



## OPEN ACCESS

EDITED BY  
Tianyi Zhao,  
Harbin Institute of Technology, China

REVIEWED BY  
Quan Zou,  
University of Electronic Science and  
Technology of China, China  
Junwei Han,  
Harbin Medical University, China

\*CORRESPONDENCE  
Zhufang Kuang  
zfkuan.cn@163.com

SPECIALTY SECTION  
This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Oncology

RECEIVED 06 July 2022  
ACCEPTED 14 September 2022  
PUBLISHED 29 September 2022

CITATION  
Duan T, Kuang Z and Deng L (2022)  
SVMMMDR: Prediction of miRNAs-drug  
resistance using support vector  
machines based on  
heterogeneous network.  
*Front. Oncol.* 12:987609.  
doi: 10.3389/fonc.2022.987609

COPYRIGHT  
© 2022 Duan, Kuang and Deng. This is  
an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# SVMMMDR: Prediction of miRNAs-drug resistance using support vector machines based on heterogeneous network

Tao Duan, Zhufang Kuang\* and Lei Deng

School of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha, China

In recent years, the miRNA is considered as a potential high-value therapeutic target because of its complex and delicate mechanism of gene regulation. The abnormal expression of miRNA can cause drug resistance, affecting the therapeutic effect of the disease. Revealing the associations between miRNAs-drug resistance can help in the design of effective drugs or possible drug combinations. However, current conventional experiments for identification of miRNAs-drug resistance are time-consuming and high-cost. Therefore, it's of pretty realistic value to develop an accurate and efficient computational method to predicting miRNAs-drug resistance. In this paper, a method based on the Support Vector Machines (SVM) to predict the association between MiRNA and Drug Resistance (SVMMMDR) is proposed. The SVMMMDR integrates miRNAs-drug resistance association, miRNAs sequence similarity, drug chemical structure similarity and other similarities, extracts path-based Hetesim features, and obtains inclined diffusion feature through restart random walk. By combining the multiple feature, the prediction score between miRNAs and drug resistance is obtained based on the SVM. The innovation of the SVMMMDR is that the inclined diffusion feature is obtained by inclined restart random walk, the node information and path information in heterogeneous network are integrated, and the SVM is used to predict potential miRNAs-drug resistance associations. The average AUC of SVMMMDR obtained is 0.978 in 10-fold cross-validation.

## KEYWORDS

miRNA, drug resistance, support vector machines, hetesim score, random walk with restart

## 1 Introduction

In recent years, the difficulty of drug target selection has led to the increase of drug development cost and the low efficiency of pharmaceutical industry. So far, it has been discovered that the human genome can encode up to 25,000 genes. But only 600 of the disease-causing proteins have targeted drugs, meaning a significant number are

“undrugable”. Therefore, the focus of target selection has now shifted to other macromolecules, such as non-coding RNA.

According to the genetic central dogma, the DNA is a carrier that carries genetic information. During growth and development, the genetic information in DNA is transcribed into RNA. Then the RNA is translated into various proteins to perform specific biological functions. There are many types of RNAs with complex functions. Research shows that only 2% of the RNA could code for proteins, and 98% couldn't. In biology, RNAs with non-coding are called non-coding RNAs (ncRNAs). Among ncRNAs, miRNA is a group of non-coding RNAs encoded by the genome with a length of about 2023 nucleotides. The miRNAs play an major role in gene expression regulation. They have a significant meaning in many biological processes such as cell differentiation, development and cellular signaling pathways.

The miRNAs play an important role in the understanding of life sciences. The miRNAs are significant in many aspects such as cellular biological processes, gene expression regulation at transcriptional and post-transcriptional levels, and others.

There are many studies on the biological mechanisms and interactions between genes, miRNAs, lncRNAs, diseases and drugs, such as the relationship between genes and diseases, miRNAs and diseases, lncRNAs and diseases, miRNAs and lncRNAs, etc.

For the association between genes and diseases, a network impulse dynamics framework based on multiple biological networks NIDM is proposed by Xiang et al. (1) to predict potential disease-gene associations. The HyMM is proposed by Xiang et al. (2) to more effectively predict disease-related genes by integrating information from the structure of multi-scale modules. The PrGeFNE is proposed by Xiang et al. (3) based on fast network embedding. An understanding of the association between genetics and disease can help understand the pathogenesis of disease.

For the association between miRNAs and diseases, a meta-path-based MDPBMP is proposed by Yu et al. (4). The information carried by the nodes is extracted and integrated through MDPBMP, and the miRNAs-disease association is predicted using embedded feature vectors. The VGAE-MDA, a deep learning framework with variational graph autoencoder, is proposed by Ding et al. (5). The MLPMDA, a miRNAs-disease association prediction method using multilayer linear projection, is proposed by Guo et al. (6). The prediction method GRPAMDA is proposed by Zhong et al. (7). The GRPAMDA combines the graph random propagation network based on DropFeature and attention network. The NIMGSA is proposed by Jin et al. (8) to predict miRNAs-disease association based on neural induction matrix completion. An ensemble learning framework with resampling method for miRNA-disease association ERMDA prediction is proposed by Dai et al. (9). A double random walk model is proposed by Zhu et al. (10). The end-to-end deep learning method PDMDA is proposed by Yan et al. (11). A

computational framework SMALF based on XGBoost is proposed by Liu et al. (12). The algorithm MSCDE is proposed by Han et al. (13) based on a variety of biological source information. The method based on tensor factorization and label propagation (TFLP) is proposed by Yu et al. (14) for multi-type miRNA-disease association prediction.

For the association between lncRNAs and diseases, a non-negative matrix factorization based on graph regularization LDGRNMF is proposed by Wang et al. (15) to predict the lncRNAs-disease association. The internal inclined random walk with restart (IIRWR) is used by Wang et al. (16) to infer potential lncRNA-disease associations. A lncRNAs-disease association prediction method GBDTLRL2D based on Gradient Boosting Decision Tree and Logistic Regression is proposed by Duan et al. (17). The GCRFLDA, a prediction method based on graph convolution matrix completion, is proposed by Fan et al. (18). An end-to-end computational method based on graph attention network GANLDA is proposed by Lan et al. (19). A method called LRWRHLDA is proposed by Wang et al. (20). The LRWRHLDA designs a multi-layer network using six known heterogeneous networks, and uses Laplace normalized random walk and restart algorithm to predict. A dual attention network is proposed by Liu et al. (21).

For the association between miRNAs and lncRNAs, the LMI-INGI, based on interactome network and graphlet interaction, is proposed by Zhang et al. (22) to predict the lncRNAs-miRNAs associations. The NALMA is proposed by Zhang et al. (23) to use network distance analysis. The DWLMI proposed by Yang et al. (24). utilizes lncRNAs-miRNAs-disease-protein-drug diagram. The structural perturbation method SPMLMI is proposed by Xu et al. (25). A logical matrix factorization with neighborhood regularized, LMFNRLMI, is proposed by Liu et al. (26).

Advances in genomics and bioinformatics have facilitated the identification of miRNAs. The miRNAs have also been found to interact with a variety of drugs. It is possible to develop resistance or sensitivity during drug treatment because of the regulation of genes by miRNAs (27). For example, scientists have found that miRNA let-7b is resistant to the drug cisplatin (28). Cisplatin is an important drug in the treatment of many diseases, such as sarcoma. Cisplatin has also been reported to down-regulate miRNA let-7b expression, lead to up-regulation of Cyclin D1, and induce resistance to cisplatin. Similarly, miRNA Mir-106a is found to enhance the sensitivity of OVCAR3/CIS cells to cisplatin (29). Since both the increase and decrease of miRNA expression level can cause diseases, miRNA-targeted therapy drugs can be divided into miRNA mimics and miRNA inhibitors. Their aim is to induce gene silencing and selective up-regulation of proteins.

There are several public databases that collate miRNAs-drug relationships. For instance, the database of miRNAs-drug interactions, pharmaco-miR is developed by Rukov et al. (30) according to the interaction between miRNA target genes and

drug proteins. The database mTD is developed by Chen et al. (31) to collect information about the impact of miRNAs during drug treatment. The ncDR is developed by Dai et al. (32) to provide information of noncoding RNAs related to drug resistance. However, the known link between miRNAs and drug resistance is limited. Because biological experiments are time-consuming and expensive, it is necessary to develop computation-based methods to predict the potential association between miRNAs and drug resistance.

