Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Special Track on AI and Health

DIAM: Adaptive Drug Repositioning with Decoupled Biological Mechanism and Instance-Aware Modulation

Kerui Xu[1], **Keyuan Xu**[2], **Shuheng Yin**[1] and **Hang Qiu**[1*]
[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[2]School of Information and Software Engineering, University of Electronic Science and Technology of China, China
{xukerui, xukeyuan, yinshuheng}@std.uestc.edu.cn, qiuhang@uestc.edu.cn

Abstract

Drug repositioning has emerged as an attractive drug development strategy with deep learning-based computational methods showing great potential in predicting Drug-Disease Associations (DDAs). However, dominant computational paradigms typically rely on Random Negative Sampling (RNS) and static embedding fusion, leading to two fundamental limitations. First, RNS treats unobserved pairs uniformly, resulting in coarse decision boundaries that fail to distinguish true associations from ambiguous candidates. Second, static fusion applies a monolithic combination of heterogeneous features, failing to adapt to the sample-specific dominance of different biological mechanisms. To address these issues, we propose DIAM, which establishes a mechanism-adaptive paradigm by explicitly decoupling structural and molecular signals. Specifically, DIAM introduces a Dual-Stream Biological Mechanism Decoupling module to construct global structural propagation and local molecular interaction views explicitly. Leveraging these views, we design a biological plausibility score to guide the hard negative sampling, enforcing finer-grained decision boundaries. Furthermore, an Adaptive Residual Gating (ARG) is devised to perform instance-aware modulation, dynamically weighing the contribution of global and local views for each specific pair. Extensive experiments on three benchmark datasets demonstrate that DIAM outperforms seven state-of-the-art methods. A case study on Alzheimer's disease further validates the model's effectiveness in identifying potential candidate drugs for practical application.

1 Introduction

Over the past few decades, the traditional model of drug research and development (R&D) has faced significant challenges, characterized by high costs, lengthy development cycles, and high failure rates [Huang *et al.*, 2025]. Statistics indicate that only about 11% of the candidate drugs that enter human clinical trials ultimately receive regulatory approval [DiMasi *et al.*, 2003; Kola and Landis, 2004]. In this context, drug repositioning, which identifies new indications for existing drugs, has emerged as a pivotal strategy to mitigate failure risks and accelerate therapeutic discovery [Lei *et al.*, 2024]. To support this strategy, computational methods, particularly deep learning, have been widely adopted to predict potential DDAs.

Despite the promising performance of recent deep learning models, the dominant computational paradigm in DDA prediction still suffers from two fundamental limitations, arising from a mismatch between simplified computational assumptions and complex biological realities: (1) Prevalent methods predominantly rely on random negative sampling [Huang *et al.*, 2024; Liu *et al.*, 2024], which indiscriminately treats unobserved drug-disease pairs as uniform negatives. Since random samples are often topologically distant from true associations, they fail to provide sufficient gradient penalties for ambiguous candidates, reducing the training process to a trivial discrimination against easy negatives. This deficiency causes the resulting decision boundary to remain insufficiently compact, compromising the model's capability to discern fine-grained discrepancies between verified interactions and ambiguous candidates. (2) Existing approaches typically employ a static, monolithic fusion logic (e.g., a fixed dot-product) to integrate heterogeneous biological features [Liu *et al.*, 2024; Wang *et al.*, 2023; Yuan *et al.*, 2024]. This one-size-fits-all strategy fails to accommodate instance-specific heterogeneity, overlooking that the dominant informative view varies dynamically between global structural propagation and local molecular interactions across different drug-disease pairs. Consequently, this rigidity leads to suboptimal representation learning for complex biological associations.

To address these challenges, we propose DIAM, a novel framework that rethinks DDA prediction from the perspective of decoupled biological mechanisms. The overall architecture of DIAM is depicted in Figure 1. Specifically, DIAM first constructs a Dual-Encoder Representation Learning backbone. We employ Heterogeneous Graph Transformer (HGT) and Hypergraph Convolutional Networks (HGCN) to capture heterogeneous structural patterns and high-order cor-

*Corresponding Author.

relations, respectively. Distinct from previous approaches that rely on implicit feature propagation, we introduce a Dual-Stream Biological Mechanism Decoupling module to explicitly quantify association evidence into topological propagation signals and target-mediated interaction signals. These decoupled signals serve two critical functions. First, we utilize them to calculate a biological plausibility score, which guides the mining of hard negative samples and forces the model to learn robust decision boundaries. Second, we design an Adaptive Residual Gating to perform instance-aware modulation. This module uses prior confidence projection to dynamically modulate the fusion of the learned representations. Additionally, we integrate a contrastive learning objective to align the semantic spaces of the dual-encoder representations. Finally, the modulated representations are fed into a predictor to infer association probabilities. The main contributions of this study are summarized as follows.

- We propose DIAM, a novel framework that orchestrates a Dual-Encoder backbone with a Dual-Stream Mechanism Decoupling module to synthesize the biological semantics underlying both global structural and local molecular views.
- We introduce two mechanism-driven strategies leveraging the decoupled signals: (1) Hard Negative Sampling, which utilizes biological plausibility scores to identify ambiguous candidates and compel the model to learn finer-grained discriminative features; and (2) Adaptive Residual Gating for instance-aware modulation, which dynamically rectifies the fusion of learned representations.
- Extensive experiments on three benchmark datasets demonstrate that DIAM outperforms the state-of-the-art methods.

2 Related Work

2.1 Network Diffusion-Based Methods

Network diffusion-based methods infer associations by simulating information propagation across heterogeneous biological graphs. Representative works, such as TL-HGBI [Wang *et al.*, 2014] and MBiRW [Luo *et al.*, 2016], utilize update algorithm or random walk algorithms to predict potential drug-disease pairs. Despite offering strong interpretability, these methods are heavily reliant on the network's existing edges. This dependency hinders their ability to capture complex drug-disease associations, particularly with sparse data, limiting their predictive performance.

2.2 Machine Learning-Based Methods

Machine learning-based methods treat drug-disease association prediction as a classification or regression problem, making predictions by learning patterns from the feature vectors of drugs and diseases. Early works exhibit diverse strategies: PREDICT [Gottlieb *et al.*, 2011] utilizes logistic regression, whereas FBLM [Ding *et al.*, 2020] adopts a fuzzy SVM to model drug-disease relationships. Subsequently, SFRLDDA [Zhao *et al.*, 2023] employs random forests, while DDA-SKF [Gao *et al.*, 2022] leverages Laplacian Regularized Least Squares (LapRLS) to integrate heterogeneous similarities. However, machine learning-based methods often rely on features extracted by shallow learning frameworks, making it difficult to capture the deep, abstract associations hidden between drugs and diseases, thereby hindering further improvement of prediction performance [Aittokallio, 2022].

