

# Multidependency Graph Convolutional Networks and Contrastive Learning for Drug Repositioning

Yanglan Gan,\* Shengnan Li, Guangwei Xu, Cairong Yan, and Guobing Zou



Cite This: <https://doi.org/10.1021/acs.jcim.4c02424>

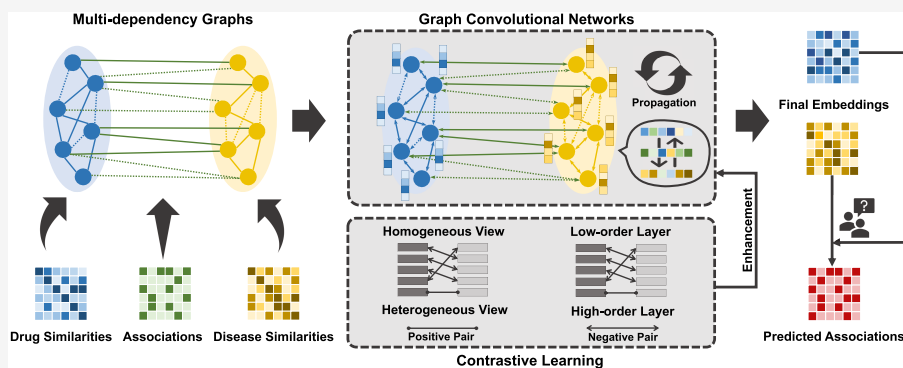


Read Online

ACCESS |

Metrics & More

Article Recommendations



**ABSTRACT:** The goal of drug repositioning is to expedite the drug development process by finding novel therapeutic applications for approved drugs. Using multifeature learning, different computational drug repositioning techniques have recently been introduced to predict possible drug–disease relationships. Nevertheless, current graph-based methods tend to model drug–disease interaction relationships without considering the semantic influence of node-specific side information on graphs. These approaches also suffer from the noise and sparsity inherent in the data. To address these limitations, we propose MDGCN, a novel drug repositioning method that incorporates multidependency graph convolutional networks and contrastive learning. Based on drug and disease similarity matrices and the drug–disease relationships matrix, this approach constructs multidependency graphs. It subsequently employs graph convolutional networks to spread side information between various graphs in each layer. Meanwhile, the weak supervision of drug–disease connections is effectively addressed by introducing cross-view and cross-layer contrastive learning to align node embedding across various views. Extensive experiments show that MDGCN performs better in drug–disease association prediction than seven advanced methods, offering strong support for investigating novel therapeutic indications for medications of interest.

## INTRODUCTION

Despite substantial investment increases in drug discovery, the approval rate of new drugs continues to stagnate. Modern innovations in genomics, biotechnology, and computational tools have made strides but have not drastically shortened the drug discovery timeline.<sup>1,2</sup> This highlights the pressing need for innovative methodologies to speed up drug development and address critical health challenges.<sup>3</sup> Recently, the concept of drug repositioning has drawn significant attention. It aims to uncover additional therapeutic potentials for approved drugs, extending their use beyond their primary medical indications.<sup>4</sup> By circumventing lengthy clinical trials, drug repositioning notably cuts down on the time, financial investment, and uncertainties inherent in conventional drug discovery processes, thereby offering a valuable alternative route for drug discovery.

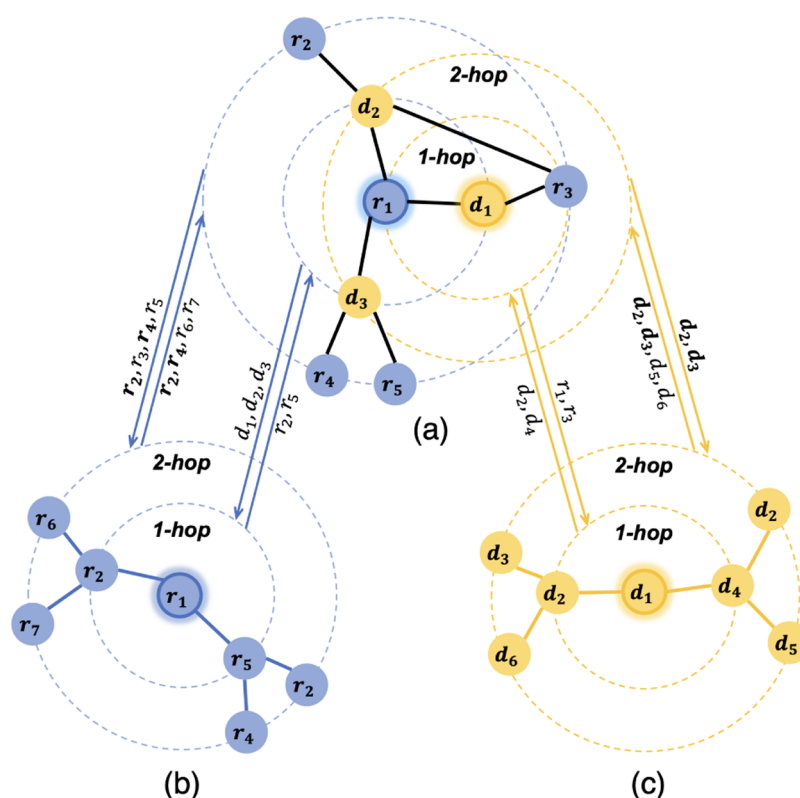
To address drug repositioning, researchers have explored a variety of computational approaches, such as machine learning, matrix factorization, and network-based techniques.<sup>5</sup> Specifi-

cally, machine learning-based drug repositioning methods extract features from existing drug–disease interaction data to predict potential new uses for existing drugs.<sup>6,7</sup> For instance, a kernel-based support vector machine is employed to predict drug repositioning based on integrated multilayer drug-related features.<sup>8</sup> DRP-VEM<sup>9</sup> adopts a voting ensemble training strategy to overcome the serious imbalance problem between positive and negative instances. However, these methods rely heavily on feature extraction and negative sample selection. Differently, the strength of the correlations between drugs and

**Received:** December 29, 2024

**Revised:** February 23, 2025

**Accepted:** March 5, 2025



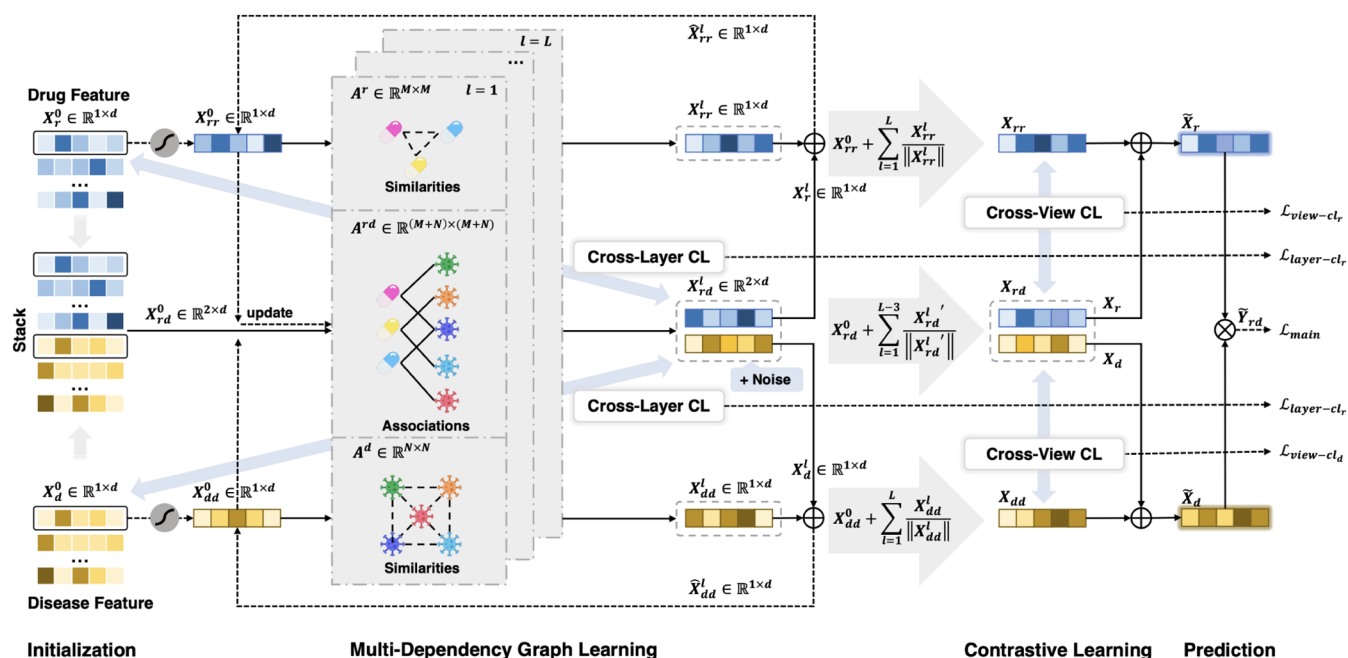
**Figure 1.** Toy example of multidependency graph convolutional networks. (a) Drug–disease bipartite graph. (b) Drug–drug graph and (c) disease–disease graph, both constructed using their respective similarity matrices and KNN algorithm with  $k = 2$ .

diseases is evaluated by using the dot product of their learned representations in matrix factorization-based techniques, which embed drugs and diseases into a common hidden space. DRIMC<sup>10</sup> develops a Bayesian inductive matrix completion technique that integrates multiple similarity matrices of drug characteristics and disease features for link prediction. Additionally, BNNR<sup>11</sup> employs a bounded nuclear norm constraint to reconstruct the drug–disease association matrix, leveraging its low-rank structural assumption for effective matrix completion. TIWMFLP<sup>12</sup> develops a two-tier interactive framework to integrate multiple similarities. Despite their effectiveness, these methods face challenges in identifying the intricate relationships within drug–disease association data and in managing complex matrix operations on a large scale. Network-based drug repositioning methods model the relationships among various biological entities, such as drugs, diseases, proteins, and genes, to explore more efficient representations of drugs and diseases.<sup>5</sup> For instance, deepDR<sup>13</sup> learns latent features from ten drug-related networks using a multimodal deep autoencoder and identifies potential drug–disease associations through a collective variational autoencoder. AMDGT<sup>14</sup> introduces dual-graph transformer modules with complex biochemical knowledge to identify new indications of medicines. Nonetheless, integrating biological data requires prior knowledge and combining large-scale and diverse data sets, such as chemical, biological, and clinical information, into a cohesive framework presents significant challenges.

