation can be obtained:

$$\text{vec}(X) = \left(Z_3 \otimes Z_2 \otimes Z_1\right)\text{vec}(G). \tag{28}$$

Let $Q = Z_3 \otimes Z_2 \otimes Z_1$, then obviously, formula (28) can also be regarded as a NMF, and thereafter, the core tensor in formula (28) can be obtained as follows [63]:

$$\text{vec}(G) \longleftarrow \text{vec}(G) * \frac{Q^T\text{vec}(T)}{Q^TQ\text{vec}(G)}$$
$$= \text{vec}\left(G * \frac{[\![T; Z_1^T, Z_2^T, Z_3^T]\!]}{[\![G; Z_1^T, Z_2^T, Z_3^T]\!]}\right), \tag{29}$$

$$G \longleftarrow G * \frac{[\![T; Z_1^T, Z_2^T, Z_3^T]\!]}{[\![G; Z_1^T, Z_2^T, Z_3^T]\!]}. \tag{30}$$

Based on above formulas, the pseudocode of our prediction model PADLMHOOI based on tensor decomposition can be described as follows:

Step 1. **Input**: T, R_1, R_2, R_3, Z_1, Z_2, Z_3, G, and the convergence threshold ε.

Step 2. **Repeat**

    For $i = 1$ to 3:

        Update Z_i according to formula (26)

    End For

    Update G according to formula (30)

    Until $\|T - [[G; Z_1, Z_2, Z_3]]\|_F^2 < \varepsilon$

Step 3. **Return** Z_1, Z_2, Z_3, G

According to above steps, we can obtain the final predicted disease-lncRNA-miRNA association tensor $T^* = GX_1Z_1X_2Z_2X_3Z_3$, and after prioritizing the disease-related lncRNA-miRNA pairs based on the entities in the tensor T^*, obviously, the top-ranked lncRNA-miRNA pairs can be regarded as more likely to be related to the corresponding disease.

3. Results and Analysis

3.1. Leave-One-Out Cross-Validation (LOOCV). In order to estimate the prediction performance of our newly proposed prediction model, the global leave-one-out cross-validation (LOOCV), 2-fold cross-validation (2-fold CV), and 10-fold cross-validation (10-fold CV) were implemented on PADLMHOOI, respectively. In the K-fold cross-validation, the initial sample will be divided into K subsample sets, and a single subsample set is retained as the data for the validation model, while the other $K-1$ samples are used to train the model. During simulation, the cross-validation will be performed K times, and each subsampling set will be verified once, and the average results of K times will be utilized to obtain a single estimation. Moreover, in order to reduce the performance deviation caused by the random sample partitioning, we divide the partition 100 times and then obtain the ROC curve and the AUC value in the same way as the LOOCV. And, as a result, from the following Table 1, it is easy to see that PADLMHOOI can achieve reliable AUCs of 0.9545, 0.9730 ± 0.0119, and 0.9626 ± 0.0150 in the frameworks of global LOOCV, 2-fold CV, and 10-fold CV, respectively. Additionally, in order to further estimate the prediction performance of PADLMHOOI, we implemented it under the framework of local LOOCV, and the simulation results of 50 predicted related diseases were illustrated in Supplementary Table 4.

3.2. Performance Comparison with Other Methods. To the best of our knowledge, up to now, PADLMP [39] is the unique model having been proposed for predicting potential associations between disease and lncRNA-miRNA pairs, in which, these three kinds of nodes such as disease nodes, lncRNA nodes, and miRNA nodes are considered simultaneously to construct a triple network. And, the major difference between PADLMP and our model PADLMHOOI is that PADLMP is based on the method of link prediction. Therefore, in order to compare PADLMP with our model PADLMHOOI, we implemented LOOCV to verify the prediction performance of these two models based on the 3047 known disease-lncRNA-miRNA associations downloaded above. In the first experiment, we set the parameters in PADLMP to their best values; specifically, the step size K is set to 2 and the attenuation coefficient γ is set to 0.01. Meanwhile, for convenience, we set the parameters in PADLMHOOI as follows: the parameters a_1, a_2, and a_3 in formula (20) are all set to 1, the parameters r_1, r_2, and r_3 in formula (21) are all set to 5, and the parameters K and α in formulas (17)–(19) are all set to 3 and 0.1 separately. And, as illustrated in Figure 2, it is easy to see that PADLMHOOI and PADLMP can achieve the AUCs of 0.9545 and 0.9318 separately, which demonstrate that the prediction performance of PADLMHOOI is superior to that of PADLMP.

