DAE-CFR: detecting microRNA-disease associations using deep autoencoder and combined feature representation

Yanling Liu[1,2], Ruiyan Zhang[1], Xiaojing Dong[1], Hong Yang[1], Jing Li[1], Hongyan Cao[1], Jing Tian[3*] and Yanbo Zhang[1,4,5*]

*Correspondence:
1105551933@qq.com;
sxmuzyb@126.com

[1] Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, China
[2] Department of Mathematics, Changzhi Medical College, Changzhi, China
[3] Department of Cardiology, First Hospital of Shanxi Medical University, Taiyuan, China
[4] Shanxi Provincial Key Laboratory of Major Diseases Risk Assessment, Taiyuan, China
[5] School of Health and Service Management, Shanxi University of Chinese Medicine, Jinzhong, China

Abstract

Background: MicroRNA (miRNA) has been shown to play a key role in the occurrence and progression of diseases, making uncovering miRNA-disease associations vital for disease prevention and therapy. However, traditional laboratory methods for detecting these associations are slow, strenuous, expensive, and uncertain. Although numerous advanced algorithms have emerged, it is still a challenge to develop more effective methods to explore underlying miRNA-disease associations.

Results: In the study, we designed a novel approach on the basis of deep autoencoder and combined feature representation (DAE-CFR) to predict possible miRNA-disease associations. We began by creating integrated similarity matrices of miRNAs and diseases, performing a logistic function transformation, balancing positive and negative samples with k-means clustering, and constructing training samples. Then, deep autoencoder was used to extract low-dimensional feature from two kinds of feature representations for miRNAs and diseases, namely, original association information-based and similarity information-based. Next, we combined the resulting features for each miRNA-disease pair and used a logistic regression (LR) classifier to infer all unknown miRNA-disease interactions. Under five and tenfold cross-validation (CV) frameworks, DAE-CFR not only outperformed six popular algorithms and nine classifiers, but also demonstrated superior performance on an additional dataset. Furthermore, case studies on three diseases (myocardial infarction, hypertension and stroke) confirmed the validity of DAE-CFR in practice.

Conclusions: DAE-CFR achieved outstanding performance in predicting miRNA-disease associations and can provide evidence to inform biological experiments and clinical therapy.

Keywords: MiRNA-disease associations, Integrated similarity, Logistic function transformation, Deep autoencoder, Combined feature representation

Liu *et al. BMC Bioinformatics*   (2024) 25:139

Page 2 of 19

Background

MiRNAs are endogenous 22 nucleotide-long noncoding RNA strands that are widely found in plants, viruses, animals and humans [1, 2]. They manipulate gene expression by base pairing with partially complementary mRNA [3, 4]. Mounting evidence points to miRNA's vital role in various bioprocesses, such as immune response [5], cell proliferation [6], tumor invasion [7], and metabolism [8]. Predicting novel miRNA-disease associations can aid understanding of complex disease mechanisms, which in turn can help to prevent, diagnose, and treat diseases [9, 10]. Additionally, understanding the role of miRNA on disease pathogenesis may contribute to the development of personalized medicines [11, 12], and advance medical progress overall. Given its significance, the identification of latent miRNA-disease interactions has become a prevalent area of academic research [13].

MiRNAs whose expression is associated with various diseases have been investigated using biological experimental methods, such as quantitative reverse transcription, microarray analysis, and deep sequencing [14]. However, biological experiments are slow, laborious, and costly, with uncertain outcomes. Using a large amount of laboratory-generated data, researchers have constructed many computing models to infer underlying miRNA-disease interactions. Of those developed thus far, these computing models can be basically summarized into two types: network-based and machine learning-based approaches. Network-based approaches are mainly based on the biological hypothesis that functionally similar miRNAs prefer to associate with phenotypically similar diseases and vice versa [15]. Chen et al. [16] presented a model named RWRMDA that used restart random walk to forecast miRNA-disease interactions. The authors applied global network similarity measurements for the first time and implemented a random walk on functional similarity network of miRNA. Gu et al. [17] designed a network consistent projection method (NCPMDA) to infer possible miRNA-disease pairs using miRNA-disease association network, miRNA similarity network and disease similarity network. Qu et al. [11] developed the KATZMDA model in which the KATZ algorithm was applied to a heterogeneous network composed of the association network and integrated similarity networks. Dai et al. [18] proposed LWBRW to infer the potential miRNA-disease interactions, a model that operated a logistic function transformation on the similarity networks and then applied bi-random walks on the miRNA and disease network. Ha [19] introduced SMAP, an efficient computational strategy for identifying miRNA-disease pairs. This approach utilized existing miRNA-disease associations to construct the matrix factorization model, incorporating comprehensive similarity measures for both miRNAs and diseases.

Network-based algorithms can mine the unknown miRNA-disease relationships by extracting topological information from association and similarity networks. While they have been proven to be effective for analyzing small-scale data, their computational complexity increases substantially as the network scale expands [20]. Therefore, it can be challenging to apply network algorithms to large-scale data, such as those involving miRNA-disease associations, which can comprise a considerable number of nodes and edges. Additionally, it is difficult to acquire an accurate prediction of the overall associations landscape because known associations are sparse in the network and limit the spread of information [21].

Machine learning-based algorithms usually use known miRNA-disease associations as positive samples, randomly select some unknown associations as negative samples, and then predict the unknown associations using training classifiers. For example, Chen et al. [22] developed the RFMDA algorithm to infer miRNA-disease interactions, which reduced dimension of sample space using a filter-based approach, and finally employed the random forest (RF) classifier for training. In another model, Zhao et al. [23] used *k*-means clustering to solve sample imbalance problem in data processing and then proposed the ABMDA model based on the Adaptive Boosting (AdaBoost) algorithm to forecast miRNA-disease interactions, which enhanced the classification accuracy. Zhou et al. [24] proposed GBDT-LR, which balanced the positive and negative samples by using *k*-means clustering, then extracted the novel features using the Gradient Boosting Decision Tree (GBDT) method, and finally used the Logistic Regression (LR) classifier to infer the scores of each miRNA-disease pair. In recent years, deep learning technology has developed rapidly in the field of bioinformatics. Liu et al. [25] presented the DFELMDA method. The authors proposed a novel feature representation strategy and then employed deep autoencoder for low-dimensional feature extraction for each miRNA and disease. Finally, the model used RF classifier to predict novel miRNA-disease pairs. Chen et al. [26] developed the DBNMDA model using deep-belief network (DBN) to infer miRNA-disease associations, which contained two parts: pre-training restricted Boltzmann machines, and fine-tuning DBN. Ha et al. [27] presented a novel approach called NCMD for predicting miRNA-disease associations. This method utilized node2vec to create low-dimensional vector representations of miRNAs and diseases. It then incorporated a deep learning framework that combined the linearity of generalized matrix factorization with the nonlinearity of a multilayer perceptron (MLP). Although these machine learning algorithms performed well, there are still several drawbacks. For instance, in previous studies, selecting negative samples was a problem; acquiring the appropriate feature representation of each miRNA-disease pair for model prediction is challenging.

Building on previous studies, we introduce more effective biological information, consider the problem of sample imbalance, and adopt a suitable feature representation strategy to enhance the model prediction ability. For the manuscript, we built a deep learning framework using Deep AutoEncoder and Combined Feature Representation (DAE-CFR) to identify hidden miRNA-disease associations. First, according to the known miRNA-disease associations, we computed the integrated similarity of miRNAs and diseases by employing the Gaussian interaction profile (GIP) kernel similarity and functional similarity for miRNAs, and GIP kernel similarity along with two kinds of semantic similarity for diseases. After applying the logistic function transformation to the two integrated similarity matrices, we addressed sample imbalance using *k*-means clustering and subsequently constructed training samples. Next, the deep autoencoder method was used to extract latent features for miRNAs and diseases, considering two types of features: the original association feature and similarity feature. Finally, we combined these latent features to form the feature representation for each miRNA-disease pair, which were then fed into the LR classifier to predict unknown associations, with model performance measured using five and tenfold cross-validation (CV). We compared DAE-CFR with six popular algorithms and nine classifiers, and the experimental results showed that

Liu *et al. BMC Bioinformatics*   (2024) 25:139

Page 4 of 19

DAE-CFR performed excellently. Additionally, our model was validated on another dataset to ensure its robustness. Case studies on myocardial infarction, hypertension and stroke further illustrated its effectiveness and practicability. The final pathway analysis confirmed its capability to identify disease-associated miRNAs, providing insights into their roles in diseases.