Recently, due to the ability of effectively extracting key patterns, graph neural networks (GNNs) have been increasingly used to model drug–disease association networks. These drug repositioning methods typically use drug similarity matrix and disease similarity matrix as features of drug–disease heteroge-

neous graph, inputting them into GNNs for further encoding node representations.<sup>15,16</sup> DRWBNCF<sup>17</sup> utilizes weighted bilinear graph convolution operations to learn unified representations of drugs and diseases. DRHGCN<sup>18</sup> introduces graph convolution operations and a mechanism to selectively emphasize different layers, enabling the extraction of inter-domain and intradomain embeddings for prediction. AdaDR<sup>19</sup> is an adaptive framework based on GCNs that employs an attention-based strategy to integrate information from node-level data and their corresponding graph structures. However, most existing GCN-based drug repositioning models primarily focus on the drug–disease interaction graph and do not fully exploit auxiliary information from homogeneous drug–drug and disease–disease dependencies. They often overlook the fact that the treatment efficacy of drugs can vary due to their unique chemical structures, which may differ from the drug–disease interaction patterns. With the advancement of GNN models,<sup>20–23</sup> more sophisticated architectures have been proposed to capture complex relationships in biomedical networks, offering new opportunities for drug repositioning. Recent research indicates that homogeneous dependency graphs can serve as side information to augment specific node representations. However, effectively integrating and fully utilizing homogeneous and heterogeneous supervised information remains challenging. Recent works<sup>24,25</sup> adopt contrastive learning for drug repositioning; they rely on single-strategy contrastive objectives, limiting their robustness to sparse and noisy interaction data.

To tackle these issues, we propose MDGCN, an innovative approach for drug repositioning that leverages multidependency graph convolutional networks and contrastive learning. Initially, we construct two homogeneous dependency graphs from drug



**Figure 2.** Overall architecture of MDGCN involves three main steps. First, two homogeneous dependency graphs are constructed from drug and disease similarity matrices, and a heterogeneous dependency graph is constructed from the drug–disease association matrix. Second, multilayer graph convolutional networks are utilized to comprehensively learn drug and disease embeddings on these dependency graphs. Third, cross-view and cross-layer contrastive learning are introduced to enhance node embeddings for drug–disease association prediction.

and disease similarity matrices alongside a heterogeneous dependency graph using the drug–disease association matrix. Unlike existing methods that primarily focus on the heterogeneous graph, MDGCN systematically incorporates multiple dependencies to propagate node-specific side information from homogeneous graph through multilayer GCNs, facilitating the learning of richer drug and disease representations with the comprehensive integration of structural and semantic information across dependency graphs. Furthermore, MDGCN employs a dual-strategy approach that simultaneously performs cross-view and cross-layer contrastive learning. Cross-view contrastive learning generates self-supervision signals for data augmentation through the consistency among diverse views, while cross-layer contrastive learning further enhances hierarchical feature interactions, preserving and refining information across different graph layers. This joint strategy not only improves representation learning but also mitigates issues related to data sparsity and noise in drug repositioning tasks. Ultimately, the predictor utilizes the optimized embeddings of drug and disease nodes to infer potential drug–disease associations.

In summary, the main contributions of this work can be outlined as follows:

- MDGCN proposes a general framework for cross-view information exchange, which constructs multidependency graphs from drug–drug, disease–disease similarities, and drug–disease associations and employs multilayer graph convolutional networks to embed node features and propagate node-specific side information across diverse graphs.
- MDGCN introduces two different contrastive learning strategies to enhance its representation learning capacity. The cross-view contrastive learning is to align node embeddings in different dependency graphs, and cross-

layer contrastive learning further coordinates high-order semantics and low-order features.

- Experiments on four benchmark data sets demonstrate that MDGCN outperforms seven state-of-the-art methods in drug–disease association prediction. Furthermore, additional experiments on a multirelational data set validate the effectiveness of MDGCN in handling complex heterogeneous graphs, further demonstrating its adaptability in drug–disease association prediction.

Figure 1 illustrates a toy example of multidependency graphs, providing a clear explanation of how the graph convolution model integrates multiple dependencies. During the graph convolution process, the collaboration between the homogeneous graphs in Figure 1(b,c) enables the target nodes  $r_1$  and  $d_1$  in Figure 1(a) to focus earlier on their respective one-hop neighbors ( $r_2, r_5$ ) and ( $d_2, d_4$ ). This facilitates the filtering of noisy information when aggregating two-hop neighbors ( $r_2, r_3, r_4, r_5$ ), as  $r_3$  and  $r_4$  may not be as important as  $r_2$  and  $r_5$ . Similarly, for target node  $d_1$  in Figure 1(a), greater attention may be paid to  $d_2$  rather than  $d_3$  among its two-hop neighbors. Furthermore, in Figure 1(b), nodes  $r_2, r_4, r_6$ , and  $r_7$  at the two-hop level might exhibit equal importance to  $r_1$ . However, through the collaboration with two-hop neighbors ( $r_2, r_3, r_4, r_5$ ) of  $r_1$  in Figure 1(a),  $r_1$  can effectively distinguish more influential nodes such as  $r_2$  and  $r_4$ . Similarly, for  $d_1$ , the two-hop neighbors  $d_2$  and  $d_3$  in Figure 1(c) may be more significant compared to  $d_5$  and  $d_6$ . The multidependency graphs collaboratively interact at each layer of the graph convolution process, enabling target nodes to suppress noisy information, focus on more relevant signals, and ultimately learn more discriminative node embeddings.

## MATERIALS AND METHODS

**Overview of the Proposed MDGCN Model.** MDGCN is a novel drug repositioning method based on multidependency graph convolutional networks and contrastive learning. It

formulates drug repositioning as a link prediction problem, determining whether a therapeutic relationship exists between a specific drug and a target disease. As illustrated in Figure 2, MDGCN involves three main steps. First, we construct drug-related and disease-related graphs based on the drug similarity matrix, disease similarity matrix, and drug–disease association matrix, initializing drug and disease embeddings according to their IDs. Then, multidependency graph learning is introduced to extract and fuse isomorphic and heteromorphic information, updates node embeddings across different dependency graphs, and aggregates layer-specific representations. Lastly, we introduce cross-view contrastive learning to align node embeddings in different dependency graphs and further utilize cross-layer contrastive learning to balance high-order semantics and low-order features.

**Construction of Multidependency Graphs.** *Homogeneous Dependency Graph.* Each benchmark data set provides a drug–drug similarity matrix  $S^r \in \mathbb{R}^{M \times M}$  and a disease–disease similarity matrix  $S^d \in \mathbb{R}^{N \times N}$ , where  $M$  is the number of drugs and  $N$  is the number of diseases. These similarity matrices provide information about drug substructures and semantic information about disease phenotypes. And  $S^r(i, j)$  represents the similarity between drug  $i$  and drug  $j$ , while  $S^d(i, j)$  represents the similarity between disease  $i$  and disease  $j$ . As the dense similarity information may lead to overly smooth features learned by the graph convolutional networks,<sup>26</sup> we first process  $S^r$  and  $S^d$  into homogeneous but sparse KNN graphs,  $\mathcal{G}^r = \{\mathcal{V}_r, \mathcal{E}_{rr}\}$  and  $\mathcal{G}^d = \{\mathcal{V}_d, \mathcal{E}_{dd}\}$ , to capture isomorphic dependency information. Here,  $\mathcal{V}_r$  and  $\mathcal{V}_d$ , respectively, represent the set of drug and disease nodes, while  $\mathcal{E}_{rr}$  and  $\mathcal{E}_{dd}$  represent the set of edges of drug pairs and disease pairs. The adjacency matrix corresponding to  $\mathcal{G}^r$  is denoted as a binary matrix  $A^r \in \mathbb{R}^{M \times M}$ , where each entry of  $A^r$  is computed based on the similarity between each pair of drugs. For convenience, the  $i$ th drug in the data set is denoted as  $r_i$ , where  $i = 1, 2, \dots, M$ . Then, the entry  $A_{ij}^r$  is defined as

$$A_{ij}^r = \begin{cases} 1, & \text{if } r_j \in \tilde{N}_k(r_i) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\tilde{N}_k(r_i) = \{r_i\} \cup N_k(r_i)$  is the extended  $k$ -nearest neighbor set of  $r_i$  including  $r_i$  itself and  $N_k(r_i)$  is the  $k$ -nearest neighbors of drug  $r_i$ . Similarly,  $d_i$  represents the  $i$ th disease in the data set, where  $i = 1, 2, \dots, N$ . The entry  $A_{ij}^d$  of the binary adjacency matrix  $A^d \in \mathbb{R}^{N \times N}$  corresponding to  $\mathcal{G}^d$  is defined as

$$A_{ij}^d = \begin{cases} 1, & \text{if } d_j \in \tilde{N}_k(d_i) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where  $\tilde{N}_k(d_i) = \{d_i\} \cup N_k(d_i)$  is the extended  $k$ -nearest neighbor set of  $d_i$  including  $d_i$  itself, and  $N_k(d_i)$  is the  $k$ -nearest neighbor of disease  $d_i$ .

**Heterogeneous Dependency Graph.** Based on the drug–disease association matrix and zero matrices, we construct the heterogeneous dependency graph. An association network between drugs and diseases is depicted by a binary matrix  $A \in \mathbb{R}^{M \times N}$ , where  $M$  and  $N$ , respectively, denote the quantity of drug nodes and disease nodes. The therapeutic relationship between a target drug  $r_i$  and disease  $d_j$  is represented by each element  $A_{ij} \in \{0, 1\}$  in the matrix. If drug  $r_i$  and disease  $d_j$  are

associated, then  $A_{ij} = 1$ ; otherwise,  $A_{ij} = 0$ . Using the validated drug–disease relationships matrix  $A \in \mathbb{R}^{M \times N}$  and the zero matrices  $O^r \in \mathbb{R}^{M \times M}$  and  $O^d \in \mathbb{R}^{N \times N}$ , we can construct the heterogeneous graph  $\mathcal{G}^{\text{rd}} = \{\mathcal{V}_r, \mathcal{V}_d, \mathcal{E}_{\text{rd}}\}$ , where the sets of drug nodes and disease nodes are represented by  $\mathcal{V}_r$  and  $\mathcal{V}_d$ , while  $\mathcal{E}_{\text{rd}}$  represents the set of edges between them. The adjacency matrix corresponding to  $\mathcal{G}^{\text{rd}}$  can be expressed as

$$A^{\text{rd}} = \begin{bmatrix} O^r & A \\ A^T & O^d \end{bmatrix} \in \mathbb{R}^{(M+N) \times (M+N)} \quad (3)$$

**Multidependency Graph Learning.** For the constructed multidependency graphs, we first initialize the node-specific embedding matrices  $X_r^0 \in \mathbb{R}^{M \times d}$  and  $X_d^0 \in \mathbb{R}^{N \times d}$  by the Xavier initializer.<sup>27</sup> For drug–disease dependency graph  $\mathcal{G}^{\text{rd}}$ , we concatenate  $X_r^0$  and  $X_d^0$  to obtain  $X_{\text{rd}}^0 = X_r^0 \| X_d^0$  as its input. For the homogeneous dependency graph  $\mathcal{G}^r$  and  $\mathcal{G}^d$ , a self-gating module<sup>28</sup> are respectively trained to derive drug dependency-aware embedding  $X_{rr}^0$  and disease dependency-aware embedding  $X_{dd}^0$  from the initial node-specific embedding. Then,  $X_{rr}^0$  and  $X_{dd}^0$  are used as the inputs of  $\mathcal{G}^r$  and  $\mathcal{G}^d$ . Specifically, the embedding  $X_{rr}^0$  and  $X_{dd}^0$  are calculated as follows:

$$\begin{cases} X_{rr}^0 = X_r^0 \odot \sigma(X_r^0 W_{g_r} + b_{g_r}) \\ X_{dd}^0 = X_d^0 \odot \sigma(X_d^0 W_{g_d} + b_{g_d}) \end{cases} \quad (4)$$

where  $X_{rr}^0 \in \mathbb{R}^{M \times d}$  and  $X_{dd}^0 \in \mathbb{R}^{N \times d}$ , and  $\sigma(\cdot)$  represents the sigmoid activation function, and  $\odot$  denotes element-wise multiplication.  $W_{g_r}$  and  $W_{g_d} \in \mathbb{R}^{d \times d}$  are transformations, while  $b_{g_r}$  and  $b_{g_d} \in \mathbb{R}^{1 \times d}$  for bias. With skip connections, the self-gating mechanism enables the representations  $X_{rr}^0$  and  $X_{dd}^0$  to incorporate dependency from drug–drug and disease–disease domains while retaining common semantic information from the initial features  $X_r^0$  and  $X_d^0$ .

