



OPEN ACCESS

EDITED BY

Sajjad Gharaghani,
University of Tehran, Iran

REVIEWED BY

Xun Wang,
China University of Petroleum, China
Razieh Sheikhpour,
Ardakan University, Iran

*CORRESPONDENCE

Ping Lu,
✉ 2011990101@t.xmut.edu.cn
Yunbing Wu,
✉ wyb5820@fzu.edu.cn
Fan Yang,
✉ yang@xmu.edu.cn

RECEIVED 14 December 2023

ACCEPTED 30 January 2024

PUBLISHED 16 February 2024

CITATION

Pan D, Lu P, Wu Y, Kang L, Huang F, Lin K and Yang F (2024), Prediction of multiple types of drug interactions based on multi-scale fusion and dual-view fusion.
Front. Pharmacol. 15:1354540.
doi: 10.3389/fphar.2024.1354540

COPYRIGHT

© 2024 Pan, Lu, Wu, Kang, Huang, Lin and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prediction of multiple types of drug interactions based on multi-scale fusion and dual-view fusion

Dawei Pan¹, Ping Lu^{2*}, Yunbing Wu^{3*}, Liping Kang⁴,
Fengxin Huang¹, Kaibiao Lin¹ and Fan Yang^{5,6*}

¹School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China, ²School of Economics and Management, Xiamen University of Technology, Xiamen, China, ³College of Computer and Big Data, Fuzhou University, Fuzhou, China, ⁴Pasteur Institute, Soochow University, Suzhou, China, ⁵Shenzhen Research Institute of Xiamen University, Shenzhen, China, ⁶Department of Automation, Xiamen University, Xiamen, China

Potential drug-drug interactions (DDI) can lead to adverse drug reactions (ADR), and DDI prediction can help pharmacy researchers detect harmful DDI early. However, existing DDI prediction methods fall short in fully capturing drug information. They typically employ a single-view input, focusing solely on drug features or drug networks. Moreover, they rely exclusively on the final model layer for predictions, overlooking the nuanced information present across various network layers. To address these limitations, we propose a multi-scale dual-view fusion (MSDF) method for DDI prediction. More specifically, MSDF first constructs two views, topological and feature views of drugs, as model inputs. Then a graph convolutional neural network is used to extract the feature representations from each view. On top of that, a multi-scale fusion module integrates information across different graph convolutional layers to create comprehensive drug embeddings. The embeddings from the two views are summed as the final representation for classification. Experiments on two real-world datasets demonstrate that MSDF achieves higher accuracy than state-of-the-art methods, as the dual-view, multi-scale approach better captures drug characteristics.

KEYWORDS

drug drug interaction prediction, graph neural network, multi-scale fusion, graph features represent learning, multi-class classification

1 Introduction

The concurrent use of multiple medications is common, as combining drugs can reduce individual dosages and toxicity (Chou and Talalay, 1983). However, drug-drug interactions can alter potency and lead to adverse reactions (Lazarou et al., 1998). Moreover, as polypharmacy rises, so does the likelihood of adverse drug-drug interactions. For example, in the United States, approximately 74,000 emergency visits and 195,000 hospitalizations annually stem from antagonistic DDIs (Percha and Altman, 2013). Consequently, accurate identification of DDI is critical. Nevertheless, traditional *in vitro* and clinical diagnosis methods for DDI detection are often expensive and time-consuming (Huang et al., 2020b).

In recent years, deep learning has shown strong performance on prediction tasks (Dong et al., 2022; Meng et al., 2022). The deep learning technique, as one of the most advanced computational methods, has demonstrated a good performance on prediction tasks (Deng et al., 2020), bringing new solutions to the DDI prediction. When applying deep learning for DDI prediction, three main information types are now used: i) drug features like chemical substructures, enzymes, and targets (Lin S. et al., 2022b); ii) knowledge graphs containing DDI information (Wang et al., 2023); iii) DDI networks (Xiong et al., 2023).

While the above-mentioned methods have yielded good results, there are still two unresolved deficiencies. First, most graph neural network-based approaches (Feng et al., 2020) rely solely on the DDI graphs as inputs, which limits their ability to capture the full spectrum of drug interactions. This is because some DDIs may not yet be identified, leading to incomplete topological features. Second, current methods do not leverage multi-scale information that emerges during the process of information propagation. In graph neural networks, feature vectors from different network layers vary in dimensionality, semantics, and other informational scales. This variability is known as multi-scale information (Peng et al., 2021). Shallow network layers yield feature vectors with low-scale, more basic information, while deeper layers provide high-scale, more semantically rich information. By integrating information from these various scales, a more comprehensive understanding can be achieved. However, existing DDI prediction models predominantly utilize feature vectors from the final network layer, which only represent high-scale information. Although this information is less complex and rich in semantic features, it overlooks the detailed insights offered by low-scale information.

To solve the above problems, this paper introduces a new DDI prediction model, named MSDF, which employs multi-scale and two-view fusion techniques. First, alongside the DDI graph, we construct a feature similarity graph to supplement topological information. We base our approach on the assumption that drug pairs likely to interact exhibit high feature similarity. By connecting these similar nodes, the feature graph includes potential, yet undiscovered DDIs, thus addressing the issue of incomplete topological data in a standalone DDI graph. However, it is important to note that not all drugs with high feature similarity will interact. Hence, during the construction of the feature graph, some non-interacting drug pairs may be erroneously linked. This issue is mitigated by combining the feature graph with the more accurate topological data from the DDI graph, allowing for a correction of any inaccuracies in the feature graph. Overall, the complementary views improve predictions when fused. Second, we extract and utilize multi-scale insights across network layers from both views, which differ in dimensionality and semantics. Attention mechanisms then fuse these multi-scale representations to create comprehensive drug embeddings. In this way, the drug node representations integrate both localized and high-level insights for comprehensive learning, and are more favorable for the final DDI prediction task. Our contributions are summarized as follows.

- (1) We propose a new DDI prediction model: the MSDF. The model introduces a multi-scale fusion module, so that the node representations of the drug can contain structural and semantic information at different scales.
- (2) MSDF fuses multi-scale information from two views (DDI topology and feature views), and experimental results show that this approach facilitates model performance.
- (3) The model in this paper accomplishes the prediction of binary classification and multi-classification on both DeepDDI and DDIMDL datasets, and achieves better performance than the baseline method on both tasks, reflecting the sophistication and comprehensiveness of the model.

The remainder of the paper is organized as follows. Section 2 reviews related work in DDI prediction. Section 3 details the methodology behind MSDF. Section 4 presents the conducted experiments. Section 5 provides a conclusion and outlines the future directions.

2 Related work

Existing deep learning-based DDI prediction methods can be basically categorized into three types: methods based on drug feature information, methods that fuse knowledge graph, and methods based on topological features.

The drug feature similarity approach assumes that drugs with potential interactions share similar characteristics. This method employs deep learning techniques to distill drug features for classification tasks. Commonly used drug features include drug category (Zhang et al., 2022), chemical structure (Vilar et al., 2012), side effects (Tatonetti et al., 2011), and profile fingerprints (Vilar et al., 2013). Initially, these methods often focused on a single feature. For example, Ryu et al. (Ryu et al., 2018) developed the DeepDDI model, which uses the chemical substructures of drug pairs to calculate structural similarity profiles (SSPs). These SSPs are then fed into a Deep Neural Network (DNN) for predicting interactions. More recent research has highlighted the benefits of integrating multiple feature sources for improved prediction accuracy (Gottlieb et al., 2012; Cheng and Zhao, 2014; Yan et al., 2020). For instance, Lee et al. (Lee et al., 2019) combined SSPs, target genes, and gene ontology, encoding each feature separately using an AutoEncoder (AE) for classification. Deng et al. (Deng et al., 2020) utilized a dataset from Drugbank (Wishart et al., 2018) with 37,264 DDI events. They employed the StanfordNLP (Qi et al., 2018) tool to categorize these events into 65 types, using this classification as labels for model predictions. The model incorporated features like chemical substructures, enzymes, and drug targets, processed through a DNN. Furthermore, Yang et al. (Yang et al., 2023) built upon Deng et al.'s (Deng et al., 2020) dataset, employing Convolutional Neural Networks (CNNs) to extract drug feature vectors, resulting in enhanced prediction outcomes.

Methods based on drug feature similarity often overlook the crucial topological information represented as the DDI graph, which indicates the likelihood of interactions between drugs. However, incorporating this topological data can significantly enhance the accuracy of DDI predictions. Additionally, the choice and combination of drug features profoundly affect the model's performance. More specifically, Deng et al. (Deng et al., 2020) observed that models using a diverse range of features tend to yield better results than those relying on a single feature. Nevertheless, it's also important to note that simply adding more

features does not always lead to improved outcomes; in some cases, it may even diminish the model's effectiveness. Thus, careful selection and integration of multi-source features are essential. This process requires a deep understanding and extensive experience in the field, as the model designer must balance the quantity and quality of features to optimize the model's performance.

In the method of fusion knowledge graph, a knowledge graph is first constructed from the dataset. This knowledge graph includes not only drug entities but also various heterogeneous nodes, such as target and transporter nodes. By extracting information from the knowledge graph, a drug node can assimilate a richer array of information from these diverse nodes. Once the knowledge graph is established, it serves as an input for predictive models. The knowledge graph fusion method typically employs traditional knowledge graph embedding techniques, such as TransE (Bordes et al., 2013), ComplEx (Trouillon et al., 2016), RotatE (Sun et al., 2018) or utilize graph neural network approaches to learn the feature vectors of drug nodes within the graph. For instance, Yu et al. (Yu et al., 2022) developed the RANEDDI model, which uses the RotatE method for initializing drug embeddings in the knowledge graph and then feeds these embeddings into a relation-aware network for DDI prediction. Similarly, Hong et al. (Hong et al., 2022) introduced LaGAT, a graph attention model that employs the knowledge graph as its input. LaGAT generates various attention paths for drug entities by aggregating information from neighboring nodes, tailored to different drug pairs. Yao et al. (Yao J. et al., 2022) took a different approach by modeling drug pairs and their side effects within a knowledge graph, using nonlinear functions for semantic transfer, thereby demonstrating effective scalability in extensive knowledge graphs. While the knowledge graph fusion method successfully integrates a broader spectrum of heterogeneous information, it also introduces certain challenges. These include the potential for increased noise in the data and the complexity involved in constructing and managing the knowledge graph.

