BMC Bioinformatics

RESEARCH Open Access

CDPMF-DDA: contrastive deep probabilistic matrix factorization for drug-disease association prediction

Xianfang Tang[1], Yawen Hou[1], Yajie Meng[1], Zhaojing Wang[1], Changcheng Lu[2], Juan Lv[3], Xinrong Hu[1], Junlin Xu[4*] and Jialiang Yang[5*]

*Correspondence: xjl@hnu.edu.cn; yangjl@geneis.cn

[1] School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan 430200, China
[2] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
[3] College of Traditional Chinese Medicine, Changsha Medical University, Changsha 410000, China
[4] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China
[5] Geneis Beijing Co., Ltd, Beijing 100102, China

Abstract

The process of new drug development is complex, whereas drug-disease association (DDA) prediction aims to identify new therapeutic uses for existing medications. However, existing graph contrastive learning approaches typically rely on single-view contrastive learning, which struggle to fully capture drug-disease relationships. Subsequently, we introduce a novel multi-view contrastive learning framework, named CDPMF-DDA, which enhances the model's ability to capture drug-disease associations by incorporating diverse information representations from different views. First, we decompose the original drug-disease association matrix into drug and disease feature matrices, which are then used to reconstruct the drug-disease association network, as well as the drug-drug and disease-disease similarity networks. This process effectively reduces noise in the data, establishing a reliable foundation for the networks produced. Next, we generate multiple contrastive views from both the original and generated networks. These views effectively capture hidden feature associations, significantly enhancing the model's ability to represent complex relationships. Extensive cross-validation experiments on three standard datasets show that CDPMF-DDA achieves an average AUC of 0.9475 and an AUPR of 0.5009, outperforming existing models. Additionally, case studies on Alzheimer's disease and epilepsy further validate the model's effectiveness, demonstrating its high accuracy and robustness in drug-disease association prediction. Based on a multi-view contrastive learning framework, CDPMF-DDA is capable of integrating multi-source information and effectively capturing complex drug-disease associations, making it a powerful tool for drug repositioning and the discovery of new therapeutic strategies.

Keywords: Drug-disease association prediction, Contrastive learning, Matrix factorization, Multiple contrastive views

Introduction

The conventional method of drug discovery is costly and laborious [1–3]. Many drug candidates that appear promising in early trials often fail in later stages due to toxicity or inefficacy [4]. In the period from 2015 to 2020, a total of only 245 drugs secured approval

from the US FDA's Center for Drug Evaluation and Research [5]. By contrast, DDA prediction leverages advancements in sequencing technology to repurpose approved drugs for new therapeutic applications beyond their original indications [6, 7]. This strategy not only mitigates risks but also significantly reduces development costs and time [8, 9]. Consequently, DDA prediction holds substantial potential and offers numerous benefits, making it an invaluable approach in drug discovery.

Matrix factorization methods have been extensively utilized in bioinformatics research [10–13]. In particular, Zhang et al. described a method for matrix factorization that includes manifold regularization, combining manifold regularization with drug feature similarity to predict drug associations [14]. Sadeghi et al. developed NMFDR, which is a complex drug-disease interaction network that integrates drug-disease associations along with disease and drug similarities, predicting the scores for unidentified drug-disease pairs was made possible by an advanced non-negative matrix factorization method [15]. Wang et al. created WNMFDDA, which uses collaborative non-negative matrix factorization with weighted graph regularization to predict interactions pertaining to drugs and diseases [16]. Yang et al. offered the MSBMF method, where the matrix representing drug-disease associations is divided into separate feature matrices for drugs and diseases using multi-similarity bilinear matrix factorization, and these matrices subsequently help in predicting potential drug-disease associations [17]. These methods map various object features to a low-dimensional space through matrix factorization, extracting latent features from complex data to predict interactions between different entities. Matrix factorization excels at extracting latent features from intricate data, revealing the structural relationships within the data [18]. However, when facing high-dimensional data or lower data quality, matrix-based methods may encounter performance issues [19].

Graph neural networks have achieved significant progress in recent years, establishing themselves as a robust approach for graph representation learning with applications in social networks, chemistry, and biology [20, 21]. Within the domain of DDA prediction, Yu et al. brought forward LAGCN, which combines known drug-disease associations and similarities between drugs and diseases, resulting in heterogeneous networks. LAGCN uses graph convolution operations to create embeddings, then combines these from multiple convolutional layers into a final representation via an attention mechanism [22]. Similarly, Zhao et al. unveiled DDAGDL, a geometric deep learning scheme that was founded on heterogeneous information networks (HINs). This framework forms a heterogeneous network by integrating complex biological information, utilizes geometric prior knowledge to comprehend smooth characteristics related to drugs and diseases, employing an attention mechanism to gather crucial neighborhood information for representation learning [23]. Sun et al. developed an adaptive GCN to generate a k-nearest neighbor graph that combines feature similarity and topological structure, extracting embeddings of drug and disease features, and optimizing the embedding weights using consistency constraints and attention mechanisms. They identified both general and specific representations of drug and disease nodes, utilizing attention mechanisms and integration modules to adaptively merge these representations for the final prediction of associations [24]. He et al. presented EDEN, a framework that merges drug-disease associations into the disease information network. By optimizing mutual information between local semantics and global structures,

this system predicts drug-disease associations [25]. Although these methods have demonstrated effectiveness in DDA prediction, they often face challenges related to their reliance on incomplete information, particularly when drug-disease associations and similarity data are incomplete or noisy, which can lead to performance degradation and limit their applicability in complex biological networks. Graph-based contrastive learning, as an emerging approach, aims to enhance the robustness and generalization ability of GNNs by generating different views [26, 27]. However, most existing contrastive learning methods are primarily based on single-view contrast, overlooking the potential of multimodal or multi-view information [28, 29]. Therefore, these methods may struggle to fully capture semantic differences and structural information in graphs, particularly when dealing with highly heterogeneous biological networks, limiting their performance.

Based on this, we propose an innovative framework named CDPMF-DDA, which utilizes probabilistic matrix factorization to construct multi-view contrastive learning, deeply exploring potential drug-disease associations. By conducting contrastive learning across different views, the method effectively captures the multi-layered information of drug-disease relationships. At the same time, the probabilistic matrix factorization is employed to reconstruct the networks, effectively filtering out noisy information and enhancing the robustness of graph structure representation learning. The framework can be summarized as follows: (1) We first construct the drug-disease association, drug similarity, and disease similarity networks. (2) By introducing probabilistic matrix factorization with Gaussian noise, these networks are reconstructed into new drug-disease association, drug similarity, and disease similarity networks. (3) Multiple contrastive views of drug-disease, drug-drug, and disease-disease are then constructed upon these networks to better capture potential feature associations. (4) CDPMF-DDA incorporates only the target node neighborhood aggregation part in the graph convolution operation, improving computational efficiency. Furthermore, CDPMF-DDA integrates contrastive loss and binary cross-entropy to compute the overall loss between view pairs, enhancing its generalization capability. We evaluate the prediction performance of CDPMF-DDA against five other baseline methods across three public datasets. In overall performance, CDPMF-DDA demonstrates superior performance compared to the other methods, as indicated by the experimental results.

