BMC Bioinformatics

RESEARCH

Open Access

Predicting drug combination side effects based on a metapath-based heterogeneous graph neural network

Leixia Tian[2,3,4†], Qi Wang[1†], Zhiheng Zhou[3,4], Xiya Liu[5], Ming Zhang[3,4] and Guiying Yan[3,4*]

†Leixia Tian and Qi Wang have contributed equally to this work, and their names are listed in alphabetical order.

*Correspondence: yangy@amss.ac.cn

[1] College of Science, China Agricultural University, Beijing 100083, China
[2] Beijing School, Beijing 100088, China
[3] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[4] University of Chinese Academy of Sciences, Beijing 100190, China
[5] Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Abstract

In recent years, combined drug screening has played a very important role in modern drug discovery. Generally, synergistic drug combinations are crucial in treatment for many diseases. However, the toxic side effects of drug combinations are probably increased with the increase of drugs numbers, so the accurate prediction of toxic side effects of drug combinations is equally important. In this paper, we built a Metapath-based Aggregated Embedding Model on Single Drug–Side Effect Heterogeneous Information Network (MAEM-SSHIN), which extracts feature from a heterogeneous information network of single drug side effects, and a Graph Convolutional Network on Combinatorial drugs and Side effect Heterogeneous Information Network (GCN-CSHIN), which transforms the complex task of predicting multiple side effects between drug pairs into the more manageable prediction of relationships between combinatorial drugs and individual side effects. MAEM-SSHIN and GCN-CSHIN provided a united novel framework for predicting potential side effects in combinatorial drug therapies. This integration enhances prediction accuracy, efficiency, and scalability. Our experimental results demonstrate that this combined framework outperforms existing methodologies in predicting side effects, and marks a significant advancement in pharmaceutical research.

Keywords: Combinatorial drugs, Side effect prediction, Metapath, Graph convolutional network, Heterogeneous information network

Introduction

The use of combinatorial drug therapy or polypharmacy has become a primary clinical approach in modern disease treatment, with most hospital visits resulting in prescriptions for more than two types of drugs [1, 2]. The absorption and metabolism of different drugs within the body can lead to various negative interactions, such as enhancing or diminishing clinical efficacy and increasing the risk of drug interactions, and potential adverse drug reactions [3–6]. Therefore, predicting the side effects of multiple drugs and adequately preparing for them is critically important.

However, as the number of drugs increases, the potential combinations sometimes reach an explosive level, making precise prediction challenging, akin to finding a needle

in a haystack. Mathematical models can effectively address this complexity by systematically analyzing and prioritizing treatment options, thereby aiding clinicians in making effective decisions and devising suitable treatment strategies [7, 8].

To address these challenges, pioneering computational methods have been developed to predict drug side effects. For instance, Atias et al. [9] first proposed a computational approach focusing on individual drugs. Building on this foundation, the field has seen advancements with models like the Decagon [10], which integrates multimodal networks including protein and drug-protein interactions to predict multiple drug side effects using a graph convolutional network and a tensor decomposition decoder. This model significantly outperforms traditional methods by 69% in accuracy. Subsequent enhancements by Xu et al. [11] with the Tripartite Information Propagation model, and Carletti et al. [12] through the incorporation of a graph attention network, have further refined the accuracy of these predictions by emphasizing different weights in side effect relationships within heterogeneous graphs. Liu et al. [13] developed SC-DDIS, introducing additional drug features into the model to optimize predictions further. Extended applications like those by Brandon et al. [14] utilize Decagon data to create drug-protein knowledge graphs employing learning frameworks such as DistMult and KBlrn for improved prediction outcomes. New methodologies like the TriVec [15] and SimVec [16] models respectively, focus on enhancing node initialization in knowledge graphs, illustrating ongoing innovation in the field. Most recently, Jinwoo et al. [17] presented a unified embedding prediction model based on knowledge graphs, setting a new direction for future drug research by improving the accuracy of multi-drug side effect predictions.

In response to these advancements, this study proposes two distinct heterogeneous graph neural network models tailored for single and combination drugs within heterogeneous information networks. These models are designed to enhance the learning of feature vector representations of drugs and side effect nodes, optimizing the prediction of combinatorial side effects. An in-depth evaluation of these models demonstrates their predictive effectiveness, underscoring the potential of heterogeneous graph embedding in drug research.

Heterogeneous graph embedding, essential for projecting nodes into a low-dimensional vector space, has evolved significantly with two main categories of models. The first category, including shallow embedding algorithms like those based on the skip-gram model, captures semantic similarities through proximity in low-dimensional space, assuming nodes with similar neighbors should be closely positioned [18–24, 57]. The second category encompasses Graph Neural Networks (GNNs), which employ complex encoders to model network structure and node features more deeply [25–33]. This category includes significant innovations such as the application of GNNs to heterogeneous graphs by models, each contributing uniquely to the field by handling different types of data and network dynamics. For our Single Drug–Side Effect Heterogeneous Information Network, integrating metapaths based on proteins and side effects, as well as their features, is crucial to better represent drug node features and enhance prediction accuracy.

In this paper, we introduce two innovative models: the Metapath-based Aggregated Embedding Model for Single Drug-Side Effect Heterogeneous Information Networks (MAEM-SSHIN) and the Graph Convolutional Network for Combinatorial Drugs and

Side Effect Heterogeneous Information Networks (GCN-CSHIN). MAEM-SSHIN effectively extracts features from a heterogeneous information network, focusing on the side effects associated with individual drugs, while GCN-CSHIN simplifies the complex task of predicting multiple side effects between drug pairs by forecasting interactions between combinatorial drugs and individual side effects. The integration of these models into a unified framework significantly enhances the accuracy, efficiency, and scalability of predicting potential side effects in combinatorial drug therapies. Our experimental results demonstrate that this combined approach substantially outperforms existing methodologies in side effect prediction, representing a major advancement in pharmaceutical research. This achievement highlights the transformative potential of advanced computational models in enhancing drug safety and efficacy.

Methods

Notations and definitions

Heterogeneous Information Graph: Let $G=(V, E,\mathcal{A}, \mathcal{R})$ be a graph, where V is the set of nodes of the graph G, E is the set of edges of the graph G, is the set of node types, each representing a distinct category of nodes based on their features or roles within the graph, and $\mathcal{R}$ is the set of edge types, each signifying a distinct kind of relationship or interaction between the nodes. If $|\mathcal{A}|+|\mathcal{R}|>2$, then the graph G is called a heterogeneous graph, which can also be referred to as a Heterogeneous Information Network (HIN).

Metapath: Given a heterogeneous graph $G=(V,E,,\mathcal{R})$, a metapath P is a sequence composed of node types and edge types, represented as $A_1 \rightarrow^{R_1} A_2 \rightarrow^{R_2} \cdots \rightarrow^{R_l} A_{l+1}$, which can be abbreviated as $A_1A_2\cdots A_{l+1}$. Here, $A_1, A_2, \ldots, A_{l+1} \in \mathcal{A}, R_1, R_2, \ldots, R_l \in \mathcal{R}$, describing a composite relationship $R = R_1 \circ R_2 \circ \cdots \circ R_l$, where $\circ$ is a composite relation operator.

Metapath Instance: Given a metapath $P = A_1A_2\cdots A_{l+1}$ in a heterogeneous graph $G = (V, E, \mathcal{A}, \mathcal{R})$, an instance of metapath P refers to a sequence of nodes in graph G that follows the pattern of metapath P, denoted as $v_{1_i}v_{2_j}\ldots v_{(l+1)_k}$. Here, $v_{1_i}, v_{2_j}, \ldots, v_{(l+1)_k} \in V$, with v_{1_i} being of type A_1, v_{2_j} of type A_2, and $v_{(l+1)_i}$ of type A_{l+1}.

