BMC Biology

RESEARCH

Open Access

HNF-DDA: subgraph contrastive-driven transformer-style heterogeneous network embedding for drug–disease association prediction

Yifan Shang[1†], Zixu Wang[2†], Yangyang Chen[1], Xinyu Yang[1], Zhonghao Ren[1], Xiangxiang Zeng[1] and Lei Xu[3*]

Abstract

Background Drug–disease association (DDA) prediction aims to identify potential links between drugs and diseases, facilitating the discovery of new therapeutic potentials and reducing the cost and time associated with traditional drug development. However, existing DDA prediction methods often overlook the global relational information provided by other biological entities, and the complex association structure between drug diseases, limiting the potential correlations of drug and disease embeddings.

Results In this study, we propose HNF-DDA, a subgraph contrastive-driven transformer-style heterogeneous network embedding model for DDA prediction. Specifically, HNF-DDA adopts all-pairs message passing strategy to capture the global structure of the network, fully integrating multi-omics information. HNF-DDA also proposes the concept of subgraph contrastive learning to capture the local structure of drug-disease subgraphs, learning the high-order semantic information of nodes. Experimental results on two benchmark datasets demonstrate that HNF-DDA outperforms several state-of-the-art methods. Additionally, it shows superior performance across different dataset splitting schemes, indicating HNF-DDA's capability to generalize to novel drug and disease categories. Case studies for breast cancer and prostate cancer reveal that 9 out of the top 10 predicted candidate drugs for breast cancer and 8 out of the top 10 for prostate cancer have documented therapeutic effects.

Conclusions HNF-DDA incorporates all-pairs message passing and subgraph capture strategies into heterogeneous network embedding, enabling effective learning of drug and disease representations enriched with heterogeneous information, while also demonstrating significant potential for applications in drug repositioning.

Keywords Drug–disease association prediction, Drug repositioning, Heterogeneous Network, Contrastive learning, Transformer

†Yifan Shang and Zixu Wang contributed equally to this work.

*Correspondence:
Lei Xu
csleixu@szpu.edu.cn
Full list of author information is available at the end of the article

Background

The development of a small molecule drug from design to market approval typically takes an average of 15 years and approximately $2 billion in investment [1]. In addition to the high costs of research and development, the clinical trial phase for new drugs has a failure rate as high as 90% [2]. Consequently, the number of newly approved drugs is insufficient to meet the needs of treating increasingly complex diseases. In 2022, the FDA approved only 37 novel drugs [3, 4]. New strategies are urgently needed to reduce costs and shorten the development cycle to enhance drug discovery efficiency. Compared to traditional drug discovery methods, drug repositioning identifies new indications for already approved clinical drugs, thereby avoiding the complex and costly drug design process and the high failure rates associated with clinical trials. Drug repositioning significantly improves drug discovery efficiency [5–11] and has been successfully applied in treating diseases such as COVID-19 [12] and Alzheimer's disease [13]. With the advancement of computer technology and the massive accumulation of biomedical data, many computational methods have been applied in the field of biomedicine [14–20], which computational virtual screening for new drug indications has gradually gained attention [21, 22]. Utilizing machine learning models to predict reliable potential drug–disease associations (DDAs) can substantially reduce the human and material costs associated with traditional experiments [23–25]. Therefore, computational methods for predicting DDAs have become crucial for accelerating drug discovery.

Computational methods for predicting DDAs can be categorized into two types: drug-based and disease-based methods and multi-source heterogeneous data-based methods [26]. The first type predicts potential DDAs by constructing a drug-disease bipartite network and leveraging known drug–disease association patterns [27–31]. For instance, NCH-DDA [27] employs single-neighborhood and multi-neighborhood feature extraction modules to extract critical features of drugs and diseases from both the drug-disease bipartite network and drug/disease similarity networks in parallel, utilizing contrastive learning to obtain common features. DRAGNN [28] uses a graph attention mechanism to obtain dynamically allocated attention coefficients for nodes, enhancing the effectiveness of information gathering for target nodes. However, these methods have a limitation: the mechanisms of drug action and disease pathology involve multiple biomolecules and signaling pathways. By focusing solely on the direct associations between drugs and diseases, these methods neglect the biological mechanisms involving other entities, such as proteins, in DDAs.

