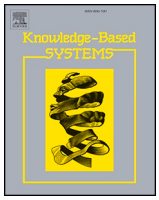
Knowledge-Based Systems 291 (2024) 111622

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Bi-SGTAR: A simple yet efficient model for circRNA-disease association prediction based on known association pair only

Shiyuan Li [a], Qingfeng Chen [a,*], Zhixian Liu [b], Shirui Pan [c], Shichao Zhang [d]

[a] School of Computer, Electronics and Information, Guangxi University, Guangxi 530004, China
[b] School of Electronics and Information Engineering, Beibu Gulf University, Qinzhou, Guangxi 535011, China
[c] School of Information and Communication Technology, Griffith University, QLD 4222, Australia
[d] School of Computer Science and Engineering, Central South University, Hunan 410083, China

ARTICLE INFO

Keywords:
Circular RNAs
circRNA-disease association
Biological network
Machine learning
Sparse gating

ABSTRACT

Identifying circRNA (circular RNA) associated with diseases holds promise as diagnostic and prognostic biomarkers, offering potential avenues for novel therapeutics. Several computational methods have been designed to predict circRNA-disease associations. Unfortunately, current computational models face issues stemming from the integration of data from multiple sources, leading to blind spots in data combination and increased model complexity. Thus, this article introduces a novel method named Bi-SGTAR (Bi-view Sparse Gating and True Association Regression). Notably, Bi-SGTAR demonstrates comparable performance to existing multi-source information fusion methods while utilizing only known circRNA-disease association pairs. In contrast to previous methods, the model divides the adjacency matrix into two views and employs an encoder with sparse gating to assess the reliability of all associations. Additionally, a supervised reconstructor is employed to define the true association probability, quantifying the truthfulness of all associations. The Encoding-Reconstruction-Regression (ERR) framework adeptly merges both reliable and truthful associations from both views. The experimental results unequivocally show that Bi-SGTAR surpasses state-of-the-art models across seven circRNA-disease datasets, one lncRNA-disease dataset, and one microbe-drug dataset, with fewer data needed.

1. Introduction

CircRNA, categorized as a non-coding RNA, distinguishes itself by lacking 3′ and 5′ polyadenylated tails. Unlike linear RNA, circRNA exhibits greater stability due to its ability to evade exonuclease digestion [1,2]. Extensive studies have revealed that circRNA, with diverse biological functions such as miRNA sponge activity and modulation of gene expression, interacts with Argonaute and RNA polymerase II proteins, influencing nuclear DNA transcription [3,4]. Its tissue-specific expression patterns underscore its potential as a biomarker [5]. A growing body of evidence points to the pivotal involvement of circRNA in the development and progression of various diseases [6], spanning cancer [5,7,8], neurodegenerative diseases [9,10], and cardiovascular diseases [11]. Hence, unraveling the intricate associations between circRNAs and diseases becomes paramount for comprehending the underlying molecular mechanisms. Moreover, this understanding is instrumental in identifying potential targets for drug development and advancing precision medicine.

The intricate nature of circRNA, coupled with the challenges and costs inherent in traditional experiments, has left numerous valuable associations undiscovered [12]. In response, computational approaches have emerged to leverage dependable circRNA-disease datasets [13–16]. These approaches fall into broad categories, including information propagation, matrix completion, machine learning (ML), and deep learning (DL). As for information propagation, Fan et al. [17] introduced KATZCDA (KATZ CircRNA-Disease Association), a model that leverages a heterogeneous network and the KATZ metric to predict associations between circRNA and diseases. Zhao et al. developed a novel computational method called IBNPKATZ, which integrates bipartite networks based on KATZ to forecast potential associations between circRNAs and diseases [18]. Based on the label propagation strategy, Zhang et al. proposed CD-LNLP. They initially constructed similarity matrices for circRNA and disease, followed by implementing label propagation based on these matrices. The circRNA and disease similarity matrices' calculation results were combined to predict new associations [19].

* Corresponding author.
E-mail addresses: shiy.li@alu.gxu.edu.cn (S. Li), qingfeng@gxu.edu.cn (Q. Chen).

https://doi.org/10.1016/j.knosys.2024.111622
Received 25 November 2023; Received in revised form 8 February 2024; Accepted 6 March 2024
Available online 7 March 2024

Strategies that directly follow social networks (such as KATZ) to predict circRNA-disease associations may encounter challenges related to data sparsity and domain bias. Matrix completion methods offer an alternative by filling in missing data. Li et al. introduced a novel method called SIMCCDA, which leverages the principles of speedup inductive matrix completion to predict disease-related circRNAs [20]. Similarly, Wei et al. developed a similar approach involving decomposing known association information using a matrix factorization algorithm, updating the circRNA-disease interaction profile, and effectively predicting associations [21]. In a separate study, Peng et al. introduced an approach known as RNMFLP for predicting circRNA-disease associations. This method combines non-negative matrix factorization with a label propagation algorithm based on circRNA similarity matrix and disease similarity matrix to enhance prediction accuracy [22]. While a variety of matrix completion methods can be used to predict associations, it is important to note that assumptions about data distribution and noise can potentially introduce risks.

Machine learning-based methods offer a promising avenue for improving the accuracy and robustness of circRNA-disease association predictions. Hongze et al. applied SVM (Support Vector Machines) to construct an RNA classification model for diagnosing lumbar disc herniation. This model was then employed to identify the competitive endogenous RNA regulatory network, shedding light on potential mechanisms associated with feature genes [23]. Furthermore, Lei et al. integrated random walk with restart and *k*-nearest neighbor methods to predict circRNA-disease associations. This method effectively leverages global network topology information to assign weights to the features, thereby enhancing prediction accuracy [24].

While various traditional machine learning methods contribute to predicting circRNA-disease associations, challenges arise with manual feature selection and the explosive amount of data. In contrast, the automatic feature extraction of Deep Learning presents a promising approach to augment circRNA-disease association predictions. Wang L et al. proposed an efficient computational model based on deep CNN (Convolution Neural Network), integrating multiple sources of information. This model exhibits robust performance on the CircR2Disease dataset, surpassing methods such as SVM classifiers applied to the same dataset [25]. In the similar vein, Deepthi et al. employed a deep autoencoder to learn more compact and high-level features. These features were then input into a deep neural network for predicting potential associations [26]. The application of GNN (Graph Neural Network) technology is noteworthy for circRNA-disease association prediction, utilizing network topology structures to extract node features from heterogeneous networks. For instance, Wang et al. developed a model that constructs a unified feature descriptor based on various similarities between circRNA and diseases, incorporating graph convolution to capture potential associations between nodes [27]. Additionally, Lan et al. devised KGANCDA, an algorithm based on the knowledge graph attention network. This algorithm employs graph attention networks to aggregate entity features and tackles the data sparsity by utilizing high-order neighbor information from multi-source associations [28].

In summary, current methods for predicting circRNA-disease associations have achieved significant results on benchmark databases. However, some limitations persist: 1) Known circRNA-disease associations are scarce, resulting in a sparse association matrix and false negative noise. Constructing similarity or heterogeneous networks based on this information may propagate false negatives. 2) While integrating multi-source data can mitigate false negatives, strict data requirements pose challenges, almost no medical institutions or public databases can consistently meet these demands. 3) Many existing methods focus extensively on integrating multi-source data, often overlooking the evaluation of data reliability and truthfulness. Hence, there is a need to consider achieving a lightweight, scalable, and cross-platform algorithm that utilizes less information while maintaining optimal performance. Addressing these challenges could pave the way for more robust circRNA-disease association prediction methods.

In this paper, we proposed Bi-SGTAR for predicting associations using single-source information. Comprising a sparse quality control (SQC) module and a true association regression module, Bi-SGTAR strategically utilizes known circRNA-disease associations (adjacency matrix) by splitting them into RNA and disease views. In the sparse quality control module, the encoder gauges the reliability of associations within each view, employing sparse gating to suppress unreliable associations. The true association regression module then establishes the true association probability (TAP) through a supervised reconstructor. The regressor utilizes the TAP to approximate and identify more truthful associations in the view. To dynamically integrate reliable and true associations across views, our proposed Encoding-Reconstruction-Regression (ERR) framework comes into play. The contributions of our work can be summarized as follows:

1) Leveraging only known association information from the adjacency matrix, thereby requiring less data.
2) Bi-SGTAR has a simple network architecture that enables reliable dynamic integration using only common ML techniques.
3) Our experimental evaluations on nine datasets demonstrate that Bi-SGTAR achieves comparable performance to multi-source information fusion models.