Methods

Methods overview

The structure of CDPMF-DDA is demonstrated in Fig. 1 and is primarily divided into three steps: (1) Using probabilistic matrix factorization, reconstruct networks of drug similarity, disease similarity, and drug-disease associations. (2) Generate comparative views for the drug-disease association network, drug similarity network, and disease similarity network. (3) Utilize the unified model to train and optimize these comparative views.

Model architecture

Probabilistic matrix factorization

Given n drugs $R = \{r_1, r_2, r_{3,} \ldots, r_n\}$ and m diseases $D = \{d_1, d_2, d_{3,} \ldots, d_m\}$, we present the known drug-disease associations using a binary matrix $Y \in \mathbb{R}^{n*m}$.The elements in the matrix $Y_{ij} \in \{0, 1\}$ indicate the association of a given drug r_i with a disease d_j. In cases where experimental evidence confirms the association between drug r_i and

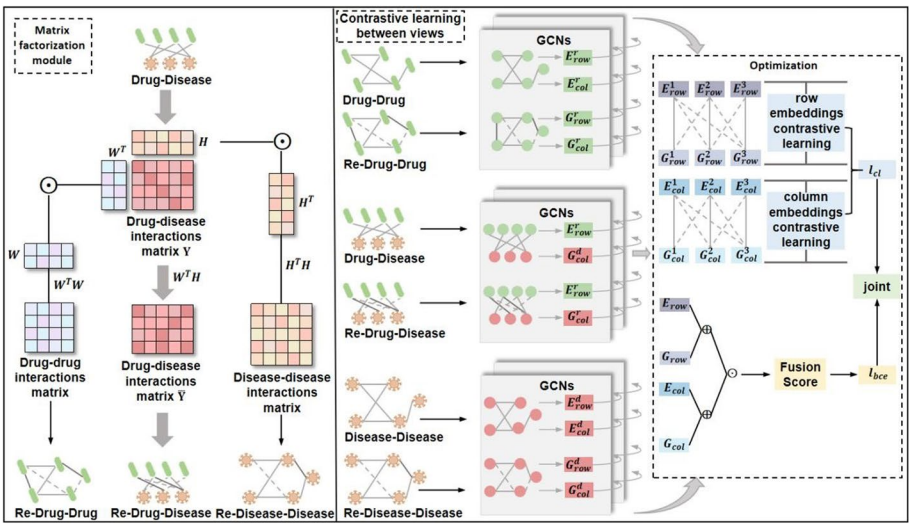

Fig. 1 The general architecture of CDPMF-DDA. (1) Matrix factorization module (2) Contrastive learning module (3) Optimization module

disease d_j, $Y_{ij} = 1$; otherwise, $Y_{ij} = 0$. Matrix factorization techniques typically project the drug-disease association matrix into latent feature matrices for drugs and diseases and reconstruct the drug-disease association matrix through their product. Let the drug latent feature matrix be expressed as $W \in \mathbb{R}^{t*n}$ and the disease latent feature matrix be $H \in \mathbb{R}^{t*m}$. Then, the reconstructed drug-disease association matrix is $\overline{Y} = W^T H$. Here, our objective is to find W and H to accurately reconstruct the drug-disease association matrix. From a probabilistic perspective, the conditional distribution of the drug-disease interaction matrix $\mathrm{Y} \in \mathbb{R}^{\mathrm{n}^*\mathrm{m}}$ is:

$$\mathrm{P}\left(\mathrm{Y}|\mathrm{W},\mathrm{H},\sigma^2\right) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left[\mathcal{N}\left(Y_{ij}|\mathrm{W}_i^T\mathrm{H}_j,\sigma^2\right)\right]^{I_{ij}} \tag{1}$$

where $\mathcal{N}\left(\chi|\mu,\sigma^2\right)$ represents the Gaussian normal distribution's probability density function, and I_{ij} is an indicator function. When the drug r_i and the disease d_j are linked, $I_{ij} = 1$; otherwise, $I_{ij} = 0$. Formula (1) serves as the generative model for the latent feature matrices of drug and disease. To further refine this model, we incorporate a zero-mean spherical Gaussian prior into the feature vectors of both drug and disease, as illustrated below:

$$P\left(W|\sigma_W^2\right) = \prod_{i=1}^{n}\mathcal{N}\left(W_i|0,\sigma_W^2 I\right) \tag{2}$$

$$P\left(H|\sigma_H^2\right) = \prod_{j=1}^{n}\mathcal{N}\left(H_j|0,\sigma_H^2 I\right) \tag{3}$$

where I is a k-dimensional unit diagonal matrix. By taking the logarithm of the posterior distribution of drug and disease characteristics and transforming it while keeping the hyperparameters unchanged, the maximization problem can be converted into

a minimization problem. This minimization procedure involves the objective function that sum of squared errors, combined with a quadratic regularization term:

$$\min_{W_i,H_j} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}\left(Y_{ij} - W_i^T H_j\right)^2 + \frac{\delta_W}{2}\sum_{i=1}^{n} \parallel W_i \parallel_F^2 + \frac{\delta_H}{2}\sum_{j=1}^{m} \parallel H_j \parallel_F^2 \tag{4}$$

where $\| \cdot \|_F$ refers to the Frobenius norm. Considering both drug and disease similarity constraints, we minimize the following loss function to ultimately learn W and H:

$$\begin{aligned}\min_{W_i,H_j} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(Y_{ij} - W_i^T H_j\right)^2 + \frac{\delta_W}{2}\sum_{i=1}^{N} \parallel W_i \parallel_F^2 + \frac{\delta_H}{2}\sum_{j=1}^{M} \parallel H_j \parallel_F^2 + \frac{\delta_1}{2} \parallel WW^T \\ - \mathbf{S}^r \parallel_F^2 + \frac{\delta_2}{2} \parallel HH^T - \mathbf{S}^d \parallel_F^2\end{aligned} \tag{5}$$

among these, $\mathbf{S}^r$ and $\mathbf{S}^d$ are the matrices representing drug similarity and disease similarity, respectively. To achieve minimization of this loss function, we apply the gradient descent method, updating W and H according to the following rule:

$$W \leftarrow W \times \frac{H\left(Y^T \odot I^T\right) + 2\mu_3 W\mathbf{S}^r}{H\left(\left(H^T W\right) \odot I^T\right) + \mu_1 W + 2\mu_3 W\left(W^T W\right)} \tag{6}$$

$$H \leftarrow H \times \frac{W\left(Y^T \odot I^T\right) + 2\mu_3 W\mathbf{S}^d}{W\left(\left(W^T H\right) \odot I\right) + \mu_2 H + 2\mu_4 H\left(H^T H\right)} \tag{7}$$

among them, I is a binary matrix indicating the positions of non-zero elements in the drug-disease association matrix Y. Concurrently, μ_1, μ_2, μ_3, and μ_4 are crucial hyperparameters that regulate model complexity and the impact of similarity measures. These update rules will be applied iteratively until a predefined number of iterations is reached. The final W and H matrices represent the latent feature matrices of the drug and disease that we aim to find. Subsequently, the iterated drug latent feature matrix W and disease latent feature matrix H are multiplied by their transposes to reconstruct the drug similarity matrix $\overline{\mathbf{S}}^r$ and disease similarity matrix $\overline{\mathbf{S}}^d$:

$$W^T W = \overline{\mathbf{S}}^r, H^T H = \overline{\mathbf{S}}^d \tag{8}$$

Eventually, we adopt the reconstructed drug-disease association matrix $\overline{Y}$, the drug similarity matrix $\overline{\mathbf{S}}^r$, and the disease similarity matrix $\overline{\mathbf{S}}^d$ to construct the drug-disease similarity network $\overline{G}$, the drug similarity network $\overline{G}^r$, and the disease similarity network $\overline{G}^d$. These networks integrate the relationships and similarities pertaining to drugs and diseases, providing a solid foundation for further analysis and prediction. A more profound analysis of these networks will help us to better understand the complexities inherent in drugs and diseases. This method effectively captures the complex associations linking drugs and diseases and reveals their internal similarity structures, providing new perspectives and a scientific basis for DDA prediction and disease treatment.