Different computational methods have been developed to identify and predict miRNAs-drug resistance. For example, an algorithm for predicting potential miRNAs-drug resistance associations through Bi-Random Walk (BiRW) is proposed by Xu et al. (33). The method SNMFSMMA based on symmetric non-negative matrix factorization and kronecker regularized least squares is developed by Zhao et al. (34) for prediction of small molecular-miRNAs association. The GCMDR proposed by Huang et al. (35) uses graph convolution to built potential factor model, learns the graph embedding feature of miRNAs and drugs, and expresses the problem of predicting miRNA-drug association as a link prediction problem involving two-miRNA-drug sensitivity associations, named LGCMDS, is proposed by Yu et al. (36). The MDIPA, a matrix factor-based method, is proposed by Jamali et al. (37) to predict the unknown interactions between miRNAs and drug resistance. Predicting associations between small molecular and microRNAs using functional similarity of miRNAs and multiple similarity measures of small molecular is proposed by Qu et al. (38). In addition, combined with clinical, chemical, and biological information, a method based on non-negative matrix factorization to predict miRNAs-small molecule relationships is developed by Luo et al. (39).

Although there are some studies on predictive tools for miRNAs-drug resistance associations, these methods cannot fully utilize the structure and semantics in heterogeneous networks to extract higher-quality information. At the same time, the accuracy and performance obtained by these methods need to be improved. To address these issue, a method for predicting miRNAs-drug resistance based on support vector machines SVMMDR is proposed in this paper. The SVMMDR considers the path information between nodes in heterogeneous networks. The hetesim measures the correlation between nodes of the same type or different types within a unified framework. At the same time, based on the search path between two nodes, the measure between node pairs is defined by following a sequence. The node information and path information in heterogeneous networks are integrated. The SVM is used to predict potential miRNAs-drug resistance associations. The contribution of our method mainly consists of the following parts:

- The SVMMDR introduces the concept of miRNA and drug groups. On this basis, a roaming network is

established. Walker is more inclined to choose the next node of the walk. The inclined diffusion feature is obtained by inclined restart random walk.

- The SVMMDR integrates node information and path information in heterogeneous networks. The data feature is obtained by combining the inclined diffusion feature and hetesim score.
- The SVMMDR algorithm improves prediction accuracy and has the highest AUC values when compared to existing algorithms.

## 2 Materials and methods

The miRNAs-drug resistance association data required in this paper are downloaded from ncDR database (32). The ncDR collected 5864 validated relationships between 145 compounds and 1039 ncRNAs (877 miRNAs and 162 lncRNAs) from approximately 900 published papers. We only need the correlation between miRNAs-drug resistance among them. After removing duplicate data, the 625 miRNAs, 85 drugs and 2301 miRNAs-drug resistance associations are obtained.

In this paper, an SVM-based method SVMMDR is proposed to predict the association between miRNAs-drug resistance. The SVMMDR integrates miRNAs-drug resistance association, miRNAs sequence similarity, drugs chemical structure similarity and other similarities. The path-based hetesim feature is obtained, and the concepts of miRNAs group and drug group are introduced to obtain the inclined diffusion feature through inclined random walk. Finally, the SVM algorithm is used to predict the association between miRNAs and drug resistance. This mainly includes the following steps:

- (1) The miRNAs-drug resistance association data set is downloaded from the ncDR, and the list of miRNAs and drugs, the matrix  $A$  of miRNAs-drug resistance association are obtained by de-duplication of the downloaded data. Then the gaussian interaction profile kernel similarity matrix of miRNAs GSM and of drug GSD are calculated.
- (2) The sequence of miRNA list is downloaded from miRBase database, and the miRNAs sequence similarity matrix SSM between miRNAs is calculated. The drug chemical structure similarity matrix ESD is obtained by using the published tool SimComp.
- (3) The miRNAs similarity network is obtained based on the GSM and SSM, and drugs similarity network is obtained based on the GSD and ESD.
- (4) The miRNA-resistance association network, miRNA similarity network, and drug similarity network are integrated to construct a heterogeneous network. In

the heterogeneous networks, the inclined diffusion feature are obtained based on the inclined random walk with restart. Then the low-dimensional inclined diffusion feature are obtained by using Singular Value Decomposition (SVD).

- (5) The hesitism scores for miRNA-drug pairs are calculated based on paths in the heterogeneous network.
- (5) The inclined diffusion feature and the hesitism score are combined to obtain the feature data set. The combined features are used in the SVM classifier to obtain the predicted scores for miRNAs-drug resistance. The flow of SVMMDR is shown in Figure 1.

## 2.1 Calculate Gaussian interaction profile kernel similarity

The matrix A of miRNAs-drug resistance association network is obtained. The number of rows of A is the number of miRNAs, and the number of columns of A is the number of drugs, as shown in the formula (1):

$$A(m_i, d_j) = \begin{cases} 1 & m_i \text{ is associated with } d_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where  $A(m_i, d_j) = 1$  indicates that there is a resistance between miRNA  $m_i$  and drug  $d_j$ .

For any given miRNA  $m_i$  and  $m_j$ , the gaussian interaction profile kernel similarity  $GSM(m_i, m_j)$  can be obtained based on A, as shown in the formula (2) and (3):

$$GSM(m_i, m_j) = \exp(-\delta_m \|A(i, :) - A(j, :)\|^2) \quad (2)$$

$$\delta_m = \delta'_m / \left( \frac{1}{nm} \sum_{i=1}^{nm} \|A(i, :)\|^2 \right) \quad (3)$$

where  $nm$  is the number of miRNAs and  $A(i, :)$  is the  $i_{th}$  row of the adjacency matrix A. The  $\delta_m$  is used to control the frequency band, it represents the normalized frequency band of Gaussian interaction profile kernel similarity based on the new frequency band parameter  $\delta'_m$ . The gaussian interaction profile kernel similarity between drugs can be obtained in the same way, represented by GSD, which is given by (4) and (5):

$$GSD(d_x, d_y) = \exp(-\delta_d \|A(:, x) - A(:, y)\|^2) \quad (4)$$

$$\delta_d = \delta'_d / \left( \frac{1}{nd} \sum_{x=1}^{nd} \|A(:, x)\|^2 \right) \quad (5)$$

where  $nd$  is the number of miRNAs and  $A(:, x)$  is the  $x_{th}$  col of the A.

## 2.2 Calculate miRNA sequence similarity and drug chemical structure similarity

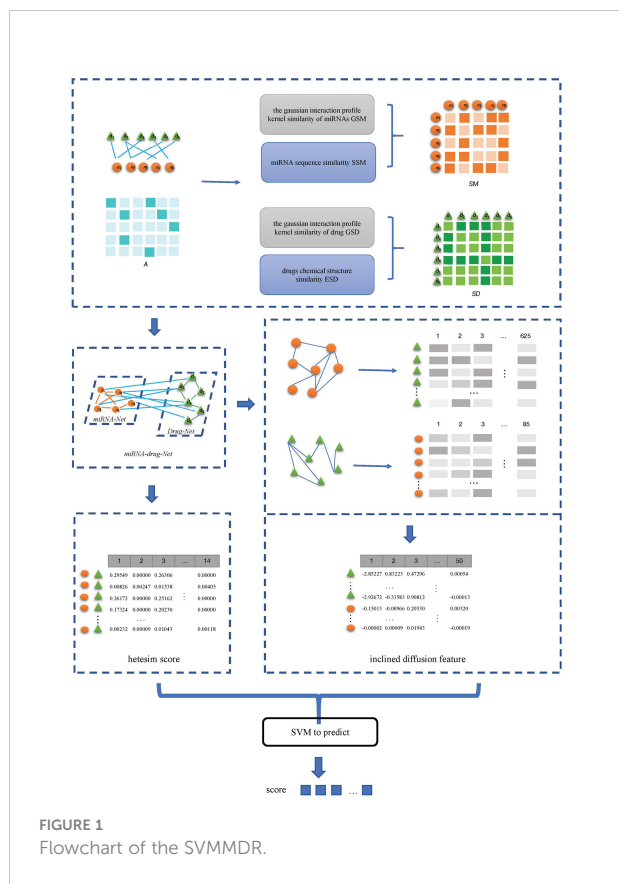
The sequences of relevant miRNAs are downloaded from the public database miRBase (<https://mirbase.org/>) (40). The miRBase database provides information including miRNAs sequence data, annotations, and predicted gene targets. The sequence similarity SSM between miRNAs is calculated as shown in the formula (6):

$$SSM(m_i, m_j) = 1 - \frac{Levenshtein(m_i, m_j)}{\text{len}(m_i) + \text{len}(m_j)} \quad (6)$$

$$0 \leq \text{Levenshtein}(m_i, m_j) \leq \text{len}(m_i) + \text{len}(m_j) \quad (7)$$

Where  $\text{len}(m_i)$  represents the length of miRNA  $m_i$  sequence,  $\text{len}(m_j)$  represents the length of miRNA  $m_j$  sequence,  $\text{Levenshtein}(m_i, m_j)$  is defined as the class editing distance of the transformation from  $m_i$  sequence to  $m_j$  sequence.