2. Proposed method

The goal of association prediction is to construct a mapping between node features $x \in \mathbb{R}^h$ and circRNA-disease associations $y \in \mathbb{R}^{R\times D}$, where h represents the feature dimension of nodes, and R and D represent the total number of RNA and disease nodes, respectively. Formally, to integrate multi-source information and mine potential associations, traditional association prediction models often train neural networks $f(x, A)\rightarrow y$ based on the known association matrix $A \in \mathbb{R}^{R\times D}$.

In order to identify reliable and truthful associations with minimal information (i.e., adjacency matrix), we introduce a lightweight circRNA-disease association prediction model termed Bi-SGTAR. Departing from conventional approaches that predominantly focus on multi-source information fusion, this paper shifts its perspective towards achieving efficient and lightweight association prediction under the constraint of single-source information. Specifically, according to the known circRNA and disease association pairs, the adjacency matrix A can be obtained. Furthermore, we use $A^c = A$ and $A^d = A^T$ as inputs in the circRNA and disease view, where $A^c \in \mathbb{R}^{R\times D}$ and $A^d \in \mathbb{R}^{D\times R}$. Then, Bi-SGTAR models association reliability (detailed in Section 2.1) and true association probability (described in Section 2.2) at the bi-view. An encoding-reconstruction-regression framework is proposed in Section 2.3. Finally, to effectively optimize Bi-SGTAR, well-crafted objective functions are introduced in Section 2.4.

2.1. Sparse quality control module

Motivations: In human social networks, the likelihood of **B** and **C** knowing each other or having an association increases if **A** knows both **B** and **C**, owing to the common neighbor node **A**'s social network [29]. As networks grow larger and associations become more intricate, the small-world theory suggests that no more than six intermediate nodes can connect any two nodes [30]. This underlines the prevalence of implicit associations in social networks and underscores the significance of implicit associations in determining potential links between unfamiliar nodes.

Building on this insight and considering the hypothesis that similar RNA molecules may be associated with the same disease [31], we posit that there are universal implicit associations in circRNA-disease biological networks. Evaluating these associations is crucial for better identifying potential associations.

Example 1: In Fig. 1, the initial association between c_1 and d_4 may

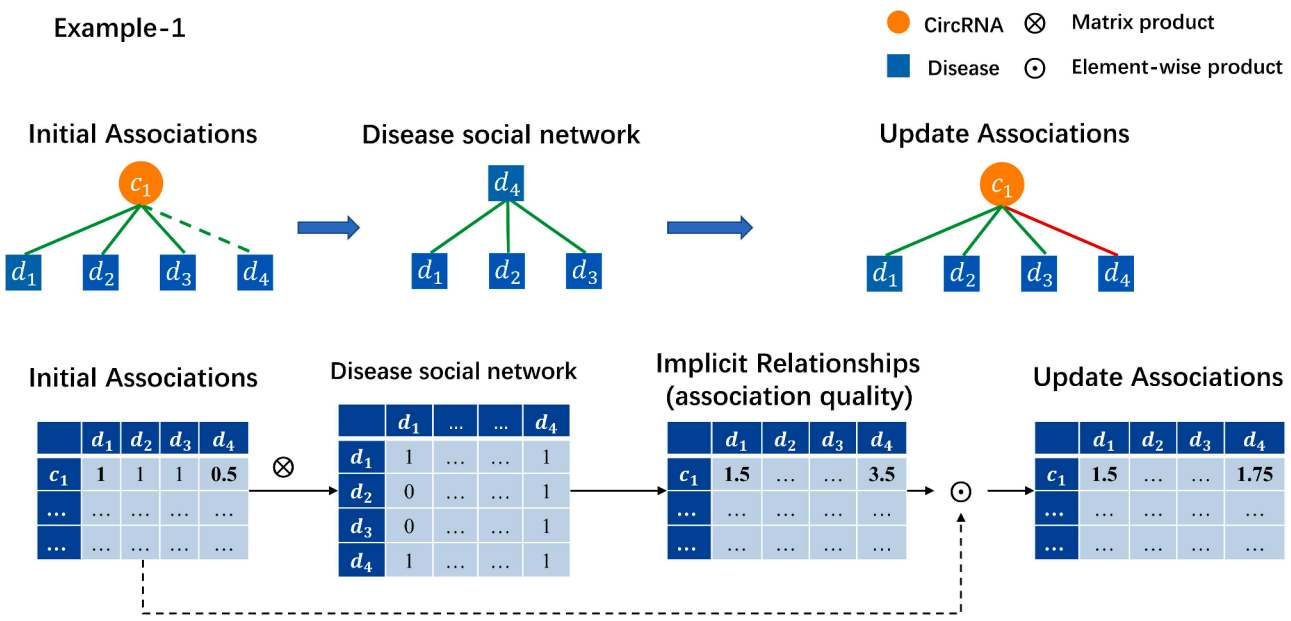


Fig. 1. Upon incorporating the implicit association into the initial pairing as the measure of association quality, the likelihood shifts, making c_1 more probable to be associated with d_4 rather than d_1.

seem weak at first glance, but when considering implicit associations mapped through the disease social network, a stronger association emerges. Updating the initial association with these implicit associations unveils a stronger underlying association between c_1 and d_4 as opposed to c_1 and d_1. Therefore, the proper construction of the disease social network proves to be crucial for accurate association assessment.

Based on the above hypothesis, we propose constructing a sparse quality control module to quantify all potential associations, as shown in Fig. 2. Specifically, considering the universality of implicit associations, all unknown samples are assigned initial values through numerical smoothing: $\widetilde{A}^v = u + (1 - 2u) * A^v$, where A^v represents the initial association matrix of view v (i.e., c and d) and u is the smoothing factor. Subsequently, we introduce a quality encoder to dynamically model node social network, quantifying the strength of implicit associations, i.e., association reliability:

$$W^v = \sigma(\widetilde{A}^v \Theta^v + b^v), \tag{1}$$

where W^v represents the association reliability, the Θ^v refers to the social network of view v, which can be learned and optimized, and the activation function σ is *sigmoid*. In addition, combining the sparse prior of known associations, we introduce a sparse gating strategy aimed at screening out more probable potential associations. Specifically, we employ L1 regularization loss for the module: $\mathscr{L}_v^{SQC} = \| \Theta^v \|_1$ to enforce sparse control. This approach contrasts with directly computing similarity based on inherent information, such as Disease Ontology (DO, i.e., DOID: 263, kidney cancer). The disease social network (Θ^d) can adaptively adjust the social relationship between diseases during the model learning process, enhancing the accuracy of potential association predictions. Likewise, the modeling of RNA social networks (Θ^c) can be approached from a disease view to yield more reliable association.

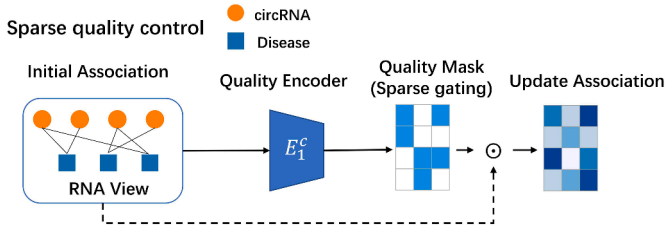


Fig. 2. The flowchart of the sparse quality control module.

2.2. True association regression module

A supervised autoencoder is instrumental in compressing noisy high-dimensional data and integrating label information to guide data reconstruction, thereby playing a crucial role in circRNA-disease association prediction. Formally, $p^v(y \mid x^v) = d^v(z^v)$, $z^v = f^v(x^v)$, where f^v and d^v representing the encoder and reconstructor, respectively. The reconstruction loss is calculated as follows:

$$\mathscr{L}_v^{REC} = -\frac{1}{|E|}\sum_{i,j\in E}\left(y_{ij}\log\left(p_{ij}^v\right) + (1 - y_{ij})\log\left(1 - p_{ij}^v\right)\right), \tag{2}$$

where, E represents the association set between circRNA and disease, and y_{ij} and p_{ij}^v represent the true and predicted value of the association between circRNA i and disease j, respectively. However, relying solely on the initial association as a supervision signal is coarse-grained. Inspired by true class probability, we define true association probability, aimed at rectifying errors in prediction results using real label information [32,33]. To elaborate, for the view v, we can derive the reconstructed association probability $p^v(y \mid x^v)$ based on the reconstructor. When considering the corresponding true association y, the true association probability (TAP) is expressed as follows:

$$TAP^v = y \odot p^v(y \mid x^v), \tag{3}$$

where $\odot$ represents the elements-wise multiplication.

