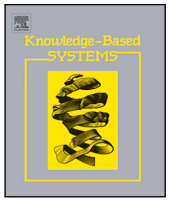
Knowledge-Based Systems 284 (2024) 111329

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



AMDGT: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction

Junkai Liu [a,b], Shixuan Guan [b,c], Quan Zou [b], Hongjie Wu [a,b,*], Prayag Tiwari [d], Yijie Ding [b]

[a] School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, 215009, China
[b] Yangtze Delta Region Institute(Quzhou), University of Electronic Science and Technology of China, Quzhou, 324003, China
[c] School of Science and Technology, University of Tsukuba, Tsukuba, 3058577, Japan
[d] School of Information Technology, Halmstad University, Sweden

ARTICLE INFO

Keywords:
Drug–disease associations
Multimodal learning
Graph transformer
Attention mechanism
Drug repositioning

ABSTRACT

Identification of new indications for existing drugs is crucial through the various stages of drug discovery. Computational methods are valuable in establishing meaningful associations between drugs and diseases. However, most methods predict the drug–disease associations based solely on similarity data, neglecting valuable biological and chemical information. These methods often use basic concatenation to integrate information from different modalities, limiting their ability to capture features from a comprehensive and in-depth perspective. Therefore, a novel multimodal framework called AMDGT was proposed to predict new drug associations based on dual-graph transformer modules. By combining similarity data and complex biochemical information, AMDGT understands the multimodal feature fusion of drugs and diseases effectively and comprehensively with an attention-aware modality interaction architecture. Extensive experimental results indicate that AMDGT surpasses state-of-the-art methods in real-world datasets. Moreover, case and molecular docking studies demonstrated that AMDGT is an effective tool for drug repositioning. Our code is available at GitHub: https://github.com/JK-Liu7/AMDGT.

1. Introduction

Conventional drug discovery is laborious and time-consuming, predominantly because of its inherently high risk of failure [1]. The journey from drug discovery to clinical utilisation typically spans over a decade and incurs costs ranging from $ 500 million to $ 2 billion [2]. Nevertheless, the clinical approval rate for novel drugs remains low, with less than 10% regulatory clearance [3]. Against the backdrop of rapid advancements in artificial intelligence, drug repositioning (DR), also known as drug repurposing, has emerged as an alternative and complementary strategy for identifying novel indications for drugs, offering advantages in expediting drug discovery and reducing extensive efforts [4]. Recently, various computation-based DR methods adopt a biomedical data-driven approach to discover novel drug–disease associations (DDAs) [5]. These methods typically extract relevant features from biochemical and medical data pertaining to drugs and diseases that are subsequently incorporated into well-established classification models to facilitate DDA prediction. Generally, they are divided into two categories: machine learning (ML)-based [6–10] and deep learning (DL)-based methods.

ML-based DDA prediction methods employ diverse ML algorithms and techniques such as matrix factorisation [11–14], support vector machines [15,16], multiple kernel learning [17–19], and neural networks [20–23] to predict novel DDAs according to the extracted embeddings. However, ML-based methods rely on shallow features and have a limited capacity to represent abstract and high-level features of drugs and diseases. To address this limitation, DL-based DDA prediction models that harness powerful representation learning capabilities have been proposed [24–30]. For example, deepDR [27] integrates various association matrices, employs a multimodal autoencoder to obtain the feature embeddings of entities effectively, and leverages a variational autoencoder to identify new diseases. HINGRL [28] and DDAGDL [29] construct heterogeneous information networks and learn feature representations with biological information using graph representation learning methods. Yu et al. [31] proposed layer attention graph convolutional networks (LAGCN), which integrates three types of similarity networks and employs graph convolutional networks (GCN) to learn embeddings. Sigmoid Kernel and Convolutional Neural Network (SKCNN) [32] was proposed as a convolutional neural network

* Corresponding author.
E-mail addresses: 2111041014@post.usts.edu.cn (J. Liu), s2336025@u.tsukuba.ac.jp (S. Guan), zouquan@nclab.net (Q. Zou), hongjiewu@mail.usts.edu.cn (H. Wu), prayag.tiwari@ieee.org (P. Tiwari), wuxi_dyj@csj.uestc.edu.cn (Y. Ding).

https://doi.org/10.1016/j.knosys.2023.111329
Received 10 August 2023; Received in revised form 10 November 2023; Accepted 20 December 2023
Available online 28 December 2023

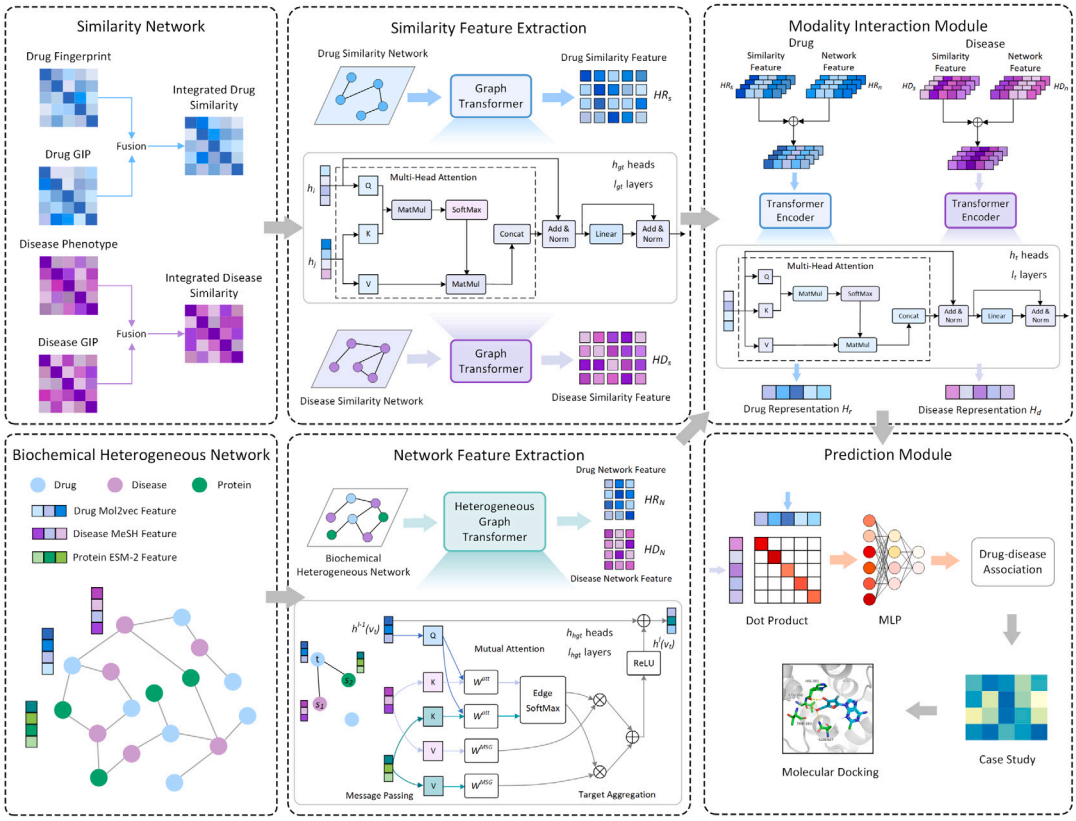


Fig. 1. The overall architecture of AMDGT.

model with a sigmoid kernel to extract drug and disease features for DDA prediction. Additionally, knowledge graphs have been used in this domain to explicitly model multiple pieces of information between different entities and mine biomedical knowledge [33,34]. Moreover, Zhao et al. [23] adopted a GCN- and meta-path-based strategy to learn lower- and higher-order biological attributes, respectively. Multi-similarities graph convolutional autoencoder (MSGCA) [35] was used to construct multiple drug and disease similarities, which were then integrated using a graph autoencoder with an attention mechanism to predict DDAs. Wang et al. [36] utilised a novel method to fuse various similarity networks with a multi-view feature projection method to obtain DDA scores.

Although these methods have proven their efficacy, their success has primarily been demonstrated on curated biomedical data and advanced DL technology [37–41]. Nevertheless, only a small number of methods simultaneously consider both drug and disease similarities and biochemical properties, hindering the model learning method from providing a high-quality representation with a comprehensive view [42,43]. In addition, deficiencies of generalisation may develop when processing unseen drugs and diseases with novel scaffolds. Currently, multimodal data are commonly used in DL-based models that exhibit satisfactory and competitive performances in bioinformatics [30,44–52]. For instance, Zhang et al. [30] proposed M2Seq2Seq, a multimodal multitask learning model based on an encoder–decoder framework with a novel attention mechanism to understand and recognise the corresponding emotion in a conversation. Prayag et al. [44] proposed QFNN model by extending [30] based on fuzzy logic and quantum neural network for multimodal sarcasm and sentiment detection. Wang et al. [49] employed graph representation approaches with pretrained embeddings to predict the interactions between drugs and proteins. Thus, we aimed to capture multimodal feature embeddings from different networks, empowering the model to learn from a comprehensive and in-depth perspective. Despite integrating embedded representations from different entities or modalities [53–56], the concatenation operation may limit capturing high-level associations and internal interactions between instances. This limitation may affect performance; however, research regarding this outcome is lacking.

To address these challenges, a novel computational method known as AMDGT was proposed, exploiting multi-modal networks and dual graph transformer frameworks to predict DDAs. First, we constructed integrated similarity networks for drugs and diseases using multisource data. Subsequently, a biochemical heterogeneous network, including drugs, diseases, and protein entities, was generated with numerous pieces of biological and chemical information. Furthermore, dual advanced graph neural network models, that is, homogeneous and heterogeneous graph transformers, are used to learn the representations from similarity data and association networks, respectively. These models have several advantages. First, the architecture of the two models facilitates the capture of common and specific patterns of node and connection information. Second, an improved and fresh attention mechanism is crucial for estimating the importance between node entities of identical and different types, which can be used to discriminate between the contributions of various types of nodes. Additionally, similarity data and biochemical information can be acquired from both neighbouring and global topological aspects simultaneously, which is beneficial for generalising our model. Subsequently, an AttentionFusion module was designed to model the relationship and interaction of multimodal networks, through which we obtained the final drug and disease representations with substantial modality and view properties. Finally, AMDGT employed a dot-product decoder with a multilayer perception (MLP) model to predict the association between drugs and diseases. The architecture of AMDGT is shown in Fig. 1. The main contributions of this study are summarised as follows.

- Multi-modal similarities from multiple perspectives were integrated together to capture the relationship between drugs and diseases. Furthermore, we innovatively introduced pretrained language model embeddings to construct a DDA network, which

Table 1
Summary of three benchmark datasets.

Dataset	Drugs	Diseases	Proteins	Drug–disease associations	Drug–protein associations	Disease–protein associations	Sparsity
B-dataset	269	598	1021	18,416	3110	5898	0.1144
C-dataset	663	409	993	2532	3773	10,734	0.0093
F-dataset	593	313	2741	1933	3243	54,265	0.0104

enabled our model to learn abundant biochemical information and enhance their generalisation.