Multi-view contrastive learning

The GCN method aggregates the neighboring node information for each node in each layer, then updates its representation and uses it as the input for the next layer. Based on the research of Xia et al. [30], a two-layer GCN is used to aggregate the information from neighboring nodes for each node in this study. The first layer's aggregation process proceeds as follows:

$$z_{il}^{(r)} = \Omega\left(p(Y_i)E_{l-1}^{(d)}\right),\ z_{jl}^{(d)} = \Omega\left(p(Y_j)E_{l-1}^{(r)}\right) \tag{9}$$

where $z_{il}^{(r)}$ and $z_{jl}^{(d)}$ denote the first-layer embeddings of drugs r_i and diseases d_j, $\Omega()$ represents the identity activation function. $p()$ signifies the edge removal in the drug-disease association matrix Y to alleviate the overfitting problem. The sum of the aggregations over all layers forms the final representation of each node:

$$Z_i^{(r)} = \Sigma z_{il}^{(r)}, Z_j^{(d)} = \Sigma z_{jl}^{(d)} \tag{10}$$

where $Z_i^{(r)}$ and $Z_j^{(d)}$ represent the aggregated embeddings of drugs r_i and diseases d_j across all layers, respectively. Similarly, at each layer, we propagate messages on the reconstructed drug-disease association graph $\overline{G}$, and the specific expression is:

$$g_{il}^{(r)} = \Omega\left(p(\overline{Y}_i)E_{l-1}^{(d)}\right),\ g_{jl}^{(d)} = \Omega\left(p(\overline{Y}_j)E_{l-1}^{(r)}\right) \tag{11}$$

here, $g_{\mathrm{il}}^{(r)}$ and $g_{jl}^{(d)}$ represent the first-level embedding of drugs r_i and diseases d_j on the reconstructed $\overline{G}$, respectively. In a similar way, the final embeddings of drugs and diseases in the $\overline{G}$ are:

$$G_i^{(r)} = \Sigma g_{il}^{(r)}, G_j^{(d)} = \Sigma g_{jl}^{(d)} \tag{12}$$

In the end, the drug and disease representations are fused for optimal predictions in the following manner:

$$Q_r = \lambda_1 * Z_i^{(r)} + (1-\lambda_1)G_i^{(r)} \tag{13}$$

$$Q_d = \lambda_2 * Z_j^{(d)} + (1-\lambda_2)G_j^{(d)} \tag{14}$$

where Q_r and Q_d represent the final embeddings of drug r_i and disease d_j, respectively. λ_1 and λ_2 are hyperparameters that regulate the importance of the embeddings from the two views. Hence, the final prediction score for disease d_j associated with drug r_i is:

$$x_{ij} = Q_r^T Q_d \tag{15}$$

To obtain a deeper insight into the sophisticated associations relevant to drugs and diseases, we systematically analyzed the original drug similarity network G^r and its reconstructed network $\overline{G}^r$ through matrix factorization, as well as the disease similarity network G^d and its reconstructed network $\overline{G}^d$ using the same graph contrastive learning method. Initially, we process these networks using the k-nearest neighbor algorithm. Specifically, for each drug node, we identify the top k drug nodes with the

highest similarity in G^r and $\overline{G}^r$. The same method applies to the similarity analysis of disease nodes in G^d and $\overline{G}^d$. After extracting these key nodes, we utilize the aforementioned message propagation rules to further extract features and learn from this information.

In summary, the multi-view contrastive learning method effectively identifies and compares changes and stability within the network, uncovering key informational markers and potential therapeutic targets. The k-nearest neighbor algorithm captures direct neighbor relationships of drug or disease nodes, reflecting broader relational patterns through the network's topology. Meanwhile, node representations in the graph convolutional layer are updated by aggregating neighboring information, capturing local connection patterns and dynamic changes. By integrating these strategies, our framework extracts deep features from static data and captures dynamic network changes, enabling more accurate and comprehensive predictions of drug-disease associations. These insights enhance our understanding of drug-disease interactions, providing a robust foundation for novel drug discovery and therapeutic strategies.

Optimization

We denote the final embeddings of the row vectors in the drug similarity network G^r, the disease similarity network G^d, and the drug-disease association network G as $Z^{(r)}$. The final embeddings of the column vectors are denoted as $Z^{(C)}$. $G^{(r)}$ and $G^{(C)}$ indicate the final embeddings of row and column vectors in the reconstructed drug similarity network $\overline{G}^r$, the reconstructed disease similarity network $\overline{G}^d$, and the reconstructed drug-disease association network $\overline{G}$.

CDPMF-DDA uses consistent optimization scheme to calculate the overall loss across the three contrastive views. Contrastive learning aims to align node embeddings across contrasting views. We consider perspectives of the same node as positive examples, whereas perspectives of different nodes are considered negative examples. Positive samples are derived from known high-confidence interaction pairs, while negative samples are generated by randomly sampling pairs that have not been verified as interactions, with careful consideration to avoid inadvertently introducing potential positive samples during the sampling process. This sampling strategy ensures the validity and rationale of the positive-to-negative sample comparison, maintaining a robust and reliable dataset for further analysis. Therefore, we apply the InfoNCE contrastive loss to enhance the alignment of positive pairs while reducing the influence of negative pairs, thereby obtaining the contrastive loss of row embeddings in the three contrastive views:

$$l_{row} = \sum_{r_i \epsilon Row} -\log \frac{\exp\left(S\left(Z_i^{(r)} G_i^{(r)}\right)/\tau\right)}{\sum_{r_i \in Row} \exp\left(S\left(Z_i^{(r)} G_i^{(r)}\right)/\tau\right)} \tag{16}$$

where $s()$ stands for the cosine similarity function, τ denotes the temperature coefficient. In a similar manner, we have the ability to obtain the contrastive loss of the column embeddings:

$$l_{col} = \sum_{c_j \epsilon Col} -\log \frac{\exp\left(S\left(Z_j^{(c)} G_j^{(c)}\right)/\tau\right)}{\sum_{c_j \in Col} \exp\left(S\left(Z_j^{(c)} G_j^{(c)}\right)/\tau\right)} \tag{17}$$