Materials and methods

Known human miRNA-disease associations

The known associations between miRNA and disease were downloaded from the HMDD v2.0 database [28], in which the associations have been experimentally validated. After data cleaning, 495 miRNAs, 383 diseases, and 5430 associations were obtained, as reported in the reference [25] (see Additional file 1). Let n_m and n_d denote the quantity of miRNAs and diseases, respectively; $A = \{a_{ij}\}_{n_m \times n_d}$ represents the association matrix, where a_{ij} is described below:

$$a_{ij} = \begin{cases} 1, & \text{miRNA } m_i \text{ is associated with disease } d_j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

GIP kernel similarity of miRNAs and diseases

The GIP kernel similarity is a widely used metric in the biomedical field [29]. The association matrix has been used to calculate the GIP kernel similarity [30]. We calculated the GIP kernel similarity between miRNAs m_i and m_j using the following formulae:

$$KM(m_i, m_j) = \exp\left(-\gamma_m \parallel IP(m_i) - IP(m_j) \parallel^2\right) \tag{2}$$

$$\gamma_m = \gamma_m' \left/ \left(\frac{1}{n_m} \sum_{i=1}^{n_m} \parallel IP(m_i) \parallel^2\right)\right. \tag{3}$$

where γ_m controls kernel bandwidth and $IP(m_i)$ denotes the i-th row of A. Similarly, for disease d_i and disease d_j, the GIP kernel similarity is computed as below:

$$KD(d_i, d_j) = \exp\left(-\gamma_d \parallel IP(d_i) - IP(d_j) \parallel^2\right) \tag{4}$$

$$\gamma_d = \gamma_d' \left/ \left(\frac{1}{n_d} \sum_{i=1}^{n_d} \parallel IP(d_i) \parallel^2\right)\right. \tag{5}$$

where γ_d controls kernel bandwidth and $IP(d_i)$ represents the i-th row of A^T. We set $\gamma_m' = 1$ and $\gamma_d' = 1$ according to the references [31, 32].

Functional similarity of miRNAs

According to the hypothesis that functionally similar miRNAs prefer associating with similar diseases, Wang et al. [33] calculated the functional similarity between miRNAs. These data can be downloaded from http://www.cuilab.cn/files/images/cuilab/misim.zip. Let *FM*

denotes miRNA functional similarity matrix, in which the element $FM(m_i, m_j)$ means the similarity value between miRNAs m_i and m_j.

Semantic similarity of diseases

We downloaded the relations of diseases from the Medical Subject Headings (MeSH) database (http://www.ncbi.nlm.nih.gov/) [33, 34]. Then, we constructed hierarchical directed acyclic graphs (DAGs), which are commonly applied to compute the disease semantic similarity. For a given disease d, $DAG(d) = (d, N(d), E(d))$, where $N(d)$ denotes the node-set containing d, and $E(d)$ represents the edge-set about d. Using two different methods from a previous study [35], we obtained two models of disease semantic similarity.

For disease t in DAG(d), its semantic contribution value to d is defined as

$$D1_d(t) = \begin{cases} 1, & if \ t = d \\ max\{\Delta * D1_d(t') \, | \, t' \in children \ of \ t\}, & if \ t \neq d \end{cases} \qquad (6)$$

where Δ is the semantic contribution factor and is often set to 0.5 [33]. After traversing all nodes in $N(d)$, the calculation of semantic value of d is shown below:

$$DV1(d) = \sum_{t \in N(d)} D1_d(t) \qquad (7)$$

For any two diseases d_i and d_j, the more shared nodes in their DAGs, the more similar they are. Then the semantic similarity between d_i and d_j is computed as follows:

$$SD1(d_i, d_j) = \frac{\sum_{t \in N(d_i) \cap N(d_j)} (D1_{d_i}(t) + D1_{d_j}(t))}{DV1(d_i) + DV1(d_j)} \qquad (8)$$

However, for a given disease d, the contribution of the diseases in the same layer of the DAG(d) is different. If the disease t in the DAG(d) appears less in other DAGs, its contribution to d is higher. Therefore, we adapted the model using another semantic similarity method [35] to represent t's semantic contribution to d:

$$D2_d(t) = -log \frac{\text{the number of DAGs including t}}{\text{the number of diseases}} \qquad (9)$$

Similar to formulae (7) and (8), we obtained the following formulae:

$$DV2(d) = \sum_{t \in N(d)} D2_d(t) \qquad (10)$$

$$SD2(d_i, d_j) = \frac{\sum_{t \in N(d_i) \cap N(d_j)} (D2_{d_i}(t) + D2_{d_j}(t))}{DV2(d_i) + DV2(d_j)} \qquad (11)$$

Finally, to better describe the disease semantic similarity, the mean value of *SD*1 and *SD*2 was calculated as follows:

$$SS(d_i, d_j) = \frac{SD1(d_i, d_j) + SD2(d_i, d_j)}{2} \qquad (12)$$

Integrated similarity of miRNAs and diseases

Using the similarity matrices mentioned above, we built the integrated similarity matrices of miRNAs and diseases, which denoted by SM and SD. SM is computed based on KM and FM. SD is computed using KD and SS. Therefore, the formulae for SM and SD are as follows:

$$SM(m_i, m_j) = \begin{cases} \frac{KM(m_i,m_j)+FM(m_i,m_j)}{2}, & if \ FM(m_i,m_j) \neq 0 \\ KM(m_i,m_j), & otherwise \end{cases} \tag{13}$$

$$SD(d_i, d_j) = \begin{cases} \frac{KD(d_i,d_j)+SS(d_i,d_j)}{2}, & if \ SS(d_i,d_j) \neq 0 \\ KD(d_i,d_j), & otherwise \end{cases} \tag{14}$$

Transformation of logistic function

Logistic function transformation has been performed successfully to adjust the similarity [18, 36, 37]. The logistic function can make the small value weaker and the large value stronger, thereby providing more differentiated similarity information for subsequent predictions. The final integrated similarity of miRNAs and diseases are defined as below:

$$LSM(m_i, m_j) = \frac{1}{1 + e^{(c \cdot SM(m_i,m_j)+d)}} \tag{15}$$

$$LSD(d_i, d_j) = \frac{1}{1 + e^{(c \cdot SD(d_i,d_j)+d)}} \tag{16}$$

where *c* and *d* are the control parameters. In this study, we set $c \in [-15, -1]$, tuned with five and tenfold CV. *d* was set to $log(9999)$ according to previous studies [18, 36, 37].

DAE-CFR for identifying unknown miRNA-disease interactions

To identify hidden miRNA-disease interactions, we developed a novel approach using deep autoencoder and combined feature representation (DAE-CFR). The entire computation process of DAE-CFR consisted of three steps (see Fig. 1).

Step 1 Data preprocessing.