Subsequently, we employ multilayer graph convolutional networks (GCNs) to learn embeddings from graph  $\mathcal{G}^r$ ,  $\mathcal{G}^d$ , and  $\mathcal{G}^{\text{rd}}$ . To broaden the information perception of each dependency graph and facilitate information interaction, we configure LightGCN<sup>29</sup> as the graph message-passing paradigm to fuse and update the embeddings across layers of these graphs. The GCNs are applied as the encoder for the three dependency graphs:

$$\begin{cases} X_{rr}^l = D_r^{-1/2} A^r D_r^{-1/2} \hat{X}_{rr}^{l-1} \\ X_{dd}^l = D_d^{-1/2} A^d D_d^{-1/2} \hat{X}_{dd}^{l-1} \\ X_{rd}^l = D_{\text{rd}}^{-1/2} A^{\text{rd}} D_{\text{rd}}^{-1/2} \hat{X}_{rd}^{l-1} \end{cases} \quad (5)$$

where  $X_{rr}^l \in \mathbb{R}^{M \times d}$  and  $X_{dd}^l \in \mathbb{R}^{N \times d}$  represent the latent features of drug and disease nodes, respectively. The joint node embeddings at the  $l_{\text{th}}$  iteration are denoted by  $X_{\text{rd}}^l \in \mathbb{R}^{(M+N) \times d}$ . The input for the  $0_{\text{th}}$  layer is  $X_{rr}^0$ ,  $X_{dd}^0$ , and  $X_{rd}^0$ . The iteration is  $1 \leq l \leq L$ , where  $L$  represents the maximum number of GCN layers. Similar to soft meta-path design,<sup>30</sup> information in each layer is aggregated from multihop neighbors. Through multiple iterations of message propagation, high-order embeddings retain multihop connected heteroge-



neous semantics. Inspired by XSimGCL,<sup>31</sup> contrastive recommendation models can be enhanced by fine-tuning the uniformity of learned representations within a specific range. After each convolution operation, noise perturbation is added to the joint embeddings of drugs and diseases to achieve effective representation-level augmentation:

$$X_{rd}^{l'} = X_{rd}^l + \Delta' \quad (6)$$

where  $\Delta'$  represents the noise vector and  $\|\Delta\|_2 = \epsilon$ , with  $\epsilon$  being a small constant. Furthermore, node embeddings are updated through the information fusion process:

$$\begin{cases} \hat{X}_{rr}^l = (X_{rr}^l + X_r^l)/2 \\ \hat{X}_{dd}^l = (X_{dd}^l + X_d^l)/2 \\ \hat{X}_{rd}^l = \hat{X}_{rr}^l \parallel \hat{X}_{dd}^l \end{cases} \quad (7)$$

where  $(X_r^l, X_d^l) = \text{split}(X_{rd}^l)$ . These fused representations are fed into the next layer of the encoder.

By capturing the outputs at each layer of the multidependency GCNs and aggregating heterogeneous and homogeneous information, we calculate the overall representation related to drugs and diseases as follows:

$$\begin{cases} X_{rr} = \text{mean}\left(X_{rr}^0 + \sum_{l=1}^L \frac{X_{rr}^l}{\|X_{rr}^l\|}\right) \\ X_{dd} = \text{mean}\left(X_{dd}^0 + \sum_{l=1}^L \frac{X_{dd}^l}{\|X_{dd}^l\|}\right) \\ X_{rd} = \text{mean}\left(X_{rd}^0 + \sum_{l=1}^{L-3} \frac{X_{rd}^{l'}}{\|X_{rd}^{l'}\|}\right) \end{cases} \quad (8)$$

where  $X_{rr}^0$ ,  $X_{dd}^0$ , and  $X_{rd}^0$  are initial embeddings, which are incorporated by skip connections. Meanwhile, we normalized the output of each GCN layer. Notably, during the multidependency GCN process, the continuous sharing and integration of information between homogeneous and heterogeneous views may gradually weaken the discriminative power of node representations. To address this, we deliberately discard the outputs of the last three layers when calculating the final joint node embeddings on the heterogeneous view, better supporting the subsequent cross-view contrastive learning.

Finally, the node embeddings on the drug–drug and disease–disease views are enhanced to optimize the embeddings encoded on the drug–disease view. The fusion process is as follows:

$$\begin{cases} \tilde{X}_r = (1 - \omega)X_r + \omega X_{rr} \\ \tilde{X}_d = (1 - \omega)X_d + \omega X_{dd} \end{cases} \quad (9)$$

where  $(X_r, X_d) = \text{split}(X_{rd})$  and  $0 \leq \omega \leq 1$  is a hyperparameter that controls the weight between the embeddings of the drug–disease view and the drug–drug (or disease–disease) view. Here, the embeddings of the homogeneous views,  $X_{rr}$  and  $X_{dd}$ , are incorporated for better optimization. The ultimate embeddings  $\tilde{X}_r \in \mathbb{R}^{M \times d}$  for drugs and  $\tilde{X}_d \in \mathbb{R}^{N \times d}$  corresponding to diseases are utilized for the drug repurposing task.

**Cross-View and Cross-Layer Contrastive Learning.** To address the challenge of label sparsity, we integrate contrastive learning into MDGCN to enhance its representation learning

capacity. Contrastive learning not only facilitates the model to capture meaningful patterns from sparsely labeled data but also enables more robust representation alignment across heterogeneous and homogeneous views. Specifically, we adopt two contrastive learning strategies for MDGCN, including cross-view and cross-layer contrastive learning. These strategies are introduced to leverage multiview embeddings and hierarchical information flow, enabling the model to effectively balance low-order and high-order feature dependencies.

**Cross-View Contrastive Learning.** The heterogeneous view of drugs and diseases reflects their complex dependencies through the association network, while the homogeneous views focus more on the internal similarities within a single group. The cross-view contrastive learning<sup>32</sup> paradigm is based on the two sets of learned embeddings for drugs and diseases to align the embeddings of these two views and obtain more robust representations:

$$\begin{cases} \mathcal{L}_{\text{view-cl}_r} = \sum_{r \in \mathcal{V}_r} -\log \frac{\exp(\text{sim}(x_r, x_{rr})/\tau)}{\sum_{r' \in \mathcal{V}_r} \exp(\text{sim}(x_r, x_{rr}')/\tau)} \\ \mathcal{L}_{\text{view-cl}_d} = \sum_{d \in \mathcal{V}_d} -\log \frac{\exp(\text{sim}(x_d, x_{dd})/\tau)}{\sum_{d' \in \mathcal{V}_d} \exp(\text{sim}(x_d, x_{dd}')/\tau)} \end{cases} \quad (10)$$

where  $x_r, x_{rr} \in \mathbb{R}^d$  are respectively embeddings from  $X_r, X_{rr}$  and  $x_d, x_{dd} \in \mathbb{R}^d$  are embeddings from  $X_d, X_{dd}$ . Additionally,  $\text{sim}(\cdot)$  denotes the similarity function. In this work, cosine similarity is employed as the measure of similarity.  $\tau$  represents the temperature coefficient, which is capable of automatically identifying difficult negative instances.  $x_{rr}'$  or  $x_{dd}'$  denotes the embedding of the negative sample whose index ( $r'$  or  $d'$ ) differs from  $r$  or  $d$ , i.e., an unrelated drug or disease node from homogeneous views.

Finally, the complete cross-view contrastive loss is

$$\mathcal{L}_{\text{view-cl}} = \alpha_r \mathcal{L}_{\text{view-cl}_r} + \alpha_d \mathcal{L}_{\text{view-cl}_d} \quad (11)$$

where  $\alpha_r$  and  $\alpha_d$  are hyperparameters used to adjust the loss weights. With the assistance of self-supervised signals, the homogeneous view contributes side information to enhance the accuracy of drug–disease interaction modeling.  $\mathcal{L}_{\text{view-cl}}$  achieves minimization by bringing closer positive instance pairs and separating negative ones, enabling the model to generate more discriminative embeddings for the drug–disease interaction prediction.

**Cross-Layer Contrastive Learning.** As indicated in ref, <sup>33</sup> contrastive learning achieves optimal performance when the mutual information between correlated views is maintained at a balanced level. In cross-view contrastive learning, the mutual information between two views of the same node tends to be significantly high as both embeddings encapsulate information from  $L$ -hop neighbors. And information fusion is applied at each layer of the encoder; the contrastive effectiveness may be insufficient. According to the prior studies,<sup>31,34</sup> we aim to further contrast the embeddings from different layers, enabling stronger information flow between layers and preventing node embeddings from becoming overly similar. Specifically, the output of the last layer for the drug–disease view is aligned with the input of the first layer of the encoder:

$$\begin{cases} \mathcal{L}_{\text{layer-cl}_r} = \sum_{r' \in \mathcal{V}_r} -\log \frac{\exp(\text{sim}(\mathbf{x}_r^L, \mathbf{x}_{r'}^0)/\tau)}{\sum_{r' \in \mathcal{V}_r} \exp(\text{sim}(\mathbf{x}_r^L, \mathbf{x}_{r'}^0)/\tau)} \\ \mathcal{L}_{\text{layer-cl}_d} = \sum_{d' \in \mathcal{V}_d} -\log \frac{\exp(\text{sim}(\mathbf{x}_d^L, \mathbf{x}_{d'}^0)/\tau)}{\sum_{d' \in \mathcal{V}_d} \exp(\text{sim}(\mathbf{x}_d^L, \mathbf{x}_{d'}^0)/\tau)} \end{cases} \quad (12)$$

where  $(\mathbf{x}_r^L, \mathbf{x}_d^L) = \text{split}(\mathbf{x}_{rd}^{L'})$  and  $\mathbf{x}_{rd}^{L'}$  is the embedding from  $X_{rd}^{L'}$ . And the complete cross-layer contrastive objective function is as follows:

$$\mathcal{L}_{\text{layer-cl}} = \mathcal{L}_{\text{layer-cl}_r} + \mathcal{L}_{\text{layer-cl}_d} \quad (13)$$

Formally,  $X_r^0$  and  $X_d^0$  retain some original low-order features, while  $X_{rd}^{L'}$  can be regarded as higher-order context embeddings, which is a combination of embeddings within the subgraphs containing neighbors from multiple hops. Through cross-layer comparison, the similarity of positive instances between the context embeddings  $X_r^L$  or  $X_d^L$  and the corresponding initial embeddings  $X_r^0$  or  $X_d^0$  is maximized, while negative instances are effectively distinguished. Benefiting from this, the model can better coordinate the high-order semantics with the low-order features to improve the information alignment effect across multiple views.