The final total contrastive loss is:

$$l_{cl} = \beta_1 * l_{row} + \beta_2 * l_{col} \tag{18}$$

where β_1 and β_2 are hyperparameters. In an effort to strengthen the model's generalization ability, we incorporate binary cross-entropy loss to optimize our model jointly:

$$l_{bce} = -\frac{1}{n \times m}\left(\alpha \times \sum_{(i,j)\in\theta_{rc}^{+}} logx_{r_i,c_j} + \sum_{(i,j)\in\theta_{rc}^{-}} \left(1 - logx_{r_i,c_j}\right)\right) \tag{19}$$

For the drug-disease comparative view, let (i,j) represent a drug-disease pair, θ_{rc}^{+} represents the collection of all known drug-disease association pairs, while θ_{rc}^{-} denotes the collection of unknown drug-disease association pairs. In the drug comparative view, (i,j) indicates a drug pair, θ_{rc}^{+} signifies the entirety of known drug association pairs, with θ_{rc}^{-} comprises those drug association pairs that remain unknown; in the disease comparative view, the letters have the same meaning as above. The number of association pairs in the sets θ_{rc}^{+} and θ_{rc}^{-} are represented by $\left|\theta_{rc}^{+}\right|$ and $\left|\theta_{rc}^{-}\right|$, respectively, α is the balance factor, and its value is the ratio of $\left|\theta_{rc}^{+}\right|$ to $\left|\theta_{rc}^{-}\right|$. The overall optimization method is as follows:

$$L = \eta * l_{cl} + l_{bce} \tag{20}$$

where η act as a hyperparameter used to balance the two loss values.

Experimental

Datasets

To assess the performance of CDPMF-DDA, we utilize three distinct datasets: Fdataset, Cdataset, and LRSSL [31]. As per Gottlieb et al. Fdataset is broadly considered the gold standard dataset for drug-disease associations [32]. Fdataset contains 593 drugs sourced from DrugBank [33], 313 diseases gathered from OMIM [34], and 1933 verified drug-disease associations. Cdataset includes 663 drugs from DrugBank, 409 diseases from OMIM, and 2532 drug-disease association [35]. LRSSL consists of 3051 drug-disease associations, involving 763 drug from DrugBank and 681 diseases cataloged in MeSH [36].

Evaluation metrics

To thoroughly assess the performance of CDPMF-DDA, we employ a tenfold cross-validation, repeated 10 times, with the average result serving as the final outcome. In this process, known drug-disease pairs serve as positive samples, whereas unknown pairs are considered negative samples. The model's performance is assessed utilizing standard bioinformatics evaluation metrics, including AUROC and AUPR.

Baseline methods

For the intent of evaluating our model's performance, we conduct comparisons between CDPMF-DDA and a set of five state-of-the-art association prediction methods:

LBMFF [37]: This method combines literature information and feature fusion for the prediction of drug-disease associations. LBMFF constructs drug similarity and disease similarity matrices building on multiple features from public databases and PubMed sources. Next, an attention-based GCN is utilized to extract structural characteristics from the combined similarity matrix along with the existing drug-disease associations.

SCMFDD [38]: This similarity-constrained matrix factorization approach predicts drug-disease associations by transforming their relationships into two reduced-rank spaces, incorporating drug feature similarities and disease semantic similarities to constrain them.

SCPMF [39]: This similarity-constrained probabilistic matrix factorization model function on the adjacency matrix of a complex drug-disease network, integrating drug-virus interactions, chemical structures, and viral genomic data. It maps the drug-virus interaction matrix into a pair of latent feature matrices, one corresponding to drugs and the other to viruses, and integrates a weighted similarity interaction matrix to constrain both.

MNGACDA [40]: This approach predicts circRNA-drug sensitivity by leveraging a multimodal network with a graph encoder and an attention mechanism. It constructs integrated networks for circRNA similarity, drug similarity, and circRNA-drug sensitivity. By embedding an attention layer at the node level into a deep graph neural network, the method extracts intrinsic node information. These representations of circRNA and drugs are then used in an internal decoder to predict sensitivity associations.

MKGCN [41]: By using GCN, this method extracts multi-layer features from a heterogeneous network comprising microorganisms and drugs. It calculates the kernel matrix at every layer and merges the various kernel matrices by employing a weighted average technique. These kernels are integrated within the microorganism and drug space by employing the dual Laplace regularized least squares approach to deduce new associations between microorganisms and drugs.

Parameters SETTING

CDPMF-DDA utilizes a two-layer of GCN for iteration processing. The configuration includes a learning rate of 0.001, an embedding size of 64, and a batch size of 5120. In the Fdataset and Cdataset, the number of epochs for the CDPMF-DDA model is set to 200, while for the LRSSL dataset, it is set to 128. The hyperparameters of LBMFF, SCMFDD, SCPMF, MNGACDA, and MKJCN are all set to their original optimum values.

Results and discussions

Performance of CDPMF-DDA in 10 times of tenfold cross-validation

In assessing CDPMF-DDA's effectiveness, we apply tenfold cross-validation, repeated ten times, to three public datasets and evaluate the performance against baseline methods. In these comparisons, all baseline models are configured using the optimal parameter settings reported in the literature to ensure consistency in the optimization conditions.

Table 1 The results of 10 times of tenfold cross-validation

Datasets	LBMFF	SCPMFDD	SCPMF	MKGCN	MNGACDA	CDPMF-DDA
AUROCs						
Fdataset	0.7953 ± 0.035	0.7740 ± 0.001	**0.8957 ± 0.001**	0.8870 ± 0.001	0.8179 ± 0.005	0.9477 ± 0.003
Cdataset	0.9069 ± 0.001	0.7937 ± 0.001	**0.9117 ± 0.002**	0.9109 ± 0.001	0.8406 ± 0.005	0.9548 ± 0.005
LRSSL	**0.9139 ± 0.002**	0.7668 ± 0.001	0.8977 ± 0.001	0.8596 ± 0.001	0.7936 ± 0.002	0.9401 ± 0.005
Avg	0.8720	0.7782	**0.9017**	0.8858	0.8173	0.9475
AUPRs						
Fdataset	0.0266 ± 0.012	0.0056 ± 0.000	0.3451 ± 0.006	**0.4471 ± 0.001**	0.2494 ± 0.001	0.5540 ± 0.005
Cdataset	0.2176 ± 0.006	0.0055 ± 0.000	0.4140 ± 0.004	**0.5647 ± 0.001**	0.3396 ± 0.001	0.5967 ± 0.005
LRSSL	0.1740 ± 0.005	0.0038 ± 0.000	0.2710 ± 0.002	**0.3518 ± 0.001**	0.1799 ± 0.001	0.3520 ± 0.005
Avg	0.1394	0.00498	0.3433	**0.4545**	0.2563	0.5009

a ± b indicates that a fluctuates within the range of b. bold values represent the second-best results across the three datasets

Results in Table 1 show that CDPMF-DDA outperforms all comparative models within the datasets, with average AUROC and AUPR of 0.9475 and 0.5009, respectively, surpassing the second-best by 5.082% and 9.999%. Figures 2 and 3 show the corresponding ROC and PR curves. These results demonstrate that CDPMF-DDA provides notable improvements in accuracy and consistency for DDA prediction.