2.3 Deep Learning-Based Methods

Deep learning-based methods leverage complex neural architectures to extract nonlinear representations from biological networks. For example, deepDR [Zeng *et al.*, 2019] employs multimodal deep autoencoders to integrate multiple drug-related network information. In another case, LAGCN [Yu *et al.*, 2021] applies a Graph Convolutional Network (GCN) to learn node embeddings from a heterogeneous network with a hierarchical attention mechanism. Beyond these, REDDA [Gu *et al.*, 2022] learns node representations on a five-element heterogeneous network containing drugs, proteins, genes, pathways, and diseases by combining GNN with multiple attention mechanisms. More recently, AMDGT [Liu *et al.*, 2024] utilizes a dual-graph transformer to learn features from both similarity networks and three-element heterogeneous networks.

3 Methodology

3.1 Network Construction

Association Networks

In this work, three types of associations are considered: drug-disease associations, drug-protein associations, and disease-protein associations. These are denoted $A^{rd} \in R^{N_r \times N_d}$, $A^{rp} \in R^{N_r \times N_p}$, $A^{dp} \in R^{N_d \times N_p}$, where N_r, N_d, and N_p represent the number of drugs, diseases, and proteins. Taking A^{rd} as an example, if drug r_i and disease d_j are confirmed to be associated, the corresponding element A^{rd}_{ij} is assigned a value of 1; otherwise, it is assigned a value of 0.

Similarity networks

For the drug similarity network S^r, we employ the RDKit tool to process the SMILES strings of drugs and compute the fingerprint similarity for each drug, denoted as S^{rf} [Steinbeck *et al.*, 2003]. Furthermore, the Gaussian Interaction Profile (GIP) kernel similarity based on known DDAs is incorporated to supplement the fingerprint similarity when it is insufficient. Specifically, the GIP similarity $S^{rg}(r_i, r_j)$ between drug r_i and r_j is calculated as follows:

$$S^{rg}(r_i, r_j) = \exp(-\gamma \|IP(r_i) - IP(r_j)\|^2), \tag{1}$$

$$\gamma_r = \frac{\gamma'_r}{\frac{1}{N_r}\sum_{k=1}^{N_r} \|IP(r_k)\|^2}, \tag{2}$$

where $IP(r_i)$ represents the corresponding column of drug r_i in A^{rd}, and γ'_r is set to 1. The final drug similarity matrix S^r is defined as:

$$S^r(r_i, r_j) = \begin{cases} \dfrac{S^{rf}(r_i, r_j) + S^{rg}(r_i, r_j)}{2}, & \text{if } S^{rf}(r_i, r_j) \neq 0, \\ S^{rg}(r_i, r_j), & \text{otherwise.} \end{cases} \tag{3}$$

Similarly, the disease similarity network S^d is computed based on Medical Subject Headings (MeSH) descriptors

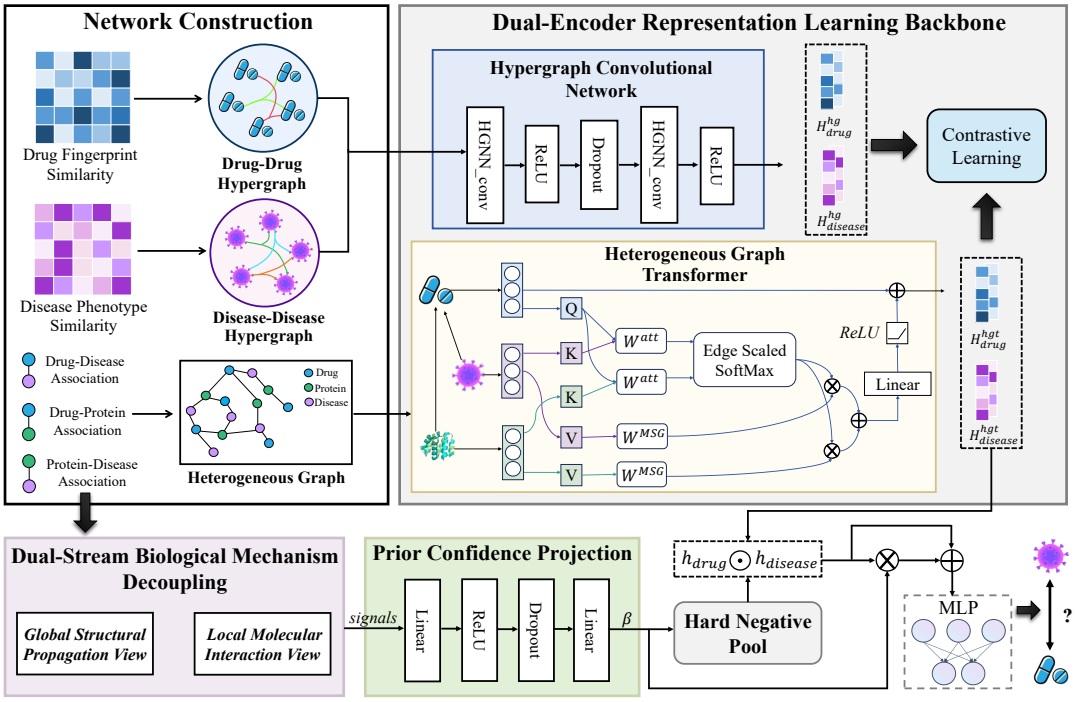

Figure 1: The workflow of DIAM for DDA prediction.

[Van Driel *et al.*, 2006] and GIP similarity. The protein similarity network integrates sequence similarity and GIP kernel similarity, where sequence similarity is calculated using the cosine similarity measure to compare 3-mer feature vectors. It is worth noting that the GIP similarity for proteins can be calculated from both drug and disease perspectives. Therefore, the two GIP matrices are averaged to derive the final GIP similarity.

3.2 Dual-Encoder Representation Learning Backbone

Heterogeneous Graph Transformer

We employ HGT to learn the interaction patterns between different types of nodes in the heterogeneous graph [Hu *et al.*, 2020; Mei *et al.*, 2022; Ni *et al.*, 2023]. HGT aggregates messages from neighbors $N(t)$ using attention weights specific to the meta-relation $\langle \tau(s), \phi(e), \tau(t) \rangle$. In particular, the intermediate representation at the $l-th$ layer is computed as:

$$\tilde{H}^{(l)}[t] = \oplus_{s \in N(t)} \big(Attention(s,e,t) \cdot Message(s,e,t) \big). \quad (4)$$

Subsequently, the node embedding is updated via a type-specific linear projection and a residual connection:

$$H^{(l)}[t] = \sigma \left(A - Linear_{\tau(t)} \tilde{H}^{(l)}[t] \right) + H^{(l-1)}[t]. \quad (5)$$

Through this process, we obtain contextually enriched representations for drug node H_r^{hgt} and disease nodes H_d^{hgt}. These representations will serve as key inputs for the subsequent association prediction task.