For clarity, we provide typical examples to explain the aforementioned formula. Considering an initial association matrix A, let's assume that the smoothed matrix $\widetilde{A}$ has $\widetilde{A}_{ij} = 0.99$ if circRNA i and disease j are associated, and $\widetilde{A}_{ij} = 0.01$ otherwise. The reconstructor d^v predicts the association probability matrix P, and the true association probability matrix $TAP_{ij} = P_{ij} \cdot \widetilde{A}_{ij}$. The efficacy of the true association probability can be demonstrated through the following example.

Example 2: For $\widetilde{A}_{ij} = 0.01$, when the reconstructor provides a prediction $P_{ij} = 0.9$, $TAP_{ij} = 0.009 < 0.9$, effectively alleviating this error in the result. For $\widetilde{A}_{ij} = 0.01$, when the reconstructor gives a prediction

$P_{ij} = 0.1$, $TAP_{ij} = 0.001 < 0.1$, bolstering confidence in the result.

Example 3: For $\widetilde{A}_{ij} = 0.99$, when the reconstructor predicts $P_{ij} = 0.1$, $TAP_{ij} = 0.099$. Even if not numerically corrected, the association probability based on true information is maximally preserved compared to true negative samples in Example 2 ($TAP_{ij} = 0.001$).

The true association is however unknown during the inference stage, so we train a parallel true association regressor (TAR): $\widetilde{d}^v(z^v) \rightarrow TAP^v$ to obtain a probabilistic approximation. The corresponding loss function is as follows:

$$\mathscr{L}_v^{TAR} = \frac{1}{N^v}||\widetilde{X}^v - TAP^v||_2^2 + \mathscr{L}_v^{REC}, \tag{4}$$

where $\widetilde{X}^v = \widetilde{d}^v(z^v)$. Prediction by regressor not only ensures the truthfulness of the predicted association but also learns fine-grained information at the probability level.

2.3. Encoding-reconstruction-regression framework

According to Sections 2.1 and 2.2, the sparse quality control (SQC) module and the true association regression (TAR) module can be established, respectively. To leverage the strengths of both modules, we opted for a nested fusion structure: encoding-reconstruction-regression, illustrated in Fig. 3. The model begins with performing sparse quality control on the known circRNA-disease association to quantify the reliability of each association. Given the smooth matrix $\widetilde{A}^v$, association reliability is computed as $W^v = \sigma(\widetilde{A}^v\Theta^v + b^v)$. Sparse gating is then applied to retain high-reliability associations and suppress low-reliability ones: $X^v = [W^v \odot \widetilde{A}^v]$. Subsequently, more uncertain associations are introduced, providing additional abundant input information for the subsequent autoencoder to reconstruct the association.

Then, we will establish the true association probability. In accordance with Section 2.2, we achieve the view-specific encoder f^v, reconstructor d^v, and regressor $\widetilde{d}^v$ to unearth the potential associations. The encoder f^v is employed to compress the information from a specific view. Formally, the low-dimensional space information representation for each view is denoted: $z^v = f^v(X^v)$. Simultaneously, the regressor utilizes $\widetilde{X}^v = \widetilde{d}^v(z^v)$ and Eq.(4) to mine view-specific truthful associations. Finally, the different views are weighted and fused to yield more comprehensive associations:

$$\widetilde{X} = \sum_{v\in V}\alpha^v\widetilde{X}^v, \ \sum_{v\in V}\alpha^v = 1, \tag{5}$$

where α^v represents the weight of view v.

2.4. Model training and optimization

In this subsection, we present the objective function designed to optimize Bi-SGTAR. This function is carefully constructed across three levels: intra-view, inter-view, and predicted loss.

For the intra-view loss, it mainly consists of two modules, SQC and TAR:

$$\mathscr{L}_v^{intra} = \beta\mathscr{L}_v^{TAR} + (1-\beta)\mathscr{L}_v^{SQC}, \tag{6}$$

where β is weight to balance the modules SQC and TAR.

Inter-view loss is designed for fusing disease with RNA views as follows:

$$\mathscr{L}^{inter} = \sum_{v\in V}\alpha^v\mathscr{L}_v^{intra}, \tag{7}$$

where α^v is the same as weight of fusing different views prediction.

In addition, in order to enhance the accuracy of the final prediction, the binary cross entropy loss is introduced as follows:

$$\mathscr{L}^{CLS} = -Ylog(\widetilde{X}) - (1-Y)\log(1-\widetilde{X}), \tag{8}$$

where $\widetilde{X}$ is the finally fusing association matrix. Then, the total loss function can be written as:

$$\mathscr{L} = (1-\zeta)\mathscr{L}^{inter} + \zeta\mathscr{L}^{CLS} + \eta||\Phi||_2^2, \tag{9}$$

where ζ represents the weight of the fusion part, Φ represents the parameter set of the whole model, η is used to control the complexity of the model to achieve tradeoff, and the model can be trained by optimizing the loss function $\mathscr{L}$.

3. Experimental results

In this section, we evaluate Bi-SGTAR through extensive experiments. Initially, we detail the experimental setup, including datasets, implementation details, and baseline methods. Subsequently, we provide the detailed experimental results on performance comparison and analytical experiments.

3.1. Experimental setup

Datasets. Given the exponential increase in circRNA numbers, researchers have contributed to a plethora of circRNA databases and computational tools, with over 20 such databases having emerged by 2023. Since Bi-SGTAR only requires known associations as input, we tested it on three different datasets: circRNA-disease, lncRNA-disease, and microbe-drug. As for circRNA-disease, we curated pertinent public datasets from a comprehensive study [31]. The datasets encompass Circ2Traits [34], CircR2Disease [35], CircR2Cancer [36], circRNADisease [37], and Circad [38]. Moreover, we retrieved 2884 known microbe-drug associations involving 1720 drugs and 140 microbes from aBioflm (http://bioinfo.imtech.res.in/manojk/abiofilm/). For lncRNA-disease associations, datasets from LncRNADisease, Lnc2Cancer, and GeneRIF [39–41] were utilized, encompassing 2697 associations between 240 lncRNAs and 412 diseases. Detailed information on all datasets used in our study is presented in Table 1. It is noteworthy that, with the exception of the CircTraits dataset, most datasets exhibit high sparsity, characterized by numerous unknown associations. This observation indicates that constructing heterogeneous networks, knowledge graphs, and node features can suffer from significant unreliability.

Evaluation metrics. This study uses accuracy (Acc), recall, precision (Pre), area under ROC curve (AUC), area under precision-recall curve (AUPR), and F1 score multiple indicators to comprehensively evaluate the performance. The definitions of various indicators are as follows:

$$TPR = Recall = \frac{TP}{TP+FN}, \tag{10}$$

$$FPR = \frac{FP}{TN+FP}, \tag{11}$$

$$Acc = \frac{TP+TN}{TP+FP+FN+TN}, \tag{12}$$

$$Pre = \frac{TP}{TP+FP}, \tag{13}$$

$$F1 = \frac{2*precision*recall}{precision+recall}, \tag{14}$$

where, TP represents the number of positive samples correctly predicted. FP represents the number of samples that are incorrectly predicted as positive. TN represents the number of negative samples correctly predicted. FN represents the number of positive samples that are incorrectly ignored.