We not only calculated similarity matrices LSM and LSD in the data preparation phase, but also constructed training samples. In this study, there were 189,585 miRNA-disease pairs consisting of 5430 known associations and 184,155 unknown associations. Here, known associations were defined as positive samples, whereas unknown associations were defined as negative samples. Since $184,155 \gg 5430$, there is a sample imbalance problem. To solve this problem, we introduced *k*-means clustering negative sampling, with $k = 23$ according to previous study [38]. First, we divided the negative samples into 23 groups, without making changes to the positive samples. Then we randomly chose 240 negative samples from each group so that the total number of all chosen negative samples was 5520, approximately equivalent to 5430. Overall, we obtained 10,950 training samples consisting of 5520 negative and 5430 positive samples.
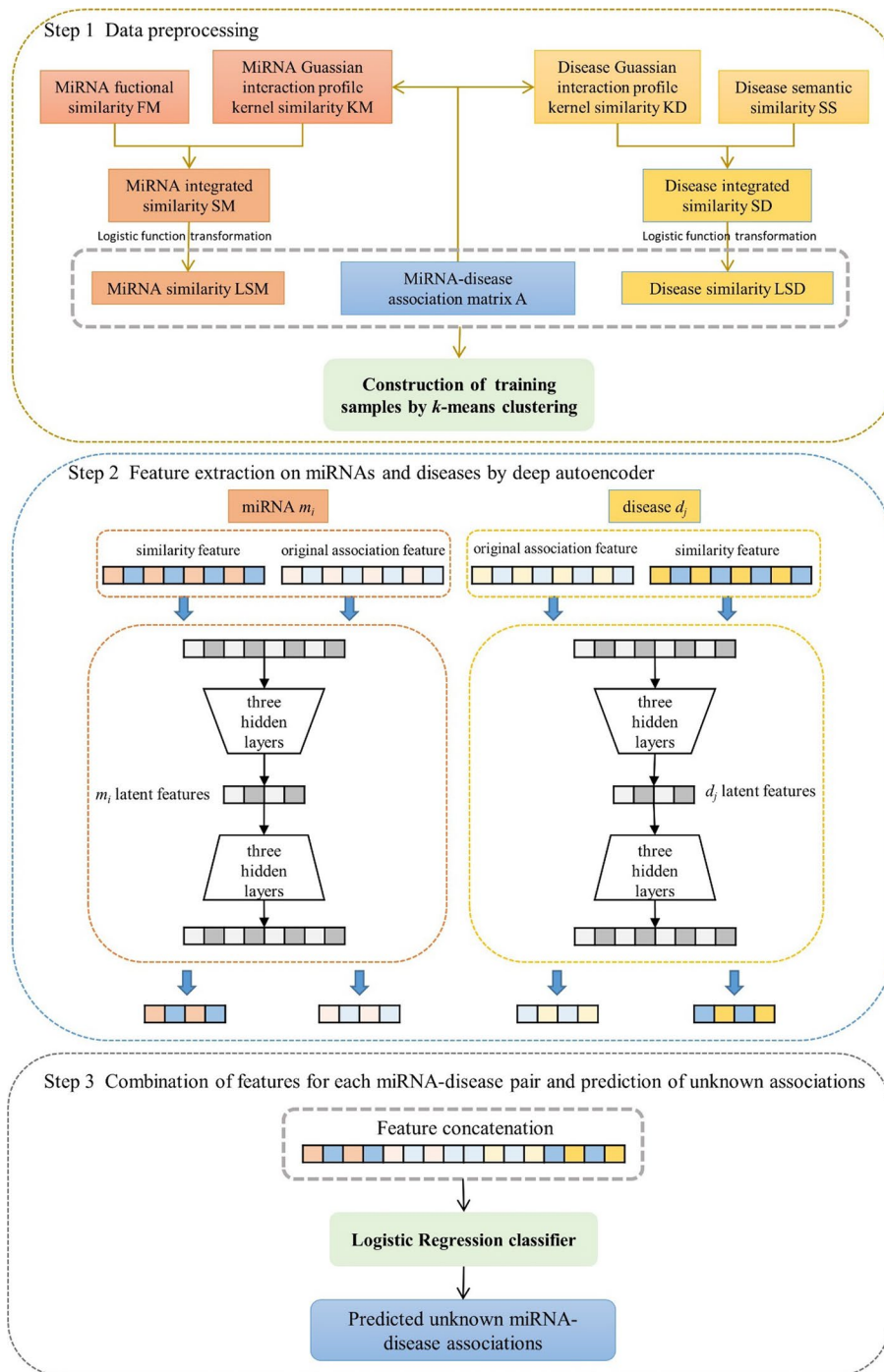
Fig. 1 Flowchart of DAE-CFR

Step 2 Feature extraction on miRNAs and diseases by deep autoencoder.

Based on the original association matrix A and the similarity matrices LSM and LSD, we obtained the feature representation. In this study, we considered two types of features for each miRNA and disease (see Table 1). The first type is the original association feature: $IP(m_i)$ represents the association information of miRNA m_i related to all diseases and

Table 1 Feature representation of each miRNA and disease

	The type of features	Feature representation	Notation	Dimension
miRNA m_i	Original association feature	The i-th row of A	$IP(m_i)$	383
	Similarity feature	The i-th row of LSM	$LSM(m_i)$	495
disease d_j	Original association feature	The j-th column of A	$IP(d_j)$	495
	Similarity feature	The j-th column of LSD	$LSD(d_j)$	383

$IP(d_j)$ denotes the association information of disease d_j related to all miRNAs. The second type is similarity feature: $LSM(m_i)$ represents the similar information of miRNA m_i with all miRNAs and $LSD(d_j)$ denotes the similar information of disease d_j with all diseases.

To represent the features more appropriately and reduce the computational complexity for subsequent prediction, we applied autoencoder to extract the low-dimensional feature representation of miRNAs and diseases. Autoencoder is an unsupervised model to recognize implicit biological patterns [39]. The autoencoder includes two phases: the encoder and the decoder [25]. In the encoding phase, the original data with high-dimensional features are compressed to low-dimensional features. In the decoding phase, the original inputs are reconstructed by mapping from the hidden layer to the output layer. The goal is to reduce the difference between the reconstructed and original data to a minimum. The autoencoder has a symmetric structure; that is, in the encoding phase, there are the same hidden layers as in the decoding phase. In this work, we used a deep autoencoder with three hidden layers to extract the latent and nonlinear features for each miRNA and disease. The deep autoencoder was implemented in the TensorFlow framework. The reduced dimensionality L of the latent features was set to 8, 16, 32, and 64, and we chose the proper dimensionality by comparing the effects of different L values. We set the batch size to 100 and used the Adam optimizer.

Step 3 Combination of features for each miRNA-disease pair and prediction of unknown associations.

The low-dimensional and latent features of miRNAs and diseases were extracted by deep autoencoder in the last step, which we denoted as $IP_1(m_i)$, $LSM_1(m_i)$, $IP_1(d_j)$ and $LSD_1(d_j)$ corresponding to the four features in Table 1, respectively. Then, we concatenated the four parts and obtained the feature representation for each miRNA-disease association as follows:

$$Vec(m_i, d_j) = [IP_1(m_i), LSM_1(m_i), IP_1(d_j), LSD_1(d_j)] \tag{17}$$

The dimension of $Vec(m_i, d_j)$ is $4L$, as each part is L-dimensional.

Finally, the above constructed features for each miRNA-disease pair were entered into the LR classifier to infer possible associations. These associations were then ranked by their predicted scores, with higher scores giving a higher rank. Pairs with higher scores are considered more likely to exist.

Results

Performance evaluation

K-fold CV has been widely used to assess model performance. In k-fold CV, the dataset is divided into k equal parts at random, with one part for testing and the residual parts

for training. Each part takes turns as a test set, and once all *k* parts have served as the test set, the average result of all *k* test sets is used as the final evaluation. Here, we used five and tenfold CV. AUC was used as the model evaluation index and AUC $\in [0, 1]$. A larger value of AUC indicates better model performance. Beyond AUC, we adopted several well-established metrics. These include the F1 score, which balances precision and recall; Accuracy (ACC), indicating the proportion of correctly predicted observations to the total observations; Area Under the Precision-Recall Curve (AUPR), reflecting both the precision and recall of the model; and the Matthews Correlation Coefficient (MCC), for comprehensive class performance assessment. Together, these metrics offer a multifaceted evaluation of the model's predictive performance and effectiveness.

