

Computational approaches for circRNA-disease association prediction: a review

Mengting NIU^{1,2}, Yaojia CHEN^{3,4}, Chunyu WANG⁵, Quan ZOU^{3,4}, Lei XU (✉)²

1 Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China

2 School of Applied Chemistry and Biological Technology, Shenzhen Polytechnic University, Shenzhen 518055, China

3 Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

4 Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China

5 Faculty of Computing, Harbin Institute of Technology, Harbin 150006, China

© The Author(s) 2024. This article is published with open access at link.springer.com and journal.hep.com.cn

Abstract Circular RNA (circRNA) is a covalently closed RNA molecule formed by back splicing. The role of circRNAs in posttranscriptional gene regulation provides new insights into several types of cancer and neurological diseases. CircRNAs are associated with multiple diseases and are emerging biomarkers in cancer diagnosis and treatment. The associations prediction is one of the current research hotspots in the field of bioinformatics. Although research on circRNAs has made great progress, the traditional biological method of verifying circRNA-disease associations is still a great challenge because it is a difficult task and requires much time. Fortunately, advances in computational methods have made considerable progress in circRNA research. This review comprehensively discussed the functions and databases related to circRNA, and then focused on summarizing the calculation model of related predictions, detailed the mainstream algorithm into 4 categories, and analyzed the advantages and limitations of the 4 categories. This not only helps researchers to have overall understanding of circRNA, but also helps researchers have a detailed understanding of the past algorithms, guide new research directions and research purposes to solve the shortcomings of previous research.

Keywords circular RNA, disease association prediction, machine learning, data mining, deep learning

1 Introduction

In 1976, Sanger et al. first identified a closed circular RNA molecule in a plant virus and named it circRNA [1]. However, because the expression of circRNA molecules is low and circRNAs are rare, they are regarded as abnormal products in the process of gene transcription. And because traditional RNA analysis methods have difficulty detecting such RNA

molecules, scientists detected very few circRNA molecules for a long time. It was not until 2012, when there was a breakthrough in high-throughput sequencing technology, the emergence of biochemical methods for enriching circRNA, and the improvement of calculation methods, that researchers discovered that circRNA molecules are ubiquitous in eukaryotic organisms and have been identified to participate in various biological processes [2–4]. Since then, there has been a new era of noncoding circRNA research.

As a special kind of noncoding RNA, circRNA forms a closed RNA molecule by reverse splicing compared with traditional long noncoding RNA (lncRNA) [5]. The characteristics of circRNA are: its closed circular structure, longer half-life than linear RNA, specific reverse splicing site, tissue expression specificity, disease specificity, etc. [6] In addition to microRNA sponge function, fruitful progress has shown that circRNAs play a critical role in gene regulation, development, and carcinogenesis, and scientists have begun to gain new insights into circRNAs [7]. Its functions can be summarized as follows: (1) Coding function. CircRNA plays significant roles in disease research, and its expression is usually not related to host gene expressions. CircRNA can bind to transcription factors to inhibit gene transcription [8]. (2) Regulating gene transcription function. CircRNAs can interact with RNA binding protein (RBP) to regulate protein function and affect the expression of parental genes. CircRNA can bind to transcription factors to inhibit gene transcription. (3) miRNA molecular sponge function. CircRNA have a great quantity of miRNA binding sites that function as miRNA sponges [9]. By inhibiting miRNA, circRNA indirectly regulates n. circRNA is a steady-state product of mRNA splicing and participates in complex gene expression regulation in a new way, with important noncoding function [10]. By inhibiting miRNA, circRNA indirectly regulates the expression of mRNA and participates in the regulation of various human tumors, which may become a new marker [11].

Received January 11, 2024; accepted April 21, 2024

E-mail: csleixu@szpu.edu.cn

CircRNAs are widely found in animals and plants. Studying the potential association can contribute to comprehending complicated disease mechanisms, discovering therapeutic targets, and further understanding the function of circRNAs in disease pathophysiology. Computational methods are indispensable tools in the study of disease associations, and algorithms can be improved according to researchers' methodological and performance needs [12]. Reviews on predicting the association between circRNA and disease already exist [13–16]. It can be found that each document summarizes the algorithm according to different classification rules. We summarized the current algorithms and refined them into 4 categories based on the properties of the prediction algorithms. In this review, we first introduced the function and characteristics of circRNA. Then, we mainly focus on computing methods for circRNA-disease association studies, and introduce the latest progress of predictive algorithms.

2 The nature of circRNAs: biogenesis and function

CircRNA is highly enriched in eukaryotes and has specific expression in different tissues at different developmental stages. CircRNA can act as an miRNA endogenous adsorbent and play the role of miRNA sponge, affect gene expression at the transcriptional or translational level, regulate gene expression (including parental gene expression), gene transcription, and RNA binding proteins, and provide a protein/peptide translation template [17].

2.1 CircRNAs act as miRNA sponges

Although the function of circRNAs is still unknown to a large extent, previous studies have shown that circRNAs can serve as miRNA sponges [18]. CDR1as is the first initially reported circRNA that functions as a miRNA sponge, and contains more than 70 miRNA-7 binding sites [19]. cSRY is found in mouse testis and contains 16 miRNA binding sites [20]. In *Drosophila*, the splicing factor MBL combines with circMb to inhibit the production of circbl, which can not only regulate selective splicing, but also act as a protein sponge [21]. Recently, circRNAs are discovered to share miRNA reaction elements, which enables them to compete for miRNA-binding sites, play the role of miRNA sponge, isolate or reduce the number of miRNAs available for targeted mRNA, and promote mRNA stability or expression. CircHIPK3 originates from the HIPK3 gene and bind with a variety of miRNAs, such as miR-124, miR-30a, and miR-558 [22]. CircRNAs, as miRNA sponges, play a major role in aging related diseases and can become latent biomarkers. Research on the biological function of circRNAs has shown that they participate in human aging related diseases, such as cancer, cardiovascular and neurodegenerative diseases, and inflammatory respiratory diseases and diabetes [23].

The regulatory mode of the mRNA target is determined by the degree of base-pairing complementarity between the “seed” region at the 5' end of mRNA and miRNA. CircRNA as endogenous miRNA sponges demonstrates a novel mode of action for this kind of ncRNA and suggests an alternative mechanism for regulating miRNA activity [24]. On the one

hand, circRNAs can participate in the proliferation, differentiation, and aging of normal cells. On the other hand, they can play significant role in etiopathogenesis through the large regulatory network of sponges [25]. The interaction between miRNA and circRNA has been verified to be very significant in cases of gastric cancer, colorectal cancer, osteoarthritis and neurodegenerative pathology. CircRNA, as a miRNA sponge, has some characteristics, namely, abundance, stability and tissue-specific expression, that are attractive in clinical research, and new methods of using circRNA to diagnose human diseases in clinical research can be explored. Moreover, as a crucial pathway, the circRNA-miRNA-mRNA axis not only provides new directions for the morbidity, diagnosis and treatment of diseases, but can also be used as a progressive molecular technology to imitate or produce therapeutic drugs [26].

2.2 CircRNAs regulate parental gene expression

Genes usually need transcriptional regulatory elements and molecular mechanisms, which interact with regulatory elements to regulate gene expression patterns. Accumulating studies supports the role of circRNAs in the transcriptional complex-mediated regulation of parental gene transcription.

A 2018 study showed that circITGA7 can repress the transcription factor RREB1 and promote the its host gene ITGA7 [27]. Chia et al. proposed that circ-DAB1 upregulates the expression of an important transcription factor recombination signal binding protein in the NOTCH pathway and activates the transcription of the parental gene DAB1 [28]. Recently, circ-STAT3 was discovered to increase the expression of STAT3 by upregulating the transcription factor Gli2 [29]. The above actions show that circRNAs can adjust the transcription of their parental genes and affect the transcription cycle.

MiRNA is a significant regulatory factor of genetic expression. CircRNAs can enhance target gene expression by inhibiting miRNA. Circ-Sirt1 combines with miR-132/212 and promotes the expression of the host gene SIRT1. Circ-Sirt1 has a beneficial protective effect on the inflammation of VSMCs, which indicates that circ-Sirt1 affects the pathogenesis of vascular diseases and may be a new biomarker for atherosclerosis [30]. cTFRC enhances the expression of its parent gene TFRC by cTFRC-miR-107-TFRC, thus achieving carcinogenesis [31]. CircFBLIM1 may regulate the expression of the parent fund FBLIM1 through miR-346, and play a regulatory role in hepatocellular carcinoma [32]. In addition to the above circRNAs, circ-ENO1 [33], circGFRA1 [34], circAmotl1 [35], circ-VANGL1 [36], and circ-TFF1 [37] have also been discovered to regulate the expression of parent genes of related cancers, thus changing the treatment of human cancers. However, in view of the low expression level of most circRNAs, it is necessary to conduct further research in different types of cells or tissues to determine whether there is sufficient evidence to indicate that circRNAs regulate the expression of parental genes as ceRNAs.

2.3 CircRNAs regulate RNA-binding proteins

Studies demonstrated that circRNAs may adsorb protein factors and regulate RNA-binding proteins. In terms of

interacting with proteins, although the main function of circRNAs is through their activity as miRNA sponges, their secondary important function is through circRNA-protein interactions. Although there are fewer circRNA-RBP interaction sites than on linear mRNAs, there are still studies supporting the interaction of RBPs with circRNAs [38,39]. First, circRNAs can regulate RBPs in various ways. CircRNAs regulate the function of RBPs and play the role of RBP sponges, RBP assembly platforms and supporter proteins. RBP-adsorbed circRNA can regulate and splice as a regulatory factor. CircRNA can act as decoy to keep RBP in specific cells. In addition, the effect of RBP on circRNA is also urgently highlighted. RBPs, as a double-stranded or single-stranded RNA-binding protein, exist throughout the entire life cycle of RNA, including transcription, metabolism, translation and degradation. RBPs participate in the generation of circRNA and affect the whole process of circRNA [40]. RBPs are produced under pathological conditions, and defects in their expression can lead to many diseases or other effects. It has been proven that the interaction of circRNA and RBP has a considerable impact on diseases such as cancer and may be a biomarker of disease [41,42].