- Dual graph transformer models were utilised to learn representations from homogeneous and heterogeneous networks, respectively. The attention mechanism in these modules assisted our model in obtaining both neighbour and global embeddings, through which AMDGT explored the attention aware representations of drugs and diseases deeply and comprehensively.
- We designed a modality interaction process based on the transformer encoder to fuse and incorporate the representations from different modalities. To the best of our knowledge, AMDGT is the first attention aware approach to model modality interactions between similarity data and topological information simultaneously in DDA predictions.
- Experimental results indicated that AMDGT is superior to state-of-the-art approaches on benchmark datasets. Discovering drug candidates for novel diseases also presented its promising and robust predictive capabilities in real-world settings.

Our paper is organised as follows. Section 2 presents the materials and methods used in this study, including datasets, data pre-processing, and detailed model-related information. Section 3 presents the experimental results and discussion, including a performance comparison under different settings, parameter sensitivity experiments, ablation experiments, case studies, and molecular docking experiments. Finally, Section 4 concludes our study findings.

2. Materials and methods

2.1. Datasets

In this study, three benchmarks were employed to evaluate performance. Each dataset included different types of entities, namely drugs, diseases, and proteins, and their association networks. B-dataset contains 269 drugs, 598 diseases, and 18,416 verified DDAs obtained from the Comparative Toxicogenomics Database (CTD) [57] by Zhang et al. [58]. The C-dataset was collected from previous studies [59] and included 663 drugs, 409 diseases, and 2532 known DDAs. The F-dataset was constructed by Gottlieb et al. [60] and comprised 593 drugs, 313 diseases, and 1933 DDAs. Among these, drugs were downloaded from the DrugBank Database [61], whereas diseases were collected from the Online Mendelian Inheritance in Man (OMIM) database [62]. These datasets were expanded by Zhao et al. [28,29] to introduce proteins and their corresponding associations into the DisGeNET database [63]. In the present study, negative samples were generated through random pairing of drugs and diseases without known associations. The dataset was carefully balanced to ensure equal numbers of negative and positive sets. Table 1 summarises the three datasets.

2.2. Similarity measures

2.2.1. Drug fingerprint similarity

Similar to Ref. [17], the similarity of drug molecular fingerprints was calculated to measure the similarity. For the Simplified Molecular Input Line Entry Specification (SMILES) of drugs, fingerprint similarity was obtained based on RDKit and Tanimoto [64]. We represented the drug fingerprint similarity as a matrix $DR_f \in R^{M\times M}$, where M denotes the drug number.

2.2.2. Disease phenotype similarity

Phenotype similarity is a measure of disease similarity calculated by text mining analysis of medical information derived from the OMIM database. Specifically, MimMiner [65] is conventionally used to compute disease phenotype similarity. The phenotype similarity matrix is denoted by $DS_p \in R^{N\times N}$, where N is the number of diseases.

2.2.3. Gaussian interaction profile (GIP) kernel similarity for drugs and diseases

We introduced GIP kernel similarity to comprehensively describe drug fingerprint similarity and disease phenotype similarity, addressing their inherent sparsity and potential information insufficiency. According to the hypothesis that drugs associated with the same disease tend to exhibit similarity, the GIP kernel similarity between drugs r_i and r_j is formulated as:

$$DR_G\left(r_i, r_j\right) = \exp\left(-\gamma_r \left\| IP\left(r_i\right) - IP\left(r_j\right)\right\|^2\right) \tag{1}$$

$$\gamma_r = \frac{\gamma_r'}{\frac{1}{M}\sum_{k=1}^{M}\left\| IP\left(r_k\right)\right\|^2} \tag{2}$$

where $DR_G \in R^{M\times M}$, $IP(r_i)$ represents the corresponding column of drug r_i in the drug–disease association matrix and γ_r' equals 1.

Similarly, the GIP kernel similarity of diseases is formulated as :

$$DS_G\left(s_i, s_j\right) = \exp\left(-\gamma_s \left\| IP\left(s_i\right) - IP\left(s_j\right)\right\|^2\right) \tag{3}$$

$$\gamma_s = \frac{\gamma_s'}{\frac{1}{N}\sum_{k=1}^{N}\left\| IP\left(s_k\right)\right\|^2} \tag{4}$$

where $DS_G \in R^{N\times N}$, $IP(s_i)$ means the corresponding row of disease d_i and γ_s' equals 1.

2.2.4. Similarity fusion

As mentioned above, we acquired drug and disease similarities through various modalities and perspectives and calculated the final comprehensive similarity matrices as follows:

$$DR_S\left(r_i, r_j\right) = \begin{cases} \dfrac{DR_f\left(r_i, r_j\right) + DR_G\left(r_i, r_j\right)}{2}, & \text{if } DR_f\left(r_i, r_j\right) \neq 0 \\ DR_G\left(r_i, r_j\right), & \text{otherwise} \end{cases} \tag{5}$$

$$DS_S\left(s_i, s_j\right) = \begin{cases} \dfrac{DS_p\left(s_i, s_j\right) + DS_G\left(s_i, s_j\right)}{2}, & \text{if } DS_p\left(s_i, s_j\right) \neq 0 \\ DS_G\left(s_i, s_j\right), & \text{otherwise} \end{cases} \tag{6}$$

where $DR_S\left(r_i, r_j\right) \in R^{M\times M}$ and $DS_S\left(s_i, s_j\right) \in R^{N\times N}$ are integrated drug and disease similarity matrices, respectively. The corresponding similarity networks of drugs and diseases generated from the matrices are denoted as DR_N and DS_N, respectively.

2.3. Similarity feature extraction

We employed a modified graph transformer (GT) network [66] to extract node embeddings from the similarity networks DR_N and DS_N. Inspired by the transformer model [67], the GT primarily learns node representations using a multi-head attention mechanism. Our approach utilised DR_N as a reference point for introducing the GT process and its principles. In defining the detailed update process for the l th GT

layer for the node features, represented by the similarity matrix DR_S, we employed the following equations:

$$h^0 = W^0 DR_S + b^0 \tag{7}$$

$$Q_{ij}^k = Q^k h_i^0,\ K_{ij}^k = K^k h_j^0,\ V_{ij}^k = V^k h_j^0 \tag{8}$$

$$A_{gt}^k\left(v_i, v_j\right) = softmax\left(\frac{Q_{ij}^k \cdot K_{ij}^k}{\sqrt{d'_{gt}}}\right) V_{ij}^k \tag{9}$$

$d'_{gt} = d_{gt}/h_{gt}$ denotes the dimension of each head; $j \in N_i$ represents the neighbouring nodes of node i; and accordingly, $h_i^0 \in R^{d_{gt}}$ and $h_j^0 \in R^{d_{gt}}$ denote the ith and jth column of h^0, respectively. Additionally, A_{gt}^k is the calculated multi-head attention score which captures the similarity and correlation between queries and keys.

$$\hat{h}_i = h_i + O_{gt}\left(Concat_{k=1}^{h_{gt}} \sum_{j \in N_i} A_{gt}^k\left(v_i, v_j\right)\right) \tag{10}$$

where $O_{gt} \in R^{d_{gt} \times d_{gt}}$ denotes the learnable matrix. Subsequently, the attention output $\hat{h}_i$ is fed into the feedforward network to obtain the final output node representation h_i of the $(l+1)$th layer. Updated after the l_{gt} layer, the drug similarity feature matrix $HR_s \in R^{M \times d_{gt}}$, that is, the final node feature of the similarity network, was obtained. Similarly, we calculated the disease similarity feature matrix $HD_s \in R^{N \times d_{gt}}$ using the aforementioned GT layers.

2.4. Biochemical heterogeneous network modelling

In addition to the aforementioned similarity networks, we constructed a heterogeneous network using biological and chemical information to mine drug- and disease-association information from a multimodal perspective. In particular, heterogeneous networks are composed of drug–disease, drug–protein, and disease–protein association networks. Biochemical information is crucial in the modelling of heterogeneous networks, which a vast number of previous studies have overlooked. In the present study, we introduced embeddings from pretrained language models to represent node element features, which enabled our biochemical network topology to be flexible and robust. For drugs, we used the mol2vec [68] features $F_{DR} \in R^{300}$ as a high-level representation of drug molecules to capture their chemical structural information. For diseases, following previous research [69], the Medical Subject Headings (MeSH) database was utilised to construct directed acyclic graphs, from which we obtained the disease node feature $F_{DI} \in R^{64}$ with rich biological and semantic information. For proteins, the latest and most superior protein language model, ESM-2 [70], was introduced to generate protein node embeddings $F_{PR} \in R^{320}$, which contain contextual structural and functional information. Finally, we modelled the biochemically heterogeneous network $A_N = \{V, E, F\}$, where V denotes the set of nodes, including drugs, diseases, and proteins; E denotes the edges, that is, the associations between the three different types of nodes; and F represents the feature vectors.

2.5. Network feature extraction

We proposed an expanded heterogeneous graph transformer (HGT) architecture to learn network embeddings and explore the relationship between drugs and diseases from A_N, which is similar to the architecture proposed by Hu et al. [71] and Mei et al. [72]. The outputs of HGT are network-level features, which are contextualised representations.

As defined in the vanilla heterogeneous graph transformer, the node v_t in V is denoted as a target node, while the node v_s is denoted as a source node. $e_{s,t}$ denotes the edge between v_s and v_t. For edge $e_{s,t}$ linked from source node v_s to target node v_t, the meta-relation is represented by $\left\langle \tau\left(v_s\right), \phi\left(e_{s,t}\right), \tau\left(v_t\right)\right\rangle$. The following sections elaborate the three steps that compose the update process for the lth layer of HGT.

2.5.1. Heterogeneous mutual attention

The node features of v_s and v_t in the lth layer are represented by $h^l\left[v_s\right]$ and $h^l\left[v_t\right]$. Similar to the vanilla transformer model, a multi-head attention mechanism with h_{hgt} heads is applied:

$$Q^k\left(v_t\right) = Q^{\,k}\text{-Linear}_{\tau(v_t)}\left(h^{(l-1)}\left[v_t\right]\right) \tag{11}$$

$$K^k\left(v_s\right) = K^{\,k}\text{-Linear}_{\tau(v_s)}\left(h^{(l-1)}\left[v_s\right]\right) \tag{12}$$

$$V^k\left(v_s\right) = V^{\,k}\text{-Linear}_{\tau(v_s)}\left(h^{(l-1)}\left[v_s\right]\right) \tag{13}$$

The function $Q^k\left(v_t\right)$ transforms v_t into the kth query vector Q^k. Similarly, the functions $K^k\left(v_s\right)$ and $V^k\left(v_s\right)$ convert v_s into the kth key and value representations, respectively. The mutual attention score is then defined as

$$A\text{-}head^k\left(v_s, e_{s,t}, v_t\right) = \left(K^k\left(v_s\right) W_{\phi(e_{s,t})}^{ATT} Q^k\left(v_t\right)^T\right) \cdot \frac{\mu_{\langle \tau(v_s), \phi(e_{s,t}), \tau(v_t)\rangle}}{\sqrt{d'_{hgt}}} \tag{14}$$

$$A_{hgt}\left(v_s, e_{s,t}, v_t\right) = softmax\left(Concat_{k=1}^{h_{hgt}} \sum_{v_s \in \mathcal{N}(v_t)} A - head^k\left(v_s, e_{s,t}, v_t\right)\right) \tag{15}$$

where $W_{\phi(e_{s,t})}^{ATT}$ is a weight parameter to extract meta-relation semantic embeddings; $d'_{hgt} = d_{hgt}/h_{hgt}$ and μ serve to represent the significance of the triplet $\left\langle \tau\left(v_s\right), \phi\left(e_{s,t}\right), \tau\left(v_t\right)\right\rangle$. The attention weights of the source and target nodes are calculated by the softmax operation and concatenation process, similar to the vanilla multi-head attention in the transformer.