Parameters analysis

There are two parameters in the overall model: $c \in \mathbb{Z}^-$ in the logistic function transformation and $L \in \mathbb{Z}^+$ in the deep autoencoder. In this work, we considered the following value ranges: $c \in [-15, -1]$ [18] and $L \in \{8, 16, 32, 64\}$, and repeated the experiment 10 times. The grid search algorithm was adopted to choose the best parameter values. After performing the calculations and comparing the results, we obtained the optimal parameters for fivefold CV framework: $c = -8$ and $L = 8$ (see Additional file 3: Table S1). In tenfold CV (see Additional file 3: Table S2), the best parameters were found to be $c = -8$ and $L = 16$. When $c = -8$ and $L = 8$, the AUC value ranked second. Therefore, for the convenience, we set $c = -8$ and $L = 8$ both in five and tenfold CV.

Comparison with other algorithms

To illustrate the excellent performance, DAE-CFR was compared with six popular algorithms: ABMDA [23], GBDT-LR [24], DFELMDA [25], KATZMDA [11] NCPMDA [17], and LWBRW [18]. We chose specific parameter settings for each model as described in the original study and all model parameters were listed in Additional file 3: Table S3.

We conducted fivefold CV on the dataset and the AUC values are shown in Fig. 2. The AUC of DAE-CFR reached 0.9691, which exceeded the AUCs of other algorithms (ABMDA: 0.8831, GBDT-LR: 0.9364, DFELMDA: 0.9479, KATZMDA: 0.9034, NCP-MDA: 0.8625, LWBRW: 0.9123). The results of all methods across various metrics were presented in Additional file 3: Table S4. The table highlighted the best-performing values for each metric in bold. From this comparison, it was evident that DAE-CFR
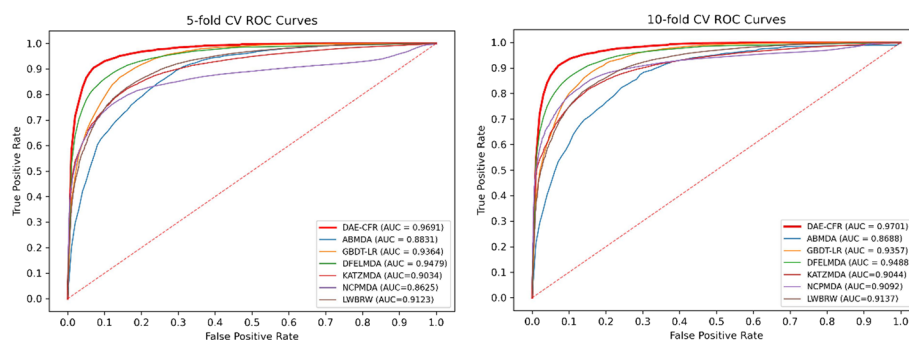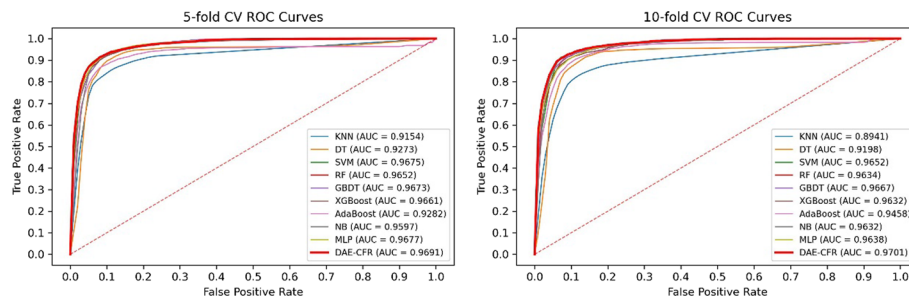

Fig. 2 ROC curves and AUCs of seven algorithms

Table 2 Average AUCs of seven algorithms with 10 repeats

	AUC (fivefold CV)	AUC (tenfold CV)
DAE-CFR	0.9692 ± 0.0014	0.9696 ± 0.0012
ABMDA	0.8776 ± 0.0080	0.8639 ± 0.0080
GBDT-LR	0.9358 ± 0.0018	0.9362 ± 0.0017
KATZMDA	0.9032 ± 0.0005	0.9042 ± 0.0002
DFELMDA	0.9478 ± 0.0013	0.9484 ± 0.0020
NCPMDA	0.8638 ± 0.0014	0.9091 ± 0.0003
LWBRW	0.9130 ± 0.0004	0.9140 ± 0.0002

Fig. 3 The ROC curves of different classifiers

outperformed all other methods evaluated. We then performed tenfold CV and obtained the AUC values of 0.9701, 0.8688, 0.9357, 0.9488, 0.9044, 0.9092, and 0.9137 for DAE-CFR, ABMDA, GBDT-LR, DFELMDA, KATZMDA, NCPMDA, and LWBRW, respectively. Therefore, our proposed DAE-CFR method exhibited excellent performance both in five and tenfold CV.

In addition, to ensure the robustness of the results, we repeated the experiment 10 times. The mean and standard deviation of AUCs for each method in five and tenfold CV were calculated and shown in Table 2. Here, DAE-CFR exhibited stable and superior performance compared to the six other methods, evidenced by low standard deviations and consistently higher AUC scores. This combination of reliability and effectiveness highlights its robustness in accurately predicting miRNA-disease associations, illustrating its value in biomedical research.

Comparison with other classifiers

In our model, we used LR classifier in the final step. To test the effectiveness of this choice, we replaced LR with the following nine common supervised learning classifiers: K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), RF, GBDT, eXtreme Gradient Boosting (XGBoost), AdaBoost, Naive Bayesian (NB) and MLP. Through calculations and subsequent comparisons, we derived the AUCs for various classifiers, as shown in Fig. 3. The computational results indicated that DAE-CFR outperformed the other nine classifier models, demonstrating that LR is particularly well-suited for DAE-CFR. We also repeated the experiment 10 times (see Table 3), and the results showed the stability of each method, further emphasizing the superiority of our method.

Table 3 Comparison of ten classifiers with 10 repeats

	AUC (fivefold CV)	AUC (tenfold CV)
LR (DAE-CFR)	**0.9692 ± 0.0014**	**0.9696 ± 0.0012**
KNN	0.9036 ± 0.0074	0.9082 ± 0.0050
DT	0.9243 ± 0.0041	0.9216 ± 0.0037
SVM	0.9649 ± 0.0014	0.9658 ± 0.0010
RF	0.9651 ± 0.0017	0.9617 ± 0.0018
GBDT	0.9644 ± 0.0022	0.9639 ± 0.0021
XGBoost	0.9642 ± 0.0015	0.9606 ± 0.0017
AdaBoost	0.9199 ± 0.0236	0.9148 ± 0.0197
NB	0.9619 ± 0.0011	0.9619 ± 0.0011
MLP	0.9435 ± 0.0231	0.9542 ± 0.0097

Ablation study

In the work, we constructed the feature representation of each miRNA and disease using two types of features: the original association feature and the similarity feature. We compared DAE-CFR with the following two models: (1) a model with only the original association feature; (2) a model with only the similarity feature. The combined information achieved the best performance in inferring the underlying miRNA-disease pairs, as depicted in Fig. 4. Furthermore, the results confirmed the stability with 10 repeats (see Table 4).