To evaluate CDPMF-DDA's performance comprehensively, we employ the same metric, Recall@k, as used by Zeng et al. [42], to further analyze CDPMF-DDA's capacity to identify positive samples. Recall@k is defined as the fraction of positive samples in the top K predictions out of the total number of positive samples within

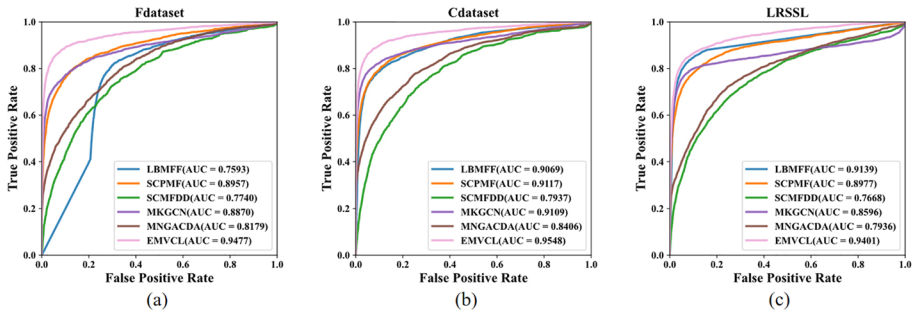

Fig. 2 AUROC with different methods for predicting potential drugs for new diseases. **a** Fdataset, **b** Cdataset, and **c** LRSSL

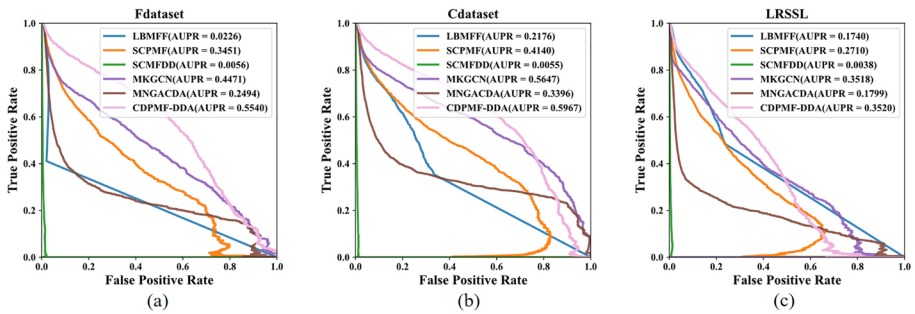

Fig. 3 AUPR with different methods for predicting potential drugs for new diseases. **a** Fdataset, **b** Cdataset, and **c** LRSSL

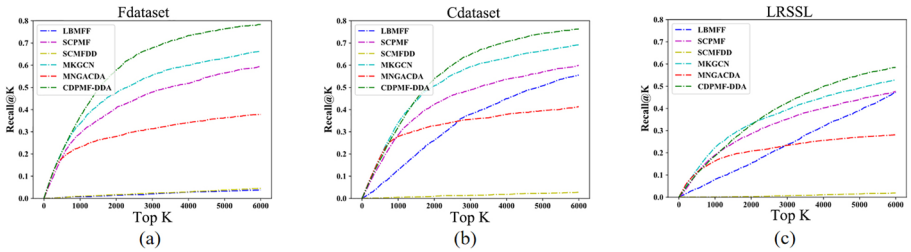

Fig. 4 Comparison different methods of Recall@k values for the top k predicted. **a** Fdataset, **b** Cdataset, and **c** LRSSL

Table 2 AUROC and AUPR of CDPMF-DDA and variants with tenfold cross validation

Variants	Fdataset	
	AUROC	**AUPR**
CDPMF-DDA -noMF	0.8420	0.4932
CDPMF-DDA -noCL	0.9453	0.5477
CDPMF-DDA	0.9477	0.5540

the dataset. During this analysis, all models are optimized using the same evaluation criteria and hyperparameter tuning strategies to guarantee fairness and consistency across the experiments. As indicated in Fig. 4, for the top 6,000 predictions on Fdataset, Cdataset, and LRSSL, the Recall@k value of CDPMF-DDA surpasses that of other models. This advantage is attributed to CDPMF-DDA's effective utilization of the probabilistic matrix factorization technique, which thoroughly explores potential drug-disease associations, thereby enhancing the identification of positive samples.

Ablation analysis

To verify the effectiveness of contrastive learning with matrix factorization, we compare two variants. The model variants are summarized as follows:

CDPMF-DDA-noMF: This variant uses random edge culling to generate contrastive views.

CDPMF-DDA-noCL: In this variant, contrastive learning between views is disabled.

Table 2 illustrates the performance comparison between CDPMF-DDA and variants across the Fdataset compared to CDPMF-DDA, CDPMF-DDA-noMF has a notably reduced performance, indicating that probabilistic matrix factorization plays a crucial role in improving the model's final performance. By using matrix factorization, the model is able to more accurately capture the latent information of the nodes while reducing noise interference, which enhances the expressiveness of the embedding representations. Additionally, the performance of CDPMF-DDA is superior to that of CDPMF-DDA-noCL, demonstrating the effectiveness of incorporating contrastive learning into representation learning between nodes. Contrastive learning enhances semantic consistency across different views. This helps the model to extract node features from multiple perspectives, improving understanding of the underlying

distribution and enhancing the generalization and robustness of node representations, boosting overall model performance.

Case studies

The practical utility of CDPMF-DDA is validated through case studies on two diseases, Alzheimer's disease and Epilepsy. For predicting candidate drugs for these diseases, CDPMF-DDA is trained using all known drug-disease pairs in the Fdataset, with unobserved pairs considered as candidate sets. Following the prediction, the probabilities are ranked from highest to lowest, and the top ten drugs for each disease are selected for subsequent investigation Authoritative data sources such as DrugBank [33], DrugCentral [43], CTD, and ClinicalTrials.gov are used to validate the accuracy of CDPMF-DDA's prediction results.

Alzheimer's disease ranks as the predominant progressive neurocognitive condition in senior citizens and is a type of neurodegenerative disorder [44]. Table 3 lists the 10 drugs predicted by CDPMF-DDA to potentially treat Alzheimer's disease. CDPMF-DDA predicts Memantine as the top potential drug for the treatment of Alzheimer's disease. Memantine is an excitatory amino acid receptor antagonist used to treat Alzheimer's disease [45], and this prediction is confirmed by ClinicalTrials.gov, CTD, alongside DrugCentral. Vitamin E, known for its antioxidant and neuroprotective effects, helps reduce inflammation and lower cholesterol levels, which are essential for maintaining brain health [46]. CDPMF-DDA's prediction of Vitamin E's relevance to Alzheimer's disease is supported by ClinicalTrials.gov and CTD. Additionally, the candidate drugs Rivastigmine and Pramipexole predicted by CDPMF-DDA are also validated by ClinicalTrials.gov and CTD, as well as DrugCentral. In summary, among the top 10 candidate drugs, 8 (achieving an 80% success ratio) are validated through reliable public datasets and clinical evidence.