The second category, multi-source heterogeneous data-based methods, integrates data from various biological entities to capture potential associations between drugs and diseases. These methods can be divided into three types based on data integration strategies: path-based, network embedding-based, and knowledge graph embedding-based. Path-based methods use walk strategies, such as random walks, to generate node sequences that capture relationships between different types of nodes and edges, thereby learning the representations of drug and disease nodes [32–34]. For example, DREAMwalk [32] proposed a "semantic multilayer association induction" method, which uses random walks guided by semantic information to generate node sequences populated by drugs and diseases. FuHLDR [25] obtains low-order features based on graph convolutional networks and high-order features based on meta-paths, then integrates these high-order and low-order representations to determine a comprehensive representation of drugs and diseases. However, these meta-path-based features often rely on local information and have limited ability to extract higher-order structures, making it difficult to capture the complex interaction mechanisms between drugs and diseases. Network embedding-based methods construct a heterogeneous network containing various biological entities and then use graph representation learning techniques to capture the network structure and learn node feature representations [35–39]. For example, PSGCN [35] proposed an end-to-end specific partner drug repositioning method based on graph convolutional networks. DDAGDL [24] incorporates complex biological information into the topology of heterogeneous networks, effectively learning smooth representations of drugs and diseases through an attention mechanism. These methods use graph convolutional networks (GCN) or graph attention networks (GAT) to integrate information from neighboring nodes but overlooks the all-pairs message passing between nodes [40]. Knowledge graph embedding-based methods view associations in the knowledge graph as transformations from source entities to target entities [41–43]. For example, RotatE [41] models relationships between entities as rotations in the complex plane. Although knowledge graph embedding techniques can map entities and relationships in the graph into a low-dimensional vector space, this representation method may lose some structural and semantic information.

Considering the limitations of the existing methods, we propose a subgraph contrastive-driven transformer-style heterogeneous network embedding model (HNF-DDA) for DDA prediction (Fig. 1). First, we construct a heterogeneous network encompassing various biological entities and employ the attribute information of

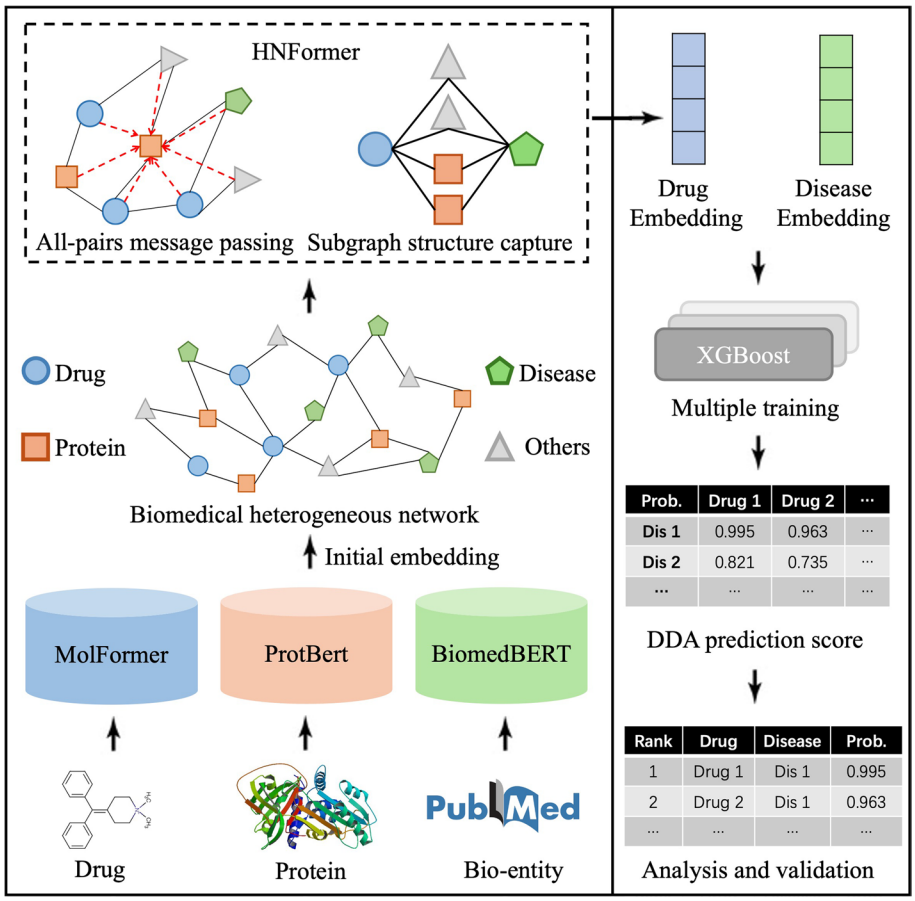

Fig. 1 The overview of the HNF-DDA framework. We input the SMILES of drugs, the sequences of proteins, and the textual descriptions of other biological entities into a biological language model to obtain the initial features of the nodes. The HNFormer module is then used to derive the embeddings of drugs and diseases. Next, we employ XGBoost for multiple independent training sessions. We average the predicted scores from these multiple runs and perform a ranking analysis based on the average scores