Drug topology feature-based approaches (Zitnik et al., 2018; Yue et al., 2020) usually use graph embedding techniques [e.g., Deepwalk (Perozzi et al., 2014), Node2Vec (Grover and Leskovec, 2016), SDNE (Wang et al., 2016), GCN (Kipf and Welling, 2023), GAT (Veličković et al., 2017), etc.] to extract topological features of drug nodes in DDI network for DDI prediction. Earlier studies, often preferred traditional graph embedding methods to extract topological features of drugs. For example, Park et al. (Park et al., 2015) used a random wandering algorithm and a restart algorithm to extract topological features of drugs on protein-protein interaction networks. However, these topological feature-based methods typically only depend on the topological feature. In other words, they tend to overlook the drug's own attributes, such as enzymes and targets, leading to incomplete analysis. Consequently, attribute graph-based methods have gained popularity. The attribute graph-based methods enhance a topology graph by adding node attributes, thereby incorporating the drug's own attributes into the analysis. This integration of additional information makes attribute graphs more effective for DDI prediction than traditional topological graphs. Graph neural network approaches, as opposed to traditional graph embedding methods like Deepwalk or Node2Vec, are better suited to handle attribute graphs. They can process both the topology and node attributes simultaneously,

extracting comprehensive drug node information through multi-layer aggregation and update operations. Recently, the trend has shifted towards combining graph neural networks with attribute graphs. A notable example is the work of Wang et al. (Wang et al., 2022) who created DDI increasing and decreasing graphs based on DDI types and employed GCN to learn drug representations using drug targets as node features. However, challenges persist in this evolving field. Wang et al. (Wang et al., 2020) noted that GCN's fusion mechanism sometimes fails to effectively integrate node features and topological information. Furthermore Yao et al. (Yao K. et al., 2022) pointed out that topological graphs might suffer from incomplete information, leading to suboptimal embeddings of learned nodes and negatively impacting downstream applications.

Therefore, in order to address the problems mentioned above, our model incorporates both a topological graph and a feature graph as inputs. By introducing a feature graph, the feature matrices of the nodes are propagated in both graphs simultaneously and the information from both graphs is complemented in order to improve the performance of the model. Moreover, in contrast to existing graph neural network-based models, which typically use only the output of the last network layer for classification tasks, we develop a multi-scale information fusion module. This module is designed to merge information from different scales, thereby enabling more accurate predictions.

3 Materials and methods

3.1 Overview

The modeling framework of MSDF is shown in Figure 1. It consists of three modules: dual view construction module, multi-scale fusion module, and model optimization module. The process begins with the dual view construction module, where two views of the drug information are generated: the topological graph (G_{adj}) and the feature graph (G_{knn}). These graphs, along with the drug feature matrix, are then fed into the multi-scale fusion module. In the multi-scale fusion module, the graphs and drug feature matrices undergo a series of operations including feature aggregation, updating, and cycling. These operations are performed using a GCN to derive multi-scale information from the drug nodes. To enhance the integration of information across different scales, the model incorporates an attention mechanism. This mechanism assigns weights, or "attention factors," to different scales based on their contribution to the final outcome. Each feature vector at a specific scale is then multiplied by its corresponding attention factor, creating a scaled vector representation. Subsequently, these scaled vector representations from different scales are concatenated using a splice operation. This results in a comprehensive final node representation for each drug node in that particular view. Once the multi-scale representations for both views are obtained, they are combined through a summation operation to form the final vector representations of the drug nodes. Finally, for the DDI classification task, the feature vectors of drug pairs are merged and inputted into a DNN. This is done based on the labeling information provided in the dataset, completing the classification process.

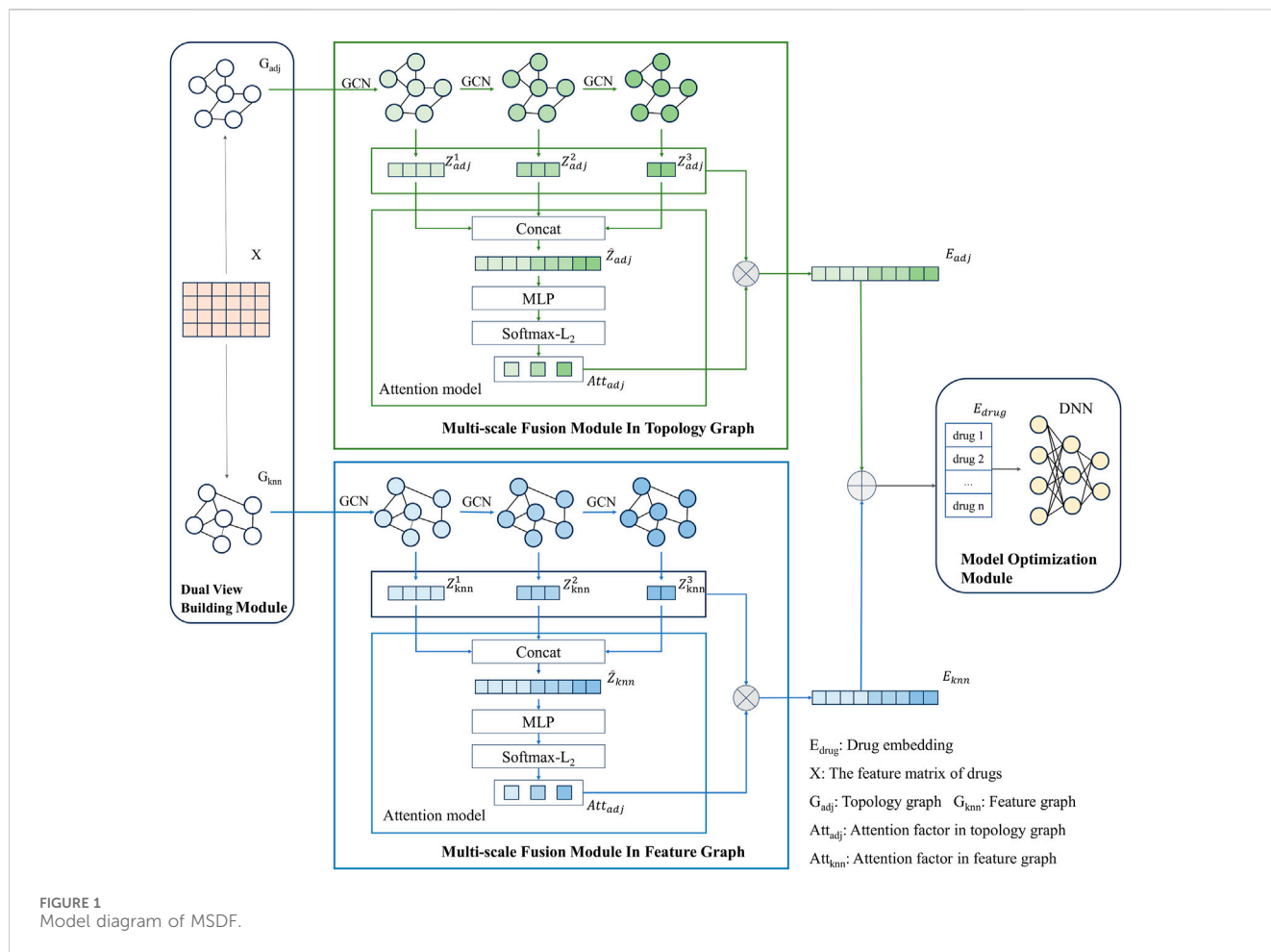


TABLE 1 Key symbols and definitions.

Notation	Definition and description
G	The input network
V	The node set
E	The edge set
X	Attribute feature of nodes
S_{ij}	The feature similarity of drug i, j vectors
$Z_y^{(l)}$	The embedding learned by GCN in view y at layer l
Att_y	Attention factor for different scales of information in view y
E_y	Embedding of nodes under y -view
E_{drug}	The unified embedding

3.2 Problem formulation

In this section, we define the problem using the notations provided in Table 1. We consider to a DDI network represented as $G = \{V, E, X\}$. Here, V denotes the drug nodes in the DDI network, E denotes the edges between the drug nodes, and X is the feature matrix associated with the drugs. The objective of the

MDSF accurately predict DDIs through both binary and multi-classification approaches. In more detail, the aim of binary classification is to determine whether a connection exists between any two given drug nodes, i.e., whether a DDI will occur. For multi-classification, the goal is to identify multi-classification is to the specific type of DDI that will occur between a pair of interacting drugs.

3.3 Data pre-processing

The features of a drug can be represented by a set of descriptors. In this representation, when a certain feature is present in a drug, the description of the drug in this feature is set to 1, and *vice versa* to 0. For instance, consider the drug Etodolac in the DDIMDL dataset, which lists 1162 potential targets. Etodolac is associated with three specific targets: P35354, P23219, and P19793. In its feature vector, the positions corresponding to these three targets are set to “1,” while the rest of the positions, representing the other 1159 targets, are set to “0.” While this descriptor-based approach accurately captures the features of a drug, it has a notable drawback: the resulting feature vectors are high-dimensional and predominantly filled with zeros. This leads to the “curse of dimensionality,” a phenomenon where the high

number of dimensions (features) negatively impacts the model's performance due to the sparsity of data. To mitigate this issue, it is essential to preprocess the drug features, aiming to reduce the dimensionality of the feature vectors.

The method of pre-processing in this paper is to compute the Jaccard similarity of drug features and then use the resulting Jaccard similarity matrix as the node features for each drug. This approach is applied to two datasets: the DDIMDL dataset extracted by Deng et al. (Deng et al., 2020) and the DeepDDI dataset extracted by Ryu et al. (Ryu et al., 2018). Detailed information about the datasets will be presented in the experimental section. For the DDIMDL dataset (Deng et al., 2020), this paper calculates the Jaccard similarity matrices of drug chemical substructures, enzymes, and targets, respectively, and splices the Jaccard similarity matrices of the three features as the feature vectors of the drugs in this dataset. For DeepDDI dataset (Ryu et al., 2018), since this dataset has only one feature of chemical substructures, this paper calculates the Jaccard similarity matrix of this feature as the feature vector of the drug in this dataset. The formula for calculating the Jaccard similarity is shown in Eq. 1. In Eq. 1, \mathbf{d}_i and \mathbf{d}_j are the initial feature vectors of drug i and drug j , \cap is the intersection operation, and \cup is the concatenation operation.

$$J(\mathbf{d}_i, \mathbf{d}_j) = \frac{|\mathbf{d}_i \cap \mathbf{d}_j|}{|\mathbf{d}_i \cup \mathbf{d}_j|} = \frac{|\mathbf{d}_i \cap \mathbf{d}_j|}{|\mathbf{d}_i| + |\mathbf{d}_j| - |\mathbf{d}_i \cap \mathbf{d}_j|} \quad (1)$$

3.4 Dual view building module

In order to represent the topological relationships between DDIs and the similarity of features between drugs, this paper constructs a topological graph of DDIs as well as a feature graph between drugs. These graphs are then utilized as inputs to the model for subsequent prediction tasks. The construction of the two views is described below.