Epilepsy is a brain disorder characterized by long-term recurrent seizures, occurring when brain cells malfunction and send electrical signals uncontrollably. We analyze the leading ten drug options predicted by CDPMF-DDA as possible treatments for Epilepsy. According to Table 4, 9 out of the 10 candidates (90% success rate) are proven effective with reliable evidence. CDPMF-DDA identifies Gabapentin as an effective treatment for Epilepsy, a discovery confirmed by data available from sources including DrugCentral

Table 3 The top ten potential drugs for Alzheimer's disease

Rank	DrugBank IDs	Candidate drugs	Evidences
1	DB01043	Memantine	ClinicalTrials.gov,CTD,DrugCentral
2	DB00163	Vitamin E	ClinicalTrials.gov,CTD
3	DB01356	Lithium cation	Unconfirmed
4	DB00989	Rivastigmine	ClinicalTrials.gov, DrugCentral
5	DB00915	Amantadine	CTD
6	DB00413	Pramipexole	ClinicalTrials.gov, DrugCentral
7	DB00382	Tacrine	DrugCentral
8	DB00822	Disulfiram	ClinicalTrials.gov
9	DB00763	Methimazole	Unconfirmed
10	DB00674	Galantamine	DB, DrugCentral

Table 4 The top ten potential drug candidates for Epilepsy

Rank	DrugBank IDs	Candidate drugs	Evidences
1	DB00996	Gabapentin	DB,DrugCentral,ClinicalTrials.gov,CTD
2	DB00794	Primidone	DB,DrugCentral,CTD
3	DB00909	Zonisamide	DB,DrugCentral,ClinicalTrials.gov,CTD
4	DB01202	Levetiracetam	DB,DrugCentral,
5	DB00898	Ethanol	ClinicalTrials.gov,CTD
6	DB01378	Magnesium cation	CTD
7	DB01019	Bethanechol	Unconfirmed
8	DB00515	Cisplatin	CTD
9	DB00653	Magnesium sulfate	ClinicalTrials.gov,CTD
10	DB00160	Alanine	CTD

Table 5 Molecular binding energies (kcal/mol) of the top 10 Alzheimer's disease candidate drugs

Drug	Docking Energy(kcal/mol)				
	1QJH	2R0Z	3K9O	4FWU	5N62
Memantine	−5.4	−5.7	−5.2	−5.8	−5.6
Vitamin E	−5.9	−5.7	−5.4	−6.2	−5.7
Lithium cation	−3.8	−4.8	−4.8	−5.4	−4.8
Rivastigmine	−5.0	−5.7	−5.0	−6.2	−6.3
Amantadine	−4.5	−5.4	−4.7	−5.0	−5.1
Pramipexole	−5.0	−5.3	−4.8	−5.1	−5.7
Tacrine	−5.8	−6.6	−5.8	−6.6	−7.0
Disulfiram	−3.8	−4.1	−3.4	−4.4	−3.9
Methimazole	−3.4	−3.2	−3.5	−3.5	−3.7
Galantamine	−6.1	−6.5	−6.2	−6.7	−6.3

and pertinent academic articles. Ethanol is used for therapeutic neurolysis of nerves or ganglia to relieve intractable chronic pain in conditions such as tics. CDPMF-DDA predicts that Ethanol could be used to treat Epilepsy, and this assertion is corroborated by information from ClinicalTrials.gov and CTD. Magnesium cation, an important cation, can also be used to treat epilepsy, as confirmed by relevant literature from CTD. Additionally, CDPMF-DDA predicts that Primidone and Levetiracetam might be promising candidates for treating Epilepsy, as validated by DB and DrugCentral. These findings underscore the robustness and reliability of CDPMF-DDA in identifying effective drug candidates for Epilepsy, demonstrating its potential utility in medical research and clinical practice.

Additionally, we conduct molecular docking experiments using Alzheimer's disease as an example. We select five target proteins associated with Alzheimer's disease and use AutoDock Vina [47] to evaluate the molecular interaction energy among the top 10 drugs forecasted by CDPMF-DDA in Fdataset and these five target proteins. The specific results are displayed in Table 5. The molecular affinity energies of the untested candidate drug lithium cation with the Alzheimer's disease target proteins, identified by the codes 1QJH, 2R0Z, 3K9O, 4FWU, and N62, are -3.8, -4.8, -4.8, -5.4, and -4.8 kcal/mol, respectively. Although lithium cation has not yet been confirmed as tied to Alzheimer's

disease, it is prescribed to treat bipolar disorder and various other mental health disorders. Hampel et al. suggested that it may have therapeutic potential for Alzheimer's disease [48]. The experimental results indicate that lithium cation may serve a potential role in the therapy for Alzheimer's disease, but further verification and research are still needed.

Subsequently, we employ DS software to visualize the molecular docking results of Methimazole and Abeta (PDB code: 2R0Z). As illustrated in Fig. 5, the small molecules interact through van der Waals forces with particular amino acid residues exemplified by VAL132, SER176, SER161, SER168, PHE166, LEU177, THR176, LEU169, LEU141, and ASN160. Additionally, several common intermolecular interactions are noticed, such as the typical hydrogen bond interaction between hydrogen atoms and THR177. There are also hydrophobic interactions, including carbon-hydrogen bond interactions between hydrogen atoms and LEU159 and SER175. According to the visualization results, the drugs predicted by CDPMF-DDA as potential treatments for Alzheimer's disease could provide worthwhile references for medical research.

Based on drug similarity, we select the top 10 drug similarity pairs from the drug similarity matrix of Fdataset and obtain 15 drug similarity sets. With CDPMF-DDA, we predict the top 30 potential therapeutic diseases for each drug, which allows us to build the corresponding drug-disease network. We divide these 15 clear community structure drugs into 6 communities and performed a visual analysis using the modularization function of Gephi software. The experimental results show that these 15 drugs exhibit a clear community structure, reflecting the clustering trend of drugs in terms of similarity and interaction. As shown in Fig. 6, Gonadorelin, Sermorelin, Cyclosporine, and Leuprolide are in the same cluster, all acting on the central nervous system. Among them, Gonadorelin and Leuprolide are gonadotropin-releasing hormone analogs that stimulate the anterior pituitary to secrete LH and FSH. Similarly, Sermorelin treats children with growth hormone deficiency or growth disorders by promoting the release of growth hormone from the pituitary. Cyclosporine, which acts as a calcineurin inhibitor, can prevent rejection in organ transplants and treat a range of inflammatory and autoimmune

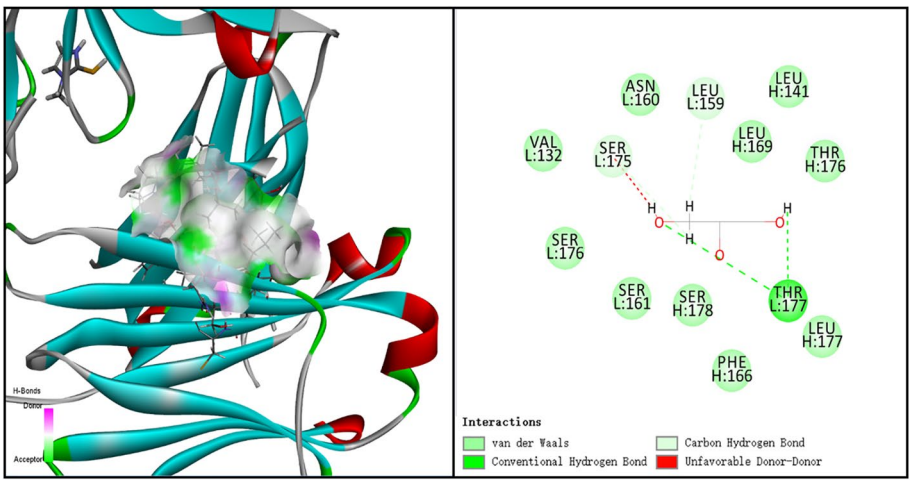

Fig. 5 Molecular docking outcomes in both 3D and 2D for Methimazole (DrugBank ID: DB00763) with Abeta (PDB code: 2R0Z)