2.5.2. Heterogeneous message passing

The message-passing module was devised to incorporate the meta-relations of edges to alleviate the distribution differences between nodes and edges of different types in the heterogeneous network. The message between nodes is calculated as follows:

$$M - head^k\left(v_s, e_{s,t}, v_t\right) = V^k\left(v_s\right) W_{\phi(e_{s,t})}^{MSG} \tag{16}$$

$$M_{hgt}\left(v_s, e_{s,t}, v_t\right) = Concat_{k=1}^{h_{hgt}} \sum_{v_s \in \mathcal{N}(v_t)} M - head^k\left(v_s, e_{s,t}, v_t\right) \tag{17}$$

where $W_{\phi(e_{s,t})}^{MSG}$ is a weight matrix similar to $W_{\phi(e_{s,t})}^{ATT}$, used to incorporate edge dependency.

2.5.3. Target specific aggregation

The next step is to aggregate the neighbour information from the source node v_s to the target node v_t.

$$\hat{h}^l\left[v_t\right] = \underset{v_s \in \mathcal{N}(v_t)}{Aggregate}\left(A_{hgt}\left(v_s, e_{s,t}, v_t\right) \cdot M_{hgt}\left(v_s, e_{s,t}, v_t\right)\right) \tag{18}$$

$$h^l\left[v_t\right] = ReLU\left(\hat{h}^l\left[v_t\right]\right) + h^{l-1}\left[v_t\right] \tag{19}$$

The representation of target node v_t of the lth layer $h^l\left[v_t\right]$ was acquired. After stacking l_{hgt} layers, we calculated the highly contextualised output embeddings of drugs $HR_N \in R^{M \times d_{hgt}}$ and diseases $HD_N \in R^{N \times d_{hgt}}$.

2.6. Modality interaction

Instead of the conventional concatenation operation used in previous studies, we designed a modality interaction module called AttentionFusion to combine information on drugs and diseases from different modalities. The transformer encoder is leveraged in this module, serving as a fusion and interaction unit to capture effective associations using the similarity and heterogeneous network aspects. Additionally, we followed the interaction process of drug embedding in implementing this module. First, we stacked the similarity and network features to

generate the initialised drug embedding $\hat{H}_r \in R^{M\times 2\times d_{gt}}$. The multi-head attention layer is the most vital module in the transformer encoder and can be represented as

$$Q^k = W_Q^k \hat{H}_r,\ K^k = W_K^k \hat{H}_r,\ V^k = W_V^k \hat{H}_r \tag{20}$$

$$A_t\left(\hat{H}_r\right) = Concat_{k=1}^{h_t}\left(\sum softmax\left(\frac{Q^k \cdot K^k}{\sqrt{d_t'}}\right) V^k\right) O_t \tag{21}$$

where $1 \le k \le h_t$, $d_t' = d_t/h_t$. Subsequently, a two-layer feed-forward network with a residue connection and activation function was implemented to obtain the final multimodal drug representation $H_r \in R^{M\times 2d_t}$. Furthermore, our model learned disease embedding $H_d \in R^{N\times 2d_t}$ using a similar transformer encoder.

2.7. Prediction module

After performing the aforementioned steps, AMDGT learned the multimodal drug representations from DR_N and A_N. Similarly, disease representations were learned from DS_N and A_N. Finally, we computed the integrated drug–disease pair embedding using an element-level dot-product decoder. After decoding, the predictive association scores were output by a three-layer MLP with dropout layers and ReLU activation functions. The cross-entropy loss was utilised for optimisation as follows:

$$L = -\sum_{(i,j)\in Y^+\cup Y^-}\left[y_{ij} ln y_{ij}' + \left(1-y_{ij}\right)\ln\left(1-y_{ij}'\right)\right] \tag{22}$$

where Y^+ and Y^- denote positive and negative training samples, respectively. (i, j) denotes a given pair of drugs i and disease j. y_{ij} and y_{ij}' denote the ground-truth label and predicted score, respectively. The training procedure for AMDGT is presented in Algorithm 1.

Algorithm 1 The training procedure of AMDGT

Input: Integrated drug similarity network DR_N; Integrated disease similarity network DS_N; Biochemical heterogeneous network A_N; Number of epochs N.
Output: Predicted score y_{ij};

for $epoch = 1 \to N$ **do**
  $HR_S = GT(DR_N)$;
  $HD_S = GT(DS_N)$;
  $HR_N,\ HD_N = HGT(A_N)$;
  $\hat{H}_r = [HR_S,\ HR_N]$;
  $\hat{H}_d = [HD_S,\ HD_N]$;
  $H_r = AttentionFusion(\hat{H}_r)$;
  $H_d = AttentionFusion(\hat{H}_d)$;
  for drug i, disease j in data **do**
    $y_{ij} = MLP(H_r(i) \cdot H_d(j))$
  end for
  Update $GT, HGT, AttentionFusion, MLP$ by descending loss in equation (22).
end for

3. Results and discussions

3.1. Experimental settings and evaluation metrics

In our study, AMDGT was implemented using Python 3.9.13 and Pytorch 1.10.0 with a Nvidia RTX 3090 GPU. The Adam optimiser [73] was used for the training process, with a learning rate of 0.0001. The number of epochs used for training was 1000. The number of negative drug–disease pairs was equal to that of positive samples in our experimental settings.

Moreover, seven commonly used evaluation metrics were utilised in our work: area under the receiver operating characteristic (ROC) curve (AUC), area under the precision–recall (PR) curve (AUPR), accuracy, precision, recall, F1-score, and Matthews correlation coefficient (MCC). ROC curves were plotted to depict the relationship between the sensitivity and specificity of the model at various thresholds. Additionally, PR curves were plotted to reflect the relationship between prediction accuracy and recall.

3.2. Performance comparison under 10-fold cross-validation (10-CV) experiments

10-CV was conducted on the benchmark datasets. We first randomly split the dataset 10-folds. In each fold of the experiment, one fold was treated as test data without repetition, whereas the remaining nine folds served as training data. We compared the AMDGT with six state-of-the-art methods: deepDR [27], HNet-DNN [74], DRHGCN [75], HINGRL [28] DRWBNCF [76], and DDAGDL [29].

- deepDR [27] adopts a multi-modal deep autoencoder and a variational autoencoder to learn features from the heterogeneous networks.
- HNet-DNN [74] utilises a deep neural network to predict DDA based on the features extracted from the drug–disease heterogeneous network.
- DRHGCN [75] is a GCN-based drug repositioning method which designs inter- and intra-domain feature extraction modules and a layer attention mechanism.
- HINGRL [28] constructs a heterogeneous information network with biological information and applies graph representation learning techniques to obtain topological and biological features.
- DRWBNCF [76] proposes a novel weighted bilinear graph convolution operation, based on the neighbourhood interaction neural collaborative filtering model to predict DDA.
- DDAGDL [29] employs a geometric DL method on a heterogeneous network, effectively learning the smoothed features based on an attention mechanism (see Tables 2–4).

For all three datasets, AMDGT demonstrated superior performance over other baseline methods in terms of AUC, AUPR, accuracy, recall, F1-score, and MCC. These results indicate that AMDGT achieves highly competitive performance, surpassing state-of-the-art methods in DDA prediction (see Fig. 2).

In addition to its superior performance, AMDGT is more robust than the other baseline methods. For instance, the performance of deepDR on AUC and AUPR was significantly better than that on the Recall, F1-score and MCC. Importantly, both deepDR and DRHGCN demonstrated higher precision scores than their recall scores, indicating that they tend to identify known DDAs as negative. Nevertheless, the performance fluctuation of our model across all metrics was much lower than that of the other baselines. We attribute the robust performance of AMDGT to the implementation of multimodal networks and AttentionFusion architecture, which promote superior and consistent performance across all metrics.

Furthermore, when compared with GCN-based methods, namely DRHGCN and DRWBNCF, AMDGT demonstrated its advantages in DDA prediction. Compared with DRHGCN, AMDGT achieved 3.31% and 2.39% improvement in terms of AUC and AUPR, respectively. Furthermore, AMDGT outperformed DRWBNCF by 4.73% and 3.29% in AUC and AUPR. The main reason for the improvement in AMDGT is the application of multimodal networks. The lack of biological and medical information in the similarity networks of DRHGCN and DRWBNCF led to unsatisfactory performance. Moreover, compared with HINGRL and DDAGDL, which construct networks using biological knowledge, our model achieved the best performance. These two methods employ classifiers, such as Random Forest and XGBoost to complete the prediction of DDA based on drug and disease feature embeddings, which ignore information interaction and correlation. AMDGT alleviated this problem through the transformer encoder module after the feature

Table 2
Comparison results with baseline methods on B-dataset.

Model	AUC	AUPR	Accuracy	Precision	Recall	F1-score	MCC
deepDR	0.8202 ± 0.001	0.8048 ± 0.003	0.6012 ± 0.001	0.8832 ± 0.001	0.2334 ± 0.001	0.3692 ± 0.002	0.2990 ± 0.001
HNet-DNN	0.8927 ± 0.002	0.8919 ± 0.001	0.8101 ± 0.001	0.7825 ± 0.001	0.8281 ± 0.002	0.8047 ± 0.001	0.6211 ± 0.002
DRHGCN	0.9092 ± 0.002	0.9106 ± 0.002	0.8268 ± 0.002	0.8678 ± 0.001	0.7711 ± 0.001	0.8166 ± 0.001	0.6577 ± 0.001
HINGRL	0.8845 ± 0.003	0.8774 ± 0.002	0.8035 ± 0.002	0.8006 ± 0.003	0.8084 ± 0.004	0.8045 ± 0.004	0.6071 ± 0.004
DRWBNCF	0.9004 ± 0.001	0.9018 ± 0.002	0.5991 ± 0.002	**0.9810 ± 0.002**	0.2021 ± 0.004	0.3352 ± 0.003	0.3260 ± 0.003
DDAGDL	0.8421 ± 0.003	0.8315 ± 0.002	0.7646 ± 0.003	0.7616 ± 0.004	0.7703 ± 0.002	0.7659 ± 0.004	0.5292 ± 0.003
AMDGT	**0.9337 ± 0.002**	**0.9309 ± 0.003**	**0.8629 ± 0.002**	0.8614 ± 0.002	**0.8650 ± 0.002**	**0.8632 ± 0.003**	**0.7258 ± 0.002**

Table 3
Comparison results with baseline methods on C-dataset.