Hypergraph Convolutional Network

To capture high-order correlations beyond pairwise connections, we construct homogeneous hypergraphs $\mathcal{G}_r$ and $\mathcal{G}_d$ using the K-Nearest Neighbors (KNN) strategy [Huang *et al.*, 2022; Ouyang *et al.*, 2024; Wang *et al.*, 2024]. Using the drug hypergraph as an example, for each drug node $v \in \mathcal{V}_r$, a hyperedge e_v is formed by grouping v with its K most similar neighbors based on S^r. Since each node generates a unique hyperedge, this results in a square incidence matrix H, defined as:

$$\mathbf{H}(v,e) = \begin{cases} 1, & \text{if } v \in e, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

In the constructed homogeneous hypergraphs, we employ HGCN to encode group-level dependencies. Its feature propagation rule at the $l-th$ layer is defined as:

$$\mathbf{X}^{(l+1)} = \sigma \left(\mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-\frac{1}{2}} \mathbf{H}^T \mathbf{D}_v^{-\frac{1}{2}} \mathbf{X}^{(l)} \theta^{(l)} \right), \quad (7)$$

where $X^{(l)}$ and $\theta^{(l)}$ denote the node feature matrix and learnable filter at the $l-th$ layer. D_v, D_e, and W represent the diagonal matrices of node degrees, hyperedge degrees, and hyperedge weights, respectively. By applying HGCN to the drug hypergraph H_r and the disease hypergraph H_d, we finally obtain high-order embeddings H_r^{hg} and H_d^{hg}.

3.3 Dual-Stream Biological Mechanism Decoupling

Global Structural Propagation View

This global structural propagation view quantifies the topological propagation signal by modeling the global reachabil-

ity. The core idea is that the association potential of a pair (r_i, d_j) can be quantified by leveraging the similarity information of its topological neighbors.

Formally, let $\mathcal{N}_r(j) = \{k \mid A^{rd}_{kj} = 1 \wedge k \neq i\}$ denote the set of drugs treating disease d_j, and $\mathcal{N}_d(i) = \{l \mid A^{rd}_{il} = 1 \wedge l \neq j\}$ denote the set of diseases treated by drug r_i. We define the topological propagation signal $C_{topo}(r_i, d_j)$ as a two-side weighted aggregation:

$$C_{topo}(r_i, d_j) = \alpha \cdot \underbrace{\frac{1}{|\mathcal{N}_r(j)|} \sum_{k \in \mathcal{N}_r(j)} S^r(r_i, r_k)}_{\text{Drug-Side Propagation}} + (1-\alpha) \cdot \underbrace{\frac{1}{|\mathcal{N}_d(i)|} \sum_{l \in \mathcal{N}_d(i)} S^d(d_j, d_l)}_{\text{Disease-Side Propagation}}, \quad (8)$$

where the hyperparameter $\alpha \in [0, 1]$ is used to balance the relative contributions of the drug-side and disease-side information.

Local Molecular Interaction View

Complementing the global topology, this view explicitly models the molecular interaction signal driven by shared protein targets. We first define the importance $I(t)$ for each protein target t based on its connectivity degree distributions in A^{rp} and A^{dp}, followed by Min-Max normalization to obtain $I^{norm}(t)$. The specific formula is as follows:

$$I(t) = \log\big(1 + \sum_{i=1}^{N_r} A^{rp}_{it} + \sum_{j=1}^{N_d} A^{dp}_{jt}\big), \quad (9)$$

$$I^{norm}(t) = \frac{I(t) - \min(I)}{\max(I) - \min(I) + \epsilon_1}, \quad (10)$$

where ϵ_1 is a small constant 10^{-8} to avoid division by zero. Finally, the molecular interaction signal $C_{target}(r_i, d_j)$ is derived by aggregating the normalized importance of all shared targets that bridge drug r_i and disease d_j:

$$C_{target}(r_i, d_j) = \sum_{t=1}^{N_p} A^{rp}_{it} \cdot I^{norm}(t) \cdot A^{dp}_{jt}. \quad (11)$$

3.4 Mechanism-Driven Learning Strategies

Prior Confidence Projection

For the purpose of integrating the prior information with implicit deep representations, we propose the Prior Confidence Projection (PCP) module. Given the normalized topological propagation signal $\hat{C}_{topo}(r_i, d_j)$ and the target-mediated interaction signal $\hat{C}_{target}(r_i, d_j)$, PCP projects these mechanism indicators into a unified latent confidence scalar β_{ij}:

$$\mathbf{x}_{ij} = \left[\widehat{C}_{topo}(r_i, d_j) \parallel \widehat{C}_{target}(r_i, d_j)\right], \quad (12)$$

$$\beta_{ij} = \sigma\left(\mathbf{W}_2 \cdot \mathrm{ReLU}\left(\mathbf{W}_1 \mathbf{x}_{ij} + \mathbf{b}_1\right) + \mathbf{b}_2\right). \quad (13)$$

Hard Negative Sampling

To enforce a finer-grained decision boundary, we design a Hard Negative Sampling strategy that identifies unobserved pairs exhibiting high biological plausibility. We first quantify this plausibility by synthesizing the decoupled signals, weighted by the projected confidence β_{ij}, to generate the biological plausibility score W:

$$W(r_i, d_j) = \big(\hat{C}_{topo}(r_i, d_j)\big)^{1-\beta_{ij}} \cdot \big(\hat{C}_{target}(r_i, d_j)\big)^{\beta_{ij}}. \quad (14)$$

Consequently, the hard negative pool $\mathcal{N}_{hard}$ is constructed by filtering unobserved pairs that exceed a plausibility threshold θ_{neg}:

$$\mathcal{N}_{hard} = \big\{(r_i, d_j) \mid W(r_i, d_j) \geq \theta_{neg} \wedge A^{rd}_{ij} = 0\big\}. \quad (15)$$

During training, negative samples are drawn from $\mathcal{N}_{hard}$ to penalize ambiguous candidates and sharpen the model's discriminative capability.

Adaptive Residual Gating

To overcome the limitations of monolithic fusion strategies and perform instance-aware signal rectification, we propose the Adaptive Residual Gating (ARG) module. ARG adopts a residual enhancement formulation modulated by the projected confidence β_{ij}.

Formally, given the base interaction embedding $\mathbf{z}_{ij} = h^{hgt}_{r_i} \odot h^{hgt}_{d_j}$, the final rectified representation $\mathbf{z}^{final}_{ij}$ is computed as:

$$\mathbf{z}^{final}_{ij} = \mathcal{M}(\mathbf{z}_{ij}, \beta_{ij}) = (1 + \lambda\beta_{ij}) \cdot \mathbf{z}_{ij}, \quad (16)$$

where λ controls the modulation intensity. This residual gain selectively amplifies biologically plausible signals, thereby enhancing the expressiveness of embeddings.