Metapath-based Neighbors: Let $G = (V, E, \mathcal{A}, \mathcal{R})$ be a heterogeneous graph and P be a metapath. Define the set of neighbors $\mathcal{N}_v^{\mathcal{P}}$ for a node v based on metapath P as:

$$\mathcal{N}_v^{\mathcal{P}} = \{u \in V : \text{there exists a path from } v \text{ to } u \text{ following instances of } P\}$$

Furthermore, if P is symmetric, then include v in $\mathcal{N}_v^{\mathcal{P}}$, i.e.,$v \in \mathcal{N}_v^{\mathcal{P}}$.

Metapath-based Graph: Given a heterogeneous graph $G = (V, E, \mathcal{A}, \mathcal{R})$ and a metapath P.

The metapath-based graph G^P is defined as $G^P = \left(V, E^P\right)$, where V is the set of all nodes in the original graph G, E^P is the set of edges in the graph G^P, specifically defined as:

$$E^P = \{(u, w) \in V \times V : u, w \in \mathcal{N}_{\sqsubseteq}^{\mathcal{P}} \text{ for some } v \in V \text{ and there exists a path following } P \text{ between } u \text{ and } w\}$$

If P is symmetric, then G^P is defined as a homogeneous graph, implying that all edges (u, w) conform uniformly to the relations specified by P.

Heterogeneous Graph Embedding: Let $G=(V, E, A, \mathcal{R})$ be a heterogeneous graph. For each node type $A_i \in \mathcal{A}$, let $V_{A_i} \subseteq V$ denote the set of nodes of type A_i, and let $X_{A_i} \in \mathbb{R}^{\left|V_{A_i}\right| \times d_{A_i}}$ be the feature matrix associated with nodes of type A_i, where d_{A_i} represents the dimensionality of the feature space for nodes of type A_i. Here, $\mathbb{R}$ denotes the real numbers.

Heterogeneous graph embedding aims to learn a d-dimensional vector representation for each node $v \in V$, denoted as $v \in R^d$. This representation should satisfy the condition that d is significantly smaller than $|V|$, the total number of nodes in the graph.

Heterogeneous information network construction

In this study, we initially constructed six network structures using data from the bioinformatics database at Stanford University BioSNAP Datasets [34]. The sources and methods of construction for these networks are detailed as follows:

Combinatorial drugs and side effect network

The TWOSIDES database [35] revealed 4,649,441 combinatorial drugs and side effect relationships, involving 63,473 drug pairs, 645 drugs, and 1,317 side effects. Here, the term 'drug' includes both combinatorial and single drugs unless specified otherwise. The frequency distribution of combinatorial drugs with side effects is shown in Fig. 1,

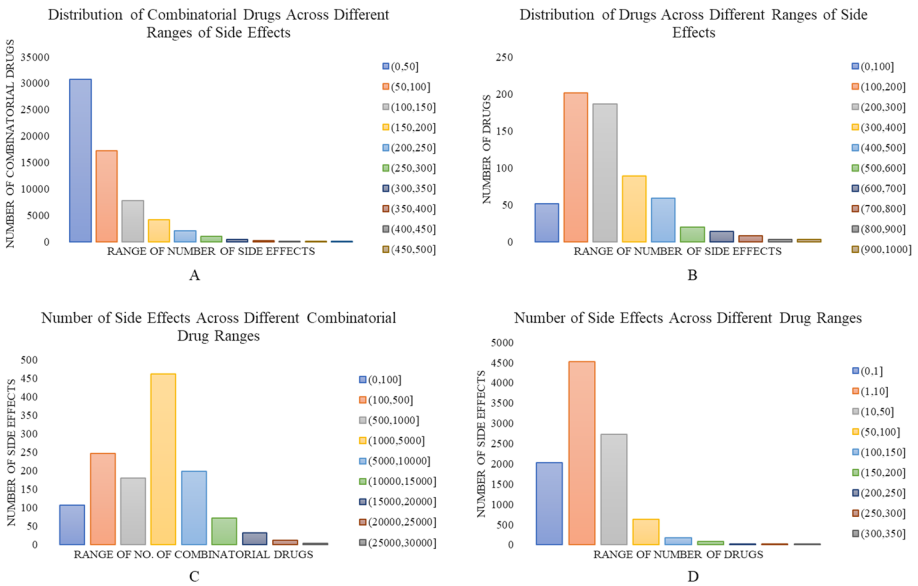

Fig. 1 Distribution of side effects across drug combinations and single drugs. **A** Distribution of side effects across different drug ranges. **B** Distribution of side effects across different combinatorial drug ranges. **C** Distribution of drugs across different ranges of side effects. **D** Distribution of combinatorial drugs across different ranges of side effects

with 964 side effects occurring in at least 500 drug combinations, forming our final dataset for experimentation.

Drug-side effect network

Individual drug side effect information was derived from the SIDER [36] and OFFSIDES [35] databases. SIDER data, includes 286,399 drug-side effect relationships involving 1556 drugs and 5868 side effects. OFFSIDES data, collected from adverse event reporting systems, consists of 487,530 drug-side effect pairs involving 1332 drugs and 10,097 side effects. Combining these databases, we compiled 174,977 drug-side effect relationships involving 639 drugs and 10,184 side effects. The frequency distribution of single drugs with side effects generally ranges from 100 to 300, up to over 1000 (Fig. 1). In contrast, single drug side effects are more specific, with 9271 side effects observed in a maximum of only 50 drugs.

These four charts illustrate the distribution of drugs and their combinations across different ranges of side effect counts, providing an in-depth analysis of the relationship between side effect frequency and the number of drug combinations.

The first chart (A) shows the distribution of side effects across different drug count ranges. It is evident that the number of side effects is concentrated in a small range of drugs, particularly in the [1, 9] range, where the number of side effects is significantly higher than in other ranges. This suggests that a small number of drug combinations are responsible for generating many side effects.

The second chart (B) displays the distribution of side effects across different drug combination count ranges. Similar to the first chart, the number of side effects is largely concentrated in the lower drug combination ranges (such as [0, 100] and [100,500]), indicating that these combinations are more likely to produce a higher number of side effects.

The third chart (C) focuses on the distribution of drugs within different ranges of side effect counts. The data shows that most drugs have a small range of side effects (e.g., [100,200] and [200,300]), while drugs with a high number of side effects are relatively rare.

The final chart (D) illustrates the distribution of drug combinations across different ranges of side effects. The results indicate that the majority of drug combinations generate fewer side effects (e.g., [0, 50] and [50, 100]), while the number of drug combinations sharply decreases as the number of side effects increases.

Drug-drug similarity network

Utilizing chemical structural information, we obtained SMILES notation for each drug from DrugBank and transformed it into 2048-dimensional binary vectors using RDKit [37, 38]. Let $\mathbf{A}$ and $\mathbf{B}$ be vectors in Z_2^{2048}, representing the corresponding binary vectors for drugs a and b. Here, Z_2 refers to the set of binary values $\{0, 1\}$, where each vector component is either 0 or 1. The similarity of their chemical structures is calculated using the following formula (1). Subsequently, a threshold c is selected. If the chemical structure similarity SimChem is greater than or equal to c, an edge is established between the nodes of drug a and drug b; otherwise, no edge is formed.

$$\text{SimChem}=\frac{2A^T B}{|A|+|B|} \tag{1}$$

Building on the combinatorial drugs and side effect data from the previous database, we analyzed the similarity distributions among 639 drugs, as summarized in Table 1, finding that most similarities were below 0.3. To more effectively leverage chemical structural similarities in our network construction, we established a similarity threshold of 0.2. Drugs exhibiting a similarity above this threshold were linked, resulting in the formation of 52,557 edges in the network.