**Prediction and Optimization.** Our predictor uses an inner product method to calculate the likelihood between the target drug and disease. The label is obtained through an activation function:  $\tilde{y}_{ij} = \sigma(\tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_j)$ , where  $\tilde{\mathbf{x}}_i$  corresponds to the drug-specific representation of  $r_i$  derived from the final fused matrix  $\tilde{X}_r$ , while  $\tilde{\mathbf{x}}_j$  denotes the disease-specific representation of  $d_j$  from the  $\tilde{X}_d$ . And  $\sigma(\cdot)$  represents the sigmoid activation function.

As our principal loss function, we utilize binary cross-entropy loss:

$$\mathcal{L}_{\text{main}} = -\sum_{i,j} (1 - y_{ij}) \log(1 - \tilde{y}_{ij}) + y_{ij} \log(\tilde{y}_{ij}) \quad (14)$$

where  $y_{ij}$  represents the actual label for the target nodes corresponding to  $r_i$  and  $d_j$ . By weighting the main loss and contrastive loss, the final loss is formulated as

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \alpha(\mathcal{L}_{\text{view-cl}} + \beta \mathcal{L}_{\text{layer-cl}}) \quad (15)$$

where  $\alpha$  is used to adjust the weights of the contrastive loss and main loss and  $\beta$  is for balancing the contribution of the cross-layer contrastive loss in the overall objective.

**Data Sets.** In recent studies,<sup>12,14,25</sup> four representative data sets, including Fdata set,<sup>35</sup> Cdata set,<sup>36</sup> LRSSL,<sup>37</sup> and Ldata set,<sup>38</sup> are widely used in drug–disease association prediction tasks. MDGCN's effectiveness is evaluated across these four benchmark data sets. Specifically, the Fdata set comprises 313 diseases from the OMIM<sup>39</sup> database and 593 drugs from the DrugBank<sup>40</sup> database, along with 1933 known associations between them. A total of 2532 validated drug–disease relationships are included in the Cdata set, in addition to 409 medications that are listed in DrugBank and 663 disorders collected from the OMIM repository. Similarly, LRSSL contains 763 drugs and 681 diseases, with 3051 validated relationships between drugs and diseases. The Ldata set is sourced from the CTD<sup>41</sup> data set, comprising 18416 known drug–disease associations involving 269 drugs and 598 diseases. This large-scale interaction network enables rigorous evaluation of MDGCN's scalability on densely connected graphs. To construct the homogeneous graph for each benchmark data

set, drug similarity features are derived from chemical substructures, and disease phenotype similarities are based on semantic features. A summary of these data sets is presented in Table 1. To further assess MDGCN's adaptability to complex

**Table 1. Summary of Four Benchmark Data Sets with Sparsity Analysis**

data sets	Fdata set	Cdata set	LRSSL	Ldata set
number of drugs	593	663	763	269
number of diseases	313	409	681	598
identified associations	1933	2532	3051	18416
sparsity	0.0104	0.0093	0.0059	0.1145

biological networks, we introduce the DB-KEGG data set from MGATRx,<sup>42</sup> which encompasses various biological relationships, as shown in Table 2. The drug and disease similarity matrices used as input for MDGCN are generated by averaging Jaccard similarity matrices computed from multiple relationship adjacency matrices.

**Table 2. Statistics of the DB-KEGG Data Set**

edge types	number of edges	size
drug–disease	8957	(4008, 2958)
disease–targets	27,245	(2958, 8722)
drug–targets	11,146	(4008, 8722)
drug–substructures	422,742	(4008, 881)
drug–sideeffects	297,999	(4008, 11744)
drug–MeSH	35,277	(4008, 2124)

**Evaluation Metrics.** We make use of two metrics that are frequently employed, namely, AUROC and AUPRC, in order to evaluate the effectiveness of various approaches related to drug repositioning. Specifically, the receiver operating characteristic, often known as the ROC, is a probability curve that makes a comparison between the true positive rate (TPR) and the false positive rate (FPR).<sup>43</sup> The area under the ROC curve is denoted by the acronym AUROC. AUPRC is obtained by calculating the area under the precision-recall curve, which measures the relationship between precision and recall at various thresholds.<sup>44</sup> The higher AUROC and AUPRC scores suggest that the performance is better. When there is a class imbalance problem, AUPRC may be more informative than AUROC. The F1-score is also computed on the DB-KEGG data set, using the optimal threshold determined by the elbow method on the AUPR curve.

$$\text{TPR(or recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (17)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (18)$$

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (19)$$

where TP and TN are, respectively, the number of correctly predicted known and unknown drug–disease associations and FP and FN are the number of incorrectly predicted known and unknown drug–disease associations, respectively.

## RESULTS

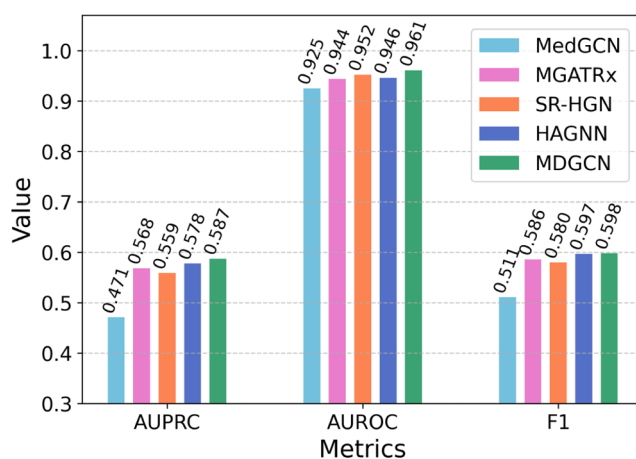
**Parameter Setting.** Within the framework of PyTorch, the MDGCN paradigm is put into action. The parameter is

**Table 3. Hyperparameter Setting of MDGCN**

hyperparameter	Fdata set	Cdata set	LRSSL	Ldata set	DB-KEGG
$\epsilon$ : the training epoch	110	60	50	20	10
$\eta$ : the decay factor	0.99	0.99	0.99	0.99	0.99
$\tau$ : the temperature coefficient	0.5	0.5	0.5	0.5	0.5
$\epsilon$ : a small constant to control noise	0.3	0.6	0.2	0.3	0.2
$\omega$ : the coefficient for embedding fusion	0.1	0.3	0.2	0.5	0.2
$lr$ : the rate of learning	0.05	0.055	0.055	0.1	0.1
$\alpha_r$ : the weight of $\mathcal{L}_{\text{view-cl}_r}$	0.068	0.068	0.08	0.064	0.08
$\alpha_d$ : the weight of $\mathcal{L}_{\text{view-cl}_d}$	0.088	0.085	0.09	0.085	0.09
$\alpha$ : the weight of all contrastive loss	0.09	0.09	0.075	0.025	0.09
$\beta$ : the weight of $\mathcal{L}_{\text{layer-cl}}$	0.01	0.01	0.01	0.01	0.01
$d$ : the dimensionality of embeddings	128	512	256	256	128
$k$ : the quantity of proximate neighbors	3	4	7	6	7

optimized using the Adam. Following empirical adjustments, we specified 4096 as the batch size across the four benchmark data sets and 20,000 for DB-KEGG. Furthermore, the values of our model's customized hyperparameters for various data sets are provided in Table 3. The exhaustive tuning procedure for the dimensionality  $d$  of learned embeddings and the quantity  $k$  of proximate neighbors on representative data sets is delineated in Hyperparameter Analysis Section. In the experiments, baseline models are configured to the default values, as specified in the literature.

**Baselines.** To assess the effectiveness of MDGCN, a comparative analysis is conducted between MDGCN and seven advanced approaches across four gold-standard data sets. Included in these baseline approaches are two models (iDrug, BNNR) that rely on matrix completion, two models (DRWBNCF, DRHGCN) utilizing graph convolutional networks, and three models (HCCF, DRGCL, AutoDR) based on the GCNs and contrastive learning. The detailed descriptions of these baselines are as follows:



**Figure 3.** Comparison of MDGCN and other models using a 10-fold cross-validation on DB-KEGG.

- iDrug<sup>45</sup> develops a unified model to obtain cross-domain representations for drug-related prediction.
- BNNR<sup>11</sup> is a bounded nuclear norm regularization technique based on matrix completion to predict new drug indications.
- DRWBNCF<sup>17</sup> adopts neural collaborative filtering techniques to infer new potential drugs for diseases based on neighborhood interaction.
- DRHGCN<sup>18</sup> learns dual-domain embeddings of nodes by applying graph convolution operations.
- HCCF<sup>46</sup> introduces a hypergraph-based self-supervised framework for capturing local and global collaborative relations through contrastive learning.
- DRGCL<sup>24</sup> employs contrastive learning through topology and semantic graphs for drug repositioning.
- AutoDR<sup>25</sup> is a collaborative learning framework that incorporates neighbor interaction levels into GNN's message-passing process.