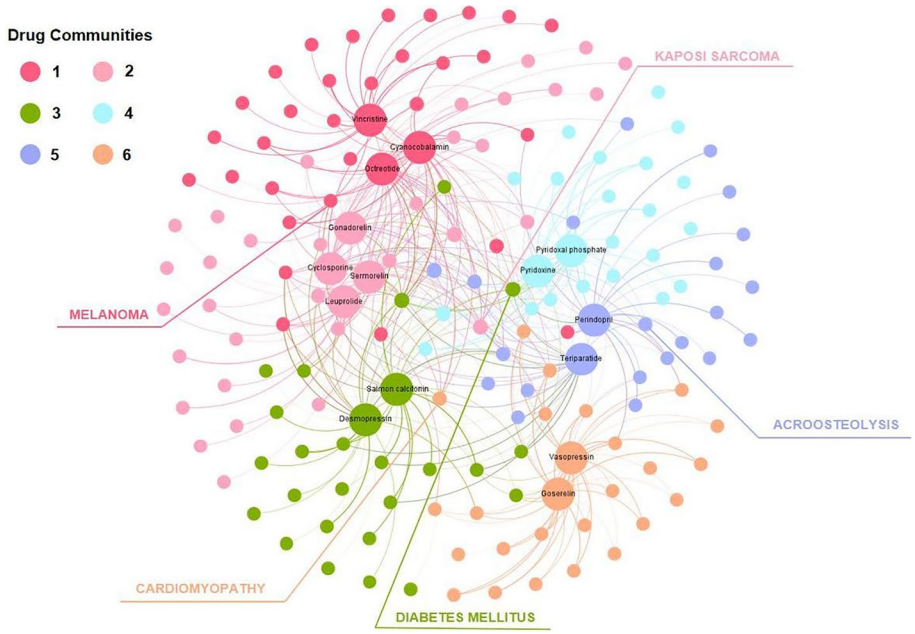

Fig. 6 The drug-disease network predicted from Fdataset. This network links 15 selected drugs (depicted by large circles) with 450 potential therapeutic diseases (represented by small circles) as predicted by CDPMF-DDA. The potential therapeutic diseases for each drug are ranked in accordance with the prediction probabilities from CDPMF-DDA, and the connections between drugs and diseases are weighted and emphasized accordingly

diseases by providing significant immunosuppressive effects on T cells. Additionally, Cyanocobalamin, Vincristine, and Octreotide are widely used in anticancer treatment, with Vincristine and Octreotide primarily used to treat tumors. Vincristine may cause bone marrow suppression, increasing the need for vitamin B12, while Octreotide may reduce the patient's vitamin B12 levels. Cyanocobalamin can correct vitamin B12 deficiency. Pyridoxal phosphate and Pyridoxine, both vitamins, can be used to treat vitamin B6 deficiency. Drug combinations can improve therapeutic effects, reduce drug toxicity and side effects, and have a significant impact on the treatment of various complex diseases [49, 50].

In summary, CDPMF-DDA can effectively identify associations between drugs and uncover shared characteristics among them, thereby providing a reliable foundation for predicting potential treatments for new diseases.

Conclusion

In this study, we introduce CDPMF-DDA, an innovative multi-view contrastive learning framework for drug-disease association prediction. The framework aims to comprehensively explore the latent feature relationships between drug-disease, drug-drug, and disease-disease interactions by integrating multi-view feature representations. This approach effectively captures complex network structures and association patterns across multiple dimensions of information. By employing probabilistic matrix factorization, CDPMF-DDA extracts hidden features across various networks, enhancing the understanding of drug-disease associations. Additionally, we design a multi-view contrastive learning mechanism that encompasses drug-disease, drug-drug, and

disease-disease networks. This mechanism strengthens the model's ability to learn rich features representations and capture key structural information. Extensive experiments on three publicly available datasets demonstrate that CDPMF-DDA outperforms five state-of-the-art models. Specifically, in 10 times of tenfold cross-validation, CDPMF-DDA achieves an average AUC of 0.9475 and AUPR of 0.5009, surpassing all comparison methods. Furthermore, CDPMF-DDA excels Recall@k for the top-k item predictions, further validating its outstanding capability in capturing critical drug-disease relationships.

CDPMF-DDA also shows promise in drug-target prediction and gene-disease association analysis. Future research will focus on optimizing matrix factorization to more effectively preserve key information from the original data and exploring adaptive multimodal data fusion strategies to improve performance.

Acknowledgements
Not applicable.

Author contributions
XT is responsible for conceptualizing the project, participating in the review and editing of the manuscript, and designing the methodology; YH is involved in the conceptualization, writes the initial draft of the study, and develops the relevant software; YM participates in the review and editing of the manuscript and is also responsible for acquiring funding; ZW and JL each conduct formal analysis of the data; CL and XH are in charge of data collection and curation; JX and JY provide supervision and manage the necessary resources.All authors reviewed the manuscript.

Funding
This work is supported by the National Natural Science Foundation of China (Grant Nos. 62302156, and 62402349), the Natural Science Foundation of Hunan Province (Grant No. 2023JJ40180), the Natural Science Foundation of Hubei Province (Grant No. 2024AFB127), and Wuhan Textile University Foundation (Grant Nos. 20230612 and 2024309).

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not Applicable.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Received: 30 October 2024 Accepted: 27 December 2024
Published online: 07 January 2025

References

1. Lau A, So H-C. Turning genome-wide association study findings into opportunities for drug repositioning. Comput Struct Biotechnol J. 2020;18:1639–50.
2. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform. 2016;17(1):2–12.
3. Ao C, Xiao Z, Guan L, Yu L. Computational Approaches for Predicting Drug-Disease Associations: A Comprehensive Review. arXiv preprint arXiv:06388.2023.
4. Giri S, Bader A. A low-cost, high-quality new drug discovery process using patient-derived induced pluripotent stem cells. Drug Discov Today. 2015;20(1):37–49.
5. Bhutani P, Joshi G, Raja N, Bachhav N, Rajanna PK, Bhutani H, et al. US FDA approved drugs from 2015–June 2020: a perspective. J Med Chem. 2021;64(5):2339–81.
6. Asselah T, Durantel D, Pasmant E, Lau G, Schinazi RF. COVID-19: discovery, diagnostics and drug development. J Hepatol. 2021;74(1):168–84.
7. Huang Z, Xiao Z, Ao C, Guan L, Yu L. Computational approaches for predicting drug-disease associations: a comprehensive review. Front Comp Sci. 2025;19(5):1–15.
8. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58.