With the Kyoto Encyclopedia of Genes and Genomes (KEGG) database entry number corresponding to drugs in the DLREFD database as the parameter, the chemical structural



similarity matrix ESD between drugs is calculated using the SimComp tool (41).

### 2.3 Integer similarity

In this section, miRNAs similarity network and drugs similarity network are constructed. The miRNAs similarity network is expressed as SM. The SM is fused by SSM and GSM, which is given by (8):

$$SM(m_i, m_j) = \begin{cases} GSM(m_i, m_j) & \text{if } SSM(m_i, m_j) = 0 \\ SSM(m_i, m_j) & \text{otherwise} \end{cases} \quad (8)$$

Similarly, denote SD as the drug similarity network, which is fused by ESD and GSD, as follows:

$$SD(d_i, d_j) = \begin{cases} GSD(d_i, d_j) & \text{if } SSM(d_i, d_j) = 0 \\ SSD(d_i, d_j) & \text{otherwise} \end{cases} \quad (9)$$

### 2.4 Obtain low-dimensional network inclined diffusion features

A global heterogeneous network is constructed by integrating the association matrix A of miRNAs-drug resistance network, the similarity matrix SM of miRNA and the similarity matrix SD of drugs. The concepts of miRNAs group and drugs group are introduced to obtain miRNA weight matrix and drug weight matrix to construct roaming network. The restart random walk is used to calculate the inclined diffusion feature on the roaming network, and the high dimensional inclined diffusion feature are obtained. Then, the SVD is used to reduce the dimension of the high-dimensional inclined diffusion feature, and the low-dimensional inclined diffusion feature is obtained. The specific sub-steps are as follows:

#### 2.4.1 Building a heterogeneous network

The heterogeneous network  $G = (V, E)$  is constructed. The dimension of the matrix G is  $(nm + nd) * (nm + nd)$ , where  $nm$  and  $nd$  is the number of miRNAs and drugs, as shown in formula (10):

$$G = \begin{bmatrix} SM & A \\ A^T & SD \end{bmatrix} \quad (10)$$

where  $A^T$  is the transpose of A.

#### 2.4.2 Obtain the weight matrix

The drugs associated with the same miRNA are regarded as a drug group. If one miRNA with high similarity to this miRNA are associated with a drug in this drug group, this miRNA is

considered to have a potential association with other drugs in the drug group.

For example, for drug  $d_i$ , miRNAs associated with  $d_i$  are regarded as a miRNA group. If  $d_j$  with high similarity to  $d_i$  is associated with miRNA in this miRNA group, then it is assumed that  $d_j$  may be associated with other miRNAs in the miRNA group. Based on the above assumptions, miRNAs weight matrix  $W_{MM}$  of  $nd * nm$  dimension and drugs weight matrix  $W_{DD}$  of  $nm * nd$  dimension are obtained, as shown in the formula (11) and (12):

$$W_{MM}(d_i, m_j) = \frac{SS_M(d_i, m_j)}{\max_{1 \leq i \leq nd} \{SS_M(d_i, m_j)\}} \quad (11)$$

$$SS_M(d_i, m_j) = \sum_{m_k \in DM(d_i)} SM(m_k, m_j) \quad (12)$$

where  $DM(d_i) = \{m_k \mid \forall m_k \in \{if(A(m_k, d_i) = 1)\}, 1 \leq k \leq nm\}$  represents the miRNA group associated with the drug  $d_i$ .  $if(A(m_k, d_i) = 1, 1 \leq k \leq nm)$  represents that miRNA  $m_k$  is associated with drug  $d_i$ .  $SM(m_k, m_j)$  is the similarity between miRNA  $m_k$  and  $m_j$ .

The drugs weight matrix  $W_{DD}$  of  $nm * nd$  dimension can also be obtained:

$$W_{DD}(m_j, d_i) = \frac{SS_D(m_j, d_i)}{\max_{1 \leq j \leq nm} \{SS_D(m_j, d_i)\}} \quad (13)$$

$$SS_D(m_j, d_i) = \sum_{d_z \in DD(m_j)} SD(d_z, d_i) \quad (14)$$

where  $DD(m_j) = \{d_z \mid \forall d_z \in \{if(A(d_z, m_j) = 1)\}, 1 \leq z \leq nd\}$  represents the drug group associated with the drug  $m_j$ .  $if(A(d_z, m_j) = 1, 1 \leq z \leq nd)$  represents that drug  $d_z$  is associated with drug  $m_j$ .  $SM(d_z, m_j)$  is the similarity between drug  $d_z$  and  $m_j$ .

#### 2.4.3 Construct roaming network

When drug  $d_x$  is a walker, it walks on the miRNAs node network.  $T_D$  is the transition probability matrix of the roaming network. For any given miRNA  $m_i$  and  $m_j$ , denote  $T_D$  as the probability of  $m_i$  transferring to  $m_j$  during the walking process.

$$T_D(m_i, m_j) = \frac{ST_D(m_i, m_j)}{\sum_{n=1}^{nm} ST_D(m_n, m_j)} \quad (15)$$

$$ST_D(m_i, m_j) = \begin{cases} W_{DD}(m_j, d_x) & \text{If } W_{DD}(m_j, d_x) > 0 \\ SM(m_i, m_j) & \text{otherwise} \end{cases} \quad (16)$$

where,  $m_i$  is the current node of migration, and  $m_j$  is the next node. If the value of  $m_j$  and  $d_x$  in the weight matrix is not 0, it means that  $m_j$  and  $d_x$  have potential correlation, namely  $ST_D(m_i, m_j) = W_{DD}(m_j, d_x)$ . Otherwise, the probability of  $m_i$  transferring to  $m_j$  is related to miRNA similar matrix,  $ST_D(m_i, m_j) = SM(m_i, m_j)$ .

When miRNA  $m_y$  is a walker, it walks on the miRNAs node network. Denote  $T_M$  as the transition probability matrix of the roaming network. For any given drug  $d_i$  and  $d_j$ ,  $T_M$  represents the probability of  $d_i$  transferring to  $d_j$  during the walking process.

$$T_M(d_i, d_j) = \frac{ST_M(d_i, d_j)}{\sum_{n=1}^{nd} ST_M(d_n, d_j)} \tag{17}$$

$$ST_M(d_i, d_j) = \begin{cases} W_{MM}(d_j, d_x) & \text{If } W_{MM}(d_j, m_y) > 0 \\ SD(d_i, d_j) & \text{otherwise} \end{cases} \tag{18}$$

### 2.4.4 Obtain inclined diffusion feature by IIRWR

Based on the transfer probability matrix  $T_D$  and  $T_M$  obtained, the drugs inclined diffusion feature  $P_D = [P^1, P^2, P^3, \dots, P^x, \dots, P^{nd}]$  can be obtained by random walk, where  $P^x$  represents the  $nm$ -dimensional inclined diffusion feature of drug node  $d_x$ . Meanwhile, the miRNAs inclined diffusion feature  $P_M = [P^1, P^2, P^3, \dots, P^y, \dots, P^{mm}]$ , where  $P^y$  denotes the  $nd$ -dimensional inclined diffusion feature of miRNA node  $m_y$ . The  $nm$  and  $nd$  denote the number of miRNA nodes and drug nodes.

When the inclined diffusion feature  $P^x$  of drug node  $d_x$  is calculated, each step of the walking is faced with two choices: randomly selecting adjacent miRNA node or returning to the starting node. The walking process is shown in the equation (19):

$$P_{t+1}^x = (1 - r) \times T_D \times P_t^x + r \times P_0^x \tag{19}$$

$$P_0^x(m_i) = \frac{A(m_i, d_x)}{\sum_{n=1}^{nm} A(m_n, d_x)} \tag{20}$$

When the inclined diffusion feature  $P^y$  of miRNA node  $m_y$  is calculated, the walking process is shown as follows:

$$P_{t+1}^y = (1 - r) \times T_M \times P_t^y + r \times P_0^y \tag{21}$$

$$P_0^y(d_j) = \frac{A(m_y, d_j)}{\sum_{n=1}^{nd} A(m_y, d_n)} \tag{22}$$

Where  $r$  is the restart probability,  $P_t^x$  is a  $nd$ -dimensional transition probability vector of node  $m_y$ , and its  $k$ -th element represents the probability of accessing node  $k$  at  $t$  step,  $k \in \{1, 2, \dots, nd\}$ .  $P_0^y$  represents the initial migration probability vector of node  $m_y$ , and  $P_0^y(d_j)$  represents the initial migration probability of  $m_y$  visiting node  $d_j$ .

After several iterations, the difference between the two iterations of  $p^x$  and  $p^y$  is less than  $10^{-10}$ . The miRNA inclined diffusion feature  $P_M$  and the drug inclined diffusion feature  $P_D$  reach a stable state, and the final inclined diffusion feature is obtained.