Model	AUC	AUPR	Accuracy	Precision	Recall	F1-score	MCC
deepDR	0.8959 ± 0.002	0.9203 ± 0.002	0.7295 ± 0.001	**0.9943 ± 0.001**	0.4616 ± 0.002	0.6305 ± 0.003	0.5435 ± 0.001
HNet-DNN	0.9460 ± 0.002	0.9399 ± 0.001	0.8838 ± 0.001	0.8778 ± 0.002	0.8820 ± 0.001	0.8799 ± 0.001	0.7674 ± 0.002
DRHGCN	0.9324 ± 0.003	0.9427 ± 0.004	0.8652 ± 0.002	0.9192 ± 0.001	0.8008 ± 0.002	0.8559 ± 0.002	0.7366 ± 0.003
HINGRL	0.9372 ± 0.004	0.9457 ± 0.005	0.8698 ± 0.002	0.8851 ± 0.004	0.8500 ± 0.004	0.8672 ± 0.003	0.7403 ± 0.002
DRWBNCF	0.9234 ± 0.004	0.9419 ± 0.004	0.8663 ± 0.004	0.8984 ± 0.002	0.8370 ± 0.004	0.8612 ± 0.004	0.7449 ± 0.003
DDAGDL	0.8693 ± 0.003	0.8935 ± 0.004	0.8168 ± 0.002	0.7874 ± 0.004	0.7721 ± 0.002	0.7797 ± 0.003	0.6230 ± 0.003
AMDGT	**0.9681 ± 0.003**	**0.9698 ± 0.003**	**0.9062 ± 0.004**	0.8903 ± 0.004	**0.9265 ± 0.004**	**0.9081 ± 0.003**	**0.8131 ± 0.005**

Table 4
Comparison results with baseline methods on F-dataset.

Model	AUC	AUPR	Accuracy	Precision	Recall	F1-score	MCC
deepDR	0.8553 ± 0.001	0.8870 ± 0.002	0.7501 ± 0.003	**0.9566 ± 0.002**	0.5241 ± 0.002	0.6771 ± 0.001	0.5609 ± 0.001
HNet-DNN	0.9188 ± 0.002	0.9157 ± 0.001	0.8426 ± 0.002	0.8502 ± 0.002	0.8413 ± 0.002	0.8457 ± 0.001	0.6851 ± 0.001
DRHGCN	0.9207 ± 0.004	0.9375 ± 0.002	0.8583 ± 0.001	0.9309 ± 0.001	0.7739 ± 0.002	0.8452 ± 0.002	0.7269 ± 0.002
HINGRL	0.9366 ± 0.006	0.9449 ± 0.004	0.8645 ± 0.005	0.8832 ± 0.004	0.8402 ± 0.003	0.8612 ± 0.006	0.7300 ± 0.004
DRWBNC	0.8958 ± 0.005	0.9200 ± 0.004	0.8296 ± 0.002	0.8752 ± 0.003	0.8237 ± 0.004	0.8341 ± 0.004	0.7232 ± 0.002
DDAGDL	0.9239 ± 0.007	0.9235 ± 0.002	0.8513 ± 0.004	0.8475 ± 0.005	0.8567 ± 0.004	0.8521 ± 0.005	0.7026 ± 0.003
AMDGT	**0.9598 ± 0.005**	**0.9617 ± 0.003**	**0.8905 ± 0.003**	0.8741 ± 0.003	**0.9128 ± 0.004**	**0.8929 ± 0.005**	**0.7819 ± 0.004**

extraction process, which is one of the merits of our model over those of previous studies. To summarise, incorporating multimodal networks and leveraging the modality interaction process of AMDGT is advantageous for the accurate and efficient capture of the complexity and structural information between drugs and diseases.

3.3. Parameter analysis

We explored the effect of the hyperparameters in AMDGT: (1) the number of neighbours of the similarity networks construction, termed k; and (2) the dimension of embeddings of drugs and diseases, termed d. We conducted a parameter analysis experiment on three datasets under 10-CV.

The parameter k represents the connection of each drug or disease to the top k drugs or diseases exhibiting the highest similarity when constructing the integrated drug and disease similarity network. Empirically, increasing the number of neighbours k can aggregate more similarity information for drugs and diseases. However, excessive k values may lead to noisy connections. Furthermore, the dimensions of the feature embeddings affect the capability of the model to learn the features. For convenience and accuracy, we set $d_{gt} = d_{hgt} = d_t$ in our experiments. To assess the sensitivity of AMDGT, we estimated the impact of two hyperparameters using grid search: the number of neighbours k ranged from $\{2, 3, 5, 10, 15\}$ and the dimension of embeddings d spanned $\{32, 64, 128, 256, 512\}$. Fig. 3 shows that the best AUC and AUPR values were observed for the B-dataset ($k = 3$, $d = 512$), C-dataset ($k = 5$, $d = 256$), and F-dataset ($k = 5$, $d = 256$).

3.4. Ablation studies

To demonstrate the function of the different modules in AMDGT, we performed ablation studies by devising five variant models, as illustrated in Table 5. The major difference between these variants is the method used to capture the feature embeddings. More specifically, we removed the similarity network, biochemical heterogeneous network, pretrained embedding features, and AttentionFusion module.

The results of ablation studies under 10-CV are presented in Fig. 4. First, we evaluated the function of the similarity network using variant 2, which uses only a heterogeneous network to extract features. This variant model exhibits the worst performance compared with the other variants. Accordingly, relying only on biochemical and medical knowledge may be insufficient to understand the complex association between drugs and diseases. Second, we evaluate the importance of the heterogeneous network in our model using variant 1. The results are presented in Fig. 4 and suggest that the heterogeneous network is necessary in our model to provide a rich and efficient link and topological information. Additionally, owing to the extreme complexity of drugs and diseases, we introduced pretrained language embeddings to accurately reflect their biological and chemical information and supplement other modalities beyond similarity. To validate the effectiveness of pretrained features in DDA prediction, we used similarity data to replace pretrained features to devise variants 3 and 5. According to the experimental results, we observed that the performances slightly decreased, demonstrating that the pretrained embeddings can provide better high-quality information as an additional supplement and extension of the similarity modality, which is beneficial for accurate DDA prediction. Furthermore, for the ablation experiments of modality interaction, we avoided employing the AttentionFusion module and only used concatenation to fuse features from the dual modals. AMDGT significantly outperformed variant 4, showcasing the remarkable improvement achieved through utilising the attention-aware modality interaction module, which allows the mining of high-level and deep information. Ultimately, AMDGT outperformed the other existing variants by combining the advantages of similarity data and heterogeneous networks with the modality interactions between them. Overall, from the ablation studies, we can suggest that the four major modules of AMDGT contributed to its superior performance.

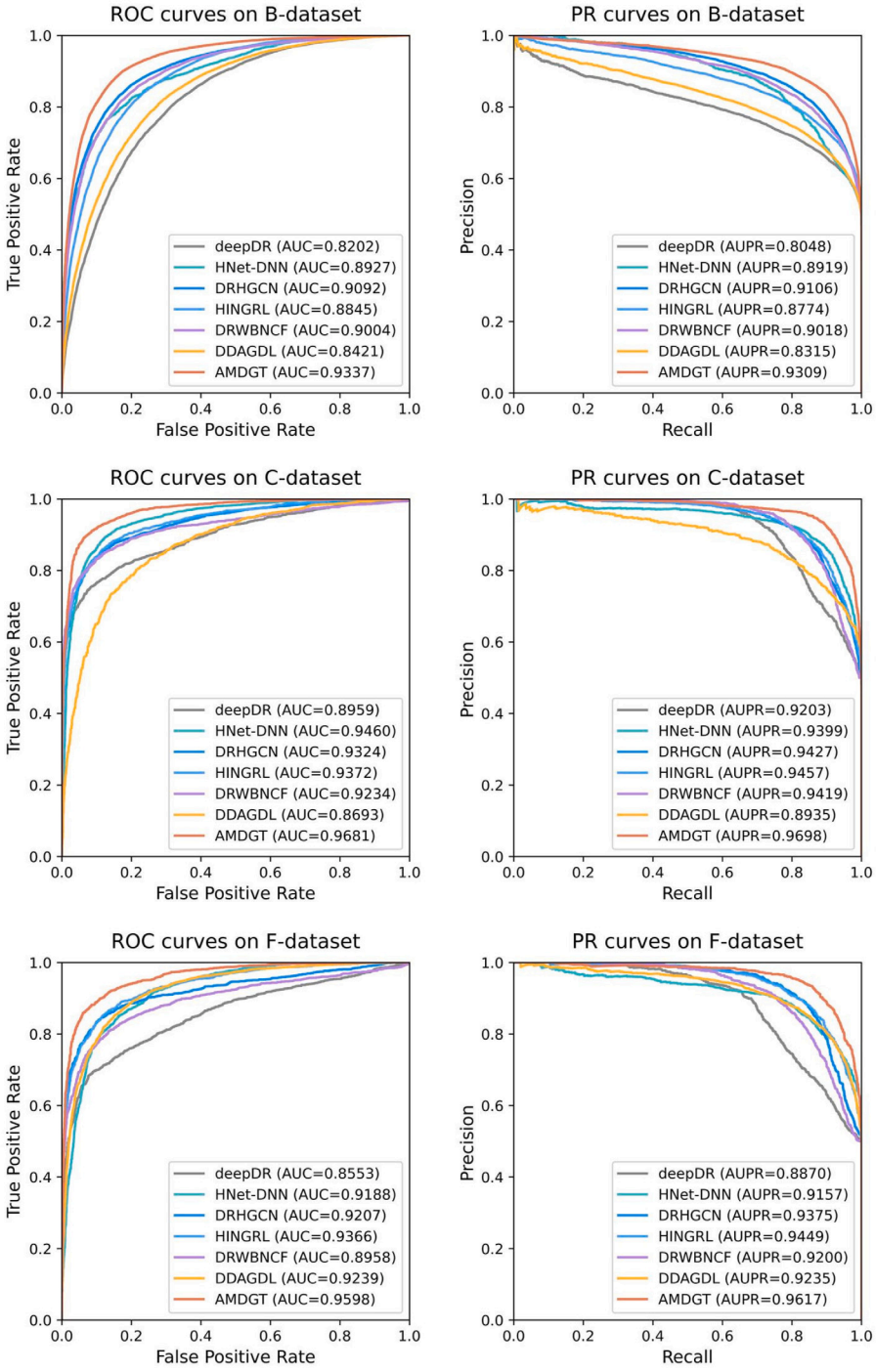


Fig. 2. The ROC and PR curves of different models for the benchmark datasets in the 10-CV experiment.

Table 5
Detail of variant models in ablation studies.