In the paper, we conducted the logistic function transformation in the similarity calculation, greatly enhancing model's performance. To illustrate this, we performed experiments without the logistic function transformation, as presented in Fig. 5, where DAE-CFR still performed best. The results with logistic function transformation in Fig. 4 surpassed those in Fig. 5, demonstrating the critical importance of the logistic function transformation in model building for better performance.
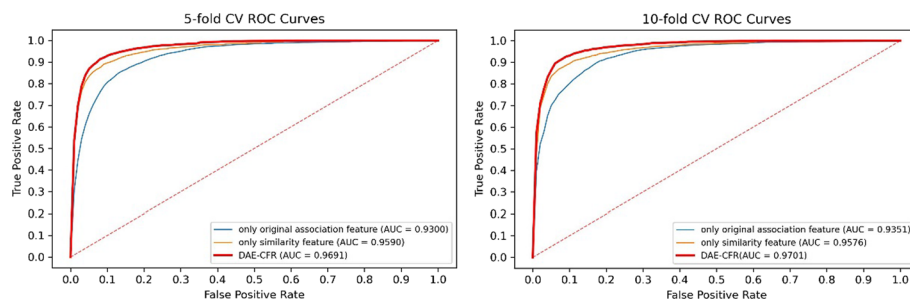
Fig. 4 The ROC curves in the ablation study

Table 4 The result of the ablation study with 10 repeats

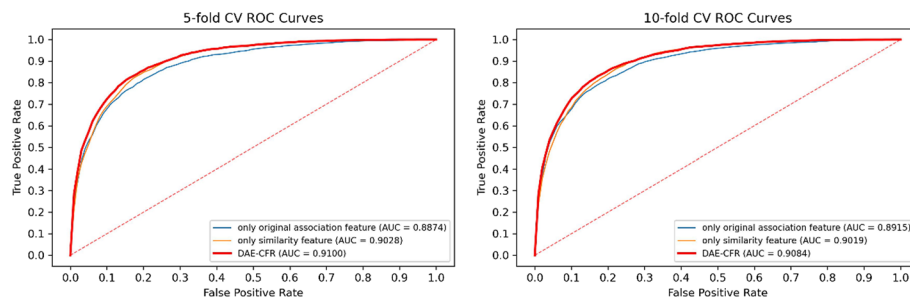|	AUC (fivefold CV)	AUC (tenfold CV)
Only original association feature	0.9293 ± 0.0050	0.9330 ± 0.0039
Only similarity feature	0.9586 ± 0.0009	0.9583 ± 0.0011
DAE-CFR	**0.9692 ± 0.0014**	**0.9696 ± 0.0012**

Fig. 5 The ROC curves without the logistic function transformation in the ablation study
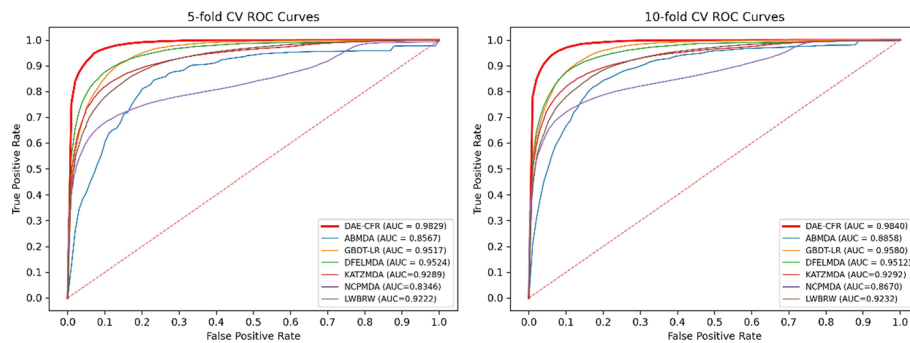

Fig. 6 The ROC curves of seven models using the HMDD v3.2 database

Robustness of DAE-CFR on another dataset

To test its ability to maintain outstanding performance, we applied the DAE-CFR model to an additional dataset. For this validation, we utilized the HMDD v3.2 database [40] to extract known interactions between miRNAs and diseases. Following the data refinement, a total of 8,968 known interactions involving 374 diseases and 788 miRNAs were selected, as detailed in the reference [41] (see Additional file 2). Implementing the same experimental setup as before, the results shown in Fig. 6 indicated that DAE-CFR achieved an AUC of 0.9829 in fivefold CV, surpassing the AUCs of ABMDA (0.8567), GBDT-LR (0.9517), DFELMDA (0.9524), KATZMDA (0.9289), NCPMDA (0.8346), and LWBRW (0.9222). In tenfold CV, DAE-CFR reached an AUC score of 0.9840, outperforming the AUCs of ABMDA (0.8858), GBDT-LR (0.9580), DFELMDA (0.9512), KATZMDA (0.9292), NCPMDA (0.8670), and LWBRW (0.9232). The superior performance achieved on the HMDD v3.2 database highlights the consistency and robustness of our model.

Case studies

To further verify the accuracy and validity of DAE-CFR, we conducted case studies on myocardial infarction (MI), hypertension (HTN) and stroke. In the study, after calculating the predicted scores, we ranked all unknown pairs and listed the top 10 miRNAs for each of the three diseases in Table 5. Subsequently, the predicted miRNAs were verified using the RNADisease database [42].

MI is a significant component of the global cardiovascular disease burden, leading to increased hospital admissions and substantial financial implications all over the world

Liu *et al. BMC Bioinformatics*   (2024) 25:139

Page 13 of 19

Table 5 Top 10 miRNAs related to three diseases predicted by DAE-CFR

Diseases	Rank	MiRNA	Predicted score	Evidence
Myocardial Infarction	1	hsa-mir-146a	0.9895	RNADisease
	2	hsa-mir-221	0.9867	RNADisease
	3	hsa-mir-125b	0.9855	RNADisease
	4	hsa-mir-145	0.9839	RNADisease
	5	hsa-mir-17	0.9784	RNADisease
	6	hsa-mir-16	0.9770	RNADisease
	7	hsa-mir-222	0.9721	RNADisease
	8	hsa-mir-200c	0.9711	RNADisease
	9	hsa-let-7a	0.9709	Unknown
	10	hsa-mir-181a	0.9664	RNADisease
Hypertension	1	hsa-mir-146a	0.9862	RNADisease
	2	hsa-mir-34a	0.9821	RNADisease
	3	hsa-mir-221	0.9815	RNADisease
	4	hsa-mir-125b	0.9801	Unknown
	5	hsa-mir-126	0.9752	RNADisease
	6	hsa-mir-16	0.9686	RNADisease
	7	hsa-mir-222	0.9623	RNADisease
	8	hsa-mir-200c	0.9609	Unknown
	9	hsa-let-7a	0.9607	Unknown
	10	hsa-mir-143	0.9603	RNADisease
Stroke	1	hsa-mir-21	0.9995	RNADisease
	2	hsa-mir-146a	0.9824	RNADisease
	3	hsa-mir-34a	0.9771	RNADisease
	4	hsa-mir-221	0.9759	RNADisease
	5	hsa-mir-125b	0.9744	RNADisease
	6	hsa-mir-126	0.9684	RNADisease
	7	hsa-mir-20a	0.9640	RNADisease
	8	hsa-mir-17	0.9624	RNADisease
	9	hsa-mir-16	0.9598	RNADisease
	10	hsa-mir-29a	0.9545	RNADisease

[43]. MiRNAs have been found to be circulating biomarkers for the diagnosis and prevention of MI [43–46]. The top 10 predicted MI-related miRNAs were shown in Table 5, with 9 out of 10 verified by the RNADisease database. The "unknown" miRNA hsa-let-7a maybe a MI-related biomarker. We conducted a literature search on PubMed and found two articles for hsa-let-7a related to MI. Du et al. [47] found that hsa-let-7a controls the expression of β1-AR and establishes a negative feedback mechanism within the β1-AR signaling pathway in cases of ischemic heart failure. This discovery offers a fresh perspective on the differences in β1-AR expression between the early and later stages of MI. According to Gan et al. [48], the circRNA-101237/let-7a-5p/IGF2BP3 axis, which plays a role in controlling cardiomyocyte death, presents potential as a promising therapeutic target for addressing cardiovascular diseases, including MI. These two clues imply that hsa-let-7a is a promising biomarker of MI and may be confirmed by further biological experiment.