3.4.1 Topology graph construction

The topology graph represents a kind of localized structural information between DDIs, denoted as G_{adj} . It is constructed by creating an adjacency matrix whose rows and columns are represented as drug nodes. In this matrix, an entry is set to "1" if there is a DDI between the corresponding drugs in a row and a column, indicating a connection. Conversely, an entry is marked as "0" if there is no interaction. Thus, the adjacency matrix can be represented as the topology graph of the DDI.

3.4.2 Feature graph construction

The feature graph represents the similarity of drug nodes in the feature space, denoted as G_{knn} . It is constructed by traversing all the drug nodes. During this process, we compute the feature similarity between each traversed drug node and all other nodes in the dataset. Upon completing this traversal, we obtain a similarity score for each pair of drug nodes. To construct the feature graph, also known as a K-Nearest Neighbors (KNN) graph, we connect each drug node to its top-K most similar nodes, forming edges based on these high similarity scores. Note that, the method used for calculating the similarity between drug nodes is the cosine similarity, as shown in Eq. 2.

$$s_{ij} = \frac{\mathbf{X}_i \cdot \mathbf{X}_j}{\|\mathbf{X}_i\| \|\mathbf{X}_j\|} \quad (2)$$

Where s_{ij} is the feature similarity score of drug i and drug j , \mathbf{X}_i is the feature vector of drug i , \mathbf{X}_j is the feature vector of drug j , and the symbol $\|\cdot\|$ denotes the Euclidean paradigm operation.

3.5 Multi-scale fusion module

After obtaining the feature matrix of the drug and the two graphical views - the topological graph and the feature graph -, our model proceeds to extract and fuse these features for the final DDI prediction task. To facilitate this, we have designed a multi-scale fusion module.

First, we use GCN to extract the drug node information in the network. For the topological graph, the output of GCN at each layer of the network is denoted as $Z_{adj}^{(l)}$, where l represents the layer number. Taking the topological graph as an example, the formulation of GCN can be represented by Eqs 3, 4.

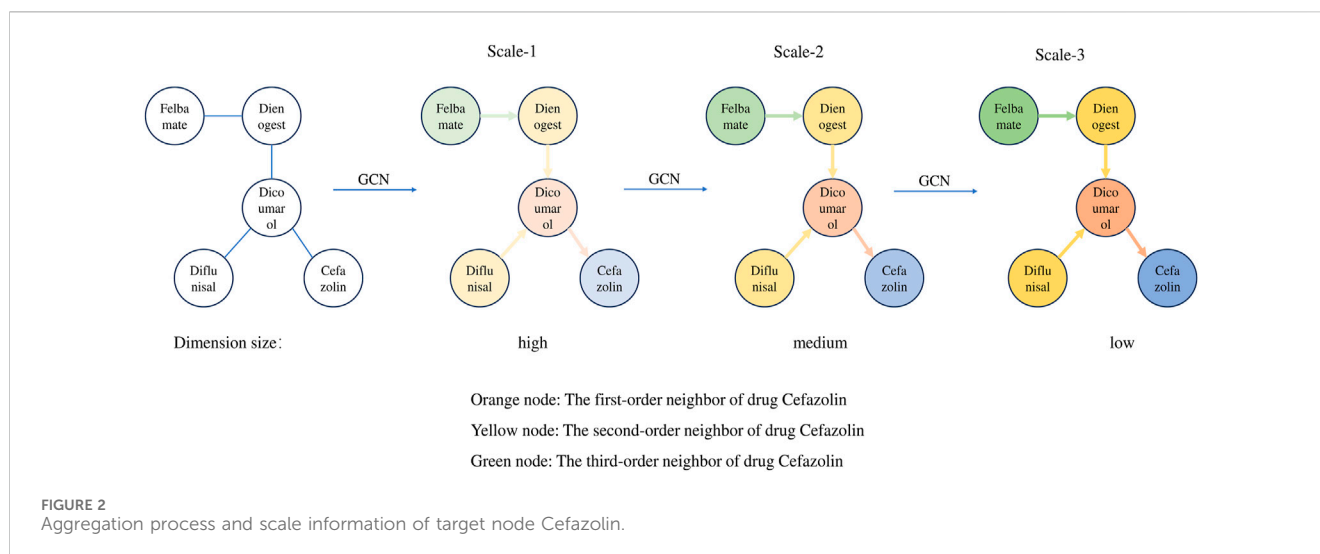
$$\tilde{A} = D^{-\frac{1}{2}} (A + I) D^{-\frac{1}{2}} \quad (3)$$

$$Z_{adj}^{(l)} = \sigma(\tilde{A} Z_{adj}^{(l-1)} W^{(l)}) \quad (4)$$

In Eq. 3, A denotes the adjacency matrix, I is the diagonal matrix of A , and D is the degree matrix of A . In Eq. 4, $Z_{adj}^{(l)}$ denotes the output of the l th layer in the GCN, $Z_{adj}^{(l-1)}$ denotes the output of the $(l-1)$ th layer in the GCN, $Z_{adj}^{(0)}$ is the node feature matrix X , $W^{(l)}$ denotes the learnable parameter matrix in the GCN, and σ is the activation function.

To illustrate the information aggregation process in GCNs, this paper selects five drugs from Drugbank to construct a DDI subgraph. In this subgraph, Cefazolin is treated as the target node, with its neighboring nodes of different orders represented by varying colors. This approach allows us to visually examine how target nodes in GCNs aggregate information from neighboring nodes at different layers. As can be seen from Figure 2, with a single GCN layer, the target node Cefazolin only aggregates information from its first-order neighbor Dicoumarol. However, as we increase the number of GCN layers, Cefazolin begins to incorporate information from higher-order neighbors. For instance, at two GCN layers, Cefazolin continues to aggregate information from its first-order neighbor Dicoumarol. But since Dicoumarol has already aggregated information from its own first-order neighbors (Dienogest and Diflunisal) in the first GCN layer, Cefazolin indirectly aggregates second-order neighbor information via the graph convolution in the second layer. Additionally, GCN serves a role in reducing the dimensionality of feature vectors. As the number of GCN layers increases, the feature vector dimensionality of the target node Cefazolin is progressively reduced from high to low. Therefore, we observe two key changes in node characteristics within the DDI network as the number of GCN layers increases:

- The target node obtains information about its higher-order neighbors as the number of GCN layers increases;
- The feature dimension of the target node decreases as the number of GCN layers increases.



With the two changes summarized above, it is known that as the number of GCN layers increases, the network generates information with different characteristics at different scales. In more detail, higher scales will contain information about higher-order neighbors. This results in feature vectors with lower dimensionality but richer in semantic content, although they may lose some detailed information. In contrast, at lower scales, the information is limited to lower-order neighbors, retaining more detailed data in the feature vectors due to their higher dimensionality, albeit with a slight increase in noise. Intuitively, feature vectors at higher scales are of superior quality as they encompass a broader scope of neighboring node information. However, unlike in other deep learning networks like CNNs, where the network depth can extend to dozens or even hundreds of layers, a GCN typically does not exceed 5 layers. The reason is the occurrence of “over-smoothing” as the number of layers increases. In higher GCN layers, the node representations tend to become overly similar, diminishing their ability to effectively perform subsequent downstream tasks. This phenomenon and its impact on node feature vectors are further corroborated by the experiments conducted in this paper.

As mentioned above, different scales of information in the GCN network have their own advantages and shortcomings for the final prediction results. By integrating information from various scales, we can maximize the benefits of each scale while compensating for their respective shortcomings. By doing so, our approach marks a departure from traditional practices in graph neural networks, where only the output from the network’s last layer is typically used as the final node vector. In contrast, our paper emphasizes the utility of outputs from each layer of the GCN, processing them collectively to enhance the accuracy of subsequent DDI predictions.

While multi-scale information contributes to the final prediction in our model, it is important to note that information from different scales has varying degrees of impact on the results. Drawing inspiration from Peng et al. (Peng et al., 2021), our paper incorporates an attention mechanism to more effectively integrate information across multiple scales. First of all, due to the varied dimensions of information at different scales, a direct approach to unifying this multi-scale information is not feasible. Instead, we

concatenate the feature vectors from different scales to create a unified input for the attention mechanism module. To be more specific, taking the topological graph as an example, we splice the feature vectors from different scales as demonstrated in Eq. 5.

$$\hat{Z}_{adj} = \left[Z_{adj}^1 \parallel Z_{adj}^2 \parallel \dots \parallel Z_{adj}^l \right] \quad (5)$$

In Eq. 5, Z_{adj}^l denotes the output of the l th layer of the GCN in the topological graph, \parallel denotes the splicing operation, and \hat{Z}_{adj} is the final vector after splicing of the multiscale feature vectors. After obtaining \hat{Z}_{adj} , we input \hat{Z}_{adj} into the fully-connected layer, the dimension of the fully-connected output layer is the number of layers l of the GCN. The output is the attention factor of the output of each layer of the GCN. After obtaining the attention factors, this paper uses the softmax function to perform the normalization operation and then obtains the final attention factors by l_2 regularization, which is calculated via Eq. 6.

$$Att_{adj} = l_2(\text{softmax}(\text{LeakyRelu}(\text{linear}(\hat{Z}_{adj})))) \quad (6)$$

In Eq. 6, Att_{adj} represents the matrix of attention factors in the topological graph, which can be expressed as $[a_{adj}^1 \parallel a_{adj}^2 \parallel \dots \parallel a_{adj}^l]$, and the data in each column of Att_{adj} represent the attention factors of the eigenvectors of the corresponding scales. By the same method, we can get the attention factor matrix in the feature graph as Att_{knn} . In this paper, the attention factors are multiplied with the corresponding scale information and spliced into the final embedding form, the embedding vector E_{adj} in the topology graph and the embedding vector E_{knn} in the feature graph can be expressed by Eqs 7, 8.

$$E_{adj} = a_{adj}^1 \cdot Z_{adj}^1 \parallel a_{adj}^2 \cdot Z_{adj}^2 \parallel \dots \parallel a_{adj}^l \cdot Z_{adj}^l \quad (7)$$

$$E_{knn} = a_{knn}^1 \cdot Z_{knn}^1 \parallel a_{knn}^2 \cdot Z_{knn}^2 \parallel \dots \parallel a_{knn}^l \cdot Z_{knn}^l \quad (8)$$

After obtaining the embedding vector E_{adj} of the drug for the topological graph and the embedding vector E_{knn} for the feature graph, we use summation to realize the fusion of the embedding vectors learned by the model in the two views, as shown in Eq. 9.

$$E_{drug} = E_{adj} + E_{knn} \quad (9)$$

TABLE 2 The way the three drug pairs were combined.