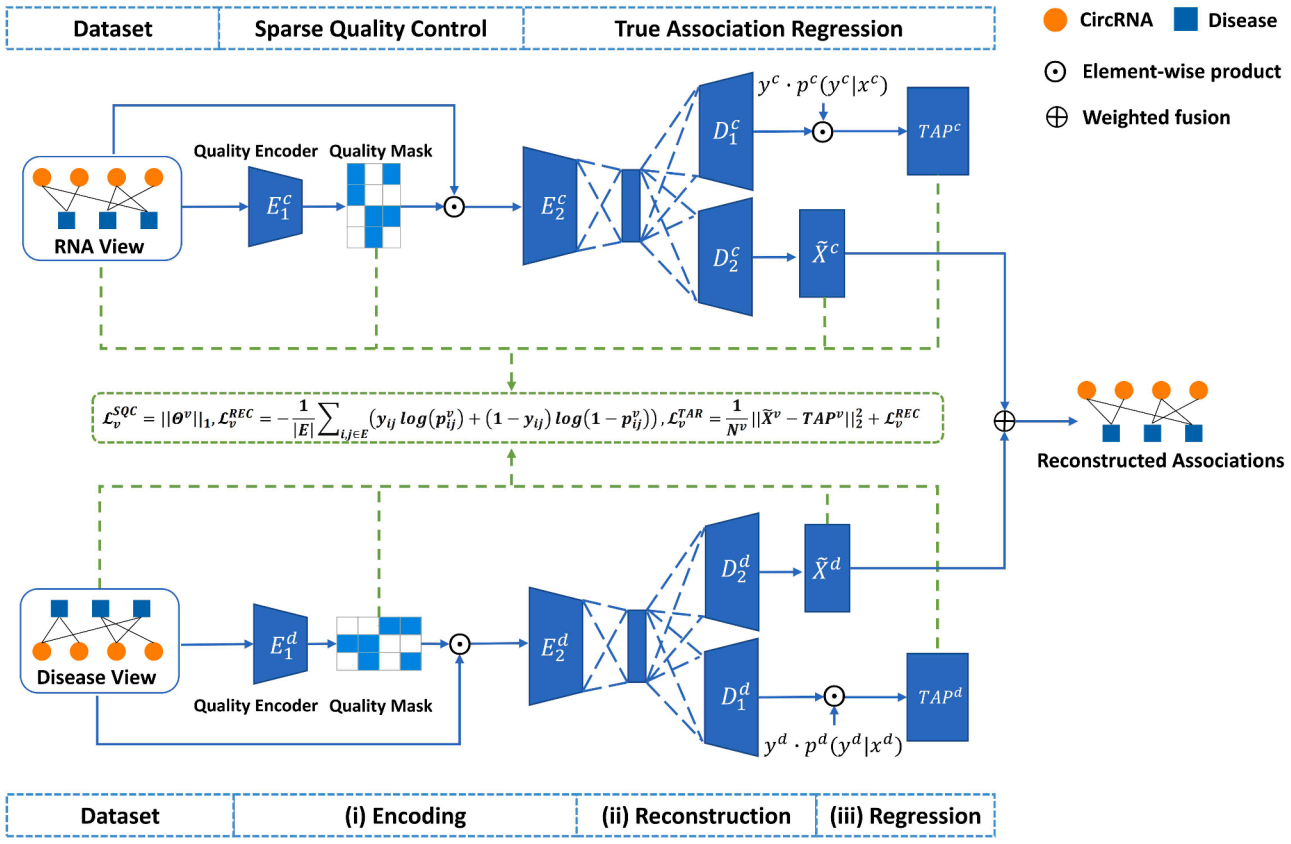


Fig. 3. Framework of Bi-SGTAR. The proposed method is mainly composed of the following steps. (i) For view v, the sparse association quality mask is obtained with the Encoder E_1^v, then a gating strategy is employed to preserve the reliable associations, and Encoder E_2^v compresses the reliable associations to the low-dimensional space for reconstruction later. (ii) The supervised reconstructor D_1^v defines the *TAP* with real labels from a low-dimensional space. (iii) The true association regressor D_2^v is used to approximate *TAP*, which is the predictive probability of the reconstructor D_1^v corresponding to the real labels.

Table 1
The details of selected datasets are in this paper.

Datasets	Data Source	ncRNAs/Microbes	Diseases/Drugs	Associations	Sparsity
Dataset 1	CircR2Disease	533	89	613	98.71%
Dataset 2	CircR2Cancer	514	62	647	97.97%
Dataset 3	circRNADisease	312	40	331	97.35%
Dataset 4	CircTraits	923	104	37,660	60.77%
Dataset 5	Circad	1265	151	1369	99.28%
Dataset 6	KGETCDA [42]	330	79	346	98.67%
Dataset 7	KGETCDA [42]	561	190	1399	98.69%
Dataset 8	LncRNADisease	240	412	2697	97.28%
Dataset 9	aBiofilm	140	1720	2884	98.81%

Implementation details. The experiments are conducted on a machine with an Intel(R) Xeon(R) Platinum 8255C CPU@2.50GHz and a single RTX 3090 GPU with 24GB GPU memory. The machine's operating system is Ubuntu 20.04. As for software versions, we use Python 3.8, Pytorch 1.11.0, and CUDA 11.3. The code is available at this URL[1]. For hyperparameter settings, we perform a specific set of grid searches to identify crucial settings for Bi-SGTAR. The search space is outlined as follows:

- Learning rate: {5e-4, 1e-3, 2e-3}
- Weight decay η: {1e-10,1e-9,1e-8}
- Hidden size: {128, 256}
- Epochs: {400, 500, 600}
- α: {0.5, 0.8}
- β: {0.4, 0.6, 0.8}
- ζ: {0.05, 0.1, 0.5, 0.8}
- Smooth factor u: {0., 1e-3, 1e-2}

State-of-the-art methods. In order to study the effectiveness of Bi-SGTAR, we compare this method with state-of-the-art algorithms in different domains. CircRNA-disease association prediction models include KATZHCDA [17], LLCDC [43], iCircDA-MF [21], RWRKNN [24], RWR [44], GMNN2CD [45], IGNSCDA [46], AE-RF [11], CD-LNLP [19], DMFCDA [47], RNMFLP [22], KGANCDA [28], KGETCDA [42]. For lncRNA-disease association prediction models, the comparison involves GAMCLDA [48], SIMCLDA [49], TPGLDA [50], SKFLDA [51], GANLDA [52], and VGAELDA [53]. Additionally, in the realm of Microbe-drug association prediction, we compared the model with six algorithms: HMDAKATZ [54], HMDA-Pred [55], NTSHMDA [56],

[1] https://github.com/Shiy-Li/Bi-SGTAR

LAGCN [57], BPNNHMDA [58], and GSAMDA [59].

3.2. Comparison with state-of-the-art methods

CircRNA-Disease: We employed a 5-fold cross-validation (CV) throughout the experiment, wherein all samples were divided into five subsets. Each subset was sequentially used as the test set, with the remaining samples forming the training set for the models. To maintain consistency in the experimental setting, GIPK (Gaussian Interaction Profile Kernel) similarity was uniformly applied in all instances involving similarity calculation, given the diverse datasets and incomplete code/data from different algorithms. Algorithm parameters were set to their recommended values, and a 5-fold CV was performed. Then, circRNA-disease association scores were ranked in descending order. The average AUC and AUPR for each method were obtained based on these scores, as shown in Fig. 4 (additional indicators are detailed in the Supplementary).

Across the three preceding small and sparse datasets, Bi-SGTAR consistently outperforms other state-of-the-art methods with AUCs of 0.8705, 0.8368, and 0.8859, respectively. Even on the two larger datasets, Dataset 4 and Dataset 5, Bi-SGTAR maintains superior performance on the sparser Dataset 5 (0.8644), while ranking third on the denser Dataset 4 (0.9409). This slight dip in performance on the dense dataset may be attributed to the strong learning ability of Bi-SGTAR leading to a mild overfitting phenomenon. Although CD-LNLP achieves the highest AUC value of 0.9654 on Dataset 4, its sensitivity to data sparsity results in poorer performance on other datasets. This underscores the significance of considering data sparsity when devising computational models for circRNA-disease association prediction, as evident in Table 1.