### 2.4.5 Calculate the low-dimensional inclined diffusion feature

The more nodes in the heterogeneous network, the higher the feature dimension obtained by the inclined restart random walk. However, when the feature dimension is high, there will be data redundancy. The sample distribution of the high-dimension space is sparse. The SVD is used to reduce the dimension of the inclined diffusion feature.

Suppose that the  $m \times n$  dimensional matrix  $P$  can be decomposed by  $P = U\Sigma V^T$ , where  $U$  is a  $m \times m$ -dimensional matrix and  $V$  is a  $n \times n$ -dimensional matrix. The  $U$  and  $V$  are left singular vectors and the right singular vectors, both unitary matrices, that is,  $UU^T = I$ ,  $VV^T = I$ . The  $m \times n$  dimensional matrix  $\Sigma$  has values only on the main diagonal, and all other elements are zero. Every element along the main diagonal is called singular value. The singular values are arranged from largest to smallest, and the decrease is extremely fast. In many cases, the sum of the first 10% or even 1% of the singular values accounts for more than 99% of the total singular values. In other words, we can also use the largest  $d$  singular values and corresponding left and right singular vectors to approximate the matrix, as follows:

$$P_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \approx U_{m \times d} \Sigma_{d \times d} V_{d \times n}^T \tag{23}$$

where  $d$  is far less than  $n$ , and the low-dimensional feature vector  $X$  can be obtained by formula (24):

$$X = U_{m \times d} (\Sigma_{d \times d})^{1/2} \tag{24}$$

The SVD is performed on  $P_D$  and  $P_M$  respectively to obtain low-dimensional node feature matrix  $X_D$  and  $X_M$ .

## 2.5 Calculate the hetesim score

In heterogeneous networks, the types of nodes are different, and the relationship between nodes has various meanings. In order to obtain the correlation between different nodes, the hetesim scores are calculated (42). The hetesim is a path-based measurement method used to measure the correlation of objects (including objects of the same type or different types) in heterogeneous networks.

(1) The transition probability matrix  $I_{MD}$  from miRNA to drug,  $I_{DD}$  from drug to drug,  $I_{MM}$  from miRNA to miRNA are obtained as follows:

$$I_{MD}(m_i, d_j) = \frac{A(m_i, d_j)}{\sum_{k=1}^{nd} A(m_i, d_k)} \tag{25}$$

$$I_{DD}(d_i, d_j) = \frac{SD(d_i, d_j)}{\sum_{k=1}^{nd} SD(d_i, d_k)} \tag{26}$$

$$I_{MM}(m_i, m_j) = \frac{SM(m_i, m_j)}{\sum_{k=1}^{nm} SM(m_i, m_k)} \quad (27)$$

(2) The N is the node, and there are only miRNA and drug node. The path with length  $l$  between any two nodes is represented by  $\rho=N_1 N_2 \dots N_{l+1}$ , and the reachable probability matrix  $PM=I_{N_1} N_2 I_{N_2} N_3 \dots I_{N_l} N_{l+1}$ . Divide the path in half, get the  $PM_{\rho_L}$  and  $PM_{\rho_R}$ .

- when  $l$  is even,  $\rho_L = N_1 N_2 \dots N_{\frac{l}{2}+1}$ ,  $\rho_R = N_{\frac{l}{2}+1} N_{\frac{l}{2}+2} \dots N_{l+1}$ . The  $PM_{\rho_L}$  and  $PM_{\rho_R}$  is calculated.
  - When  $l$  is odd,  $\rho_{L_1} = N_1 N_2 \dots N_{\frac{l+1}{2}}$ ,  $\rho_{L_2} = N_1 N_2 \dots N_{\frac{l+1}{2}}$ ,  $\rho_{R_1} = N_{\frac{l+1}{2}} N_{\frac{l+1}{2}+1} \dots N_{l+1}$ ,  $\rho_{R_2} = N_{\frac{l+1}{2}} N_{\frac{l+1}{2}+1} \dots N_{l+1}$ .
- Then  $PM_{\rho_L} = \frac{PM_{\rho_{L_1}} + PM_{\rho_{L_2}}}{2}$ ,  $PM_{\rho_R} = \frac{PM_{\rho_{R_1}} + PM_{\rho_{R_2}}}{2}$ .

(3) The  $PM_{\rho_L}$  and  $PM_{\rho_R^{-1}}$  are calculated, where  $\rho_R^{-1}$  represents the reverse of  $\rho_R$ , for example, if  $\rho_R = MMMDD$ , the  $\rho_R^{-1} = DDDMM$ .

$$Hetesim(a, b | \rho) = \frac{PM_{\rho_L} (PM_{\rho_R^{-1}})^T}{\|PM_{\rho_L}\|_2 * \|PM_{\rho_R^{-1}}\|_2} \quad (28)$$

where  $Hetesim(a, b | \rho)$  represents the hetesim score of the node  $a$  reaching the node  $b$  through path  $\rho$ . As shown in Table 1, there are 14 different paths from a miRNA to a drug when the  $l < 5$ . So, the 14-dimensional hetesim feature between each miRNA-drug node pair in the heterogeneous network is obtained.

TABLE 1 The paths from a miRNA to a drug in the heterogeneous network when  $l < 5$ .

id	path	meaning
1	MMD	miRNA-miRNA-drug
2	MDD	miRNA-drug-drug
3	MMMD	miRNA-miRNA-miRNA-drug
4	MDMD	miRNA-drug-miRNA-drug
5	MMDD	miRNA-miRNA-drug-drug
6	MDDD	miRNA- drug-drug-drug
7	MMMMD	miRNA-miRNA-miRNA-miRNA-drug
8	MMMDD	miRNA-miRNA-miRNA-drug-drug
9	MMDMD	miRNA-miRNA-drug-miRNA-drug
10	MMDDD	miRNA-miRNA-drug-drug-drug
11	MDMMD	miRNA-drug-miRNA-miRNA-drug
12	MDMDM	miRNA-drug-miRNA- drug-drugdrug
13	MDDMD	miRNA-drug-drug-miRNA-drug
14	MDDDD	miRNA-drug-drug-drug-drug

## 2.6 Training the support vector machine classifier

Feature data are obtained by combining inclined diffusion feature and hetesim score. For each pair of drug and miRNA sample in the calculated Hetesim score matrix, the 50-dimensional inclined diffusion feature of the corresponding miRNA and drug are obtained respectively, and 114-dimensional feature is obtained. For example, sample drug  $d_i$  and miRNA  $m_j$ , the 14-dimensional HeteSim score of  $d_i - m_j$  pair is combined with the 50-dimensional inclined diffusion feature of the corresponding drug  $d_i$  and the 50-dimensional inclined diffusion feature of miRNA  $m_j$ , namely, the  $i$ -th row of the  $X_D$  and the  $j$ -th row of the matrix  $X_M$ , to obtain the 114-dimensional feature of drug  $d_i$  and miRNA  $m_j$ . The 114-dimension feature data of all sample pair are obtained by a similar method. The obtained feature data are used for SVM classifier to predict the miRNAs-drug resistance relationship.

The SVM is an effective classification method and has been widely used in the classification of biological data (43–45). The SVM can transform sample space into high-dimensional or even infinite-dimensional feature space (46). The goal of SVM is to find a hyperplane so that the sample points close to the hyperplane can have a larger distance. The steps of SVM for the algorithm are as follows:

(1) The kernel function  $K(x_i, x_j)$  and punish parameter  $C$  need to be selected first. The optimization problem is constructed and solved.

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \end{aligned} \quad (29)$$

where  $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$ . The punish function  $C = 64$ . The optimal solution is obtained as  $\alpha^* = \{\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*\}^T$  (2). A positive component of  $a^*$  is selected,  $0 \leq \alpha_j^* \leq C$ :

$$b^* = y_j - \sum_{i=1}^N \alpha_i^* y_i K(x_i, x_j) \quad (30)$$

(3) The decision function is constructed.

$$f(x) = \text{sign} \left( \sum_{i=1}^N \alpha_i^* y_i K(x_i, x_j) + b^* \right) \quad (31)$$

## 2.7 The SVMMDR algorithm

In this section, Algorithm 1 describes the implementation details of SVMMDR. In lines 2 to 15 of Algorithm 1, the low-dimensional inclined diffusion feature matrix  $X_{Ma}$  and  $X_{Da}$  are

obtained by using the inclined random walk with restart and SVD. The hetesim score between any two nodes in a heterogeneous network is obtained from lines 16 to 41. In lines 42 to 45, the combined features is obtained and used to train the SVM classifier. Then the final prediction score is obtained.