Combination method	Dimensionality	Description
Concatenation	2d	$\Phi(d_i, d_j) = \Phi(d_i) \parallel \Phi(d_j)$
Average	d	$\Phi(d_i, d_j) = 0.5 \times [\Phi(d_i) + \Phi(d_j)]$
Hadamard	d	$\Phi(d_i, d_j) = \Phi(d_i) \cdot \Phi(d_j)$

3.6 Model optimization module

Once obtaining the final embedding vector E_{drug} for each drug, our next step is to predict the DDI based on the labeling information in the dataset. To do this, we first combine the embeddings of two drugs that are likely to interact. This combination is achieved using a specific function, as outlined in Table 2. We express the combined drug pairs as shown in Eq. 10, where aggregation() represents the combination function, E_{drug_i} is the embedding vector of drug i, E_{drug_j} is the embedding vector of drug j, and the resultant drug pairs are denoted as DPs. For the binary classification task, the activation function of the last layer of the DNN is the sigmoid function, and its prediction can be expressed as Eq. 11. For the multi-classification task, the activation function of the last layer of the DNN is a softmax function, and its prediction result can be expressed as Eq. 12.

$$DPs = \text{aggregation}(E_{drug_i}, E_{drug_j}) \quad (10)$$

$$\hat{y}_{ij}^b = \text{sigmoid}(DNN(DPs)) \quad (11)$$

$$\hat{y}_{ij}^m = \text{softmax}(DNN(DPs)) \quad (12)$$

In this paper, we address both binary and multi-classification problems, necessitating the use of different loss functions to train the model effectively for each task. For binary classification, we employ the binary cross-entropy loss function. The formula for this function is detailed in Eq. 13. In contrast, for multi-classification tasks, the model uses the cross-entropy loss function, with its formula provided in Eq. 14.

$$Loss_b = - \sum_{(d_i, d_j) \in \epsilon} (y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})) \quad (13)$$

$$Loss_m = - \sum_{(d_i, d_j, r) \in \epsilon} \sum_{c=1}^R y_m^c \log(\hat{y}_m^{(c)}) \quad (14)$$

In Eq. 13 y_{ij} denotes the true label and \hat{y}_{ij} denotes the predicted outcome. In Eq. 14, c is the response type of the drug pair, y_m^c denotes the true labeling of the drug pair, and $\hat{y}_m^{(c)}$ denotes the labeling of the predicted outcome of the model.

4 Results and discussion

4.1 Datasets

In this paper, experiments were conducted using two datasets of different sizes, and the details of the datasets are shown in Table 3. The larger dataset is the one provided by DeepDDI, in which there are 1710 drugs and 192284 DDI events. These events are categorized into 86 different reaction types, which are used as labels for the

TABLE 3 Information on the dataset.

Dataset	Size	Drugs	Interaction	DDI type
DB1	Small	572	37624	65
DB2	Large	1710	192284	86

models. The drugs in this dataset are characterized by their medicinal chemical substructures. Within the DeepDDI dataset, we conduct experiments for both multi-classification and binary classification tasks. The smaller dataset, obtained from DDIMDL, consists of 572 drugs and 37,264 DDI events, which are classified into 65 reaction types. In the DDIMDL dataset, a total of molecular structures, enzymes, targets, and channels were collected as features, and it was verified that the model achieved the best results when using molecular structures, enzymes, and targets as the drug features according to the experiments of Deng et al. (Deng et al., 2020). Therefore, in this paper, these three features were chosen as the features of the model.

4.2 Baseline methods

In order to verify the effectiveness of the MSDF method proposed in this paper, we have chosen the most advanced DDI prediction algorithms for comparison. The methods chosen are those based on the drug's own feature: DNN, DeepDDI, and DDIMDL, and attribute graph based methods: SkipGNN, DM-DDI, MDFA, and AM-GCN. Among the methods based on attribute graphs, SkipGNN, MDFA, and AM-GCN are all multi-view methods. Note that, the purpose of introducing multi-view methods is to verify whether the method proposed in this paper is superior under the same multi-view condition. The type information of the baseline method is shown in Table 4, and the details of the baseline methods are shown below:

- DNN: This paper inputs the drug's own features directly into a DNN to perform both binary and multi-classification tasks for DDI prediction.
- DeepDDI (Ryu et al., 2018): DeepDDI takes the chemical substructures of drugs as features. Moreover, it employs Principal Component Analysis (PCA) to reduce the dimensionality of the feature matrix, which is then input into a DNN for prediction.
- DDIMDL (Deng et al., 2020): As part of a multimodal DDI prediction framework, DDIMDL uses chemical substructures, enzymes, and targets as drug features. It reduces feature sparsity by calculating Jaccard similarity for each feature,

TABLE 4 Information on the type of baseline methods.

Method	Feature type			Task	
	Self-feature	Network	Multi-view	Binary prediction	Multi-class prediction
DNN	√			√	√
DeepDDI	√			√	√
DDIMDL	√				√
SkipGNN	√	√	√	√	
AM-GCN	√	√	√	√	√
DM-DDI	√	√			√
MFDA	√	√	√		√
MSDF	√	√	√	√	√

followed by inputting these features into a DNN for prediction.

- SkipGNN (Huang et al., 2020a): SkipGNN is a multi-view graph neural network model, which takes into account the existence of jump similarity of nodes in a DDI network. It includes a two-hop neighbor graph in addition to the original topology graph and employs an iterative fusion method for information integration between the two views.
- AM-GCN (Wang et al., 2020): AM-GCN constructs a KNN graph based on node features and designs three channels for extracting information from topological and KNN graphs. The first two channels use GCNs with different parameters for extracting information in topological and KNN graphs. The third channel uses GCNs with shared parameters to extract common information in both graphs and learns the attention factors of the embedding vectors from the three channels using a self-attention mechanism for subsequent downstream tasks.
- DM-DDI (Kang et al., 2022): DM-DDI learns drug node features through AutoEncoder (AE) and topological features through GCN. It inputs outputs from each AE layer into corresponding GCN layers using a deep fusion strategy, which helps mitigate the over-smoothing problem in deep GCNs. The model then uses a self-attention mechanism to fuse topological and node features.
- MFDA (Lin K. et al., 2022a): MFDA constructs three views and uses a cross-fertilization strategy to fuse topological information from the graph and feature information of the drug itself. It introduces dual attention mechanisms at both the node-level and view-level, fusing the topological and feature information of nodes at the node-level, and combining embedding vectors from three views at the view-level for comprehensive drug embeddings.

4.3 Assessment of indicators

In order to evaluate the prediction ability of our model, different evaluation metrics are employed for binary and multi-classification tasks in our experiments. For binary classification, the model's performance is measured using three key indicators: accuracy rate, Area Under the Curve (AUC) of the Receiver Operating

Characteristic (ROC) curve, and the Area Under the Precision-Recall Curve (AUCPR).

In multi-classification experiments, a broader set of metrics is utilized, including accuracy, F1 score, precision, recall, AUC curve, and AUCPR curve. Among these metrics, we use macro-averaged measures for F1 score (F1_macro), precision (Pre_macro), and recall (Recall_macro). Conversely, for AUC and AUCPR, we employ micro-averaged metrics, denoted as AUC_micro and AUCPR_micro, respectively. It is noteworthy that in the context of multi-classification tasks, the micro-averaged precision (Pre_micro), recall (Recall_micro), and F1 score (F1_micro) are equivalent to the overall accuracy. Therefore, in our experiments, we opt for macro-averaged metrics for precision, recall, and F1 score to provide a more comprehensive evaluation.

4.4 Experimental settings

In this paper, we conducted 5-fold cross-validation, dividing the dataset into 5 equal parts, training on 4 parts and testing on the remaining part. This process was repeated 5 times, with each part used as the test set once. The final model performance was the average across the 5 folds. For the experiments of binary classification, in this paper, unconnected drug pairs are randomly selected as negative samples, and the ratio of positive and negative samples for the experiments is 1:1.

Regarding the model's parameter settings, we align the DNN parameters with those used in DDIMDL. This includes the incorporation of a batch normalization layer to accelerate convergence and a dropout layer with a rate of 0.3 to prevent overfitting. The activation function used here is the ReLU function. For the final layer, the softmax function is used in multi-classification tasks, while the sigmoid function is applied in binary classification tasks. Specifically for the DDIMDL dataset, the batch size is set to 1000, the epoch is set to 100, and the learning rate is set to 0.001. For the multi-classification task on the DeepDDI dataset, the batch size is set to 512, the epoch is set to 50, and the learning rate is set to 0.001. For the binary classification task on the DeepDDI dataset and the DDIMDL dataset, the batch size is set to 1000, the epoch is set to 50, and the learning rate is set to 0.0001.

TABLE 5 Prediction results for binary classification.

Dataset	Method	ACC	AUCPR	AUC
DeepDDI	DNN	0.869	0.932	0.943
	DeepDDI	0.878	0.942	0.946
	SkipGNN	0.947	0.989	0.989
	AM-GCN	0.928	0.982	0.982
	MSDF	0.959	0.993	0.993
DDIMDL	DNN	0.925	0.973	0.974
	DeepDDI	0.864	0.931	0.936
	SkipGNN	0.897	0.971	0.969
	AM-GCN	0.930	0.984	0.982
	MSDF	0.963	0.993	0.993

The best results are shown in bold.

4.5 Predicted results of DDI binary classification

The experimental results of our model and other baseline models for the task of binary classification are shown in Table 5. It is important to note that the original SkipGNN article, there is did not use the drug's own features as the node features, and the drug's node features are set as one-hot encoding in the original article. However, in our experiments, we have employed the drug's own features as node features in SkipGNN.