Drug-protein network

Relationships between drugs and proteins were obtained from the STITCH database [39], focusing solely on experimentally verified interactions between small molecular compounds and target proteins. This dataset includes 8,083,600 drug-protein interactions, involving 8934 proteins and 519,022 compounds. In the resulting network, each protein and each compound are represented as nodes. Edges between nodes represent the verified interactions from the STITCH database.

Protein–protein interaction network

We utilized the human PPI network compiled by Menche et al. [40] and Chatr-Aryamontri et al. [41], combined with additional PPI data from STRING [42] and Rolland et al. [43]. This network includes physically interacting proteins as recorded in human experiments, unweighted and undirected, comprising 19,085 proteins and 719,402 interactions.

Side effect-protein network

While no database directly records side effect-protein interactions, indirect relationships can be established through associated diseases. We extracted side effect-protein relationships from disease-protein data sourced from the DisGeNET database [44], which

Table 1 Statistics of chemical structure similarity of 582 drugs

Similarity	Combinatorial drug pairs
0	5607
(0, 0.1]	23,737
(0.1, 0.2]	89,072
(0.2, 0.3]	43,642
(0.3, 0.4]	5571
(0.4, 0.5]	864
(0.5, 0.6]	273
(0.6, 0.7]	165
(0.7, 0.8]	99
(0.8, 0.9]	64
(0.9, 1]	13

Table 2 Experimental data for predicting combination drug side effects

Networks	No. of interaction	No. of combinatorial drugs	No. of drug	No. of protein	No. of side effect
Combinatorial Drugs-Side Effect	4,576,287	63,472	645	–	964
Drug-Side Effect	169,906	–	611	–	9167
Drug-Drug	169,071	–	582	–	–
Drug-Protein	18,632	–	276	3632	–
Protein–Protein	460,803	–	–	17,861	–
Side Effect-Protein	5489	–	–	3039	125

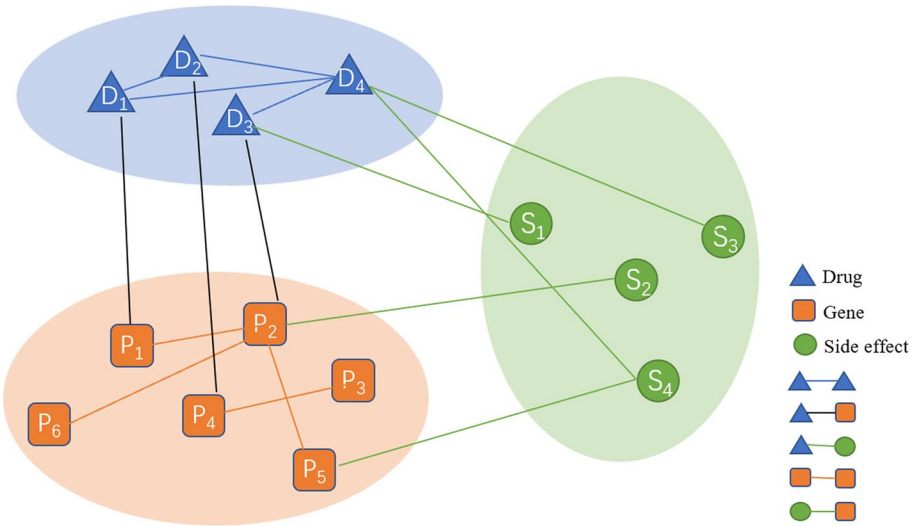

Fig. 2 Single drug–side effect heterogeneouss information network

includes 21,357 relationships between 519 diseases and 7294 proteins. In this network, nodes represent both proteins and side effects. Edges are established based on a documented association between a protein and a disease that manifests a given side effect.

In this study, we focused on predicting side effects related to combinatorial drug therapies. We selected 964 side effects that appeared in at least 500 drug combinations as our experimental dataset. Detailed dataset and network information is in Table 2.

Utilizing the above network information, we integrated and constructed two heterogeneous information networks as follows:

The first network is the Single Drug–Side Effect Heterogeneous Information Network (SSHIN), depicted in Fig. 2. This network is composed of five sub-networks: the drug-drug chemical structure similarity network, the drug-side effect network, the drug-protein network, the side effect-protein network, and the protein–protein interaction network. It is formally represented as the heterogeneous graph $G_{SSHIN} = (V_1, E_1, \{D, P, S\}, \{D-D, D-P, P-P, D-S, S-P\})$. Here, D (Drug) denotes the set of drug nodes, P (Protein) denotes the set of protein nodes, and S (Side effect) denotes the set of side effect nodes. To analyze and interpret this complex

network, we utilized a metapath-based graph neural network model, which was specifically designed to learn feature vector representations of drugs, thereby laying the groundwork for subsequent predictive tasks.

The second network we constructed is the Combinatorial Drugs and Side Effect Heterogeneous Information Network (CSHIN), as illustrated in Fig. 3. This network establishes a bipartite graph structure between combinations of drugs and their corresponding side effects. An edge in this network indicates that a specific combination of drugs is associated with a particular side effect. However, it is important to note that the edges between drug combinations and side effects in Fig. 3 do not imply that individual drugs necessarily cause these side effects (i.e., this result is not derived directly from the information in Fig. 2); rather, it is the combination of different individual drugs that may lead to these side effects (independent of Fig. 2's information). It is formally represented as the heterogeneous graph $G_{CSHIN} = (V_2, E_2, \{C, S\}, \{CD - S\})$, where C (Combinatorial drugs) represents nodes of drug combinations, and S (Side effect) represents nodes of side effects. For analysis and prediction within this network, we utilized advanced graph convolutional neural network models specifically designed for edge prediction tasks, effectively identifying potential adverse effects arising from drug combinations.

Computational models

Inspired by the metapath-based models [33, 45], we propose the Metapath-based Aggregated Embedding Model on Single Drug–Side Effect Heterogeneous Information Network (MAEM-SSHIN), to learn embeddings of drug and side effect nodes in SSHIN.

In this paper, within the Combinatorial Drugs and Side Effect Heterogeneous Information Network, we transform the relationships between combinatorial drug pairs and side effects into an edge prediction problem. To address this, we employ a two-layer graph convolutional neural network model, which is specifically designed for predicting new connections. This model utilizes feature vectors of drug nodes and side effect nodes, which are learned from the SSHIN (Single Drug–Side Effect Heterogeneous Information

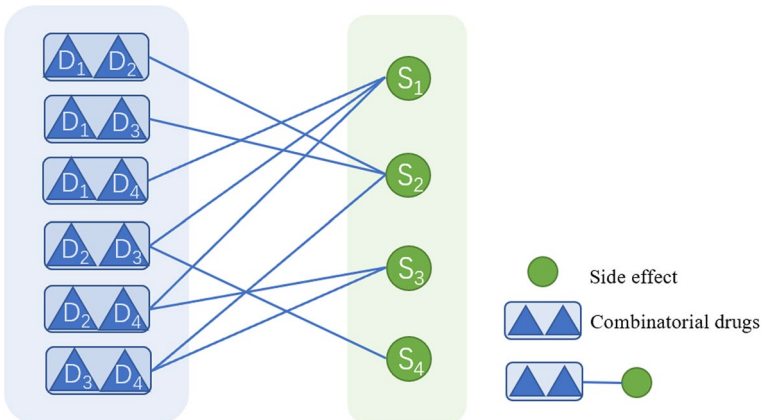

Fig. 3 Combinatorial drugs and side effect heterogeneous information network

Network). We denote this model as GCN-CSHIN (Graph Convolutional Network Model on Combinatorial Drugs and Side Effect Heterogeneous Information Network), emphasizing its tailored application for intricate network dynamics and interactions.