3.5 Model Optimization

Contrastive Learning

To jointly optimize the heterogeneous graph and the homogeneous hypergraph views, we adopt the InfoNCE loss [Oord *et al.*, 2018] for dual-view contrastive learning. Considering drug nodes as an example, let $h^{hgt}_{r_i}$ and $h^{hg}_{r_i}$ denote the embedding vectors of drug r_i in the two views. The cross-view contrastive loss is designed to maximize the mutual information across views:

$$\mathcal{L}\_cross^{hgt}_{r_i} = -\log \frac{\sum_{j \in P_{r_i}} \exp\big(sim(h^{hgt}_{r_i}, h^{hg}_{r_j})/\tau\big)}{\sum_{k \in \{P_{r_i} \cup N_{r_i}\}} \exp\big(sim(h^{hgt}_{r_i}, h^{hg}_{r_k})/\tau\big)}, \quad (17)$$

$$\mathcal{L}\_cross^{hg}_{r_i} = -\log \frac{\sum_{j \in P_{r_i}} \exp\big(sim(h^{hg}_{r_i}, h^{hgt}_{r_j})/\tau\big)}{\sum_{k \in \{P_{r_i} \cup N_{r_i}\}} \exp\big(sim(h^{hg}_{r_i}, h^{hgt}_{r_k})/\tau\big)}, \quad (18)$$

$$\mathcal{L}\_cross_{r_i} = \frac{1}{|V_r|} \sum_{i \in V_r} \left[\gamma_c \cdot \mathcal{L}\_cross^{hgt}_{r_i} + (1-\gamma_c) \cdot \mathcal{L}\_cross^{hg}_{r_i}\right], \quad (19)$$

where P_{r_i} and N_{r_i} are the positive and negative sample sets for r_i, and τ is the temperature parameter. γ_c is the balancing coefficient for the cross-view loss.

Complementarily, the intra-view contrastive loss for drug nodes is formulated as follows:

$$\mathcal{L}\_intra^{hgt}_{r_i} = -\log \frac{\sum_{j \in P_{r_i}} \exp\big(sim(h^{hgt}_{r_i}, h^{hgt}_{r_j})/\tau\big)}{\sum_{k \in \{P_{r_i} \cup N_{r_i}\}} \exp\big(sim(h^{hgt}_{r_i}, h^{hgt}_{r_k})/\tau\big)}, \quad (20)$$

Dataset	Drugs	Diseases	Proteins	Drug-disease associations	Drug-protein associations	Disease-protein associations	Sparsity
Cdataset	663	409	993	2352	3773	10734	0.0093
Bdataset	269	598	1021	18416	3110	5898	0.1145
Fdataset	593	313	2741	1933	3243	54265	0.0104

Table 1: Summary of datasets.

$$\mathcal{L}\_intra_{r_i}^{hg} = -\log \frac{\sum_{j \in P_{r_i}} \exp\left(sim(h_{r_i}^{hg}, h_{r_j}^{hg})/\tau\right)}{\sum_{k \in \{P_{r_i} \cup N_{r_i}\}} \exp\left(sim(h_{r_i}^{hg}, h_{r_k}^{hg})/\tau\right)}, \quad (21)$$

$$\mathcal{L}\_intra_{r_i} = \frac{1}{|V_r|} \sum_{i \in V_r} \left[\gamma_i \cdot \mathcal{L}\_intra_{r_i}^{hgt} + (1-\gamma_i) \cdot \mathcal{L}\_intra_{r_i}^{hg}\right]. \quad (22)$$

The total contrastive loss for drugs is defined as $\mathcal{L}_r = \mathcal{L}\_cross_r + \mathcal{L}\_intra_r$. Similarly, the contrastive loss for diseases $\mathcal{L}_d$ can be derived.

Joint Optimization Objective

We employ a Multi-Layer Perceptron (MLP) as the decoder, optimizing the supervised task via the cross-entropy loss $\mathcal{L}_{cls}$. To simultaneously leverage the benefits of self-supervised alignment and supervised association mining, the overall objective function is defined as:

$$\mathcal{L}_{total} = \eta \cdot (\mathcal{L}_r + \mathcal{L}_d) + (1-\eta) \cdot \mathcal{L}_{cls}, \quad (23)$$

where $\eta \in [0, 1]$ is a hyperparameter that balances the contribution of the contrastive learning task and the classification task to the model's optimization.

4 Experiments

4.1 Experimental Setup

Datasets

We evaluate our proposed DIAM on three public benchmark datasets, each comprising three types of biological entities: drugs, diseases, and proteins, along with their association networks. Table 1 summarizes the dataset statistics. Cdataset[Luo *et al.*, 2016], sourced from DrugBank [Knox *et al.*, 2024] and Online Mendelian Inheritance in Man (OMIM) [Hamosh *et al.*, 2005], contains 663 drugs, 409 diseases, and 2532 drug-disease associations. Bdataset[Zhang *et al.*, 2018], extracted from the Comparative Toxicogenomics Database (CTD) [Davis *et al.*, 2023], includes 269 drugs, 598 diseases, and 18416 drug-disease associations. Fdataset[Gottlieb *et al.*, 2011], obtained from DrugBank and OMIM, comprises 593 drugs, 313 diseases, and 1933 drug-disease associations.

Evaluation metrics

To ensure a comprehensive and fair performance evaluation, we compare DIAM against all baselines using six standard metrics: Area Under the Receiver Operating Characteristic Curve (AUC), Area Under the Precision-Recall Curve (AUPR), Accuracy (ACC), Precision, Recall, and F1-Score. The parameter settings for all baseline models are kept consistent with their papers. All results are reported as the mean $\pm$ standard deviation with different random seeds to ensure stability and reproducibility.

Baselines

We compare DIAM with seven state-of-the-art methods, including DRHGCN [Cai *et al.*, 2021], DRWBNCF [Meng *et al.*, 2022], SMGCL [Gao *et al.*, 2023], AMDGT [Liu *et al.*, 2024], MilGNet [Gu *et al.*, 2025], DFDRNN [Zhu *et al.*, 2025], SIGDR [Cui *et al.*, 2025].

4.2 Performance Comparison

We conduct 10-CV on benchmark datasets and compare DIAM with seven state-of-the-art baseline methods. As shown in Table 2, DIAM achieves optimal performance on the majority of evaluation metrics across all three datasets, consistently attaining the best results in AUC, AUPR, Precision, Recall, and F1-Score. For instance, on the Cdataset, DIAM achieves an AUC of 0.9948 and an AUPR of 0.9954, representing relative improvements of 2.61% and 2.49%, respectively, over the second-best model. Similarly, on the Fdataset, it attains an AUC of 0.9935 and an AUPR of 0.9945, outperforming the strongest baselines by 3.18% in AUC and 3.00% in AUPR. On the more challenging Bdataset, DIAM ranks first across all six metrics, achieving an AUC of 0.9861 and an AUPR of 0.9890, which surpass the second-best model by 5.85% and 6.43%, respectively, demonstrating strong robustness even in complex scenarios.

Notably, on both the Cdataset and Fdataset, although DFDRNN attains the highest ACC, its AUPR and F1-Score are markedly lower. This suggests that its high ACC may stem from correctly classifying a large number of negative samples, while it performs poorly in identifying the critical yet scarce positive samples. In contrast, DIAM's superior AUPR and F1-Score demonstrate its robust capability to effectively identify the positive samples. This superiority is quantitatively evident, with DIAM surpassing the second-best baseline by over 2.49% in AUPR and 4.71% in F1-Score on both datasets.

