Prioritizing CircRNA–Disease Associations With Convolutional Neural Network Based on Multiple Similarity Feature Fusion

Chunyan Fan¹, Xiujuan Lei¹* and Yi Pan²*

¹ School of Computer Science, Shaanxi Normal University, Xi’an, China, ² Department of Computer Science, Georgia State University, Atlanta, GA, United States

Accumulating evidence shows that circular RNAs (circRNAs) have significant roles in human health and in the occurrence and development of diseases. Biological researchers have identified disease-related circRNAs that could be considered as potential biomarkers for clinical diagnosis, prognosis, and treatment. However, identification of circRNA–disease associations using traditional biological experiments is still expensive and time-consuming. In this study, we propose a novel method named MSFCNN for the task of circRNA–disease association prediction, involving two-layer convolutional neural networks on a feature matrix that fuses multiple similarity kernels and interaction features among circRNAs, miRNAs, and diseases. First, four circRNA similarity kernels and seven disease similarity kernels are constructed based on the biological or topological properties of circRNAs and diseases. Subsequently, the similarity kernel fusion method is used to integrate the similarity kernels into one circRNA similarity kernel and one disease similarity kernel, respectively. Then, a feature matrix for each circRNA–disease pair is constructed by integrating the fused circRNA similarity kernel and fused disease similarity kernel with interactions and features among circRNAs, miRNAs, and diseases. The features of circRNA–miRNA and disease–miRNA interactions are selected using principal component analysis. Finally, taking the constructed feature matrix as an input, we used two-layer convolutional neural networks to predict circRNA–disease association labels and mine potential novel associations. Five-fold cross validation shows that our proposed model outperforms conventional machine learning methods, including support vector machine, random forest, and multilayer perception approaches. Furthermore, case studies of predicted circRNAs for specific diseases and the top predicted circRNA–disease associations are analyzed. The results show that the MSFCNN model could be an effective tool for mining potential circRNA–disease associations.

Keywords: circRNA–disease associations, circRNA–miRNA interaction, similarity kernel fusion, feature matrix, convolutional neural network
INTRODUCTION

Circular RNAs (circRNAs) are a type of endogenous non-coding RNA with continuous covalently closed loop structures, which are produced by back-splicing or lariat events in genes (Barrett et al., 2015). Recently, with the development of high-throughput sequencing techniques and other technologies, a large number of circRNAs have been found in various organisms, including protists, plants, and metazoans (Danan et al., 2012; Memczak et al., 2013; Tang et al., 2018). The main functions of circRNAs include sequestration of microRNAs (miRNAs) and proteins (Salmena et al., 2011), regulation of transcription and splicing (Zhang et al., 2013; Conn et al., 2017), and even translation to produce polypeptides (Yang et al., 2017; Sun and Li, 2019). Accumulating evidence implicates mutation or alteration in expression of circRNAs in the initiation and progression of numerous diseases. For example, Chioccarelli et al. (2019) identified the differentially expressed circRNAs in human spermatozoa, and found that circRNAs are related to spermatozoa quality. By comparing the expression profiles of circRNAs in disease-specific tissues or cell lines with those in normal samples, significantly increased or decreased circRNAs can be identified. In addition, the intrinsic characteristics of circRNAs indicate they are stable both inside cells and in extracellular plasma (Bahn et al., 2015; Li et al., 2015; Memczak et al., 2015). Therefore, disease-associated circRNAs are considered to be promising novel biomarkers for diseases.

Recently, several studies have analyzed the roles of circRNAs in various samples, and further explore their diversity, expression patterns, co-expression network, and so on. circAtlas integrates the most comprehensive circRNAs, their expression, and functional profiles in vertebrates (Wu et al., 2020). MiOncoCirc is a cancer-focused circRNA resource to be generated from an extensive array of tumor tissues (Vo et al., 2019). Ji et al. (2019) identifies full-length transcripts and evolutionarily conserved circRNAs, and infers circRNA functions on a global scale. Ruan et al. (2019) characterizes circRNAs expression profiles, and explores the potential mechanism of circRNA biogenesis as well as its therapeutic implications. exoRBase integrates and visualizes the RNA expression profiles both normal individuals and patients with different diseases (Li et al., 2018). These studies will trigger functional implication for human diseases and benefit biomedical research community.

The de-regulated circRNAs in diseases can be identified for validation using low-throughput biological methods such as quantitative real-time PCR, northern blotting, and so on. However, these traditional experiments are costly and time-consuming. Therefore, computational approaches are important for exploring potential disease-causing circRNAs and understanding the associated mechanisms of pathogenicity. Several models have been proposed to forecast circRNA–disease associations; most of these approaches are based on the assumption that circRNAs with similar functions are likely to be associated with the same or similar diseases. Lei et al. (2018) developed a path-weighted model to predict circRNA–disease associations based on circRNA semantic similarity and disease functional similarity (Lei et al., 2018). KATZHCDA was used to calculate the number of walks between nodes and walk lengths for circRNA–disease associations, based on a priori knowledge of the circRNA expression similarity and disease phenotype similarity (Fan et al., 2018b). DWNN-RLS predicted circRNA–disease associations using regularized least squares of the Kronecker product kernel (Yan et al., 2018). Xiao et al. (2019) proposed a weighted dual-manifold regularized low-rank approximation model for disease-related circRNA prediction, called MRLDC (Xiao et al., 2019). Another model, iCircDA-MF, incorporated circRNA–gene, gene–disease, and circRNA–disease associations, together with disease semantic information, and used non-negative matrix factorization to predict circRNA–disease associations (Wei and Liu, 2019). Zhao et al. (2019) integrated the bipartite network projection algorithm and KATZ measure algorithm to explore novel circRNA–disease associations (Zhang et al., 2019). Deng et al. (2019) combined circRNAs, proteins, and diseases to predict circRNA–disease associations using the KATZ algorithm (Deng et al., 2019). Ge et al. (2019) developed the LLCDC model for prediction of human disease-associated circRNAs using locality-constrained linear coding and a label propagation algorithm (Ge et al., 2019). CD-LNLP calculated circRNA similarity and disease similarity using linear neighborhood similarity based on known associations, and then used the label propagation algorithm to mine circRNA–disease associations (Zhang et al., 2019). Wang Y. et al. (2019) used a graph-based recommendation algorithm, PersonalRank, to predict disease-related circRNAs based on circRNA expression profiles and functional similarities (Wang Y. et al., 2019). Lei and Fang (2019) used a gradient boosting decision tree with multiple biological data fusion for circRNA–disease prediction (Lei and Fang, 2019). Ding et al. (2020) developed the RWLR model based on the random walk and the logistic regression to predict circRNA–disease associations. iCDA-CGR quantified the sequence nonlinear relationship of circRNA by chaos game representation technology based on the biological sequence position information (Zheng et al., 2020). Lei and Bian (2020) integrated the random walk with restart and k-nearest neighbors to predict the associations between circRNAs and diseases. Although these computational models have achieved encouraging results, they represent the tip of the iceberg with respect to predicting circRNA–disease associations.

