

RESEARCH

Open Access



Fusing graph transformer with multi-aggregate GCN for enhanced drug–disease associations prediction

Shihui He^{1,2}, Lijun Yun^{1,2*} and Haicheng Yi^{3*}

*Correspondence:
yunlijun@ynnu.edu.cn;
haichengyi@gmail.com

¹ School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China

² Engineering Research Center of Computer Vision and Intelligent Control Technology, Department of Education, Kunming 650500, China

³ School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China

Abstract

Background: Identification of potential drug–disease associations is important for both the discovery of new indications for drugs and for the reduction of unknown adverse drug reactions. Exploring the potential links between drugs and diseases is crucial for advancing biomedical research and improving healthcare. While advanced computational techniques play a vital role in revealing the connections between drugs and diseases, current research still faces challenges in the process of mining potential relationships between drugs and diseases using heterogeneous network data.

Results: In this study, we propose a learning framework for fusing Graph Transformer Networks and multi-aggregate graph convolutional network to learn efficient heterogeneous information graph representations for drug–disease association prediction, termed WMAGT. This method extensively harnesses the capabilities of a robust graph transformer, effectively modeling the local and global interactions of nodes by integrating a graph convolutional network and a graph transformer with self-attention mechanisms in its encoder. We first integrate drug–drug, drug–disease, and disease–disease networks to construct heterogeneous information graph. Multi-aggregate graph convolutional network and graph transformer are then used in conjunction with neural collaborative filtering module to integrate information from different domains into highly effective feature representation.

Conclusions: Rigorous cross-validation, ablation studies examined the robustness and effectiveness of the proposed method. Experimental results demonstrate that WMAGT outperforms other state-of-the-art methods in accurate drug–disease association prediction, which is beneficial for drug repositioning and drug safety research.

Keywords: Drug repositioning, Drug–disease associations, Graph transformer, Graph neural networks, Neural collaborative filtering

Background

Identification and characterization of potential interactions between drugs and diseases are crucial challenges for drug discovery and disease treatment. Conventional methods for validating drug–disease associations rely on costly and time-consuming experimental



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

procedures. The average cost of bringing a new drug to market exceeds about 2 billion dollars and 10–15 years before it can reach the pharmacy shelf [1–3]. Therefore, finding new indications for existing drugs, also known as drug repositioning, is an economically viable and time-saving strategy [4, 5]. Computational methods for drug repositioning facilitate the identification of potential drug–disease associations by screening large-scale data sources, enabling more rational design of clinical trials. Such strategies can accelerate the drug discovery pipeline and increase the availability of new treatments [6].

The utilization of computational methods for drug repositioning in drug discovery has become widespread [7, 8]. During this time, a growing array of methodologies has emerged. For example, methods rooted in matrix decomposition, like the one by Cui et al. [9] utilize dual-network $L_{2,1}$ -collaborative matrix factorization for predicting novel drug–disease interactions. Fu et al. [10] introduced MFLDA, a method that decomposes heterogeneous data sources' matrices into low-rank forms using matrix tri-factorization, thus exploring and exploiting their inherent and shared structure. MFLDA facilitates the selection and integration of these data sources by assigning varying weights to each source. Matrix factorization diminishes data dimensionality by transforming matrices into low-rank structures, extracting crucial features and patterns. However, with extensive or densely high-dimensional matrices, its computational demands might become excessive, resulting in reduced efficacy in managing considerable noise.

Network-based drug repositioning models have emerged to confront this challenge, striving to capitalize on intricate relational networks among biological entities such as drugs and diseases. These models amalgamate varied information sources, encompassing protein–protein interaction networks, gene expression data, and drug compound information, to anticipate potential novel drug–disease associations. For instance, Zhang et al. [11] proposed NTSIM to predict unobserved drug–disease associations and extended it to NTSIM-C for classifying therapeutic associations. Zhang et al. [12] proposed SCMFDD, projecting drug–disease associations into two low-rank spaces, revealing latent features, and introducing feature-based similarity and semantic constraints. Lu et al. [13] proposed heterogeneous information network (HIN) based model, namely HINGRL. Zhou et al. [14] introduced NEDD, using varied-length metapaths to explicitly capture internal relationships within drugs and diseases and obtain low-dimensional representation vectors. Martínez et al. [15] developed DrugNet, a network-based method predicting new drug uses and treatments for diseases. It utilizes a heterogeneous network formed from disease, drug, and target information, identifying novel associations by information propagation.

Despite the commendable interpretability inherent in network-based methodologies, their performance is deemed unsatisfactory. However, the drug–disease association network naturally has a graph structure, which enables techniques that leverage graph neural networks to adeptly preserve essential information, eliminate noise, and extract pivotal patterns and features. Therefore, this enhances the accuracy of information available for prediction and analysis in graph-based data scenarios. Some methods have been proposed to exploit the advantages of graph neural networks for drug–disease association prediction. Yu et al. [16] proposed LAGCN, a method that integrates heterogeneous networks, employs graph convolutional operations, and incorporates an attention mechanism. Yang et al. [17] introduced a model that infers drug–disease associations

by applying network-embedding algorithms alongside a random forest classification approach. Gu et al. [18] introduced REDDA, a heterogeneous graph neural network with three attention mechanisms for sequential drug-disease representation learning. Wu et al. [19] proposed EMP-SVD, a novel framework that predicts drug-disease associations by integrating multiple meta-paths and singular value decomposition. Li et al. [20] proposed the NIMCGCN method, which integrates Graph Convolutional Networks (GCN) with Neural Inductive Matrix Completion (NIMC) models to discover associations between miRNA and diseases. VGAE [21], introduced by Kipf et al., is a graph neural network model based on the Variational Autoencoder (VAE) framework, which models node features through an encoder-decoder structure, mapping them to a latent space distribution. Meng et al. [22] leveraged deep learning within a heterogeneous network framework to identify potential drugs related to diseases. Their model, DRWBNCF, employs weighted bilinear graph convolution operations, intricately fusing information about drug-disease associations and drug-disease similarity networks. Tang et al. [23] proposed the DRGBCN model, which exploits the embedding of graph convolutional layers and local interactions between drugs and diseases, thus significantly improving the accuracy and reliability of predictions. Ghasemian et al. [24] applied meta-learning methods in network analysis to develop a stacked model that integrates complex prediction algorithms from various domains, effectively mitigating changes in link prediction. Additionally, many studies [25, 26] have indicated that collaborative drug combination prediction is widely applied in drug repositioning. For instance, the SNRMPACDC model proposed by Li et al. [27] combines Siamese convolutional networks and random matrix projection to predict collaborative combinations of anticancer drugs. NLLSS [28] is a semi-supervised learning-based model that focuses on predicting collaborative drug combinations, enhancing the model's predictive performance through methods involving non-negative low-rank and sparse structures. They play a crucial role in revealing the associations between drugs and biomolecules, as well as in drug repositioning.