In summary, the outstanding performance of DIAM underscores the effectiveness of our mechanism-adaptive paradigm. Unlike conventional approaches that depend on random negative sampling, DIAM employs hard negative sampling to enhance the model's discriminative capacity. For embedding fusion, DIAM incorporates Adaptive Residual Gating (ARG) for instance-aware modulation. These targeted designs enable the model to more effectively uncover drug-disease associations, ultimately leading to superior and robust predictive performance across benchmark datasets.

4.3 Ablation Study

To verify the contribution of each component in DIAM, we compare the full model against eight variants with AUC as

Dataset	Model	AUC	AUPR	ACC	Precision	Recall	F1 Score
Cdataset	DRHGCN	$0.9636 \pm 6 \times 10^{-3}$	$0.6213 \pm 2 \times 10^{-2}$	$0.9932 \pm 4 \times 10^{-4}$	$0.6575 \pm 4 \times 10^{-2}$	$0.5814 \pm 4 \times 10^{-2}$	$0.6148 \pm 2 \times 10^{-2}$
	DRWBNCF	$0.9382 \pm 1 \times 10^{-2}$	$0.5631 \pm 2 \times 10^{-2}$	$0.9926 \pm 8 \times 10^{-4}$	$0.6415 \pm 7 \times 10^{-2}$	$0.5063 \pm 3 \times 10^{-2}$	$0.5623 \pm 2 \times 10^{-2}$
	SMGCL	$0.9323 \pm 8 \times 10^{-3}$	$0.4744 \pm 1 \times 10^{-2}$	$0.9879 \pm 5 \times 10^{-4}$	$0.4046 \pm 2 \times 10^{-2}$	$0.6185 \pm 3 \times 10^{-2}$	$0.4892 \pm 2 \times 10^{-2}$
	AMDGT	$0.9687 \pm 5 \times 10^{-3}$	$0.9705 \pm 4 \times 10^{-3}$	$0.9056 \pm 1 \times 10^{-2}$	$0.8860 \pm 1 \times 10^{-2}$	$0.9313 \pm 1 \times 10^{-2}$	$0.9080 \pm 1 \times 10^{-2}$
	MilGNet	$0.8941 \pm 1 \times 10^{-2}$	$0.2014 \pm 3 \times 10^{-2}$	$0.9839 \pm 4 \times 10^{-3}$	$0.2945 \pm 2 \times 10^{-2}$	$0.2628 \pm 5 \times 10^{-2}$	$0.2777 \pm 6 \times 10^{-2}$
	DFDRNN	$0.9590 \pm 6 \times 10^{-3}$	$0.6547 \pm 2 \times 10^{-2}$	$\mathbf{0.9937 \pm 2 \times 10^{-4}}$	$0.7030 \pm 3 \times 10^{-2}$	$0.5699 \pm 2 \times 10^{-2}$	$0.6289 \pm 6 \times 10^{-3}$
	SIGDR	$0.9635 \pm 8 \times 10^{-3}$	$0.9658 \pm 8 \times 10^{-3}$	$0.9177 \pm 9 \times 10^{-3}$	$0.9085 \pm 2 \times 10^{-2}$	$0.9293 \pm 1 \times 10^{-2}$	$0.9186 \pm 9 \times 10^{-3}$
	DIAM	$\mathbf{0.9948 \pm 1 \times 10^{-3}}$	$\mathbf{0.9954 \pm 2 \times 10^{-3}}$	$0.9667 \pm 9 \times 10^{-3}$	$\mathbf{0.9909 \pm 9 \times 10^{-3}}$	$\mathbf{0.9420 \pm 1 \times 10^{-2}}$	$\mathbf{0.9657 \pm 9 \times 10^{-3}}$
Bdataset	DRHGCN	$0.8950 \pm 3 \times 10^{-3}$	$0.6042 \pm 9 \times 10^{-3}$	$0.8964 \pm 5 \times 10^{-3}$	$0.5445 \pm 2 \times 10^{-2}$	$0.6043 \pm 2 \times 10^{-2}$	$0.5718 \pm 6 \times 10^{-3}$
	DRWBNCF	$0.8819 \pm 2 \times 10^{-3}$	$0.5737 \pm 1 \times 10^{-2}$	$0.8910 \pm 7 \times 10^{-3}$	$0.5241 \pm 3 \times 10^{-2}$	$0.5841 \pm 2 \times 10^{-2}$	$0.5510 \pm 7 \times 10^{-3}$
	SMGCL	$0.8794 \pm 3 \times 10^{-3}$	$0.5744 \pm 9 \times 10^{-3}$	$0.8892 \pm 5 \times 10^{-3}$	$0.5157 \pm 2 \times 10^{-2}$	$0.5777 \pm 2 \times 10^{-2}$	$0.5443 \pm 6 \times 10^{-3}$
	AMDGT	$0.9276 \pm 4 \times 10^{-3}$	$0.9247 \pm 4 \times 10^{-3}$	$0.8549 \pm 5 \times 10^{-3}$	$0.8543 \pm 5 \times 10^{-3}$	$0.8557 \pm 1 \times 10^{-2}$	$0.8549 \pm 6 \times 10^{-3}$
	MilGNet	$0.8596 \pm 5 \times 10^{-3}$	$0.5111 \pm 9 \times 10^{-3}$	$0.8730 \pm 6 \times 10^{-3}$	$0.5034 \pm 6 \times 10^{-3}$	$0.4571 \pm 2 \times 10^{-2}$	$0.4791 \pm 2 \times 10^{-2}$
	DFDRNN	$0.8947 \pm 1 \times 10^{-3}$	$0.6204 \pm 6 \times 10^{-3}$	$0.8965 \pm 1 \times 10^{-2}$	$0.5541 \pm 6 \times 10^{-2}$	$0.5904 \pm 6 \times 10^{-2}$	$0.5666 \pm 6 \times 10^{-3}$
	SIGDR	$0.9212 \pm 4 \times 10^{-3}$	$0.9186 \pm 5 \times 10^{-3}$	$0.8522 \pm 6 \times 10^{-3}$	$0.8452 \pm 1 \times 10^{-2}$	$0.8626 \pm 1 \times 10^{-2}$	$0.8537 \pm 6 \times 10^{-3}$
	DIAM	$\mathbf{0.9861 \pm 2 \times 10^{-3}}$	$\mathbf{0.9809 \pm 1 \times 10^{-3}}$	$\mathbf{0.9514 \pm 4 \times 10^{-3}}$	$\mathbf{0.9796 \pm 4 \times 10^{-3}}$	$\mathbf{0.9220 \pm 1 \times 10^{-2}}$	$\mathbf{0.9499 \pm 4 \times 10^{-3}}$
Fdataset	DRHGCN	$0.9491 \pm 1 \times 10^{-2}$	$0.5454 \pm 3 \times 10^{-2}$	$0.9910 \pm 6 \times 10^{-4}$	$0.5740 \pm 4 \times 10^{-2}$	$0.5380 \pm 4 \times 10^{-2}$	$0.5535 \pm 2 \times 10^{-2}$
	DRWBNCF	$0.9228 \pm 1 \times 10^{-2}$	$0.4790 \pm 3 \times 10^{-2}$	$0.9897 \pm 6 \times 10^{-4}$	$0.5089 \pm 3 \times 10^{-2}$	$0.4775 \pm 4 \times 10^{-2}$	$0.4912 \pm 3 \times 10^{-2}$
	SMGCL	$0.9190 \pm 1 \times 10^{-2}$	$0.4441 \pm 2 \times 10^{-2}$	$0.9867 \pm 7 \times 10^{-4}$	$0.3974 \pm 2 \times 10^{-2}$	$0.5447 \pm 3 \times 10^{-2}$	$0.4595 \pm 2 \times 10^{-2}$
	AMDGT	$0.9617 \pm 8 \times 10^{-3}$	$0.9645 \pm 6 \times 10^{-3}$	$0.8952 \pm 2 \times 10^{-2}$	$0.8777 \pm 2 \times 10^{-2}$	$0.9193 \pm 2 \times 10^{-2}$	$0.8977 \pm 2 \times 10^{-2}$
	MilGNet	$0.9017 \pm 2 \times 10^{-2}$	$0.3520 \pm 5 \times 10^{-2}$	$0.9867 \pm 2 \times 10^{-3}$	$0.4096 \pm 4 \times 10^{-2}$	$0.3902 \pm 6 \times 10^{-2}$	$0.3997 \pm 5 \times 10^{-2}$
	DFDRNN	$0.9612 \pm 4 \times 10^{-3}$	$0.7192 \pm 1 \times 10^{-2}$	$\mathbf{0.9941 \pm 1 \times 10^{-4}}$	$0.7645 \pm 2 \times 10^{-2}$	$0.6275 \pm 3 \times 10^{-2}$	$0.6884 \pm 1 \times 10^{-2}$
	SIGDR	$0.9594 \pm 6 \times 10^{-3}$	$0.9630 \pm 5 \times 10^{-3}$	$0.9043 \pm 1 \times 10^{-2}$	$0.8964 \pm 3 \times 10^{-2}$	$0.9157 \pm 2 \times 10^{-2}$	$0.9054 \pm 1 \times 10^{-2}$
	DIAM	$\mathbf{0.9935 \pm 3 \times 10^{-3}}$	$\mathbf{0.9945 \pm 2 \times 10^{-3}}$	$0.9614 \pm 1 \times 10^{-2}$	$\mathbf{0.9902 \pm 1 \times 10^{-2}}$	$\mathbf{0.9322 \pm 1 \times 10^{-2}}$	$\mathbf{0.9602 \pm 1 \times 10^{-2}}$