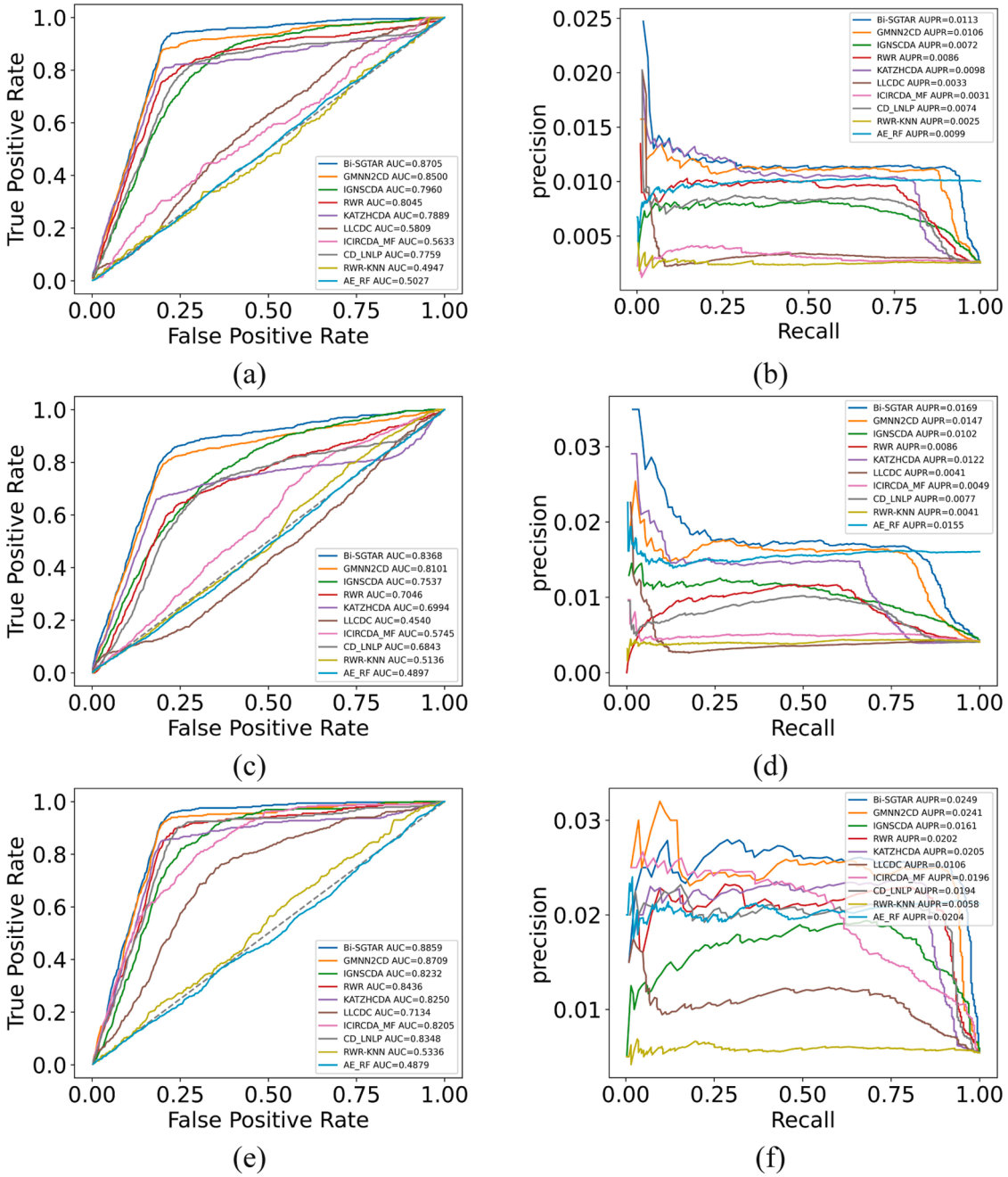


Fig. 4. The AUC and AUPR of the Bi-SGTAR and state-of-the-art model. (a), (b): Dataset 1. (c), (d): Dataset 2. (e), (f): Dataset 3. (g), (h): Dataset 4. (i), (j): Dataset 5.

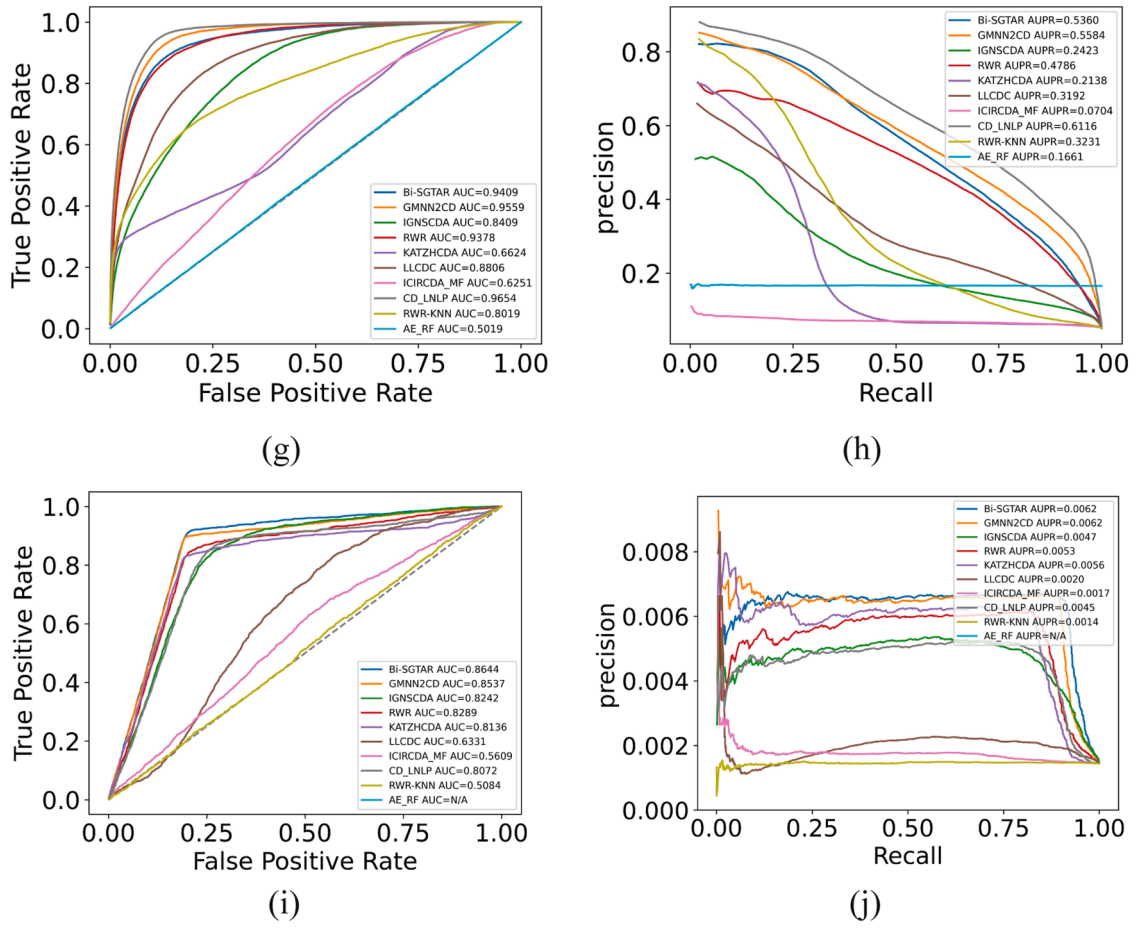


Fig. 4. (*continued*).

Moreover, to assess the substantial enhancement in model performance achieved by Bi-SGTAR, we employed a non-parametric Wilcoxon signed rank test. This test was utilized to ascertain the statistical significance of the difference between Bi-SGTAR and other methods. The null hypothesis (H_0) posits no statistical significance between the compared models, while the alternative hypothesis (H_1) suggests statistical significance. With a default P-value threshold of 0.05, the results, as detailed in Supplementary Tables 1–4, conclusively established the statistical significance of the difference, affirming the significant improvement in prediction accuracy achieved by Bi-SGTAR.

In addition, Table 2 shows the comparison of Bi-SGTAR on Datasets 6 and 7 with other methods, including knowledge graph-based approaches like KGANCDA and KGETCDA [28,42]. As can be seen from Table 2, Bi-SGTAR secures a runner-up position on Dataset 6 and surpasses all other methods on Dataset 7. It's noteworthy that the results in Table 2 are sourced from [42], indicating that all the mentioned methods utilized available multi-source data, whereas Bi-SGTAR solely relied on known circRNA-disease associations.

To summarize, Bi-SGTAR consistently achieves comparable or superior performance to more complex methods, highlighting the effectiveness of its simple, multi-layer fully connected network architecture.

Other Datasets: For the lncRNA-disease and Microbe-Drugs datasets, Table 3 presents a comprehensive comparison of AUC and AUPR values between Bi-SGTAR and other state-of-the-art models. The AUC and AUPR values for the methods on the lncRNA-Disease dataset are sourced from the study [53], with Bi-SGTAR and VGAELDA following the same experimental methodology. Similarly, the Microbe-Drugs data are obtained from the study [59], and Bi-SGTAR is compared using the same experimental approach. As illustrated in Table 3, Bi-SGTAR exhibits superior performance, surpassing other advanced methods with the highest AUC and AUPR values (0.9491, 0.0591) in Dataset 8. Furthermore, it achieves the top AUC (0.9384) on Dataset 9. Notably, in the comparative experiments on Dataset 8 and Dataset 9, all models

Table 2
The AUC and AUPR of 9 algorithms are based on circRNA-disease datasets.