2.4 The relationship between circRNAs and diseases

Studies found that circRNAs have the characteristics of broadness, conservation and stability; that they play a role in the occurrence of a variety of human complex diseases such as tumor, cardiovascular diseases, neurological diseases, autoimmune diseases, and genetic diseases; and that they can provide important research ideas and targets for disease [43].

In 2021, Peng et al. found that the circRNA molecule ciRS-7 was highly expressed in tumor tissues of renal cell carcinoma and could be used as a prognostic marker [44]. Chen et al. found that circRNA can act as an oncoagenic stimulus or tumor suppressor in cancer and is enriched and stable in extracellular fluid [45]. siRNA interference with circRNA can restore cells to a normal state. Hansen et al. found that the circRNA molecule ciRS-7 can act on cancer-related proteins by inhibiting and releasing miRNA [46]. Akhter et al. summarized associations between circRNAs and neurodegenerative diseases [47]. For example, in Alzheimer's disease, ciRS-7, a typical circRNA molecule in the human brain, inhibits miRNAs from the brains of Alzheimer's patients-7 expression. Hong et al. found that circCLK3 may act as a molecular target of cervical cancer (CC), and by controlling its expression, it can inhibit or promote the growth and metastasis of endogenous CC [48]. Ashwal et al. found that the high expression of the CDR1as hinders the function of miRNA-7, increases insulin levels, and participates in the signal pathway, which is closely related to diabetes [49]. Ji et al. discovered that circ_001621 was remarkably upregulated in osteosarcoma and could promote the proliferation and migration of osteosarcoma. circ_001621 can become a therapeutic target for advanced osteosarcoma [50]. CircRNA showed abnormal expression in many types of cancers (such as colon cancer, liver cancer and pancreatic cancer), atherosclerosis, vascular disease risk, nervous system diseases, prions, osteoarthritis, and diabetes, indicating that circRNA

may be a new biomarker, providing a new therapeutic target for tumor treatment and a new method for drug research and development.

3 Relational databases

3.1 Databases related to circRNA

Currently, the number of circRNAs identified in the human transcriptome has reached 14,807. To manage multispecies, multifunctional and huge circRNA data, multiple circRNA databases have been developed.

(1) The circBase database summarizes and collects circRNAs from multiple species, including human, mouse, *Caenorhabditis elegans*, *Drosophila melanogaster* and other species, and has collected more than 90,000 circRNA transcripts in total [51]. CircBase assembles and annotates circRNA information in a standardized format, and concordances with the UCSC genome browser and NCBI database.

(2) ExoRBase [52] aggregates protein-encoding circRNA data, including annotations, expression levels, and original tissue data of 58,330 circRNAs in the human blood exocrine body. Data were collected from experimental validation in the published literature. Based on the RNA-seq data, the database integrates and visualizes the RNA expression profile. ExoRBase will help researchers identify the molecular characteristics of blood exosomes and help to discover new biomarkers of exosomes.

(3) CIRCPedia [53] comes from the three major RNA-seq databases, namely, GEO, ENCODE, and EMBL-EBI, which collected circRNA data from more than 180 datasets from 6 different species and identified 262,782 circRNAs. The database also contains conservation analysis of circRNAs and provides computational tools to compute circRNA expression.

(4) CSCD is a cancer-specific circRNA database, that collects 272,152 cancer-specific circRNAs from cancer and normal cell lines, and provides miRNA target sites, RBP binding sites, potential circRNAs in cancer-specific circRNAs and open reading frame [54]. CSCD is the first cancer-specific circRNA database, which can make significant contributions to functional research on cancer-related circRNAs.

(5) CircRic systematically characterized the expression profile of circRNA in 935 cancer cell lines, and analyzed the relationship between circRNA and mRNA, protein and mutation, revealing the specific expression pattern of circRNAs [55]. CircRic also analyzed the key genes involved in the biogenesis of circRNA. As a user-friendly data website, CircRic can help biomedical research.

(6) CCRDB is a cancer circRNA-related database [56]. CCRDB collected 10 samples of experimental data from 5 patients with hepatocellular carcinoma (HCC), and found 11,501 circRNAs. CCRDB can be used in combination with data from exoRNA databases such as circBase. CCRDB is the first circRNA database providing analysis and comparison functions, which effectively shows the relationship between circRNAs and HCC. CCRDB can help researchers explore the causes of circRNAs in disease discovery and study the target genes of HCC.

(7) CircExp sorted out the number on 11 databases,

including 48 expression profile datasets of 18 cancers and 860,751 circRNA expression records [57]. CircExp included 189,193 differentially expressed circRNAs with significant differences between normal and cancer samples. It has been developed and proven to be useful for circRNAs with potential diagnostic and prognostic significance for various types cancer. It provides precalculated expression data of circRNA and its parent genes, as well as data browsing and searching.

(8) CircFunBase records more than 7,000 manually planned functional circRNA entries [58]. CircFunBase also provides a visual circRNA-miRNA interaction network and genomic context information of circRNA. CircFunBase will contribute to research on circRNAs and promote research on circRNAs and their function.

(9) CircInteractome analyzes the miRNA and RBP sites on the connection and connection flanking sequences by searching the public circRNA database [59]. Users can use the circRNA sequence to query the RBP binding site. The combination of joint analysis and computational analysis of large-scale transcriptome data provides a profundity method for clarifying the possible biological role of ribonucleoprotein complexes.

(10) circRNADb [60] is the first RNA database that aggregates codable proteins, collecting approximately 30,000 records.

3.2 Databases related to disease

(1) OMIM is the online human Mendelian genetic database [61], which is an online human gene and genetic disease database built and maintained by the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, USA. The OMIM database contains all known genetic diseases and information on more than 15,000 genes, with more emphasis on the relationship between disease phenotypes and their therapeutic genes.

(2) MalaCards integrates 72 disease databases, which include a general overview of diseases, the relationship between diseases, the pathways involved, and the differences in gene expression [62].

(3) PubMed is a network-based comprehensive biomedical information retrieval system developed by the national center for biotechnology information (NCBI), which provides

resource services such as links to various comprehensive molecular biology databases [63].

3.3 CircRNA regulation-disease relationship databases

In addition to some databases of basic biological information about circRNAs and diseases, researchers have built multiple databases by collecting circRNAs to provide data support for predicting candidate circRNAs potentially associated with diseases through computational methods. The databases associated with circRNA-disease are summarized in Table 1.

(1) CircR2Disease is a manually managed database that collects circRNAs related to various diseases [64]. CircR2Disease collected 725 associations between 661 circRNAs and 100 diseases. Each entry records information about circRNA and a disease, including the circRNA name, coordinates and gene symbols, disease name, expression mode of circRNA, experimental technology, brief description of circRNA disease relationship, year of publication and PubMed ID.

(2) The circRNA disease database manually screens the data from the national biotechnology information center, and the key words to be searched were “circRNA disease” and “circRNA cancer” [65]. The time limit was before November 2017. The circRNA disease database records 354 associations between 330 circRNAs and 48 diseases. This database contains information about circRNAs and diseases, including the circRNA ID and expression mode, disease, detection technology, circRNA biological function description, reference, and annotation information.

(3) Circ2Disease collects 273 associations (237 circRNAs and 54 diseases) [66]. These associations come from the experimental results of 120 studies. This database records function description of the circRNA, circRNA name, disease name and expression mode. With the rapid increase of circRNA, Circ2Disease will significantly enhance the understanding of circRNA functions in disease and provide a useful resource for studying circRNA regulation.

(4) circMine collects 136,871 circRNA (87 diseases and 120 circRNA) [67]. These data were collected from 1,107 samples of 31 body positions. To improve the universality of the database, circMine standardized each dataset and labeled the relevant clinical information. The database includes 13 analysis functions, which allow users to set parameters for

Table 1 Databases for circRNA-disease associations

Database	circRNA	Disease	Association	Description	URL
CircR2Disease	661	100	725	It includes the circRNA and disease name, coordinates and gene symbols, expression mode.	See bioinfo.snnu.edu.cn/CircR2Disease/ website.
circRNA disease	330	48	354	It includes circRNAs and diseases, including the circRNA ID and expression mode, disease, detection technology, circRNA biological function description.	See cgga.org.cn:9091/circRNADisease/ website.
Circ2Disease	237	54	273	It includes circRNA and disease name, expression pattern, experimental method, and brief functional description of the CircRNA-disease relationship.	See bioinformatics.zju.edu.cn/Circ2Disease/circRNAgroup.html website.
circMine	136,871	87	1,107	circMine offers 13 online analytical capabilities to assess the clinical and biological significance of circRNA.	See biomedical-web.com/circmine/ website.
Circad	720	150	1,388	It lists validation methods, experimental validation states, primer information, and uses standardized nomenclature to standardize interpretations of associations.	See clingen.igib.res.in/circad/ website.
CircR2Cancer	1,135	82	1,439	It includes information about cancer in the disease ontology as well as basic biological information about circRNAs.	See biobdlab.cn:8000 website.

analyses so that they can use their own circRNA transcriptome to perform these analyses. Furthermore, circMine developed three additional tools to study circRNA- miRNA interactions.

(5) In addition to being a database of circRNA and diseases, Circad's main innovation is that it provides ready-made primers to use circRNA as a biomarker or perform functional studies [68]. In addition, more PCR primer details are provided in the database, which can be used as a ready reference and used for statistical significance.