As time went by, we found that some databases have been updated. Hence, in order to further demonstrate the advancement of PADLMHOOI, we once again collected the latest disease-lncRNA correlations from the databases lnc2cancer v2.0, lncRNADisease 2.0 [64], and MNDR v2.0 [48], collected the latest disease-miRNA associations from the database HMDD v3.0, and collected the latest lncRNA-miRNA associations from the database RAID v2.0 [65] separately. And thereafter, we reconstructed the triple network based on these newly collected latest datasets. In the newly constructed triple network, the numbers of disease nodes, lncRNA nodes, and miRNA nodes are 42, 234, and 251 respectively; the number of known associations between diseases and lncRNA-miRNA pairs is 3,768; the number of known associations between diseases and lncRNAs is 733; and the number of known associations between diseases and miRNAs is 674. Then, based on the new triple network, we compared our model PADLMHOOI with PADLMP once more. And, in this second experiment, we set the parameters K and α to 10 and 0.5, respectively, in PADLMHOOI and kept other parameters unchanged as in the first experiment. And, as illustrated in Figure 3, simulation results show that PADLMHOOI and PADLMP can achieve AUCs of 0.9026 and 0.9013, respectively, which demonstrate that the prediction performance of PADLMHOOI outperforms that of PADLMP markedly.

Additionally, the interesting point is that our model can infer potential disease-lncRNA associations and disease-miRNA associations incidentally, while predicting potential associations between diseases and lncRNA-miRNA pairs. Hence, it is reasonable as well to compare our model PADLMHOOI with prediction models for inferring potential disease-lnRNA or disease-miRNA associations. Therefore, in this section, we would compare PADLMHOOI with some state-of-the-art computational prediction models

Table 1: Performance of PADLMHOOI in global LOOCV, 2-fold cross-validation, and 10-fold cross-validation.

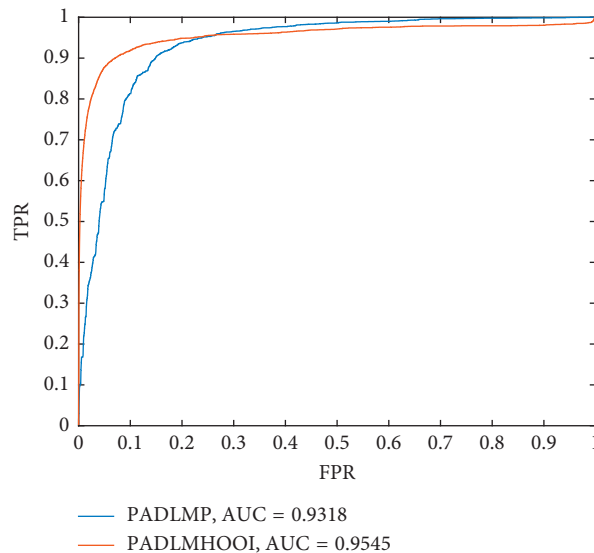| Global LOOCV	2-fold CV	10-fold CV
0.9545	0.9730 ± 0.0119	0.9626 ± 0.0150

Figure 2: Performance comparison between PADLMHOOI and PADLMP in terms of ROC curves and AUCs based on the 3047 known disease-lncRNA-miRNA associations.
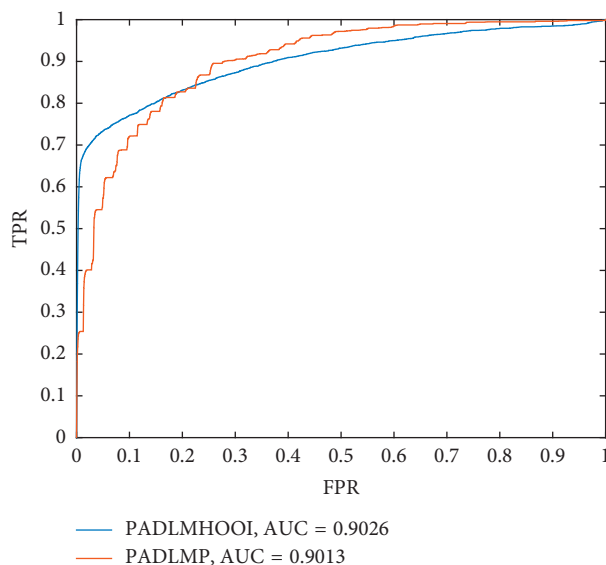

Figure 3: Performance comparison between PADLMHOOI and PADLMP in terms of ROC curves and AUCs based on the latest 3678 known disease-lncRNA-miRNA associations. Here, comparing with the AUCs in Figure 2, the reason that the AUCs of our model decline in Figure 3 is that the values of parameters K and α are different. In Figure 2, $K = 3$ and $\alpha = 0.1$, while in Figure 3, $K = 10$ and $\alpha = 0.5$.