To further evaluate the adaptability and performance of MDGCN on multirelational biomedical networks, we benchmark it against four representative models, including MedGCN, MGATRx, SR-HGN, and HAGNN. Specifically, MedGCN<sup>47</sup> leverages a multiview graph convolutional network on the complex associations between multiple medical entities for medication recommendation. MGATRx<sup>42</sup> utilizes a multiview graph attention mechanism to aggregate relevant neighbors for drug repositioning. SR-HGN<sup>48</sup> proposes node-level and type-

**Table 4. Comparison of MDGCN and Seven Other Models Using 10-Fold Cross-Validation**

data sets	iDrug	BNNR	DRWBNCF	DRHGCN	HCCF	DRGCL	AutoDR	MDGCN
AUPRC								
Fdata set	0.167 ± 0.027	0.328 ± 0.029	0.484 ± 0.027	0.490 ± 0.041	0.483 ± 0.036	0.583 ± 0.042	0.570 ± 0.041	0.602 ± 0.041
Cdata set	0.250 ± 0.027	0.431 ± 0.020	0.559 ± 0.021	0.580 ± 0.035	0.562 ± 0.031	0.671 ± 0.034	0.631 ± 0.043	0.694 ± 0.028
LRSSL	0.070 ± 0.009	0.226 ± 0.021	0.349 ± 0.034	0.384 ± 0.022	0.364 ± 0.031	0.468 ± 0.056	0.401 ± 0.043	0.501 ± 0.044
Ldata set	0.086 ± 0.004	0.142 ± 0.007	0.419 ± 0.006	0.498 ± 0.012	0.527 ± 0.008	0.566 ± 0.009	0.530 ± 0.014	0.573 ± 0.009
average	0.143	0.282	0.453	0.488	0.484	0.572	0.533	0.593
AUROC								
Fdata set	0.905 ± 0.019	0.937 ± 0.010	0.923 ± 0.013	0.948 ± 0.011	0.900 ± 0.010	0.955 ± 0.008	0.951 ± 0.005	0.961 ± 0.010
Cdata set	0.926 ± 0.010	0.952 ± 0.010	0.941 ± 0.011	0.964 ± 0.005	0.914 ± 0.012	0.969 ± 0.004	0.963 ± 0.009	0.972 ± 0.006
LRSSL	0.900 ± 0.008	0.922 ± 0.012	0.935 ± 0.011	0.961 ± 0.006	0.885 ± 0.011	0.953 ± 0.009	0.948 ± 0.010	0.965 ± 0.008
Ldata set	0.838 ± 0.005	0.866 ± 0.004	0.824 ± 0.005	0.851 ± 0.007	0.864 ± 0.003	0.879 ± 0.003	0.864 ± 0.004	0.881 ± 0.003
average	0.892	0.919	0.906	0.931	0.891	0.939	0.932	0.945

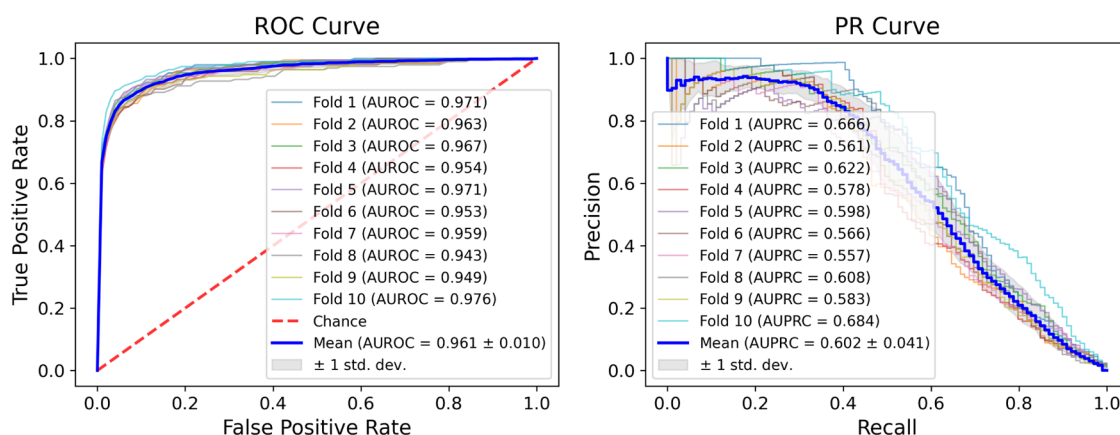


Figure 4. Performance of MDGCN on Fdata set using 10-fold cross-validation.

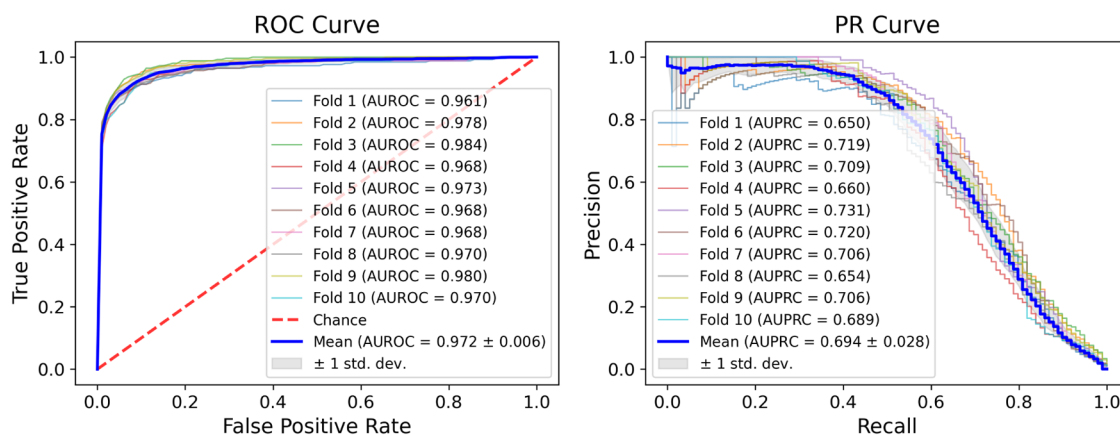


Figure 5. Performance of MDGCN on the Cdata set using 10-fold cross-validation.

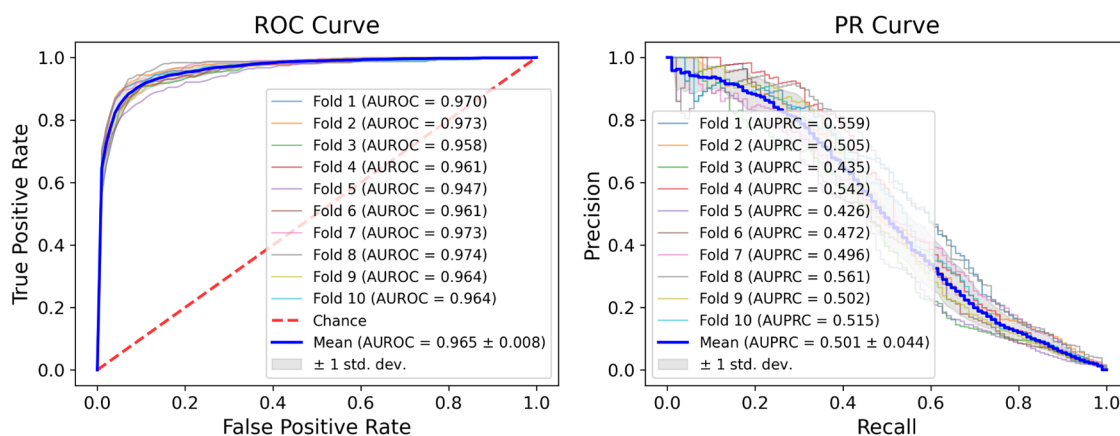


Figure 6. Performance of MDGCN on LRSSL using 10-fold cross-validation.

level aggregation to hierarchically learn node representations in heterogeneous information networks. And HAGNN<sup>49</sup> introduces the hybrid aggregation based on the fused meta-path graph to exploit the semantic information for link prediction.

**Performance Comparison with Baselines on Cross-Validation.** To ensure that the performance of MDGCN is appropriately evaluated, we carry out a 10-fold cross-validation assessment. Following previous studies,<sup>18,19</sup> positive and negative instances are first shuffled independently and then divided into ten subsets, respectively. The test set consists of one positive subset and the corresponding negative subset for each

fold, while the training set consists of all of the remaining subsets. Additionally, to mitigate potential data biases in cross-validation, MDGCN and the comparison models are evaluated over 10 repetitions. The results on the four benchmark data sets (Fdata set, Cdata set, LRSSL, Ldata set) are presented in Table 4.

We observe that MDGCN consistently outperforms all of the compared models across the four data sets in the 10-fold cross-validation, achieving superior performance in both AUPRC and AUROC metrics. Notably, AUPRC is particularly valuable for imbalanced data sets, as it directly reflects the model's ability to



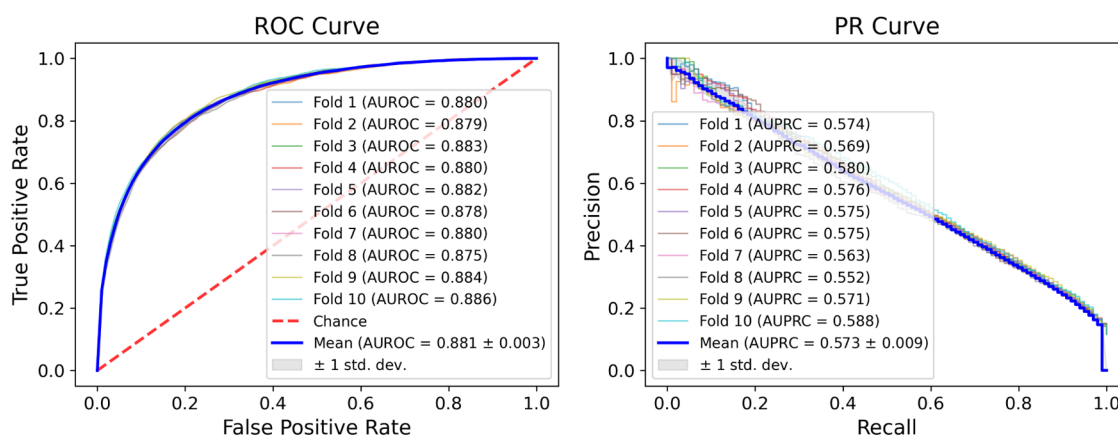


Figure 7. Performance of MDGCN on Ldata set using 10-fold cross-validation.