Although these methods can extract edge information through extensive informatics learning, they often struggle to fully mine the complex interactions between nodes in heterogeneous graphs, which may affect the accuracy of predictions. To address these challenges, we propose a heterogeneous information graph representation learning method for predicting drug-disease associations, named WMAGT, which utilizes weighted multi-aggregate graph convolutional network and graph transformer to exploit discriminative node representations. The workflow of the proposed method is demonstrated in Fig. 1. Specifically, the proposed WMAGT approach first integrates drug-drug similarity networks, disease-disease similarity networks, and validated drug-disease association networks to construct a comprehensive heterogeneous information network. Then, graph transformer combined with weighted multi-aggregate graph convolutional neural network are used to learn efficient characterizations of drugs and diseases from this heterogeneous information network. Prior to the predictor, we integrated domain embeddings and interaction embeddings through neural collaborative filtering for the final link prediction scoring. To evaluate the performance of the proposed method, we cross-validated the predictive performance of WMAGT on three benchmark datasets and compared it with four state-of-the-art methods while conducting ablation experiments. Experimental results

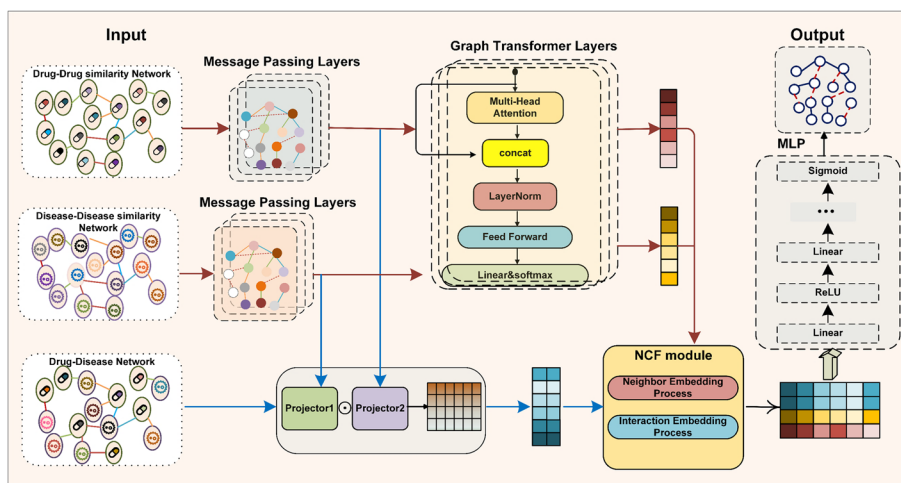


Fig. 1 The overall architecture of the proposed WMAGT. WMAGT involves three main steps. First, drug and disease similarity networks are jointly encoded using GCN and graph transformer for representation projection. In the second step, matrix operations project drug and disease representations in the network, generating new information. Lastly, the domain information from the first step and interactive information from the second step are utilized in the NCF module, and multiple loss functions along with MLP are employed to comprehensively model the drug–disease relationship

provide strong evidence of the effectiveness of WMAGT in discovering drug indications, which is important for advancing drug repurposing and reducing adverse drug reactions in the field of drug discovery. The main contributions and advantages of WMAGT include:

1. Proposing a representation learning method based on heterogeneous information graphs, which fully utilizes the multi-source information of drugs and diseases and considers their multi-level relationships.
2. Adopting a representation learning framework of Graph Transformer and Weighted Multi-Aggregation Graph Convolutional Neural Network, effectively eliminating the impact of heterogeneity and capturing relationships to learn more effective node representations.
3. Demonstrating through experiments on three public datasets that this method has broad application prospects in the field of drug discovery. Our approach not only outperforms existing models in predictive performance but also shows significant improvement in understanding complex biological networks.

Methods

In this study, we propose a novel computational method, WMAGT, for drug repositioning, aiming to discover new indications for existing drugs by inferring potential drug–disease associations. First, we build a heterogeneous network that incorporates various types of relations in the data set, such as drug–drug similarity network, disease–disease similarity network, and drug–disease association network. Then, we utilize an end-to-end model to learn the latent features of the network and predict the unknown associations.

Benchmark datasets

To explore heterogeneous network prediction methods for drug–disease associations, we utilized three publicly available real datasets to assess the efficacy of our model. The first dataset, Fdataset [29], comprises 313 diseases from the OMIM database [30] and 553 drugs from the DrugBank database [31], along with 1933 known associations between them. Another dataset, termed as Cdataset [32], consists of 663 drugs from the DrugBank database and 409 diseases from the OMIM database, encompassing 2532 established associations between drugs and diseases. The third dataset is LRSSL [33], which comprises 763 drugs from the DrugBank database, 681 diseases from the MeSH database, and a collection of 3051 validated associations between drugs and diseases. The essential statistical information of these three datasets is presented in Table 1.

The construction of heterogeneous information graph

To predict potential drug–disease associations, this research employed network analysis methods based on a known drug–disease association network denoted as G . G is represented by an $n \times m$ binary matrix A , where n and m represent the number of drugs and diseases, respectively. The matrix A_{ij} holds a value of 1 or 0, indicating the presence or absence of an experimentally validated association between drug r_i and disease d_j .

Two additional similarity networks were constructed: a drug–drug similarity network G_r and a disease–disease similarity network G_d . These networks are represented by $n \times n$ and $m \times m$ matrices A_r and A_d , respectively. The values $A_r(i, j)$ and $A_d(i, j)$ represent the similarities between drug r_i and drug r_j , and between disease d_i and disease d_j , respectively. These similarities were computed based on various characteristics including chemical, pharmacological, therapeutic, phenotypic, genetic, and environmental properties of drugs or diseases.