Several circRNAs can bind with the corresponding miRNAs and participate in multiple biological processes synchronously (Qu et al., 2018). Based on this theory, Fang and Lei (2019) used an improved random walk algorithm to predict circRNA–miRNA associations, named KRWRMC (Fang and Lei, 2019). As miRNAs have been implicated in various diseases, we consider that miRNA information should be included in the identification of circRNA–disease associations. However, there have been few studies of circRNA–miRNA interactions, and deep interaction patterns are rarely considered in prediction of circRNA–disease associations. In this work, we take circRNA–miRNA interactions and miRNA–disease associations into account, and capture the complex miRNA-based interaction features of circRNAs and diseases, respectively.
In recent years, deep learning architectures have attracted increasing attention in various fields, including image analysis (Yang and Xu, 2020), speech recognition (Graves et al., 2013), and bioinformatics (Min et al., 2017), etc. The convolutional neural network (CNN) is a well-known feed-forward artificial neural network inspired by biological processes that simulates the cognition function of human neural systems (LeCun et al., 2015). CNN architectures have the ability to automatically learn the meaning of combinations of features from the input data and simplify the process of manual feature selection (Liu et al., 2017). Recent applications of CNN-based methods indicate their effectiveness in computational biology (Liu et al., 2018), including in circRNA research. Wang and Wang (2019) developed the DeepCircCode model to discover the sequence code of back-splicing for circRNA formation, and sequence motifs were also extracted. The CSCRSites model was proposed to predict cancer-specific protein binding sites on circRNAs based on CNNs. The features learned by the CSCRSites model are converted to sequence motifs, some of which are involved in human diseases (Wang Z. et al., 2019). Inspired by the superior prediction performance of this approach, we used CNN architecture to detect combinations of features and predict potential circRNA–disease associations.

In this study, we present a novel computational model to predict potential associations between circRNAs and diseases, named MSFCNN. The main attributes of the MSFCNN model are as follows. (1) Four circRNA similarity kernels and seven disease similarity kernels are constructed using multiple biological and topological information, such as circRNA expression profiles, circRNA sequence information, disease-miRNA interactions, etc. (2) Whereas some existing methods simply use linear weighting to integrate the similarity kernels into one kernel, we considered that this may lead to information loss and noise. Hence, we used the similarity kernel fusion (SKF) method to fuse four circRNA similarity kernels and seven disease similarity kernels, thereby retaining the original information of each similarity kernel. A weight matrix is used to reduce the noise in the fused similarity kernel. (3) A feature matrix is constructed based on the fused circRNA similarity kernel, fused disease similarity kernel, and interactions and features among circRNAs, miRNAs, and diseases. Multiple biological premises are used to construct the feature matrix. On the one hand, two circRNAs (or diseases) are more similar if their circRNA–disease associations are more similar. On the other hand, circRNA–miRNA and miRNA–disease associations are also integrated, and the interaction features are captured using principal component analysis. (4) A two-layer CNN architecture is used to process the feature matrix and predict potential circRNA–disease associations. Five-fold cross-validation (CV) is used to assess the prediction performance of the MSFCNN model. The results indicate that the MSFCNN model outperforms several conventional machine learning classifiers. Furthermore, case studies of breast cancer, colorectal cancer, hepatocellular carcinoma, and acute myeloid leukemia indicate that MSFCNN could be an effective tool to infer potential circRNA–disease associations.

MATERIALS AND METHODS

A flow chart illustrating MSFCNN, our novel approach to predict potential circRNA–disease associations is shown in Figure 1. First, four circRNA similarity kernels and seven disease similarity kernels are computed based on their biological and topological properties. Then, these kernel similarities are combined into one circRNA similarity kernel and one disease similarity kernel by applying a similarity kernel fusion strategy. Subsequently, the feature matrix can be constructed based on the fused similarity kernels, and interactions and features among circRNAs, miRNAs, and diseases. Finally, we use a CNN to process the feature matrix and predict final scores for prediction of potential circRNA–disease associations.

Construction of the CircRNA–Disease, CircRNA–miRNA, and Disease–miRNA Networks

In this study, circRNA–disease associations, circRNA–miRNA associations, and disease–miRNA associations were used to predict circRNA–disease associations. Known circRNA–disease associations were downloaded from the CircR2Disease database (Fan et al., 2018a), which contained 739 entries including 725 experimentally validated circRNA–disease associations from four species. Only human circRNA–disease associations were used in this work. Interactions that did not correspond to circRNAs IDs in the circBase database and disease names were not recorded in the disease ontology database were removed (Glazar et al., 2014; Schriml et al., 2019). Thus, we retained 325 circRNAs, 53 diseases, and 371 circRNA–disease associations as the positive dataset. The circRNA–miRNA interactions were obtained from the CircBank database (Liu et al., 2019), and interactions overlapping with disease-related circRNAs were extracted. Thus, 24745 interactions between 322 circRNAs and 2545 miRNAs were obtained. In addition, the disease–miRNA associations that matched circRNA-related diseases were selected from the human microRNA disease database (Huang et al., 2019), and 4970 associations between 37 diseases and 873 miRNAs were obtained. Finally, all of these associations contained three types of nodes including 325 circRNAs, 53 diseases, and 3175 miRNAs.