(6) CircR2Cancer mainly stores the association data between circRNA and cancer [69]. The data were collected from existing literature and databases, and included 1,439 associations between 1,135 circRNAs and 82 types of cancer. The database provides basic information queried from circBase and cancer information extracted from Disease Ontology.

4 Bioinformatics tools for predicting circRNA-disease associations

Compared with traditional methods of biological experiments plus clinical experiments, the development of effective algorithms and tools can save much time and effort to quickly simulate prognosis and predict associations. This section provides a brief overview of some computational methods and tools for prediction of circRNA-disease associations. Currently, there are many calculation methods. When building a computational model to predict circRNA-disease association tasks, first, integrate circRNA similarity information, disease phenotype or semantic information, and circRNA-disease association information to build an initial similarity network. Then, a classifier is trained to predict the affinity score between disease and circRNA. In this review, we mainly introduce several classical methods, including those based on network links, recommendation systems, matrix completion, machine learning (ML) and deep learning (DL). The classification and presentation of calculation methods are shown in Fig. 1. The literature classification of computational models is shown in Fig. 2. For each category of methods, we select typical methods and list their details, advantages, AUCs (Table 2).

4.1 Prediction method based on biological networks

CircRNAs with similar functions may be related to diseases with similar functions. In line with this hypothesis, researchers have proposed several biological network-based methods to predict disease-associated circRNAs. Network-based prediction usually needs to combine data such as gene expression profiles and logical associations of various information to construct a network. The method is relatively complicated. After the construction is completed, the candidate circRNAs are sorted according to the correlation score between circRNAs and diseases to predict disease-causing genes.

Deng et al. constructed a circRNA similarity network, circRNA-protein interaction network, disease similarity network, and heterogeneous network between these molecules, employed the KATZ to calculate associations and constructed prediction model called KATZCPDA [70].

KATZCPDA passed the leave-one-out validation (LOOV) and ten-fold cross-validation (TFCV) and AUCs are 0.959 and 0.958, respectively. The method not only validates existing circRNA disease associations, but also predicts unknown associations with significantly higher performance than previously developed web-based models.

Qu et al. built a computational model based on graph regularization and mixed norm constraints to predict latent associations [71]. Qu et al. has developed a graph-based multilabel learning intrasilicon method for prediction of potential associations. By making full use of the network characteristics of circRNA and disease, the model maintains the local structural integrity of circRNA and disease. The average AUC of Five-fold cross-validation (FFCV) is 0.893.

Xiao et al. proposed a new adaptive subspace learning method based on network embedding to discover the potential relationship [72]. The research makes full use of different data sources to calculate disease similarity and circRNA similarity and uses a network embedding method to learn low-dimensional node representation and an adaptive subspace learning model. At the same time, the regularization term of the integrated weighted graph is introduced to maintain the local geometric structure of the data space, and L1 and L2 norm constraints are introduced into the model to achieve smoothness and sparsity of the projection matrix. The AUC value of FFCV is 0.926 ± 0.015 .

Based on the biological networks, Lei et al. proposed a novel computational path weighting model and constructed a new computational path weighting method using the depth-first search algorithm, PWCDA, to predict circRNA-disease associations [73]. A heterogeneous network was constructed based on known associations, and an association score was calculated according to the path to determine whether that circRNA disease pair was associated. The AUC for the LOOV was 0.89. This method considers the sparsity of the similarity subnetworks, moreover, only paths within three steps are used to reduce the noise information.

Li et al. identify novel circRNA-disease associated based on network consensus projections [74]. NCPCDA utilizes multiview similarity data to construct circRNA similarity and disease similarity. Then, circRNA and disease space are separately projected on the circRNA and disease interaction network. Finally, combining the above two spatial projection scores, we can obtain a association score matrix. The AUC value of FFCV is 0.884. NCPCDA can fully utilize the topological information of heterogeneous networks, and at the same time, as a nonparametric algorithm, it can simplify the prediction process and improve efficiency. However, the final composite score obtained by this method is obtained by averaging the circRNA and disease spatial projection, which can lead to suboptimal predictions.

By integrating the known circRNA-disease associations, disease similarities and circRNA similarities, Xiao et al. employed a graph-based learning algorithm to prioritize the prediction model to guide cumbersome clinical trials. The author developed an integrated framework iCDA-CMG to predict the potential relationship [75]. The AUC is 0.862.

Shu et al. used different learning representation methods to

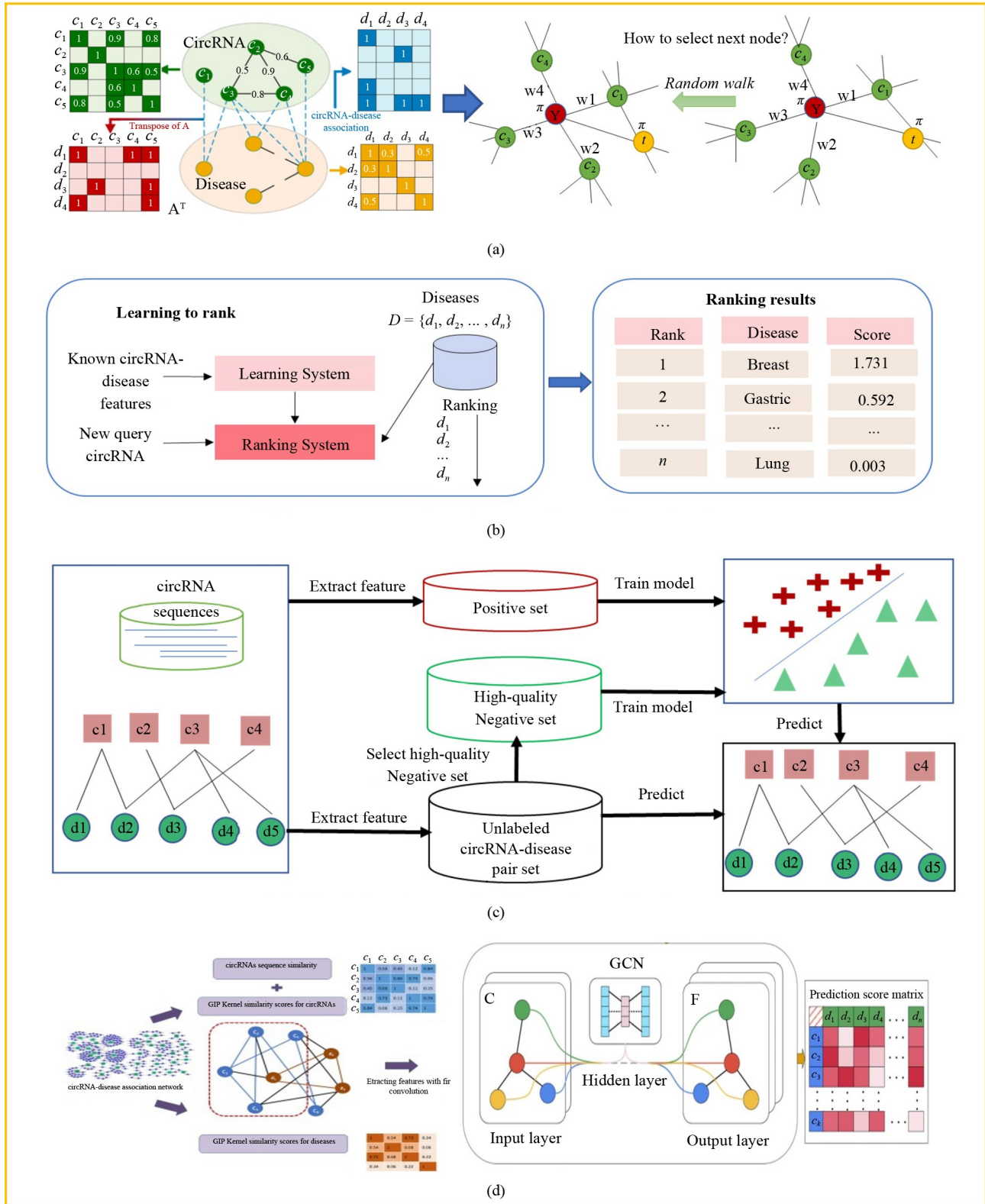


Fig. 1 Classification and representation of computational methods for circRNA-disease prediction and grouping them based on an underlying computational model consisting of biological network, recommendation algorithms, machine learning, and deep learning. (a) Biological network-based methods; (b) recommender system-based methods; (c) machine learning-based methods; (d) deep learning-based methods

construct three similar networks, and then integrated the three networks to obtain the final functional biological features. A functionally similar algorithm MSCFS for calculating circRNA based on a new integrated biological

network [76]. CircRNAs associated with the same miRNA have high similarity. The similarity of circRNA coexpression was positively correlated with the predicted results of the model. The study also concluded that circRNAs with high