Model	Similarity network	Heterogeneous network	Pretrained features	Modality interaction
Variant1	✓	×	×	×
Variant2	×	✓	×	×
Variant3	✓	✓	×	×
Variant4	✓	✓	✓	×
Variant5	✓	✓	×	✓
AMDGT	✓	✓	✓	✓

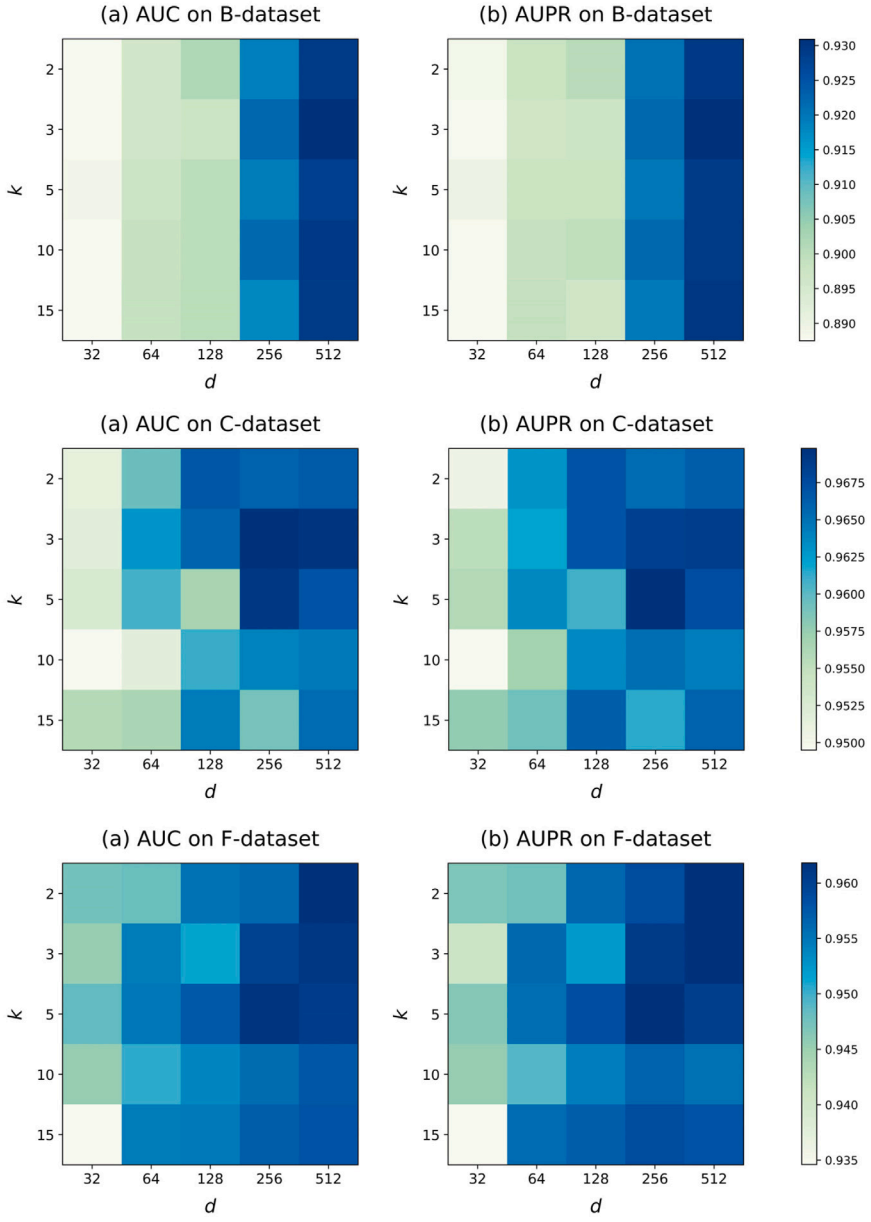


Fig. 3. The parameter setting analysis of the number of neighbours k and the dimension of embeddings d.

3.5. Discovering drug candidates for new diseases

To evaluate the ability of AMDGT to predict potential drugs for new diseases, we conducted a leave-one-out cross-validation (LOOCV) experiment using the F-dataset. In particular, for each disease d_i, all known DDAs involving that specific disease were removed to form the test data, and all the remaining samples were used as the training sets. As shown in Table 6, our AMDGT achieved an improvement of 9.90% in the AUC compared to state-of-the-art approaches and was only inferior to DRWBNCF in AUPR, demonstrating impressive performance in predicting potential drug candidates for unseen diseases. These results indicate that GCN-based models (DRHGCN and DRWBNCF) rely more on similarities, leading to an insufficient discovery of drugs for new diseases. However, AMDGT introduces high-quality pretrained embeddings with biochemical and medical information, which improves the generalisation and robustness of our model when encountering unknown data during real-world application. In addition, compared with approaches implementing biological information, namely HINGRL and DDAGDL, AMDGT also achieved better performance in terms of AUC and AUPR. The major reason for this phenomenon is attention-based integration and interaction, which allows our model to better capture association information to prioritise novel drugs for unknown diseases. In essence, when faced with a novel disease lacking known associations in the training data, AMDGT harnesses information from a dual perspective to effectively predict potential therapeutic drug candidates.

In addition, we also performed a LOOCV experiment on our variant models to evaluate the predictive ability for novel diseases. As displayed in Fig. 5, AMDGT performed satisfactorily under LOOCV settings, outperforming the other five variants in AUC and AUPR. In this experimental setting, the variants were insufficient to prioritise drugs for disease testing, mainly because of the absence of associations between new diseases and confirmed drugs. Moreover, we confirmed that the number of positive samples was successfully recovered from the top k candidates. Under the threshold of 1000, the number of positive associations identified by AMDGT was significantly higher than that of the other variant models, reaching almost 30% for the top 1000

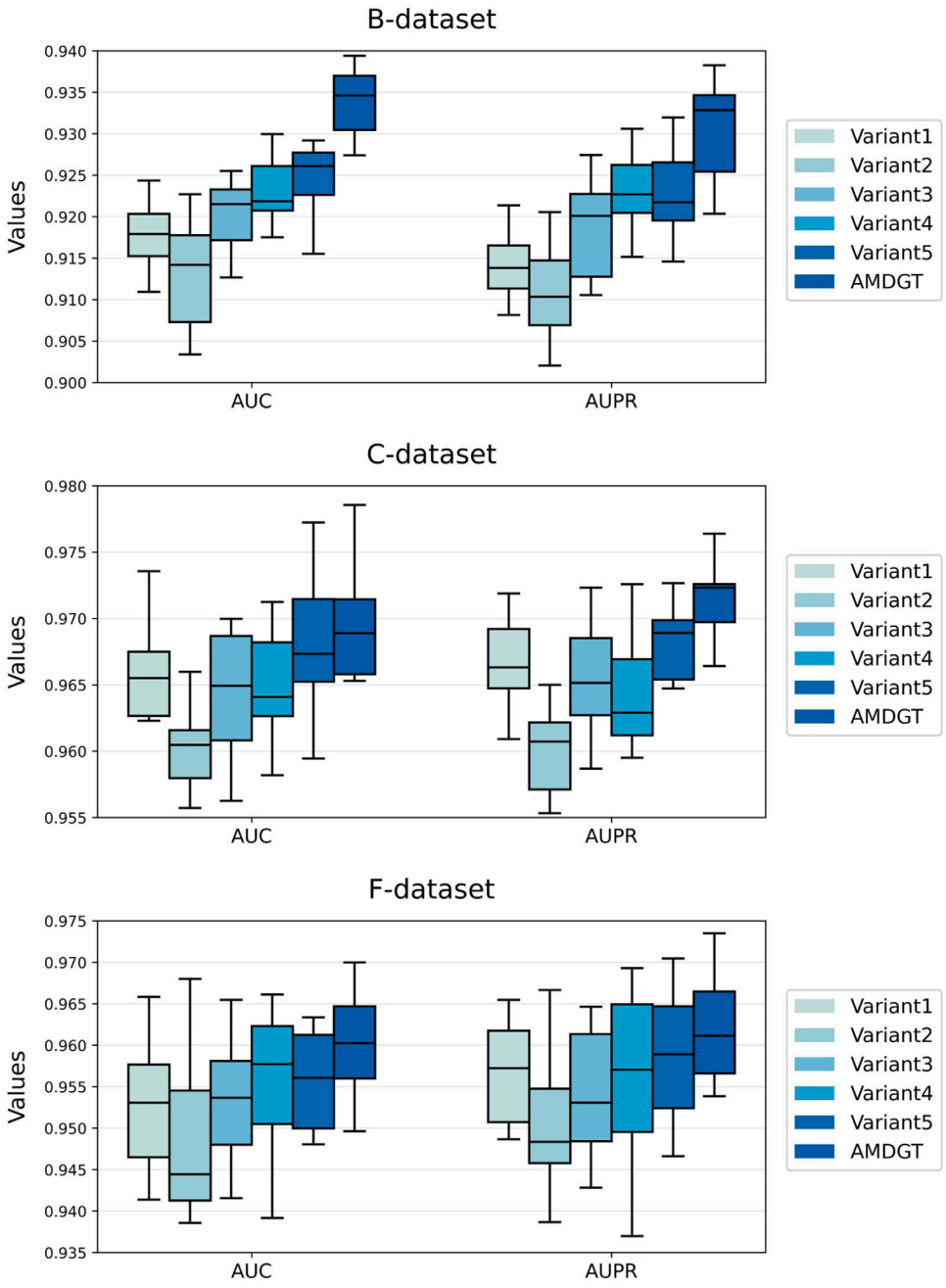


Fig. 4. The performance of variant models for the benchmark datasets in ablation studies.

Table 6
Performance comparison with the baseline methods using the F-dataset in the LOOCV experiment.

Model	AUC	AUPR
deepDR	0.5303 ± 0.152	0.0320 ± 0.025
HNet-DNN	0.7256 ± 0.125	0.1025 ± 0.114
DRHGCN	0.8110 ± 0.135	0.1180 ± 0.102
HINGRL	0.6298 ± 0.148	0.0327 ± 0.017
DRWBNCF	0.7750 ± 0.107	**0.1740 ± 0.112**
DDAGDL	0.6197 ± 0.158	0.0386 ± 0.027
AMDGT	**0.9100 ± 0.098**	0.1591 ± 0.127

predictions in the F-dataset, which again indicates the outstanding capability of discovering potential disease-related drugs.

3.6. Case studies

To investigate the ability of AMDGT to discover unknown DDAs in realistic settings, we performed additional experiments on the F-dataset. Specifically, we used all known DDAs in the F-dataset as the training data and chose the unknown associations for identification. We aimed to discover novel potential drugs for Alzheimer's disease (AD) and Parkinson's disease (PD). The candidate drugs for each disease were ranked in descending order according to the association probabilities, which are the output scores predicted by AMDGT.

The top 10 AMDGT-predicted drugs for AD are listed in Table 7. Among these, seven drugs have been validated in relevant medical literature. For example, Frovatriptan is a triptan drug used for the treatment of migraine headaches and is a potential β-secretase-1 enzyme inhibitor, which plays a key role in the production of beta-Amyloid protein (Aβ). Acamprosate, a drug used to maintain alcohol abstinence,

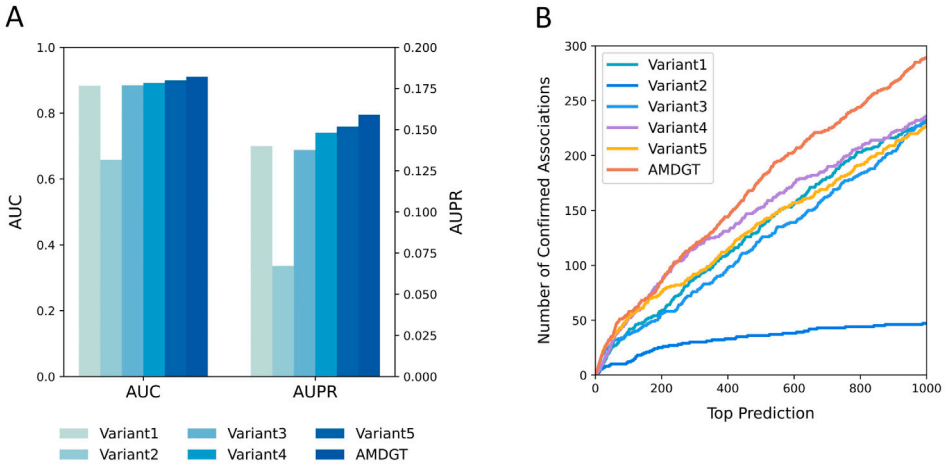


Fig. 5. The performance of variant models for the benchmark datasets in ablation studies.