Based on the circRNA–disease associations, an adjacency matrix $A(i,j)$ was constructed to represent associations between $n_c$ circRNAs and $n_d$ diseases; $A(i,j)$ was assigned a value of 1 if circRNA $c(i)$ was found to be related to disease $d(j)$, and 0 otherwise. Similarly, a circRNA–miRNA matrix $Y(i,j)$ was constructed to represent the associations between $n_c$ circRNAs and $n_m$ miRNAs, and the associations between $n_d$ diseases and $n_m$ miRNAs were represented by matrix $O(i,j)$. $Y(i,j)$ was set to 1 when there was an association between circRNA $c(i)$ and miRNA $m(j)$, and 0 otherwise. If disease $d(i)$ interacted with miRNA $m(j)$, $O(i,j)$ was set to 1, otherwise it was set to 0.
FIGURE 1 | Flow chart of the MSFCNN approach. Step 1: Four circRNA similarity kernels and seven disease similarity kernels are measured, respectively. Step 2: The similarity kernels for circRNAs (or diseases) is fused with SKF method. Step 3: The feature matrix for each circRNA–disease pair is constructed by integrating the fused similarity kernels, interactions and features among circRNAs, miRNAs, and diseases. Step 4: a CNN architecture is used to train MSFCNN approach and predict latent circRNA–disease associations.

Representation of CircRNA Similarity Kernels

CircRNA Sequence Similarity
The 325 circRNA sequences were obtained from the circBase database (Glazar et al., 2014), and the sequence similarity of each circRNA–circRNA pair was calculated using a modification of the Needleman–Wunsch algorithm with the Emboss-stretcher tool (Rice et al., 2000). Therefore, the circRNA sequence similarity score \( SC_{Seq}(c_i, c_j) \) could be obtained by setting the parameters as follows: Matrix = EDNAFULL, Gap open = 16, Gap extend = 4.

CircRNA Regulatory Similarity
Based on the assumption that circRNAs associated with the same miRNAs tend to have similar biological regulatory functions, we used the miRNA–circRNA interactions to measure the circRNA regulatory similarity (Huang et al., 2018). Given the two sets of
miRNAs, $M_i$ and $M_j$, that had relationships with circRNAs $c_i$ and $c_j$, respectively, the circRNA regulatory similarity kernel was calculated as follows:

$$SC_{RG}(c_i, c_j) = \frac{\text{card}(M_i \cap M_j)}{\sqrt{\text{card}(M_i)} \cdot \sqrt{\text{card}(M_j)}}$$ (1)

**CircRNA Expression Similarity**

The circRNA expression profiles were derived from the exoRBase database (Li et al., 2018). Each circRNA record had 90 dimensions, representing the expression levels of a single type of circRNA. By extracting the common circRNAs between the CircR2Disease and exoRBase databases, circRNA expression profiles were obtained for calculation of the circRNA similarity kernel. We used the Pearson correlation coefficient to measure circRNA expression similarity, and let $SC_{EP}(c_i, c_j)$ represent the expression similarity score between circRNAs $c_i$ and $c_j$. The expression similarity kernel of the circRNAs was computed as follows:

$$SC_{EP}(c_i, c_j) = \frac{\sum_{i=1}^{N}(x_i - ȳ)(y_i - ȳ)}{\sqrt{\sum_{i=1}^{N}(x_i - ȳ)^2 \sum_{i=1}^{N}(y_i - ȳ)^2}}$$ (2)

where $N$ represents the number of properties of the expression profiles, and $x_i$ and $y_i$ denote the expression values in different tissues. In general, a pair of circRNAs with a higher correlation score are considered to be more similarly expressed.

**GIP Kernel Similarity for CircRNAs**

The Gaussian interaction profile (GIP) kernel similarity was used to measure the similarity between circRNAs, based on the assumption that similar circRNAs are more likely exhibit a similar interaction or non-interaction pattern with miRNAs (Van Laarhoven et al., 2011). GIP kernel similarity for circRNAs was measured based on circRNA–miRNA interactions and defined as:

$$SC_{GIP}(c_i, c_j) = \exp(-\gamma_c \parallel c(i) - c(j) \parallel^2)$$

$$\gamma_c = \frac{1}{\sum_{i=1}^{n_c} \parallel c(i) \parallel^2}$$ (3)

where the circRNA interaction profiles are represented by $c(i)$, a binary vector that encodes the interaction between circRNA $i$ and all miRNAs, i.e., the $i$-th row of the circRNA–miRNA interaction matrix $Y$. The parameter $\gamma_c$ controls the kernel bandwidth, and $n_c$ is the number of circRNAs.

**Representation of Disease Similarity Kernels**

**Disease Symptom Similarity**

According to the co-occurrence of disease and symptom terms recorded in the PubMed bibliography, Zhou et al. (2014) considered that diseases are connected if they have a positive symptom similarity (Zhou et al., 2014). Thus, the disease similarity could be measured and a symptom-based human disease network was constructed. Here, the symptom-based disease similarity $SD_{Sym}$ was obtained from the symptom profiles of diseases.

**Disease Semantic Similarity**

According to Medical Subject Headings descriptions, diseases can be described by a hierarchical directed acyclic graph (DAG). Here, disease semantic similarity is calculated using the method of Wang et al. (2007). DAG$_d$ = ($d$, $T_d$, $E_d$) represents the DAG of a disease, in which $T_d$ denotes node $d$ and its ancestor nodes, and $E_d$ denotes the direct edges from a parent node to child nodes within $T_d$. Therefore, the semantic contribution of parent node $t$ to $d$ is defined as follows:

$$D_d(t) = \begin{cases} 1, & \text{if } t = d \\ \max\{\Delta \ast D_d(d') \mid d' \in \text{children of } t, \text{ if } t \neq d \} \\ \end{cases}$$ (4)

where $\Delta$ represents the semantic contribution decay factor ($\Delta$ is set as 0.5). The semantic value of disease $d$ can be calculated as follows:

$$DV(d) = \sum_{t \in T_d} D_d(t)$$ (5)

If two diseases share a larger part of DAGs, they tend to have higher similarity. The similarity score between $d_i$ and $d_j$ is defined as:

$$SD_{DSS}(d_i, d_j) = \frac{\sum_{t \in T_d \cap T_j} (D_d(t) + D_d(t))}{DV(d_i) + DV(d_j)}$$ (6)

**GIP Kernel Similarity for Diseases**

Similar to the calculation of GIP kernel similarity for circRNAs, the disease GIP kernel similarity was measured based on disease–miRNA interaction profiles. It is defined as:

$$SD_{GIP}(d(i), d(j)) = \exp(-\gamma_d \parallel d(i) - d(j) \parallel^2)$$

$$\gamma_d = \frac{1}{n_d \sum_{i=1}^{n_d} \parallel d(i) \parallel^2}$$ (7)

where the disease interaction profiles are represented by $d(i)$, a binary vector that encodes the interaction between disease $i$ and each miRNA, i.e., the $i$-th row of association matrix $O$. The parameter $\gamma_d$ is also used to control the kernel bandwidth, and $n_d$ is the number of diseases.