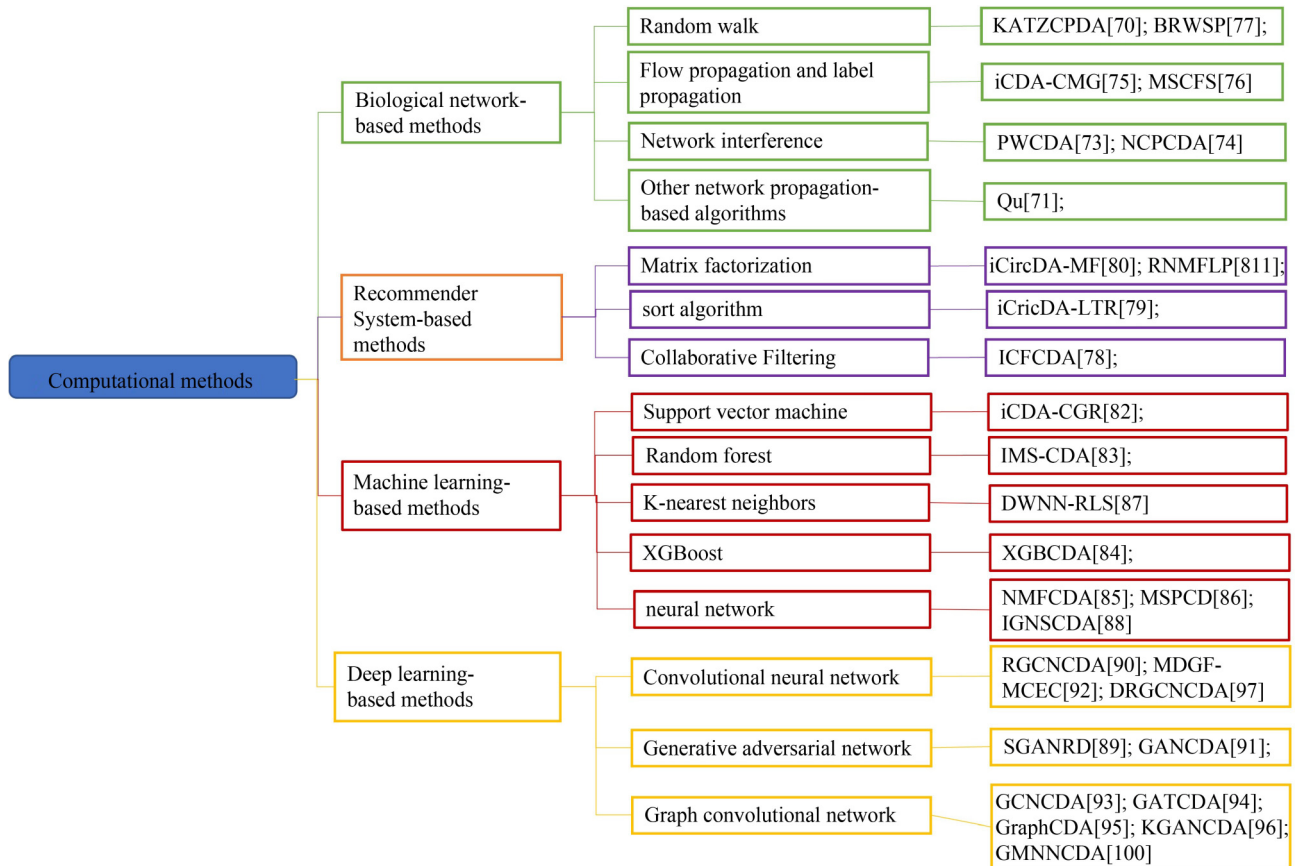


Fig. 2 Computational models for predicting circRNA-disease associations

similarity are also similar in disease association. Research on the functional similarity of circRNAs can promote the research progress of researchers on more potential functions and associations of circRNAs.

Based on the biased random walk (RW) algorithm, Lei et al. searched paths on multiple heterogeneous networks [77]. First, BRWSP built multiple heterogeneous network using circRNAs, diseases and genes. Based on multiple heterogeneous networks, RW is used to search the path between circRNAs and diseases. Subsequently, the score for a particular association is then calculated by using these search paths. Finally, latent associations were recommended according to the score. The AUC of the BRWSP is 0.8675.

Computing model based on a biological network depends on the topology of the network. Incomplete information about network topology may lead to degraded prediction performance. Many data sources are often used to build biological networks. We can make full use of different types of data resources to learn more heterogeneous biological networks to represent the relationship between diseases and circRNAs. However, how to extract the feature representation of each node in the network and use adaptive learning methods to automatically integrate multimodal biometrics is still a computational challenge.

4.2 Prediction method based on recommender system

The recommendation algorithm has achieved good recommendation results in e-commerce, social applications, and news applications. It has been proven that the

recommendation algorithm can obtain the correlation between data. Therefore, researchers have gradually applied recommendation algorithms to the prediction of the associations. Recommendation algorithms used for disease association prediction include: collaborative filtering recommendation algorithm (CFR) and matrix factorization (MF). The basic principle of CFR is to work with the feedback of circRNA to filter massive information and select the information of interest. Although CFR is currently recognized as the most classic recommendation algorithm, the co-occurrence matrix is often very sparse, and the process of finding similar circRNAs is not accurate when there are few known associations. Therefore, CFR was improved and moment MF was proposed to enhance the ability to handle sparse matrices.

Lei et al. present a computational method ICFCDA using a collaborative filtering recommendation algorithm (CFRA), which addressed the “cold start” problem. ICFCDA calculated the characteristics of circRNA-like networks and disease-like networks based on data from multiple databases [78]. ICFCDA is the first to be applied to research on circRNA-disease association prediction. The paper used LOOV and the AUC is 0.946. In addition, to demonstrate the capability of ICFCDA, case studies were carried out.

Wei et al. constructed a prediction algorithm iCricDA-LTR [79]. Different from existing prediction models, iCricDA-LTR adopts ranking algorithm to query the global ranking association. A learned ranking algorithm is employed for

Table 2 Different computational methods to predict circRNA-disease associations

Calss	Model	Data	Description	AUC	Code
Biological network-based methods	KATZCPDA	CircR2Disease	KATZ	0.959, 0.958	NA
	NSL2CD	CircR2Disease	network embedding-based adaptive subspace learning	0.926	NA
	PWCDA	CircR2Disease	computational path weighting model using the depth-first search algorithm	0.89	NA
	NCPCDA	CircR2Disease	based on network consensus projection	0.884	See github.com/ghli16/NNPCD website.
	iCDA-CMG BRWSP	CircR2Disease, CircFunBase CircR2Disease	graph-based learning algorithm random walk	0.862 0.8675	NA NA
Recommender System-based method	iCircDA-MF	CircR2Disease	matrix factorization	0.918	NA
	RNMFLP	CircR2Disease, CircRNADisease, Circ2Disease	matrix factorization	0.9512	See github.com/biohuster/RNMFLP website.
	iCircDA-LTR IFCDA	CircFunBase CircR2Disease	learning to rank algorithm CFR	0.928 0.946	See bliulab.net/iCircDA-LTR/ website. NA
Machine learning-based methods	iCDA-CGR	CircR2Disease, circRNADisease, circFunBase, Circ2Disease	support vector machine	0.8533	See github.com/look0012/iCDA-CGR website.
	IMS-CDA	CircR2Disease	random forest classifier	0.8808	See github.com/look0012/IMS-CDA/ website.
	DWNN-RLS	CircR2Disease	k-nearest neighbors	0.8854, 0.9205, 0.9701	NA
	XGBCDA	CircR2Disease	XGBoost	0.9860	See github.com/Q1DT/XGBCDA website.
	NMFCDA	CircR2Disease	neural network Pseudoinverse Learning	0.9278	See github.com/look0012/NMFCDA website.
	MSPCD IGNSCDA	CircFunBase CircR2Disease	neural networks multilayer perceptron	0.9904 0.829	See github.com/dayunliu/MSPCD website. NA
Deep learning-based methods	RGCNCDA	CircR2Disease	map convolutional networks	0.8478	NA
	MDGF-MCEC	CircR2Disease	map convolutional network	0.9744	See github.com/ABard0/MDGF-MCEC website.
	DRGCNCDA	circR2Cancer	convolutional networks	0.9399	NA
	SGANRDA	CircR2Disease	generative adversarial network	0.9411, 0.9223	See github.com/look0012/SGANRDA/ website.
	GANCDA	CircR2Disease	generation confrontation network	0.906	NA
	GCNCDA	CircR2Disease	fastGCN	0.912, 0.9278	See github.com/look0012/GCNCDA/ website.
	GATCDA	CircR2Diseas, CircAtlas2.0, Circ2Disease	graph attention network	0.9011	NA
	GraphCDA	CircR2Disease	graph convolutional network	0.9548	See github.com/Ziqiang-Liu/Predict website.
	KGANCDA	circR2Cancer	knowledge graph attention network	0.8847	See github.com/lanbiolab/KGANCDA website.
	GMNN2CD	CircR2Disease	graph convolutional network	0.9634	See github.com/nmt315320/GMNN2CD website.
	iCircDA-NEAE CDHGNN CLCDA	CircR2Disease CircR2Disease CircR2Disease	convolutional autoencoder graph neural network model graph autoencoder	0.8962 0.886 0.998	See github.com/nathanyl/iCircDA-NEAE website. NA See github.com/Lxinmeng/CLCDA website.

supervised ranking of associations. The prediction results on two independent test sets demonstrate that iCircDA-LTR is superior to other methods, especially in predicting diseases associated with new circRNA.

Wei et al. proposed a new calculation method iCircDA-MF [80]. iCircDA-MF first computes feature networks for circRNA genes, gene diseases, and circRNA diseases. Then, to correct false-negative associations, circRNA-disease interaction profiles were updated using neighbor interaction profiles. Finally, the updated circRNA-disease interaction profiles were matrix decomposed. The AUC value of FFCV is 0.9178. The iCircDA-MF remodeled the sparse associated adjacency matrix with neighbor interaction profiles, which

could correct false-negatives in the original associated adjacency matrix. iCircDA-MF computes association scores based on matrix factorization (MF) and is able to detect meaningful potential features from sparse matrices.

Peng et al. proposed RNMFLP [81]. RNMFLP is implemented based on the robust nonnegative MF (RNMF) and label propagation algorithm. First, to reduce the impact of false-negative data, the original association matrix was updated by matrix multiplication. Subsequently, potential circRNA-disease pairs were captured from the association matrix using the RNMF algorithm, obtaining a restricted latent space. Finally, tag propagation algorithms were utilized. The AUC is 0.9599.