MAEM-SSHIN model

This section outlines the Metapath-based Aggregated Embedding Model in the Single Drug–Side Effect Heterogeneous Information Network (MAEM-SSHIN). The model includes the following components: metapath selection, node feature transformation, intra-metapath information aggregation, metapath inter-information aggregation, and node feature vector embedding learning.

Metapath selection In the construction of the single drug and side effect network, our objective is to elucidate the relationships among drugs in terms of chemical structure, target protein, and side effects, as well as the connections between side effects, proteins, and individual drugs. To achieve this, we begin by setting the maximum length of metapaths at 3, allowing for comprehensive exploration within the network. This constraint includes two types of nodes: firstly, the first-order neighboring nodes of drug nodes (or side effect nodes); secondly, the nodes at the second and third orders of adjacency from drug nodes (or side effect nodes), essentially limiting the selection to metapaths that start and end at the same type of node. This strategic selection results in 11 types of metapaths for drug nodes ($DD, DP, DS, DPD, DSD, DDPD, DDSD, DPDD, DSDD, DPPD, DPSD$), and 8 types for side effect nodes ($SD, SP, SDS, SPS, SDDS, SPPS, SPDS, SDPS$), effectively mapping the intricate interconnections within the network. Conversely, longer metapaths (e.g., length of 4) tend to introduce excessive noise, potentially diminishing the accuracy of predictions. Our approach to restrict the path length aims to balance computational complexity and the need for predictive accuracy, ensuring the model remains efficient without being overly complex.

Node feature transformation In heterogeneous graphs, which contain nodes with diverse features, different node types may have feature vectors—numerical representations capturing the attributes or characteristics of nodes—of varying dimensions, meaning the number of elements in each vector can differ. Even when these vectors share the same dimensionality, they may still originate from distinct feature spaces, which are mathematical spaces defined by the possible values and relationships of the features. To effectively process entities like drugs, target proteins, and side effects within a unified analytical framework, it is essential to map the initial feature vectors of various node types into a common latent feature space. This mapping is accomplished through type-specific linear transformations, ensuring that disparate data types are harmonized for subsequent analysis and modeling [32].

In our model, the initial feature vectors of all nodes are encoded as one-hot vectors based on their type, and subsequently projected into the same feature space using a node type-specific transformation matrix W_l. The transformation is represented by the equation:

$$h'_v = W_l \cdot x_v, v \in V \tag{2}$$

where h'_v and x_v are the post-transformation and original feature vectors of node v, respectively, with $l \in \{D, P, S\}$ denoting the node's type.

Intra-metapath information aggregation This section is dedicated to elucidating the process of information aggregation. It uses metapaths originating from drug nodes as a primary example to demonstrate how this process unfolds.

Given a metapath related to a drug node P_m, where m belongs to the set {1,2, . . . , 11}, and considering $p_m(v,u), u \in N_{P_m}^v$ as the metapath instance for node v, the feature information of node v is transmitted along the metapath instance $p_m(v,u)$ as follows:

$$h_{p_m(v,u)} = f_\theta\left(h'_t,\right), \forall t \in V_{p_m}(v,u) \tag{3}$$

In this equation, $h_{p_m}(v,u) \in \mathbb{R}^{d'}$ is a vector of dimension d', where $V_{p_m}(v,u)$ signifies all nodes present in the metapath instance $p_m(v,u)$, and $f_\theta(\cdot)$ represents an information transfer function connected with the parameter θ. The metapath instances are encoded into feature vectors, following which a graph attention mechanism is utilized to assign weights to diverse metapath instances for the target protein node v. The information from all metapath instances is aggregated through a weighted summation approach.

$$\begin{aligned} e_{p_m(v,u)} &= \text{LeakyReLU}\left(a_{P_m}^T \cdot \left[h'_v \parallel h_{p_m(v,u)}\right]\right) \\ \alpha_{p_m(v,u)} &= \frac{\exp\left(e_{p_m(v,u)}\right)}{\sum_{s\in\mathcal{N}_v^{P_m}} \exp\left(e_{p_m(v,s)}\right)} \\ h_v^{P_m} &= \sigma\left(\sum_{u\in\mathcal{N}_v^{P_m}} \alpha_{p_m(v,u)} \cdot h_{p_m(v,u)}\right) \end{aligned} \tag{4}$$

Here, $\boldsymbol{a}_{Pm} \in \mathbb{R}^{2d'}$ represents the attention parameter vector for metapath P_m, with '$\parallel$' denoting the concatenation of two vectors. The term $e_{p_m(v,s)}$ signifies the importance of the metapath instance $p_m(v,u)$ for node v. This importance is then normalized using the softmax function to obtain the weight $\alpha_{p_m(v,u)}$ for the metapath instance $p_m(v,u)$. Consequently, all metapath instances are aggregated to derive the feature vector $h_v^{P_m}$ for node v on metapath P_m. The symbol $\sigma(\cdot)$ denotes the activation function. To enhance the model's expressive power and stabilize the learning process, we employ a multi-head attention strategy, executing K independent attention mechanisms and subsequently concatenating their outputs as the final result.

$$h_v^{P_m} = \parallel_{k=1}^{K} \sigma\left(\sum_{u\in\mathcal{N}_v^{P_m}} \left[\alpha_{p_m(v,u)}\right]_k \cdot h_{p_m(v,u)}\right) \tag{5}$$

Here, $\left[\alpha_{p_m(v,u)}\right]_k$ denotes the weight of the metapath instance $p_m(v,u)$ under the k^{th} attention mechanism.

The methodology for metapaths starting from side effects is analogous and will not be elaborated on further.

Metapath inter-information aggregation After aggregating node and edge data within each metapath, we further aggregate the features across different metapaths to achieve a comprehensive feature expression for all nodes across all metapaths. Here is an example for drug nodes, which are representations of pharmaceutical compounds in the network, each characterized by their vector feature expressions. While this

example focuses on drug nodes, the same aggregation process is applied to side effect nodes, which also involve assembling vector feature expressions to capture the comprehensive impact of potential adverse effects associated with the drugs.

For a node v in a graph G, after the steps previously described, we obtain feature vector expressions for a drug node v across metapath P_m, represented as $h_v^{P_m}$, where $m \in \{1, 2, \cdots, M\}$.

Next, we assign a different weight to each metapath to aggregate all metapath information comprehensively. Specifically, we calculate the average of the feature vectors for drug node v, represented in V_D, across each metapath P_m, where $m \in \{1, 2, \cdots, M\}$. This average reflects the consolidated information from all instances of node v on metapath P_m, capturing the integrated effects of various relational paths in the network.

$$\boldsymbol{s}_{Pm} = \frac{1}{|V_D|} \sum_{v \in V_D} \tan\ h\left(M_D \cdot h_v^{Pm} + b_D\right) \tag{6}$$

where, $M_D \in \mathbb{R}^{d_D \times d'}$ and $b_D \in \mathbb{R}^{d_D}$ are the parameters to be learned.