**Other Disease Similarities**

Besides disease symptom similarity, disease semantic similarity, and GIP kernel similarity, disease similarities can also be measured using the Lin (1998), PSB (Mathur and Dinakarpandian, 2012), Resnik (1995), and SemFunSim (Cheng et al., 2014) methods based on the DincRNA database (Cheng et al., 2018). Four disease similarity kernels were constructed using these methods and denoted $SD_{Lin}$, $SD_{PSB}$, $SD_{Resnik}$, and $SD_{SemFunSim}$, respectively.

**Similarity Kernel Fusion**

Next, we used the similarity kernel fusion method to integrate four circRNA similarity kernels and seven disease similarity kernels (Jiang et al., 2018). Let $S_{c,m}$ ($m = 1, 2, \ldots, 4$) represent the four circRNA similarity kernels and $S_{d,n}$ ($n = 1, 2, \ldots, 7$) the seven disease similarity kernels, respectively.
First, each original similarity kernel for circRNAs was normalized using Eq. (8):

$$\text{NS}_{c,m}(c_i, c_j) = \frac{S_{c,m}(c_i, c_j)}{\sum_{c_k \in C} S_{c,m}(c_k, c_j)}$$  \hspace{1cm} (8)

where $\text{NS}_{c,m}$ denotes a normalized similarity kernel for circRNAs that satisfies $\sum_{c_k \in C} \text{NS}_{c,m}(c_k, c_j) = 1$.

Then, a sparse kernel for each circRNA similarity kernel was constructed using Eq. (9):

$$F_{c,m}(c_i, c_j) = \begin{cases} S_{c,m}(c_i, c_j) / \sum_{c_k \in N_i} S_{c,m}(c_i, c_k) & c_j \in N_i \\ 0 & c_j \notin N_i \end{cases}$$  \hspace{1cm} (9)

where $F_{c,m}$ is a sparse kernel satisfying $\sum_{c_j \in C} F_{c,m}(c_k, c_j) = 1$, and $N_i$ is a set of $c_i$'s neighbors including $c_i$ itself.

The four circRNA similarity kernels were computed using Eq. (10):

$$S_{c,r}^{t+1} = \alpha \left( F_{c,m} \times \frac{\sum_{r \neq 1} S_{c,r}^t}{2} \times F_{c,m}^T \right) + (1 - \alpha) \left( \frac{\sum_{r \neq 1} S_{c,r}^0}{2} \right) \quad \alpha \in (0, 1)$$  \hspace{1cm} (10)

where $S_{c,r}^{t+1}$ is the status matrix of $m$-th circRNA similarity kernel after $t+1$ iterations, and $S_{c,r}^0$ denotes the initial status of $S_{c,r}$.

After $t+1$ steps, the overall kernel for circRNAs was calculated using Eq. (11):

$$S_c = \frac{1}{4} \sum_{m=1}^4 S_{c,m}^{t+1}$$  \hspace{1cm} (11)

FIGURE 2 | Establishment of the feature matrix of circRNA $c_1$ and disease $d_2$. Based on three premises, $c_1$-$d_2$ feature matrix is constructed by combing fused similarities and associations among circRNAs, diseases and miRNAs.
Furthermore, a weight matrix $w_c$ was used to eliminate the noise in matrix $S_c$, and the fused circRNA similarity kernel was computed using Eq. (12):

$$S^*_c = w_c \circ S_c$$

Similarly, the seven disease similarity kernels were fused to form one disease similarity kernel, denoted by $S^*_d$.

Construction of the Feature Matrix

The feature matrix for each circRNA–disease pair was constructed by incorporating the fused circRNA similarity, fused disease similarity, circRNA–miRNA interactions, circRNA–disease associations, and disease–miRNA associations (Figure 2).

**Figure 3** Graphical illustration of the MSFCNN architecture. The feature matrix of circRNA $c_1$ and disease $d_2$ is input to the convolution neural network model to learn global deep representation between $c_1$ and $d_2$. 

$$w_c (c_i, c_j) = \begin{cases} 
1 & \text{if } c_i \in N_j \text{ and } c_j \in N_i \\
0 & \text{if } c_i \notin N_j \text{ and } c_j \notin N_i \\
0.5 & \text{otherwise}
\end{cases} (13)$$
In the construction process of the feature matrix, three biological premises were used. Here, we take the construction of the $c_1$–$d_2$ feature matrix as an example. Based on the premise that the circRNAs should be more similar that have interaction with circRNA similarities and circRNA–disease associations, the first part of the feature matrix consists of the similarity between $c_1$ and all circRNAs, and the associations of $d_2$ with all circRNAs. If circRNA $c_1$ and $c_2$ or other circRNAs have similar functions, and at the same time $d_2$ has been shown to be associated with these circRNAs, $c_1$ has a large probability associated with $d_2$. The dimension of the first part of the feature matrix is $2 \times n_c$. Similarly, based on the premise that diseases should be more similar that have interaction with disease similarities and circRNA–disease associations, we integrate the associations between circRNA $c_1$ and all diseases, as well as the similarities between disease $d_2$ and all diseases. The second part of the feature matrix has dimension $2 \times n_d$. In addition, circRNA–miRNA and miRNA–disease is integrated to capture the relation features. When $c_1$ and $d_2$ have interactions with common miRNAs, they are more likely to be associated with each other. The interactions between $c_1$ and various miRNAs, as well as the associations between $d_2$ and miRNAs, are integrated to construct a matrix with dimension $2 \times n_m$. However, the matrix is very sparse, so we perform principal component analysis (PCA) to obtain miRNA-based features for the $c_1$–$d_2$ pair with dimension $2 \times n_p$ ($n_p$ is set as 50). Finally, we concatenate these three matrices to form the feature matrix of circRNA $c_1$ and disease $d_2$ with dimension $2 \times (n_c+n_d+n_p)$.