these entities to obtain initial node embeddings using a biological large language model. Second, to learn the embeddings of drugs and diseases, HNF-DDA employs an all-pairs message passing heterogeneous network embedding model to capture global signal transmission between any nodes. A subgraph capture strategy is proposed to extract high-order semantic structures within the heterogeneous network. Finally, an eXtreme Gradient Boosting (XGBoost) classifier [44, 45] is employed to predict the association probabilities between drugs and diseases. Experiments conducted on real-world datasets demonstrate that HNF-DDA outperforms existing methods in AUROC, AUPR, and Accuracy. Results from experiments with different dataset splitting schemes indicate that HNF-DDA has superior generalization capability for new drug and disease categories. Therefore, HNF-DDA not only effectively learns the representations of drugs and diseases that contain heterogeneous information but also shows greater potential for application in drug repositioning. This study makes the following contributions:

- To obtain multi-source biological entity information, we employ a large-scale biological language model to generate initial embeddings for drug structures, protein sequences, diseases, and other biological entity attributes.
- To achieve global information transmission in heterogeneous networks, we utilize an all-pairs message-passing Transformer-style network embedding model that simulates signal transmission between any nodes, enabling adaptive integration of various biological entity information.
- To better capture the complex association mechanisms between drugs and diseases, we propose a drug-disease subgraph contrastive strategy that ensures better connections between drugs and diseases in the embedding space.

- Experimental results demonstrate that HNF-DDA outperforms state-of-the-art methods. Additionally, split experiment results and case studies on breast cancer and prostate cancer confirm the model's generalization and reliability, offering new insights for drug repositioning.

Results and discussions

Datasets

We evaluated our model on two benchmark datasets: KEGG [46] and HetioNet [34]. Both datasets contain drug, protein, disease, pathway entities and multi-type association information. The statistics of the two datasets are shown in Table 1.

Baselines

In this study, we compared HNF-DDA with 10 state-of-the-art DDA prediction methods:

- *RotatE* [41]: This model introduces a new knowledge graph embedding method capable of modeling and inferring various relational patterns, including symmetric/antisymmetric, inversion, and composition, for learning drug and disease embeddings.
- *QuatE* [47]: This method introduces a more expressive hypercomplex representation to model entities and relationships in knowledge graph embeddings, learning node embeddings.
- *WalkPool* [48]: This algorithm combines the expressive power of topology-based heuristic algorithms with the feature learning capabilities of neural networks.
- *SEAL* [49]: The model proposes a novel decaying heuristic theory that unifies a broad range of heuristic algorithms within a single framework. It demonstrates that all these heuristic algorithms can be well-approximated from local subgraphs, which retain rich information about the existence of links.
- *ComplEx* [50]: This method demonstrates that using the asymmetric Hermitian product as a relational operation can automatically understand the structural knowledge of large knowledge bases and address the link prediction problem.
- *DTi2vec* [51]: The model constructs a heterogeneous network and employs node embedding techniques to automatically generate features for each drug and target, subsequently using ensemble learning techniques to identify drug-target interactions.
- *NEWMIN* [33]: This method proposes a network embedding framework within multiple networks to predict synergistic drug combinations.
- *DDAGDL* [24]: This method incorporates complex biological information into the topology of heterogeneous networks, effectively learning smooth representations of drugs and diseases through an attention mechanism.
- *DREAMwalk* [32]: This model proposes a "semantic multi-layer guilt-by-association" method, which predicts DDAs at the drug-gene-disease level using the relational guilt principle "similar genes share similar functions."
- *FuHLDR* [25]: This methods propose a novel graph representation learning model for drug repositioning by fusing higher and lower-order biological information.

The baselines we selected can be categorized into random walk-based, graph neural network-based, and knowledge graph-based link prediction models. The random walk-based models include FuHLDR, DREAMwalk, NEWMIN, and DTi2vec; the graph neural network-based models include DDAGDL, WalkPool, and SEAL; the knowledge graph embedding models, including ComplEx, RotatE, and QuatE.

Table 1 The statistics of KEGG and HetioNet

Datasets	Nodes	Numbers	Edges	Numbers
KEGG	Drugs	6008	Drug–disease	2272
	Diseases	1963	Disease–gene	6319
	Genes	14,496	Drug–gene	11,860
	Pathways	461	Gene–pathway	43,226
			Pathway–pathway	2129
			Disease–pathway	2573
			Drug–pathway	10,274
HetioNet	Drugs	1552	Drug–disease	755
	Diseases	137	Disease–gene	27,977
	Genes	20,945	Drug–gene	51,429
	Pathways	1822	Gene–Gene	474,526
			Gene–pathway	84,372

Experimental setting and evaluation metrics

We use known DDAs as positive samples and randomly sample an equal number of negative drug-disease pairs as negative samples. Then, we evaluate the performance of HNF-DDA and other methods on the two datasets using tenfold cross-validation repeated 10 times, with different dataset splits for each experiment. Since the dimensions of the initial features of different biological entities are different, we first convert the initial embedding to the same size of the input embedding, set the size of the input embedding to 512, and set the size of the hidden layer embedding to 32; the number of layers for the all-pair message-passing encoder is 2 for the KEGG dataset

and 1 for the HetioNet dataset; the weights of the learning objectives, α and β, are both 0.01 for KEGG, and 1.0 and 0.1 for HetioNet, respectively.