Table 7
The top 10 potential drugs identified by AMDGT for Alzheimer's disease.

Disease	Drug	DrugBank ID	Score	Evidence (PMID)
Alzheimer's disease	Frovatriptan	DB00998	0.989	37175873
	Vinblastine	DB00570	0.982	25390692
	Ulobetasol	DB00596	0.980	N/A
	Meprobamate	DB00371	0.971	7988408
	Estramustine	DB01196	0.970	18077176
	Acamprosate	DB00659	0.968	25566747
	Clofarabine	DB00631	0.964	N/A
	Fenoprofen	DB00573	0.963	30328325
	Levetiracetam	DB01202	0.951	34570177
	Ciprofloxacin	DB00537	0.950	N/A

Table 8
The top 10 potential drugs identified by AMDGT for Parkinson's disease.

Disease	Drug	DrugBank ID	Score	Evidence (PMID)
Parkinson's disease	Sotalol	DB00489	0.997	20558393
	Dopamine	DB00988	0.986	18781671
	Dexamethasone	DB01234	0.985	2296253
	Zileuton	DB00744	0.981	N/A
	Disulfiram	DB00822	0.980	4646433
	Zolmitriptan	DB00315	0.975	N/A
	Amlodipine	DB00381	0.972	19276553
	Repaglinide	DB00912	0.968	36538285
	Lorazepam	DB00186	0.964	N/A
	Ciprofloxacin	DB00537	0.963	N/A

was predicted to be associated with AD by our model. This prediction was verified by previous studies reporting that acamprosate protects neurones and endothelial structures in vitro against Aβ.

In Table 8, six potential drugs, supported in the literature were identified for the effective treatment of PD. For instance, dexamethasone demonstrated benefits in protecting against neuronal damage. Additionally, AMDGT predicted repaglinide, an insulinotropic antidiabetic, as a candidate drug for PD. This prediction indicated a potential drug–disease association which has revealed that repaglinide mitigates neuroinflammation in PD. In conclusion, our artificial intelligence approach has suggested several promising treatments for AD and PD.

Furthermore, additional analysis was conducted to explain how AMDGT successfully identified verified candidate drugs whose associations were unknown in the F-dataset. Specifically, we compared the embedding representations of the predicted drugs with those of approved drugs for AD and PD in the F-dataset. We measured the similarity between these two types of drugs by calculating the Pearson coefficients of their final embeddings, H_r obtained using our model. Fig. 6 presents the similarity results for AD and PD. From the similarity distribution region, we observed that each predicted drug was similar to one or more approved drugs in certain aspects. This phenomenon scientifically demonstrates the rationale of AMDGT for predicting candidate drugs with high probabilities from biochemical and medical perspectives. Overall, AMDGT is a promising tool for discovering new drugs for the treatment of known diseases.

3.7. Molecular docking experiments

A molecular docking experiment was conducted to further evaluate and analyse the generalisability of practical applications of AMDGT. Conventional docking-based drug repositioning is marred by several drawbacks, such as false-positive rates and time-consuming processes. To overcome these problems, AMDGT simultaneously learns dual-modality information which can improve features from different perspectives. Using an attention-aware multimodal architecture, the proposed method can complete large-scale accurate DDA predictions at a reasonable time. Therefore, in the case of AD, the docking energies between the top ten predicted drugs and five important target proteins

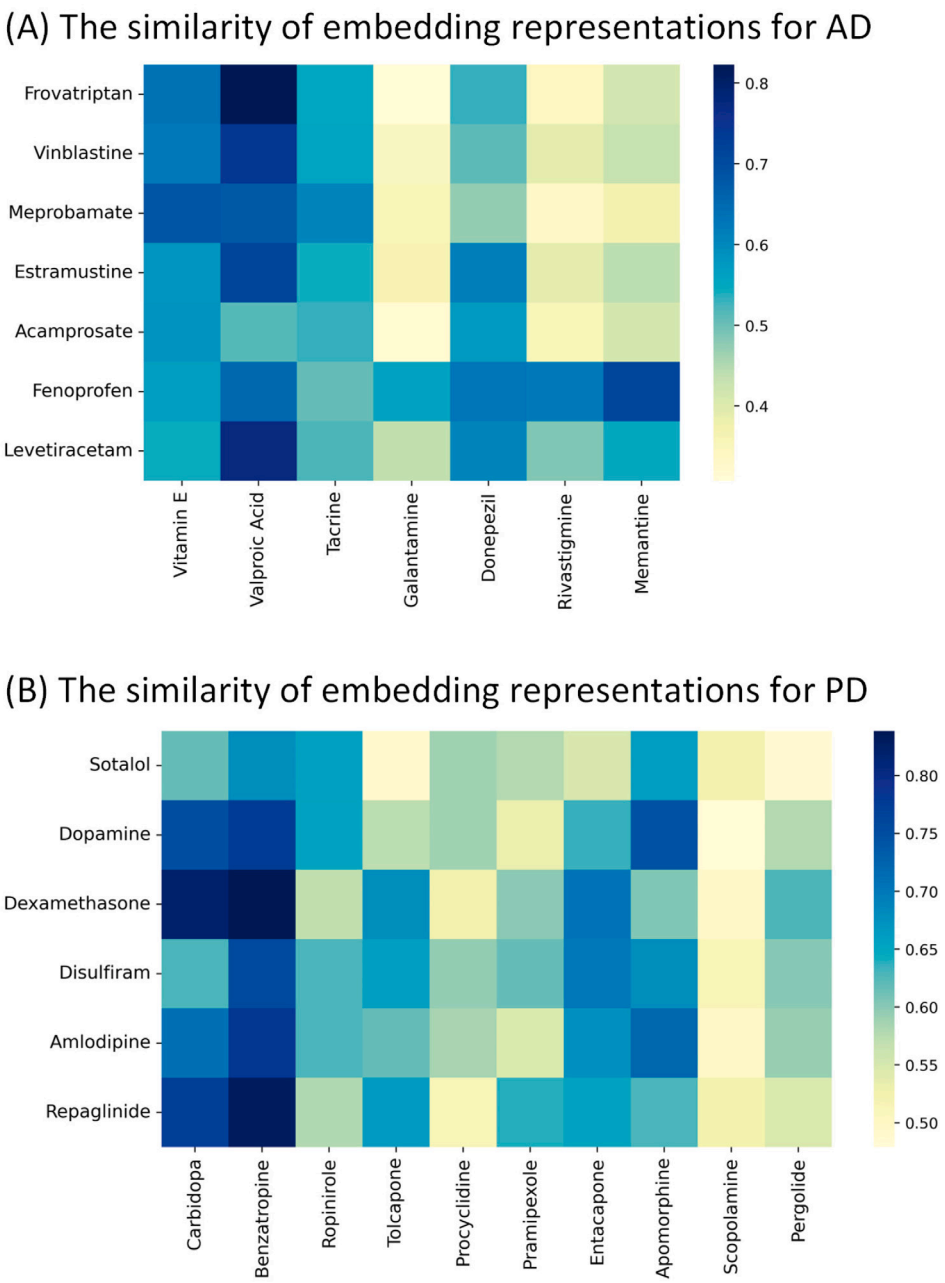


Fig. 6. The similarity of embeddings between predicted and confirmed drugs for Alzheimer's disease (AD) and Parkinson's disease (PD) in the F-dataset. The horizontal axis denotes the approved drugs, while the vertical axis denotes the predicted drugs.

were calculated [77]. AutoDock Vina [78,79] was used to perform molecular docking experiments. For unconfirmed associations between drug candidates and AD, we chose four as examples, presented in Fig. 7. As shown, the binding energies between the selected drugs and the corresponding targets are relatively low, indicating that these drug molecules have a high binding affinity for several targets of AD. Notably, the experimental results of molecular docking only imply a reasonable speculation of unapproved drugs exhibiting therapeutic effects on AD; thus, further laboratory experiments are required to verify their practical effects. In summary, owing to the promising performance of AMDGT, we believe that it is a reliable DR tool for researchers to discover novel DDAs.

4. Conclusions

In this study, we propose a novel computational method, AMDGT, for predicting DDAs. To overcome the challenges of DR, AMDGT first established multi-view similarity networks and heterogeneous association networks with pretrained language embeddings; thus, multimodal information can be learned from a dual perspective. Subsequently, homogeneous and heterogeneous graph transformer models were used to learn representations from similarity and association networks, respectively. AMDGT finally interacted with and integrated the embedding representations from different modalities based on the transformer encoder and finalised the DDA prediction task using the MLP. The experimental performance on benchmarks indicated that AMDGT is

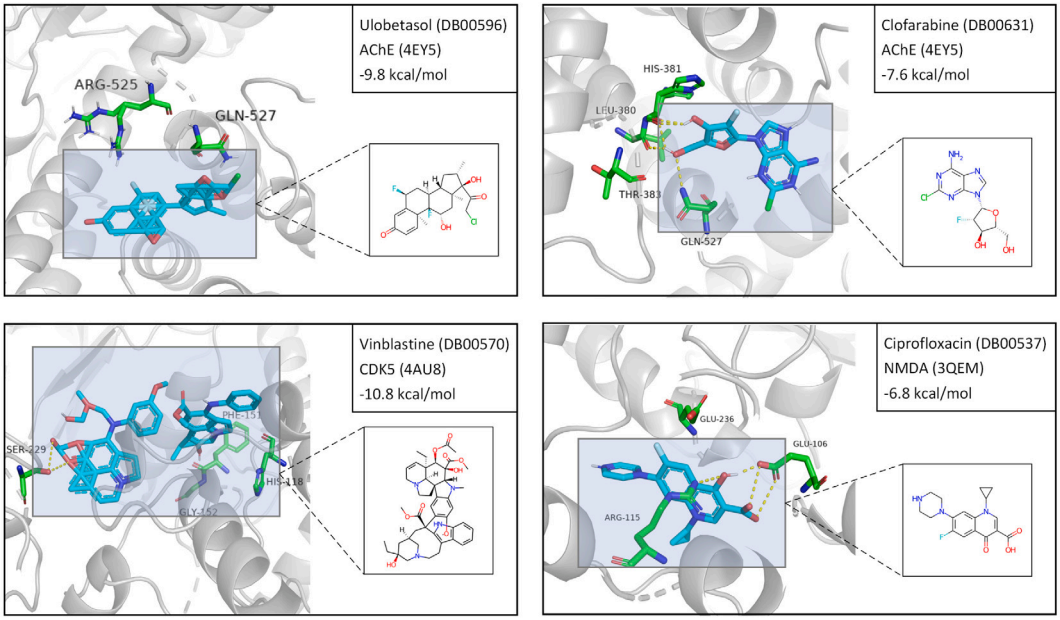


Fig. 7. The molecular docking results for four chosen drug candidates and corresponding targets of AD.

superior to the baseline approaches in terms of accuracy and robustness. Furthermore, extensive experiments of LOOCV and case studies indicated that AMDGT demonstrated strong generalisability in realistic applications, rendering AMDGT as a promising tool for predicting novel DDAs.