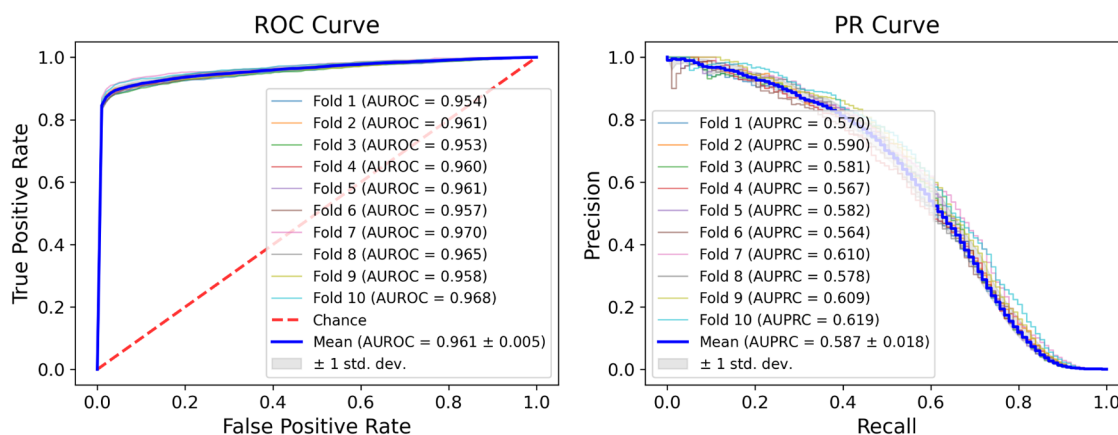


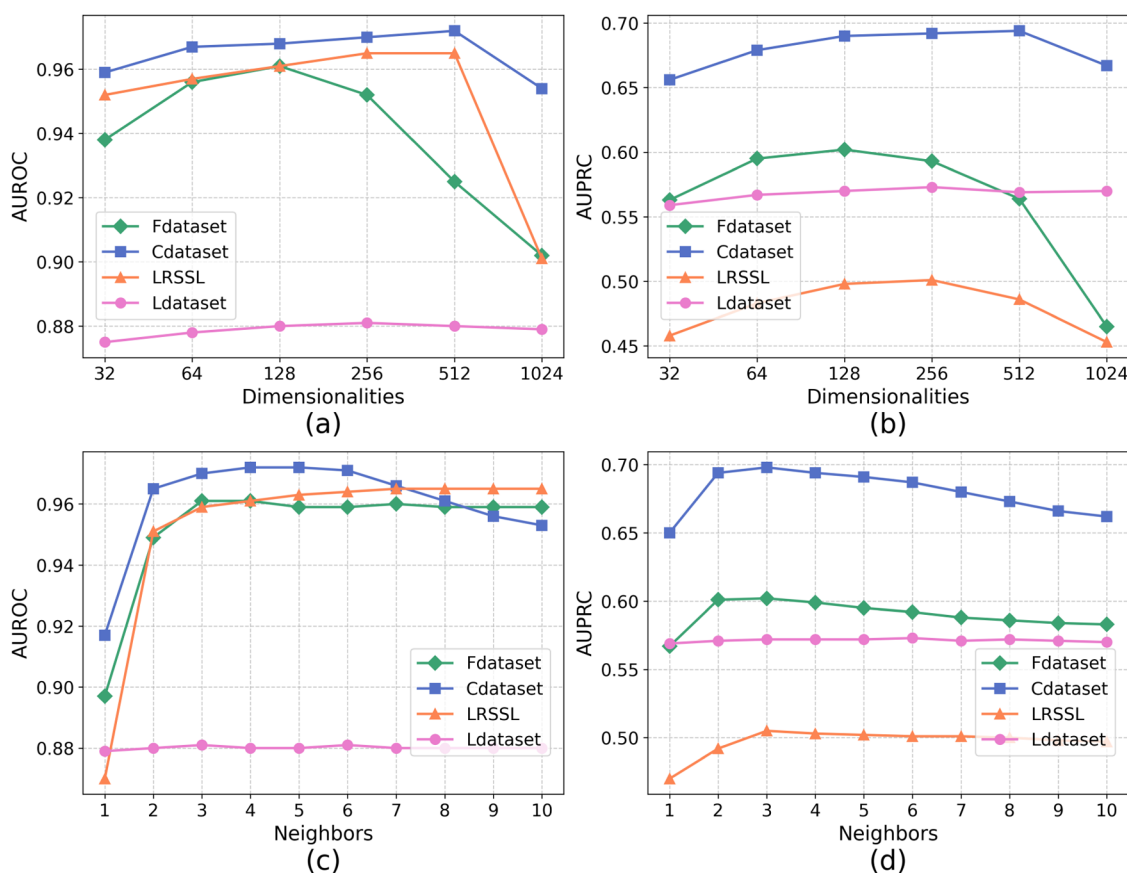
Figure 8. Performance of MDGCN on DB-KEGG using 10-fold cross-validation.

distinguish positive samples from negatives without being influenced by the true negative rate. MDGCN achieves the best AUPRC performance with an average value of 0.593, outperforming the second-best model by 1.9, 2.3, 3.3, and 0.7% across the Fdata set, Cdata set, LRSSL, and Ldata set data sets, respectively. However, when the overall model performance is assessed, AUROC serves as a more widely adopted and reliable metric. By simultaneously considering sensitivity and specificity, AUROC evaluates the global ranking of both positive and negative samples, providing a measure of the model's overall discriminative ability. A higher AUROC indicates the stronger capability of MDGCN in identifying positive samples while reducing the misclassification rate of negative ones. Specifically, MDGCN achieves the highest average AUROC value of 0.945, surpassing the suboptimal model DRGCL by 0.6% and outperforming AutoDR and DRHGCN by 1.3 and 1.4%, respectively, highlighting its superior overall performance. Although the improvement in AUROC is relatively modest, MDGCN demonstrates a substantial advantage in AUPRC, with an average increase of 2.1, 6, and 10.5% compared to DRGCL, AutoDR, and DRHGCN, indicating its remarkable ability in identifying the minority class. The consistent improvement across both AUROC and AUPRC confirms the effectiveness and reliability of MDGCN. Compared to other GCN-based models such as DRHGCN and DRWBNCF, MDGCN clearly outperforms them due to its effective integration of multiple dependency graphs and contrastive learning for learning robust node embeddings. Moreover, when compared with other

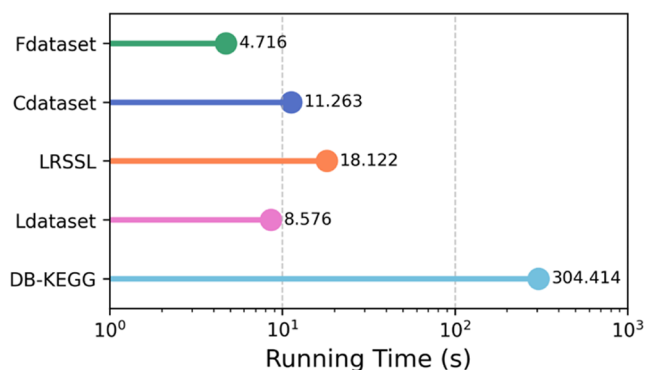
models that also incorporate contrastive learning, such as HCCF, DRGCL, and AutoDR, MDGCN achieves superior performance through its dual-strategy approach, which combines both cross-view and cross-layer contrastive learning.

To further evaluate the adaptability of MDGCN on the multirelational data set, we also evaluate MDGCN on the DB-KEGG data set. As shown in Figure 3, MDGCN achieves the highest scores across all metrics. Compared to the suboptimal model, MDGCN achieves an improvement of 0.9% in AUPRC, 0.9% in AUROC, and 0.1% in F1-score, demonstrating its adaptivity and performance to effectively capture complex dependencies in multirelational graphs. Figures 4, 5, 6, 7 and 8 illustrate the average ROC and PR curves obtained through 10-fold cross-validation of MDGCN across various data sets, showcasing the model's consistent performance over multiple folds. These results strongly demonstrate the superiority and robustness of MDGCN for drug repositioning tasks.

**Hyperparameter Analysis.** In the proposed framework, we leverage multilayer graph convolutional networks to derive informative embeddings of drugs and diseases based on multidependency graphs. Initially, to prevent overly smooth features resulting from dense similarity matrices, the drug and disease similarity matrices are transformed into sparse KNN graphs. Subsequently, multilayer GCNs are applied to learn embeddings. To identify the optimal hyperparameters of MDGCN, we assess the impact of embedding dimensions and the quantity of proximate neighbors across four representative data sets.



**Figure 9.** (a, b) Performance of MDGCN with different dimensionalities of learned embedding. (c, d) Performance of MDGCN with varying numbers of KNN neighbors.



**Figure 10.** Running time of MDGCN per-epoch training across data sets with varying scales.

**Dimensionality of Learned Embeddings.** We experiment with various dimensionalities, specifically {32, 64, 128, 256, 512, and 1024}, across four data sets. As illustrated in Figure 9(a,b), increasing the embedding dimensionality initially improves model performance. However, beyond a certain threshold, the performance begins to decline. The optimal dimensionality differs across data sets, suggesting that the dimensionality should be tailored to the data sparsity and structural characteristics. The best performance is achieved respectively with the dimensionality of 128, 512, 256, and 256 for the Fdata set, the Cdata set, LRSSL, and the Ldata set data sets. This indicates that data sets with lower sparsity may benefit from larger embedding dimensions to capture more comprehensive features.

**Quantity of Proximate Neighbors.** In light of the fact that the design of the KNN graph is essential for determining the degree of similarity between nodes, we examine how the quantity of proximate neighbors influences the model's ability to make accurate predictions. The quantity of proximate neighbors is between one and ten. Increasing the quantity of proximate neighbors effectively enriches the information on the target node, as seen in Figure 9(c,d). The quality of the target node and model performance are lowered as the quantity of neighbors increases because low-similarity node information is introduced, which may be irrelevant or behave as noise. An ideal neighbor selection greatly improves model performance according to experimental findings. For the Fdata set and Cdata set, optimal performance is achieved with 3 and 4 neighbors, respectively. Due to the inherent data sparsity, the LRSSL data set tends to benefit from more neighbors, particularly 7, while the Ldata set achieves the best performance with 6 neighbors.

**Time Complexity Analysis.** The computational complexity of each training iteration arises from two components. (1) Multidependency graph learning, where three LightGCN models perform message passing on different graphs. Each convolution layer involves normalizing adjacency matrices, contributing complexities of  $O(|\mathcal{E}_{rr}|)$ ,  $O(|\mathcal{E}_{dd}|)$ , and  $O(|\mathcal{E}_{rd}|)$ , respectively. As described in eq 5, the core computation involves multiplying sparse adjacency matrix  $A$  with dense feature matrix  $X$ , resulting in complexities of  $O(|\mathcal{E}_{rr}|d)$ ,  $O(|\mathcal{E}_{dd}|d)$ , and  $O(|\mathcal{E}_{rd}|d)$ . Since  $|\mathcal{E}_{rd}| \gg |\mathcal{E}_{rr}|$  and  $|\mathcal{E}_{rd}| \gg |\mathcal{E}_{dd}|$ , the overall complexity is dominated by  $O(L|\mathcal{E}_{rd}|d)$ . (2) Contrastive learning, where pairwise similarity computations within each batch require

Table 5. Performance Comparison of MDGCN and Its Six Variants

variants	Fdata set	Cdata set	LRSSL	Ldata set
AUPRC				
w/o dd	0.599 ± 0.038	0.655 ± 0.019	0.485 ± 0.044	0.483 ± 0.009
w/o rr	0.568 ± 0.042	0.636 ± 0.039	0.465 ± 0.039	0.499 ± 0.010
w/o view-cl	0.538 ± 0.042	0.619 ± 0.035	0.414 ± 0.049	0.573 ± 0.009
w/o layer-cl	0.584 ± 0.042	0.693 ± 0.030	0.493 ± 0.048	0.573 ± 0.009
w/o noise	0.607 ± 0.045	0.694 ± 0.029	0.502 ± 0.045	0.571 ± 0.010
w full layers	0.595 ± 0.040	0.688 ± 0.028	0.493 ± 0.046	0.572 ± 0.008
MDGCN	0.602 ± 0.041	0.694 ± 0.028	0.501 ± 0.044	0.573 ± 0.009
AUROC				
w/o dd	0.915 ± 0.015	0.931 ± 0.014	0.882 ± 0.019	0.842 ± 0.004
w/o rr	0.879 ± 0.019	0.908 ± 0.016	0.902 ± 0.012	0.852 ± 0.004
w/o view-cl	0.898 ± 0.013	0.923 ± 0.010	0.876 ± 0.010	0.880 ± 0.003
w/o layer-cl	0.942 ± 0.010	0.960 ± 0.007	0.944 ± 0.010	0.881 ± 0.003
w/o noise	0.959 ± 0.011	0.969 ± 0.007	0.963 ± 0.009	0.880 ± 0.003
w full layers	0.944 ± 0.010	0.958 ± 0.007	0.945 ± 0.008	0.880 ± 0.003
MDGCN	0.961 ± 0.010	0.972 ± 0.006	0.965 ± 0.008	0.881 ± 0.003

