

RESEARCH

Open Access



# XGBCDA: a multiple heterogeneous networks-based method for predicting circRNA-disease associations

Siyuan Shen<sup>1</sup>, Junyi Liu<sup>2</sup>, Cheng Zhou<sup>2</sup>, Yurong Qian<sup>1\*</sup> and Lei Deng<sup>1,2\*</sup>

From 17th RECOMB Satellite Conference on Comparative Genomics Montpellier, France. 1-4 October 2019

## Abstract

**Background:** Biological experiments have demonstrated that circRNA plays an essential role in various biological processes and human diseases. However, it is time-consuming and costly to merely conduct biological experiments to detect the association between circRNA and diseases. Accordingly, developing an efficient computational model to predict circRNA-disease associations is urgent.

**Methods:** In this research, we propose a multiple heterogeneous networks-based method, named XGBCDA, to predict circRNA-disease associations. The method first extracts original features, namely statistical features and graph theory features, from integrated circRNA similarity network, disease similarity network and circRNA-disease association network, and then sends these original features to the XGBoost classifier for training latent features. The method utilizes the tree learned by the XGBoost model, the index of leaf that instance finally falls into, and the 1 of K coding to represent the latent features. Finally, the method combines the latent features from the XGBoost with the original features to train the final model for predicting the association between the circRNA and diseases.

**Results:** The tenfold cross-validation results of the XGBCDA method illustrate that the area under the ROC curve reaches 0.9860. In addition, the method presents a striking performance in the case studies of colorectal cancer, gastric cancer and cervical cancer.

**Conclusion:** With fabulous performance in predicting potential circRNA-disease associations, the XGBCDA method has the promising ability to assist biomedical researchers in terms of circRNA-disease association prediction.

**Keywords:** Association prediction, circRNA, XGBoost

## Introduction

CircRNA is a covalently closed loop structure [1], and its downstream 5' splice site is connected to the upstream 3' splice site [2]. In recent decades, the researches regarding

circRNA have entered into a stage of rapid development. Emerging evidence indicates that plenty of circRNAs are related to critical biological processes. Among these processes, one of the significant aspects is the associations between circRNA and diseases, with the gradually increasing numbers of circRNA-disease associations verified by biological experiments. Jelenia et al. discovered that circRNA plays a paramount role in the evolution of cancer. Specifically, their study manifested that

\*Correspondence: qyr@xju.edu.cn; leideng@csu.edu.cn

<sup>1</sup> School of Software, Xinjiang University, Wulumuqi 830091, China

<sup>2</sup> School of Computer Science and Engineering, Central South University, Changsha 410075, China



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

cancer-related chromosomal translocations cause fusion circRNA (f-circRNA), and F-circRNAs show tumor-promoting effects in vivo models [3]. Wang et al. conducted a study showing that heart-related circRNA (HRCR) is an antihypertrophic molecule that can inhibit cardiac hypertrophy and heart failure by targeting miR-233 and ARC [4]. Liu et al. detected a new circRNA involved in the process of cartilage damage, and further proposed that circRNA-CER may be used as a potential target for osteoarthritis OA [5]. Moreover, circRNA also has a close relationship with bladder cancer, colorectal adenocarcinoma, esophageal squamous cell carcinoma, lung adenocarcinoma and other cancers [6–9]. Although circRNA has become a marker for the diagnosis of specific diseases, traditional experiments cost substantial time and resources. Thus, a fast and economical method to detect the connection between circRNA and human diseases is of great significance.

To start the analysis of the association between circRNA and diseases, it is necessary to establish a circRNA database first. Currently, multiple databases storing circRNA information have been constructed. The circBase database collects information such as the sequence, gene and genome location of circRNA and its latest update was in July 2017 [10]. The Circ2Traits database is the first disease-circRNA association database [11]. The CircNet database accumulates expression profiles, genome annotations and sequences of circRNA subtypes, and provides circRNA-miRNA gene regulatory networks [12]. The CircR2Disease gathers experimentally verified circRNA-disease associations and contains 725 associations between 661 circRNAs and 100 diseases in its latest version [13]. The CircInteractome database includes a search function for possible interactions between circRNA and RBP and miRNA [14]. The exoRBase database visualizes the collection of circRNA, lncRNA and mRNA derived from the analysis of human blood exosomal RNA-seq data [15]. The CSCD database developed by Xia et al. is designed to study the function of cancer-specific circRNA [16].

There are many methods proposed to predict circRNA-disease associations. For example, Deng et al. predicted circRNA-disease associations based on the KATZ method and the integration between circRNA, protein and disease [17]. Lu et al. proposed a method for predicting circRNA-disease associations based on sequence and ontology representations of convolutional neural networks and recurrent neural networks [18]. Li et al. used a deep learning method called DeepWalk to extract features, and then used a network consistent projection method for circRNA-disease association prediction [19]. Wang et al. used stacked autoencoders to extract features, and carousel forest (RF) classifiers for circRNA-disease association prediction [20]. Zheng et al. proposed

the iCDA-CGR model to predicate circRNA-disease associations based on chaotic game representation [21]. Wang et al. proposed a calculation method based on multi-source information combined with deep convolutional neural network (CNN) to predict circRNA-disease association [22].