	Dataset 6		Dataset 7	
Method	AUC	AUPR	AUC	AUPR
AE-RF	0.8259	0.0083	0.5895	0.0042
RWR	0.8126	0.0110	0.6403	0.0052
CD-LNLP	0.7743	0.0082	0.6384	0.0061
DMFCDA	0.6029	0.0030	0.6361	0.0048
GMNN2CD	0.8760	0.0124	0.7189	0.0073
KATZHCDA	0.7718	0.0105	0.594	0.0054
RNMFLP	0.8668	0.0117	0.5233	0.0057
KGANCDA	0.8252	0.0151	0.6395	0.0050
KGETCDA	**0.9213**	**0.0302**	0.7149	0.0081
Bi-SGTAR	0.8937	0.0125	**0.7543**	**0.0130**

Table 3
The AUC and AUPR of 7 algorithms are based on other datasets.

Dataset 8 LncRNA-disease (10-fold)			Dataset 9 Microbe-Drug (5-fold)		
Method	AUC	AUPR	Method	AUC	AUPR
GAMCLDA	0.9071	0.0378	HMDAKATZ	0.9015	0.3143
RWRlncD	0.8666	0.0258	LAGCN	0.8750	**0.3781**
MFLDA	0.6465	0.0051	NTSHMDA	0.8632	0.2040
SIMCLDA	0.8236	0.0227	HMDA-Pred	0.8094	0.0290
GANLDA	0.8834	0.0458	BPNNHMDA	0.8624	0.0543
VGAELDA	0.9287	0.0495	GSAMDA	0.9146	0.1145
Bi-SGTAR	**0.9491**	**0.0591**	**Bi-SGTAR**	**0.9384**	0.2759

utilized complete data according to their respective papers. In contrast, Bi-SGTAR solely relies on known associations, showcasing its remarkable performance even under the constraints of minimum information.

3.3. Ablation study

To gain deeper insights into the contribution of each module in Bi-SGTAR to the association prediction process, ablation experiments were conducted across multiple datasets (additional indicators are detailed in the Supplementary). Three variants were considered: Bi-SGTAR^0, Bi-SGTAR^1, and Bi-SGTAR^2. Bi-SGTAR^0 employs only one fully connected layer to model node social network, Bi-SGTAR^1 utilizes only sparse quality control, and Bi-SGTAR^2 relies solely on the true association probability regressor. Table 4 highlights the significance of the sparse quality control and true association probability regression modules in the prediction process. However, the fusion model (Bi-SGTAR) consistently demonstrates superior and more stable performance overall. Interestingly, Bi-SGTAR^0, which solely uses fully connected layers, performs well across the five circRNA-disease datasets, especially excelling in Dataset 4, where its performance is second only to the fusion model. This suggests that if the disease or RNA social network is accurately constructed, the resulting implicit associations between the nodes are already relatively accurate. The results also underscore the richness of the initial association information and emphasize the importance of utilizing it effectively in the fusion process.

3.4. De novo experiments

To assess the model's performance in predicting diseases without known circRNA-disease associations, de novo experiments were conducted. In this experiment, all known associations of each disease are removed in turn, and the remaining associations of other diseases will be used as training samples. For example, consider a disease d, its known associations were excluded as test samples, while the remaining associations were employed for training. Then, the model predicted each association of disease d, and the predicted scores were sorted in descending order, with evaluation metrics calculated accordingly. Taking three circRNA-disease association datasets as examples, Bi-SGTAR was compared with other advanced methods, and some experimental results are presented in Table 5 (additional details are provided in the Supplementary). As shown in Table 5, Bi-SGTAR achieves the highest AUC values of 0.9569 and 0.9057 on Dataset 1 and Dataset 2, respectively. Notably, most models relying on multi-source feature information suffer substantial information loss, resulting in a significant degradation of performance. Interestingly, GMNN2CD and IGNSCDA still achieve better performance. On dataset 3, GMNN2CD outperforms Bi-SGTAR, suggesting that many methods heavily rely on additional information while overlooking the deep mining of existing association information. And the multi-source information fusion does not seem to have played a real role yet. Despite the challenge posed by limited information in the de novo experiment, Bi-SGTAR maintains competitive performance compared to other methods. This highlights the effectiveness and robustness of Bi-SGTAR in predicting associations, even in scenarios with limited prior knowledge.

Table 4
The AUC, AUPR, and F1 of the ablation study are based on the circRNA-disease datasets.

Dataset	Methods	F1	AUC	AUPR
Dataset 1	Bi-SGTAR^0	0.0112	0.801	0.0098
	Bi-SGTAR^1	0.0112	0.7998	0.0096
	Bi-SGTAR^2	0.0116	0.8379	0.0101
	Bi-SGTAR	**0.0122**	**0.8718**	**0.0112**
Dataset 2	Bi-SGTAR^0	0.0161	0.7262	0.0135
	Bi-SGTAR^1	0.0161	0.7262	0.0135
	Bi-SGTAR^2	0.0162	0.7598	0.0129
	Bi-SGTAR	**0.0182**	**0.8354**	**0.0164**
Dataset 3	Bi-SGTAR^0	0.0246	0.8402	0.0229
	Bi-SGTAR^1	0.0247	0.8394	0.0229
	Bi-SGTAR^2	0.0243	0.8551	0.0213
	Bi-SGTAR	**0.0257**	**0.8835**	**0.0247**
Dataset 4	Bi-SGTAR^0	0.1983	0.9051	0.4922
	Bi-SGTAR^1	0.1686	0.8196	0.3229
	Bi-SGTAR^2	0.1865	0.8883	0.3592
	Bi-SGTAR	**0.2072**	**0.9382**	**0.5301**
Dataset 5	Bi-SGTAR^0	0.0068	0.8367	**0.0062**
	Bi-SGTAR^1	0.0068	0.8366	**0.0062**
	Bi-SGTAR^2	0.0068	0.855	0.0061
	Bi-SGTAR	**0.0069**	**0.8597**	**0.0062**

Table 5
The AUC and AUPR of 10 algorithms are based on a de novo experiment.

Method	Dataset 1		Dataset 2		Dataset 3	
	AUC	AUPR	AUC	AUPR	AUC	AUPR
KATZHCDA	0.1849	0.0148	0.2191	0.0213	0.1539	0.0195
CD-LNLP	0.4624	0.0119	0.5083	0.0193	0.5911	0.0279
LLCDC	0.1861	0.0129	0.2259	0.023	0.1606	0.0203
iCircDA-MF	0.5192	0.0192	0.5376	0.0206	0.6845	0.0778
RWRKNN	0.4668	0.012	0.5179	0.0198	0.5909	0.0279
RWR	0.4624	0.0119	0.5083	0.0193	0.5911	0.0279
GMNN2CD	0.9257	**0.3285**	0.898	**0.2425**	**0.9619**	0.3114
IGNSCDA	0.9224	0.3064	0.8834	0.2	0.9283	0.2126
AE-RF	0.5186	0.0122	0.4506	0.0192	0.5192	0.0257
Bi-SGTAR	**0.9569**	0.2451	**0.9057**	0.2027	0.9495	**0.3217**

3.5. Case study

To evaluate Bi-SGTAR's ability to predict unknown associations, case studies focusing on two diseases, lung cancer and bladder cancer, were conducted using Dataset 3. In these experiments, all known circRNA-disease associations provided by Dataset 3 were utilized to train Bi-SGTAR, which subsequently predicted unknown associations. The top 10 circRNA-disease associations with the highest prediction scores were extracted and validated against recent literature. Table 6 summarizes the results of the case study for lung cancer and bladder cancer. It is important to note that associations not confirmed by current literature do not necessarily negate their potential existence. The case study results underscore Bi-SGTAR as a promising tool for exploring implicit circRNA-disease associations.