**Input:** miRNAs set, drugs set, The association matrix of the miRNA-drug resistance,  $A$ ;  
**Output:** The gaussian interaction profile kernel similarity matrixs,  $GSM$  and  $GSD$ . The miRNAs sequence similarity matrix,  $SSM$ . The chemical structural similarity matrix,  $ESD$ . The similarity matrix  $SM$  and  $SD$ . Prediction score.

Construct the adjacency matrix  $G$ ;

Obtain the weight matrix  $W_{MM}$  and  $W_{DD}$ ;

Initialize the global transition probability matrix  $T_D$  and  $T_M$ ;

Initialize the transition probability vector for each node

$$P_0^x(m_i) = \frac{A(m_i, d_x)}{\sum_{n=1}^{mn} A(m_n, d_x)}, P_0^y(d_j) = \frac{A(m_y, d_j)}{\sum_{n=1}^{nd} A(m_y, d_j)}$$

**while**  $P_{t+1}^x - P_t^x > 10^{-10}$  **do**:

Obtain the updated probability vector:

$$P_{t+1}^y = (1 - r) * T_D * P_t^x + r * P_0^y$$

**end while**

$$P_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \approx U_{m \times d} \Sigma_{d \times d} V_{d \times n}^T X / U_{n \times d} \Sigma_{d \times d}^{1/2}$$

Get low-dimensional inclined diffusion feature  $X_M$  and  $X_D$

Calculate  $I_{MD}(m_i, d_j)$ ,  $I_{DD}(d_i, d_j)$ ,  $I_{MM}(m_i, m_j)$

$$I_{MD}(m_i, d_j) = \frac{A(m_i, d_j)}{\sum_{k=1}^{nd} A(m_i, d_k)}$$

$$I_{DD}(d_i, d_j) = \frac{SD(d_i, d_j)}{\sum_{k=1}^{nd} SD(d_i, d_k)}$$

$$I_{MM}(m_i, m_j) = \frac{SM(m_i, m_j)}{\sum_{k=1}^{nm} SM(m_i, m_k)}$$

**for**  $l=1 \rightarrow 5$  **do**

Divide the path into two parts.

**if**  $l \% 2 == 0$  **then**

$$\rho_L = N_1 N_2 \dots N_{\frac{l}{2}+1}$$

$$\rho_R = N_{\frac{l}{2}+1} N_{\frac{l}{2}+2} \dots N_{l+1}$$

$$PM_{\rho_L} = I_{N_1 N_2} I_{N_2 N_3} \dots I_{N_{\frac{l}{2}+1} N_{\frac{l}{2}+2}}$$

$$PM_{\rho_R} = I_{N_{\frac{l}{2}+1} N_{\frac{l}{2}+2}} I_{N_{\frac{l}{2}+2} N_{\frac{l}{2}+3}} \dots I_{N_l N_{l+1}}$$

**end if**

**if**  $l \% 2 != 0$  **then**

$$\rho_{L_1} = N_1 N_2 \dots N_{\frac{l+1}{2}}$$

$$\rho_{L_2} = N_1 N_2 \dots N_{\frac{l+3}{2}}$$

$$\rho_{R_1} = N_{\frac{l+1}{2}} N_{\frac{l+1}{2}+1} \dots N_{l+1}$$

$$\rho_{R_2} = N_{\frac{l+3}{2}} N_{\frac{l+3}{2}+1} \dots N_{l+1}$$

$$PM_{\rho_{L_1}} = I_{N_1 N_2} I_{N_2 N_3} \dots I_{N_{\frac{l+1}{2}} N_{\frac{l+1}{2}+1}}$$

$$PM_{\rho_{L_2}} = I_{N_1 N_2} I_{N_2 N_3} \dots I_{N_{\frac{l+1}{2}} N_{\frac{l+3}{2}}}$$

$$PM_{\rho_{R_1}} = I_{N_{\frac{l+1}{2}} N_{\frac{l+1}{2}+1}} I_{N_{\frac{l+1}{2}+1} N_{\frac{l+1}{2}+2}} \dots I_{N_l N_{l+1}}$$

$$PM_{\rho_{R_2}} = I_{N_{\frac{l+3}{2}} N_{\frac{l+3}{2}+1}} I_{N_{\frac{l+3}{2}+1} N_{\frac{l+3}{2}+2}} \dots I_{N_l N_{l+1}}$$

$$PM_{\rho_L} = \frac{PM_{\rho_{L_1}} + PM_{\rho_{L_2}}}{2}$$

$$PM_{\rho_R} = \frac{PM_{\rho_{R_1}} + PM_{\rho_{R_2}}}{2}$$

**end if**

$$Hetesim(a, b | \rho) = \frac{PM_{\rho_L} (PM_{\rho_R}^{-1})^T}{\|PM_{\rho_L}\|_2 * \|PM_{\rho_R}^{-1}\|_2}$$

**end for**

Combined with the inclined diffusion feature and HeteSim score to get the data set

$$D_{train} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

$$D_{test} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

Use  $D_{train}$  to train the Support Vector Machines (SVM) as classifier

ALGORITHM 1  
 SVMMDR algorithm.

### 3 Result and discussion

#### 3.1 Data sets

The miRNAs-drug resistance association data are downloaded from ncDR database. After deduplication, 85 drugs and 625 miRNAs are obtained, and 2301 miRNAs-drug resistance known association are obtained, all as positive samples. Negative samples are randomly selected from unknown associations with three times the number of positive samples. The final sample dataset is constructed from 2301 positive samples and 6903 negative samples.

#### 3.2 Performance measures

The 10-fold Cross-Validation(10-CV) is performed to evaluate the performance of SVMMDR. The process of 10-CV is as follows: the sample data is equally divided into 10 groups. The 9 group of data is used as the training set, and the remaining group is used as the validation set. After ten times of the above process, each of the 10 groups in turn is used as a validation data to obtain 10 performance results. The final performance evaluation is obtained by averaging the 10 performance results. Multiple measures are used to evaluate performance, such as the area under the receiver operating characteristic curves (AUC), recall (REC), accuracy (ACC), F1-score and Matthews Correlation Coefficient (MCC). They can be presented as below:

$$Recall = \frac{TP}{TP + FN}, \tag{32}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{33}$$



$$F1 - Score = \frac{2 \times TP}{2TP + FP + FN}, \tag{34}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{35}$$

where the *TP* is the number of samples that are correctly classified as positive, the *FP* is the number of samples that are misclassified as positive, the *TN* represents the number of samples that are correctly classified as negative, and the *FN* is the number of samples that are misclassified as negative.

### 3.3 Performance comparison with existing machine learning methods

In order to reflect the performance of SVM, the proposed SVMMDR methods will be compared with the following solution, including using logistic regression (LR) as a classifier,

the use of random forests (RF) used as a classifier, K nearest neighbor (KNN) as a classifier. The same features of the same training sample are used to train the corresponding classifiers. To get performance, the 10-fold cross-validation is applied. For KNN classifier, the 10 nearest neighbors and leaf size of 20 point is used. The RF builds a number of decision tree classifiers trained on a set of randomly selected samples of the benchmark to improve the performance. For LR, the maxiter and tol parameters are optimized to 500 and 0.001, respectively.

Figure 2 indicates the ROC curves of SVMMDR using other classifiers. The AUC of SVMMDR, KNN, RF and LR are 0.978, 0.939, 0.892 and 0.948. Furthermore, Table 2 shows the values of performance measures such as ACC, Pre, Recall, F1-score, MCC. The results show that the AUC value obtained by SVMMDR is the highest. The value of performance measure is also better than other classifiers. The SVM classifier can achieve effective classification by mapping features to higher dimensional space through kernel function changes. At the same time, the optimal solution is obtained with constraints, which can make the classification more accurate.