**Identification of CircRNA–Disease Associations Based on CNN**

The MSFCNN architecture consists of an input layer, two convolutions, and an activation layer, polling layer, fully connected layer, and softmax layer (Figure 3). The feature matrix $X$ of node pairs is used as an input to the CNN architecture to learn the representations of node-pair circRNAs and diseases. The MSFCNN can be summarized as:

$$\text{Out} = \text{Softmax} \circ \text{Fully Connected} \circ \text{Global Max Pool} \circ \text{Conv2D} \circ \text{ReLU} (X)$$

(14)

where $X$ is the feature matrix that is fed to the two-dimensional convolution (Conv2D) layer. In the first convolutional layer, if the number of filters is $n_{conv1}$, the width of the kernel is $n_w$, and its length is set as $n_l$. The convolution filters are indicated as $W_{conv1} \in \mathbb{R}^{n_{conv1} \times n_w \times n_l}$, and the feature maps are $Z_{conv1} \in \mathbb{R}^{n_{conv1} \times (2-nw+1) \times (nc+n_d+n_p-nl+1)}$. The convolution process can be described as follows:

$$X_{k,i,j} = X(i + n_w, j + n_l) \quad X_{k,i,j} \in \mathbb{R}^{n_w \times n_l}$$

(15)

$$Z_{conv1,k}(i,j) = g(W_{conv1}(k, :, :) \ast X_{conv1,i,j} + b_{conv1}(k)) \quad k \in [1, n_{conv1}], i \in [1, 2], j \in [1, n_c + n_d + n_p - n_l + 1],$$

(16)

where $X(i,j)$ is the element of matrix $X$ in the $i$-th row and $j$-th column, and $X_{k,i,j}$ represents the region in the filter where the $k$-th filter slides to the position $X(i,j)$. $g$ is a rectified linear units (relu) function (Nair and Hinton, 2010), $b_{conv1}$ is the bias vector, $*$ represents the convolution operation, and $Z_{conv1,k}(i,j)$ represents the convolution result of the $k$-th filter sliding to the $j$-th column of the $i$-th row.

Similarly, the second Conv2D layer is also used to learn the higher-level features. To compress data and reduce over-fitting, the polling layer is used to obtain robust features. Here, the max-pooling operation is employed for each feature map (Collobert et al., 2011). Then, the outputs of the pooling layer are concatenated together from all kernels into one feature vector and input into the fully connected layer. The nonlinear softmax activation function is used to perform the task of classification.

To avoid over-fitting, a dropout layer is implemented before the output, in which the output of every neuron is set to zero with a probability of 0.5. The dropped-out neurons are not included in the forward pass or the back-propagation (Hinton et al., 2012).

**Prediction of Novel CircRNA–Disease Associations**

Next, we used all the positive and negative circRNA–disease association samples to train the MSFCNN architecture. Then, MSFCNN was used to score the unlabeled associations between circRNAs and diseases. Owing to the different negative samples used to train the model in each iteration of the five-fold cross validation (five-fold CV), we scored the candidate associations 10 times. Finally, we calculated the average scores for the candidate associations, and the candidate circRNAs related to specific diseases were analyzed using case studies.

**RESULTS**

**Performance Evaluation**

The performance of MSFCNN and other conventional machine learning-based methods for predicting circRNA–disease associations was evaluated using five-fold CV. If the circRNA $c(i)$ was found to be related to disease $d(j)$, the node pair $c_i$-$d_j$ was considered as a positive example. Hence, the validated circRNA–disease associations were regarded as the positive set. However, because of the unavailability of a dataset for negative samples, we randomly selected a negative set from unobserved associations that was the same size as the positive set. All the positive samples were divided into five subsets of equal size, and each subset was tested once. For each CV, we took four positive subsets and the same number of negative subsets from five subsets to train the models; the remaining one positive subset and one negative subset were used for testing to evaluate the prediction performance. To lessen the bias resulting from sample division, we performed 10 repetitions of five-fold CV and obtained the average values of five experiments.

Receiver operating characteristic (ROC) curves were plotted to show the prediction performance by calculating the true positive rate and false positive rate. The area under the curve (AUC) was calculated to evaluate the overall performance. In addition, five metrics, precision ($\text{Prec}$), sensitivity ($\text{Sen}$), accuracy ($\text{Acc}$), $F_1$-score, and Matthews’ correlation coefficient (MCC) were used.
to evaluate the capability of the MSFCNN model. The detailed calculation of these metrics was as follows:

$$Pre = \frac{TP}{TP + FP}$$  \hspace{1cm} (17)

$$Sen = \frac{TP}{TP + FN}$$  \hspace{1cm} (18)

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$  \hspace{1cm} (19)

$$F1 - score = \frac{2 \times Sen \times Pre}{Sen + Pre}$$  \hspace{1cm} (20)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN) \times (TP + FP) \times (TN + FN) \times (TN + FP)}}$$  \hspace{1cm} (21)

where $TP$ and $TN$ represent the number of true positives and true negatives, respectively, and $FP$ and $FN$ represent the number of positives and negatives, respectively, that were wrongly predicted.

**Parameter Setting**

Convergence and parameter selection are important factors in the SKF method, that is, the number of iterations and two parameters, $\alpha$ and the size of neighbors. Following a previous study (Jiang et al., 2018), we set these two parameters to 0.1 and 36, respectively. As the number of iterations is important for the convergence of the SKF method, we also analyzed whether the number of iterations was sufficient for convergence in the four circRNA similarity kernels and seven disease similarity kernels. The relative error of the process of iteration was denoted $EC_t$ and $ED_t$ for circRNA similarity fusion and disease similarity fusion, respectively. The number of iterations ranged from 1 to 25 with steps of 1, and $EC_t$ and $ED_t$ were computed after every iteration. The convergence processes of the four circRNA similarity kernels and seven disease similarity kernels are shown in Figure 4. The results indicate that the convergence process was fast, and the $EC_t$ and $ED_t$ values reached $10^{-10}$ after 10 iterations. Therefore, we set the number of iterations to 10 for both circRNA similarity fusion and disease similarity fusion.

$$EC_t = \frac{\| SC^{t+1}_{c,m} - SC^t_{c,m} \|}{\| SC^t_{c,m} \|}$$  \hspace{1cm} (22)

$$ED_t = \frac{\| SD^{t+1}_{d,n} - SD^t_{d,n} \|}{\| SD^t_{d,n} \|}$$  \hspace{1cm} (23)

In the convolution operation of the MSFCNN model, the number of filters was set to 8. The kernel size was set to $2 \times 32$ in the first convolutional layer and $1 \times 16$ in the second convolutional layer. We implemented the MSFCNN model using the Keras 2.2.4 library in Python 3.7.3.