Then, an attention mechanism is employed to aggregate the node vectors of node v across all metapaths.

$$\begin{aligned} e_{Pm} &= \boldsymbol{q}_D^T \cdot \boldsymbol{s}_{Pm} \\ \beta_{Pm} &= \frac{\exp\left(e_{Pm}\right)}{\sum_{i=1}^{M} \exp\left(e_{Pi}\right)} \\ h_v &= \sum_{i=1}^{M} \beta_{Pm} \cdot h_v^{Pm} \end{aligned} \tag{7}$$

where, $\mathrm{q_D} \in \mathbb{R}^{\mathrm{d_D}}$ is the attention parameter for drug nodes to be learned, and h_v is the final aggregated embedding vector for node v.

Node feature vector embedding learning To predict drug-drug and drug-side effect linkages, we employ the Binary Cross Entropy loss (BCE) as the loss function to assess the discrepancy between the predicted values and the actual outcomes. The formula is:

$$\mathcal{L} = -\frac{1}{B} \sum_{b=1}^{B} \left[y_b \log\left(y_b'\right) + \left(1 - y_b\right) \log\left(1 - y_b'\right) \right] \tag{8}$$

Here, B represents the number of training data per batch, y_b denotes the actual value of the b^{th} edge prediction sample, where the presence of an edge is indicated by 1, and absence by 0. $y_b' = F(h_u, h_v)$ signifies the predicted probability of an edge existing between node pair (u, v) in the model, where $F(\cdot)$ is a function calculating the similarity between two feature vectors. Ultimately, when the loss function falls below a set threshold, the resulting h_v is the final embedding vector for node v.

With this, the complete process of the MAEM-SSHIN model is thoroughly described. The specific flowchart is presented in Fig. 4.

GCN-CSHIN model

The goal of GCN-CSHIN model is to predict whether an edge exists between combinatorial drugs and side effects in a heterogeneous graph. An edge signifies the emergence

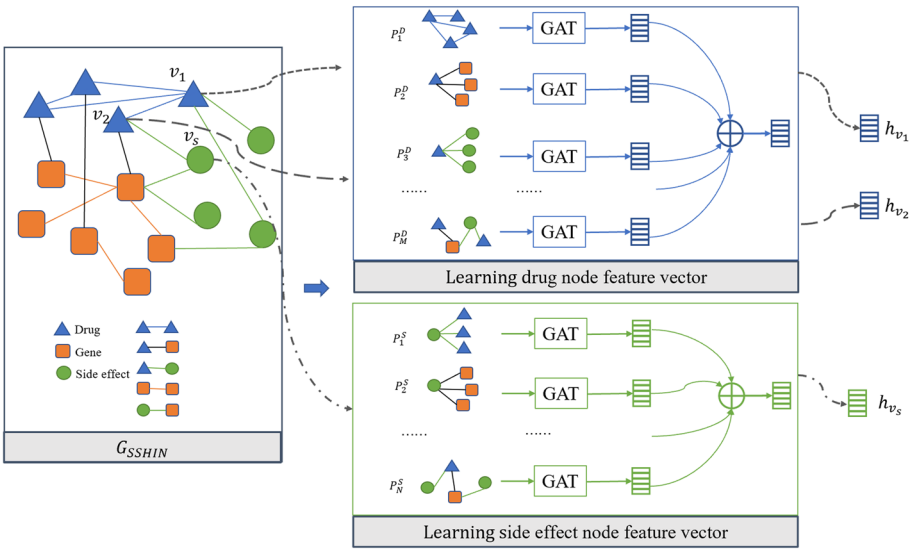

Fig. 4 Flowchart of MAEM-SSHIN

of a particular side effect due to the combinatorial drugs. We employ a Graph Convolutional Network (GCN) model to aggregate structural information of combinatorial drugs and side effect nodes for this edge prediction task.

Leveraging the MAEM-SSHIN model, we obtain feature vector representations for each drug node. By combining the feature vectors of two drug nodes, we derive the initial feature $f_0^c = h_{v_i} + h_{v_j}$ for the combinatorial drugs node in the heterogeneous graph G_{CSHIN}, where $c \in V_C$ and $v_i, v_j \in V_D$. This additive approach is chosen to mitigate the influence of the drug combination's sequence on prediction outcomes. The initial feature of side effect nodes is as obtained from the MAEM-SSHIN model, i.e., $f_0^s = h_{v_s}$, where $v_s \in V_S$.

In the GCN model, the node information propagation formula for each layer (illustrated here for combinatorial drugs nodes) is:

$$f_{k+1}^c = \phi\left(\sum_{u \in N(c)} \gamma_{cu} W_k f_k^u + f_k^c\right) \tag{9}$$

Here, f_k^c and f_{k+1}^c denote the feature representations of node c at the k^{th} and $(k+1)^{th}$ layer, respectively, $N(c)$ is the set of neighboring nodes of c, $\gamma_{cu} = 1/|N(c)|$, W_k is the parameter matrix, and $\phi(\cdot)$ is the activation function. We use a two-layer GCN model to ultimately derive the feature representations f_c and f_s for combinatorial drugs and side effect nodes in the aggregated CSHIN graph, where $c \in V_C and v_s \in V_S$. The node pair combines as $f = f_c || f_s$ and inputs into a Multilayer Perceptron (MLP) model to yield the prediction result:

$$y = \sigma\left(W_f f\right) \tag{10}$$

Here, W_f represents the trainable parameters, and $\sigma(\cdot)$ is the activation function. The Binary Cross-Entropy loss is used to evaluate the difference between predicted and actual results:

$$\mathcal{L} = -\frac{1}{B}\sum_{b=1}^{B}\left[y_b \log(y_b\prime) + (1 - y_b)\log(1 - y_b\prime)\right] \tag{11}$$

In this formula, B is the number of training data per batch, $y_b = 1$ indicates the presence of an edge between the b^{th} pair of combinatorial drugs and side effect nodes, and $y_b\prime$ denotes the model's predicted probability of an edge existing between them. For the flowchart of the GCN-CSHIN process, refer to Fig. 5.

Results

Evaluation metrics

To assess the performance of our model, we utilized three commonly employed metrics in the field of multi-drug side effect prediction:

- *AUROC (Area under the curve ROC)* This is the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve's vertical axis represents the True Positive Rate (TPR), defined as $TPR = Recall = \frac{TP}{TP+FN}$, where the horizontal axis is the False Positive Rate (FPR), defined as $FPR = \frac{FP}{FP+TN}$.

- *AUPRC (area under the curve PRC)* This is the area under the Precision-Recall Curve (PRC). The PRC curve's vertical axis is Precision, defined as $Precision = \frac{TP}{TP+TN}$, and the horizontal axis is Recall, defined as $Recall = \frac{TP}{TP+FN}$.

Parameter settings

Parameter settings were determined using a grid search approach for evaluating candidate values, while ensuring the consistency of the training, testing, and validation nodes. The key settings for the MAEM-SSHIN model include employing a dual attention mechanism and a two-layer neural network architecture, with each layer having a dimension of 128. Negative sample selection is based on random walk sampling, and the loss function is set to binary cross-entropy. Similarly, the GCN-CSHIN model utilizes a two-layer graph neural network, where each layer is also 128-dimensional.