The impact of HTN on public health and the economy is far beyond the scope of HTN treatment [49]. HTN, as a common chronic disease that affects the aging population

Liu *et al. BMC Bioinformatics*   (2024) 25:139

Page 14 of 19

[50], is a risk factor for many diseases including cardiovascular disease [51], chronic kidney disease [52] and so on, which severely threatens human life and health. Several miRNAs have been identified as potential HTN biomarkers [53, 54]. In the study, 7 out of 10 HTN-related miRNAs confirmed by the RNADisease database, as shown in Table 5. The "unknown" miRNAs (hsa-mir-125b, hsa-mir-200c and hsa-let-7a) may be novel biomarkers. We conducted a search on PubMed and identified one paper on hsa-let-7a related to HTN. Through an investigation into the roles of brain microvascular pericyte-derived extracellular vesicle miRNAs in HTN, Wu et al. [55] identified specific miRNAs like miR-21-5p, let-7c-5p, and let-7a-5p that showed abnormal expression in spontaneously hypertensive rats compared to normotensive rats. This study sheds light on the connection between brain microvascular pericytes and HTN. It suggests that hsa-let-7a is a more likely biomarker for HTN.

Stroke is a significant global cause of both mortality and disability, affecting people worldwide [56]. Notably, 87% of all strokes are ischemic in nature [57]. Researchers have explored the potential of miRNAs as biomarkers for diagnosing, predicting outcomes, and assessing brain injury in ischemic strokes [58–60]. Here, we employed DAE-CFR to identify miRNAs associated with stroke and selected the top 10 candidates. Consequently, all of the 10 miRNAs were confirmed by the RNADisease database (see Table 5).

Pathway analysis

Inspired by references [61] and [19], we recognized the importance of in-depth pathway analysis in understanding the role of miRNAs in disease incidence. Therefore, we employed DIANA-miRPath v4.0 [62], an online platform for miRNA target and pathway analysis, to explore the regulatory functions of miRNAs and their impact on various pathways, shedding light on their connections to diseases. Specifically, we illustrated, using stroke as an example, how the majority of miRNA targets identified through the DAE-CFR method are linked to biological processes and functionalities relevant to stroke. Details of the top 10 enrichment results for stroke-associated candidate miRNAs were listed in Table 6. Research has underscored the neuroprotective role of the PI3K-Akt signaling pathway in ischemic stroke [63]. Furthermore, an association has been found between hepatitis B virus infection and a decreased risk of ischemic stroke [64]. In Fig. 7, a heatmap was presented, created using miRpathDB v2.0 [65], to show

Table 6 TOP 10 Enrichment results for Stroke-related candidate miRNAs

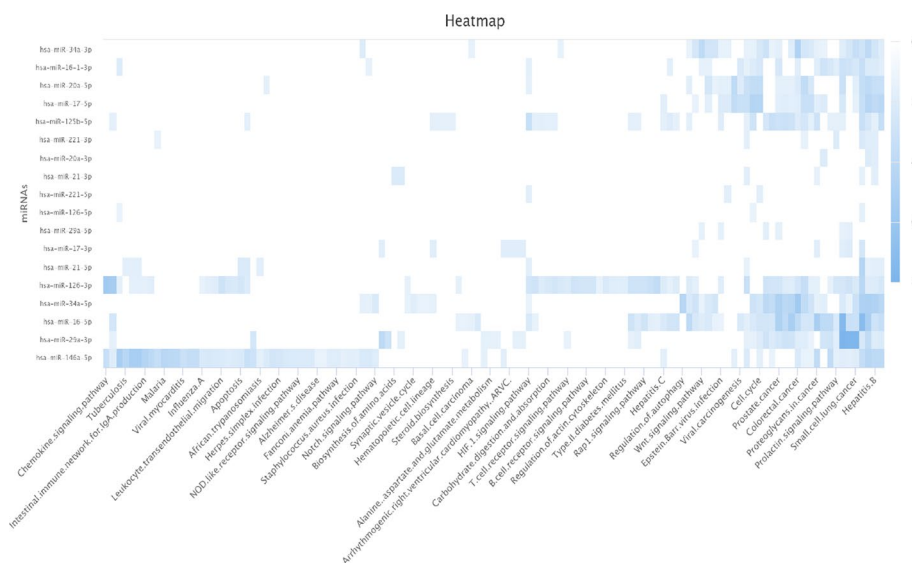| KEGG pathway	*p*-value
Pathways in cancer	1.24e-126
MicroRNAs in cancer	1.22e-86
PI3K-Akt signaling pathway	1.05e-72
Breast cancer	1.14e-70
Hepatitis B	1.22e-70
Pancreatic cancer	1.23e-70
AGE-RAGE signaling pathway in diabetic complications	1.27e-70
EGFR tyrosine kinase inhibitor resistance	1.08e-67
Proteoglycans in cancer	9.42e-67
Prostate cancer	3.76e-65

Fig. 7 The heatmap of Stroke-related miRNAs

the associations between miRNA targets and their respective pathways, where a darker shade indicated a stronger association with pathway functions. This pathway analysis not only validated the reliability and efficiency of the DAE-CFR method in identifying disease-related candidate miRNAs but also provided valuable insights into the role of miRNAs in diseases.

Discussion

In this study, we developed a novel deep learning framework DAE-CFR for predicting hidden associations between miRNAs and diseases. The computational results indicated that DAE-CFR achieved outstanding performance in both five and tenfold CV. There are several reasons why DAE-CFR performed excellently. First, we introduced more biological information into the model. We computed integrated similarity of miRNAs and diseases using two sources of miRNA similarity and three types of disease similarity, respectively. Second, we applied the logistic function transformation to the two integrated similarity matrices to improve the discrimination of the similarity information. Third, we adopted *k*-means clustering negative sampling, which is simple and efficient for large-scale datasets. Fourth, the use of deep autoencoder achieved feature dimension reduction and improved computational efficiency. We considered both the original association feature and similarity feature for miRNAs and diseases and found that the combined features outperformed the singular features in the ablation study. Due to all of these factors, our model significantly enhanced forecasting performance.

Despite its numerous strengths, DAE-CFR has limitations that require further exploration and refinement. First, GIP kernel calculations are essentially based on current miRNA-disease associations. The limited number of known interactions could bias the predictive results. This scarcity of confirmed associations might lead models to overvalue the few recognized links, potentially neglecting unexplored or novel interactions. Second, we treated known associations as positive samples and considered all unknown

associations as negative samples, facing the challenge of acquiring credible negative samples, inherently difficult to obtain in biological systems. This strategy may risk neglecting or misclassifying potential interactions, further illustrating the limitations of relying only on known miRNA-disease relationships. Third, in our method, we did not consider how changes in the relationships between miRNAs and diseases might affect model performance. This consideration is crucial for ensuring that our model remains effective as association data are updated. Forth, our analysis has focused solely on miRNAs as the biological determinant in disease pathogenesis, overlooking the roles of other biological entities, such as proteins and lncRNAs, which also influence disease mechanisms. Future research should not only extend this methodological framework to more miRNA-disease association datasets and more prediction challenges but also aim to include sensitivity analysis to evaluate how variations in these associations affect model performance. Additionally, we plan to broaden the investigative scope to encompass a wider array of biological entities and their interactions. This comprehensive approach will enrich our understanding and enhance the accuracy of our predictions.