3.6. Hyperparameter analysis

In addition to the ablation experiments, we visualized the effects of varying β, α, and ζ values (ranging from 0 to 1) on AUC using Datasets 1,

Table 6
Top 10 ranked associations for two cancers.

circRNAs-Lung cancer	Evidence	circRNAs-Bladder cancer	Evidence
hsa_circ_0067934	**PMID:** 33832139	ciRS-7	**PMID:** 36426933
ciRS-7	**PMID:** 33390857	circFUT8	**PMID:** 35296022
hsa_circ_0018289	Unconfirmed	circIPO11	Unconfirmed
circZKSCAN1	**PMID:** 32010565	circAmotl1	Unconfirmed
hsa_circRNA_100338	Unconfirmed	circR-284	Unconfirmed
circRNA_000839	Unconfirmed	hsa_circ_0018289	Unconfirmed
hsa_circRNA_104566	Unconfirmed	circ-ITCH	**PMID:** 29386015
hsa_circ_0001649	**PMID:** 31599076	circRNA_100290	Unconfirmed
circMTO1	**PMID:** 30975029	circHIAT1	Unconfirmed
hsa_circRNA_104075	Unconfirmed	circFoxo3	**PMID:** 32612392

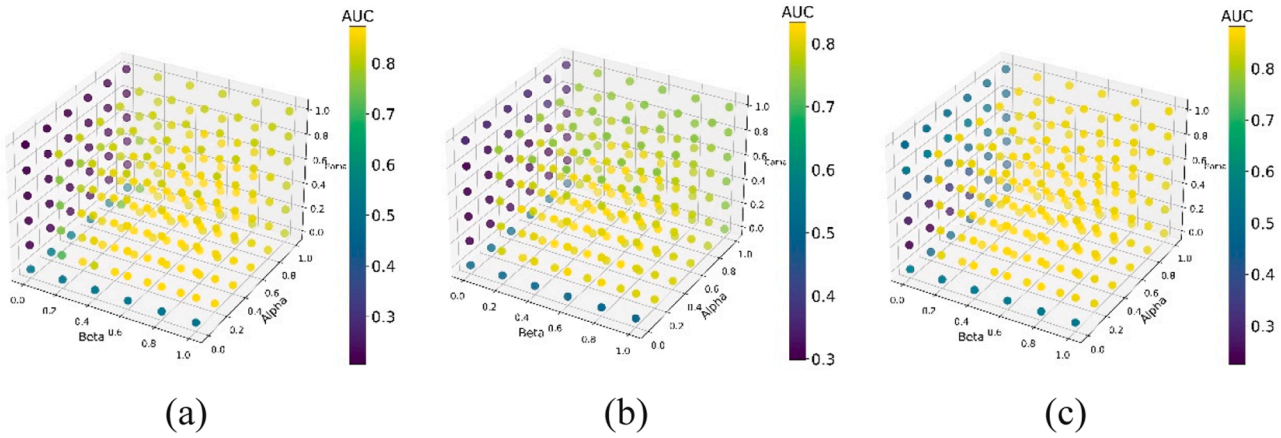


Fig. 5. The 3D heat map visualization of hyperparameter effects on model AUC. (a): Dataset 1. (b): Dataset 2. (c): Dataset3.

2, and 3 as examples. Fig. 5 illustrates that the model exhibits a lower AUC when β is small, indicating that solely focusing on sparsity control hinders the model's ability to find the optimal parameter space. Furthermore, Bi-SGTAR performs better when the α value is centered, emphasizing the effectiveness of bi-view fusion. Optimal performance is usually achieved when $\beta = 0.4$, $\alpha = 0.5$, and $\zeta = 0.05$.

4. Conclusions

In recent years, extensive research efforts have been directed towards predicting circRNA-disease associations, often emphasizing complex models that integrate multiple data sources and employ advanced neural architectures. In contrast, our work focuses on a different perspective, exploring whether simple and traditional ML techniques can achieve equivalent predictive performance. Instead of the complex GNNs and advanced neural techniques (e.g., attention), the presented Bi-SGTAR model is built upon dropout, a fully connected neural network, and L1 regularization—the common ML techniques. Comparative experiments with 25 state-of-the-art models underscored the effectiveness of Bi-SGTAR, suggesting that models amalgamating data from diverse sources should prioritize the quality and complementarity of each data source rather than redundancy to achieve superior performance. A noteworthy aspect of Bi-SGTAR is its minimal data requirement, relying solely on a basic data-adjacency matrix. This characteristic is valuable for researchers working with constrained computational resources and data. Consequently, we posit that the design philosophy behind Bi-SGTAR offers a fresh perspective for advancing the prediction of circRNA-disease associations, which can positively impact future research.

Given the low data requirements of the model, there is promising potential for its application in various fields related to non-coding RNA (ncRNA)-disease or drug association identification. While our initial investigations on lncRNA-disease and microbe-drug datasets have been encouraging, it is crucial to recognize that the current success of Bi-SGTAR in these specific domains does not necessarily imply its seamless efficiency in diverse association prediction tasks across multiple domains. Currently, Bi-SGTAR operates with bi-views, limiting its scope. As we encounter associations from multiple heterogeneous sources, expanding the number of views and merging them may enhance the accuracy of potential associations. Future endeavors may involve extending the model to form a biomolecular-disease association prediction framework, catering to a broader spectrum of association prediction domains. Additionally, our focus will be directed towards designing a trustworthy model that can offer more reliable screening results for subsequent biological experiments. This emphasis on trustworthiness is critical in ensuring the robustness and practical applicability of the model across various research contexts.

CRediT authorship contribution statement

Shiyuan Li: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Qingfeng Chen:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Zhixian Liu:** Writing – review & editing, Validation, Formal analysis, Data curation. **Shirui Pan:** Writing – original draft, Methodology, Investigation, Data curation. **Shichao Zhang:** Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The work reported in this paper was partially supported by a National Natural Science Foundation of China project 61963004 and a Specific Research Project of Guangxi for Research Bases and Talents 2023AC11022.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.knosys.2024.111622.

References

[1] M. Danan, S. Schwartz, S. Edelheit, R. Sorek, Transcriptome-wide discovery of circular RNAs in Archaea, Nucleic Acids Res. 40 (7) (2012) 3131–3142.
[2] W.R. Jeck, N.E. Sharpless, Detecting and characterizing circular RNAs, Nat. Biotechnol. 32 (5) (2014) 453–461.
[3] A.C. Panda, Circular RNAs act as miRNA sponges. Circular RNAs: Biogenesis and Functions, 2018, pp. 67–79.
[4] X. Xu, et al., CircRNA inhibits DNA damage repair by interacting with host gene, Mol. Cancer 19 (2020) 1–19.
[5] H. Zhang, L. Jiang, D. Sun, J. Hou, Z. Ji, CircRNA: a novel type of biomarker for cancer, Breast Cancer 25 (2018) 1–7.
[6] K. Abdelmohsen, et al., Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1, RNA Biol. 14 (3) (2017) 361–369.
[7] Q. Shang, Z. Yang, R. Jia, S. Ge, The novel roles of circRNAs in human cancer, Mol. Cancer 18 (1) (2019) 1–10.