Table 2: Performance comparison of the prediction results between DIAM and baselines on benchmark datasets.

the primary metric. The variants are defined as follows: V1 (w/o Decoupling) removes the decoupling module; V2 (w/o Local) and V3 (w/o Global) exclude molecular and structural views, respectively; V4 (w/o HNS) uses random sampling; V5 (w/o ARG) replaces ARG with a standard dot product; V6 (w/o PCP) replaces the PCP with a fixed weight; V7 (w/o CL) removes the contrastive learning objective; and V8 (w/o HGT) replaces the HGT encoder with GAT. The results are shown in Figure 2.

Impact of Mechanism Decoupling (V1-V3). V1 suffers the sharpest degradation across all datasets, confirming that decoupled signals are foundational to the model's performance. V2 and V3 further validate that biological signal dominance varies dynamically across datasets.

Efficacy of Sampling and Modulation (V4-V6). The significant drop in V4 confirms that mining hard negatives is essential for sharpening decision boundaries against ambiguous candidates. Moreover, V5 and V6 demonstrate that instance-aware gating effectively enhances the expressiveness of embeddings.

Backbone and Auxiliary Tasks (V7, V8). V7 and V8 confirm the value of cross-view alignment and HGT's superiority in modeling multi-relational dependencies, respectively.

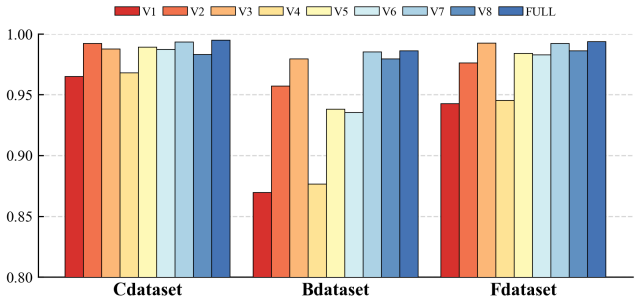

Figure 2: Ablation study results on three benchmark datasets.

4.4 Parameter sensitivity analysis

We perform a parameter sensitivity analysis on three key hyperparameters of DIAM using grid search. The model is trained for 300 epochs with a consistent learning rate of 5e-3 across all datasets. Optimal values are selected based on AUC. The detailed sensitivity analysis for these three parameters is presented in Figure 3.

Impact of embedding dimension (d)**.** The embedding dimension d determines the capacity of the node feature representations. It defines both the output dimension of the HGT and the input dimension of the final MLP predictor. We evaluate its impact by testing values from the set {32, 64, 128, 256, 512}. The optimal dimension is 128 for Cdataset and 64 for both Bdataset and Fdataset.

Impact of the structural propagation coefficient (α)**.** The coefficient α balances the relative contributions of the drug-side and disease-side information when computing C_{topo}. It is evaluated over {0.1, 0.3, 0.5, 0.7, 0.9}, where the optimal values are 0.3, 0.7, and 0.5 for Cdataset, Bdataset, and Fdataset, respectively.

Impact of the negative sampling threshold (θ_{neg})**.** The negative sampling threshold θ_{neg} is set to a value that filters the top percentile of unobserved pairs as hard negatives. We assess its impact over the range {50, 60, 70, 80, 90}. We select 90 as the optimal threshold, since further increases drastically shrink the negative pool and heighten the risk of overfitting on a small subset of samples.

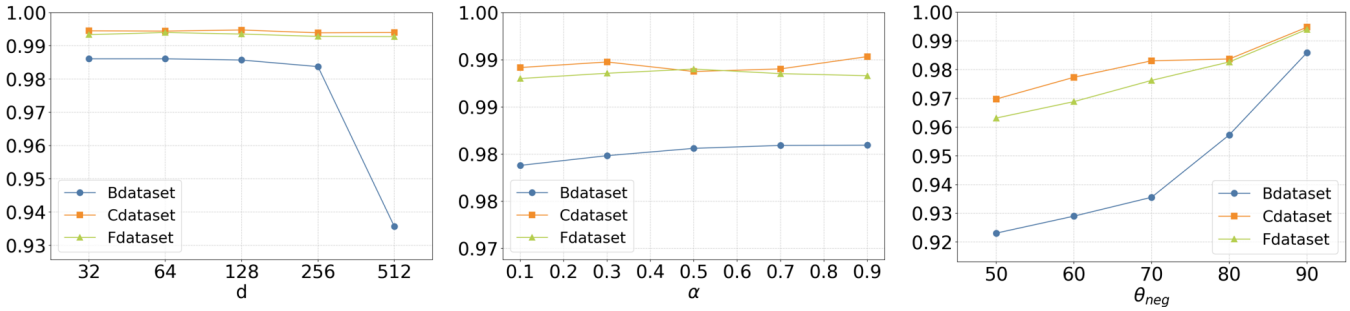

Figure 3: Parameter sensitivity analysis on three benchmark datasets.