**Evaluation of Prediction Performance**

To assess the performance of the MSFCNN model for prediction of circRNA–disease associations, we used five-fold CV with 10 experiments (see Table 1 and Figure 5 for details). MSFCNN achieved average precision, sensitivity, $F1$-score, $Acc$, $MCC$, and AUC values of 0.9030, 0.9464, 0.9240, 0.9220, 0.8452, and 0.9525, with standard deviations of 0.0360, 0.0256, 0.0292, 0.0305, 0.0605, and 0.0202, respectively. Furthermore, the ROC curves for the MSFCNN model were at the upper left of the picture. These results indicate that our proposed model performs well in prediction of circRNA–disease associations.

**Comparison With Average Kernel Fusion Strategy**

In the MSFCNN model, the SKF method is used to fuse the four circRNA similarity kernels and seven disease similarity kernels into one circRNA similarity kernel and one disease similarity kernel, respectively. We compared the performance of the SKF method when integrating several similarity kernels with that of an average kernel fusion strategy. The average fusion strategy calculated the average similarity scores for four circRNA similarity matrices or seven disease similarity matrices, respectively. Five-fold CV was performed 10 times for predicting

![FIGURE 4 | Relative errors of the SKF method with various numbers of iteration for the four circRNA similarity matrices and seven disease similarity matrices.](image-url)
TABLE 1 | Evaluation metrics for performance of the MSFCNN approach.

| Times | Pre   | Sen   | F1-score | Acc   | MCC   |
|-------|-------|-------|----------|-------|-------|
| 1     | 0.9573| 0.9677| 0.9625   | 0.9623| 0.9246|
| 2     | 0.8488| 0.9380| 0.8912   | 0.8854| 0.7752|
| 3     | 0.9251| 0.9326| 0.9289   | 0.9286| 0.8572|
| 4     | 0.8660| 0.9057| 0.8854   | 0.8827| 0.7663|
| 5     | 0.9203| 0.9650| 0.9421   | 0.9407| 0.8824|
| 6     | 0.9010| 0.9568| 0.9281   | 0.9259| 0.8534|
| 7     | 0.9258| 0.9757| 0.9501   | 0.9488| 0.8989|
| 8     | 0.8641| 0.9084| 0.8857   | 0.8827| 0.7665|
| 9     | 0.8835| 0.9407| 0.9112   | 0.9084| 0.8184|
| 10    | 0.9377| 0.9730| 0.9950   | 0.9642| 0.9090|
| Average| 0.9030| 0.9464| 0.9240   | 0.9220| 0.8452|

FIGURE 5 | ROC curves of MSFCNN model for the task of circRNA–disease association prediction.

circRNA–disease associations. The average kernel fusion-based MSFCNN model had an average AUC of 0.8628 (Figure 6); by comparison, the SKF-based MSFCNN model had an AUC of 0.9525 (an improvement of 0.0897). Other evaluation metrics also indicated that the SKF method performs better than the average kernel fusion strategy in MSFCNN (Table 2). Hence, the SKF method is an effective fusion strategy for prediction of circRNA–disease associations.

Comparison With Conventional Machine Learning Approaches

To demonstrate the reliability and robustness of the MSFCNN method, we made comparisons with state-of-the-art machine learning approaches: support vector machine (SVM), random forest (RF), and multilayer perception (MLP). For each of these machine learning approach, the feature matrix fed into the model was consistent with that used for MSFCNN to ensure the fairness of the experiments. As shown in Figure 7, the average AUC of the MSFCNN model in the five-fold CV was 0.9179 higher than those of the SVM, RF, and MLP methods. In addition, MSFCNN achieved higher precision, sensitivity, F1-score, Acc, and MCC values than the other machine learning approaches (Table 3). Therefore, the proposed method is more suitable than these conventional approaches for the task of circRNA–disease association prediction.
TABLE 3 | Evaluation metrics for performance of the MSFCNN and other machine learning methods.

| Methods  | Pre  | Sen  | F1-score | Acc  | MCC  |
|----------|------|------|----------|------|------|
| MSFCNN   | 0.8468 | 0.8491 | 0.8479   | 0.8477 | 0.6954 |
| SVM      | 0.6166 | 0.6415 | 0.6288   | 0.6213 | 0.2428 |
| RF       | 0.6851 | 0.5337 | 0.6000   | 0.6442 | 0.2967 |
| MLP      | 0.6455 | 0.6577 | 0.6515   | 0.6482 | 0.2965 |

TABLE 4 | Candidate circRNAs predicted by the MSFCNN model for four diseases.

| Diseases                    | circRNAs               | Rank | Evidence   |
|-----------------------------|------------------------|------|------------|
| Acute myeloid leukemia      | hsa_circ_0000677       | 3    | Circ2Traits|
|                             | hsa_circ_0000175       | 6    | Circ2Traits|
| Breast cancer               | hsa_circ_0000677       | 8    | Circ2Traits|
|                             | hsa_circ_0000175       | 11   | Circ2Traits|
|                             | hsa_circ_0001417       | 25   | Circ2Traits|
| Colorectal cancer           | hsa_circ_0001417       | 16   | Circ2Traits|
|                             | hsa_circ_0000175       | 19   | Circ2Traits|
|                             | hsa_circ_0001283       | 40   | Circ2Traits|
|                             | hsa_circ_0000615       | 56   | Circ2Traits|
| Hepatocellular              | hsa_circ_0000677       | 10   | Circ2Traits|
|                             | hsa_circ_0001417       | 24   | Circ2Traits|
|                             | hsa_circ_0001283       | 48   | Circ2Traits|

FIGURE 8 | Top 20 predicted circRNA–disease associations.