Table 5 demonstrates that in the DeepDDI dataset, methods relying solely on a drug's own features (DNN and DeepDDI) underperform compared to those utilizing attribute graphs across all metrics. Conversely, in the DDIMDL dataset, DNN's performance aligns with the attribute graph-based AMGNN, while DeepDDI lags. This discrepancy arises because the DeepDDI dataset benefits from topological information for a holistic drug representation, whereas the DDIMDL dataset is more reliant on intrinsic drug features. Notably, DeepDDI only processes chemical substructures, while DNN, incorporating three drug features, surpasses DeepDDI in the DDIMDL dataset. Across both datasets, our proposed method consistently yields optimal results, substantiating MSDF's efficacy in binary classification tasks.

4.6 Predicted results of DDI multi-classification

Since multi-classification DDI prediction provides a more valuable reference for medical practitioners, it is the focus of this paper. The performance comparison of our model with other state-of-the-art models is shown in Tables 6, 7.

Table 6 reveals the performance of various methods on the DDIMDL dataset. DNN, DDIMDL, and DeepDDI, which solely rely on the drug's own features without incorporating topological features of the drug network, exhibit lower performance across all metrics compared to methods that utilize graph embedding techniques. Among the methods based on graph embedding

TABLE 6 Comparison results of MSDF with other models in DDIMDL dataset.

Method	ACC	AUCPR	AUC	F1	Precision	Recall
DNN	0.880	0.913	0.996	0.722	0.805	0.703
DDIMDL	0.885	0.921	0.998	0.759	0.847	0.718
DeepDDI	0.837	0.890	0.996	0.685	0.728	0.661
AM-GCN	0.912	0.968	0.999	0.810	0.854	0.795
DM-DDI	0.908	0.964	0.999	0.852	0.879	0.839
MFDA	0.902	0.963	0.998	0.851	0.857	0.853
MSDF	0.943	0.981	0.999	0.863	0.889	0.854

The best results are shown in bold.

TABLE 7 Comparison results of MSDF with other models in DeepDDI dataset.

Method	ACC	AUCPR	AUC	F1	Precision	Recall
DNN	0.840	0.913	0.992	0.455	0.510	0.441
DDIMDL	0.939	0.968	0.999	0.903	0.918	0.900
DeepDDI	0.905	0.943	0.998	0.797	0.824	0.795
AM-GCN	0.924	0.979	0.999	0.880	0.904	0.875
DM-DDI	0.923	0.977	0.999	0.887	0.928	0.867
MFDA	0.940	0.986	0.999	0.908	0.928	0.900
MSDF	0.972	0.995	0.999	0.946	0.957	0.941

The best results are shown in bold.

techniques, AM-GCN performs poorly in general. This is because AM-GCN only uses the output of GCN as the embedding vector of the drug. However, GCN, in this context, can not effectively fuse topological and node features. In contrast, both DM-DDI and MFDA implement a cross-fertilization strategy, allowing for a more comprehensive integration of the drug's topological and node features. Therefore, they lead to superior performance in most metrics. Although both DM-DDI and MFDA achieve good results, the method MSDF proposed in this paper performs even better. The superior performance of MSDF is due to the introduction of the multi-scale fusion module, which effectively combines information from different scales to enhance model results. This hypothesis is further supported by subsequent ablation experiments.

While MSDF generally outperforms the baseline method, the evaluation of multi-classification models requires not only observing the overall classification effectiveness, but also assessing the classification performance of individual categories. Thus, in our analysis of the DDIMDL dataset, we comprehensively evaluate the predictive performance across all categories using two metrics: AUCPR and F1 score. These results are visually depicted in the radar chart of Figure 3. In the radar plot of Figure 3, we compare the category-wise prediction performance of MSDF against three methods that demonstrate overall superior performance: AM-GCN, DM-DDI, and MFDA. Each spoke of the radar chart represents a different label category, arranged in descending order of label frequency occurrence. For example, the first spoke (number 1) represents

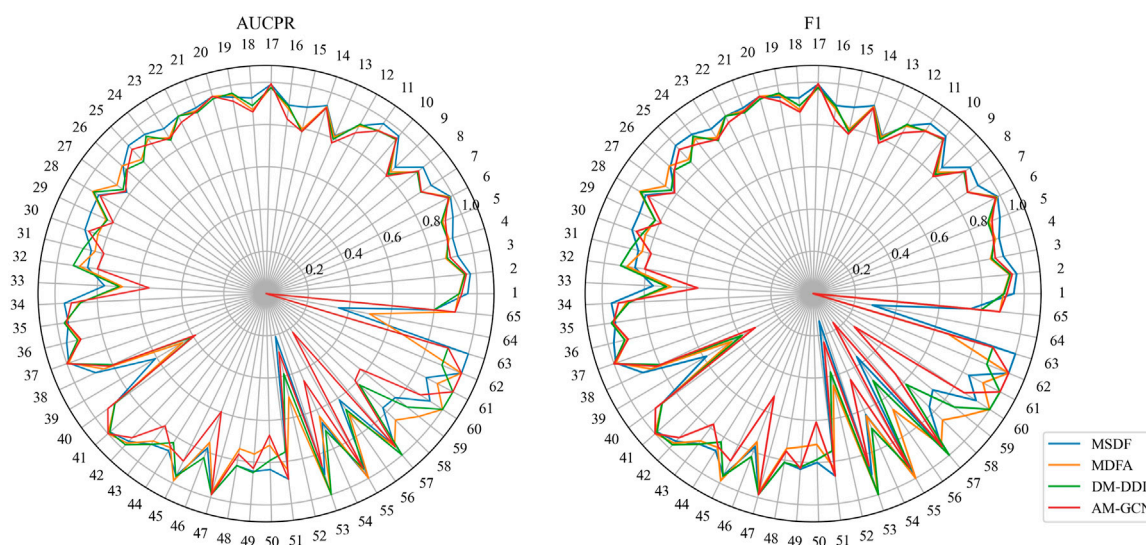


FIGURE 3 Comparison of AUCPR and F1 predicted for each event in the DDIMDL dataset.

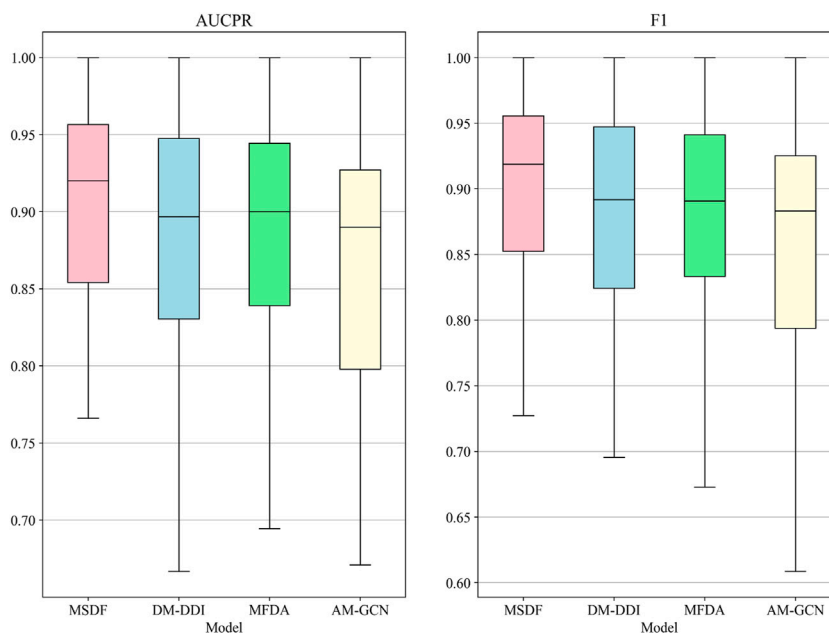
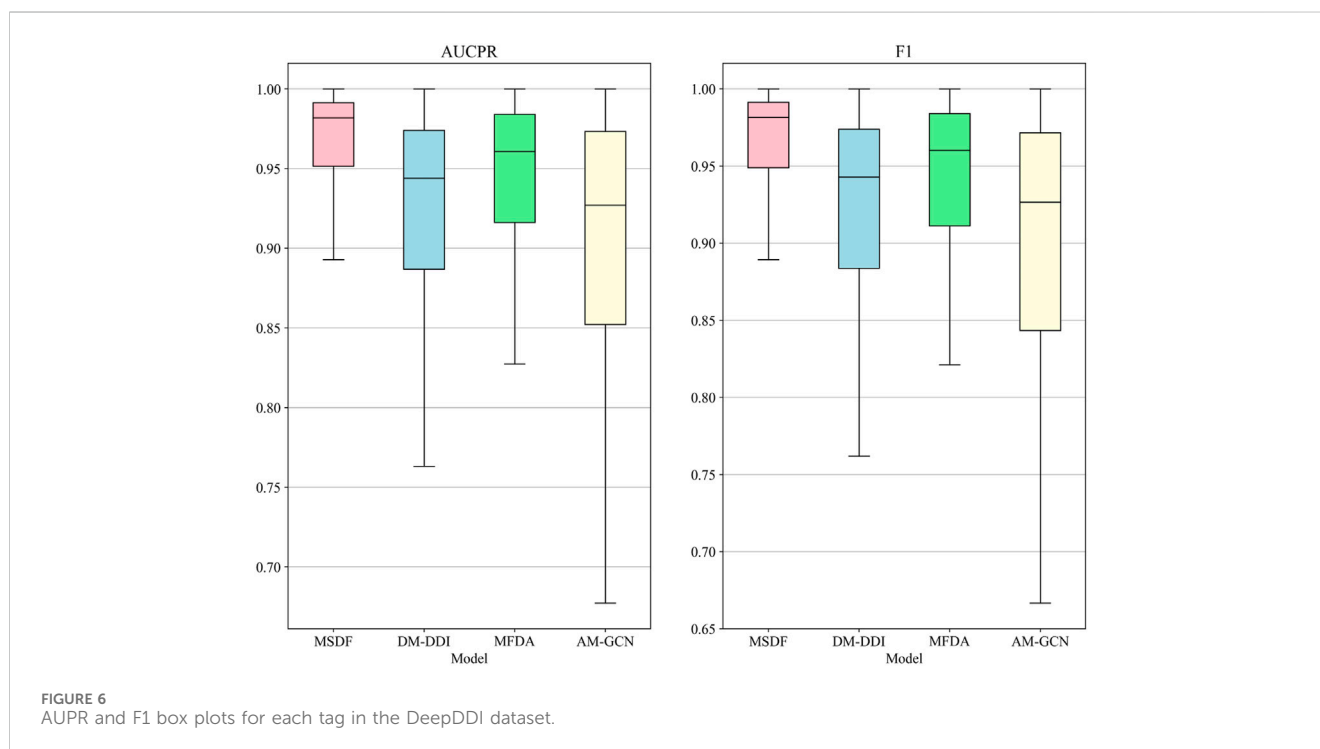
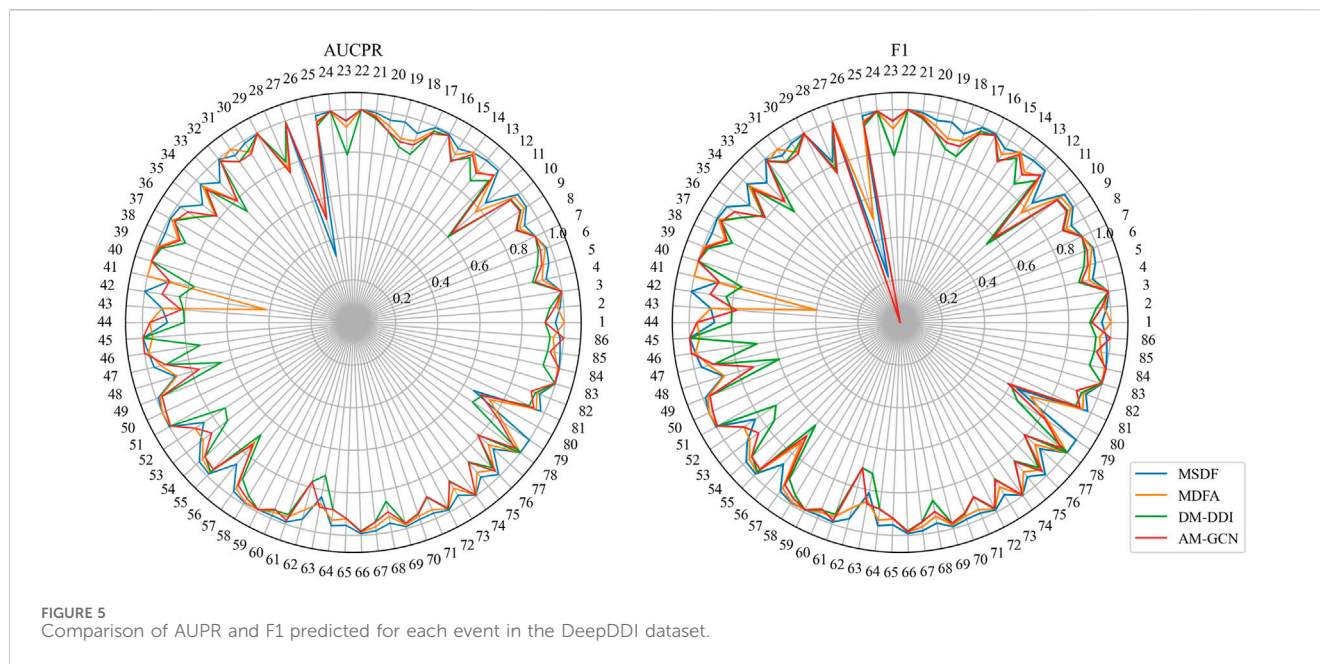