such as the LRLSLDA [26], NBCLAD [25], WBSMDA [66], and RLSMDA [67]. Among them, LRLSLDA is a semi-supervised learning-based prediction model for inferring potential lncRNA-disease associations; NBCLAD is

a probabilistic model for predicting potential associations between diseases and lncRNAs; WBSMDA is a prediction model for predicting potential associations between diseases and miRNAs; and RLSMDA is a prediction model for predicting disease-related miRNAs based on the framework of regularized least squares. In addition, while comparing with LRSLDA, known disease-lncRNA associations were obtained from the triple disease-lncRNA-miRNA network; however, the parameters in LRSLDA are set to the same values given in the literature. Moreover, while comparing with NBCLDA, considering that there are four kinds of nodes such as diseases, lncRNAs, miRNAs, and genes included in NBCLDA, there are three kinds of nodes such as diseases, lncRNAs, and miRNAs in our model PAD-LMHOOI. Hence, for the sake of fairness, we only compared PADLMHOOI with the submethod NBCLDA-GN1-SD. And, as illustrated in Figure 4, simulation results show that PADLMHOOI, NBCLDA-G1-SD, and LRSLDA can achieve AUCs of 0.9568, 0.7928, and 0.5924 separately, which demonstrate that PADLMHOOI thoroughly defeats both NBCLDA-G1-SD and LRSLDA. In addition, while comparing with WBSMDA and RLSMDA, 674 known disease-miRNA associations were obtained from the triple disease-lncRNA-miRNA network; however, the parameters in both WBSMDA and RLSMDA are set to the same values given in the literatures. And, as illustrated in Figure 5, simulation results show that PADLMHOOI, WBSMDA, and RLSMDA can achieve AUCs of 0.9157, 0.8544, and 0.8991, respectively, which demonstrate that PADLMHOOI outperforms both WBSMDA and RLSMDA thoroughly as well.

3.3. Recall Ratio Analysis. In this section, in order to further evaluate the prediction performance of PADLMHOOI, we compared the recall value of PADLMHOOI and other state-of-the-art models. It is well known that the higher recall ratio of all selected diseases in a top k ranking list means that the more positive testing samples (real disease-related lncRNA-miRNA pairs) have been identified successfully. And, as a result, Figure 6 illustrates the recall rate of all selected diseases in different top k ranking lists. Moreover, we further listed the recall rate of some given diseases associated with at least 80 verified lncRNA-miRNA associations in Supplementary Table 5.

3.4. Case Studies. In this section, case studies of breast neoplasms, colon neoplasms, and prostate neoplasms were conducted to further verify the capability of PADLMHOOI to detect novel associations between diseases and lncRNA-miRNA pairs separately. And, among these three kinds of case studies, breast cancer is the second leading cause of female cancer death and comprises 22% of all cancers in women [68, 69]. The related literature has suggested that lncRNAs and miRNAs play an important role in the formation of many diseases, and the formation of breast cancer may be more relevant to them [70, 71]. Predicting breast cancer-associated lncRNA-miRNA pairs and identifying lncRNAs and miRNAs as biomarkers may make a significant contribution to better diagnosis and treatment of breast
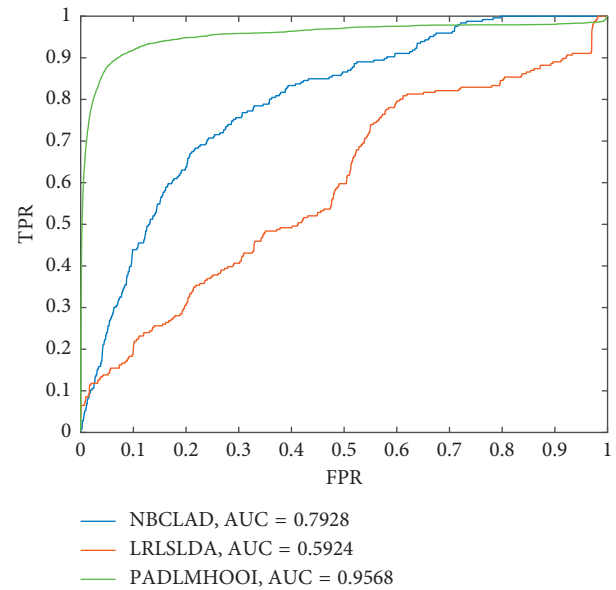

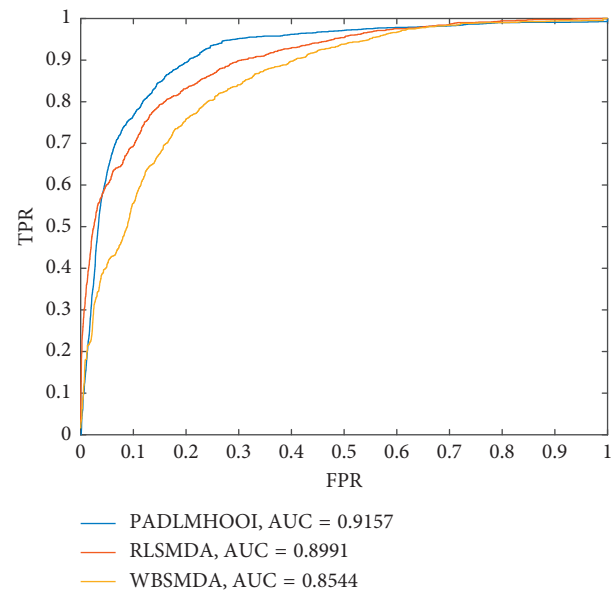FIGURE 4: The comparison results between PADLMHOOI and LRLSLDA and NBCLAD.