To enhance accuracy and reduce noise, a k -nearest neighbor approach was employed. It considered only the k most similar neighbors for each drug or disease. The extended k -nearest neighbor sets of drugs or diseases, represented as \tilde{N}_k , comprised the individual entities along with their k nearest neighbors. $A_r(i, j)$ and $A_d(i, j)$ illustrate the similarities among drugs or diseases, considering their extended k -nearest neighbor sets \tilde{N}_k . Mathematical representation:

$$G = (A_{ij})_{n \times m} \quad (1)$$

$$G_r = (A_r(i, j))_{n \times n} \quad (2)$$

$$G_d = (A_d(i, j))_{m \times m} \quad (3)$$

Table 1 Details of the three benchmark datasets

Datasets	Drugs	Diseases	Associations
Fdataset	593	313	1933
Cdataset	663	409	2532
LRSSL	763	681	3051

Considering two sets representing drugs (R) and diseases (D), where each $r \in R$ and $d \in D$ introduces an association label $Y_{r,d}$ signifying the presence $Y_{r,d}=1$ or absence $Y_{r,d}=0$ of an association between drug r and disease d . Consequently, inferring the association label $Y_{r,d}$ for a given drug r and disease d relies on known associations within the sets. The expression for the association label $Y_{r,d}$ remains defined as:

$$Y_{r,d} = \begin{cases} 1 & \text{if } r \text{ is associated with } d \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

This representation aims to establish the foundation of the drug repositioning problem, framing it as a task of predicting association labels.

Graph convolutional network module

In a drug–disease heterogenous graph, nodes represent various drugs and diseases. Typically, each node contains its own similarity information, and the edges connecting two nodes represent the relationship between them. We employ Graph Convolutional Networks (GCN) [34–37] to integrate node information, which usually consists of aggregation functions and update functions. Aggregation functions are applied to each node/edge to gather information from their neighbors, while update functions generate new representations for each node/edge based on the collected information and the previous representation. The update function is defined as follows:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (5)$$

Here, $H^{(l)}$ represents the input features at layer l in the GCN. $H^{(l+1)}$ signifies the output features at layer $l+1$ after the convolution operation. σ is the activation function (commonly ReLU or Leaky ReLU). $W^{(l)}$ is the learnable weight matrix at layer l . $\tilde{A} = A + I_n$ represents the adjacency matrix of the graph, where A is the original adjacency matrix, and I_n is the identity matrix. \tilde{D} is the diagonal node degree matrix of \tilde{A} .

Node attentions in graph transformer module

Recently, the Transformer model has extended its application beyond the field of natural language processing to include a wide range of tasks, including link prediction. In the information integration module, we have incorporated both the graph transformer [38–41] and GCN, thereby enhancing the model's flexibility and performance. The two fundamental components of the Transformer are the dot-product attention mechanism and the feedforward network, playing crucial roles in link prediction tasks.

The graph attention formula is:

$$h_i^{(l+1)} = \sum_{j=1}^n \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)} \quad (6)$$

where $h_j^{(l)}$ is the i -th node's feature vector in layer l , $W^{(l)}$ is the layer's weight matrix, n is the graph size, and $\alpha_{ij}^{(l)}$ is the attention weight between nodes i and j in layer l , computed by

$$\alpha_{ij}^{(l)} = \frac{\exp(e_{ij}^{(l)})}{\sum_{k=1}^n \exp(e_{ik}^{(l)})} \quad (7)$$

where $e_{ij}^{(l)}$ is the similarity between nodes i and j in layer l , computed by:

$$e_{ij}^{(l)} = a^{(l)}(W^{(l)}h_i^{(l)}, W^{(l)}h_j^{(l)}) \quad (8)$$

where $a^{(l)}$ is a differentiable similarity function in layer l , such as dot product, bilinear, multilayer perceptron, etc.

Layer-wise transformation At each layer of the Graph Transformer, the hidden states of nodes are updated using multi-head self-attention and feedforward neural networks. The transformation can be summarized as [42]:

$$h_i^{(l)} = \text{MultiHeadAttention}(h_i^{(l-1)}) + \text{FeedForward}(h_i^{(l-1)}) \quad (9)$$

Aggregation across heads The outputs from multiple attention heads are aggregated to obtain the final node representations:

$$h_i^{(l)} = \text{concat}(h_i^{(l,1)}, h_i^{(l,2)}, \dots, h_i^{(l,H)}) \quad (10)$$

where H represents the number of attention heads.

Overview of the proposed WMAGT model

As shown in Fig. 1, this section will delve into the detailed description of the model, delineating its architecture, methodologies employed, and the intricate components contributing to its predictive capability.

Graph representation learning with mixed aggregation parameters

In the context of learning node neighborhood information, a hybrid approach is employed utilizing mixed parameters, integrating two distinct graph convolution operations to acquire meaningful representations of graph data. The fundamental idea of hybrid parameters involves a weighted combination of the outputs of two graph convolution operations, thereby generating the final node features.

$$\text{out} = \beta \cdot \text{ReLU}(\text{Pool}(XW, A)) + \alpha \cdot \text{ReLU}\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} XW\right) \quad (11)$$

here, α and β control the relative influence of the two aggregation methods. The Rectified Linear Unit function (ReLU) [43] is employed as activation function, while *Pool* represents a customized pooling operation involving specific manipulations of the adjacency matrix. This process involves the product of the adjacency matrix and node feature matrix, square operations, and some matrix operations. The entire operation can be expressed mathematically as follows: $Z = (A \cdot XW)^2 - (A^2 \odot XW^2)$. Here, A is the adjacency matrix of the graph, XW is the node feature matrix, and Z represents the new node representation matrix obtained after the graph pooling operation.

This operation introduces additional information into the graph structure, aiming to better capture the relationships between nodes. In general, the introduction of hybrid parameters imparts adaptability to the model, allowing it to determine the relative contributions of different graph convolution operations during the learning process and thus better adapt to diverse graph structures.