In this article, we propose an effective method, named XGBCDA, to predict circRNA-disease associations. Initially, we construct a circRNA similarity matrix composed of circRNA expression profile similarity and Gaussian interaction profile kernel similarity, and a disease similarity matrix composed of disease semantic similarity and Gaussian interaction profile kernel similarity. Besides, we also integrate the circRNA similarity network, the disease similarity network and the known circRNA-disease association network. Then, we utilize the aforementioned data to calculate original features, namely statistical features and graph theory features, and send extracted original features to the XGBoost classifier to obtain latent features. Finally, we input the fused features into the XGBoost classifier again to predict the circRNA-disease association. As a result, our method achieves outstanding performance on the circR2disease dataset, and with the tenfold cross-validation, the area under the curve (AUC) is 0.9860. Figure 1 illustrates the flowchart of our method.

## Methods

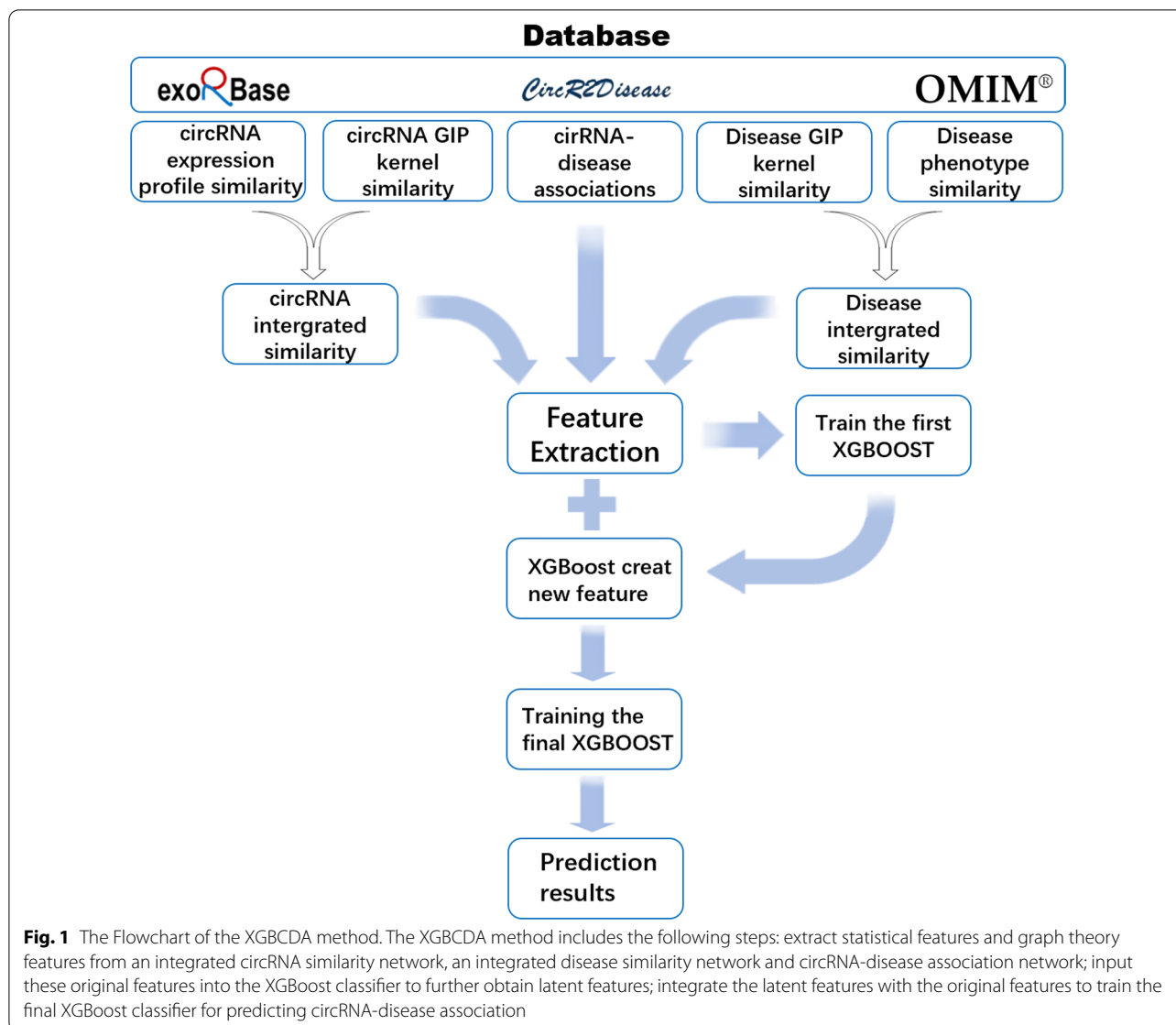
### Human circRNA-disease associations

In this study, we obtain human circRNA-disease associations dataset from the CircR2Disease database, including 660 circRNA-disease associations between 604 circRNAs and 88 diseases. CircR2disease provides experimentally verified circRNA-disease associations, which is of great help to our further research in this field. Here, we use adjacency matrix  $A$  to represent the circRNA-disease association. If a certain circRNA  $c_i$  is related to the disease  $d_j$ , then we assign the element  $A(c_i, d_j)$  to 1, otherwise to 0.