Recommended algorithms such as collaborative filtering recommendation and matrix decomposition have been applied in disease-related circRNA prediction, but most of the recommended algorithms may predict suboptimal results with sparse and single data.

4.3 Prediction method based on ML

Typically, ML algorithms train a learning model based on the known circRNA-disease association and then apply the learned model to make predictions on novel associations. Such methods require the integration of sequence information, structure, and expression data before predictively integrated annotation of novel circRNA-disease associations can be performed. Unlike recommended methods, ML methods employ various supervised classification algorithms to distinguish whether target circRNAs are associated with some specific diseases.

Zheng et al. proposed a computational approach iCDA-CGR using quantified localization and nonlinear information to predict associations [82]. iCDA-CGR incorporates circRNA sequence, gene-circRNA association information and disease semantic information, and then uses the SVM classifier for prediction. ICDA-CGR introduced circRNA sequence information for the first time, extracted biological position information and nonlinear relationship of circRNA, using chaos game representation. ACC is 0.9518 and AUC is 0.8645.

Based on feature matrices of disease and circRNAs, Wang et al. extracted hidden features using a stacked autoencoder algorithm, used the RF for prediction, and proposed a new prediction model IMS-CDA [83]. The CircR2Disease dataset resulted in an AUC of 88.08% and the accuracy of 88.36%. The overall performance of IMS-CDA is best compared to SVM and K-nearest neighbor models. More concretely, IMS-CDA incorporated information on disease semantic similarity, disease Jaccard and Gaussian interaction contour nuclear similarity. In the case study, 8 of the top 15 associations with the highest prediction score were validated by the literature.

Shen et al. proposed the XGBCDA method based on multiple heterogeneous networks to predict latent associations [84]. This method first extracted statistical features and graph theoretical features. The method used trees learned by XGBoost, coded with $1(k)$ to represent the latent features. The AUC is 0.9860.

Wang et al. predicted potential associations based on neural network pseudo inverse learning (PIL) by randomization and nonnegative matrix decomposition [85]. This model NMFCDAs began by extracting sequence features and similar features of circRNAs, semantic features of diseases, and similar features. Fusion of circRNAs and disease signatures was then performed; Key feature extraction was subsequently performed using nonnegative matrix factorization. The potential associations were finally predicted using a randomization based PIL for the global optimal solution. The accuracy of 92.56% and an AUC of 0.9278.

Deng et al. presented MSPCD to infer circRNA-disease associations [86]. MSPCD computes biological information such as circRNA-miRNA associations and circRNA-gene

ontology associations, and then uses neural networks to extract higher-order features of circRNAs and diseases. Ultimately, MSPCD employed DNN to predict unknown associations. The AUC is 0.9904.

Yan et al. developed a Kronecker product kernel regularized least squares model (DWNN-RLS) [87]. DWNN-RLS utilized Kronecker to compute a kernel similarity feature for circRNA disease pairs, and then calculated initial relationship scores for novel circRNAs to disease using decreasing weights k-nearest neighbors. The AUC values in FFCV, TFCV, and LOOV are 0.8854, 0.9205, and 0.9701, respectively.

Lan et al. proposed the multilayer perceptual machine-based association prediction model IGNSCDA [88]. A heterogeneous network was constructed based on known circRNA-disease associations, and IGNSCDA began by constructing an improved graph convolutional network to obtain feature vectors. Then, a multilayer perceptron was employed to predict the association of circRNAs with diseases. In addition, to reduce the effect of noisy samples, IGNSCDA employs a negative sampling method to select negative samples based on the expression profile similarity and Gaussian interaction profile kernel similarity of circRNAs. The performance of the proposed method was evaluated by the IGNSCDA using FFCV, which yielded an AUC of 0.829.

For supervised learning algorithms, it needs to be assumed that the disease associated circRNAs and unrelated circRNAs are isolated; however, the number of circRNAs that are proven to be disease associated is still low. Therefore, this study developed some models based on semi-supervised learning. The biggest challenge facing ML algorithms is collecting more data and selecting effective biological features for classifiers. Therefore, we can get richer features by collecting multiple data to improve performance. However, irrelevant biological data may be redundant and even degrade the performance of classification models. At the same time, different classification algorithms may work well for specific data. Therefore, better prediction performance can be obtained by utilizing multiple learning algorithms.

4.4 Prediction method based on DL

DL has been widely applied to solve various problems of bioinformatics and has achieved good results. Moreover, initial circRNA disease association prediction studies largely emphasized particular emphasis on identifying potential associations based on shallow learning methods. These methods cannot extract deep feature representations. Therefore, researchers have attempted to predict potential associations by constructing different DL networks.

Based on fused circRNAs and disease features, Wang et al. presented a computational method SGANRD employing deep generative adversarial network (GAN) algorithm to predict associations [89]. First, the authors fused multisource information such as disease semantic similarity, then used GAN to extract features of the fused information objectively and effectively, and finally sent it to a logistic model tree classifier. On the Circ2Disease dataset, the AUC value of FFCV was 90.6%. Nine of the 15 associations were confirmed

in the case analysis.

Chen et al. proposed RGCNCDA, a computational method based on relationship map graph convolutional networks (R-GCNs) for prediction [90]. RGCNCDA first constructed a global heterogeneous network of circRNAs, miRNAs and diseases. Then, RGCNCDA employed RW and principal component analysis to learn higher-order information as topological features from heterogeneous networks. Finally, RGCNCDA based on the R-GCNs encoder and the DistMult decoder to predict. The AUC value of the FFCV result was 0.8478.

Yan et al. proposed the model GANCDNA based on a deep GAN to predict disease associated circRNAs [91]. GANCDNA achieved an AUC of 90.6% for FFCV on the Circ2Disease dataset. In addition, the predicted effect of GANCDNA was also validated, supported by biological experiments.

Wu et al. proposed MDGF-MCEC to predict the association between circRNAs and diseases [92]. MDGF-MCEC mainly uses multiview double attention map convolutional network to build a model with ensemble learning. First, MDGF-MCEC constructed a circRNA and disease relationship map based on distinct similarities. Then, representation learning is performed using the multiview graph convolutional network (GCN), and the attentional mechanism is introduced to learn high-dimensional valid features. Finally, the authors constructed a multiperspective synergistic ensemble classifier to predict and experiment on the Circ2Disease database with an AUC value of 0.9744.

Wang et al. presented a prediction framework GCNCDA based on rapid learning with GCN to predict potential associations [93]. Specifically, this method calculated a unified descriptor based on known circRNA disease associations. The high-level features in the descriptors were then objectively extracted using the fastGCN algorithm. Finally, associations are predicted by RF. The FFCV at GCNCDA had an AUC of 90.90%. Case study experiments have been conducted in breast cancer, glioma, colorectal cancer and other diseases. Among the top 20 candidate circRNAs, number 16, 15, and 17 were verified by the literature and databases.

Bian et al. constructed a new computational framework, called GATCDA, utilizing the graph attention network (GAT) [94]. GAT uses an attention mechanism to learn the representation of nodes on a graph by assigning the weights of different nodes. Among them, GATCDA adopted features of circRNA-miRNA interactions and disease-mRNA interactions, considering the effect of the circRNA-miRNA-mRNA axis in the occurrence of diseases. The AUC value of GATCDA was 0.9011 by FFCV.

Based on the circRNA similar network and disease similar network, Dai et al. constructed a hybrid graph embedded model combining the graph convolutional network and graph attention network GraphCDA [95] to identify circRNA disease associations while learning the feature characterization of circRNAs and diseases. However, it also, experimentally demonstrates GraphCDA's excellent performance on several public databases. The average AUC of the fivefold cross validation was 0.9458.

Lan et al. proposed a novel knowledge graph attention network-based computational approach (KGANCDNA) to predict circRNA-disease associations [96]. By collecting data on multiple relationships among circRNAs, diseases, miRNAs, and lncRNAs, KGANCDNA constructed a circRNA disease knowledge map. Then, a knowledge map is designed by paying attention to the network to obtain the embedded value by distinguishing the importance information. KGANCDNA also captures higher-order neighbor information, alleviating the problem of data rarefaction. The AUC was 0.8847.

Lan et al. proposed DRGCNCDA, based on convolutional networks of entanglement diagrams [97]. A circRNA disease multiple relationship map was constructed by collecting data on multiple relationships among circRNAs, diseases, miRNAs, and lncRNAs. Then, the feature vectors were obtained using the deconvoluted relationship map convolutional network. Finally, a knowledge map model was applied to predict an affinity score based on the embedding of circRNAs with disease. The AUC value was 0.9399.

Yuan et al. introduced circRNA expression profile and Jaccard similarity and developed a new DL model (iCircDA-NEAE) [98]. iCircDA-NEAE is based on accelerated attribute network embedding and dynamic smooth autoencoder extraction. This method can fully extract biometric information to represent the association features. The ACC and AUC were obtained to be 0.8735 and 0.8962 respectively.

Lu et al. proposed CDHGNN based on edge weighted graph attention and heterogeneous graph neural network [99]. CDHGNN first builds a heterogeneous network based on multi-source data. In order to reflect the association probability between network nodes, a weight mechanism is introduced, and CDHGNN designs an edge-weighted graph attention network model. And achieved an ACC of 0.824 and an AUC of 0.886.

Based on the similarity network of circRNA and diseases, Niu et al. proposed a computational model GMNN2CD based on graph Markov neural network, which integrates variational inference and graph autoencoder to achieve circRNA-disease prediction [100]. A feature inference network is designed to infer representations from circRNA and disease features, and a label propagation network is designed to propagate labels from known circRNA-disease associations. The variational expectation maximization algorithm is used to alternately train two autoencoders. Finally, 5-fold cross-validation and case analysis are used to prove the performance of GMNN2CD.