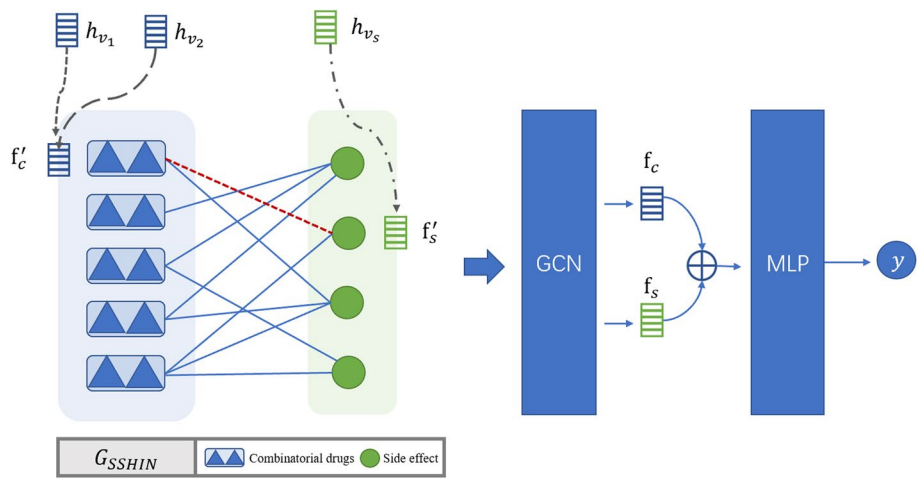

Fig. 5 Flowchart of GCN-CSHIN

We conduct a fivefold cross-validation, ensuring that the training, validation, and test nodes are consistent across both MAEM-SSHIN and GCN-CSHIN. During each fold of cross-validation, we keep the node division fixed to ensure consistency of the train, validation, and test nodes. For each fold, 80% of the positive and negative samples are randomly selected as the training set, and the remaining samples are used as the test set. All models are trained and evaluated on the same partitioned datasets. The learning rate is a dynamic learning rate, where the initial rate is 1e-3 and decreases by 0.05 every 100 steps. The batch sizes for training and testing differ, with the training batch size being 1024 and the test batch size being 4096. The smaller batch size during training allows for more frequent updates and better memory management, while the larger batch size during testing enables faster and more efficient evaluation of the model's performance.

Case study

To validate whether our model is capable of predicting side effects for drug combinations that currently lack recorded side effect information, we designed a case study specifically to test the model's predictive capabilities.

We conducted experiments using a dataset containing 915,583 edges. Given the large size of the dataset, we adopted a batch processing approach, predicting 4098 positive edges per batch. Through multiple experiments, the model generated over 200 top-10 prediction results. Among all the predicted results, the side effect of "anaemia" was predicted the most frequently, appearing in 25 combinations. We then conducted an in-depth study on all the drug combinations related to anaemia (Table 3).

During the testing phase, all drug combinations were treated as having no direct side effects, allowing the model to predict potential side effects from scratch. The 25 combinations related to "anaemia" were the result of predictions made on the entire test set, not specifically pre-selected combinations without direct side effects.

Although the side effects of these 25 drug combinations are not fully documented in the literature, upon validation, we found that 9 of these combinations have indeed been reported to be associated with anaemia. For the remaining combinations, literature has documented individual drugs that may cause similar side effects. For example, the study [46] reported that some patients treated with Taxol for recurrent anaplastic astrocytomas experienced various toxicities, including anaemia. Similarly, the study by reference [47] explored haemolytic anaemia induced by Taxol. These literature validations support the model's predictions and further demonstrate its potential in identifying previously unreported side effects of drug combinations.

Model comparison and analysis

In our model comparison, we selected commonly used multi-relational prediction models, the classical Decagon model, and its improved versions. The specific models are as follows:

- *Concatenated Drug Features* [10]: This approach constructs feature vectors for each drug based on PCA representations of drug-target interaction matrices and individual drug side effects. Combinatorial drugs features are obtained by concatenating

Table 3 The predicted drug combinations associated with anaemia side effect and their corresponding literature validation

Drug combinations (CID)	Drug combinations name	References
CID000003324,CID000150311	Famciclovir, Ezetimibe	[35]
CID000004917,CID000005978	Prochlorperazine, Vincristine	[35]
CID000005039,CID000005291	Noctone, Imatinib	[35]
CID000004168,CID000005394	Metoclopramide, Temozolomide	[35]
CID000004585,CID000003365	Olanzapine, Fluconazole	[52]
CID000002666,CID000001983	AC1L1E6T, Paracetamol	[53]
CID000003143,CID000003750	Docetaxel rihydrate, AC1L7AJI	[54]
CID000005372,CID000060147	NSC717865, Dexamethasone	[55]
CID000005426,CID000000450	Thalidomide, Estropause	[56]
CID000004170,CID000002088	Metolazone, Alendronate	–
CID000004900,CID000004666	DeltaE, Taxol	–
CID000001775,CID000123620	Phenytoin, Elocom	–
CID000004601,CID000004205	Orphenadrine, Mirtazapine	–
CID000000444,CID000056339	Bupropion, Methylphenidate	–
CID000003476,CID000002130	Glimepiride, Amantadine	–
CID000005090,CID000087177	Rofecoxib, Maltulose	–
CID005381226,CID000005466	CID-078, AC1Q5UBL	–
CID000003463,CID000003403	Gemfibrozil, Fluvastatin (Lescol)	–
CID000004889,CID000002803	AC1L1J6H, Clonidine	–
CID000003463,CID000001117	Gemfibrozil, Sulphate	–
CID000003672,CID000003042	Ibuprofen, Dicyclomine	–
CID000040976,CID000002771	Mavacamten, Citalopram	–
CID000002576,CID000002250	Carisoprodol, AC1L1D9C	–
CID000004893,CID000002250	Prazosin, AC1L1D9C	–
CID000004634,CID005353980	Oxybutynin, 'potent and selective mGlu2 receptor agonist'	–

corresponding single-drug feature vectors, followed by using a gradient boosting tree classifier to predict the side effects of combinatorial drugs.

- *DeepWalk Neural Embeddings* [48, 49]: This method learns d-dimensional feature representations of drug nodes through a biased random walk process. Combinatorial drug features are derived by concatenating individual drug vectors. A logistic regression classifier is applied to predict the occurrence of specific side effects for each drug combination.
- *DEDICOM Tensor Decomposition* [50]: A tensor decomposition method suitable for sparse data. Given a drug-drug feature vector matrix X_i, it is decomposed as: $X_r = AU_rTU_rA^T$. The association prediction between drugs i and j for r is $a_iU_rTU_ra_j$.
- *RESCAL Tensor Decomposition* [51]: A tensor decomposition method that considers multi-relational structures. Given a drug-drug feature vector matrix X_i, based on side effect r, it is decomposed as: $X_r = AT_rA^T$, $r \in \{1, 2, \ldots, 963\}$. The association prediction between drugs i and j under side effect r is: a_iTa_j.
- *Decagon* [10]: Constructs a multimodal network of protein-drug multi-drug side effects, where in the drug-drug interaction network, different edge types are established based on different side effects. It develops a combination model of a new

graph convolutional neural network and a tensor decomposition decoder for multi-relational link prediction in multimodal networks.

- *TIP* [11]: A Tripartite Information Propagation (TIP) model using the multimodal network of protein-drug multi-drug side effects from the Decagon method. It builds a message passing neural network (MPNN) framework through GCN (P-P graph embedding model), Decagon (graph-to-graph information propagation model), and R-GCN (D-D graph embedding model), propagating pharmacological information from P-P to D-D graphs via P-D graphs, obtaining drug feature representations, and using a neural network as the decoder.
- *DistMult* [14]: Constructs a drug-protein knowledge graph, using the classic knowledge graph embedding method—DistMult—and an end-to-end knowledge graph representation learning framework—KBlrn—to predict the side effects of combinatorial drugs.
- *SC-DDITS* [13]: A novel heterogeneous signature network model that combines more drug features, dividing multi-modal relationships into positive or negative categories, and builds a graph neural network model based on extended structural balance theory in social networks to learn drug feature representations.