### circRNA similarity

#### circRNA expression profile similarity

We download 49 human circRNA expression profile data from the exoRbase database [13], whose current version contains 58,330 circRNAs. Then we unify the circRNA id in exoRbase with the circRNA id in the aforementioned circR2disease. Next, we use the person correlation coefficient to calculate the similarity of the expression profile between two circRNAs, represented as element  $CS\_EP(X, Y)$ . If the person correlation coefficient of circRNA  $X$  and circRNA  $Y$  is higher than the threshold, the element  $CS\_EP(X, Y)$  is assigned to 1, otherwise 0. In this



method, we assign the threshold to 0.4. The similarity of two circRNA is defined as follows:

$$CS_{EP}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

**circRNA GIP kernel similarity**

Based on the hypothesis that similar diseases may be related to similar circRNAs, we calculate the similarity of the Gaussian interaction profile kernel of circRNAs [23]. The Gaussian kernel function is a scalar function that is symmetric along the radial direction and it is widely used in constructing the kernel with eigenvectors [24]. In 1964, Aizermann et al. applied

this approach to machine learning to study the potential function method [25]. The specific formula is as follows:

$$CS_{GS}(c_i, c_j) = \exp\left(-\gamma_c \left\| y_{c_i} - y_{c_j} \right\|^2\right) \quad (2)$$

The parameter  $\gamma_c$  has impact on adjusting the calculated kernel bandwidth. Here we define the value of  $\gamma_c$  as follows:

$$\gamma_c = \frac{\gamma'_c}{\left(\frac{1}{n_c} \sum_{i=1}^{n_c} |y_{c_i}|^2\right)} \quad (3)$$

where  $n_c$  represents the number of all circRNAs.

### circRNA similarity integration

Finally, we integrate the obtained circRNA expression profile similarity with the circRNA Gaussian interaction profile kernel similarity, using the following formula:

$$CS(i, j) = \begin{cases} CS_{EP}(c_i, c_j), & \text{if } CS_{EP}(c_i, c_j) \neq 0 \\ CS_{GS}(c_i, c_j), & \text{otherwise} \end{cases} \quad (4)$$

### Disease similarity

#### Disease functional similarity

We gather the phenotypic similarity moment data of diseases from Zhang et al. [17]. And we extract the diseases names from the circRNA-disease association in the circR2disease database and employ them to search for the most similar phenotype ID for each disease within the OMIM database. For the sake of ensuring the accuracy of the data, we delete the diseases that do not match the disease phenotype ID in the OMIM database. Eventually, we collect the qualified phenotypic similarity data of the diseases.

#### Disease GIP Kernel similarity

The computational process of disease GIP kernel similarity is analogous to that of disease Gaussian interaction profile kernel similarity. Based on the hypothesis that similar diseases may constantly be related to similar circRNAs [23], we calculate the kernel similarity of the Gaussian interaction profile kernel of a certain disease by following formula:

$$DS_{GS}(d_i, d_j) = \exp\left(-\gamma_d \left\| \gamma_{d_i} - \gamma_{d_j} \right\|^2\right) \quad (5)$$

The parameter  $\gamma_d$  limits the bandwidth. Here we define the value of  $\gamma_d$  as follows:

$$\gamma_d = \frac{\gamma'_d}{\left(\frac{1}{n_d} \sum_{i=1}^{n_d} |\gamma_{d_i}|^2\right)} \quad (6)$$

where  $n_d$  represents the number of all diseases.

#### Disease similarity integration

We utilize a similar way, as depicted in the integration of circRNA similarity, to integrate the obtained disease semantic similarity with the disease Gaussian interaction profile kernel similarity by the following formula:

$$DS(i, j) = \begin{cases} DS_{SS}(d_i, d_j), & \text{if } DS_{SS}(d_i, d_j) \neq 0 \\ DS_{GS}(d_i, d_j), & \text{otherwise} \end{cases} \quad (7)$$

### XGBCDA method

In the XGBCDA method, we construct three matrices, the integrated circRNA similarity matrix CS, the

integrated disease similarity matrix DS, and the circRNA-disease association matrix A. Inspired by Tong He et al.'s research [26], we calculate the statistical characteristics of each circRNA/disease similarity score, including the histogram distribution and the mean of similarity scores, according to the circRNA similarity matrix CS and the disease similarity matrix DS respectively. Besides, we construct a network whose nodes are circRNA/disease, according to the circRNA/disease similarity matrix. In the network, if the similarity score between two nodes is higher than the average similarity score, then there is an edge between two nodes. We also calculate the number of neighbors that each node has, and nodes' graph theory characteristics, namely degree centrality, closeness centrality, betweenness centrality. Then, we select the 10 nodes closest to the node's similarity score as neighbors, and calculate the average and histogram distribution of their similarity scores. In addition, we design a network whose nodes are circRNA and disease, according to the circRNA-disease association matrix A, and use the NMF (Non-Negative Matrix Factorization) algorithm to calculate the latent vector. We then combine the above features to construct a composite feature vector to train the XGBoost model. Subsequently, we use the tree learned by the XGBoost model to form new features. Finally, these new features accompanied with the original features are added to the model for training. After finishing all the procedures, we put the trained XGBoost model into predicting potential circRNA-disease associations. The complete process is illustrated in Fig. 2.