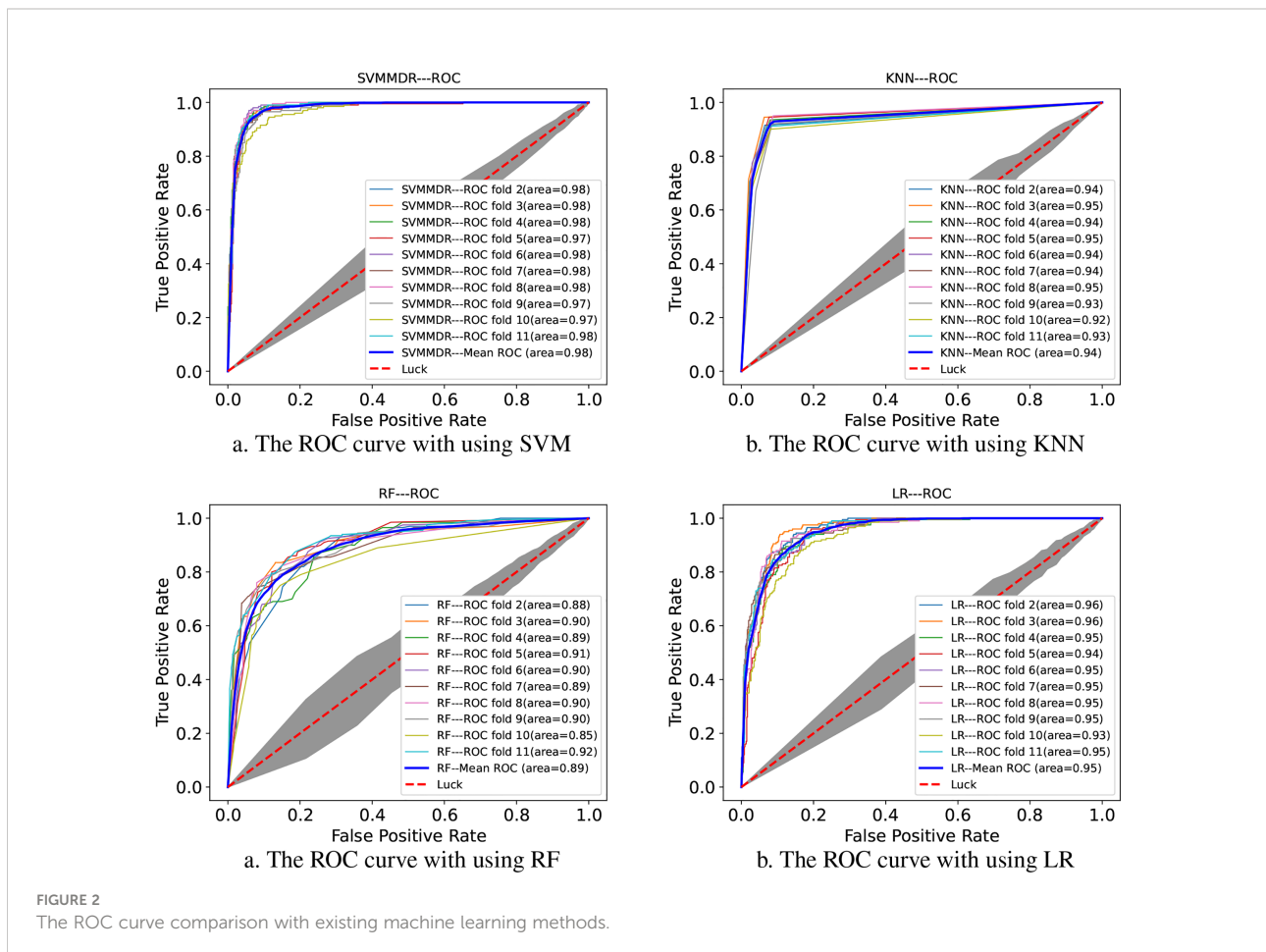


TABLE 2 Performance comparison of existing machine learning methods.

Method	ACC	Pre	RECALL	F1-score	MCC
SVMMMDR	0.9393	0.8705	0.8905	0.8800	0.8399
RF	0.8381	0.81733	0.4625	0.5704	0.5236
LR	0.8925	0.8020	0.7565	0.7785	0.7082
KNN	0.9080	0.8936	0.7175	0.7957	0.7448

### 3.4 Performance comparison with different topological features

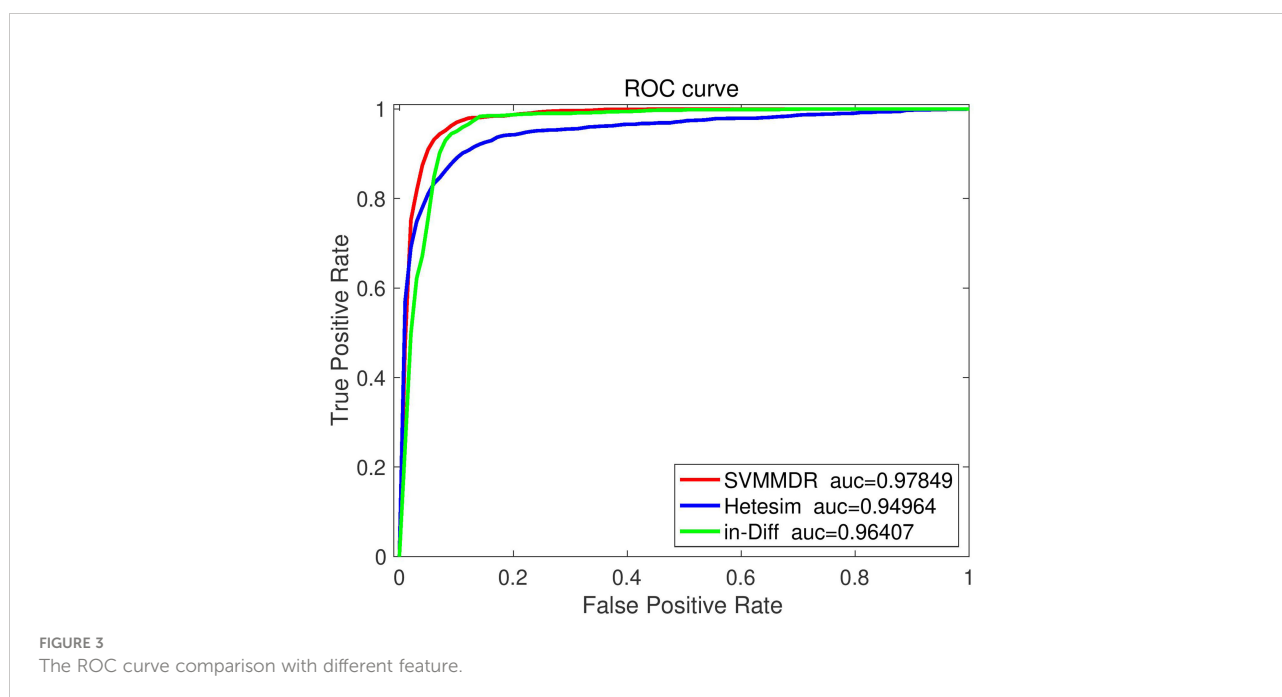
To demonstrate the advantages of combining features in SVMMMDR, different feature groups (hetesim+ inclined diffusion feature, hetesim feature, and inclined diffusion feature) are used for comparison experiment. The comparison results are shown in Figures 3 and 4. Denote “SVMMMDR”, “Hetesim” and “in-Diff” as the combination feature, hetesim feature and inclined diffusion feature. Figure 3 shows the ROC curves of different feature groups. It can be seen that the combination of hetesim and inclined diffusion obtained a higher AUC than the two separate feature, and the AUC obtained by inclined diffusion feature alone is higher than that obtained by hetesim alone. Figure 4 represents the performance achieved for the different feature groups. It can also see that the combination of the two features has best performance. Although the AUC of inclined diffusion feature reaches 0.96, the Pre, F1 and MCC are all relatively low. The combination of inclined diffusion feature and hetesim feature can solve this problem and improve performance.

### 3.5 Performance comparison with existing methods

To further illustrate the superiority of the proposed method, the SVMMMDR is compared with existing miRNA-resistance prediction algorithms, such as GCMDR, MDIPA and BiRWMD, all of which use the sample set in this paper. Performance measures are obtained by performing 10-fold cross-validation.

GCMDR (35): Data from multiple data sources are fused. The latent factor method are constructed using graph convolution to learn the graph embedding feature of miRNAs and drugs, and end-to-end prediction method are built.

MDIPA (37): The identification of potential miRNAs-drug interactions is seen as a matrix completion problem, the unknown associations are predicted based on weighted non-negative matrix factorization. The path-based miRNAs similarity matrix and drugs similarity matrix based on drugs structure information are obtained, which are combined with extracted drugs and miRNAs neighbor information for prediction.



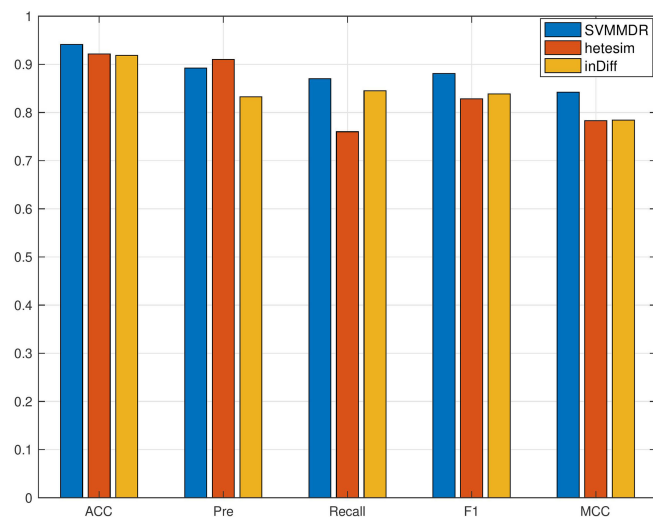


FIGURE 4 The performance comparison with different feature.