Computing weighted matrices for drug and disease nodes

To compute the weighted matrices for drugs and diseases, we utilize the input feature matrix $X \in \mathbb{R}^{N \times d}$, where N represents the number of nodes and d denotes the feature dimension. The weight matrix, denoted as $W \in \mathbb{R}^{d \times d'}$, corresponds to the output feature dimension d' . The graph's adjacency matrix, $A \in \{0, 1\}^{N \times N}$, demonstrates connections between nodes in a symmetric matrix form. The weighted feature matrix is obtained from this process, which can be represented as:

$$XW = X \cdot W \quad (12)$$

the iterative graph convolution concludes with the normalization of the resulting feature matrix using a normalization matrix represented as:

$$out = norm \cdot out \quad (13)$$

Furthermore, an element-wise addition of a bias term is performed to further refine the output matrix.

$$out = out + self.bias \quad (14)$$

Compute the element-wise product of drug and disease embeddings

Given an input drug embedding matrix as $D \in \mathbb{R}^{N_{drug} \times d}$, and a Disease Embedding matrix as $E \in \mathbb{R}^{N_{disease} \times d}$, where N_{drug} and $N_{disease}$ represent the quantities of drugs and diseases respectively, and d represents the embedding dimension.

Projection of Drug and Disease via Linear Mapping:

$$y = x_1 \cdot P_1 \quad (15)$$

$$y = x_2 \cdot P_2 \quad (16)$$

this involves computing the element-wise product of drug and disease embeddings. For instance, let D_{ij} represent the row and j th column element of matrix D , and E_{ij} represent the i th row and j th column element of matrix E . The element-wise product P can be obtained as: $P_{ij} = D_{ij} \times E_{ij}$. The resulting matrix P captures the element-wise products of the drug and disease embeddings. This process facilitates the exploration of interactions between drugs and diseases within a feature space defined by their embeddings.

To normalize the association matrix P , L_2 normalization is applied row-wise post element-wise product computation. Each row's L_2 norm is computed, and its elements are divided by this norm, ensuring unit L_2 norm per row. The normalization formula is: $P_{normij} = \frac{P_{ij}}{\sqrt{\sum_{k=1}^N P_{ik}^2}}$, Where N denotes the column count, and the summation extends

over the row's elements. This yields a matrix P_{normij} with standardized rows, enhancing association representation and minimizing dataset bias.

Neural collaborative filtering for drug and disease expression

Neural Collaborative Filtering (NCF) [44] is a neural network-based collaborative filtering algorithm designed for learning relationships between users and items for recommendation purposes. The implementation of NCF in our model involves key components: Neighbor Embedding Process, defines the neighbor embedding process, and integrating information from neighbors of drugs and diseases to better capture relationships between nodes. Interaction Embedding Process, defines the interaction information between drugs and diseases. This section mainly involves calculating interaction embedding through element-wise multiplication and normalization operations. Decoding Process, defines the decoder, transforming embedded node representations into final prediction scores. This process primarily involves linear transformations and non-linear activation functions.

In summary, in the forward method of the model, node embedding representations are first obtained through processes such as neighbor embedding and interaction embedding. Subsequently, the decoder yields the final prediction scores. The core idea of Neural Collaborative Filtering involves learning implicit relationships between drugs and diseases through processes such as embedding, neighbor embedding, interaction embedding, and decoding.

Neighbor-weighted interaction decoding

In this module, the descriptions of drug–disease associations, drug proximity, and disease proximity are amalgamated into a unified vector $\tilde{h}_{r,d}$ using the concatenation operation \oplus , defined as:

$$\tilde{h}_{r,d} = h_{r,d} \oplus h_{r,d} \oplus \tilde{h}_r \oplus \tilde{h}_d \quad (17)$$

Here, the operator \oplus signifies concatenation, facilitating the formation of an encompassing representation that merges established associations with contextual information drawn from drug and disease proximities.

Subsequently, linear transformations and ReLU activation were utilized in processing the hidden layers. In each hidden layer i where i ranges from 1 to the length of *hidden_dims*, the use of linear transformations and ReLU activation generated z_i . Afterwards, at the output layer, linear transformations and Sigmoid activation were applied to handle the outputs from the hidden layer $z_{\text{len}(\text{hidden_dims})}$, producing the output y . The overall model output Y can be interpreted as probabilities for specific categories.

MLP-based prediction

The introduction of Multilayer Perceptron (MLP) is motivated by its ability to capture intricate nonlinear relationships, extract advanced features, manage sparse data, and exhibit a flexible architecture adaptable to various data traits. Within drug–disease association studies, integrating MLP aims to enhance the accurate prediction and interpretation of complex drug–disease associations, thereby providing deeper insights into correlation studies within the pharmaceutical domain.

Forward propagation in an MLP involves multiple layers, each with numerous neurons. Assuming inputs X , H neurons in the hidden layer, output Y , weight parameters W , and biases b , the forward propagation can be represented as:

$$Y = \sigma(W_{\text{output}} \cdot \sigma(W_{\text{hidden}} \cdot X + b_{\text{hidden}}) + b_{\text{output}}) \quad (18)$$

Parameters setting

Hyperparameter settings are crucial for fine-tuning the neural collaborative filtering model, covering dimensions like node embedding, neighbor embedding, and decoder hidden layers. Specifically, the node embedding dimension is set at 64, the neighbor embedding dimension at 32, and the decoder hidden layer dimension is specified as (64, 32). The learning rate is set to $5e-4$, and the dropout rate is 0.3. Additionally, a comprehensive set of loss functions is utilized, encompassing binary cross-entropy loss, focal loss, mean squared error loss, and ranking loss. For the focal loss, parameters are configured with α set to 0.5 and γ set to 2.0. The graph transformer network parameter is defined as $\lambda=0.8$. Throughout the training process, a holistic consideration of these loss functions is conducted, aiming to comprehensively optimize the model. These configurations are designed to strike a balance between model complexity and performance, ensuring optimal predictive outcomes across diverse facets.

Loss Function Formula:

$$\text{Focal Loss} = -\alpha \cdot (1 - p)^\gamma \cdot \log(p) \quad (19)$$

Here, α controls the balance of weights between positive and negative samples, and γ regulates the focus of the focal loss. We use the Adam [45] optimizer to update model parameters, ensuring efficient training. A cyclic learning rate scheduler dynamically adjusts the learning rate, enhancing training effectiveness. Additionally, the model incorporates two graph neural network layers (Graph Transformer and Graph Convolution Network), employing different neighbor sampling quantities during training.