[8] R. Li, J. Jiang, H. Shi, H. Qian, X. Zhang, W. Xu, CircRNA: a rising star in gastric cancer, Cellul. Mol. Life Sci. 77 (2020) 1661–1680.
[9] Y. Ma, Y. Liu, Z. Jiang, CircRNAs: a new perspective of biomarkers in the nervous system, Biomed. Pharmacother. 128 (2020) 110251.
[10] L. Xie, M. Mao, K. Xiong, B. Jiang, Circular RNAs: a novel player in development and disease of the central nervous system, Front. Cell Neurosci. 11 (2017) 354.
[11] K. Deepthi, A.S. Jereesh, Inferring potential CircRNA–disease associations via deep autoencoder-based classification, Mol. Diagn. Ther. 25 (2021) 87–97.
[12] Y. Zhao, et al., NONCODE 2016: an informative and valuable data source of long non-coding RNAs, Nucleic Acids Res. 44 (D1) (2016) D203–D208.
[13] Z. Huang, et al., HMDD v3. 0: a database for experimentally supported human microRNA–disease associations, Nucleic Acids Res. 47 (D1) (2019) D1013–D1017.
[14] L.L. Zheng, et al., deepBase v2. 0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data, Nucleic Acids Res. 44 (D1) (2016) D196–D202.
[15] L. Chen, et al., The bioinformatics toolbox for circRNA discovery and analysis, Brief Bioinf. 22 (2) (2021) 1706–1728.
[16] X. Chen, J. Yin, J. Qu, L. Huang, MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction, PLoS Comput. Biol. 14 (8) (2018) e1006418.
[17] C. Fan, X. Lei, F.X. Wu, Prediction of CircRNA-disease associations using KATZ model based on heterogeneous networks, Int. J. Biol. Sci. 14 (14) (2018) 1950.
[18] Q. Zhao, Y. Yang, G. Ren, E. Ge, C. Fan, Integrating bipartite network projection and KATZ measure to identify novel CircRNA-disease associations, IEEE Trans. Nanobiosci. 18 (4) (2019) 578–584.
[19] W. Zhang, C. Yu, X. Wang, F. Liu, Predicting CircRNA-disease associations through linear neighborhood label propagation method, IEEE Access 7 (2019) 83474–83483.
[20] M. Li, M. Liu, Y. Bin, J. Xia, Prediction of circRNA-disease associations based on inductive matrix completion, BMC. Med. Genomics 13 (2020) 1–13.
[21] H. Wei, B. Liu, iCircDA-MF: identification of circRNA-disease associations based on matrix factorization, Brief Bioinf. 21 (4) (2020) 1356–1367.
[22] L. Peng, C. Yang, L. Huang, X. Chen, X. Fu, W. Liu, RNMFLP: predicting circRNA–disease associations based on robust nonnegative matrix factorization and label propagation, Brief Bioinf. 23 (5) (2022) bbac155.
[23] H. Chang, et al., Integrating multiple microarray dataset analysis and machine learning methods to reveal the key genes and regulatory mechanisms underlying human intervertebral disc degeneration, PeerJ 8 (2020) e10120.
[24] X. Lei, C. Bian, Integrating random walk with restart and k-Nearest Neighbor to identify novel circRNA-disease association, Sci. Rep. 10 (1) (2020) 1–9.
[25] L. Wang, Z.H. You, Y.A. Huang, D.S. Huang, K.C. Chan, An efficient approach based on multi-sources information to predict circRNA–disease associations using deep convolutional neural network, Bioinformatics 36 (13) (2020) 4038–4046.
[26] K. Deepthi, A.S. Jereesh, An ensemble approach for CircRNA-disease association prediction based on autoencoder and deep neural network, Gene 762 (2020) 145040.
[27] L. Wang, Z.H. You, Y.M. Li, K. Zheng, Y.A. Huang, GCNCDA: a new method for predicting circRNA-disease associations based on graph convolutional network algorithm, PLoS Comput. Biol. 16 (5) (2020) e1007568.
[28] W. Lan, et al., KGANCDA: predicting circRNA-disease associations based on knowledge graph attention network, Brief Bioinf. 23 (1) (2022) bbab494.
[29] G. Simmel, The Sociology of Georg Simmel, Vol. 92892, Simon and Schuster, 1950.
[30] S. Milgram, The small world problem, Psychol. Today 2 (1) (1967) 60–67.
[31] W. Lan, et al., Benchmarking of computational methods for predicting circRNA-disease associations, Brief Bioinf. 24 (1) (2023) bbac613.
[32] C. Corbière, N. Thome, A. Bar-Hen, M. Cord, P. Pérez, Addressing failure prediction by learning model confidence, Adv. Neural Inf. Process. Syst. 32 (2019).
[33] Z. Han, F. Yang, J. Huang, C. Zhang, J. Yao, Multimodal dynamics: dynamical fusion for trustworthy multimodal classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 20707–20717.
[34] S. Ghosal, S. Das, R. Sen, P. Basak, J. Chakrabarti, Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits, Front. Genet. 4 (2013) 283.
[35] C. Fan, X. Lei, Z. Fang, Q. Jiang, F.X. Wu, CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases, Database 2018 (2018).
[36] W. Lan, et al., CircR2Cancer: a manually curated database of associations between circRNAs and cancers, Database 2020 (2020).
[37] Z. Zhao, et al., circRNA disease: a manually curated database of experimentally supported circRNA-disease associations, Cell Death. Dis. 9 (5) (2018) 475.
[38] M. Rophina, D. Sharma, M. Poojary, V. Scaria, Circad: a comprehensive manually curated resource of circular RNA associated with diseases, Database 2020 (2020).
[39] G. Chen, et al., LncRNADisease: a database for long-non-coding RNA-associated diseases, Nucleic Acids Res. 41 (D1) (2012) D983–D986.
[40] S. Ning, et al., Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers, Nucleic Acids Res. 44 (D1) (2016) D980–D985.
[41] Z. Lu, K. Bretonnel Cohen, L. Hunter, GeneRIF quality assurance as summary revision, in: Biocomputing 2007, World Scientific, 2007, pp. 269–280.
[42] J. Wu, Z. Ning, Y. Ding, Y. Wang, Q. Peng, L. Fu, KGETCDA: an efficient representation learning framework based on knowledge graph encoder from transformer for predicting circRNA-disease associations, bioRxiv. (2023) 2023–3003.
[43] E. Ge, Y. Yang, M. Gang, C. Fan, Q. Zhao, Predicting human disease-associated circRNAs based on locality-constrained linear coding, Genomics 112 (2) (2020) 1335–1342.
[44] H. Vural, M. Kaya, R. Alhajj, A model based on random walk with restart to predict circRNA-disease associations on heterogeneous network, in: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019, pp. 929–932.
[45] M. Niu, Q. Zou, C. Wang, GMNN2CD: identification of circRNA–disease associations based on variational inference and graph Markov neural networks, Bioinformatics 38 (8) (2022) 2246–2253.
[46] W. Lan, et al., IGNSCDA: predicting CircRNA-disease associations based on improved graph convolutional network and negative sampling, IEEE/ACM. Trans. Comput. Biol. Bioinform. 19 (6) (2021) 3530–3538.
[47] C. Lu, M. Zeng, F. Zhang, F.X. Wu, M. Li, J. Wang, Deep matrix factorization improves prediction of human circRNA-disease associations, IEEE J. Biomed. Health Inform. 25 (3) (2020) 891–899.
[48] X. Wu, W. Lan, Q. Chen, Y. Dong, J. Liu, W. Peng, Inferring LncRNA-disease associations based on graph autoencoder matrix completion, Comput. Biol. Chem. 87 (2020) 107282.
[49] C. Lu, et al., Prediction of lncRNA–disease associations based on inductive matrix completion, Bioinformatics 34 (19) (2018) 3357–3364.
[50] L. Ding, M. Wang, D. Sun, A. Li, TPGLDA: novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph, Sci. Rep. 8 (1) (2018) 1065.
[51] G. Xie, T. Meng, Y. Luo, Z. Liu, SKF-LDA: similarity kernel fusion for predicting lncRNA-disease association, Mol. Therapy-Nucleic Acids 18 (2019) 45–55.
[52] W. Lan, X. Wu, Q. Chen, W. Peng, J. Wang, Y.P. Chen, GANLDA: graph attention network for lncRNA-disease associations prediction, Neurocomputing 469 (2022) 384–393.
[53] Z. Shi, H. Zhang, C. Jin, X. Quan, Y. Yin, A representation learning model based on variational inference and graph autoencoder for predicting lncRNA-disease associations, BMC Bioinf. 22 (1) (2021) 1–20.
[54] L. Zhu, G. Duan, C. Yan, J. Wang, Prediction of microbe-drug associations based on Katz measure, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019, pp. 183–187.
[55] Y. Fan, M. Chen, Q. Zhu, W. Wang, Inferring disease-associated microbes based on multi-data integration and network consistency projection, Front. Bioeng. Biotechnol. 8 (2020) 831.
[56] J. Luo, Y. Long, NTSHMDA: prediction of human microbe-disease association based on random walk by integrating network topological similarity, IEEE/ACM. Trans. Comput. Biol. Bioinform. 17 (4) (2018) 1341–1351.
[57] Z. Yu, F. Huang, X. Zhao, W. Xiao, W. Zhang, Predicting drug–disease associations through layer attention graph convolutional network, Brief Bioinf. 22 (4) (2021) bbaa243.
[58] H. Li, et al., Identifying microbe-disease association based on a novel back-propagation neural network model, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (6) (2020) 2502–2513.
[59] Y. Tan, et al., GSAMDA: a computational model for predicting potential microbe–drug associations based on graph attention network and sparse autoencoder, BMC Bioinf. 23 (1) (2022) 492.