FIGURE 4 AUCPR and F1 box plots for each tag in the DDIMDL dataset.

the label with the highest frequency of occurrence, while the last spoke (number 65) corresponds to the label with the 65th frequency of occurrence.

In the radar chart, it can be seen that MSDF achieves better or comparable results to other methods in predicting labels up to number 51. This is particularly notable for labels up to number 18, where MSDF outperforms comparison methods. For instance, in predicting label number 39, other methods show significantly lower results, falling below 50% in both AUCPR and F1 scores. In contrast,

MSDF’s predictions for this label are considerably higher, with AUCPR at 60.9% and F1 at 58.5%. These observations suggest that it can be seen that MSDF excels in predicting labels that occur more frequently. However, for labels numbered 51 and beyond MSDF does not fetch better performance. An analysis of label distribution reveals that labels after number 51 appear fewer than 10 times in the dataset, indicating extremely limited data. Labels after number 60 appear only 5 times. According to Kang et al. (Kang et al., 2022), GCN cannot achieve better performance in category-imbalanced



data, which might explain MSDF's lower performance in these rarer label predictions. To further analyze the predictive performance of the models, box plots of different models on AUCPR and F1 under 65 classifications are plotted in this paper, as shown in Figure 4. From the box plots, it can be clearly seen that MSDF achieves better performance from the statistical level.

In the DeepDDI dataset, we obtain similar conclusions as in the DDIMDL dataset, i.e., our proposed method outperforms the other models in all metrics. However, by comparing the results of the two

datasets, we can see that on the DeepDDI dataset, the metrics of each model except DNN are significantly improved compared to those on the DDIMDL dataset, and this paper speculates that this is because as the dataset grows larger, the information input to the model is richer and therefore more effective. Interestingly, the performance of AM-GCN on DeepDDI does not lag behind DM-DDI, unlike in the DDIMDL dataset. This paper suggests that this difference may be attributed to the sparser network structure in DDIMDL dataset, where the drug's inherent features have a more significant impact on

TABLE 8 Results of ablation experiments.

Dataset	Method	ACC	AUCPR	AUC	F1	Precision	Recall
DDIMDL	Without multi-scale fusion	0.889	0.954	0.998	0.748	0.804	0.724
	Without topological view	0.931	0.975	0.998	0.838	0.869	0.824
	Without feature view	0.928	0.977	0.999	0.853	0.887	0.837
	MSDF full model	0.943	0.981	0.999	0.863	0.889	0.854
DeepDDI	Without multi-scale fusion	0.954	0.990	0.999	0.913	0.940	0.901
	Without topological view	0.962	0.992	0.999	0.919	0.935	0.912
	Without feature view	0.966	0.994	0.999	0.926	0.939	0.921
	MSDF full model	0.972	0.995	0.999	0.946	0.957	0.941

The best results are shown in bold.

prediction outcomes. In this paper, we also evaluate the classification performance of individual categories on the DeepDDI dataset and the results are shown in Figures 5, 6. Unlike the DDIMDL dataset, the labels in the DeepDDI dataset are not ordered by the frequency of label occurrence. In Figure 5, it can be seen that MSDF achieves optimal results in most categories. From the box plots in Figure 6, it can be seen that MSDF achieves the best performance in the DeepDDI dataset at the statistical level as well.

4.7 Ablation experiments

To verify the effectiveness of the dual-view and multi-scale fusion modules in MSDF, this paper conducts ablation experiments. The results of the ablation experiments are shown in Table 8. Among them, “without multi-scale fusion” refers to experiments where only the dual-view approach is implemented, omitting the multi-scale fusion module, “without topological graph” indicates that only the feature graph and multi-scale fusion module are introduced in the experiments, Conversely, “without feature graph” means that the experiments were conducted with just the topological graph and the multi-scale fusion module. Based on the results in Table 8, the following conclusions can be drawn: (1) The removal of the multi-scale fusion module resulted in the largest decrease in model effectiveness, so the fusion of information from different scales is crucial. (2) The performance of the model decreases with the removal of either topological or feature view, which indicates that the introduction of two views for prediction is effective. (3) The model can be optimized by introducing two views with the multi-scale fusion module.

In this paper, we not only conduct experiments to validate the overall functionality of our model but also specifically investigate how information from different scales affects DDI prediction results. For this purpose we individually select and utilize information from each scale for DDI prediction. In our representation, “Scale-*n*” denotes the scale used, where “*n*” indicates the scale number. For instance, “Scale-1” refers to information from the first GCN layer, while “MSDF” represents the complete model that fuses information from all three layers for prediction. The outcomes of these focused experiments are detailed in Table 9. From Table 9, the following conclusions can be drawn: (1) Relatively effective predictions can be

obtained from information at different scales; (2) The prediction effect is better for information at lower scales, indicating that as the number of GCN layers increases, the phenomenon of over-smoothing occurs, which reduces the prediction performance; (3) The best results can be obtained by fusing the three kinds of information, because the information is more comprehensive.

4.8 Parameter sensitivity analysis

In this section, we investigate the impact of three key parameters on our model’s performance: the number of GCN layers, the method of splicing drug pairs, and the number of neighbors in the feature graph. We employ AUPR and F1 scores as our primary evaluation metrics, as they are particularly relevant in multi-classification experiments. Figures 7, 8 illustrate the results of the study for these specific parameters in the DDIMDL and DeepDDI datasets, respectively.

In this paper, we first validate the effect of the number of GCN layers on the model. The experiment increases the number of GCN layers from 1 to 5, with the dimensions of each layer set as (256, 128, 64, 32, 16). After each increase in the layer count, we record the model’s performance. The results of the DDIMDL dataset are shown in Figure 7A, and results of the DeepDDI dataset are shown in Figure 8A. It can be seen that the model achieves the best performance when the number of GCN layers is 3. Beyond this, adding more layers does not further enhance the results, suggesting that three layers suffice for adequate information gathering.