4.5 Effectiveness of Adaptive Residual Gating

To validate the superiority of the proposed ARG module, we compare DIAM against two representative fusion strategy: -concat, which directly concatenates drug and disease embeddings, and -attention, which employs a standard attention mechanism to adaptively assign weights.

As shown in Table 3, DIAM consistently achieves the highest AUC across all datasets. The primary limitation of -concat and -attention is that they fail to incorporate biological semantics for instance-specific modulation. In contrast, DIAM leverages the decoupled signals from both the global structural and local molecular views to guide the fusion, ensuring that the final embedding is adaptively rectified by the biological signals for each drug-disease pair.

Dataset	-concat	-attention	**ARG**
Cdataset	$0.9871 \pm 4 \times 10^{-3}$	$0.9831 \pm 5 \times 10^{-3}$	$\mathbf{0.9948 \pm 1 \times 10^{-3}}$
Bdataset	$0.9397 \pm 5 \times 10^{-3}$	$0.9296 \pm 5 \times 10^{-3}$	$\mathbf{0.9861 \pm 2 \times 10^{-3}}$
Fdataset	$0.9873 \pm 3 \times 10^{-3}$	$0.9846 \pm 3 \times 10^{-3}$	$\mathbf{0.9935 \pm 3 \times 10^{-3}}$

Table 3: Comparison of different fusion strategies (AUC).

4.6 Case Study

To assess the model's practical application, we conduct a case study on Alzheimer's Disease (AD). AD is a progressive neurodegenerative disorder with limited therapeutic options [Livingston *et al.*, 2024]. Drug repositioning provides an important avenue for exploring new treatment strategies.

In this study, we rank the candidate drugs from high to low based on their prediction scores and select the top 10 for in-depth analysis. As shown in Table 4, 8 of these top 10 candidates have demonstrated therapeutic potential for AD, according to existing literature evidence.

Tacrolimus ranks first in our predictions. Recent preclinical studies have confirmed it as a strong candidate drug for AD [PMID: 40412360]. As a calcineurin inhibitor (CNI), it targets key AD features like neuroinflammation and glutamatergic disruptions. Preclinical studies have demonstrated its neuroprotective effects: in an acute $A\beta_{1-42}$ model, tacrolimus prevents cognitive impairment and neurodegeneration. Furthermore, in chronic APP/PS1 mice, it improves social interaction deficits, partially reduces microgliosis, and restores impaired glutamate release and calcium levels.

Rank	Drugs	DrugBank ID	Evidence (PMID)
1	Tacrolimus	DB00864	40412360
2	Progesterone	DB00396	39663597
3	Sertraline	DB01104	36638595
4	Metamfetamine	DB01577	32304732
5	Prednisone	DB00635	8835883
6	Vigabatrin	DB01080	NA
7	Epoprostenol	DB01240	NA
8	Zileuton	DB00744	24973121
9	Cyclosporine	DB00091	40750957
10	Dextroamphetamine	DB00721	18591575

Table 4: Top-10 candidate drugs for AD.

Progesterone ranks second in our list. Existing literature demonstrates its multifaceted neuroprotective potential, such as upregulating BDNF and activating the PI3K/Akt signaling pathway [PMID: 39663597]. However, as this literature also points out, current clinical research conclusions regarding the direct treatment of AD with Progesterone are inconsistent and have not reached a consensus, suggesting its complex mechanisms of action require further investigation. Overall, the fact that 80% of the top-ranked candidates are supported by recent studies validates DIAM's reliability in prioritizing biologically significant candidates for drug repositioning.

5 Conclusions

In this paper, we present DIAM, a mechanism-adaptive framework that rectifies the mismatch between static computational paradigms and dynamic biological realities in drug repositioning. Specifically, we design a Dual-Stream Biological Mechanism Decoupling module to explicitly quantify distinct signals from global structural propagation and local molecular interaction views. Leveraging these decoupled signals, we implement Hard Negative Sampling to enforce finer-grained decision boundaries, and an Adaptive Residual Gating module to perform instance-aware modulation. Extensive experiments on three benchmark datasets demonstrate DIAM's consistent superiority over state-of-the-art methods, while case studies on Alzheimer's Disease validate its reliability in prioritizing biologically significant candidates. In future work, we will extend the depth of this paradigm by integrating multi-omics modalities to construct a comprehensive multi-view biological landscape. Code and datasets are available at https://github.com/Eyre68/DIAM.

Acknowledgements

This work was supported in part by the National Major Science and Technology Projects of China (2024ZD0523903), and the National Natural Science Foundation of China (72342014).

References

[Aittokallio, 2022] Tero Aittokallio. What are the current challenges for machine learning in drug discovery and repurposing? *Expert opinion on drug discovery*, 17(5):423–425, 2022.

[Cai *et al.*, 2021] Lijun Cai, Changcheng Lu, Junlin Xu, Yajie Meng, Peng Wang, Xiangzheng Fu, Xiangxiang Zeng, and Yansen Su. Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Briefings in bioinformatics*, 22(6), 2021.

[Cui *et al.*, 2025] Hai Cui, Meiyu Duan, Jianyuan Yuan, Ziwei Ma, and Yijia Zhang. Sign-aware graph contrastive learning for drug repositioning. *IEEE Journal of Biomedical and Health Informatics*, 2025.

[Davis *et al.*, 2023] Allan Peter Davis, Thomas C Wiegers, Robin J Johnson, Daniela Sciaky, Jolene Wiegers, and Carolyn J Mattingly. Comparative toxicogenomics database (ctd): update 2023. *Nucleic acids research*, 51(D1):D1257–D1262, 2023.

[DiMasi *et al.*, 2003] Joseph A DiMasi, Ronald W Hansen, and Henry G Grabowski. The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185, 2003.

[Ding *et al.*, 2020] Yijie Ding, Jijun Tang, and Fei Guo. Identification of drug–target interactions via fuzzy bipartite local model. *Neural Computing and Applications*, 32(14):10303–10319, 2020.

[Gao *et al.*, 2022] Chu-Qiao Gao, Yuan-Ke Zhou, Xiao-Hong Xin, Hui Min, and Pu-Feng Du. Dda-skf: predicting drug–disease associations using similarity kernel fusion. *Frontiers in Pharmacology*, 12:784171, 2022.

[Gao *et al.*, 2023] Zihao Gao, Huifang Ma, Xiaohui Zhang, Yike Wang, and Zheyu Wu. Similarity measures-based graph co-contrastive learning for drug–disease association prediction. *Bioinformatics*, 39(6):btad357, 2023.

[Gottlieb *et al.*, 2011] Assaf Gottlieb, Gideon Y Stein, Eytan Ruppin, and Roded Sharan. Predict: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology*, 7(1):496, 2011.