As mentioned above, although deep learning-based methods have made great progress in prediction, each method has advantages and limitations. Classic deep learning technology mainly extracts high-level features from circRNA and disease attributes (such as semantic information) through models such as CNN. A classifier is then designed to predict circRNA-disease pair scores. Its main advantage is automatic nonlinear representation learning, however, most nonlinear representations extracted based on classical deep learning models have poor interpretability. Furthermore, it is unreasonable to rely on one kind of similarity information or simple integration of multiple similarities, which may ignore

the complementary information of multiple biological knowledges. Furthermore, the available biological data is limited, which results in the constructed network being sparse, which in turn further limits the performance of deep learning networks.

5 Challenges and conclusion

Since their discovery, circRNAs have been a challenging research focus in biological sciences. From the discovery of circRNAs to the present, researchers around the world have contributed to the well-established field of circRNA biogenesis, structure, and function. An increasing number of studies have proven that circRNAs are closely related to a variety of diseases, and the prediction of the association is a new research direction. Studying the relationship will help to deepen the understanding of the pathogenic mechanism of diseases and lay the foundation for subsequent disease diagnosis, treatment, and prevention. Compared with traditional biological experiments and clinical experiments, effective algorithms and tools can save time and energy, and achieve the purpose of predicting association. Different methods of predicting associations based on biological characteristics can help us further explore the mystery of circRNA and promote the study of circRNA function [16,101–103].

This paper provides a brief overview of some available public databases, algorithms and tools developed to predict circRNA-disease associations. Although these algorithms and tools have achieved certain results, there are still some shortcomings, and more in-depth research and improvement are needed in the future. (1) There is an imbalance of circRNA disease association positive samples versus negative samples. There are two main solutions. 1) Follow the progress of the study to obtain new data. 2) The same number of positive and negative samples can be extracted to solve the imbalance of the data. (2) Most of the current computational methods predict circRNA disease associations that are covered in known association datasets; they simply predict fewer novel circRNA disease associations and do not detect diseases associated with novel circRNAs well. Therefore, we will adopt more biological data to overcome this weakness. (3) Existing computational models are mostly based on incompletely related biological information, there is redundancy and noise between data, and circRNA similarity and disease similarity cannot be sufficiently fused. (4) The generalizable performance of the prediction model was not validated on other circRNA disease association datasets, and computational models cannot infer new circRNAs without any relationship to disease.

Prediction accuracy is likely to improve further as data accumulate in the future. Integrating different types of heterogeneous networks, such as circRNA or disease-related miRNA/gene interaction networks, can also improve the performance of potential association predictions. In the future, constructing efficient computational models and rationally fusing similarity scores with different sources of biological information may also become research hotspots.

We will also point out some important issues in current

research and outline several research directions worthy of advancing computational associations. (1) Model interpretability. The most critical challenge of the new deep learning approaches remains interpretability. (2) Computational and experimental integration. Bioinformatics analysis has been widely used in life science research, which collects, stores, identifies and analyzes DNA information or protein functions of various diseases with the help of computers. It is used to explore the pathogenic mechanisms of various diseases, find key genes and diagnostic molecular markers, and is widely used in clinical early diagnosis and new drug research. In recent years, machine learning has continued to develop. Based on biological information data, it can verify and predict genes, present complex biological information in a simple and intuitive way, and greatly promote the development of precision medicine. Establishing accurate and reliable prediction models and verifying the effectiveness of the models through biological validation is of great significance to the development of computational biology. Although multiple prediction models for the association between circRNA and diseases have been constructed. However, there have been few follow-up studies on computational models, and biological validation based on computational models is even lacking. Therefore, as a next step, we should also consider combining computational models with biological validation and using more biological data to verify the effectiveness of the model. (3) There are still few disease association pairs related to circRNA and its biological experimental verification, and there are no clear negative sample relationship pairs. How to effectively expand positive sample correlation pairs, select high-quality negative sample correlation pairs, reduce the negative impact of sample noise, and ensure that the computing model fully learns the correlation pattern is an important engineering optimization approach to improve model prediction performance.

Acknowledgements The work was supported by the National Natural Science Foundation of China (Grant Nos. 62231013, 62201129, 62303328, 62302341, 62271329, 62372332), the National Key R&D Program of China (2022ZD0117700), the National funded postdoctoral researcher program of China (GZC20230382), the Shenzhen Polytechnic University Research Fund (6024310027K, 6022310036K, 6023310037K), the Key Field of Department of Education of Guangdong Province (2022ZDZX2082), and the Special Science Foundation of Quzhou (2023D036). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests The authors declare that they have no competing interests or financial conflicts to disclose.

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.

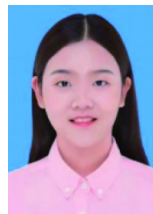
To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Sanger H L, Klotz G, Riesner D, Gross H J, Kleinschmidt A K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proceedings of the National Academy of Sciences of the United States of America, 1976, 73(11): 3852–3856
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak S D, Gregersen L H, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, Le Noble F, Rajewsky N. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature, 2013, 495(7441): 333–338
- Qu S, Yang X, Li X, Wang J, Gao Y, Shang R, Sun W, Dou K, Li H. Circular RNA: a new star of noncoding RNAs. Cancer Letters, 2015, 365(2): 141–148
- Ye C Y, Chen L, Liu C, Zhu Q H, Fan L. Widespread noncoding circular RNAs in plants. New Phytologist, 2015, 208(1): 88–95
- Hsiao K Y, Sun H S, Tsai S J. Circular RNA—new member of noncoding RNA with novel functions. Experimental Biology and Medicine, 2017, 242(11): 1136–1141
- Jeck W R, Sharpless N E. Detecting and characterizing circular RNAs. Nature Biotechnology, 2014, 32(5): 453–461
- Meng S, Zhou H, Feng Z, Xu Z, Tang Y, Li P, Wu M. CircRNA: functions and properties of a novel potential biomarker for cancer. Molecular Cancer, 2017, 16(1): 94
- Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J, Chen D, Gu J, He X, Huang S. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. Cell Research, 2015, 25(8): 981–984
- Verduci L, Strano S, Yarden Y, Blandino G. The circRNA—microRNA code: emerging implications for cancer diagnosis and treatment. Molecular Oncology, 2019, 13(4): 669–680
- Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Luo Y, Lyu D, Li Y, Shi G, Liang L, Gu J, He X, Huang S. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nature Communications, 2016, 7(1): 11215
- Verduci L, Tarcitano E, Strano S, Yarden Y, Blandino G. CircRNAs: role in human diseases and potential use as biomarkers. Cell Death & Disease, 2021, 12(5): 468
- Wang Y, Zhang X, Ju Y, Liu Q, Zou Q, Zhang Y, Ding Y, Zhang Y. Identification of human microRNA-disease association via low-rank approximation-based link propagation and multiple kernel learning. Frontiers of Computer Science, 2024, 18(2): 182903
- Wang C-C, Han C-D, Zhao Q, Chen X. Circular RNAs and complex diseases: from experimental results to computational models. Briefings in Bioinformatics, 2021, 22(6): bbab286
- Lan W, Dong Y, Zhang H, Li C, Chen Q, Liu J, Wang J, Chen Y P P. Benchmarking of computational methods for predicting circRNA-disease associations. Briefings in Bioinformatics, 2023, 24(1): bbac613
- Chen Y, Wang J, Wang C, Liu M, Zou Q. Deep learning models for disease-associated circRNA prediction: a review. Briefings in Bioinformatics, 2022, 23(6): bbac364
- Xiao Q, Dai J, Luo J. A survey of circular RNAs in complex diseases: databases, tools and computational methods. Briefings in Bioinformatics, 2022, 23(1): bbab444
- Belousova E A, Filipenko M L, Kushlinskii N E. Circular RNA: new regulatory molecules. Bulletin of Experimental Biology and Medicine, 2018, 164(6): 803–815
- Gao J-L, Chen G, He H-Q, Wang J. CircRNA as a new field in human disease research. China Journal of Chinese Materia Medica, 2018, 43(3): 457–462
- Lou J, Hao Y, Lin K, Lyu Y, Chen M, Wang H, Zou D, Jiang X, Wang R, Jin D, Lam E W F, Shao S, Liu Q, Yan J, Wang X, Chen P, Zhang B, Jin B. Circular RNA *CDRIAs* disrupts the p53/MDM2 complex to inhibit Gliomagenesis. Molecular Cancer, 2020, 19(1): 138
- Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. Cell, 1993, 73(5): 1019–1030
- Pamudurti N R, Patop I L, Krishnamoorthy A, Bartok O, Maya R, Lerner N, Ashwall-Fluss R, Konakondla J V V, Beatus T, Kadener S. circMbl functions in *cis* and in *trans* to regulate gene expression and physiology in a tissue-specific fashion. Cell Reports, 2022, 39(4): 110740
- Panda A C. Circular RNAs act as miRNA sponges. Circular RNAs: Biogenesis and Functions, 2018: 67-79
- Su M, Xiao Y, Ma J, Tang Y, Tian B, Zhang Y, Li X, Wu Z, Yang D, Zhou Y, Wang H, Liao Q, Wang W. Circular RNAs in Cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. Molecular Cancer, 2019, 18(1): 90
- Hansen T B, Jensen T I, Clausen B H, Bramsen J B, Finsen B, et al. Natural RNA circles function as efficient microRNA sponges. Nature, 2013, 495(7441): 384-388
- Gupta S K, Garg A, Bär C, Chatterjee S, Foinquinos A, Milting H, Streckfuß-Bömeke K, Fiedler J, Thum T. Quaking inhibits doxorubicin-mediated cardiotoxicity through regulation of cardiac circular RNA expression. Circulation Research, 2018, 122(2): 246–254
- Chen Y-J, Chen C-Y, Mai T-L, Chuang C-F, Chen Y-C, Gupta S K, Yen L, Wang Y D, Chuang T J. Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNA-microRNA-mRNA regulatory axes in autism. Genome Research, 2020, 30(3): 375–391
- Zhang F, Zhang R, Zhang X, Wu Y, Li X, Zhang S, Hou W, Ding Y, Tian J, Sun L, Kong X. Comprehensive analysis of circRNA expression pattern and circRNA-miRNA-mRNA network in the pathogenesis of atherosclerosis in rabbits. Aging, 2018, 10(9): 2266–2283
- Chia W, Liu J, Huang Y-G, Zhang C. A circular RNA derived from DAB1 promotes cell proliferation and osteogenic differentiation of BMSCs via RBPJ/DAB1 axis. Cell Death & Disease, 2020, 11(5): 372
- Liu Y, Song J, Liu Y, Zhou Z, Wang X. Transcription activation of circ-STAT3 induced by Gli2 promotes the progression of hepatoblastoma via acting as a sponge for miR-29a/b/c-3p to upregulate STAT3/Gli2. Journal of Experimental & Clinical Cancer Research, 2020, 39(1): 101
- Kong P, Yu Y, Wang L, Dou Y-Q, Zhang X-H, Cui Y, Wang H-Y, Yong Y-T, Liu Y-B, Hu H-J, Cui W, Sun S-G, Li B-H, Zhang F, Han M. circ-Sirt1 controls NF- κ B activation via sequence-specific interaction and enhancement of SIRT1 expression by binding to miR-132/212 in vascular smooth muscle cells. Nucleic Acids Research, 2019, 47(7): 3580–3593
- Liang W-C, Wong C-W, Liang P-P, Shi M, Cao Y, Rao S-T, Tsui S K W, Wayne M M Y, Zhang Q, Fu W-M, Zhang J-F. Translation of the circular RNA circ β -catenin promotes liver cancer cell growth through activation of the Wnt pathway. Genome Biology, 2019, 20(1): 84
- Bai N, Peng E, Qiu X, Lyu N, Zhang Z, Tao Y, Li X, Wang Z. circFBLIM1 act as a ceRNA to promote hepatocellular cancer progression by sponging miR-346. Journal of Experimental & Clinical Cancer Research, 2018, 37(1): 172
- Zhou J, Zhang S, Chen Z, He Z, Xu Y, Li Z. CircRNA-ENO1 promoted glycolysis and tumor progression in lung adenocarcinoma through upregulating its host gene ENO1. Cell Death & Disease, 2019, 10(12): 885