Conclusion

Previous studies have found that miRNAs are critical in disease processes. Inferring unknown miRNA-disease interactions can increase our understanding of the pathogenesis of complex human diseases, contributing to their prevention and therapy. In recent years, the identification of miRNA-disease associations has increased significantly owing to the growth of experimental technologies. However, laboratory methods can be time-consuming and laborious. Therefore, many researchers have developed algorithms to forecast potential miRNA-disease pairs. In this research, we proposed the DAE-CFR method, which employs deep autoencoder for complex feature extraction and utilizes a combined feature representation technique. First, we computed the integrated similarity of miRNAs and diseases using GIP kernel similarity, miRNA functional similarity, and two types of disease semantic similarity; applied the logistic function transformation to obtain the final integrated similarity; balanced the positive and negative samples by *k*-means clustering and then constructed training samples. Second, we used deep autoencoder to extract latent features from two types of feature representation for each miRNA and disease: the original association feature and the similarity feature. Finally, we combined these latent features to form the feature representation for each miRNA-disease pair, and then applied the LR classifier to forecast unknown pairs. To verify the superiority of DAE-CFR, we compared it with six other popular models: ABMDA, GBDT-LR, DFELMDA, KATZMDA, NCPMDA, and LWBRW in five and tenfold CV frameworks, finding that DAE-CFR showed the best results with AUCs of 0.9691 and 0.9701, respectively. In our model, we chose LR classifier in the final prediction, which was superior to the other nine common classifiers. Subsequently, the robustness was affirmed through its validation on another dataset. Furthermore, we conducted case studies on three diseases and found that the accuracy of the top 10 predicted miRNAs for MI, HTN and Stroke was 90%, 70% and 100%, respectively. The final pathway analysis validated the DAE-CFR method's effectiveness in identifying disease-related miRNAs and offered insights into miRNAs' roles in diseases, enhancing our model's predictive accuracy and biological understanding. In summary, DAE-CFR presented powerful

performance in identifying miRNA-disease associations, demonstrating its significant potential in the field.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-024-05757-y.

Additional file 1. Known miRNA-disease associations collected from HMDD v2.0 database.

Additional file 2. Known miRNA-disease associations collected from HMDD v3.2 database.

Additional file 3. Supplementary Tables. This file includes four tables: Table S1. The AUC values of 5-fold CV in parameters analysis with 10 repeats; Table S2. The AUC values of 10-fold CV in parameters analysis with 10 repeats; Table S3. Parameter settings for all methods; Table S4. The performance comparison of different methods on 5-fold CV.

Author contributions

YL conceived the study, designed the study protocol, and wrote the code. YL, RZ and XD wrote the manuscript. HY, JL, HC participated in the data analysis. JT and YZ came up with the original concept for the study, oversaw the data analysis, and revised the paper. All authors contributed to the article and approved the final manuscript.

Availability of data and materials

The datasets used in this study are included in the article/Supplementary Material.

Declarations

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–97.
2. Kabekkodu SP, Shukla V, Varghese VK, Adiga D, Vethil Jishnu P, Chakrabarty S, Satyamoorthy K. Cluster miRNAs and cancer: diagnostic, prognostic and therapeutic opportunities. Wiley Interdiscip Rev RNA. 2020;11(2): e1563.
3. Vasudevan S, Tong YC, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. Science. 2007;318(5858):1931–4.
4. Andres-Leon E, Gonzalez Pena D, Gomez-Lopez G, Pisano DG. miRGate: a curated database of human, mouse and rat miRNA-mRNA targets. Database (Oxford). 2015;2015:bav035.
5. Taganov KD, Boldin MP, Kuang-Jung C, David B. NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proc Natl Acad Sci United States Am. 2006;103(33):12481–6.
6. Miska EA. How microRNAs control cell division, differentiation and death. Curr Opin Genet Dev. 2005;15(5):563–8.
7. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. Gastroenterology. 2007;133(2):647–58.
8. Izreig S, Samborska B, Johnson RM, Sergushichev A, Ma EH, Lussier C, Loginicheva E, Donayo AO, Poffenberger MC, Sagan SM, Vincent EE, Artyomov MN, Duchaine TF, Jones RG. The miR-17 approximately 92 microRNA cluster is a global regulator of tumor metabolism. Cell Rep. 2016;16(7):1915–28.
9. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. Int J Biochem Cell Biol. 2010;42(8):1273–81.
10. Tricoli JV, Jacobson JW. MicroRNA: potential for cancer detection, diagnosis, and prognosis. Cancer Res. 2007;67(10):4553–5.
11. Qu Y, Zhang HX, Liang C, Dong X. KATZMDA: prediction of miRNA-disease associations based on KATZ model. IEEE Access. 2018;6:3943–50.

Liu *et al. BMC Bioinformatics*   (2024) 25:139

Page 18 of 19

12. Matulic M, Grskovic P, Petrovic A, Begic V, Harabajsa S, Korac P. miRNA in molecular diagnostics. Bioeng (Basel). 2022;9(9):459.

13. Wu Y, Zhu DH, Wang XF, Zhang S. An ensemble learning framework for potential miRNA-disease association prediction with positive-unlabeled data. Comput Biol Chem. 2021;95: 107566.

14. Wang YT, Wu QW, Gao Z, Ni JC, Zheng CH. MiRNA-disease association prediction via hypergraph learning based on high-dimensionality features. BMC Med Inform Decis Mak. 2021;21(Suppl 1):133.

15. Jiang QH, Hao YY, Wang GH, Juan L, Zhang TJ, Teng MX, Liu YL, Wang YD. Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst Biol. 2010;4(Suppl 1):S2.

16. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. Mol Biosyst. 2012;8(10):2792–8.

17. Gu C, Liao B, Li X, Li K. Network consistency projection for human miRNA-disease associations inference. Sci Rep. 2016;6(1):36054.

18. Dai LY, Liu JX, Zhu R, Wang J, Yuan SS. Logistic weighted profile-based bi-random walk for exploring MiRNA-disease associations. J Comput Sci Technol. 2021;36(2):276–87.

19. Ha J. SMAP similarity-based matrix factorization framework for inferring miRNA-disease association. Knowl-Based Syst. 2023;263:110295.

20. Wang L, You ZH, Huang YA, Huang DS, Chan KCC. An efficient approach based on multi-sources information to predict circRNA-disease associations using deep convolutional neural network. Bioinformatics. 2020;36(13):4038–46.

21. Liu DY, Huang YB, Nie WJ, Zhang JX, Deng L. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGBoost. BMC Bioinformatics. 2021;22(1):219.

22. Chen X, Wang CC, Yin J, You ZH. Novel human miRNA-disease association inference based on random forest. Molecular Therapy Nucleic Acids. 2018;13:568–79.

23. Zhao Y, Chen X, Yin J. Adaptive boosting-based computational model for predicting potential miRNA-disease associations. Bioinformatics. 2019;35(22):4730–8.

24. Zhou S, Wang SL, Wu Q, Azim R, Li W. Predicting potential miRNA-disease associations by combining gradient boosting decision tree with logistic regression. Comput Biol Chem. 2020;85: 107200.

25. Liu W, Lin H, Huang L, Peng L, Tang T, Zhao Q, Yang L. Identification of miRNA-disease associations via deep forest ensemble learning based on autoencoder. Brief Bioinform. 2022;23(3):bbac104.

26. Chen X, Li TH, Zhao Y, Wang CC, Zhu CC. Deep-belief network for predicting potential miRNA-disease associations. Brief Bioinform. 2021;22(3):bbac186.

27. Ha J, Park S. NCMD: Node2vec-based neural collaborative filtering for predicting MiRNA-disease association. IEEE/ACM Trans Comput Biol Bioinf. 2023;20(2):1257–68.

28. Li Y, Qiu CX, Tu J, Geng B, Yang JC, Jiang TZ, Cui QH. HMDD v20: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. 2014;42:D1070–4.

29. Sun PC, Yang SY, Cao Y, Cheng RJ, Han SY. Prediction of potential associations between miRNAs and diseases based on matrix decomposition. Front Genet. 2020;11: 598185.

30. Van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics. 2011;27(21):3036–43.

31. Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics. 2013;29(20):2617–24.

32. Xu JL, Cai LJ, Liao B, Zhu W, Wang P, Meng YJ, Lang JD, Tian G, Yang JL. Identifying potential miRNAs-disease associations with probability matrix factorization. Front Genet. 2019;10:1234.

33. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26(13):1644–50.