Evaluation metrics

We adopted six widely used indicators to measure the predictive performance of the proposed model, including accuracy (Acc), Area Under the Precision-Recall Curve (AUPR), Area Under the Receiver Operating Characteristic Curve (AUC), F1 score, Precision and Recall. Since AUPR and F1 are more sensitive to severe imbalances data. Micro metrics are used for AUPR and AUC, while macro metrics are used for other measurements. The definitions of these indicators can be described as follows:

$$\text{Acc} = \frac{TP + TN}{TN + TP + FN + FP} \quad (20)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (21)$$

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (23)$$

where the TN , PN , FN and FP denote the number of correctly predicted positive and negative samples, wrongly predicted positive and negative samples, respectively. In addition, we use the *Micro* mode to calculate AUC and Recall, which treats each element of the label indicator matrix as a label. In contrast, F1 calculates each label in a *Macro* mode and finds their unweighted average.

Baseline methods

NIMCGCN [20]. This study introduces a novel approach named Neural Inductive Matrix Completion with Graph Convolutional Network (NIMCGCN), amalgamating Graph Convolutional Networks (GCNs) and Neural Inductive Matrix Completion (NIMC) models to forecast the association between miRNAs and diseases. By optimizing parameters through supervised learning and demonstrating its superiority in prediction accuracy and forecasting new diseases during experimental validation, the method serves as an effective computational tool for swiftly identifying disease-associated miRNAs.

DRWBNCf [22]. This study introduces a new method called DRWBNCf for drug repositioning, addressing limitations of traditional latent factor models. Leveraging deep learning techniques and a heterogeneous network framework, DRWBNCf infers potential drugs for diseases. By amalgamating drug–disease association information and drug–disease similarity networks, employing a weighted bilinear graph convolution operation, and utilizing a multi-layer perceptron combined with α -balanced focal loss function and graph regularization, DRWBNCf demonstrates effectiveness in predicting unknown drug–disease associations.

Ghasemian 's model [24]. Ghasemian et al. employed a meta-learning approach within network analysis to devise a stacked model, amalgamating various sophisticated prediction algorithms. This approach successfully mitigated the variations observed in link prediction across diverse domains of networks.

VAGE [21]. Variational Graph Auto-Encoders (VGAE) is a graph neural network model built upon the framework of Variational Autoencoders (VAE). VGAE integrates the encoder-decoder structure of VAE, modeling node features into latent space distributions and reconstructing them back to the original feature space. Key features include probabilistic modeling, representing node embeddings as Gaussian distributions using reparameterization techniques and KL divergence, while also considering graph structure through graph convolutional networks (GCNs) to efficiently capture local structural information.

DRGBCN [23]. DRGBCN presents an approach that utilizes bilinear attention networks and local interactive learning to improve performance in drug repositioning tasks. Significant performance gains are achieved by emphasizing local association and deep learning applications in the medical domain.

Table 2 Performance of WMAGT and other compared methods on three benchmark datasets

Datasets	NIMCGCN	DRWBNCF	Ghasemian's model	VAGE	DRGBCN	WMAGT
<i>AUROC</i>						
Fdataset	0.7428 ± 0.0276	0.8781 ± 0.0192	0.8902 ± 0.0328	0.9163 ± 0.1052	0.9326 ± 0.013	0.9353 ± 0.012
Cdataset	0.7928 ± 0.0248	0.8928 ± 0.0154	0.9114 ± 0.0292	0.9551 ± 0.0842	0.9454 ± 0.0091	0.9458 ± 0.0114
LRSSL	0.8661 ± 0.0165	0.8297 ± 0.0161	0.8791 ± 0.0359	0.8856 ± 0.0536	0.9437 ± 0.005	0.9434 ± 0.0083
Average	0.8006	0.8669	0.8936	0.9189	0.9405	0.9415
<i>AUPR</i>						
Fdataset	0.0558 ± 0.0106	0.4638 ± 0.0548	0.4046 ± 0.0683	0.0589 ± 0.0429	0.4087 ± 0.0281	0.5231 ± 0.0487
Cdataset	0.0751 ± 0.0138	0.5801 ± 0.0332	0.4881 ± 0.1047	0.0608 ± 0.0355	0.4517 ± 0.0423	0.6 ± 0.0429
LRSSL	0.1807 ± 0.0204	0.4033 ± 0.0201	0.4925 ± 0.1166	0.0381 ± 0.0144	0.2558 ± 0.033	0.3651 ± 0.026
Average	0.1039	0.4824	0.4617	0.0526	0.3721	0.4961

The bold indicates the best performing method on each metric

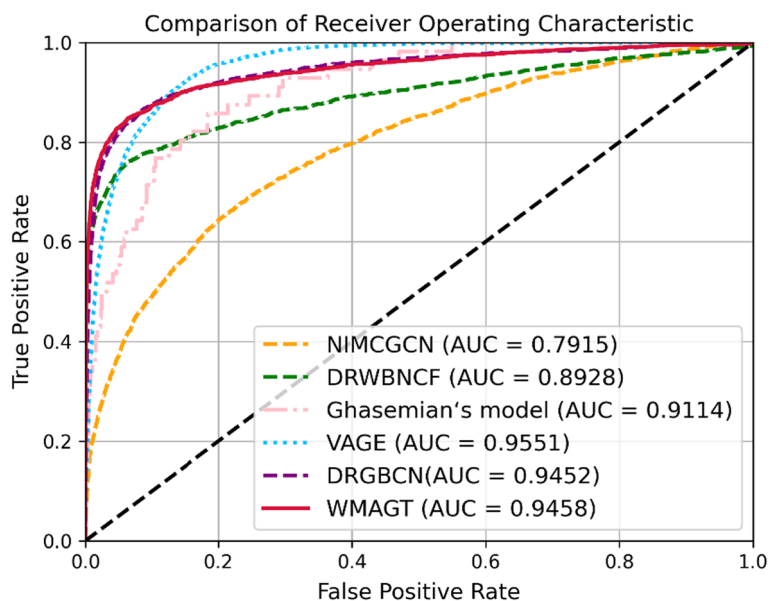


Fig. 2 The performance of WMAGT and other compared methods under tenfold cross-validation on Cdataset

Results and discussion

Comparison of WMAGT and state-of-the-art methods under tenfold cross-validation

To evaluate the performance of the WMAGT model, we conducted extensive experiments on three benchmark datasets, comparing WMAGT with five state-of-the-art methods under tenfold cross-validation. Table 2, Figs. 2, 3 and 4 present the performance of comparison models, including NIMCGCN, DRWBNCF, Ghasemian's model, VAGE, WMAGT and DRGBCN, across different datasets (Fdataset, Cdataset, LRSSL).