[Gu *et al.*, 2022] Yaowen Gu, Si Zheng, Qijin Yin, Rui Jiang, and Jiao Li. Redda: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. *Computers in biology and medicine*, 150:106127, 2022.

[Gu *et al.*, 2025] Yaowen Gu, Si Zheng, Bowen Zhang, Hongyu Kang, Rui Jiang, and Jiao Li. Deep multiple instance learning on heterogeneous graph for drug–disease association prediction. *Computers in Biology and Medicine*, 184:109403, 2025.

[Hamosh *et al.*, 2005] Ada Hamosh, Alan F Scott, Joanna S Amberger, Carol A Bocchini, and Victor A McKusick. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, 33(suppl_1):D514–D517, 2005.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, pages 2704–2710, 2020.

[Huang *et al.*, 2022] Weihong Huang, Zhong Li, Yanlei Kang, Xinghuo Ye, and Wenming Feng. Drug repositioning based on the enhanced message passing and hypergraph convolutional networks. *Biomolecules*, 12(11):1666, 2022.

[Huang *et al.*, 2024] Shuhan Huang, Minhui Wang, Xiao Zheng, Jiajia Chen, and Chang Tang. Hierarchical and dynamic graph attention network for drug-disease association prediction. *IEEE Journal of Biomedical and Health Informatics*, 28(4):2416–2427, 2024.

[Huang *et al.*, 2025] Zhaoyang Huang, Zhichao Xiao, Chunyan Ao, Lixin Guan, and Liang Yu. Computational approaches for predicting drug-disease associations: a comprehensive review. *Frontiers of Computer Science*, 19(5):195909, 2025.

[Knox *et al.*, 2024] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun Chin, Seth A Strawbridge, et al. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic acids research*, 52(D1):D1265–D1275, 2024.

[Kola and Landis, 2004] Ismail Kola and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716, 2004.

[Lei *et al.*, 2024] Song Lei, Xiujuan Lei, Ming Chen, and Yi Pan. Drug repositioning based on deep sparse autoencoder and drug–disease similarity. *Interdisciplinary Sciences: Computational Life Sciences*, 16(1):160–175, 2024.

[Liu *et al.*, 2024] Junkai Liu, Shixuan Guan, Quan Zou, Hongjie Wu, Prayag Tiwari, and Yijie Ding. Amdgt: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction. *Knowledge-Based Systems*, 284:111329, 2024.

[Livingston *et al.*, 2024] Gill Livingston, Jonathan Huntley, Kathy Y Liu, Sergi G Costafreda, Geir Selbæk, Suvarna Alladi, David Ames, Sube Banerjee, Alistair Burns, Carol Brayne, et al. Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. *The lancet*, 404(10452):572–628, 2024.

[Luo *et al.*, 2016] Huimin Luo, Jianxin Wang, Min Li, Junwei Luo, Xiaoqing Peng, Fang-Xiang Wu, and Yi Pan. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics*, 32(17):2664–2671, 2016.

[Mei *et al.*, 2022] Xin Mei, Xiaoyan Cai, Libin Yang, and Nanxin Wang. Relation-aware heterogeneous graph transformer based drug repurposing. *Expert Systems with Applications*, 190:116165, 2022.

[Meng *et al.*, 2022] Yajie Meng, Changcheng Lu, Min Jin, Junlin Xu, Xiangxiang Zeng, and Jialiang Yang. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings in bioinformatics*, 23(2):bbab581, 2022.

[Ni *et al.*, 2023] Qin Ni, Tingjiang Wei, Jiabao Zhao, Liang He, and Chanjin Zheng. Hhskt: A learner–question interactions based heterogeneous graph neural network model for knowledge tracing. *Expert Systems with Applications*, 215:119334, 2023.

[Oord *et al.*, 2018] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[Ouyang *et al.*, 2024] Jihong Ouyang, Chang Xuan, Bing Wang, and Zhiyao Yang. Aspect-based sentiment classification with aspect-specific hypergraph attention networks. *Expert Systems with Applications*, 248:123412, 2024.

[Steinbeck *et al.*, 2003] Christoph Steinbeck, Yongquan Han, Stefan Kuhn, Oliver Horlacher, Edgar Luttmann, and Egon Willighagen. The chemistry development kit (cdk): An open-source java library for chemo-and bioinformatics. *Journal of chemical information and computer sciences*, 43(2):493–500, 2003.

[Van Driel *et al.*, 2006] Marc A Van Driel, Jorn Bruggeman, Gert Vriend, Han G Brunner, and Jack AM Leunissen. A text-mining analysis of the human phenome. *European journal of human genetics*, 14(5):535–542, 2006.

[Wang *et al.*, 2014] Wenhui Wang, Sen Yang, Xiang Zhang, and Jing Li. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics*, 30(20):2923–2930, 2014.

[Wang *et al.*, 2023] Ying Wang, Ying-Lian Gao, Juan Wang, Feng Li, and Jin-Xing Liu. Msgca: Drug-disease associations prediction based on multi-similarities graph convolutional autoencoder. *IEEE Journal of Biomedical and Health Informatics*, 27(7):3686–3694, 2023.

[Wang *et al.*, 2024] Ying Wang, Yingji Li, Yue Wu, and Xin Wang. Exploring multiple hypergraphs for heterogeneous graph neural networks. *Expert Systems with Applications*, 236:121230, 2024.

[Yu *et al.*, 2021] Zhouxin Yu, Feng Huang, Xiaohan Zhao, Wenjie Xiao, and Wen Zhang. Predicting drug–disease associations through layer attention graph convolutional network. *Briefings in bioinformatics*, 22(4):bbaa243, 2021.

[Yuan *et al.*, 2024] Yongna Yuan, Jiahui Liu, Xiaohang Pan, Ruisheng Zhang, and Wei Su. Heterogenous biological network multi-task learning model for ncrna-disease-drug association prediction. *Knowledge-Based Systems*, 300:112222, 2024.

[Zeng *et al.*, 2019] Xiangxiang Zeng, Siyi Zhu, Xiangrong Liu, Yadi Zhou, Ruth Nussinov, and Feixiong Cheng. deepdr: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics*, 35(24):5191–5198, 2019.

[Zhang *et al.*, 2018] Wen Zhang, Xiang Yue, Weiran Lin, Wenjian Wu, Ruoqi Liu, Feng Huang, and Feng Liu. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics*, 19(1):233, 2018.

[Zhao *et al.*, 2023] Bo-Wei Zhao, Xiao-Rui Su, Yue Yang, Dong-Xu Li, Guo-Dong Li, Peng-Wei Hu, Yong-Gang Zhao, and Lun Hu. Drug-disease association prediction using semantic graph and function similarity representation learning over heterogeneous information networks. *Methods*, 220:106–114, 2023.

[Zhu *et al.*, 2025] Enqiang Zhu, Xiang Li, Chanjuan Liu, and Nikhil R Pal. Boosting drug-disease association prediction for drug repositioning via dual-feature extraction and cross-dual-domain decoding. *Journal of Chemical Information and Modeling*, 65(11):5745–5757, 2025.