9. He S, Yun L, Yi H. Fusing graph transformer with multi-aggregate GCN for enhanced drug–disease associations prediction. BMC Bioinf. 2024;25(1):79.
10. Chen X, Yin J, Qu J, Huang L. MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. PLoS Comput Biol. 2018;14(8):e1006418.
11. Rohani N, Eslahchi C, Katanforoush A. ISCMF: Integrated similarity-constrained matrix factorization for drug–drug interaction prediction. Netw Model Anal Health Inf Bioinform. 2020;9:1–8.
12. Guan N-N, Zhao Y, Wang C-C, Li J-Q, Chen X, Piao X. Anticancer drug response prediction in cell lines using weighted graph regularized matrix factorization. Mol Therapy-Nucleic Acids. 2019;17:164–74.
13. Liu W, Tang T, Lu X, Fu X, Yang Y, Peng L. MPCLCDA: predicting circRNA–disease associations by using automatically selected meta-path and contrastive learning. Brief Bioinf. 2023;24(4):bbad227.
14. Zhang W, Chen Y, Li D, Yue X. Manifold regularized matrix factorization for drug-drug interaction prediction. J Biomed Inform. 2018;88:90–7.
15. Sadeghi S, Lu J, Ngom A. A network-based drug repurposing method via non-negative matrix factorization. Bioinformatics. 2022;38(5):1369–77.
16. Wang M-N, Xie X-J, You Z-H, Ding D-W, Wong L. A weighted non-negative matrix factorization approach to predict potential associations between drug and disease. J Transl Med. 2022;20(1):552.
17. Yang M, Wu G, Zhao Q, Li Y, Wang J. Computational drug repositioning based on multi-similarities bilinear matrix factorization. Brief bioinf. 2021;22(4):267.
18. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Comput Sci Rev. 2009;42(8):30–7.
19. Meng Y, Wang Y, Xu J, Lu C, Tang X, Peng T, et al. Drug repositioning based on weighted local information augmented graph neural network. Brief Bioinf. 2024;25(1):bbad431.
20. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? arXiv preprint https://arxiv.org/abs/1810.00826.
21. You J, Ying Z, Leskovec J. Design space for graph neural networks. Adv Neural Inf Process Syst. 2020;33:17009–21.
22. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug–disease associations through layer attention graph convolutional network. Brief Bioinf. 2021;22(4):bbaa243.
23. Zhao B-W, Su X-R, Hu P-W, Ma Y-P, Zhou X, Hu L. A geometric deep learning framework for drug repositioning over heterogeneous information networks. Brief Bioinf. 2022;23(6):bbac384.
24. Sun X, Jia X, Lu Z, Tang J, Li M. Drug repositioning with adaptive graph convolutional networks. Bioinformatics. 2024;40(1):btad748.
25. He C, Duan L, Zheng H, Song L, Huang M. An explainable framework for drug repositioning from disease information network. Neurocomputing. 2022;511:247–58.
26. Gao Z, Ma H, Zhang X, Wang Y, Wu Z. Similarity measures-based graph co-contrastive learning for drug–disease association prediction. Bioinformatics. 2023;39(6):btad357.
27. Jia X, Sun X, Wang K, Li M, Informatics H. DRGCL: drug repositioning via semantic-enriched graph contrastive learning. IEEE J Biomed. 2024. https://doi.org/10.1109/JBHI.2024.3372527.
28. Gao Y, Li X, Yan H. Rethinking graph contrastive learning: an efficient single-view approach via instance discrimination. IEEE Trans Multimed. 2023. https://doi.org/10.1109/TMM.2023.3313267.
29. Bae S, Kim S, Ko J, Lee G, Noh S, Yun S-Y, editors. Self-contrastive learning: single-viewed supervised contrastive framework using sub-network. Proceedings of the AAAI Conference on Artificial Intelligence; 2023.
30. Xia L, Huang C, Xu Y, Zhao J, Yin D, Huang J, editors. Hypergraph contrastive collaborative filtering. Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval; 2022.
31. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A weighted bilinear neural collaborative filtering approach for drug repositioning. Brief Bioinf. 2022;23(2):bbab581.
32. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol. 2011;7(1):496.
33. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36(supp1):D901–6.
34. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(suppl1):D514–7.
35. Yi H-C, You Z-H, Wang L, Su X-R, Zhou X, Jiang T-H. In silico drug repositioning using deep learning and comprehensive similarity measures. BMC Bioinf. 2021;22:1–15.
36. Liang X, Zhang P, Yan L, Fu Y, Peng F, Qu L, et al. LRSSL: predict and interpret drug–disease associations based on data integration using sparse subspace learning. Bioinformatics. 2017;33(8):1187–96.
37. Kang H, Hou L, Gu Y, Lu X, Li J, Li Q. Drug–disease association prediction with literature based multi-feature fusion. Front Pharmacol. 2023;14:1205144.
38. Zhang W, Yue X, Lin W, Wu W, Liu R, Huang F, et al. Predicting drug-disease associations by using similarity constrained matrix factorization. BMC Bioinf. 2018;19:1–12.
39. Meng Y, Jin M, Tang X, Xu J. Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. Appl Soft Comput. 2021;103:107135.
40. Yang B, Chen H. Predicting circRNA-drug sensitivity associations by learning multimodal networks using graph auto-encoders and attention mechanism. Brief Bioinf. 2023;24(1):bbac596.
41. Yang H, Ding Y, Tang J, Guo F. Inferring human microbe–drug associations via multiple kernel fusion on graph neural network. Knowl-Based Syst. 2022;238:107888.
42. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics. 2019;35(24):5191–8.
43. Avram S, Bologa CG, Holmes J, Bocci G, Wilson TB, Nguyen D-T, et al. DrugCentral 2021 supports drug discovery and repositioning. Nucleic Acids Res. 2021;49(D1):D1160–9.
44. Wu Q, Su S, Cai C, Xu L, Fan X, Ke H, et al. Network Proximity-based computational pipeline identifies drug candidates for different pathological stages of Alzheimer's disease. Comput Struct Biotechnol J. 2023;21:1907–20.

45. Murakawa-Hirachi T, Mizoguchi Y, Ohgidani M, Haraguchi Y, Monji A. Effect of memantine, an anti-Alzheimer's drug, on rodent microglial cells in vitro. Sci Rep. 2021;11(1):6151.
46. Lloret A, Esteve D, Monllor P, Cervera-Ferri A, Lloret A. The effectiveness of vitamin E treatment in Alzheimer's disease. Int J Mol Sci. 2019;20(4):879.
47. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61.
48. Hampel H, Lista S, Mango D, Nistico R, Perry G, Avila J, et al. Lithium as a treatment for Alzheimer's disease: the systems pharmacology perspective. J Alzheimers Dis. 2019;69(3):615–29.
49. Tang X, Zhou C, Lu C, Meng Y, Xu J, Hu X, et al. Enhancing drug repositioning through local interactive learning with bilinear attention networks. IEEE J Biomed. 2023. https://doi.org/10.1109/JBHI.2023.3335275.
50. Chen X, Ren B, Chen M, Wang Q, Zhang L, Yan G. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. PLoS Comput Biol. 2016;12(7):e1004975.

Publisher's Note