34. Li J, Zhang S, Liu T, Ning C, Zhang Z, Zhou W. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. Bioinformatics. 2020;36(8):2538–46.

35. Ding YL, Tian LP, Lei XJ, Liao B, Wu FX. Variational graph auto-encoders for miRNA-disease association prediction. Methods. 2020;192:25–34.

36. Wen YQ, Han GS, Anh VV. Laplacian normalization and bi-random walks on heterogeneous networks for predicting lncRNA-disease associations. BMC Syst Biol. 2018;12(S9):122.

37. Yin MM, Cui Z, Gao MM, Liu JX, Gao YL. LWPCMF: logistic weighted profile-based collaborative matrix Factorization for predicting MiRNA-disease associations. IEEE/ACM Trans Comput Biol Bioinform. 2021;18(3):1122–9.

38. Rayhan F, Ahmed S, Shatabda S, Farid DM, Mousavian Z, Dehzangi A, Rahman MS. iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting. Sci Rep. 2017;7(1):17731.

39. Chicco D, Sadowski P, Baldi P. (2014) Deep autoencoder neural networks for gene ontology annotation predictions. In: the 5th ACM conference on bioinformatics, computational biology, and health informatics, pp 533–40.

40. Huang Z, Shi JC, Gao YX, Cui CM, Zhang S, Li JW, Zhou Y, Cui QH. HMDD v3.0: a database for experimentally supported human microRNA-disease. Nucleic Acids Res. 2019;47(D1):D1013–7.

41. Lou Z, Cheng Z, Li H, Teng Z, Liu Y, Tian Z. Predicting miRNA–disease associations via learning multimodal networks and fusing mixed neighborhood information. Brief Bioinf. 2022;23(5):bbac159.

42. Chen J, Lin JH, Hu YF, Ye M, Yao L, Wu L, Zhang W, Wang M, Deng T, Guo F, Huang Y, Zhu B, Wang D. RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. Nucleic Acids Res. 2023;51(D1):D1397–404.

43. Lock MC, Tellam RL, Botting KJ, Wang KCW, Selvanayagam JB, Brooks DA, Seed M, Morrison JL. The role of miRNA regulation in fetal cardiomyocytes, cardiac maturation and the risk of heart disease in adults. J Physiol. 2018;596(23):5625–40.

44. Zhang L, Zhang Y, Yu F, Li X, Gao H, Li P. The circRNA-miRNA/RBP regulatory network in myocardial infarction. Front Pharmacol. 2022;13: 941123.

45. Krzywińska O, Bracha M, Jeanniere C, Recchia E, Kędziora Kornatowska K, Kozakiewicz M. Meta-analysis of the potential role of miRNA-21 in cardiovascular system function monitoring. Biomed Res Int. 2020;2020:4525410.

46. Sharma AK, Bisht P, Gupta B, Sayeed Akhtar MD, Shaik Alavudeen S, Afzal O, Sa AA. Investigating miRNA subfamilies: can they assist in the early diagnosis of acute myocardial infarction? Drug Discovery Today. 2023;28(10): 103695.

47. Du Y, Zhang MY, Zhao W, Shu Y, Lu Y. Let-7a regulates expression of β1-adrenoceptors and forms a negative feedback circuit with the β1-adrenoceptor signaling pathway in chronic ischemic heart failure. Oncotarget. 2017;8(5):8752–64.

48. Gan JT, Yuan J, Liu Y, Lu ZD, Xue Y, Shi L, Zeng H. Circular RNA-101237 mediates anoxia/reoxygenation injury by targeting let-7a-5p/IGF2BP3 in cardiomyocytes. Int J Mol Med. 2020;45(2):451–60.

49. Hodgson TA, Cai L. Medical care expenditures for hypertension, its complications, and its comorbidities. Med Care. 2001;39(6):599–615.

50. Sekar D, Shilpa BR, Das AJ. Relevance of microRNA 21 in different types of hypertension. Curr Hypertens Rep. 2017;19(7):57.

51. Lawes CM, Vander Hoorn S, Rodgers A. International Society of H Global burden of blood-pressure-related disease. Lancet. 2008;371(9623):1513–8.

52. Harjutsalo V, Groop PH. Epidemiology and risk factors for diabetic kidney disease. Adv Chronic Kidney Dis. 2014;21(3):260–6.

53. Romaine SP, Charchar FJ, Samani NJ, Tomaszewski M. Circulating microRNAs and hypertension–from new insights into blood pressure regulation to biomarkers of cardiovascular risk. Curr Opin Pharmacol. 2016;27:1–7.

54. Zhang XY, Wang XY, Wu J, Peng J, Deng X, Shen Y, Yang CJ, Yuan J, Zou YZ. The diagnostic values of circulating miR-NAs for hypertension and bioinformatics analysis. Biosci Rep. 2018;38(4):BSR20180525.

55. Wu QB, Yuan XC, Li BW, Yang JN, Han RQ, Zhang HG, Xiu RJ. Differential miRNA expression analysis of extracellular vesicles from brain microvascular pericytes in spontaneous hypertensive rats. Biotech Lett. 2020;42(3):389–401.

56. Chen HS, He YC, Chen S, Qi SH, Shen JG. Therapeutic targets of oxidative/nitrosative stress and neuroinflammation in ischemic stroke: applications for natural product efficacy with omics and systemic biology. Pharmacol Res. 2020;158: 104877.

57. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, Commodore-Mensah Y, Elkind MSV, Evenson KR, Eze-Nliam C, Fugar S, Generoso G, Heard DG, Hiremath S, Ho JE, Kalani R, Kazi DS, Ko D, Levine DA, Liu J, Ma J, Magnani JW, Michos ED, Mussolino ME, Navaneethan SD, Parikh NI, Poudel R, Rezk-Hanna M, Roth GA, Shah NS, St-Onge M-P, Thacker EL, Virani SS, Voeks JH, Wang N-Y, Wong ND, Wong SS, Yaffe K, Martin SS. Heart disease and stroke statistics—2023 update: a report from the american heart association. Circulation. 2023;147(8):e93–621.

58. Wicik Z, Eyileten C, Jakubik D, Simões SN, Martins DC, Pavão R, Siller-Matula JM, Postula M. ACE2 interaction networks in COVID-19: a physiological framework for prediction of outcome in patients with cardiovascular risk factors. J Clin Med. 2020;9(11):3743.

59. Bao Y, Li SJ, Ding YY, Du X, Zhang M, Tang W, Zhou S. MiRNA: a potential target for gene diagnosis and treatment of atherosclerotic stroke. Int J Neurosci. 2020;131(3):283–8.

60. Gugliandolo A, Silvestro S, Sindona C, Bramanti P, Mazzon E. MiRNA: involvement of the MAPK pathway in ischemic stroke. Promis Therapeutic Target Med. 2021;57(10):1053.

61. Ha J. MDMF: predicting miRNA–disease association based on matrix factorization with disease similarity constraint. J Personal Med. 2022;12(6):885.

62. Tastsoglou S, Skoufos G, Miliotis M, Karagkouni D, Koutsoukos I, Karavangeli A, Kardaras F, Hatzigeorgiou A. DIANA-miRPath v4.0: expanding target-based miRNA functional analysis in cell-type. Nucleic Acids Res. 2023;51(W1):W154–9.

63. Wang C, Wei Z, Jiang G, Liu H. Neuroprotective mechanisms of miR-124 activating PI3K/Akt signaling pathway in ischemic stroke. Exp Ther Med. 2017;13(6):3315–8.

64. Tseng CH, Muo CH, Hsu CY, Kao CH. Association of hepatitis B virus infection with decreased ischemic stroke. Acta Neurol Scand. 2016;134(5):339–45.

65. Kehl T, Kern F, Backes C, Fehlmann T, Stöckel D, Meese E, Lenhof HP, Keller A. miRPathDB 2.0: a novel release of the miRNA pathway dictionary database. Nucleic Acids Res. 2020;48(D1):D142–7.

Publisher's Note