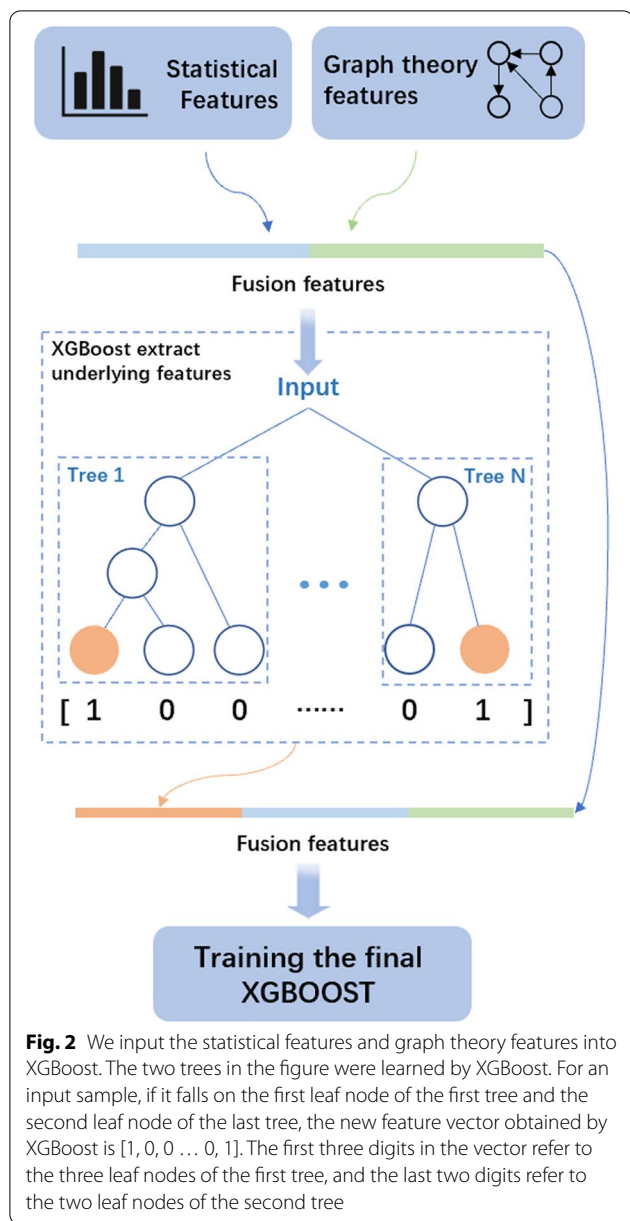
## Results

### Performance evaluation

In order to comprehensively assess the prediction performance of our method, we implement the method on the CIRC2Disease dataset by fivefold cross-validation. Our data set contains positive samples, namely all 660 pairs of known circRNA-disease associations, and negative samples, namely the same amount of unknown associations. Based on the fivefold cross-validation, the area under the curve (AUC) of our method is 0.9935, 0.9913, 0.9996, 0.9968 and 0.9660 respectively, and the average AUC is 0.9861. The experimental results are summarized in Fig. 3.

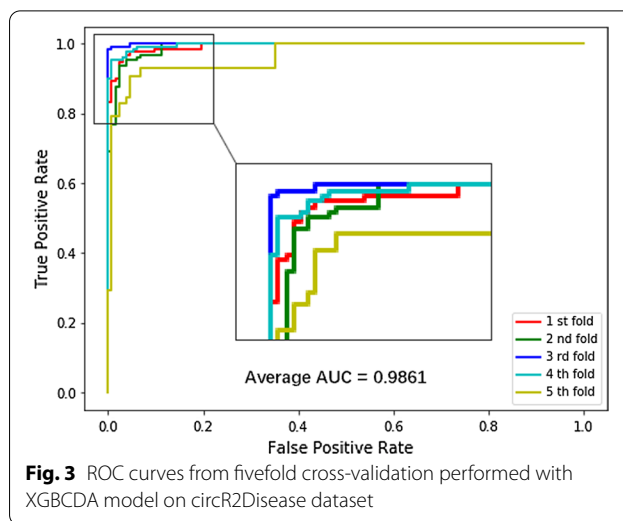
### Comparison with different classifiers

To verify the XGBoost classifier's performance in the model, we compared it with other four popular classifier models (SVM, Decision Tree, KNN, Naive Bayes). These five classifiers all share the same data set, and to ensure the validity of the comparison, we use the default parameters for training and prediction. The evaluation



**Fig. 2** We input the statistical features and graph theory features into XGBoost. The two trees in the figure were learned by XGBoost. For an input sample, if it falls on the first leaf node of the first tree and the second leaf node of the last tree, the new feature vector obtained by XGBoost is [1, 0, 0 ... 0, 1]. The first three digits in the vector refer to the three leaf nodes of the first tree, and the last two digits refer to the two leaf nodes of the second tree

criteria includes accuracy(ACC), Area under the ROC curve(AUC), precision(PRE), recall(REC). With tenfold cross-validation, all parameters of the XGBoost model are ahead of other classifier models, and the verification results of the remaining four classifier models were shown in Table 1. For an apparent comparison, we present the results of these five models in the form of the histogram. From Fig. 4, it is evident that the XGBoost exhibits the first-rate competence in the evaluation. The comparative experiment results fully prove that the XGBoost classifier is superior to other classifier models in every aspect.