Notably, while VAGE achieved a slightly higher AUROC of 0.9551 on the Cdataset compared to our proposed model's 0.9458, and DRGBCN attained an AUROC of 0.9437 compared to our proposed model's 0.9434 on the LRSSL dataset, WMAGT model consistently demonstrated the highest average AUROC value 0.9415 across all datasets. Despite DRGBCN exhibiting a secondary performance in AUROC 0.9405, WMAGT

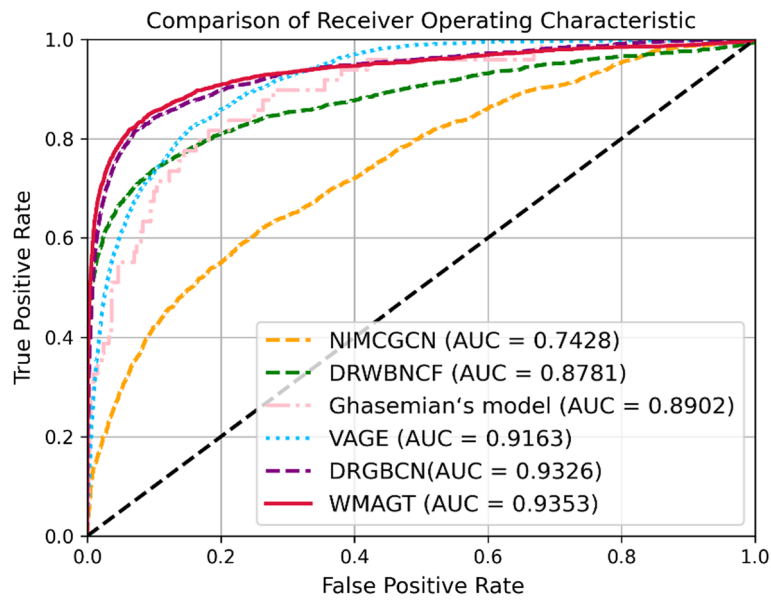


Fig. 3 The performance of WMAGT and other compared methods under tenfold cross-validation on Fdataset

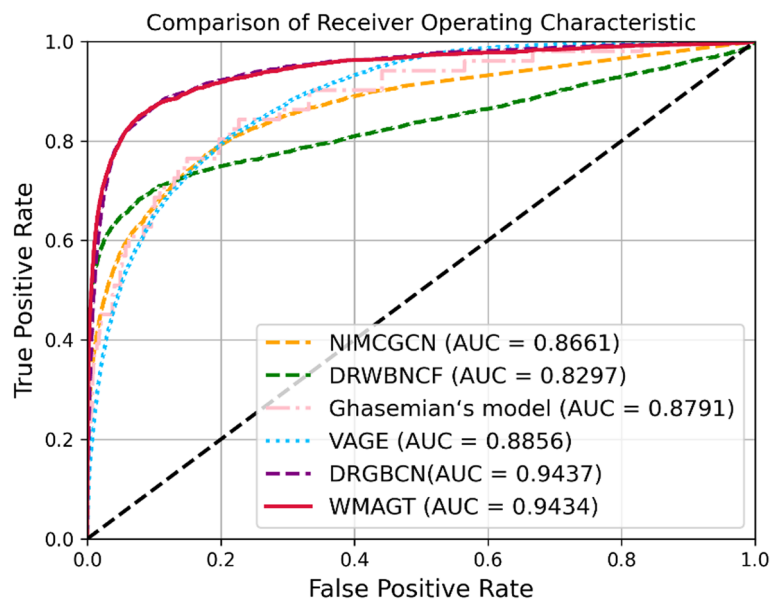


Fig. 4 The performance of WMAGT and other compared methods under tenfold cross-validation on LRSSL

surpasses DRGBCN by more than 10% in terms of AUPR on each dataset. VAGE, Ghasemian's model, DRWBNCF, and NIMCGCN models secured the third, fourth, fifth and sixth positions with AUROC values of 0.9189, 0.8936, 0.8669, 0.8006, respectively. AUPR, particularly sensitive to imbalanced datasets of positive and negative samples, serves as an indispensable evaluation metric. The WMAGT model excelled in AUPR performance, boasting the highest average AUPR value 0.4961, indicating its robust performance under the precision-recall curve. In contrast, the AUPR performances of the other five models were as follows: NIMCGCN 0.1039, DRWBNCF 0.4824, Ghasemian's

model 0.4617, VAGE 0.0526 and DRGBCN 0.3721. Average performance is a crucial indicator for assessing the overall effectiveness of models. In this regard, WMAGT demonstrated relatively superior average performance in both AUROC and AUPR, highlighting its effectiveness across various datasets. After statistical testing and analysis, the proposed WMAGT shows significant performance improvement compared to compared models.

Ablation study

In this section, we delve deeply into the far-reaching impacts of two pivotal modules on our experimental framework:

- 'w/o Transformer': Our investigation goes beyond, scrutinizing the specific effects of excluding the transformer mechanism on model performance. This involves understanding how the model handles information, learns representations, and ultimately predicts drug–disease relationships.
- 'w/o NCF': Further discourse is dedicated to the model's performance in the absence of collaborative filtering. This decision plays a crucial role in determining the model's effectiveness in handling user-item associations, particularly in our specific application scenario.

In WMAGT model, we employed a simplified approach, omitting the steps of neighbor embedding and interaction embedding, directly feeding the node representations obtained from the graph convolution module into the decoder. The rationale behind this decision and its implications on model performance necessitate a broader contextual understanding. The results of the ablation study in Fig. 5 showcase the consequences of these decisions. Notably, both the transformer module and the NCF module contribute significantly to enhancing the performance in predicting drug–disease relationships, with the NCF module being particularly noteworthy. This indicates that, when considering multiple embeddings and the transformer comprehensively, the model can more accurately capture latent relationships between drugs and diseases, thereby improving the predictive accuracy of drug–disease relationships. This finding provides profound insights for future model optimization and further research endeavors.