34. He R, Liu P, Xie X, Zhou Y, Liao Q, Xiong W, Li X, Li G, Zeng Z, Tang H. circGFRA1 and GFRA1 act as ceRNAs in triple negative breast cancer by regulating miR-34a. *Journal of Experimental & Clinical Cancer Research*, 2017, 36(1): 145
35. Ou R, Lv J, Zhang Q, Lin F, Zhu L, Huang F, Li X, Li T, Zhao L, Ren Y, Xu Y. circAMOTL1 motivates AMOTL1 expression to facilitate cervical cancer growth. *Molecular Therapy Nucleic Acids*, 2020, 19: 50–60
36. Yang L, Zeng Z, Kang N, Yang J C, Wei X, Hai Y. Circ-VANGL1 promotes the progression of osteoporosis by absorbing miRNA-217 to regulate RUNX2 expression. *European Review for Medical and Pharmacological Sciences*, 2019, 23(3): 949–957
37. Wan L, Han Q, Zhu B, Kong Z, Feng E. Circ-TFF1 facilitates breast cancer development via regulation of miR-338-3p/FGFR1 Axis. *Biochemical Genetics*, 2022, 60(1): 315–335
38. Lu M. Circular RNA: functions, applications and prospects. *ExRNA*, 2020, 2(1): 1
39. Geng X, Lin X, Zhang Y, Li Q, Guo Y, Fang C, Wang H. Exosomal circular RNA sorting mechanisms and their function in promoting or inhibiting cancer. *Oncology Letters*, 2020, 19(5): 3369–3380
40. Auffiero S, Reckman Y J, Pinto Y M, Creemers E E. Circular RNAs open a new chapter in cardiovascular biology. *Nature Reviews Cardiology*, 2019, 16(8): 503–514
41. Xu Z, Song L, Liu S, Zhang W. DeepCRBP: improved predicting function of circRNA-RBP binding sites with deep feature learning. *Frontiers of Computer Science*, 2024, 18(2): 182907
42. Guo Y, Lei X, Liu L, Pan Y. circ2CBA: prediction of circRNA-RBP binding sites combining deep learning and attention mechanism. *Frontiers of Computer Science*, 2023, 17(5): 175904
43. Zhou R, Wu Y, Wang W, Su W, Liu Y, Wang Y, Fan C, Li X, Li G, Li Y, Xiong W, Zeng Z. Circular RNAs (circRNAs) in cancer. *Cancer Letters*, 2018, 425: 134–142
44. Peng L, Yuan X Q, Li G C. The emerging landscape of circular RNA ciRS-7 in cancer. *Oncology Reports*, 2015, 33(6): 2669–2674
45. Chen B, Huang S. Circular RNA: an emerging non-coding RNA as a regulator and biomarker in cancer. *Cancer Letters*, 2018, 418: 41–50
46. Hansen T B, Kjems J, Damgaard C K. Circular RNA and miR-7 in cancer. *Cancer Research*, 2013, 73(18): 5609–5612
47. Akhter R. Circular RNA and Alzheimer's disease. In: Xiao J, ed. *Circular RNAs: Biogenesis and Functions*. Singapore: Springer, 2018, 239–243
48. Hong H, Zhu H, Zhao S, Wang K, Zhang N, Tian Y, Li Y, Wang Y, Lv X, Wei T, Liu Y, Fan S, Liu Y, Li Y, Cai A, Jin S, Qin Q, Li H. The novel circCLK3/miR-320a/FoxM1 axis promotes cervical cancer progression. *Cell Death & Disease*, 2019, 10(12): 950
49. Ashwal-Fluss R, Meyer M, Pamudurti N R, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. circRNA biogenesis competes with pre-mRNA splicing. *Molecular Cell*, 2014, 56(1): 55–66
50. Ji X, Shan L, Shen P, He M. Circular RNA circ_001621 promotes osteosarcoma cells proliferation and migration by sponging miR-578 and regulating VEGF expression. *Cell Death & Disease*, 2020, 11(1): 18
51. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA*, 2014, 20(11): 1666–1670
52. Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, Zheng Q, Li Y, Wang P, He X, Huang S. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Research*, 2018, 46(D1): D106–D112
53. Dong R, Ma X-K, Li G-W, Yang L. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics, Proteomics & Bioinformatics*, 2018, 16(4): 226–233
54. Pan X, Xiong K, Anthon C, Hyttel P, Freude K K, Jensen L J, Gorodkin J. WebCircRNA: classifying the circular RNA potential of coding and noncoding RNA. *Genes*, 2018, 9(11): 536
55. Ruan H, Xiang Y, Ko J, Li S, Jing Y, Zhu X, Ye Y, Zhang Z, Mills T, Feng J, Liu C J, Jing J, Cao J, Zhou B, Wang L, Zhou Y, Lin C, Guo A Y, Chen X, Diao L, Li W, Chen Z, He X, Mills G B, Blackburn M R, Han L. Comprehensive characterization of circular RNAs in ~ 1000 human cancer cell lines. *Genome Medicine*, 2019, 11(1): 55
56. Liu Q, Cai Y, Xiong H, Deng Y, Dai X. CCRDB: a cancer circRNAs-related database and its application in hepatocellular carcinoma-related circRNAs. *Database*, 2019, 2019: baz063
57. Zhao M, Liu Y, Qu H. circExp database: An online transcriptome platform for human circRNA expressions in cancers. *Database*, 2021, 2021: baab045
58. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. *Database*, 2019, 2019: baz003
59. Dudekula D B, Panda A C, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biology*, 2016, 13(1): 34–42
60. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Scientific Reports*, 2016, 6(1): 34985
61. Hamosh A, Scott A F, Amberger J S, Bocchini C A, McKusick V A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 2005, 33(S1): D514–D517
62. Rappaport N, Nativ N, Stelzer G, Twik M, Guan-Golan Y, Stein T I, Bahir I, Belinky F, Morrey C P, Safran M, Lancet D. MalaCards: an integrated compendium for diseases and their annotation. *Database*, 2013, 2013: bat018
63. Canese K, Weis S. PubMed: the bibliographic database. 2002 Oct 9 [Updated 2013 Mar 20]. In: *The NCBI Handbook*[Internet]. 2nd edn. Bethesda (MD): National Center for Biotechnology Information (US). 2013, Available from the website of ncbi.nlm.nih.gov/books/NBK153385/
64. Zhu L, Ren T, Zhu Z, Cheng M, Mou Q, Mu M, Liu Y, Yao Y, Cheng Y, Zhang B, Cheng Z. Thymosin-β4 mediates hepatic stellate cell activation by interfering with CircRNA-0067835/miR-155/FoxO3 signaling pathway. *Cellular Physiology and Biochemistry*, 2018, 51(3): 1389–1398
65. Zhao Z, Wang K, Wu F, Wang W, Zhang K, Hu H, Liu Y, Jiang T. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death & Disease*, 2018, 9(5): 475
66. Yao D, Zhang L, Zheng M, Sun X, Lu Y, Liu P. Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Scientific Reports*, 2018, 8(1): 11018
67. Zhang W, Liu Y, Min Z, Liang G, Mo J, Ju Z, Zeng B, Guan W, Zhang Y, Chen J, Zhang Q, Li H, Zeng C, Wei Y, Chan G C F. circMine: a comprehensive database to integrate, analyze and visualize human disease-related circRNA transcriptome. *Nucleic Acids Research*, 2022, 50(D1): D83–D92
68. Rophina M, Sharma D, Poojary M, Scaria V. Circad: a comprehensive manually curated resource of circular RNA associated with diseases. *Database*, 2020, 2020: baaa019
69. Lan W, Zhu M, Chen Q, Chen B, Liu J, Li M, Chen Y P P. CircR2Cancer: a manually curated database of associations between circRNAs and cancers. *Database*, 2020, 2020: baaa085
70. Deng L, Zhang W, Shi Y, Tang Y. Fusion of multiple heterogeneous networks for predicting circRNA-disease associations. *Scientific Reports*, 2019, 9(1): 9605
71. Xiao Q, Yu H, Zhong J, Liang C, Li G, Ding P, Luo J. An *in-silico*