In terms of limitations and future work, we aim to extend our study from several perspectives. More specific instances and features, such as the drug molecular graph and protein contact map, can be added to the biochemically heterogeneous network to explore the mechanisms between drug molecules and diseases, which is expected to provide more expressive information. Moreover, we intend to investigate the interpretability in DDA prediction tasks which has been ignored by multiple current approaches. Understanding the mechanisms underlying the decisions made by the model may help biological researchers develop this field. In addition, considering the significant value of target proteins in predicting the association between drugs and diseases, introducing datasets with numerical values for the affinity between drugs and proteins and predicting both this value and DDAs to form a multitask prediction may further improve performance and enhance interpretability.

Code availability

The code is freely available at GitHub:https://github.com/JK-Liu7/AMDGT.

CRediT authorship contribution statement

Junkai Liu: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shixuan Guan:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Quan Zou:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Funding acquisition. **Hongjie Wu:** Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Prayag Tiwari:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation. **Yijie Ding:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Data curation.

Declaration of competing interest

The authors declare no conflct of lnterests.

Data availability

Github link is shared in the paper.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (62073231, 62176175, 62172076), National Research Project (2020YFC2006602), Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, China (KJS2166), Opening Topic Fund of Big Data Intelligent Engineering Laboratory of Jiangsu Province, China (SDGC2157), Postgraduate Research and Practice Innovation Program of Jiangsu Province, China, Zhejiang Provincial Natural Science Foundation of China (Grant No. LY23F020003), and the Municipal Government of Quzhou, China (Grant No. 2023D038).

References

[1] S. Whitebread, J. Hamon, D. Bojanic, L. Urban, Keynote review: In vitro safety pharmacology profiling: An essential tool for successful drug development, Drug Discov. Today 10 (21) (2005) 1421–1433, http://dx.doi.org/10.1016/S1359-6446(05)03632-9.

[2] Z. Tanoli, U. Seemab, A. Scherer, K. Wennerberg, J. Tang, M. Vähä-Koskela, Exploration of databases and methods supporting drug repurposing: A comprehensive survey, Brief. Bioinform. 22 (2) (2020) 1656–1678, http://dx.doi.org/10.1093/bib/bbaa003.

[3] T.T. Ashburn, K.B. Thor, Drug repositioning: Identifying and developing new uses for existing drugs, Nat. Rev. Drug Discov. 3 (8) (2004) 673–683, http://dx.doi.org/10.1038/nrd1468.

[4] S. Pushpakom, F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, et al., Drug repurposing: Progress, challenges and recommendations, Nat. Rev. Drug Discov. 18 (1) (2019) 41–58, http://dx.doi.org/10.1038/nrd.2018.168.

[5] Z. Tanoli, U. Seemab, A. Scherer, K. Wennerberg, J. Tang, M. Vähä-Koskela, Exploration of databases and methods supporting drug repurposing: A comprehensive survey, Brief. Bioinform. 22 (2) (2020) 1656–1678, http://dx.doi.org/10.1093/bib/bbaa003.

[6] C. Wang, Q. Zou, A machine learning method for differentiating and predicting human-infective coronavirus based on physicochemical features and composition of the spike protein, Chin. J. Electron. 30 (2021) 815–823.

[7] Y. Ding, F. Guo, P. Tiwari, Q. Zou, Identification of drug-side effect association via multi-view semi-supervised sparse model, IEEE Trans. Artif. Intell. (2023) 1–12, http://dx.doi.org/10.1109/TAI.2023.3314405.

[8] Y. Ding, P. Tiwari, Q. Zou, F. Guo, H.M. Pandey, C-loss based higher order fuzzy inference systems for identifying DNA N4-methylcytosine sites, IEEE Trans. Fuzzy Syst. 30 (11) (2022) 4754–4765, http://dx.doi.org/10.1109/TFUZZ.2022.3159103.

[9] Y. Ding, P. Tiwari, F. Guo, Q. Zou, Shared subspace-based radial basis function neural network for identifying ncRNAs subcellular localization, Neural Netw. 156 (2022) 170–178, http://dx.doi.org/10.1016/j.neunet.2022.09.026.

[10] Y. Ding, W. He, J. Tang, Q. Zou, F. Guo, Laplacian regularized sparse representation based classifier for identifying DNA N4-methylcytosine sites via $l_{2,1/2}$l2,1/2-matrix norm, IEEE/ACM Trans. Comput. Biol. Bioinform. 20 (1) (2023) 500–511, http://dx.doi.org/10.1109/TCBB.2021.3133309.

[11] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, Nat. Commun. 8 (1) (2017) 573, http://dx.doi.org/10.1038/s41467-017-00680-8.

[12] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, J. Wang, Computational drug repositioning using low-rank matrix approximation and randomized algorithms, Bioinformatics 34 (11) (2018) 1904–1912, http://dx.doi.org/10.1093/bioinformatics/bty013.

[13] H. Wang, J. Tang, Y. Ding, F. Guo, Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment, Brief. Bioinform. 22 (5) (2021) bbaa409, http://dx.doi.org/10.1093/bib/bbaa409.

[14] W. Zhang, H. Xu, X. Li, Q. Gao, L. Wang, DRIMC: An improved drug repositioning approach using Bayesian inductive matrix completion, Bioinformatics 36 (9) (2020) 2839–2847, http://dx.doi.org/10.1093/bioinformatics/btaa062.

[15] Y. Ding, J. Tang, F. Guo, Identification of drug-target interactions via fuzzy bipartite local model, Neural Comput. Appl. 32 (2020) 10303–10319.

[16] M. Sun, P. Tiwari, Y. Qian, Y. Ding, Q. Zou, MLapSVM-LBS: Predicting DNA-binding proteins via a multiple Laplacian regularized support vector machine with local behavior similarity, Knowl.-Based Syst. 250 (2022) 109174, http://dx.doi.org/10.1016/j.knosys.2022.109174.

[17] H. Yang, Y. Ding, J. Tang, F. Guo, Drug–disease associations prediction via multiple kernel-based dual graph regularized least squares, Appl. Soft Comput. 112 (2021) 107811, http://dx.doi.org/10.1016/j.asoc.2021.107811.

[18] Y. Ding, J. Tang, F. Guo, Identification of drug–target interactions via dual Laplacian regularized least squares with multiple kernel fusion, Knowl.-Based Syst. 204 (2020) 106254, http://dx.doi.org/10.1016/j.knosys.2020.106254.

[19] Y. Ding, J. Tang, F. Guo, Identification of drug-target interactions via multi-view graph regularized link propagation model, Neurocomputing 461 (2021) 618–631, http://dx.doi.org/10.1016/j.neucom.2021.05.100.

[20] X. Wang, B. Xin, W. Tan, Z. Xu, K. Li, F. Li, W. Zhong, S. Peng, DeepR2cov: Deep representation learning on heterogeneous drug networks to discover anti-inflammatory agents for COVID-19, Brief. Bioinform. 22 (6) (2021) bbab226, http://dx.doi.org/10.1093/bib/bbab226.

[21] T.N. Jarada, J.G. Rokne, R. Alhajj, SNF–CVAE: Computational method to predict drug–disease interactions using similarity network fusion and collective variational autoencoder, Knowl.-Based Syst. 212 (2021) 106585, http://dx.doi.org/10.1016/j.knosys.2020.106585.

[22] Y. Ding, P. Tiwari, F. Guo, Q. Zou, Multi-correntropy fusion based fuzzy system for predicting DNA N4-methylcytosine sites, Inf. Fusion 100 (2023) 101911, http://dx.doi.org/10.1016/j.inffus.2023.101911.

[23] B.-W. Zhao, L. Wang, P.-W. Hu, L. Wong, X.-R. Su, B.-Q. Wang, Z.-H. You, L. Hu, Fusing higher and lower-order biological information for drug repositioning via graph representation learning, IEEE Trans. Emerg. Top. Comput. (2023) 1–14, http://dx.doi.org/10.1109/TETC.2023.3239949.

[24] J.-L. Yu, Q.-Q. Dai, G.-B. Li, Deep learning in target prediction and drug repositioning: Recent advances and challenges, Drug Discov. Today 27 (7) (2022) 1796–1814, http://dx.doi.org/10.1016/j.drudis.2021.10.010.

[25] X. Pan, X. Lin, D. Cao, X. Zeng, P.S. Yu, L. He, R. Nussinov, F. Cheng, Deep learning for drug repurposing: Methods, databases, and applications, WIREs Comput. Mol. Sci. 12 (4) (2022) e1597, http://dx.doi.org/10.1002/wcms.1597.

[26] X. Su, L. Hu, Z. You, P. Hu, B. Zhao, Attention-based knowledge graph representation learning for predicting drug-drug interactions, Brief. Bioinform. 23 (3) (2022) bbac140, http://dx.doi.org/10.1093/bib/bbac140.

[27] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, F. Cheng, deepDR: A network-based deep learning approach to in silico drug repositioning, Bioinformatics 35 (24) (2019) 5191–5198, http://dx.doi.org/10.1093/bioinformatics/btz418.

[28] B.-W. Zhao, L. Hu, Z.-H. You, L. Wang, X.-R. Su, HINGRL: Predicting drug–disease associations with graph representation learning on heterogeneous information networks, Brief. Bioinform. 23 (1) (2021) bbab515, http://dx.doi.org/10.1093/bib/bbab515.

[29] B.-W. Zhao, X.-R. Su, P.-W. Hu, Y.-P. Ma, X. Zhou, L. Hu, A geometric deep learning framework for drug repositioning over heterogeneous information networks, Brief. Bioinform. 23 (6) (2022) bbac384, http://dx.doi.org/10.1093/bib/bbac384.

[30] Y. Zhang, J. Wang, Y. Liu, L. Rong, Q. Zheng, D. Song, P. Tiwari, J. Qin, A multitask learning model for multimodal sarcasm, sentiment and emotion recognition in conversations, Inf. Fusion 93 (2023) 282–301, http://dx.doi.org/10.1016/j.inffus.2023.01.005.

[31] Z. Yu, F. Huang, X. Zhao, W. Xiao, W. Zhang, Predicting drug–disease associations through layer attention graph convolutional network, Brief. Bioinform. 22 (4) (2020) bbaa243, http://dx.doi.org/10.1093/bib/bbaa243.

[32] H.J. Jiang, Z.H. You, Y.A. Huang, Predicting drug-disease associations via sigmoid kernel-based convolutional neural networks, J. Transl. Med. 17 (1) (2019) 1–11, http://dx.doi.org/10.1186/s12967-019-2127-5.

[33] F. Gong, M. Wang, H. Wang, S. Wang, M. Liu, SMR: Medical knowledge graph embedding for safe medicine recommendation, Big Data Res. 23 (2021) 100174, http://dx.doi.org/10.1016/j.bdr.2020.100174.

[34] X. Su, Z. You, D. Huang, L. Wang, L. Wong, B. Ji, B. Zhao, Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction, IEEE Trans. Knowl. Data Eng. 35 (6) (2023) 5640–5651, http://dx.doi.org/10.1109/TKDE.2022.3154792.

[35] Y. Wang, Y.-L. Gao, J. Wang, F. Li, J.-X. Liu, MSGCA: Drug-disease associations prediction based on multi-similarities graph convolutional autoencoder, IEEE J. Biomed. Health Inf. 27 (7) (2023) 3686–3694, http://dx.doi.org/10.1109/JBHI.2023.3272154.

[36] S. Wang, J. Li, D. Wang, D. Xu, J. Jin, Y. Wang, Predicting drug-disease associations through similarity network fusion and multi-view feature projection representation, IEEE J. Biomed. Health Inf. 27 (10) (2023) 5165–5176, http://dx.doi.org/10.1109/JBHI.2023.3300717.