Case study

To assess the practical applicability of WMAGT, a case study was conducted with the aim of predicting drug candidates for Parkinson's disease. Specifically, the model was trained using all known drug–disease associations in the F dataset, and a descending order ranking was performed after obtaining the probabilities of all drug–disease associations. In this process, the top 10 drug candidates associated with Parkinson's disease were selected for in-depth investigation. Parkinson's disease is a chronic neurological disorder typically characterized by symptoms such as movement disorders, muscle stiffness, and tremors. The primary cause of this disease is the loss of dopamine-producing neurons in the brain, where dopamine functions as a neurotransmitter controlling movement. Currently, the treatment focus for Parkinson's disease primarily revolves around alleviating symptoms, and the exploration of new drug treatment directions has

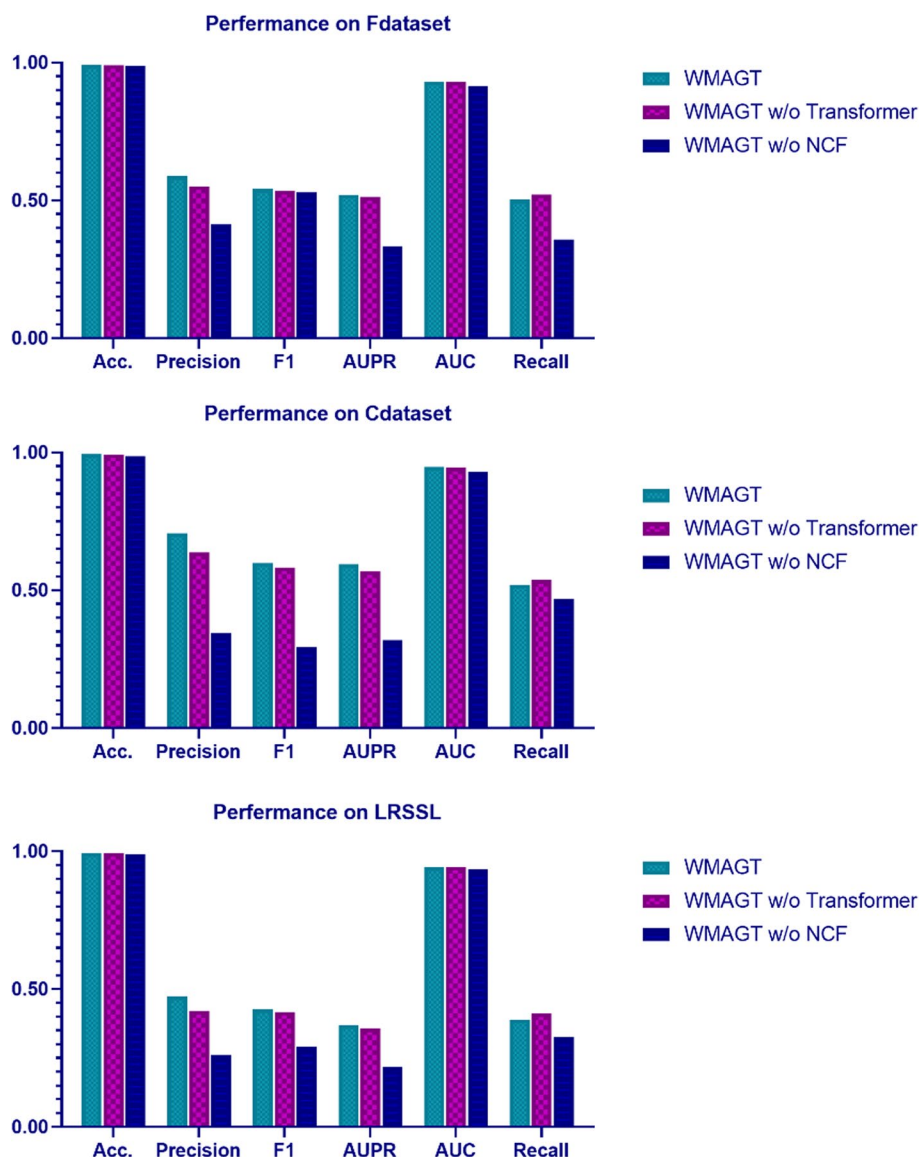


Fig. 5 The performance of WMAGT and other variants under tenfold cross-validation on three benchmark datasets

been a crucial area of scientific research. Encouragingly, the relevance of seven of these drugs was further confirmed by additional literature, as depicted in Table 3. This discovery not only enhances the reliability of our model but also indicates that WMAGT successfully identifies potential drug–disease pairs by learning multi-source information about drugs and diseases.

Conclusions

In this study, we propose a heterogenous information graph-based method for predicting drug–disease associations, named WMAGT. WMAGT innovatively integrates Graph Transformer Networks and Neural Collaborative Filtering, with a core improvement lying in the deep aggregation of local neighbors around nodes to enhance traditional

Table 3 The top 10 WMAGT-predicted candidate drugs for Parkinson's disease

Rank	Candidate drugs (DrugBank IDs)	Evidence (PMID)
1	Bupivacaine (DB00297)	NA
2	Hydromorphone(DB00327)	NA
3	Clotrimazole(DB00257)	12679339
4	Methylphenidate(DB00422)	18978488
5	Modafinil(DB00745)	12489899
6	Atenolol(DB00335)	NA
7	Ropinirole(DB00268)	9270567
8	Metformin(DB00331)	32854858
9	Guanidine(DB00536)	9548197
10	Olanzapine (DB00334)	11815682

graph convolution operations. Simultaneously, the model autonomously learns to select weights for different types of convolutional networks, resulting in a significant performance improvement compared to a singular graph convolution network. Extensive experiments were conducted to thoroughly assess the performance and robustness of WMAGT. WMAGT exhibited superior performance on three benchmark datasets, better than other compared state-of-the-art models. Ablation studies further verified the importance of different modules introduced in the proposed framework. In addition, the case study show that WMAGT has high practical predictive power, e.g., in Parkinson's potential drug mining, 7 of the top 10 drugs we predicted have been relevantly demonstrated. This study not only introduces methodological refinements but also substantiates their feasibility and superiority through rigorous experimentation and empirical validation. It's anticipated that these results can serve as valuable references for fostering further drug development and disease treatment.

Author contributions

S.-H.H. and L.-J.Y. conceived the algorithm, carried out analyses, prepared the data sets, carried out experiments, and wrote the manuscript. S.-H.H., L.-J.Y. and H.-C.Y. wrote the manuscript and analyzed experiments. All authors read and approved the final manuscript.

Funding

This work was supported in part by the Fundamental Research Funds for the Central Universities, under the Grant No. D5000230193, and in part by Natural Science Basic Research Program of Shaanxi (Program No. 2024JC-YBQN-0614).