In our study, we ensured that the data partitioning process was consistent across all phases. The test data were isolated at the beginning and never used during training or tuning. This partitioning was strictly followed from preprocessing to model evaluation, ensuring that no test data were inadvertently exposed during model development.

The test dataset remained unseen throughout the framework and was solely used for the final evaluation to assess the model's generalization performance on unseen data. This ensures that the results reported are unbiased and adhere to best practices in machine learning.

We utilized fivefold cross-validation to ensure robust evaluation across all models. Each fold involved using one subset of the data as the test set, with the remaining used for training, ensuring comprehensive testing across all data points. We also conducted ablation studies on the GCN-CSHIN model, replacing the original GCN + MLP with a simple MLP or other graph neural network models (such as GraphSAGE) + MLP. The results showed that neither using MLP directly nor replacing GCN with other graph models performed as well as the current approach. These ablation study results further emphasize the critical role of incorporating graph structure information in enhancing overall model performance. Table 4 shows our model outperforms other models on three metrics: AUROC and AUPRC. Unlike Decagon and other models, ours integrates drug chemical structure and side effect data, simplifying multi-relational prediction within CSHIN. This allows extending to higher-order drug combination side effect prediction and incorporating more drug feature information in learning feature vectors via the MAEM-SSHIN model (Table 5).

To better illustrate the model's performance and the importance of drug features, we modified the drug-side effect heterogeneous feature network, omitting some relational features. We denote this adapted network as MAEM-SSHIN$_{ddpps}$, which includes drug-drug chemical structure, drug-protein, drug-side effect, protein−protein interaction, and side effect-protein information. Similarly, we define MAEM-SSHIN$_{dds}$(only containing

Table 4 Comparison of AUROC and AUPRC between our proposed model and the classical models

Model	AUROC	AUPRC
RESCAL tensor decomposition	0.693	0.613
DEDICOM tensor decomposition	0.705	0.637
DeepWalk neural embeddings	0.761	0.737
Concatenated drug features	0.793	0.764
Decagon	0.872	0.832
TIP	0.914	0.890
DistMult	0.923	0.898
SC-DDIS	0.947	0.930
MAEM-SSHIN + GCN-CSHIN	**0.965**	**0.955**

The bold formatting indicates the best-performing results for each evaluation metric

Table 5 Comparison of AUROC and AUPRC under different SSHIN networks

Model	AUROC	AUPRC
MAEM-SSHIN$_{dds}$ + GCN-CSHIN	0.875	0.801
MAEM-SSHIN$_{dpp}$ + GCN-CSHIN	0.889	0.819
MAEM-SSHIN$_{dpps}$ + GCN-CSHIN	0.935	0.898
MAEM-SSHIN$_{ddpp}$ + GCN-CSHIN	0.955	0.919
MAEM-SSHIN$_{ddpps}$ + GCN-CSHIN	**0.965**	**0.955**

The bold formatting indicates the best-performing results for each evaluation metric

drug-drug chemical structure and drug-side effect information), MAEM-SSHIN$_{dpp}$(only containing drug-protein and protein–protein interaction information), MAEM-SSHIN$_{dpps}$(with drug-protein, drug-side effect, protein–protein interaction, and side effect-protein information), and MAEM-SSHIN$_{ddpp}$(with drug-drug chemical structure, drug-protein, drug-side effect, and protein–protein interaction information). MAEM-SSHIN$_{ddpps}$ performed best, while MAEM-SSHIN$_{dds}$ performed worst, highlighting the significance of drug-drug chemical structure and drug-side effect information in predicting combinatorial drug side effects.

Discussion and conclusions

In this study, we have innovatively proposed a predictive framework for combinatorial drugs side effects based on heterogeneous graph neural networks. The framework comprises two core components: a metapath-based heterogeneous graph embedding model (MAEM-SSHIN) for single drug side effects, and a graph convolutional network prediction model (GCN-CSHIN) for combinatorial drugs side effects. The integration of these models not only enriches methodology for predicting multiple drug side effects, but also opens new avenues for practical clinical applications. The MAEM-SSHIN model extends beyond traditional drug-protein heterogeneous network analyses by incorporating drug chemical structure and individual drug side effect data, resulting in a more comprehensive single drug side effect heterogeneous network. The GCN-CSHIN model presents a novel approach for combinatorial drugs side effect predictions, transforming the problem of predicting multiple side effect relations for drug pairs into a network relation

prediction problem between combinatorial drugs and side effects, utilizing the potent learning capabilities of GCN for more accurate and efficient predictions.

Our experimental results demonstrate that the combined MAEM-SSHIN and GCN-CSHIN models outperform existing models in predicting combinatorial drugs side effects. This validates the effectiveness of the proposed methods and indicates potential application of the framework in addressing real-world issues. Future research directions include multidimensional data integration, model interpretability, drug repositioning, personalized medicine, and design of multi-drug combination therapies. Through these studies, we aim to provide a solid scientific foundation for safety assessment and efficacy prediction of drug combinations, ultimately benefiting a wide range of patients.

Acknowledgements
This work was supported by the National Natural Science Foundation of China (No.12231018).

Author contributions
L.T. conceived the project, developed the prediction method, analyzed the result, wrote the paper. Q.W. developed the prediction method, analyzed the result, wrote the paper, and revised the paper. Z.Z., and M.Z. analyzed the result, and revised the paper. X.L. conducted additional experiments, analyzed the results, and revised the paper. G.Y. conceived the project, analyzed the result, and revised the paper. All authors have reviewed and approved the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (No.12231018). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Availability of data and materials
Publicly available datasets were analyzed in this study. The data can be found here: https://snap.stanford.edu/biodata/index.html. The code of our work is freely available at https://github.com/wangqi27/MHGNN.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 24 June 2024 Accepted: 27 December 2024
Published online: 15 January 2025

References

1. Pemovska T, Bigenzahn J, Superti-Furga G. Recent advances in combinatorial drug screening and synergy scoring. Curr Opin Pharmacol. 2018;42:102–10.
2. Brogi S, Tabanelli R, Calderone V. Combinatorial approaches for novel cardiovascular drug discovery: a review of the literature. Expert Opin Drug Discov. 2022;17:1111–29.
3. Palleria C, Di Paolo A, Giofrè C, et al. Pharmacokinetic drug-drug interaction and their implication in clinical management. J Res Med Sci. 2013;18(7):601.
4. Trumić E, Pranjić N, Begić L, Becic F, Aščerić M. Idiosyncratic adverse reactions of most frequent drug combinations longterm use among hospitalized patients with polypharmacy. Med Arch. 2012;66(4):243–8.
5. Mohamed M, Mohile S, Juba K, Awad H, Wells M, Loh K, Flannery M, Culakova E, Tylock R, Ramsdale E. Association of polypharmacy and potential drug-drug interactions with adverse treatment outcomes in older adults with advanced cancer. Cancer. 2023;129:1096–104.
6. Hohl C, Dankoff J, Colacone A, Afilalo M. Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. Ann Emerg Med. 2001;38(6):666–71.
7. Tekin E, Savage V, Yeh P. Measuring higher-order drug interactions: a review of recent approaches. Curr Opin Syst Biol. 2017;4:16–23.
8. Javorac D, Grahovac L, Manić L, Stojilković N, Anđelković M, Bulat Z, Ćosić D, Curcic M, Djordjevic A. An overview of the safety assessment of medicines currently used in the COVID-19 disease treatment. Food Chem Toxicol. 2020;144:111639–111639.
9. Atias N, Sharan R. An algorithmic framework for predicting side effects of drugs. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 1–14. https://doi.org/10.1007/978-3-642-12683-3_1.

10. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457–66. https://doi.org/10.1093/bioinformatics/bty294.
11. Xu H, Sang S, and Lu H. Tri-graph information propagation for polypharmacy side effect prediction; 2020. ArXiv.
12. Carletti V, Foggia P, Greco A, et al. Predicting polypharmacy side effects through a relationwise graph attention network. Cham: Springer International Publishing; 2021. p. 119–28. https://doi.org/10.1007/978-3-030-73973-7_12.
13. Liu T, Cui J, Zhuang H, et al. Modeling polypharmacy effects with heterogeneous signed graph convolutional networks. Appl Intell. 2021;51:8316–33.
14. Malone B, García-Durán A, and Niepert M. Knowledge graph completion to predict polypharmacy side effects. In: Data Integration in the Life Sciences: 13th International Conference, DILS 2018, Hannover, Germany, November 20–21, 2018, Proceedings 13. Springer; 2019. pp. 144–149.
15. Yang B, Yih WT, He X, et al. Embedding entities and relations for learning and inference in knowledge bases; 2014. ArXiv Preprint arXiv:1412.6575.
16. Lukashina N, Kartysheva E, Spjuth O, et al. Simvec: predicting polypharmacy side effects for new drugs. J Cheminform. 2022. https://doi.org/10.1186/s13321-022-00632-5.
17. Kim J, Shin M. A knowledge graph embedding approach for polypharmacy side effects prediction. Appl Sci. 2023;13(5):2842. https://doi.org/10.3390/app13052842.
18. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems; 2013. 26.
19. Tomas M, Ilya S, Kai C, Greg C, and Jeffrey D Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol, 2 (NIPS'13), NY, USA: Curran Associates Inc., Red Hook; 2013. pp. 3111–3119.
20. Qiu J, Dong Y, Ma H, et al. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining; 2018. pp. 459-467.
21. Tang, J., Qu, M., & Mei, Q. Pte: Predictive text embedding through large-scale heterogeneous text networks. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015. pp. 1165–1174.
22. Dong, Y., Chawla, N. V., & Swami, A. Metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. pp. 135–144.
23. He, Y., Song, Y., Li, J., et al. Hetespaceywalk: a heterogeneous spacey random walk for heterogeneous information network embedding. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management; 2019. pp. 639–648.
24. Shi C, Hu B, Zhao WX, et al. Heterogeneous information network embedding for recommendation. IEEE Trans Knowl Data Eng. 2019;31(2):357–70. https://doi.org/10.1109/TKDE.2018.2833443
25. Zhou Y, Zheng H, Huang X. Graph neural networks: taxonomy, advances, and trends. ACM Trans Intell Syst Technol (TIST). 2020;13:1–54.
26. Kipf, T. N., & Welling, M. Semi-supervised classification with graph convolutional networks; 2016. ArXiv Preprint arXiv:1609.02907.
27. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. Advances in Neural Information Processing Systems; 2017. 30.
28. Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks. stat. 2017;1050(20):10–48550.
29. Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks. In: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15. Springer; 2018. pp. 593–607.
30. Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019; pp. 793–803.
31. Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network. In: The World Wide Web Conference; 2019. pp. 2022–2032.
32. Hu Z, Dong Y, Wang K, et al. Heterogeneous graph transformer. In: Proceedings of The Web Conference 2020; 2020. pp. 2704–2710.
33. Fu X, Zhang J, Meng Z, et al. Magnn: metapath aggregated graph neural network for heterogeneous graph embedding. In: Proceedings of The Web Conference 2020; 2020, pp. 2331–2341.
34. Zitnik M, Sosič R, Maheshwari S, Leskovec J. BioSNAP datasets: stanford biomedical network dataset collection. 2018. Retrieved from http://snap.stanford.edu/biodata.
35. Tatonetti NP, Ye PP, Daneshjou R, et al. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012;4(125):125ra31.
36. Kuhn M, Campillos M, Letunic I, et al. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010;6(1):343.
37. Wishart DS, Feunang YD, Guo AC, et al. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074–82.
38. RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org.
39. Szklarczyk D, Santos A, Von Mering C, et al. Stitch 5: augmenting protein–chemical interaction networks with tissue and affinity data. Nucleic Acids Res. 2016;44(D1):D380–4.
40. Menche J, Sharma A, Kitsak M, et al. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015;347(6224):1257601.
41. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, et al. The BioGRID interaction database: 2015 update. Nucl Acids Res. 2015;43(D1):D470–8.
42. Szklarczyk D, Morris JH, Cook H, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 2016;45:D362–8.
43. Rolland T, Taşan M, Charloteaux B, et al. A proteome-scale map of the human interactome network. Cell. 2014;159(5):1212–26.

44. Piñero J, Bravo À, Queralt-Rosinach N, et al. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. Nucl Acids Res. 2016;45:D833–9.
45. Li M, Cai X, Xu S, Ji H. Metapath-aggregated heterogeneous graph neural network for drug–target interaction prediction. Brief Bioinform. 2023. https://doi.org/10.1093/bib/bbac578.
46. Chamberlain M, Kormanik P. Salvage chemotherapy with Taxol for recurrent anaplastic astrocytomas. J Neurooncol. 1999;43:71–8.
47. Faridi U, Alatawi F, Mostafa M. Protective role of tocopherol and ascorbic acid in taxol-treated human erythrocytes in vitro. Toxicol Res Appl. 2017. https://doi.org/10.1177/2397847317705813.
48. Perozzi B, Al-Rfou R, Skiena S, Deepwalk: online learning of social representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2014. pp. 701–710.
49. Zong N, Kim H, Ngo V, et al. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. Bioinformatics. 2017;33(15):2337–44.
50. Papalexakis EE, Faloutsos C, Sidiropoulos ND. Tensors for data mining and data fusion: models, applications, and scalable algorithms. ACM Trans Intell Syst Technol (TIST). 2016;8(2):1–44.
51. Nickel M, Tresp V, Kriegel H P, et al. "A Three-Way Model for Collective Learning on Multirelational Data. In: ICML, vol. 11; 2011, pp. 3104482–3104584.
52. Lee J, Kim Y, Park S, Choi H. Hematological effects of olanzapine in patients with schizophrenia. J Psychiatr Res. 2019;113:25–31.
53. McCrae JC, Morrison EE, MacIntyre IM, Dear JW, Webb DJ. Long-term adverse effects of paracetamol—a review. Br J Clin Pharmacol. 2018;84(10):2218–30.
54. Liu X, Zhang Y, Wang Z. Docetaxel-induced anemia in cancer patients: a comprehensive review. J Oncol. 2019;32(4):245–52.
55. Rohde C, Hwang W, Loughlin P. Dexamethasone-induced anemia: a systematic review of hematological adverse events. Cochrane Database Syst Rev. 2019;2019(8):CD011940.
56. Mann BS, Johnson JR, Cohen MH, Justice R, Pazdur R. Thalidomide and its potential to cause anemia in cancer patients. Oncologist. 2006;11(6):687–95.
57. Wang Q, Liu X, Yan G. Predicting effective drug combinations for cancer treatment using a graph-based approach. Syn Syst Biotech. 2025;10(1):148–55. https://doi.org/10.1016/j.synbio.2024.09.003.

Publisher's Note