**Fig. 3** ROC curves from fivefold cross-validation performed with XGBCDA model on circR2Disease dataset

**Table 1** Compare with other classifier models in tenfold cross-validation on the same dataset

Method	ACC	AUC	PRE	REC	F1
XGBoost	0.9541	0.9860	0.9844	0.9307	0.9517
SVM	0.7626	0.8696	0.8995	0.5519	0.7217
Decision Tree	0.8242	0.9146	0.8266	0.8124	0.8138
KNN	0.8618	0.9147	0.8885	0.8247	0.8509
Naive Bayes	0.7031	0.7469	0.7214	0.6485	0.6766

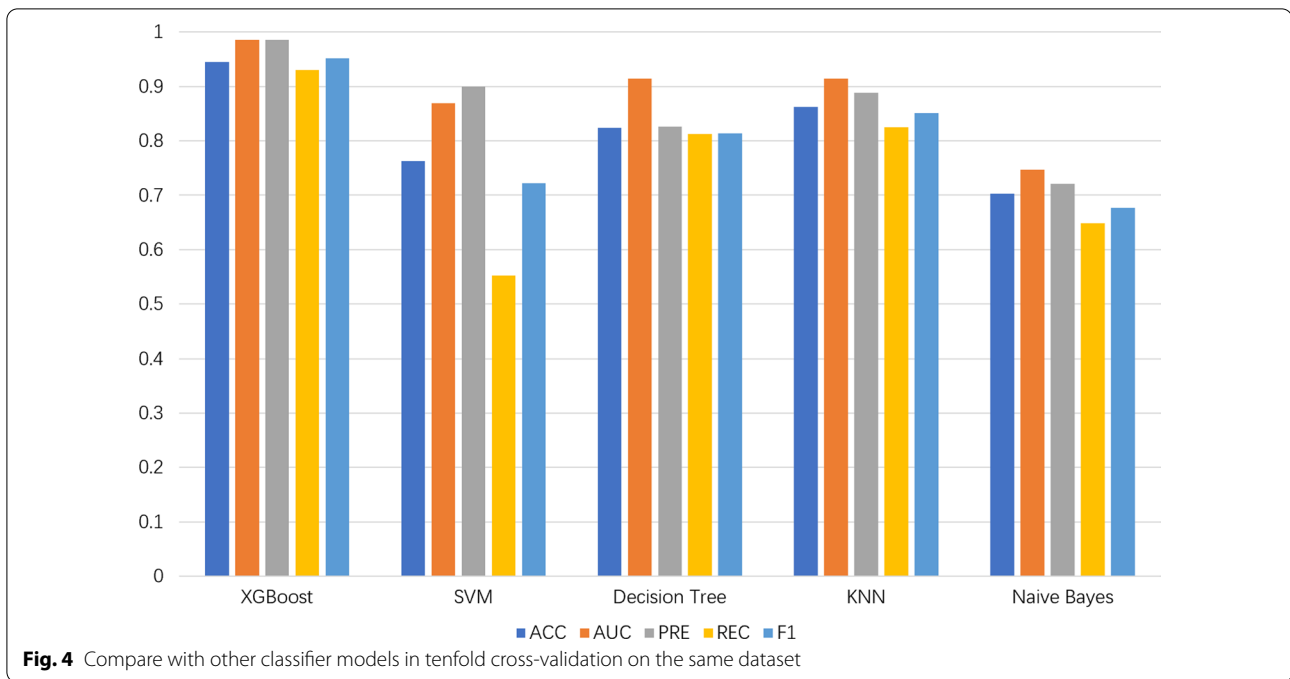
**Selection of optimal parameter values**

In order to further understand the robustness of our proposed method, we analyze the optimal values of 5 parameters in the XGBoost classifier that have the main impact on the performance of tenfold CV, including learning\_rate, n\_estimators, max\_depth, min\_child\_weight and gamma. We use the cv function in the python package of xgboost to calculate the best values of the learning\_rate and n\_estimators parameters, which are 0.1 and 463, respectively. We apply the grid search method to determine the parameters max depth and min child weight to be 5 and 4, respectively. We try 5 representative values to test the optimal value of gamma, which are 1e-5, 1e-2, 0.1, 1, 100. Table 2 below proves that 1 is the best value of gamma.