FIGURE 5: The comparison results between PADLMHOOI and RLSMDA and WBMDA.

cancer [71]. In Supplementary Table 6, we have listed the top 30 candidate lncRNA-miRNA pairs related to breast cancer. And, in Supplementary Table 6, the column of lncRID and miRID denotes lncRNA ID and miRNA ID, respectively. Evi1 and Evi2 denote some authority database or published literature containing verified disease-lncRNA or disease-miRNA associations separately. "#" and "∗" stand for databases of lncRNADisease and MNDR v2.0, respectively, which consist of known disease-lncRNA associations or contain published literatures to support the association between predicted lncRNAs and breast cancer. "!," "&," and "+" stand for databases of HMDD, miR2Disease, and miRCancer, respectively, which consist of known disease-
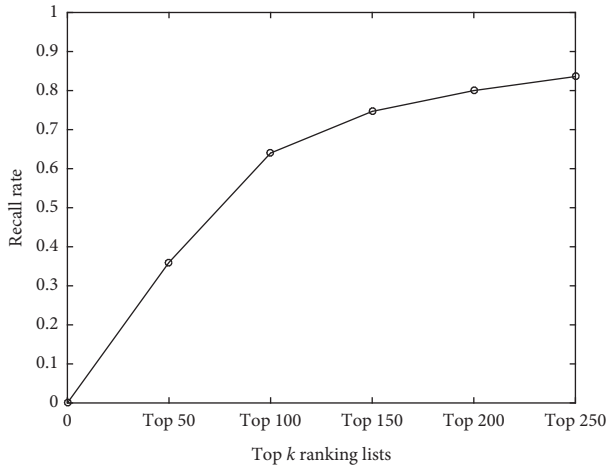
Figure 6: The recall rate of all the selected diseases in different top k ranking lists.

miRNA associations or contain published literature to support the association between predicted miRNAs and breast cancer. Particularly, "Nan" indicates that there is no database or no published literature to support the predicted results. From Supplementary Table 6, it is easy to see that all candidate disease-lncRNA associations have been verified in databases of the lncRNADisease and MNDR v2.0 or published papers containing these databases. And, in addition, there are 42 out of 50 candidate disease-miRNA associations having been reported by HMDD, miR2Disease, and miR-Cancer or published paper containing these databases. Moreover, we discovered that those novel miRNAs with miRID 35, 51, 73, 164, and 186 are related to some important factors affecting the development of breast neoplasms. Hence, it is obvious that we infer that these lncRNA-miRNA pairs may be associated with breast cancer.

In addition, colonic tumors are a type of malignancy that is common in the rectum and sigmoid borders [72]. Early colon cancer is difficult to detect because of its insignificant symptoms [73]. Unfortunately, the related literature reports that its incidence has been on the rise in recent years [74]. Therefore, predicting potential miRNAs and lncRNAs associated with colon tumors is of great significance for the diagnosis of early colon cancer. In Supplementary Table 7, we have listed the top 30 candidate lncRNA-miRNA pairs predicted to be associated with colon tumors. Moreover, all of these candidate lncRNAs and most of these candidate miRNAs have been verified by lncRNADisease database and MNDR v2.0, respectively.

Moreover, prostate neoplasm is one of the most common cancers in white and African-American men, and it is reported that there are about one in six white men and one in five African-American men having prostate cancer in their lifetime. Recent researches have shown that prostate neoplasm is caused by the malignancy of prostate epithelial cells [75], its formation includes many factors such as age, family history, and race [76], and particularly, some miRNAs such as has-let-7a-5p and lncRNAs such as XIST have been found to be involved in the formation of prostate neoplasms successively. Hence, it is interesting to infer potential

miRNAs and lncRNAs associated with prostate neoplasms. In Supplementary Table 8, we have listed the top 30 prostate neoplasm-related candidate lncRNA-miRNA pairs. Moreover, all of these candidate lncRNAs and most of these candidate miRNAs have been verified by lncRNADisease and MNDR v2.0, respectively.

3.5. Parameter Sensitivity Analysis. Considering that there are some key parameters such as K and α, which may be significant to the performance of our prediction model PADLMHOOI, in this section, we will further estimate the effects of these key parameters to the prediction performance of PADLMHOOI. Firstly, we varied K from 1 to 10 during simulation. And, as a result, Table 2 illustrates the impacts of parameter K on the performance of PADLMHOOI. By observing Table 2, it is obvious that PADLMHOOI can achieve the maximum AUC value of 0.9708 while $K = 8$. And additionally, as for the impacts of the parameter α, considering the time costs, we set $K = 3$ and varied α from 0.1 to 0.9 during simulation. And as a result, Table 3 illustrates the impacts of parameter α on the performance of PAD-LMHOOI. By observing Table 3, it is obvious that PAD-LMHOOI can achieve the maximum AUC value of 0.9591 while $\alpha = 0.7$.