- method with graph-based multi-label learning for large-scale prediction of circRNA-disease associations. *Genomics*, 2020, 112(5): 3407–3415
72. Xiao Q, Fu Y, Yang Y, Dai J, Luo J. NSL2CD: identifying potential circRNA-disease associations based on network embedding and subspace learning. *Briefings in Bioinformatics*, 2021, 22(6): bbab177
 73. Lei X, Fang Z, Chen L, Wu F-X. PWCDA: path weighted method for predicting circRNA-disease associations. *International Journal of Molecular Sciences*, 2018, 19(11): 3410
 74. Li G, Yue Y, Liang C, Xiao Q, Ding P, Luo J. NCPEDA: network consistency projection for circRNA-disease association prediction. *RSC Advances*, 2019, 9(57): 33222–33228
 75. Xiao Q, Zhong J, Tang X, Luo J. iCDA-CMG: identifying circRNA-disease associations by federating multi-similarity fusion and collective matrix completion. *Molecular Genetics and Genomics*, 2021, 296(1): 223–233
 76. Shu L, Zhou C, Yuan X, Zhang J, Deng L. MSCFS: inferring circRNA functional similarity based on multiple data sources. *BMC Bioinformatics*, 2021, 22(10): 371
 77. Lei X, Zhang W. BRWSP: predicting circRNA-disease associations based on biased random walk to search paths on a multiple heterogeneous network. *Complexity*, 2019, 2019: 5938035
 78. Lei X, Fang Z, Guo L. Predicting circRNA-disease associations based on improved collaboration filtering recommendation system with multiple data. *Frontiers in Genetics*, 2019, 10: 897
 79. Wei H, Xu Y, Liu B. iCircDA-LTR: identification of circRNA-disease associations based on Learning to Rank. *Bioinformatics*, 2021, 37(19): 3302–3310
 80. Wei H, Liu B. iCircDA-MF: identification of circRNA-disease associations based on matrix factorization. *Briefings in Bioinformatics*, 2020, 21(4): 1356–1367
 81. Peng L, Yang C, Huang L, Chen X, Fu X, Liu W. RNMFLP: predicting circRNA-disease associations based on robust nonnegative matrix factorization and label propagation. *Briefings in Bioinformatics*, 2022, 23(5): bbac155
 82. 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 Computational Biology*, 2020, 16(5): e1007872
 83. Wang L, You Z H, Li J Q, Huang Y A. IMS-CDA: prediction of CircRNA-disease associations from the integration of multisource similarity information with deep stacked autoencoder model. *IEEE Transactions on Cybernetics*, 2021, 51(11): 5522–5531
 84. Shen S, Liu J, Zhou C, Qian Y, Deng L. XGBCDA: a multiple heterogeneous networks-based method for predicting circRNA-disease associations. *BMC Medical Genomics*, 2022, 13(1): 196
 85. Wang L, You Z-H, Zhou X, Yan X, Li H-Y, Huang Y-A. NMFEDA: Combining randomization-based neural network with non-negative matrix factorization for predicting CircRNA-disease association. *Applied Soft Computing*, 2021, 110: 107629
 86. Deng L, Liu D, Li Y, Wang R, Liu J, Zhang J, Liu H. MSPCD: predicting circRNA-disease associations via integrating multi-source data and hierarchical neural network. *BMC Bioinformatics*, 2022, 23(3): 427
 87. Yan C, Wang J, Wu F-X. DWNN-RLS: regularized least squares method for predicting circRNA-disease associations. *BMC Bioinformatics*, 2018, 19(19): 520
 88. Lan W, Dong Y, Chen Q, Liu J, Wang J, Chen Y P P, Pan S. IGNSCDA: predicting CircRNA-disease associations based on improved graph convolutional network and negative sampling. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, 19(6): 3530–3538
 89. Wang L, Yan X, You Z-H, Zhou X, Li H-Y, Huang Y-A. SGANRDA: semi-supervised generative adversarial networks for predicting circRNA-disease associations. *Briefings in Bioinformatics*, 2021, 22(5): bbab028
 90. Chen Y, Wang Y, Ding Y, Su X, Wang C. RGCNCDA: Relational graph convolutional network improves circRNA-disease association prediction by incorporating microRNAs. *Computers in Biology and Medicine*, 2022, 143: 105322
 91. Yan X, Wang L, You Z-H, Li L-P, Zheng K. GANCDA: a novel method for predicting circRNA-disease associations based on deep generative adversarial network. *International Journal of Data Mining and Bioinformatics*, 2020, 23(3): 265–283
 92. Wu Q, Deng Z, Pan X, Shen H-B, Choi K-S, Wang S, Wu J, Yu D J. MDGF-MCEC: a multi-view dual attention embedding model with cooperative ensemble learning for CircRNA-disease association prediction. *Briefings in Bioinformatics*, 2022, 23(5): bbac289
 93. Wang L, You Z-H, Li Y-M, Zheng K, Huang Y-A. GCNCDA: a new method for predicting circRNA-disease associations based on graph convolutional network algorithm. *PLoS Computational Biology*, 2020, 16(5): e1007568
 94. Bian C, Lei X-J, Wu F-X. GATCDA: predicting circRNA-disease associations based on graph attention network. *Cancers*, 2021, 13(11): 2595
 95. Dai Q, Liu Z, Wang Z, Duan X, Guo M. GraphCDA: a hybrid graph representation learning framework based on GCN and GAT for predicting disease-associated circRNAs. *Briefings in Bioinformatics*, 2022, 23(5): bbac379
 96. Lan W, Dong Y, Chen Q, Zheng R, Liu J, Pan Y, Chen Y P P. KGANCDA: predicting circRNA-disease associations based on knowledge graph attention network. *Briefings in Bioinformatics*, 2022, 23(1): bbab494
 97. Lan W, Zhang H, Dong Y, Chen Q, Cao J, Peng W, Liu J, Li M. DRGCNCDA: Predicting circRNA-disease interactions based on knowledge graph and disentangled relational graph convolutional network. *Methods*, 2022, 208: 35–41
 98. Yuan L, Zhao J, Shen Z, Zhang Q, Geng Y, Zheng C-H, Huang D-S. iCircDA-NEAE: Accelerated attribute network embedding and dynamic convolutional autoencoder for circRNA-disease associations prediction. *PLoS Computational Biology*, 2023, 19(8): e1011344
 99. Lu C, Zhang L, Zeng M, Lan W, Wang J. Identifying disease-associated circRNAs based on edge-weighted graph attention and heterogeneous graph neural network. *bioRxiv*, 2022: 2022.05.04.490565
 100. Niu M, Zou Q, Wang C. GMNN2CD: identification of circRNA-disease associations based on variational inference and graph Markov neural networks. *Bioinformatics*, 2022, 38(8): 2246–2253
 101. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-Pro: support bio-sequence machine for proteins. 2023, arXiv preprint arXiv: 2308.10275
 102. Fan C, Lei X, Wu F-X. Prediction of CircRNA-disease associations using KATZ model based on heterogeneous networks. *International Journal of Biological Sciences*, 2018, 14(14): 1950–1959
 103. Fan C, Lei X, Pan Y. Prioritizing CircRNA-disease associations with convolutional neural network based on multiple similarity feature fusion. *Frontiers in Genetics*, 2020, 11: 540751



Mengting Niu is a postdoctoral fellow at University of Electronic Science and Technology of China and Shenzhen Polytechnic University, China. Her research interests include bioinformatics, data mining, and biomedicine.



Yaojia Chen is a PhD candidate at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, China. Her research interests include machine learning and bioinformatics.

of bioinformatics, machine learning, and parallel computing. Several related works have been published by Science, Briefings in Bioinformatics, Bioinformatics, the IEEE/ACM Transactions on Computational Biology and Bioinformatics, etc. He is the editor-in-chief of Current Bioinformatics, associate editor of IEEE Access, and an editorial board member of Computers in Biology and Medicine, Genes, Scientific Reports, etc.



Chunyu Wang is a professor at Faculty of Computing, Harbin Institute of Technology, China. His research fields include computational biology and machine learning, especially on the structure and function prediction of biomolecules, artificial intelligence-assisted drug discovery, high-throughput sequence data analysis etc.



Lei Xu is an associate professor at the School of Electronic and Communication Engineering, Shenzhen Polytechnic, China. She received her BSc and MSc from the School of Computer Science and Technology in Harbin Institute of Technology, China in 2006 and 2008, respectively. She got her PhD degree from the Department of Computing, The Hong Kong Polytechnic University, China in 2013. Her research interests are focused on bioinformatics, pattern recognition.



Quan Zou received the BSc, MSc, and the PhD degrees in computer science from the Harbin Institute of Technology, China in 2004, 2007, and 2009, respectively. He is currently a professor with the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. His research is in the areas