Table 6. Top Ten Potential Drugs Predicted by MDGCN for PD (OMIM: 168600)

rank	DrugBank IDs	candidate medicines	sources
1	DB00989	rivastigmine	CTD, DrugBank, DC, PubChem, ClinicalTrials
2	DB00190	carbidopa	CTD, DrugBank, DC, PubChem, ClinicalTrials
3	DB00268	ropinirole	CTD, DrugBank, DC, PubChem, ClinicalTrials
4	DB00413	pramipexole	CTD, DrugBank, DC, PubChem, ClinicalTrials
5	DB00387	procyclidine	CTD, DrugBank, DC, PubChem
6	DB01186	pergolide	CTD, DrugBank, DC, PubChem, ClinicalTrials
7	DB00245	benzatropine	CTD, DrugBank, DC, PubChem
8	DB00747	scopolamine	CTD
9	DB00810	biperiden	CTD, DrugBank, DC, PubChem
10	DB00376	trihexphenidyl	CTD, DrugBank, DC, PubChem

Table 7. Top Ten Potential Drugs Predicted by MDGCN for AD (OMIM: 104300)

rank	DrugBank IDs	candidate medicines	sources
1	DB00989	rivastigmine	CTD, DrugBank, DC, PubChem, ClinicalTrials
2	DB00163	vitamin E	CTD, ClinicalTrials
3	DB00313	valproic acid	CTD, ClinicalTrials
4	DB00181	baclofen	CTD, ClinicalTrials
5	DB04844	tetrabenazine	Unconfirmed
6	DB01043	memantine	CTD, DrugBank, DC, PubChem, ClinicalTrials
7	DB00843	donepezil	CTD, DrugBank, DC, PubChem, ClinicalTrials
8	DB00190	carbidopa	Unconfirmed
9	DB00413	pramipexole	CTD, ClinicalTrials
10	DB00382	tacrine	CTD, DrugBank, DC, PubChem

$O(B^2d)$  and  $B$  represents the batch size. Consequently, the overall contrastive learning complexity per epoch is given by  $O\left(\frac{MN}{B} \times B^2d\right)$ , which simplifies to  $O(MNBd)$ . Since  $B \gg L$  and  $MN \gg |\mathcal{E}_{rd}|$ , the final overall training complexity can be further approximated as  $O(MNBd)$ , primarily influenced by batch size, drug nodes, disease nodes, and feature dimensionality. Figure 10

illustrates the average per-epoch training time of MDGCN across data sets with varying scales, indicating that the overall training time is primarily dominated by contrastive learning, particularly for data sets with a larger number of nodes such as DB-KEGG.

**Ablation Study.** On the four benchmark data sets, we compare MDGCN with its six variants to investigate the contributions of different model components on prediction performance.

- w/o rr: MDGCN without homogeneous graph  $\mathcal{G}^r$ .
- w/o dd: MDGCN without homogeneous graph  $\mathcal{G}^d$ .
- w/o view-cl: MDGCN without cross-view contrastive loss  $\mathcal{L}_{\text{view-cl}}$ .
- w/o layer-cl: MDGCN without cross-layer contrastive loss  $\mathcal{L}_{\text{layer-cl}}$ .
- w/o noise: MDGCN without noise perturbation after each convolution operation.
- w full layers: MDGCN with full layers when calculating the final joint node embeddings  $X_{rd}$ .

Table 5 presents the performance of MDGCN and its six variants. MDGCN consistently outperforms all ablation variants, affirming the essential role of each component. The removal of either homogeneous graph  $\mathcal{G}^r$  or  $\mathcal{G}^d$  leads to significant performance degradation, demonstrating the importance of leveraging side information derived from drug–drug and disease–disease relationships. Both contrastive learning and the exclusion of the last three layers' outputs significantly affect the performance, while the impact of the noise perturbation is relatively minor. Specifically, removing cross-view contrastive learning causes substantial declines in AUROC and AUPRC, emphasizing its role in optimizing node embeddings by aligning similar nodes across views and distinguishing unrelated ones. Furthermore, eliminating cross-layer contrastive learning results in a noticeable performance degradation, highlighting the role of hierarchical contrastive constraints in strengthening the representational capacity of the multilayer network structure. The impact of noise perturbation is limited, yet it still contributes to mitigating overfitting. The full-layer case underperforms the MDGCN model, underscoring the importance of the synergistic integration of all components for optimal performance.

**Case Studies.** Two case studies on Parkinson's disease (PD) and Alzheimer's disease (AD) are conducted to further validate the proposed model's reliability and efficacy in predicting their candidate drugs. We systematically eliminate the labels pertaining to PD and AD from all established drug–disease relationships in the Fdata set, utilizing the removed associations as the test set and the remainder as the training set. Upon forecasting all absent drug–disease connections, we arrange the outcomes in descending order and concentrate on the top 10 medications identified for PD and AD. To ensure the accuracy of these predictions, we consult trustworthy sources such as CTD, DrugBank, DC, PubChem, and ClinicalTrials. Tables 6 and 7 report ten leading candidate medications for PD and AD as predicted by MDGCN, respectively.

All ten drugs for PD have been validated by diverse sources, indicating a 100% success rate. Previous research has shown that levodopa–carbidopa intestinal gel (LCIG) is widely utilized for the treatment of advanced Parkinson's disease (PD), especially in patients suffering from significant motor fluctuations and dyskinesia.<sup>50,51</sup> Furthermore, subcutaneous scopolamine is documented as an alternate therapy for a terminal cancer patient experiencing severe tremors, who was incapable of ingesting oral drugs for PD.<sup>52</sup>

Alzheimer's disease (AD) represents the most prevalent type of dementia and is a neurodegenerative condition that progressively deteriorates. Our findings indicate that 8 out of 10 drugs, representing an 80% success rate, have been validated by various forms of evidence. For example, the study<sup>53</sup> confirms memantine's safety and effectiveness in relieving AD symptoms. According to previous research,<sup>54</sup> donepezil is a combined acetylcholinesterase inhibitor that helps individuals with mild to severe Alzheimer's disease with their cognitive abilities and overall clinical performance. Although tetraabenazine and carbidopa have not yet been confirmed for the treatment of AD, recent studies are actively exploring their potential therapeutic effects.<sup>55,56</sup> These case studies demonstrate that the predictions of MDGCN have practical significance and scientific basis, providing valuable references for further drug development and disease treatment.

## CONCLUSIONS

This study introduces a novel drug repositioning method, termed MDGCN, which is based on multidependency graph convolutional networks and contrastive learning. MDGCN innovatively constructs multidependency graphs using drug and disease similarity matrices along with the drug–disease association matrix. By employing multilayer graph convolutional networks, it embeds node features and propagates node-specific side information across diverse graphs. Meanwhile, MDGCN introduces two different contrastive learning strategies to enhance its representation learning capacity. Specifically, cross-view contrastive learning is to align node embeddings in different dependency graphs, and cross-layer contrastive learning further coordinates high-order semantics and low-order features. Extensive experiments are conducted on four benchmark data sets to thoroughly assess the model's performance. Additionally, experiments on a multirelational data set further validate the model's adaptability in handling complex heterogeneous graphs. The results indicate that MDGCN exhibits a superior performance. The contributions of each module added to the framework are further confirmed by ablation experiments. Furthermore, case studies demonstrate that this framework possesses significant predictive capabilities,

rendering it an effective instrument for forecasting drug–disease relationships. In future work, we expect to construct dependency graphs based on density to address the issue that the KNN algorithm is less effective for outliers or noisy data.

## ASSOCIATED CONTENT

### Data Availability Statement

The code and data sets can be accessed at <https://github.com/DHUBlab/MDGCN>.

## AUTHOR INFORMATION

### Corresponding Author

Yangan Gan – School of Computer Science and Technology, Donghua University, Shanghai 201620, China; [orcid.org/0000-0001-5931-9006](https://orcid.org/0000-0001-5931-9006); Email: [ylgan@dhu.edu.cn](mailto:ylgan@dhu.edu.cn)

### Authors

Shengnan Li – School of Computer Science and Technology, Donghua University, Shanghai 201620, China; [orcid.org/0009-0004-4290-1808](https://orcid.org/0009-0004-4290-1808)

Guangwei Xu – School of Computer Science and Technology, Donghua University, Shanghai 201620, China

Cairong Yan – School of Computer Science and Technology, Donghua University, Shanghai 201620, China

Guobing Zou – School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.4c02424>

### Author Contributions

Y.G. conceptualized and structured the study, supervised the work, and edited the manuscript. S.L. developed and executed the method, conducted the experiments, analyzed the results, and wrote the manuscript. G.X. and C.Y. conceptualized the study and revised the manuscript. G.Z. managed the project and revised the manuscript. All authors read and approved the final manuscript.

### Funding

This work was sponsored in part by the National Natural Science Foundation of China (62172088, 62272290) and the Shanghai Natural Science Foundation (21ZR1400400).

### Notes

The authors declare no competing financial interest.

## REFERENCES

- (1) Luo, H.; Li, M.; Yang, M.; Wu, F.-X.; Li, Y.; Wang, J. Biomedical data and computational models for drug repositioning: a comprehensive review. *Briefings Bioinf.* **2021**, *22*, 1604–1619.
- (2) Park, J.-H.; Cho, Y.-R. Computational drug repositioning with attention walking. *Sci. Rep.* **2024**, *14*, No. 10072.
- (3) Katukam, L. P.; Padakanti, A. P.; Chella, N. Drug Repurposing: History, Significance, Benefits, Approaches, and Challenges. In *Drug Repurposing: Innovative Approaches to Drug Discovery and Development*; Springer, 2024; pp 1–11.
- (4) Dabhade, P.; Bharadwaj, A.; Bagde, A.; Shende, P. Computational drug designing and drug discovery approach in drug repositioning: A review. In *AIP Conference Proceedings*; AIP Publishing, 2024.
- (5) Li, Y.; Yang, Y.; Tong, Z.; Wang, Y.; Mi, Q.; Bai, M.; Liang, G.; Li, B.; Shu, K. A comparative benchmarking and evaluation framework for heterogeneous network-based drug repositioning methods. *Briefings Bioinf.* **2024**, *25*, No. bbae172.
- (6) Tanoli, Z.; Vähä-Koskela, M.; Aittokallio, T. Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin. Drug Discovery* **2021**, *16*, 977–989.