BiRW-WD (33): The multiple similarity networks and miRNAs-drugs association network are integrated to construct a heterogeneous network. In the heterogeneous networks, the bi-directional random walk (BiRW) are used to predict potential miRNAs-drug effect associations.

Figure 5 illustrates the comparison results. It can be seen that the proposed SVMMDR method achieves the best performance. The reasons are as follows: (1) The drug group and miRNA group are introduced. When restart random walk is used to obtain

diffusion feature, the walker is more inclined to select the node of the next walk. The inclined diffusion feature contribute to the prediction accuracy. (2) The hetesim score is obtained from the path information of two nodes in the heterogeneous network. Regardless of the same or different node types, the hetesim measures their correlation within a unified framework. At the same time, according to the search path between two nodes, the measure between node pairs is defined by following a sequence. (3) The SVM with high accuracy is used as the classifier.

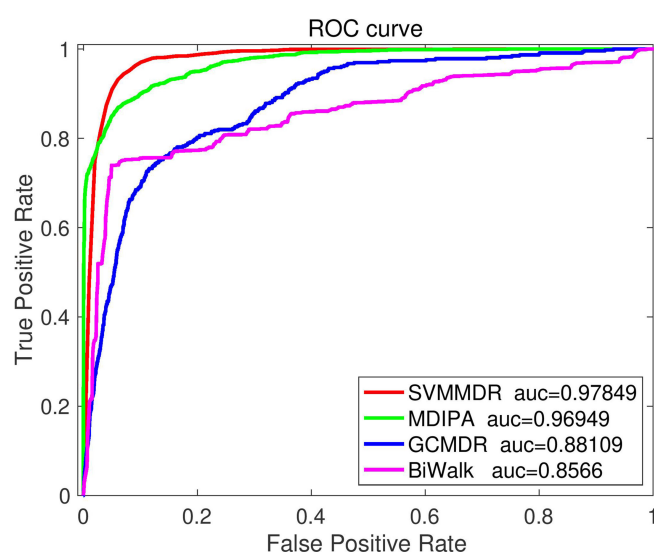


FIGURE 5 The ROC curve comparison with existing methods.

TABLE 3 The top 20 predicted miRNA related to 5-FU.

miRNA	PMID	miRNA	PMID
miR-3665	25117811	let-7b	22382630
miR-3619	25117811	miR-99a	24304648
miR-3141	25117811	miR-4465	25117811
miR-1234	25117811	miR-4484	25117811
miR-1275	25356050	miR-887	25117811
miR-29b-2	23167930	miR-4750	25117811
miR-423	25117811	miR-221	25117811
miR-107	22382630	miR-191	24304648
miR-296	20485139	miR-23a	24249161
miR-483	24304648	miR-21	24275137

### 3.6 Case study

In order to illustrate the effectiveness of the proposed method, we present a case study of the drug 5-fluorouracil (5-FU). The 5-FU is an antimetabolite drug widely used in cancer treatment, especially colorectal cancer (CRC) Longley et al. (47). There are 244 miRNAs related to 5-FU in ncDR database. We remove these associations in the association matrix A and use the rest as test data. The SVMMDR algorithm proposed in this paper is used for prediction, and 174 miRNAs with prediction scores greater than 0.95 are obtained. For the top 20 predicted miRNAs, we verify whether predicted miRNAs-drug resistance associations are confirmed by searching the PubMed literature. Table 3 indicates the miRNAs and the PMIDs of publications mentioning the association between miRNAs and 5-FU. For example, miR-21 expression levels are confirmed to lead to 5-Fluorouracil resistance Tomimaru et al. (48). The miR-23a enhances 5-FU resistance in microsatellite instability (MSI) CRC cells through targeting ABCF1 Li et al. (49).

## 4 Conclusion

More and more evidence indicates that miRNA expression level is related to drug resistance, affecting the therapeutic effect of disease. Predicting the association between miRNA-drug resistance can help to select more appropriate drugs for clinical treatment and promote the cure of disease. However, there are also very few computation-based predictive tools for miRNA- drug resistance.

Therefore, in this paper, a method based on the Support Vector Machines to predict the relationship between MiRNA and Drug Resistance (SVMMDR) is proposed. The SVMMDR integrates miRNAs-drug resistance association, miRNAs

sequence similarity, drug chemical structure similarity and other similarities, extracts path-based hetesim features, and obtains inclined diffusion features through inclined restart random walk. The machine learning algorithm SVM is used to predict the association between miRNAs and drug resistance.

The 10-fold cross-validation is used to assess the performance of SVMMDR. The area under the ROC curve AUC is used as a measure of performance. The AUC of SVMMDR reaches 0.978. The results show that SVMMDR has a significant performance advantage.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.jianglab.cn/ncDR/>.

## Author contributions

TD, ZFK, and LD conceived this work and designed the experiments. TD and ZFK carried out the experiments. TD and ZFK collected the data and analyzed the results. TD, ZFK and LD wrote, revised, and approved the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported in part by the National Natural Science Foundation of China under Grants Nos. 62072477, 61309027, 61702562 and 61702561, the Hunan Provincial Natural Science Foundation of China under Grants No.2018JJ3888, the Scientific Research Fund of Hunan Provincial Education Department under Grant No.18B197, the

National Key R&D Program of China under Grant No.2018YFB1700200, the Open Research Project of Key Laboratory of Intelligent Information Perception and Processing Technology (Hunan Province) under Grant No.2017KF01, the Hunan Key Laboratory of Intelligent Logistics Technology 2019TP1015.

## Acknowledgments

We would like to thank the Experimental Center of School of Computer and Information Engineering, Central South University of Forestry and Technology, for providing computing resources.