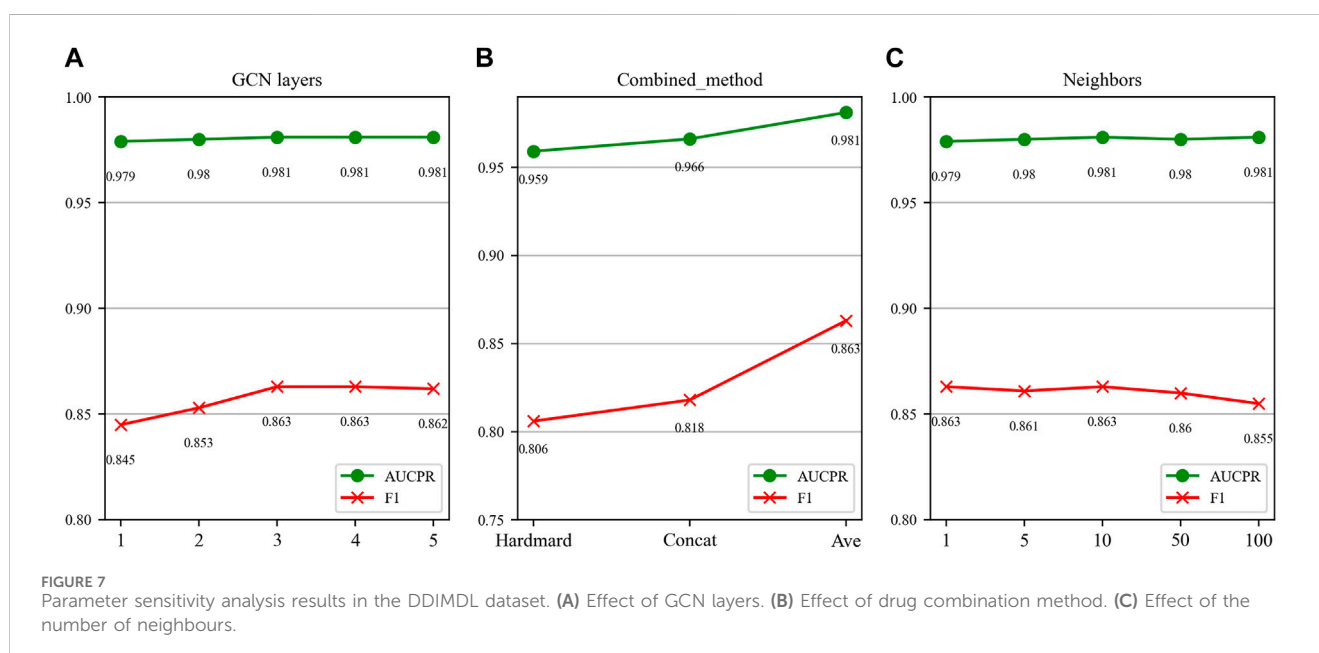
Next, we assess how different methods of splicing drug pairs affect the model. This experiment reveals varying impacts on performance, as detailed in Figures 7B, 8B. In the DDIMDL dataset, the combined “averaging” method proved to be the most effective, while in the DeepDDI dataset, the combined “concat” method proved to be the most effective.

Lastly, we explore the influence of the number of neighbors in feature graph construction. For this, this paper conduct experiments with varying number of neighbors set (1, 5, 10, 50, 100). As can be seen in Figures 7C, 8C, when the number of neighbors is less than 10, the parameter has a negligible impact on the results. However, when the number of neighbors reaches 100, the model’s performance significantly declines. This

TABLE 9 Effect of different scale information on the results.

Dataset	Method	ACC	AUCPR	AUC	F1	Precision	Recall
DDIMDL	Scale-1	0.939	0.979	0.999	0.845	0.872	0.833
	Scale-2	0.923	0.972	0.999	0.812	0.855	0.797
	Scale-3	0.889	0.954	0.998	0.748	0.804	0.724
	MSDF full model	0.943	0.981	0.999	0.863	0.889	0.854
DeepDDI	Scale-1	0.971	0.995	0.999	0.936	0.952	0.927
	Scale-2	0.964	0.993	0.999	0.925	0.944	0.917
	Scale-3	0.954	0.990	0.999	0.913	0.940	0.901
	MSDF full model	0.972	0.995	0.999	0.946	0.957	0.941

The best results are shown in bold.



decrease indicates that accumulating too many neighbors with low similarity introduces noise. Consequently, in our feature graph construction, we select the 10 nodes with the highest similarity to the target node as its neighbors.

4.9 Case study

To demonstrate the practical utility of the model, we trained it using known DDIs from the DeepDDI dataset and the DDIMDL dataset, respectively. After training, the model was used to predict the likelihood of DDIs for drug pairs not present in the dataset. We then selected the top 10 drug pairs with the highest probability of predicting DDI in each of the two datasets and searched for evidence of these interactions in the Drugbank database. The results of this case study are presented in Table 10 the method proposed in this paper achieves a good accuracy rate with 14 out of 20 predictions are validated in the Drugbank dataset. For example, an interaction between the drug

Digoxin and the drug Roflumilast is described as “Roflumilast may decrease the excretion rate of Digoxin which could result in a higher serum level.” Therefore, the model proposed in this paper is of practical utility. However, among the twenty DDIs with the highest predictive scores, six pairs: “Estrone-Drospirenone,” “Olopatadine-Quinine,” “Benzphetamine-Cholecalciferol,” “Acetylsalicylicacid-Bortezomib,” “Halothane-Azithromycin,” and “Fingolimod-Amrubicin” failed to find the corresponding evidence in drugbank. In this paper, we argue that this could be the result of an error in the model predictions, but it could also be a DDI that is currently unrecognized by people. Hence, extra attention is needed for these six pairs of DDIs.

5 Conclusion

In this paper, we propose a novel DDI prediction model, MSDF, which takes topology and feature graphs as inputs and fully integrates the

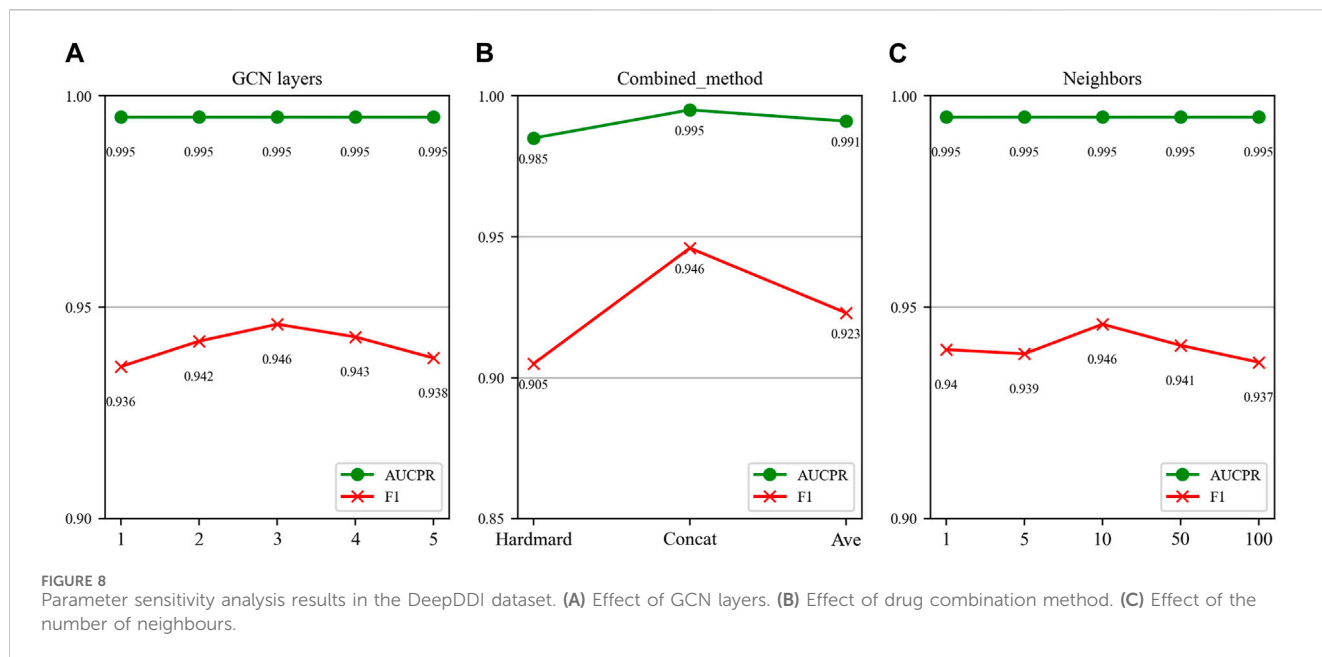


TABLE 10 The 20 DDIs with the highest predicted probabilities.

Dataset	Rank	DrugA	DrugB	Source	Description
DeepDDI	1	Pramipexole	Desvenlafaxine	Drugbank	Desvenlafaxine may increase the sedative activities of Pramipexole
	2	Warfarin	Probenecid	Drugbank	The metabolism of Warfarin can be increased when combined with Probenecid
	3	Digoxin	Roflumilast	Drugbank	Roflumilast may decrease the excretion rate of Digoxin which could result in a higher serum level
	4	Phenytoin	Nicergoline	Drugbank	The metabolism of Nicergoline can be decreased when combined with Phenytoin
	5	Estrone	Drospirenone	N.A.	N.A.
	6	Olopatadine	Quinine	N.A.	N.A.
	7	Triptorelin	Ivabradine	Drugbank	The risk or severity of QTc prolongation can be increased when Triptorelin is combined with Ivabradine
	8	Benzphetamine	Cholecalciferol	N.A.	N.A.
	9	Bosentan	Thioridazine	Drugbank	Thioridazine may decrease the antihypertensive activities of Bosentan
	10	Methysergide	Carbamazepine	Drugbank	The metabolism of Methysergide can be increased when combined with Carbamazepine
DDIMDL	1	Dronabinol	Desvenlafaxine	Drugbank	Dronabinol may increase the central nervous system depressant (CNS depressant) activities of Desvenlafaxine
	2	Dronabinol	Escitalopram	Drugbank	Dronabinol may increase the central nervous system depressant (CNS depressant) activities of Escitalopram
	3	Mifepristone	Ceritinib	Drugbank	The serum concentration of Ceritinib can be decreased when it is combined with Mifepristone
	4	Acetylsalicylicacid	Bortezomib	N.A.	N.A.
	5	Bortezomib	Etodolac	Drugbank	The metabolism of Etodolac can be decreased when combined with Bortezomib
	6	Citalopram	Losartan	Drugbank	Losartan may increase the QTc-prolonging activities of Citalopram
	7	Chlorpromazine	Doxorubicin	Drugbank	The metabolism of Doxorubicin can be increased when combined with Chlorpromazine
	8	Halothane	Azithromycin	N.A.	N.A.
	9	Fingolimod	Amrubicin	N.A.	N.A.
	10	Fluvoxamine	Amantadine	Drugbank	The risk or severity of serotonin syndrome can be increased when Fluvoxamine is combined with Amantadine

information from the outputs of the GCN layers through an attention mechanism to obtain multi-scale drug embedding vectors. The advantages of this method are that the limitation of utilising only

one view is alleviated by introducing a dual view, and the information of drug feature vectors containing multi-scale information is more comprehensive. However, the method proposed

in this paper still has a shortcoming: there is no obvious advantage in the prediction performance in the extremely rare category. However, on the whole, the method proposed in this paper achieves better results than the baseline method on datasets of different scales, reflecting the superiority of the MSDF model for DDI prediction. Meanwhile, there is still room for improvement in the model of this paper, firstly, the binary classification and multi-classification tasks are completed in the experiments of this paper, but the multi-label problem is not investigated. Secondly, this paper uses GCN to aggregate the information of drug nodes, but GCN suffers from the problem of poor performance in category unbalanced data and excessive smoothing, so the introduction of a novel graph neural network technique to aggregate the information of drug neighboring nodes is also an idea to improve the effectiveness of the model.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/YifanDengWHU/DDIMDL> <https://bitbucket.org/kaistsystemsbiology/deepddi/src/master/>.