**Comparison with other methods**

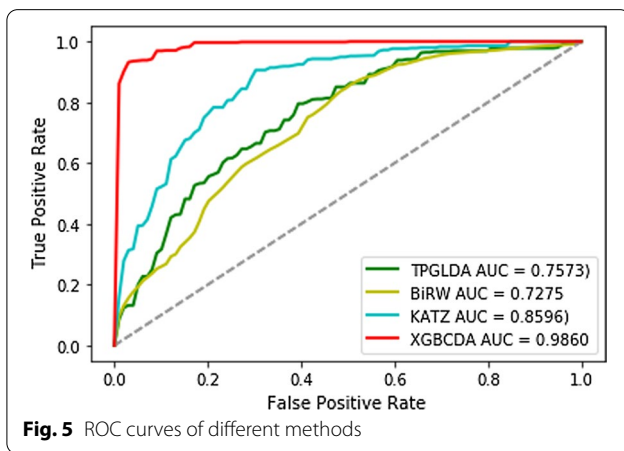
To thoroughly confirm the best performance of the proposed model, we compare XGBCDA with other state-of-art methods. In comparison with LncRDNetFlow [27], TPGLDA and BiRW [28] and KATZ [29], we use all





**Table 2** The tenfold CV prediction performance of various parameter values ranging from 1e−5 to 100 for gamma

gamma	1e−5	1e−2	0.1	1	100
AUC	0.9849	0.9852	0.9860	0.9844	0.9018



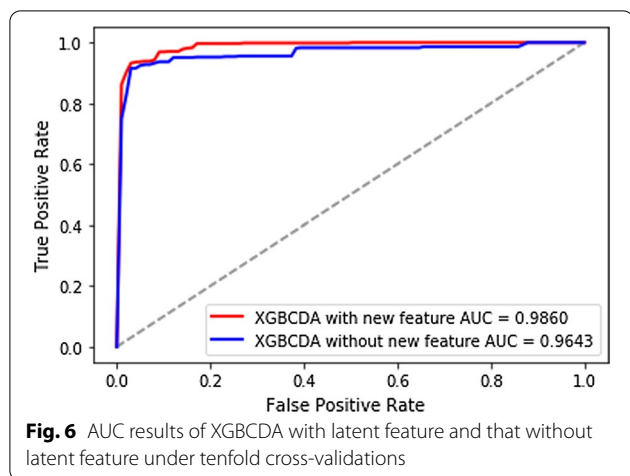
human circRNA-disease associations in the circR2disease database, defined as positive samples, and the same number of unproven circRNA-disease, defined as negative samples, to form the data set. The Fig. 5 presents that under tenfold cross-validation, the performance of our method significantly exceeds that of the other four methods, and the AUC of our method is 0.9860.

**Latent features extracted from XGBoost**

We compare the model that uses XGBoost to generate new features with the model that does not. XGBoost is also known as eXtreme Gradient Boosting package [30], and has applied to handle multiple tasks, such as regression, classification, and sorting. Furthermore, its advantages involve fast training speed and marvelous prediction performance. Given the aforesaid traits and the work of He et al. [31], we used XGBoost to extract latent features based on original features. We consider each tree as a classification feature and use the leaf index that the instance finally falls into as a value. And the ultimate latent features are coded by 1 Of K coding. Figure 6 depicts that based on tenfold cross-validation, the model using the latent features generated by XGBoost has better performance.

**Case studies**

To further evaluate the performance of our method in predicting potential circRNA-disease associations, we select the top 20 associations by prediction scores for verification. The results are presented in Table 3. In addition, we choose three diseases, which are rectal cancer, gastric cancer and cervical cancer, to conduct case studies. We pick 660 known human circRNA-disease associations from circR2Disease as training data. In terms of prediction results, the prediction scores of potential circRNA-disease associations range from 0 to 1, where 1 refers to the highest possibility of the association, and 0 refers to the lowest. In the method, we assume that



**Table 3** Based on the known association predictions in the circR2disease database, the 20 circRNA-disease pairs with the highest scores

Disease	circRNA	Source
Coronary artery disease	has_circRNA6510_1	CircR2Disease
Coronary artery disease	hsa_circRNA11783_2	CircR2Disease
Coronary artery disease	hsa_circRNA11806_28	CircR2Disease
Osteosarcoma	hsa_circ_0092509	CircR2Disease
Hepatocellular carcinoma	circC3P1	CircR2Disease
Osteosarcoma	hsa_circ_0009910	CircR2Disease
Major depressive disorder	hsa_circ_0001410	CircR2Disease
Major depressive disorder	hsa_circ_0001907	CircR2Disease
Major depressive disorder	hsa_circ_0005620	CircR2Disease
Major depressive disorder	hsa_circ_0056048	CircR2Disease
Major depressive disorder	hsa_circ_0005620	CircR2Disease
Bladder cancer	hsa_circ_0041103	CircR2Disease
Bladder cancer	hsa_circ_0007158	CircR2Disease
Bladder cancer	hsa_circ_0082582	CircR2Disease
Bladder cancer	hsa_circ_0072088	CircR2Disease
Hepatocellular carcinoma	circRNA_000839	CircR2Disease
Bladder cancer	hsa_circ_0061265	CircR2Disease
Hepatocellular carcinoma	hsa_circRNA_104135	CircR2Disease
Primary hepatic carcinoma	hsa_circRNA_100571	CircR2Disease
Coronary artery disease	hsa_circRNA5974_1	CircR2Disease
Osteosarcoma	hsa_circ_0056288	CircR2Disease
Hepatocellular carcinoma	hsa_circ_005075	CircR2Disease

circRNA-disease associations with a score higher than 0.9 have a high degree of confidence, and we select all circRNA-disease associations, which are not included in the circR2disease database, with predictive scores higher than 0.9 in the three diseases of rectal cancer, gastric cancer and cervical cancer. Among the obtained ten pairs of associations, three pairs of circRNA-disease associations

have been confirmed in the literature. However, it is worth noting that this does not mean that the other 7 circRNA-disease pairs must not be related. The results are summarized in Table 4.