Case Study
To further demonstrate the ability of the MSFCNN model to discover potential circRNA–disease associations, we scored unlabeled associations between circRNAs and diseases using the trained model. Average scores were obtained from 10 applications of the MSFCNN model, and candidate circRNA–disease associations were identified based on their ranked scores. Case studies were performed for breast cancer, colorectal cancer, hepatocellular carcinoma, and acute myeloid leukemia. Some of the predicted specific disease-related circRNAs were found in the Circ2Traits database (Ghosal et al., 2013), which collects circRNAs and miRNAs related to diseases and traits (Table 4). In addition, we plotted the top 20 predicted circRNA–disease associations; the results show that these circRNAs may be related to the same diseases, and the diseases may also be associated with the same circRNAs (Figure 8). Hence, these results show that the MSFCNN model could be an effective tool for the prediction of circRNA–disease associations.

CONCLUSION
Prioritizing potential disease-related circRNAs based on various types of prior information is beneficial to understanding disease mechanisms, diagnosis, and treatment. In this study, we developed a novel computational method named MSFCNN to predict potential circRNA–disease associations, using a two-layer two-dimensional CNN and integrating multiple biological data. First, one of the crucial technical points for predicting circRNA–disease associations is the similarity calculation for circRNA–circRNA and disease–disease pairs. Therefore, we calculated four circRNA similarity kernels and seven disease similarity kernels based on multiple biological and topological information. In addition, similarity kernel fusion was used to integrate various similarity kernels into one circRNA similarity kernel and one disease similarity kernel. Based on these fused similarity kernels and interactions/features among circRNAs, miRNA, and diseases, a feature matrix was constructed for each circRNA–disease pair. Finally, a two-layer CNN architecture was used to predict circRNA–disease associations. The MSFCNN approach showed good performance based on the five-fold CV, outperforming the SVM, RF, and MLP classifiers. Furthermore, case studies of breast cancer, colorectal cancer, hepatocellular carcinoma, and acute myeloid leukemia demonstrated that the MSFCNN framework could be an effective tool for successfully inferring potential circRNA–disease associations and providing a basis for biological validation.

The good performance of MSFCNN method mainly conclude following aspects. Firstly, multiple similarity kernels for circRNAs and diseases are effectively introduced to measure the biological and topological features of circRNAs and diseases. Secondly, the relationships of circRNA–miRNA and disease–miRNA are also used to construct the feature matrix for each circRNA–disease pair. Furthermore, the application of CNN architecture guarantees the effectiveness of learning the meaning of combinations of features from the feature matrix. Hence, MSFCNN method is an effective biomedical resource to predict the circRNA–disease associations.

Despite its promising prediction performance, the MSFCNN approach has some limitations. First, incomplete and noisy circRNA–disease associations were used as positive samples, and negative samples are randomly selected, limiting the prediction performance. This could be improved as more associations are discovered. Furthermore, more reliable biological information should be considered, such as circRNA coding potential and circRNA functional information, as well as disease phenotypes and functional information, etc. In addition, optional similarity measurements would be integrated based on comparing the prediction results of different similarity measures. Therefore,
more data sources should be collected, and a more effective model needs to be developed to address the above limitations.

**DATA AVAILABILITY STATEMENT**

Publicly available datasets were analyzed in this study. This data can be found here: http://bioinfo.snnu.edu.cn/CircR2Disease/, http://www.circbank.cn/, https://disease-ontology.org/, http://bioanotation.cn:18080/DircRNAClient/#/Home, https://www.nlm.nih.gov/mesh/, http://www.cuilab.cn/hmdd, and http://www.exorbase.org.

**REFERENCES**

Bahn, J. H., Zhang, Q., Li, F., Chan, T.-M., Lin, X., Kim, Y., et al. (2015). The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem. 61*, 221–230. doi: 10.1373/clinchem.2014.230433

Barrett, S. P., Wang, P. L., and Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *eLife* 4:e07540. doi: 10.7554/eLife.07540

Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018). DircRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* 34, 1953–1956. doi: 10.1093/bioinformatics/bty002

Cheng, L., Li, J., Ju, P., Peng, J., and Wang, Y. (2014). SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association. *PLoS One* 9:e99415. doi: 10.1371/journal.pone.0099415

Chioccarelli, T., Manfrevola, F., Ferraro, B., Sellitto, C., Cobelli, G., Migliaccio, M., et al. (2012). Expression patterns of circular RNAs in high quality and poor quality human spermatozoa. *Front. Endocrinol.* 10:435. doi: 10.3389/fendo.2019.00435

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12, 2493–2537.

Conn, V. M., Hugouvieux, V., Nayak, A., Conos, S. A., Capovilla, G., Cildir, G., et al. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *eLife* 4:e07540. doi: 10.7554/eLife.07540

**AUTHOR CONTRIBUTIONS**

XL and YP conceptualized the study. CF and XL performed the data collection, designed the method, and drafted the manuscript. All authors read and approved the final version of the manuscript.

**FUNDING**

This work was supported by the National Natural Science Foundation of China (61972451, 61672334, and 61902230) and the Fundamental Research Funds for the Central Universities, Shaanxi Normal University (GK201901010).

Glazar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. RNA 20, 1666–1670. doi: 10.1261/rna.043687.113

Graves, A., Mohamed, A.-R., and Hinton, G. (2013). “Speech recognition with deep recurrent neural networks,” in Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, 6645–6649.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv [Preprint] Available online at: arXiv.org*. 1207.0580 (accessed July 3, 2012).

Huang, Y.-A., Chan, K. C. C., and You, Z.-H. (2018). Constructing prediction models from expression profiles for large scale IncRNA-miRNA interaction profiling. *Bioinformatics* 34, 812–819. doi: 10.1093/bioinformatics/btx672

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010

Ji, P., Wu, W., Chen, S., Zheng, Y., Zhou, L., Zhang, J., et al. (2019). Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep.* 26, 3444–3460.e5. doi: 10.1016/j.celrep.2019.02.078

Jiang, L., Ding, Y., Tang, J., and Guo, F. (2018). MDA-SKF similarity kernel fusion for accurately discovering miRNA-disease association. *Front. Genet.* 9:618. doi: 10.3389/fgene.2018.00618

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.

Lei, X., and Bian, C. (2020). Integrating random walk with restart and k-Nearest Neighbor to identify novel circRNA-disease association. *Sci. Rep.* 10:1943.