Author contributions

DP: Writing–review and editing, Methodology, Writing–original draft. PL: Writing–review and editing, Conceptualization. YW: Writing–review and editing. LK: Writing–review and editing. FH: Data curation, Writing–review and editing. KL: Writing–review and editing. FY: Writing–review and editing.

References

- Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Adv. neural Inf. Process. Syst.* 26. doi:10.5555/2999792.2999923
- Cheng, F., and Zhao, Z. (2014). Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J. Am. Med. Inf. Assoc.* 21 (e2), e278–e286. doi:10.1136/amiainl-2013-002512
- Chou, T.-C., and Talalay, P. (1983). Analysis of combined drug effects: a new look at a very old problem. *Trends Pharmacol. Sci.* 4, 450–454. doi:10.1016/0165-6147(83)90490-x
- Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., and Liu, S. (2020). A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 36 (15), 4316–4322. doi:10.1093/bioinformatics/btaa501
- Dong, J., Zhao, M., Liu, Y., Su, Y., and Zeng, X. (2022). Deep learning in retrosynthesis planning: datasets, models and tools. *Briefings Bioinforma.* 23 (1), bbab391. doi:10.1093/bib/bbab391
- Feng, Y.-H., Zhang, S.-W., and Shi, J.-Y. (2020). DPDDI: a deep predictor for drug–drug interactions. *BMC Bioinforma.* 21 (1), 419–515. doi:10.1186/s12859-020-03724-x
- Gottlieb, A., Stein, G. Y., Oron, Y., Rupp, E., and Sharan, R. (2012). INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol. Syst. Biol.* 8 (1), 592. doi:10.1038/msb.2012.26
- Grover, A., and Leskovec, J. (2016). “node2vec: scalable feature learning for networks,” in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco California USA, August, 2016.
- Hong, Y., Luo, P., Jin, S., and Liu, X. (2022). LaGAT: link-aware graph attention network for drug–drug interaction prediction. *Bioinformatics* 38 (24), 5406–5412. doi:10.1093/bioinformatics/btac682
- Huang, K., Xiao, C., Glass, L. M., Zitnik, M., and Sun, J. (2020a). SkipGNN: predicting molecular interactions with skip-graph networks. *Sci. Rep.* 10 (1), 21092–21116. doi:10.1038/s41598-020-77766-9
- Huang, K., Xiao, C., Hoang, T., Glass, L., and Sun, J. (2020b). “Caster: predicting drug interactions with chemical substructure representation,” in Proceedings of the AAAI conference on artificial intelligence, Washington DC, USA, February, 2020.
- Kang, L. P., Lin, K. B., Lu, P., Yang, F., and Chen, J. P. (2022). Multitype drug interaction prediction based on the deep fusion of drug features and topological relationships. *Plos one* 17 (8), e0273764. doi:10.1371/journal.pone.0273764
- Kipf, T. N., and Welling, M. (2023). Semi-Supervised classification with graph convolutional networks. <https://arxiv.org/abs/1609.02907>.
- Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama* 279 (15), 1200–1205. doi:10.1001/jama.279.15.1200
- Lee, G., Park, C., and Ahn, J. (2019). Novel deep learning model for more accurate prediction of drug–drug interaction effects. *BMC Bioinforma.* 20 (1), 415–418. doi:10.1186/s12859-019-3013-0
- Lin, K., Kang, L., Yang, F., Lu, P., and Lu, J. (2022a). MFDA: multiview fusion based on dual-level attention for drug interaction prediction. *Front. Pharmacol.* 13, 1021329. doi:10.3389/fphar.2022.1021329
- Lin, S., Chen, W., Chen, G., Zhou, S., Wei, D. Q., and Xiong, Y. (2022b). MDDI-SCL: predicting multi-type drug–drug interactions via supervised contrastive learning. *J. Cheminform* 14 (1), 81. doi:10.1186/s13321-022-00659-8
- Meng, Y., Lu, C., Jin, M., Xu, J., Zeng, X., and Yang, J. (2022). A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings Bioinforma.* 23 (2), bbab581. doi:10.1093/bib/bbab581
- Park, K., Kim, D., Ha, S., and Lee, D. (2015). Predicting pharmacodynamic drug–drug interactions through signaling propagation interference on protein–protein interaction networks. *Plos one* 10 (10), e0140816. doi:10.1371/journal.pone.0140816

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is partially supported by the National Natural Science Foundation of China under Grant No. 62173282, the Natural Science Foundation of Xiamen City, China under Grant No. 3502Z20227180, the Xiamen Science and Technology Bureau Production Project (No. 2023CXYO415).

Acknowledgments

We acknowledge reviewers for the valuable comments on the original manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

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

- Peng, Z., Liu, H., Jia, Y., and Hou, J. (2021). "Attention-driven graph clustering network," in Proceedings of the 29th ACM SIGKDD international conference on multimedia, Virtual Event China, October, 2021.
- Percha, B., and Altman, R. B. (2013). Informatics confronts drug–drug interactions. *Trends Pharmacol. Sci.* 34 (3), 178–184. doi:10.1016/j.tips.2013.01.006
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). "Deepwalk: online learning of social representations," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York New York USA, August, 2014.
- Qi, P., Dozat, T., Zhang, Y., and Manning, C. D. (2018). Universal dependency parsing from scratch. *CoNLL* 2018, 160. doi:10.18653/v1/K18-2016
- Ryu, J. Y., Kim, H. U., and Lee, S. Y. (2018). Deep learning improves prediction of drug–drug and drug–food interactions. *Proc. Natl. Acad. Sci.* 115 (18), E4304–E4311. doi:10.1073/pnas.1803294115
- Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. (2018). "RotatE: knowledge graph embedding by relational rotation in complex space," in Paper presented at the International Conference on Learning Representations, Vancouver Canada, April, 2018.
- Tatonetti, N. P., Denny, J., Murphy, S., Fernald, G., Krishnan, G., Castro, V., et al. (2011). Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin. Pharmacol. Ther.* 90 (1), 133–142. doi:10.1038/clpt.2011.83
- Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. (2016). "Complex embeddings for simple link prediction," in Paper presented at the International conference on machine learning, New York City, NY, USA, June, 2016.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2017). Graph attention networks. <https://arxiv.org/abs/1710.10903>.
- Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., and Friedman, C. (2012). Drug–drug interaction through molecular structure similarity analysis. *J. Am. Med. Inf. Assoc.* 19 (6), 1066–1074. doi:10.1136/amiajnl-2012-000935
- Vilar, S., Uriarte, E., Santana, L., Tatonetti, N. P., and Friedman, C. (2013). Detection of drug–drug interactions by modeling interaction profile fingerprints. *Plos one* 8 (3), e58321. doi:10.1371/journal.pone.0058321
- Wang, D., Cui, P., and Zhu, W. (2016). "Structural deep network embedding," in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco California USA, August, 2016.
- Wang, F., Lei, X., Liao, B., and Wu, F.-X. (2022). Predicting drug–drug interactions by graph convolutional network with multi-kernel. *Briefings Bioinforma.* 23 (1), bbab511. doi:10.1093/bib/bbab511
- Wang, J., Zhang, S., Li, R., Chen, G., Yan, S., and Ma, L. (2023). Multi-view feature representation and fusion for drug–drug interactions prediction. *BMC Bioinforma.* 24 (1), 93. doi:10.1186/s12859-023-05212-4
- Wang, X., Zhu, M., Bo, D., Cui, P., Shi, C., and Pei, J. (2020). AM-GCN: adaptive multi-channel graph convolutional networks. <https://arxiv.org/abs/2007.02265>.
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids Res.* 46 (D1), D1074–D1082. doi:10.1093/nar/gkx1037
- Xiong, Z., Liu, S., Huang, F., Wang, Z., Liu, X., Zhang, Z., et al. (2023). "Multi-Relational contrastive learning graph neural network for drug–drug interaction event prediction," in Proceedings of the AAAI Conference on Artificial Intelligence, Ljubljana Slovenia, April, 2023.
- Yan, C., Duan, G., Zhang, Y., Wu, F.-X., Pan, Y., and Wang, J. (2020). Predicting drug–drug interactions based on integrated similarity and semi-supervised learning. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 19 (1), 168–179. doi:10.1109/TCBB.2020.2988018
- Yang, Z., Tong, K., Jin, S., Wang, S., Yang, C., and Jiang, F. (2023). CNN-Siam: multimodal siamese CNN-based deep learning approach for drug–drug interaction prediction. *BMC Bioinforma.* 24 (1), 110. doi:10.1186/s12859-023-05242-y
- Yao, J., Sun, W., Jian, Z., Wu, Q., and Wang, X. (2022a). Effective knowledge graph embeddings based on multidirectional semantics relations for polypharmacy side effects prediction. *Bioinformatics* 38 (8), 2315–2322. doi:10.1093/bioinformatics/btac094
- Yao, K., Liang, J., Liang, J., Li, M., and Cao, F. (2022b). Multi-view graph convolutional networks with attention mechanism. *Artif. Intell.* 307, 103708. doi:10.1016/j.artint.2022.103708
- Yu, H., Dong, W., and Shi, J. (2022). RANEDDI: relation-aware network embedding for drug–drug interaction prediction. *Inf. Sci.* 582, 167–180. doi:10.1016/j.ins.2021.09.008
- Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., et al. (2020). Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics* 36 (4), 1241–1251. doi:10.1093/bioinformatics/btz718
- Zhang, C., Lu, Y., and Zang, T. (2022). CNN-DDI: a learning-based method for predicting drug–drug interactions using convolution neural networks. *BMC Bioinforma.* 23 (Suppl. 1), 88. doi:10.1186/s12859-022-04612-2
- Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34 (13), i457–i466. doi:10.1093/bioinformatics/bty294