4. Discussion and Conclusion

Researches on prediction of potential associations between lncRNA-miRNA pairs and diseases not only are helpful in understanding the disease mechanisms on lncRNA and miRNA levels but also play an important role in the detection of disease biomarkers, diagnosis, prognosis, and prevention. However, to our knowledge, although there are many researches having demonstrated that lncRNA-miRNA interactions are associated with the development of complex diseases, up to now, there are few models having been proposed for large-scale forecasting potential associations between diseases and lncRNA-miRNA pairs. Since traditional biological experiments are quite expensive and time-consuming, in this paper, based on the existing disease-miRNA associations, disease-lncRNA associations, lncRNA-miRNA interactions, and the assumption that genes with similar functions are often associated with similar diseases; we firstly constructed a three-order tensor T by adopting the method of WKNNP, and then, based on the method of tensor factorization, we further proposed a prediction model called PADLMHOOI to infer potential relations between diseases and lncRNA-miRNA pairs. And thereafter, simulation results under the frameworks of global and local LOOCV, 2-fold CV, and 10-fold CV, all confirmed the superiority of PADLMHOOI. Moreover, case studies of breast neoplasms, colon neoplasms, and prostate neoplasms further demonstrate that our model PADLMHOOI is an effective method for predicting potential disease-associated lncRNA-miRNA pairs. Certainly, there are still some limitations in PADLMHOOI. For example, although a large number of datasets have been integrated in PADLMHOOI, the amount of data available is still not enough; it is obvious

Table 2: Impacts of the parameter K on the performance of PADLMHOOI.

K	1	2	3	4	5	6	7	8	9	10
AUC	0.9660	0.9649	0.9591	0.9607	0.9666	0.9657	0.9675	0.9708	0.9703	0.9703

Table 3: Impacts of the parameter α on the performance of PADLMHOOI.

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AUC	0.9545	00.9565	0.9582	0.9586	0.9583	0.9585	0.9591	0.9587	0.9539

that the prediction performance of PADLMHOOI will be better if more datasets can be collected. And in addition, in this paper, we only predicted the association between disease and a single lncRNA-miRNA pair. In the future, we will further modify PADLMHOOI to predict potential associations between diseases and multiple lncRNA-miRNA pairs.

Abbreviations

PADLMHOOI: Prediction of potential associations between diseases and lncRNA-miRNA pairs based on the higher-order orthogonal iteration.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

Supplementary Materials

Supplementary 1. File 1: introduction to tensor and optimization of objective function and update of factor matrix and core tensor.

Supplementary 2. Table 1: 2048 known disease-lncRNA associations.

Supplementary 3. Table 2: 4041 known disease-miRNA associations.

Supplementary 4. Table 3: 20324 known lncRNA-miRNA associations.

Supplementary 5. Table 4: AUC score for 50 diseases in the framework of local LOOCV.

Supplementary 6. Table 5: recall ratio of some important diseases.

Supplementary 7. Table 6: the candidate lncRNA-miRNA pairs associated with breast cancer. In addition, the LncRNADisease and MNDR v2.0 databases have confirmed that these lncRNAs or miRNAs are associated with breast cancer.

Supplementary 8. Table 7: the candidate lncRNA-miRNA pairs associated with colon cancer. In addition, the LncRNADisease and MNDR v2.0 databases have confirmed that these lncRNAs or miRNAs are associated with colon cancer.

Supplementary 9. Table 8: the candidate lncRNA-miRNA pairs associated with pprostate cancer. In addition, the LncRNADisease and MNDR v2.0 databases have confirmed that these lncRNAs or miRNAs are associated with colon cancer.

References

[1] The ENCODE Project Consortium, "Identification and analysis of functional elements in 1% of the human genome by encode pilot project," *Nature*, vol. 447, pp. 799–816, 2007, https://ci.nii.ac.jp/naid/10029378996/en/.

[2] The FANTOM Consortium, P. Carninci, T. Kasukawa et al., "The transcriptional landscape of the mammalian genome," *Science*, vol. 309, no. 5740, pp. 1559–1563, 2005, http://science.sciencemag.org/content/309/5740/1559.

[3] P. Kapranov, J. Cheng, S. Dike et al., "RNA maps reveal new RNA classes and a possible function for pervasive transcription," *Science*, vol. 316, no. 5830, pp. 1484–1488, 2007, http://science.sciencemag.org/content/316/5830/1484.

[4] Y. Okazaki, M. Furuno, T. Kasukawa et al., "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs," *Nature*, vol. 420, no. 6915, pp. 563–573, 2002.