Lei, X., and Fang, Z. (2019). GBTDCA: predicting circRNA-disease associations based on gradient boosting decision tree with multiple biological data fusion. *Int. J. Biol. Sci.* 15, 2911–2924. doi: 10.7150/ijbs.33806

Lei, X., Fang, Z., Chen, L., and Wu, F. X. (2018). PWCDCA: path weighted method for predicting circRNA-disease associations. *Int. J. Mol. Sci.* 19:3410. doi: 10.3390/ijms19113410

Li, S., Li, Y., Chen, B., Zhao, J., Yu, S., Tang, Y., et al. (2018). exorBase: a database of circRNA, lncRNA and miRNA in human blood exosomes. *Nucleic Acids Res.* 46, D106–D112. doi: 10.1093/nar/gks891

Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., et al. (2015). Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 25, 981–984. doi: 10.1261/rna.043687.113

Lin, D. (1998). “An information-theoretic definition of similarity,” in Proceedings of the Fifteenth International Conference on Machine Learning, Manitoba, 296–304.

Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., et al. (2018). Applications of deep learning to MRI images: a survey. *Big Data Min. Anal.* 1, 1–18. doi: 10.26599/bdma.2018.9020001

Liu, M., Wang, Q., Shen, J., Yang, B. B., and Ding, X. (2019). Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.* 16, 899–905. doi: 10.1080/15476286.2019.1600395

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing* 234, 11–26. doi: 10.1016/j.neucom.2016.12.038

Mathur, S., and Dinakarpandian, D. (2012). Finding disease similarity based on implicit semantic similarity. *J. Biomed. Inform.* 45, 363–371. doi: 10.1016/j.jbi.2011.11.017
Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature12198

Memczak, S., Papavasileiou, P., Peters, O., and Rajewsky, N. (2015). Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* 10:e0141214. doi: 10.1371/journal.pone.0141214

Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869. doi: 10.1093/bib/bbw068

Nair, V., and Hinton, G. E. (2010). “Rectified linear units improve restricted boltzmann machines,” in Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, 807–814.

Qu, S., Liu, Z., Yang, X., Zhou, J., Yu, H., Zhang, R., et al. (2018). The emerging functions and roles of circular RNAs in cancer. *Cancer Lett.* 414, 301–309. doi: 10.1016/j.canlet.2017.11.022

Renkly, P. (1995). “Using information content to evaluate semantic similarity in a taxonomy,” in Proceedings of the 14th International Joint Conference on Artificial Intelligence, Adelaide, 448–453.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/s0168-9525(00)02024-2

Ruan, H., Xiang, Y., Ko, J., Li, S., Jing, Y., Zhu, X., et al. (2019). Comprehensive characterization of circular RNAs in 1000 human cancer cell lines. *Genome Med.* 11:55.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Schriml, L. M., Mitra, E., Munro, J., Tauber, B., Schor, M., Nickle, L., et al. (2019). Human disease ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* 47, D955–D962. doi: 10.1093/nar/gky1032

Sun, P., and Li, G. (2019). CircCode: a powerful tool for identifying circRNA coding ability. *Front. Genet.* 10:981. doi: 10.3389/fgen.2019.00981

Tang, B., Hao, Z., Zhu, Y., Zhang, H., and Li, G. (2018). Genome-wide identification and functional analysis of circular RNAs in *Zea mays*. *PLoS One* 13:e0202375. doi: 10.1371/journal.pone.0202375

Van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500

Vo, J. N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., et al. (2019). The landscape of circular RNA in cancer. *Cell* 176, 869–881.e13. doi: 10.1016/j.cell.2018.12.021

Wang, J., and Wang, L. (2019). Deep learning of the back-splicing code for circular RNA formation. *Bioinformatics* 35, 5235–5242. doi: 10.1093/bioinformatics/btz382

Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., and Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23, 1274–1281. doi: 10.1093/bioinformatics/btm087

Wang, Y., Nie, C., Zang, T., and Wang, Y. (2019). Predicting circRNA-disease associations based on circRNA expression similarity and functional similarity. *Front. Genet.* 10:832. doi: 10.3389/fgen.2019.00832

Wang, Z., Lei, X., and Wu, F.-X. (2019). Identifying cancer-specific circRNA-RBP binding sites based on deep learning. *Molecules* 24:4035. doi: 10.3390/molecules24224035

Wei, H., and Liu, B. (2019). iCircDA-MF: identification of circRNA-disease associations based on matrix factorization. *Brief. Bioinform.* 21, 1356–1367. doi: 10.1093/bib/bbz057

Wu, W., Ji, P., and Zhao, F. (2020). CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* 21:101. doi: 10.1186/s13059-020-02018-y

Xiao, Q., Luo, J., and Dai, J. (2019). Computational prediction of human disease-associated circRNAs based on manifold regularization learning framework. *IEEE J. Biomed. Health Inform.* 23, 2661–2669. doi: 10.1109/jbi.2019.2891779

Yan, C., Wang, J., and Wu, F. X. (2018). DWNN-RLS: regularized least squares method for predicting circRNA-disease associations. *BMC Bioinformatics* 19(Suppl. 19):520. doi: 10.1186/s12859-018-2522-6

Yang, M., and Xu, S. (2020). A novel patch-based nonlinear matrix completion algorithm for image analysis through convolutional neural network. *Neurocomputing* 389, 56–82. doi: 10.1016/j.neucom.2020.01.037

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res.* 27, 626–641. doi: 10.1038/cr.2017.31

Zhang, W., Yu, C., Wang, X., and Liu, F. (2019). Predicting CircRNA-disease associations through linear neighborhood label propagation method. *IEEE Access* 7, 83474–83483. doi: 10.1109/access.2019.2920942

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., et al. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806. doi: 10.1016/j.molcel.2013.08.017

Zhao, Q., Yang, Y., Ren, G., Ge, E., and Fan, C. (2019). Integrating bipartite network projection and KATZ measure to identify novel CircRNA-disease associations. *IEEE Trans. Nanobiosci.* 18, 578–584. doi: 10.1109/tnb.2019.2922214

Zheng, K., You, Z. H., Li, J. Q., Wang, L., Guo, Z. H., and Huang, Y. A. (2020). iCDA-CGR: identification of circRNA-disease associations based on chaos game representation. *PLoS Comput. Biol.* 16:e1007872. doi: 10.1371/journal.pcbi.1007872

Zhou, X., Menche, J., Barabasi, A.-L., and Sharma, A. (2014). Human symptoms-disease network. *Nat. Commun.* 5:4212. doi: 10.1038/ncomms5212

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Fan, Lei and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*