### Discussion

We suppose that one of the possible approaches to improve the performance is utilizing other biological information as bridge, given the fact that the researches of the direct association between the circRNA and disease are in the infant stage. For instance, with the growing researches of circRNA-miRNA associations and miRNA-diseases associations, it is worth trying to use miRNA as an intermediary to enhance the performance of our method. Moreover, because the circRNA-RBP data increases exponentially, RBP may be another domain for us to explore.

### Conclusion

In this paper, we proposed an effective method to predict circRNA-disease associations by integrating the semantic similarity of diseases, the similarity of circRNA expression profiles, and the Gaussian interaction profile kernel similarity of circRNA and disease, and using XGBoost to construct latent features. Based on the circR2disease data set, we predict ten pairs of unknown circRNA-disease associations, of which three pairs have been confirmed in the literature. Although our method has achieved extraordinary performance, there is scope for improvement in the future. With the continuous development of ncRNA research by researchers, circRNA-disease associations and lncRNA-disease associations have been gradually discovered, and we can use the valuable information to develop circRNA-disease association predictions.

**Table 4** Validation results of circRNA-disease associations, which are not included in circR2disease, with predicted scores of rectal cancer, stomach cancer, and cervical cancer greater than 0.9 points

Disease	circRNA	Score	Source
Colorectal cancer	has_circ_0000504	0.9990	Unknown
	hsa_circ_0001821	0.979671	PMID: 31616472
Gastric cancer	hsa_circ_0001313	0.990974	PMID: 32253030
	hsa_circ_0001141	0.990562	Unknown
Cervical cancer	hsa_circ_0001649	0.965188	Unknown
	hsa_circ_0001313	0.958534	Unknown
	hsa_circ_0001445	0.942731	PMID: 30575898
	hsa_circ_0001946	0.911142	Unknown
	hsa_circ_0001821	0.903885	Unknown
	hsa_circ_0001141	0.903885	Unknown

## Abbreviations

circRNAs: Circular RNAs; GIP: Gaussian interaction profiles; LOOCV: Leave-one-out cross validation; fivefold CV: 5-Fold cross validation; tenfold CV: 10-Fold cross validation; OMIM: Online Mendelian Inheritance in Man; AUC: Area under the curve; XGBGDA: Our proposed computational method; XGBOOST: EXtreme Gradient Boosting.

## Acknowledgements

The authors are very grateful to the anonymous reviewers for their constructive comments which have helped significantly in revising this work. We would like to thank the Experimental Center of School of Computer Science and Engineering of Central South University, for providing computing resources.

## About this supplement

This article has been published as part of BMC Medical Genomics Volume 14 Supplement 3 2021: Selected articles from the 19th Asia Pacific Bioinformatics Conference (APBC 2021): medical genomics. The full contents of the supplement are available at <https://bmcmgenomics.biomedcentral.com/articles/supplements/volume-14-supplement-3>.

## Authors' contributions

LD and YQ conceived the project. L. SS and CZ designed the experiments. SS carried out the experiments. SS and CZ collected the data and analyzed the results. LD, SS and JL wrote and revised the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by National Natural Science Foundation of China under Grants Nos. 61972422 and 61672541. Publication costs are funded by National Natural Science Foundation of China under Grant No. 61972422. The funding body has not played any roles in the design of the study and collection, analysis and interpretation of data in writing the manuscript.

## Availability of data and materials

The experiment-supported circRNA-disease associations were obtained from circR2disease database (<http://bioinfo.snnu.edu.cn/>). The code and datasets are available at <https://github.com/Q1DT/XGBGDA>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 5 July 2021 Accepted: 29 July 2021