[5] T. Tuschl, M. Lagos-quintana, W. Lendeckel, J. Dammann, and R. Rauhut, "Identification of novel genes coding for small temporal RNAs," *Science*, vol. 294, no. 5543, pp. 853–858, 2015.

[6] L. J. Robin, T. A. Yario, and J. A. Steitz, "Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR," *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9667–9672, 2007.

[7] D. P. Bartel, "MicroRNAs: genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, 2004, http://www.sciencedirect.com/science/article/pii/S0092867404000455.

[8] J. P. G. Sluijter, A. van Mil, P. van Vliet et al., "MicroRNA-1 and -499 regulate differentiation and proliferation in human-derived cardiomyocyte progenitor cells," *Arteriosclerosis,*

Thrombosis, and Vascular Biology, vol. 30, no. 4, pp. 859–868, 2010.

[9] J. Kota, R. R. Chivukula, K. A. O'Donnell et al., "Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model," *Cell*, vol. 137, no. 6, pp. 1005–1017, 2009.

[10] L. Cliff Ji-Fan, G. Hong-Yi, T. Hung-Chia, W. Wei-Lun, and W. Jen-Leih, "miR-122 targets an anti-apoptotic gene, Bcl-w, in human hepatocellular carcinoma cell lines," *Biochemical and Biophysical Research Communications*, vol. 375, no. 3, pp. 315–320, 2008.

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

[12] X. Chen, C. C. Yan, X. Zhang, and Z. H. You, "Long non-coding RNAs and complex diseases: from experimental results to computational models," *Briefings in Bioinformatics*, vol. 18, no. 4, pp. 558–576, 2017.

[13] X. Chen, D. Xie, Q. Zhao, and Z. H. You, "MicroRNAs and complex diseases: from experimental results to computational models," *Briefings in Bioinformatics*, vol. 23, no. 2, pp. 525–539, 2019.

[14] L. Yanhui, W. Changliang, M. Zhengqiang et al., "ViRBase: a resource for virus-host ncRNA-associated interactions," *Nucleic Acids Research*, vol. 43, no. D1, pp. 578–582, 2014.

[15] W. Deng, H. Yan, K. Juanjuan et al., "ncRDeathDB: a comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system," *Autophagy*, vol. 11, no. 10, pp. 1917–1926, 2015.

[16] T. Zhang, P. Tan, L. Wang et al., "RNALocate: a resource for RNA subcellular localizations," *Nucleic Acids Research*, vol. 45, no. D1, pp. D135–D138, 2017.

[17] X. Chen, L. Wang, J. Qu, N.-N. Guan, and J.-Q. Li, "Predicting miRNA-disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, 2018.

[18] X. Chen, D. Xie, L. Wang, Q. Zhao, Z.-H. You, and H. Liu, "BNPMDA: bipartite network projection for MiRNA-disease association prediction," *Bioinformatics*, vol. 34, no. 18, pp. 3178–3186, 2018.

[19] X. Chen and L. Huang, "LRSSLMDA: laplacian regularized sparse subspace learning for MiRNA-disease association prediction," *PLoS Computational Biology*, vol. 13, no. 12, article e1005912, 2017.

[20] Z.-H. You, Z.-A. Huang, Z. Zhu et al., "PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction," *PLoS Computational Biology*, vol. 13, no. 3, article e1005455, 2017.

[21] Q. Zou, J. Li, L. Song, X. Zeng, and G. Wang, "Similarity computation strategies in the microRNA-disease network: a survey," *Briefings in Functional Genomics*, vol. 15, no. 1, pp. 55–64, 2016.

[22] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks," *Briefings in Bioinformatics*, vol. 17, no. 2, pp. 193–203, 2016.

[23] X. Zeng, L. Liu, L. Lü, and Q. Zou, "Prediction of potential disease-associated microRNAs using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.

[24] W. Tang, S. Wan, Z. Yang, A. E. Teschendorff, and Q. Zou, "Tumor origin detection with tissue-specific miRNA and DNA methylation markers," *Bioinformatics*, vol. 34, no. 3, pp. 398–406, 2017.

[25] J. Yu, P. Ping, L. Wang, L. Kuang, X. Li, and Z. Wu, "A novel probability model for LncRNA-disease association prediction based on the naïve bayesian classifier," *Genes*, vol. 9, no. 7, p. 345, 2018.

[26] C. Xing and Y. Gui-Ying, "Novel human lncRNA-disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, 2013.

[27] R. Gernapudi, B. Wolfson, Y. Zhang et al., "miR-140 promotes expression of long non-coding RNA NEAT1 in adipogenesis," *Molecular and Cellular Biology*, vol. 36, no. 1, p. 30, 2015.

[28] B. K. Dey, K. Pfeifer, and A. Dutta, "The h19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration," *Genes & Development*, vol. 28, no. 5, pp. 491–501, 2014.