Availability of data and materials

The code and datasets are freely available at: <https://github.com/ShiHHe/WMAGT>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 19 December 2023 Accepted: 14 February 2024

Published online: 20 February 2024

References

1. Chan HS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. *Trends Pharmacol Sci.* 2019;40(8):592–604.
2. Mak K-K, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today.* 2019;24(3):773–80.
3. Berdgaliev N, Aljofan M. An overview of drug discovery and development. *Future Med Chem.* 2020;12(10):939–47.
4. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, Schutjens M-HDB: drug repositioning and repurposing: terminology and definitions in literature. *Drug Discov Today.* 2015;20(8):1027–34.
5. Novac N. Challenges and opportunities of drug repositioning. *Trends Pharmacol Sci.* 2013;34(5):267–72.
6. Jarada TN, Rokne JG, Alhadj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform.* 2020;12(1):1–23.
7. Dudley JT, Deshpande T, Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. *Brief Bioinform.* 2011;12(4):303–11.
8. Jarada TN, Rokne JG, Alhadj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform.* 2020;12(1):46.
9. Cui Z, Gao Y-L, Liu J-X, Wang J, Shang J, Dai L-Y. The computational prediction of drug–disease interactions using the dual-network L 2, 1-CMF method. *BMC Bioinform.* 2019;20:1–10.
10. Fu G, Wang J, Domeniconi C, Yu G. Matrix factorization-based data fusion for the prediction of lncRNA–disease associations. *Bioinformatics.* 2018;34(9):1529–37.
11. Zhang W, Yue X, Huang F, Liu R, Chen Y, Ruan C. Predicting drug–disease associations and their therapeutic function based on the drug–disease association bipartite network. *Methods.* 2018;145:51–9.
12. Zhang W, Yue X, Lin W, Wu W, Liu R, Huang F, Liu F. Predicting drug–disease associations by using similarity constrained matrix factorization. *BMC Bioinform.* 2018;19:1–12.
13. Lu L, Yu H. DR2Dl: a powerful computational tool for predicting novel drug–disease associations. *J Comput Aided Mol Des.* 2018;32:633–42.
14. Zhou R, Lu Z, Luo H, Xiang J, Zeng M, Li M. NEDD: a network embedding based method for predicting drug–disease associations. *BMC Bioinform.* 2020;21(13):387.
15. Martinez V, Navarro C, Cano C, Fajardo W, Blanco A. DrugNet: network-based drug–disease prioritization by integrating heterogeneous data. *Artif Intell Med.* 2015;63(1):41–9.
16. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug–disease associations through layer attention graph convolutional network. *Brief Bioinform.* 2020;22(4):66.
17. Yang Y, Chen L. Identification of drug–disease associations by using multiple drug and disease networks. *Curr Bioinform.* 2022;17(1):48–59.
18. Gu Y, Zheng S, Yin Q, Jiang R, Li J. REDDA: integrating multiple biological relations to heterogeneous graph neural network for drug–disease association prediction. *Comput Biol Med.* 2022;150: 106127.
19. Wu G, Liu J, Yue X. Prediction of drug–disease associations based on ensemble meta paths and singular value decomposition. *BMC Bioinform.* 2019;20(3):1–13.
20. Li J, Zhang S, Liu T, Ning C, Zhang Z, Zhou W. Neural inductive matrix completion with graph convolutional networks for miRNA–disease association prediction. *Bioinformatics.* 2020;36(8):2538–46.
21. Kipf TN, Welling M. Variational graph auto-encoders. *arXiv preprint arXiv:161107308* 2016.
22. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief Bioinform.* 2022;23(2):bbab581.
23. Tang X, Zhou C, Lu C, Meng Y, Xu J, Hu X, Tian G, Yang J. Enhancing drug repositioning through local interactive learning with bilinear attention networks. *IEEE J Biomed Health Inform.* 2023;6:66.
24. Ghasemian A, Hosseinmardi H, Galstyan A, Airolidi EM, Clauset A. Stacking models for nearly optimal link prediction in complex networks. *Proc Natl Acad Sci.* 2020;117(38):23393–400.
25. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, Zhang Y. Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform.* 2016;17(4):696–712.
26. Wang C-C, Zhao Y, Chen X. Drug-pathway association prediction: from experimental results to computational models. *Brief Bioinform.* 2021;22(3):bbaa061.
27. Li T-H, Wang C-C, Zhang L, Chen X. SNRMPACDC: computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction. *Brief Bioinform.* 2023;24(1):bbac503.
28. 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.
29. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics.* 2018;34(11):1904–12.
30. Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2002;30(1):52–5.
31. Wishart DS, Knox C, Guo AC, 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(1):D668–72.
32. 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(17):2664–71.
33. 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(8):1187–96.

34. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:160902907](https://arxiv.org/abs/1609.02907) 2016.
35. Ma Y, Wang S, Aggarwal CC, Tang J. Graph convolutional networks with eigenpooling. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; 2019. p. 723–31.
36. Wu F, Souza A, Zhang T, Fifty C, Yu T, Weinberger K. Simplifying graph convolutional networks. In: International conference on machine learning; 2019. PMLR. p. 6861–71.
37. Gao H, Wang Z, Ji S. Large-scale learnable graph convolutional networks. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining; 2018. p. 1416–24.
38. Yun S, Jeong M, Kim R, Kang J, Kim HJ. Graph transformer networks. *Adv Neural Inf Process Syst.* 2019;32:66.
39. Cai D, Lam W. Graph transformer for graph-to-sequence learning. In: Proceedings of the AAAI conference on artificial intelligence; 2020. p. 7464–71.
40. Dwivedi VP, Bresson X. A generalization of transformer networks to graphs. arXiv preprint [arXiv:201209699](https://arxiv.org/abs/2012.09699) 2020.
41. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, Shen Y, Liu T-Y. Do transformers really perform badly for graph representation? *Adv Neural Inf Process Syst.* 2021;34:28877–88.
42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inform Process Syst.* 2017;30:66.
43. Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint [arXiv:180308375](https://arxiv.org/abs/1803.08375) 2018.
44. He X, Liao L, Zhang H, Nie L, Hu X, Chua T-S. Neural collaborative filtering. In: Proceedings of the 26th international conference on World Wide Web; 2017. p. 173–82.
45. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint [arXiv:14126980](https://arxiv.org/abs/1412.6980) 2014.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.