[37] C. Jimenez-Mesa, J. Ramirez, J. Suckling, J. Vöglein, J. Levin, J.M. Gorriz, A non-parametric statistical inference framework for deep learning in current neuroimaging, Inf. Fusion 91 (2023) 598–611, http://dx.doi.org/10.1016/j.inffus.2022.11.007.

[38] X. Zhu, H. Li, H.T. Shen, Z. Zhang, Y. Ji, Y. Fan, Fusing functional connectivity with network nodal information for sparse network pattern learning of functional brain networks, Inf. Fusion 75 (2021) 131–139, http://dx.doi.org/10.1016/j.inffus.2021.03.006.

[39] Z. Gao, H. Ma, X. Zhang, Y. Wang, Z. Wu, Similarity measures-based graph co-contrastive learning for drug–disease association prediction, Bioinformatics 39 (6) (2023) btad357, http://dx.doi.org/10.1093/bioinformatics/btad357.

[40] L. Hu, Y. Yang, Z. Tang, Y. He, X. Luo, FCAN-MOPSO: An improved fuzzy-based graph clustering algorithm for complex networks with multi-objective particle swarm optimization, IEEE Trans. Fuzzy Syst. (2023) 1–16, http://dx.doi.org/10.1109/TFUZZ.2023.3259726.

[41] X. Wang, W. Yang, Y. Yang, Y. He, J. Zhang, L. Wang, L. Hu, PPISB: A novel network-based algorithm of predicting protein-protein interactions with mixed membership stochastic blockmodel, IEEE/ACM Trans. Comput. Biol. Bioinform. 20 (2) (2023) 1606–1612, http://dx.doi.org/10.1109/TCBB.2022.3196336.

[42] Z. Lou, Z. Cheng, H. Li, Z. Teng, Y. Liu, Z. Tian, Predicting miRNA–disease associations via learning multimodal networks and fusing mixed neighborhood information, Brief. Bioinform. 23 (5) (2022) bbac159, http://dx.doi.org/10.1093/bib/bbac159.

[43] J. Wen, X. Zhang, E. Rush, V.A. Panickan, X. Li, T. Cai, D. Zhou, Y.-L. Ho, L. Costa, E. Begoli, C. Hong, J.M. Gaziano, K. Cho, J. Lu, K.P. Liao, M. Zitnik, T. Cai, Multimodal representation learning for predicting molecule–disease relations, Bioinformatics 39 (2) (2023) btad085, http://dx.doi.org/10.1093/bioinformatics/btad085.

[44] P. Tiwari, L. Zhang, Z. Qu, G. Muhammad, Quantum fuzzy neural network for multimodal sentiment and sarcasm detection, Inf. Fusion 103 (2024) 102085, http://dx.doi.org/10.1016/j.inffus.2023.102085.

[45] S.A. Hooshmand, M. Zarei Ghobadi, S.E. Hooshmand, S. Azimzadeh Jamalkandi, S.M. Alavi, A. Masoudi-Nejad, A multimodal deep learning-based drug repurposing approach for treatment of COVID-19, Mol. Diversity 25 (2021) 1717–1730, http://dx.doi.org/10.1007/s11030-020-10144-9.

[46] Z. Xiong, F. Huang, Z. Wang, S. Liu, W. Zhang, A multimodal framework for improving in silico drug repositioning with the prior knowledge from knowledge graphs, IEEE/ACM Trans. Comput. Biol. Bioinform. 19 (5) (2022) 2623–2631, http://dx.doi.org/10.1109/TCBB.2021.3103595.

[47] B. Yang, H. Chen, Predicting circRNA-drug sensitivity associations by learning multimodal networks using graph auto-encoders and attention mechanism, Brief. Bioinform. 24 (1) (2023) bbac596, http://dx.doi.org/10.1093/bib/bbac596.

[48] P. Hu, Y.a. Huang, J. Mei, H. Leung, Z.h. Chen, Z.m. Kuang, Z.h. You, L. Hu, Learning from low-rank multimodal representations for predicting disease-drug associations, BMC Med. Inf. Decis. Mak. 21 (2021) 1–13, http://dx.doi.org/10.1186/s12911-021-01648-x.

[49] P. Wang, S. Zheng, Y. Jiang, C. Li, J. Liu, C. Wen, A. Patronov, D. Qian, H. Chen, Y. Yang, Structure-aware multimodal deep learning for drug–protein interaction prediction, J. Chem. Inf. Model. 62 (5) (2022) 1308–1317, http://dx.doi.org/10.1021/acs.jcim.2c00060.

[50] Y. Zhang, D. Song, X. Li, P. Zhang, P. Wang, L. Rong, G. Yu, B. Wang, A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis, Inf. Fusion 62 (2020) 14–31, http://dx.doi.org/10.1016/j.inffus.2020.04.003.

[51] K.R.M. Fernando, C.P. Tsokos, Deep and statistical learning in biomedical imaging: State of the art in 3D MRI brain tumor segmentation, Inf. Fusion 92 (2023) 450–465, http://dx.doi.org/10.1016/j.inffus.2022.12.013.
[52] D. Bang, S. Lim, S. Lee, S. Kim, Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers, Nature Commun. 14 (1) (2023) 3570, http://dx.doi.org/10.1038/s41467-023-39301-y.
[53] B.-M. Liu, Y.-L. Gao, D.-J. Zhang, F. Zhou, J. Wang, C.-H. Zheng, J.-X. Liu, A new framework for drug–disease association prediction combing light-gated message passing neural network and gated fusion mechanism, Brief. Bioinform. 23 (6) (2022) bbac457, http://dx.doi.org/10.1093/bib/bbac457.
[54] B.-W. Zhao, Z.-H. You, L. Wong, P. Zhang, H.-Y. Li, L. Wang, MGRL: Predicting drug-disease associations based on multi-graph representation learning, Front. Genet. 12 (2021) http://dx.doi.org/10.3389/fgene.2021.657182.
[55] P. Ladosz, L. Weng, M. Kim, H. Oh, Exploration in deep reinforcement learning: A survey, Inf. Fusion 85 (2022) 1–22, http://dx.doi.org/10.1016/j.inffus.2022.03.003.
[56] Y. Ding, J. Tang, F. Guo, Q. Zou, Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization, Brief. Bioinform. 23 (2) (2022) bbab582, http://dx.doi.org/10.1093/bib/bbab582.
[57] A.P. Davis, C.J. Grondin, R.J. Johnson, D. Sciaky, B.L. King, R. McMorran, J. Wiegers, T.C. Wiegers, C.J. Mattingly, The comparative toxicogenomics database: update 2017, Nucleic Acids Res. 45 (D1) (2016) D972–D978, http://dx.doi.org/10.1093/nar/gkw838.
[58] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, F. Liu, Predicting drug-disease associations by using similarity constrained matrix factorization, BMC Bioinformatics 19 (2018) 1–12, http://dx.doi.org/10.1186/s12859-018-2220-4.
[59] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, Y. Pan, Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm, Bioinformatics 32 (17) (2016) 2664–2671, http://dx.doi.org/10.1093/bioinformatics/btw228.
[60] A. Gottlieb, G.Y. Stein, E. Ruppin, R. Sharan, PREDICT: A method for inferring novel drug indications with application to personalized medicine, Mol. Syst. Biol. 7 (1) (2011) 496, http://dx.doi.org/10.1038/msb.2011.26.
[61] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: A major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (D1) (2017) D1074–D1082, http://dx.doi.org/10.1093/nar/gkx1037.
[62] A. Hamosh, A.F. Scott, J.S. Amberger, C.A. Bocchini, V.A. McKusick, Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, Nucleic Acids Res. 33 (suppl1) (2005) D514–D517.
[63] J. Pinero, A. Bravo, N. Queralt Rosinach, A. Gutierrez Sacristan, J. Deu-Pons, E. Centeno, J. Garcia Garcia, F. Sanz, L.I. Furlong, DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants, Nucleic Acids Res. 45 (D1) (2016) D833–D839.
[64] R. Guha, Chemical informatics functionality in R, J. Stat. Softw. 18 (5) (2007) 1–16, http://dx.doi.org/10.18637/jss.v018.i05.
[65] M.A. Van Driel, J. Bruggeman, G. Vriend, H.G. Brunner, J.A. Leunissen, A text-mining analysis of the human phenome, Eur. J. Hum. Genetics 14 (5) (2006) 535–542, http://dx.doi.org/10.1038/sj.ejhg.5201585.
[66] V.P. Dwivedi, X. Bresson, A generalization of transformer networks to graphs, 2020, arXiv:2012.09699.
[67] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017, arXiv:1706.03762.
[68] S. Jaeger, S. Fulle, S. Turk, Mol2vec: Unsupervised machine learning approach with chemical intuition, J. Chem. Inf. Model. 58 (1) (2018) 27–35, http://dx.doi.org/10.1021/acs.jcim.7b00616.
[69] Z.-H. Guo, Z.-H. You, D.-S. Huang, H.-C. Yi, K. Zheng, Z.-H. Chen, Y.-B. Wang, MeSHHeading2vec: A new method for representing MeSH headings as vectors based on graph embedding algorithm, Brief. Bioinform. 22 (2) (2020) 2085–2095, http://dx.doi.org/10.1093/bib/bbaa037.
[70] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (6637) (2023) 1123–1130, http://dx.doi.org/10.1126/science.ade2574.
[71] Z. Hu, Y. Dong, K. Wang, Y. Sun, Heterogeneous graph transformer, 2020, arXiv:2003.01332.
[72] X. Mei, X. Cai, L. Yang, N. Wang, Relation-aware heterogeneous graph transformer based drug repurposing, Expert Syst. Appl. 190 (2022) 116165, http://dx.doi.org/10.1016/j.eswa.2021.116165.
[73] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv:1412.6980.
[74] H. Liu, W. Zhang, Y. Song, L. Deng, S. Zhou, HNet-DNN: Inferring new drug–disease associations with deep neural network based on heterogeneous network features, J. Chem. Inf. Model. 60 (4) (2020) 2367–2376, http://dx.doi.org/10.1021/acs.jcim.9b01008.
[75] L. Cai, C. Lu, J. Xu, Y. Meng, P. Wang, X. Fu, X. Zeng, Y. Su, Drug repositioning based on the heterogeneous information fusion graph convolutional network, Brief. Bioinform. 22 (6) (2021) bbab319, http://dx.doi.org/10.1093/bib/bbab319.
[76] Y. Meng, C. Lu, M. Jin, J. Xu, X. Zeng, J. Yang, A weighted bilinear neural collaborative filtering approach for drug repositioning, Brief. Bioinform. 23 (2) (2022) bbab581, http://dx.doi.org/10.1093/bib/bbab581.
[77] H. Xie, H. Wen, M. Qin, J. Xia, D. Zhang, L. Liu, B. Liu, Q. Liu, Q. Jin, X. Chen, In silico drug repositioning for the treatment of Alzheimer's disease using molecular docking and gene expression data, RSC Adv. 6 (2016) 98080–98090, http://dx.doi.org/10.1039/C6RA21941A.
[78] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009) 2785–2791, http://dx.doi.org/10.1002/jcc.21256.
[79] O. Trott, A.J. Olson, AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461, http://dx.doi.org/10.1002/jcc.21334.