[29] Y. Yilong, M. Jun, X. Yixue et al., "Knockdown of long non-coding RNA xist exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152," *Cancer Letters*, vol. 359, no. 1, pp. 75–86, 2015.

[30] W. Xinyu, L. Meng, W. Zhiqiong et al., "Silencing of long noncoding RNA MALAT1 by miR-101 and miR-217 inhibits proliferation, migration, and invasion of esophageal squamous cell carcinoma cells," *Journal of Biological Chemistry*, vol. 290, no. 7, pp. 3925–3935, 2015.

[31] Z. Er-Bao, K. Rong, Y. Dan-Dan et al., "Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a," *Oncotarget*, vol. 5, no. 8, pp. 2276–2292, 2014.

[32] J. You, Y. Zhang, B. Liu et al., "MicroRNA-449a inhibits cell growth in lung cancer and regulates long noncoding RNA nuclear enriched abundant transcript 1," *Indian Journal of Cancer*, vol. 51, no. 7, p. 77, 2014.

[33] S. Emmrich, A. Streltsov, F. Schmidt, V. Thangapandi, D. Reinhardt, and J.-H. Klusmann, "LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia," *Molecular Cancer*, vol. 13, no. 1, p. 171, 2014.

[34] L. Amy, T. Candi, J. Wen et al., "Novel long noncoding RNAs are regulated by angiotensin ii in vascular smooth muscle cells," *Circulation Research*, vol. 113, no. 3, pp. 266–278, 2013.

[35] Z. Miaojun, C. Qin, L. Xin et al., "lncRNA H19/miR-675 axis represses prostate cancer metastasis by targeting TGFBI," *FEBS Journal*, vol. 281, no. 16, pp. 3766–3775, 2015.

[36] V. Huang and L.-C. Li, "miRNA goes nuclear," *RNA Biology*, vol. 9, no. 3, pp. 269–273, 2012.

[37] L. M. Vaschetto, "miRNA activation is an endogenous gene expression pathway," *RNA Biology*, vol. 15, no. 6, pp. 1–3, 2018.

[38] R. Ramchandran and P. Chaluvally-Raghavan, "miRNA-mediated RNA activation in mammalian cells," in *Advances in Experimental Medicine and Biology*, Springer, Singapore, 2017.

[39] S. Zhou, Z. Xuan, L. Wang, P. Ping, and T. Pei, "A novel model for predicting associations between diseases and lncRNA-miRNA pairs based on a newly constructed bipartite network," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 6789089, 11 pages, 2018.

[40] H. Peng, C. Lan, Y. Zheng, G. Hutvagner, D. Tao, and J. Li, "Cross disease analysis of co-functional microRNA pairs on a reconstructed network of disease-gene-microRNA tripartite," *BMC Bioinformatics*, vol. 18, no. 1, p. 193, 2017.

[41] H. Zhao, L. Kuang, L. Wang et al., "Prediction of microRNA-disease associations based on distance correlation set," *BMC Bioinformatics*, vol. 19, no. 1, p. 141, 2018.

[42] P. Ping, L. Wang, L. Kuang, S. Ye, M. F. B. Iqbal, and T. Pei, "A novel method for lncRNA-disease association prediction based on an lncRNA-disease association network,"

IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 99, p. 1, 2018.

[43] W. Lei, P. Ping, L. Kuang, S. Ye, F. M. B. Lqbal, and T. Pei, "A novel approach based on bipartite network to predict human microbe-disease associations," *Current Bioinformatics*, vol. 13, no. 2, pp. 141–148, 2018.

[44] S. Yamaguchi, S. Imoto, and M. Satoru, "Network-based predictions and simulations by biological state space models: search for drug mode of action," *Journal of Computer Science & Technology*, vol. 25, no. 1, pp. 131–153, 2010.

[45] L. Xin, S. Ulf, S. K. Gupta et al., "Computational analysis of target hub gene repression regulated by multiple and co-operative miRNAs," *Nucleic Acids Research*, vol. 40, no. 18, pp. 8818–8834, 2012.

[46] C. Geng, W. Ziyun, W. Dongqing et al., "LncRNADisease: a database for long-non-coding RNA-associated diseases," *Nucleic Acids Research*, vol. 41, no. D1, pp. D983–D986, 2013.

[47] Y. Wang, L. Chen, B. Chen et al., "Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network," *Cell Death & Disease*, vol. 4, no. 8, p. e765, 2013.

[48] T. Cui, L. Zhang, Y. Huang et al., "MNDR v2.0: an updated resource of ncRNA-disease associations in mammals," *Nucleic Acids Research*, vol. 46, no. D1, pp. D371–D374, 2018.

[49] S. Ning, Z. Jizhou, P. Wang et al., "Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers," *Nucleic acids research*, vol. 44, no. D1, pp. D980–D985, 2015.