- (7) Tiwari, S. W.; Joshi, M.; Chand, G.; Gwasikoti, J.; Mittal, A.; Sharma, P. K. In *Transforming Pharmacological Research: Utilizing AI and Machine Learning in Drug Discovery and Development*, 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS); IEEE, 2024; pp 75–78.
- (8) Napolitano, F.; Zhao, Y.; Moreira, V. M.; Tagliaferri, R.; Kere, J.; D'Amato, M.; Greco, D. Drug repositioning: a machine-learning approach through data integration. *J. Cheminf.* **2013**, *5*, No. 30.
- (9) Ghorbanali, Z.; Zare Mirakabad, F.; Mohammadpour, B. DRP-VEM: Drug Repositioning Using Voting Ensemble Model AUT *J. Math. Comput.* **2024**.
- (10) Zhang, W.; Xu, H.; Li, X.; Gao, Q.; Wang, L. DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion. *Bioinformatics* **2020**, *36*, 2839–2847.
- (11) Yang, M.; Luo, H.; Li, Y.; Wang, J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* **2019**, *35*, i455–i463.
- (12) Liu, T.; Wang, S.; Zhang, Y.; Li, Y.; Liu, Y.; Huang, S. TIWMFLP: Two-Tier Interactive Weighted Matrix Factorization and Label Propagation Based on Similarity Matrix Fusion for Drug-Disease Association Prediction. *J. Chem. Inf. Model.* **2024**, *64*, 8641–8654.
- (13) 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*, 5191–5198.
- (14) Liu, J.; Guan, S.; Zou, Q.; Wu, H.; Tiwari, P.; Ding, Y. AMDGT: Attention aware multi-modal fusion using a dual graph transformer for drug-disease associations prediction. *Knowl.-Based Syst.* **2024**, *284*, No. 111329.
- (15) Wang, Y.; Gao, Y.-L.; Wang, J.; Li, F.; Liu, J.-X. MSGCA: Drug-Disease Associations Prediction Based on Multi-Similarities Graph Convolutional Autoencoder. *IEEE J. Biomed. Health Inf.* **2023**, *27*, 3686–3694.
- (16) Luo, Y.; Deng, L. DPMGCDA: Deciphering circRNA-Drug Sensitivity Associations with Dual Perspective Learning and Path-Masked Graph Autoencoder. *J. Chem. Inf. Model.* **2024**, *64*, 4359–4372, DOI: 10.1021/acs.jcim.4c00573.
- (17) Meng, Y.; Lu, C.; Jin, M.; Xu, J.; Zeng, X.; Yang, J. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings Bioinf.* **2022**, *23*, No. bbab581.
- (18) Cai, L.; Lu, C.; Xu, J.; Meng, Y.; Wang, P.; Fu, X.; Zeng, X.; Su, Y. Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Briefings Bioinf.* **2021**, *22*, No. bbab319.
- (19) Sun, X.; Jia, X.; Lu, Z.; Tang, J.; Li, M. Drug repositioning with adaptive graph convolutional networks. *Bioinformatics* **2024**, *40*, No. btad748.
- (20) Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; Welling, M. In *Modeling Relational Data with Graph Convolutional Networks*, The Semantic Web (ESWC 2018; Springer, 2018; pp 593–607.
- (21) Li, M.; Micheli, A.; Wang, Y. G.; Pan, S.; Lió, P.; Gnecco, G. S.; Sanguineti, M. Guest editorial: deep neural networks for graphs: theory, models, algorithms, and applications. *IEEE Trans. Neural Networks and Learn. Syst.* **2024**, *35*, 4367–4372.
- (22) Li, M.; Zhang, L.; Cui, L.; Bai, L.; Li, Z.; Wu, X. BLoG: Bootstrapped graph representation learning with local and global regularization for recommendation. *Pattern Recognit.* **2023**, *144*, No. 109874.
- (23) Bai, L.; Cui, L.; Wang, Y.; Li, M.; Li, J.; Philip, S. Y.; Hancock, E. R. HAQJSK: Hierarchical-Aligned Quantum Jensen-Shannon Kernels for Graph Classification. *IEEE Trans. Knowledge Data Eng.* **2024**, *36*, 6370–6384, DOI: 10.1109/TKDE.2024.3389966.
- (24) Jia, X.; Sun, X.; Wang, K.; Li, M. DRGCL: Drug Repositioning via Semantic-enriched Graph Contrastive Learning. *IEEE J. Biomed. Health Inf.* **2024**, 1656–1667, DOI: 10.1109/jbhi.2024.3372527.
- (25) Wang, Y.; Meng, Y.; Zhou, C.; Tang, X.; Zeng, P.; Pan, C.; Zhu, Q.; Zhang, B.; Xu, J. Automatic collaborative learning for drug repositioning. *Eng. Appl. Artif. Intell.* **2025**, *139*, No. 109653.
- (26) Li, Q.; Han, Z.; Wu, X.-M. Deeper insights into graph convolutional networks for semi-supervised learning *Proc. the AAAI Conf. Artif. Int.* **2018**; Vol. 32 DOI: 10.1609/aaai.v32i1.11604.
- (27) Glorot, X.; Bengio, Y. In *Understanding the Difficulty of Training Deep Feedforward Neural Networks*, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics PMLR; 2010; pp 249–256.
- (28) Dauphin, Y. N.; Fan, A.; Auli, M.; Grangier, D. In *Language modeling with gated convolutional networks*, Proceedings of the 34th International Conference on Machine Learning; PLMR, 2017; pp 933–941.
- (29) He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. In *LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation*, SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval; ACM, 2020; pp 639–648.
- (30) Hu, Z.; Dong, Y.; Wang, K.; Sun, Y. In *Heterogeneous Graph Transformer*, Proceedings of The Web Conference; ACM, 2020; pp 2704–2710.
- (31) Yu, J.; Xia, X.; Chen, T.; Cui, L.; Hung, N. Q. V.; Yin, H. XSimGCL: Towards extremely simple graph contrastive learning for recommendation. *IEEE Trans. Knowledge Data Eng.* **2023**, *36*, 913–926.
- (32) Yu, J.; Yin, H.; Li, J.; Wang, Q.; Hung, N. Q. V.; Zhang, X. In *Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation*, Proceedings of the Web Conference; ACM, 2021; pp 413–424.
- (33) Tian, Y.; Sun, C.; Poole, B.; Krishnan, D.; Schmid, C.; Isola, P. In *What Makes for Good Views for Contrastive Learning?*, Advances in Neural Information Processing Systems 33 (NeurIPS 2020); NeurIPS, 2020; pp 6827–6839.
- (34) Lin, Z.; Tian, C.; Hou, Y.; Zhao, W. X. In *Improving Graph Collaborative Filtering with Neighborhood-Enriched Contrastive Learning*, Proceedings of the ACM Web Conference; ACM, 2022; pp 2320–2329.
- (35) Gottlieb, A.; Stein, G. Y.; Ruppini, E.; Sharan, R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*, No. 496.
- (36) Luo, H.; Wang, J.; Li, M.; Luo, J.; Peng, X.; Wu, F.-X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics* **2016**, *32*, 2664–2671.
- (37) Liang, X.; Zhang, P.; Yan, L.; Fu, Y.; Peng, F.; Qu, L.; Shao, M.; Chen, Y.; Chen, Z. LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning. *Bioinformatics* **2017**, *33*, 1187–1196.
- (38) Yu, Z.; Huang, F.; Zhao, X.; Xiao, W.; Zhang, W. Predicting drug-disease associations through layer attention graph convolutional network. *Briefings Bioinf.* **2021**, *22*, No. bbab243.
- (39) Hamosh, A.; Scott, A. F.; Amberger, J. S.; Bocchini, C. A.; McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **2004**, *33*, D514–D517.
- (40) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.
- (41) Davis, A. P.; Grondin, C. J.; Johnson, R. J.; Sciaky, D.; Wiegiers, J.; Wiegiers, T. C.; Mattingly, C. J. Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138–D1143.
- (42) Yella, J. K.; Jegga, A. G. MGATRx: discovering drug repositioning candidates using multi-view graph attention. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2022**, *19*, 2596–2604.
- (43) Davis, J.; Goadrich, M. In *The relationship between Precision-Recall and ROC Curves*, Proceedings of the 23rd International Conference on Machine Learning; ACM, 2006; pp 233–240.
- (44) Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* **2015**, *10*, No. e0118432.

- (45) Chen, H.; Cheng, F.; Li, J. iDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput. Biol.* **2020**, *16*, No. e1008040.
- (46) Xia, L.; Huang, C.; Xu, Y.; Zhao, J.; Yin, D.; Huang, J. et al. In *Hypergraph Contrastive Collaborative Filtering*, Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval; ACM, 2022; pp 70–79.
- (47) Mao, C.; Yao, L.; Luo, Y. MedGCN: Graph convolutional networks for multiple medical tasks **2019**. arXiv:1904.00326. arXiv.org e-Print archive. <https://arxiv.org/abs/1904.00326>.
- (48) Wang, Z.; Yu, D.; Li, Q.; Shen, S.; Yao, S. SR-HGN: Semantic- and relation-aware heterogeneous graph neural network. *Expert Syst. Appl.* **2023**, *224*, No. 119982.
- (49) Zhu, G.; Zhu, Z.; Chen, H.; Yuan, C.; Huang, Y. HAGNN: Hybrid aggregation for heterogeneous graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.* **2024**, 1–5, DOI: [10.1109/TNNLS.2024.3519427](https://doi.org/10.1109/TNNLS.2024.3519427).
- (50) Wirdefeldt, K.; Odin, P.; Nyholm, D. Levodopa-carbidopa intestinal gel in patients with Parkinson's disease: a systematic review. *CNS Drugs* **2016**, *30*, 381–404.
- (51) Galli, S.; De Carolis, L.; Bianchini, E.; Alborghetti, M.; Calì, B.; Pacilio, P.; Fanciulli, A.; Pontieri, F. E.; Rinaldi, D. Effects of levodopa/carbidopa intestinal gel infusion on autonomic symptoms in advanced Parkinson's disease: a systematic review *Clin. Auton. Res.* **2024** DOI: [10.1007/s10286-024-01090-9](https://doi.org/10.1007/s10286-024-01090-9).
- (52) Pérez, L. M.; Farriols, C.; Puente, V.; Planas, J.; Ruiz, I. The use of subcutaneous scopolamine as a palliative treatment in Parkinson's disease. *Palliative Med.* **2011**, *25*, 92–93.
- (53) Rogawski, M. A.; Wenk, G. L. The neuropharmacological basis for the use of memantine in the treatment of Alzheimer's disease. *CNS Drug Rev.* **2003**, *9*, 275–308.
- (54) Dooley, M.; Lamb, H. M. Donepezil: a review of its use in Alzheimer's disease. *Drugs Aging* **2000**, *16*, 199–226.
- (55) Savva, K.; Zachariou, M.; Kynigopoulos, D.; Fella, E.; Vitali, M.-I.; Kosofidou, X.; Spyrou, M.; Sargiannidou, I.; Panayiotou, E.; Dietis, N.; Spyrou, G. M. Preliminary In Vitro and In Vivo Insights of In Silico Candidate Repurposed Drugs for Alzheimer's Disease. *Life* **2023**, *13*, No. 1095.
- (56) Reed, A. M.; Duff, K.; Dibble, L. E.; Paul, S. S.; Hooyman, A.; Schaefer, S. Y. Examining the Diagnostic Accuracy of a Novel Performance-Based Test for Alzheimer's Disease Screening. *J. Prev. Alzheimer's Dis.* **2024**, *11*, 903–907.