## References

- Xiang J, Zhang J, Zheng R, Li X, Li M. NIDM: network impulsive dynamics on multiplex biological network for disease-gene prediction. *Briefings Bioinf* (2021) 22. doi: 10.1093/bib/bbab080
- Xiang J, Meng X, Wu F-X, Li M. HyMM: Hybrid method for disease-gene prediction by integrating multiscale module structures. *bioRxiv* (2021) 23. doi: 10.1101/2021.04.30.442111
- Xiang J, Zhang N-R, Zhang J-S, Lv X-Y, Li M. PrGeFNE: Predicting disease-related genes by fast network embedding. *Methods* (2021) 192:3–12. doi: 10.1016/j.ymeth.2020.06.015
- Yu L, Zheng Y, Gao L. miRNA–disease association prediction based on meta-paths. *Briefings Bioinf* (2022) 23. doi: 10.1093/bib/bbab571
- Ding Y, Tian L-P, Lei X, Liao B, Wu F-X. Variational graph auto-encoders for miRNA-disease association prediction. *Methods* (2021) 192:25–34. doi: 10.1016/j.ymeth.2020.08.004
- Guo L, Shi K, Wang L. MLPMDA: Multi-layer linear projection for predicting miRNA- disease association. *Knowledge-Based Syst* (2021) 214:106718. doi: 10.1016/j.knsys.2020
- Zhong T, Li Z, You Z-H, Nie R, Zhao H. Predicting miRNA–disease associations based on graph random propagation network and attention network. *Briefings Bioinf* (2022). doi: 10.1093/bib/bbab589
- Jin C, Shi Z, Lin K, Zhang H. Predicting miRNA-disease association based on neural inductive matrix completion with graph autoencoders and self-attention mechanism. *Biomolecules* (2022) 12. doi: 10.3390/biom12010064
- Dai Q, Wang Z, Liu Z, Duan X, Song J, Guo M. Predicting miRNA-disease associations using an ensemble learning framework with resampling method. *Briefings Bioinf* (2022) 23:bbab543. doi: 10.1093/bib/bbab543
- Zhu Q, Fan Y, Pan X. Fusing multiple biological networks to effectively predict miRNA-disease associations. *Curr Bioinf* (2021) 16:371–84. doi: 10.2174/1574893615999200715165335
- Yan C, Duan G, Li N, Zhang L, Wu F-X, Wang J. PDMDA: predicting deep-level miRNA–disease associations with graph neural networks and sequence features. *Bioinformatics* (2022) 508:2226–34. doi: 10.1093/bioinformatics/btac077
- Liu D, Huang Y, Nie W, Zhang J, Deng L. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGBoost. *BMC Bioinf* (2021) 22:219. doi: 10.1186/s12859-021-04135-2
- Han G, Kuang Z, Deng L. MSCNE:predict miRNA-disease associations using neural network based on multi-source biological information. *IEEE/ACM Trans Comput Biol Bioinf* (2021) 1–1. doi: 10.1109/TCBB.2021.3106006
- Yu N, Liu Z-P, Gao R. Predicting multiple types of MicroRNA-disease associations based on tensor factorization and label propagation. *Comput Biol Med* (2022) 146:105558. doi: 10.1016/j.compbiomed.2022.105558
- Wang M-N, You Z-H, Wang L, Li L-P, Zheng K. LDGRNMF: LncRNA-disease associations prediction based on graph regularized non-negative matrix factorization. *Neurocomputing* (2021) 424:236–45. doi: 10.1016/j.neucom.2020.02.062
- Wang L, Xiao Y, Li J, Feng X, Li Q, Yang J. IIRWR: Internal inclined random walk with restart for LncRNA-disease association prediction. *IEEE Access* (2019) 7:54034–41. doi: 10.1109/ACCESS.2019.2912945
- Duan T, Kuang Z, Wang J, Ma Z. GBDTLRL2D: Predicts LncRNA-disease associations using MetaGraph2Vec and K-means based on heterogeneous network. *Front Cell Dev Biol* (2021) 9:753027. doi: 10.3389/fcell.2021.753027
- Fan Y, Chen M, Pan X. GCRFLDA: scoring LncRNA-disease associations using graph convolution matrix completion with conditional random field. *Briefings Bioinf* (2021) 23. doi: 10.4231093/bib/bbab361.Bbab361
- Lan W, Wu X, Chen Q, Peng W, Wang J, Chen YP. GANLDA: Graph attention network for LncRNA-disease associations prediction. *Neurocomputing* (2022) 469:384–93. doi: 10.1016/j.neucom.2020.09.094
- Wang L, Shang M, Dai Q, He P-A. Prediction of lncRNA-disease association based on a laplace normalized random walk with restart algorithm on heterogeneous networks. *BMC Bioinf* (2022) 23:5. doi: 10.1186/s12859-021-04538-1
- Liu Y, Yu Y, Zhao S. Dual attention mechanisms and feature fusion networks based method for predicting LncRNA-disease associations. *Interdiscip Sciences: Comput Life Sci* (2022) 14:358–71. doi: 10.1007/s12539-021-00492-x
- Zhang L, Liu T, Chen H, Zhao Q, Liu H. Predicting lncRNA–miRNA interactions based on interactome network and graphlet interaction. *Genomics* (2021) 113:874–80. doi: 10.1016/j.ygeno.2021.02.002
- Zhang L, Yang P, Feng H, Zhao Q, Liu H. Using network distance analysis to predict lncRNA–miRNA interactions. *Interdiscip Sciences: Comput Life Sci* (2021) 13:535–45. doi: 10.1007/s12539-021-00458-z
- Yang L, Li L-P, Yi H-C. DeepWalk based method to predict lncRNA-miRNA associations via lncRNA-miRNA-disease-protein-drug graph. *BMC Bioinf* (2022) 22:621. doi: 10.1186/s12859-022-04579-0
- Xu M, Chen Y, Lu W, Kong L, Fang J, Li Z, et al. SPMLMI: predicting lncRNA–miRNA interactions in humans using a structural perturbation method. *PeerJ* (2021) 9:e11426. doi: 10.7717/peerj.11426
- Liu H, Ren G, Chen H, Liu Q, Yang Y, Zhao Q. Predicting lncRNA–miRNA interactions based on logistic matrix factorization with neighborhood regularized. *Knowledge-Based Syst* (2020) 191:105261. doi: 10.1016/j.knsys.2019.105261
- Li Z, Rana TM. Therapeutic targeting of microRNAs: current status and future challenges. *Nat Rev Drug Discovery* (2014) 13:622–38. doi: 10.1038/nrd4359
- Guo Y, Yan K, Fang J, Qu Q, Zhou M, Chen F. Let-7b expression determines response to chemotherapy through the regulation of cyclin D1 in glioblastoma. *J Exp Clin Cancer Res* (2013) 32:1–10. doi: 10.1186/1756-9966-32-41
- Li H, Xu H, Shen H. microRNA-106a modulates cisplatin sensitivity by targeting PDCC4 in human ovarian cancer cells. *Oncol Lett* (2014) 7:183–8. doi: 10.3892/ol.2013.1644
- Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N. PharmacomiR: linking microRNAs and drug effects. *Briefings Bioinform* (2014) 15:648–59. doi: 10.1093/bib/bbs082

## 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

31. Chen X, Xie W-B, Xiao P-P, Zhao X-M, Yan H. mTD: a database of microRNAs affecting therapeutic effects of drugs. *J Genet Genomics* (2017) 44:269–71. doi: 10.1016/j.jgg.2017.04.003
32. Dai E, Yang F, Wang J, Zhou X, Song Q, An W, et al. ncDR: a comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics* (2017) 33:4010–1. doi: 10.1093/bioinformatics/btx523
33. Xu P, Wu Q, Rao Y, Kou Z, Fang G, Liu W, et al. Predicting the influence of MicroRNAs on drug therapeutic effects by random walking. *IEEE Access* (2020) 8:117347–53. doi: 10.1109/ACCESS.2020.3004512
34. Zhao Y, Chen X, Yin J, Qu J. SNMFSSMA: using symmetric nonnegative matrix factorization and kronecker regularized least squares to predict potential small molecule-microRNA association. *RNA Biol* (2020) 17:281–91. doi: 10.1080/15476286.2019.1694732
35. Huang Y-a, Hu P, Chan KC, You Z-H. Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* (2020) 36:851–8. doi: 10.1093/bioinformatics/btz621
36. Yu S, Xu H, Li Y, Liu D, Deng L. (2021). LGCMDS: Predicting miRNA-drug sensitivity based on light graph convolution network, in: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. New York: IEE. pp. 217–22. doi: 10.1109/BIBM52615.2021.9669566
37. Jamali AA, Kusalik A, Wu F-X. MDIPA: a microRNA–drug interaction prediction approach based on non-negative matrix factorization. *Bioinformatics* (2020) 36:5061–7. doi: 10.1093/bioinformatics/btaa577
38. Qu J, Chen X, Sun Y-Z, Zhao Y, Cai S-B, Ming Z, et al. In silico prediction of small molecule-miRNA associations based on the HeteSim algorithm. *Mol Therapy-Nucleic Acids* (2019) 14:274–86. doi: 10.1016/j.omtn.2018.12.002
39. Luo J, Shen C, Lai Z, Cai J, Ding P. Incorporating clinical, chemical and biological information for predicting small molecule-microRNA associations based on non-negative matrix factorization. *IEEE/ACM Trans Comput Biol Bioinf* (2021) 18:2535–45. doi: 10.1109/TCBB.2020.2975780
40. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* (2019) 47:D155–62. doi: 10.1093/nar/gky1141
41. Hattori M, Tanaka N, Kanehisa M, Goto S. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res* (2010) 38:W652–6. doi: 10.1093/nar/gkq367
42. Shi C, Kong X, Huang Y, Philip SY, Wu B. Hetesim: A general framework for relevance measure in heterogeneous networks. *IEEE Trans Knowledge Data Eng* (2014) 26:2479–92. doi: 10.1109/TKDE.2013.2297920
43. Zou Y, Wu H, Guo X, Peng L, Ding Y, Tang J, et al. MK-FSVM-SVDD: a multiple kernel-based fuzzy SVM model for predicting DNA-binding proteins via support vector data description. *Curr Bioinf* (2021) 16:274–83. doi: 10.2174/1574893615999200607173829
44. Wang M, Liang Y, Hu Z, Chen S, Shi B, Heidari AA, et al. Lupus nephritis diagnosis using enhanced moth flame algorithm with support vector machines. *Comput Biol Med* (2022) 145:105435. doi: 10.1016/j.combiomed.2022.105435
45. Ao C, Yu L, Zou Q. Prediction of bio-sequence modifications and the associations with diseases. *Briefings Funct Genomics* (2020) 20:1–18. doi: 10.1093/bfgp/ela023
46. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. "Support vector machines," in *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, July–Aug. 1998, doi: 10.1109/5254.708428
47. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* (2003) 3:330–8. doi: 10.1038/nrc1074
48. Tomimaru Y, Eguchi H, Nagano H, Wada H, Tomokuni A, Kobayashi S, et al. MicroRNA-21 induces resistance to the anti-tumour effect of interferon- $\alpha$ /5-fluorouracil in hepatocellular carcinoma cells. *Br J Cancer* (2010) 103:1617–26. doi: 10.1038/sj.bjc.6605958
49. Li X, Li X, Liao D, Wang X, Wu Z, Nie J, et al. Elevated microRNA-23a expression enhances the chemoresistance of colorectal cancer cells with microsatellite instability to 5-fluorouracil by directly targeting ABCF1. *Curr Protein Pept Sci* (2015) 16:301–9. doi: 10.2174/138920371604150429153309