[50] Y. Jiang, Q. Wang, Y. Hao et al., "miR2Disease: a manually curated database for microRNA deregulation in human disease," *Nucleic Acids Research*, vol. 37, pp. D98–D104, 2009.

[51] L. Yang, Q. Chengxiang, T. Jian et al., "HMDD v2.0: a database for experimentally supported human microRNA and disease associations," *Nucleic Acids Research*, vol. 42, no. D1, pp. D1070–D1074, 2014.

[52] X. Boya, D. Qin, H. Hongjin, and W. Di, "miRCancer: a microRNA-cancer association database constructed by text mining on literature," *Bioinformatics*, vol. 29, no. 5, pp. 638–644, 2013.

[53] L. Jun-Hao, L. Shun, Z. Hui, Q. Liang-Hu, and Y. Jian-Hua, "starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D92–D97, 2014.

[54] K. Ana and G. J. Sam, "miRBase: annotating high confidence microRNAs using deep sequencing data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D68–D73, 2014.

[55] S. Lynn Marie, A. Cesar, N. Suvarna et al., "Disease ontology: a backbone for disease semantic integration," *Nucleic Acids Research*, vol. 40, no. D1, pp. D940–D946, 2012.

[56] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, 2010.

[57] X. Chen, Z. H. You, G. Y. Yan, and D. W. Gong, "IRWRLDA: improved random walk with restart for lncRNA-disease association prediction," *Oncotarget*, vol. 7, no. 36, pp. 57919–57931, 2016.

[58] Y. A. Huang, X. Chen, Z. H. You, D. S. Huang, and K. C. C. Chan, "ILNCSIM: improved lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 18, pp. 25902–25914, 2016.

[59] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[60] L. De Lathauwer and B. De Moor, "A multi-linear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[61] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensor," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[62] T. G. Kolda and T. Gibson, "Multilinear operators for higher-order decompositions," Technical Reports, Office of Scientific & Technical Information, Oak Ridge, TN, USA, 2006.

[63] H. S. D. D. Lee, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[64] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, and D. Dong, "LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1034–D1037, 2019.

[65] Y. Yi, Y. Zhao, C. Li et al., "RAID v2.0: an updated resource of RNA-associated interactions across organisms," *Nucleic Acids Research*, vol. 45, no. D1, pp. D115–D118, 2016.

[66] X. Chen, C. C. Yan, and X. Zhang, "WBSMDA within and between score for miRNA-disease association prediction," *Scientific Reports*, vol. 6, no. 1, p. 21106, 2016.

[67] X. Chen and G. Y. Yan, "Semi-supervised learning for potential human microRNA-disease associations inference," *Scientific Reports*, vol. 4, no. 1, p. 5501, 2014.

[68] H. J. Donahue and D. C. Genetos, "Genomic approaches in breast cancer research," *Briefings in Functional Genomics*, vol. 12, no. 5, pp. 391–396, 2013.

[69] K. Karagoz, R. Sinha, and K. Y. Arga, "Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways," *OMICS: A Journal of Integrative Biology*, vol. 19, no. 2, pp. 115–130, 2015.

[70] J. Meng, P. Li, Q. Zhang, Z. Yang, and S. Fu, "A four-long non-coding RNA signature in predicting breast cancer survival," *Journal of Experimental & Clinical Cancer Research*, vol. 33, no. 1, p. 84, 2014.

[71] Y. Wang, X. Feng, R. Jia et al., "Microarray expression profile analysis of long non-coding RNAs of advanced stage human gastric cardia adenocarcinoma," *Molecular Genetics and Genomics*, vol. 289, no. 3, pp. 291–302, 2014.

[72] A. I. Phipps, N. M. Lindor, M. A. Jenkins et al., "Colon and rectal cancer survival by tumor location and microsatellite instability," *Diseases of the Colon & Rectum*, vol. 56, no. 8, pp. 937–944, 2013.

[73] S. Pita-Fernández, S. Pértega-Díaz, B. López-Calviño et al., "Diagnostic and treatment delay, quality of life and satisfaction with care in colorectal cancer patients: a study protocol," *Health and Quality of Life Outcomes*, vol. 11, no. 1, p. 117, 2013.

[74] V. H. Chong, M. S. Abdullah, P. U. Telisinghe, and A. Jalihal, "Colorectal cancer: incidence and trend in Brunei Darussalam," *Singapore Medical Journal*, vol. 50, no. 11, pp. 1085–1089, 2009.

[75] G. A. Gmyrek, M. Walburg, C. P. Webb et al., "Normal and malignant prostate epithelial cells differ in their response to hepatocyte growth factor/scatter factor," *American Journal of Pathology*, vol. 159, no. 2, pp. 579–590, 2001.

[76] P. C. Walsh and A. W. Partin, "Family history facilitates the early diagnosis of prostate carcinoma," *Cancer*, vol. 80, no. 9, pp. 1871–1874, 2015.