Published online: 03 November 2022

## References

- Chen L-L, Yang L. Regulation of circrna biogenesis. *RNA Biol*. 2015;12(4):381–8.
- Li X, Yang L, Chen L-L. The biogenesis, functions, and challenges of circular rnas. *Mol Cell*. 2018;71(3):428–42.
- Guarnerio J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, Lo-Coco F, Tay Y, Beck AH, Pandolfi PP. Oncogenic role of fusion-circrnas derived from cancer-associated chromosomal translocations. *Cell*. 2016;165(2):289–302.
- Wang K, Long B, Liu F, Wang J-X, Liu C-Y, Zhao B, Zhou L-Y, Sun T, Wang M, Yu T, et al. A circular rna protects the heart from pathological hypertrophy and heart failure by targeting mir-223. *Eur Heart J*. 2016;37(33):2602–11.
- Liu Q, Zhang X, Hu X, Dai L, Fu X, Zhang J, Ao Y. Circular rna related to the chondrocyte ecm regulates mmp13 expression by functioning as a mir-136 'sponge' in human cartilage degradation. *Sci Rep*. 2016;6(1):1–11.
- Li Y, Zheng F, Xiao X, Xie F, Tao D, Huang C, Liu D, Wang M, Wang L, Zeng F, et al. Circchipk 3 sponges mir-558 to suppress heparanase expression in bladder cancer cells. *EMBO Rep*. 2017;18(9):1646–59.
- Huang G, Zhu H, Shi Y, Wu W, Cai H, Chen X. cir-itch plays an inhibitory role in colorectal cancer by regulating the wnt/ $\beta$ -catenin pathway. *PLoS ONE*. 2015;10(6):0131225.
- Wan L, Zhang L, Fan K, Cheng Z-X, Sun Q-C, Wang J-J. Circular rna-itch suppresses lung cancer proliferation via inhibiting the wnt/ $\beta$ -catenin pathway. *BioMed Res Int*. 2016;2016:66.
- Li F, Zhang L, Li W, Deng J, Zheng J, An M, Lu J, Zhou Y. Circular rna itch has inhibitory effect on escc by suppressing the wnt/ $\beta$ -catenin pathway. *Oncotarget*. 2015;6(8):6001.
- Glažar P, Papavasileiou P, Rajewsky N. circbase: a database for circular rnas. *RNA*. 2014;20(11):1666–70.
- Ghosal S, Das S, Sen R, Basak P, Chakrabarti J. Circ2traits: a comprehensive database for circular rna potentially associated with disease and traits. *Front Genet*. 2013;4:283.
- Liu Y-C, Li J-R, Sun C-H, Andrews E, Chao R-F, Lin F-M, Weng S-L, Hsu S-D, Huang C-C, Cheng C, et al. Circnet: a database of circular rnas derived from transcriptome sequencing data. *Nucleic Acids Res*. 2016;44(D1):209–15.
- Fan C, Lei X, Fang Z, Jiang Q, Wu F-X. Circr2disease: a manually curated database for experimentally supported circular rnas associated with various diseases. *Database* 2018 (2018)
- Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. Circinteractome: a web tool for exploring circular rnas and their interacting proteins and micrnas. *RNA Biol*. 2016;13(1):34–42.
- Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, Zheng Q, Li Y, Wang P, He X, et al. exorbase: a database of circrna, lncrna and mrna in human blood exosomes. *Nucleic Acids Res*. 2018;46(D1):106–12.
- Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, Jin Y, Gao Y, Xia L, Chang H, et al. Cscd: a database for cancer-specific circular rnas. *Nucleic Acids Res*. 2018;46(D1):925–9.
- Deng L, Zhang W, Shi Y, Tang Y. Fusion of multiple heterogeneous networks for predicting circrna-disease associations. *Sci Rep*. 2019;9(1):1–10.
- Lu C, Zeng M, Wu F-X, Li M, Wang J. Improving circrna-disease association prediction by sequence and ontology representations with convolutional and recurrent neural networks. *Bioinformatics*. 2020;6:66.
- Li G, Luo J, Wang D, Liang C, Xiao Q, Ding P, Chen H. Potential circrna-disease association prediction using deepwalk and network consistency projection. *J Biomed Inform*. 2020;112:103624.
- Wang L, You Z-H, Li J-Q, Huang Y-A. lms-cda: prediction of circrna-disease associations from the integration of multisource similarity information with deep stacked autoencoder model. *IEEE Trans Cybernet*. 2020;6:66.
- Zheng K, You Z-H, Li J-Q, Wang L, Guo Z-H, Huang Y-A. icda-cgr: identification of circrna-disease associations based on chaos game representation. *PLoS Comput Biol*. 2020;16(5):1007872.
- Wang L, You Z-H, Huang Y-A, Huang D-S, Chan KC. An efficient approach based on multi-sources information to predict circrna-disease associations using deep convolutional neural network. *Bioinformatics*. 2020;36(13):4038–46.
- van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*. 2011;27(21):3036–43.
- Aizerman MA. Theoretical foundations of the potential function method in pattern recognition learning. *Autom Remote Control*. 1964;25:821–37.
- Dheeba J, Singh NA, Selvi ST. Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. *J Biomed Inform*. 2014;49:45–52.
- He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. Simboost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *J Cheminform*. 2017;9(1):1–14.
- Ding L, Wang M, Sun D, Li A. Tpglda: Novel prediction of associations between lncrnas and diseases via lncrna-disease-gene tripartite graph. *Sci Rep*. 2018;8(1):1–11.
- Xie M, Hwang T, Kuang R. Prioritizing disease genes by bi-random walk. In: Pacific-Asia conference on knowledge discovery and data mining. Springer; 2012. pp 292–303.
- Qu Y, Zhang H, Liang C, Dong X. Katzmda: prediction of mirna-disease associations based on Katz model. *IEEE Access*. 2017;6:3943–50.



30. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, et al. xgboost: extreme gradient boosting; 2017. R package version 0.6-4; 2015.
31. He X, Pan J, Jin O, Xu T, Liu B, Xu T, Shi Y, Atallah A, Herbrich R, Bowers S, et al. Practical lessons from predicting clicks on ads at facebook. In: Proceedings of the eighth international workshop on data mining for online advertising; 2014. pp. 1–9.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