The evaluation metrics include AUROC, AUPR, and Accuracy.

$$TPR = Recall = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP is the number of samples correctly classified as positive, TN is the number of samples correctly classified as negative, FP is the number of samples incorrectly classified as positive, and FN is the number of samples incorrectly classified as negative. Accuracy is the proportion of all samples that are correctly predicted. AUROC is the Area Under the TPR-FPR Curve plotted at different thresholds, and AUPR is the Area Under the Precision-Recall Curve plotted at different thresholds. We comprehensively evaluate the performance of HNF-DDA using AUROC, AUPR, and Accuracy.

Performance comparison

To evaluate the performance of HNF-DDA, we compared it with baselines on two datasets. Figure 2 shows the results of HNF-DDA and other baselines using ten-fold cross-validation 10 times. HNF-DDA achieved an average accuracy of 0.8897, AUROC of 0.9507, and AUPR of 0.9491 on both biomedical heterogeneous network datasets, outperforming the best-performing baselines, DREAMwalk (average accuracy of 0.8704, AUROC of 0.9382, AUPR of 0.9353), and FuHLDR (average accuracy of 0.8823, AUROC of 0.9462, AUPR of 0.9445). Therefore, HNF-DDA outperformed all compared baselines, achieving higher performance across all evaluation metrics. It is noteworthy that the KEGG dataset contains more drug, disease, and DDA data than the HetioNet dataset. HNF-DDA's performance improvement over the DREAMwalk and FuHLDR model in KEGG (accuracy by 3.5% and 0.13%, AUROC by 2.2% and 0.185%, AUPR by 2.0% and 0.195%) and in HetioNet (accuracy by 0.8% and 1.59%, AUROC by 1.1% and 0.14%, AUPR by 0.9% and 0.13%).

The best baselines DREAMwalk and FuHLDR both generate meta-paths based on the idea of random walks, thereby capturing the topological information of nodes in the network. However, since random walks tend to frequently visit nodes that are close to each other, while the probability of visiting distant nodes is low, the captured topological structure is more biased towards locality. Moreover, random walks only rely on the topological structure of the network and cannot directly capture high-order semantics. The all-pair message-passing mechanism proposed by HNF-DDA can capture the potential relationship between any nodes and learn the global information of nodes in the network; the subgraph contrastive learning module proposed by HNF-DDA can capture the high-order semantic information of the drug-disease subgraph and learn the local information of nodes in the network; in addition, the individual attribute information of nodes is learned using a biological large language model. Therefore, HNF-DDA integrates multi-source heterogeneous information from multiple perspectives, captures the potential association between drugs and diseases, and improves the prediction performance of DDA.

Predictive potential for unknown drug/disease classes

To evaluate the effectiveness of the model in real-world drug repositioning scenarios, we compare HNF-DDA with DDAGDL, DREAMwalk, and FuHLDR model, the best-performing baselines, through DDA split prediction experiments on the KEGG dataset (as shown in Fig. 3). First, we classified all drug and disease entities: drugs are classified according to their ATC codes, and diseases are classified according to the highest MeSH (Medical Subject Headings) term category. Then, based on the categories of drugs or diseases, we divide the DDAs into train, validation, and test sets in an approximate ratio of 8:1:1. This forced the model to predict the DDA probability for unknown drug or disease categories. We repeated the division 10 times to ensure that the data sets differed in each split.

As shown in Fig. 3, DREAMwalk outperforms the DDAGDL and FuHLDR in the split experiment. DREAMwalk with average accuracy of 0.7818, AUROC of 0.8868, and AUPR of 0.8976. HNF-DDA with average accuracy of 0.7961, AUROC of 0.9014, and AUPR of 0.8935. In the disease split experiment, DREAMwalk with average accuracy of 0.6190, AUROC of 0.6900, and AUPR of 0.7009. HNF-DDA with average accuracy of 0.6648, AUROC of 0.7955, and AUPR of 0.7829. These results demonstrate that HNF-DDA has greater potential to accurately predict unknown drug or disease categories in real-world scenarios compared to DREAMwalk. Additionally, as shown in Fig. 3, the distribution of prediction results across 10 experiments indicates that HNF-DDA has better stability in prediction performance. Due to the insufficient number of

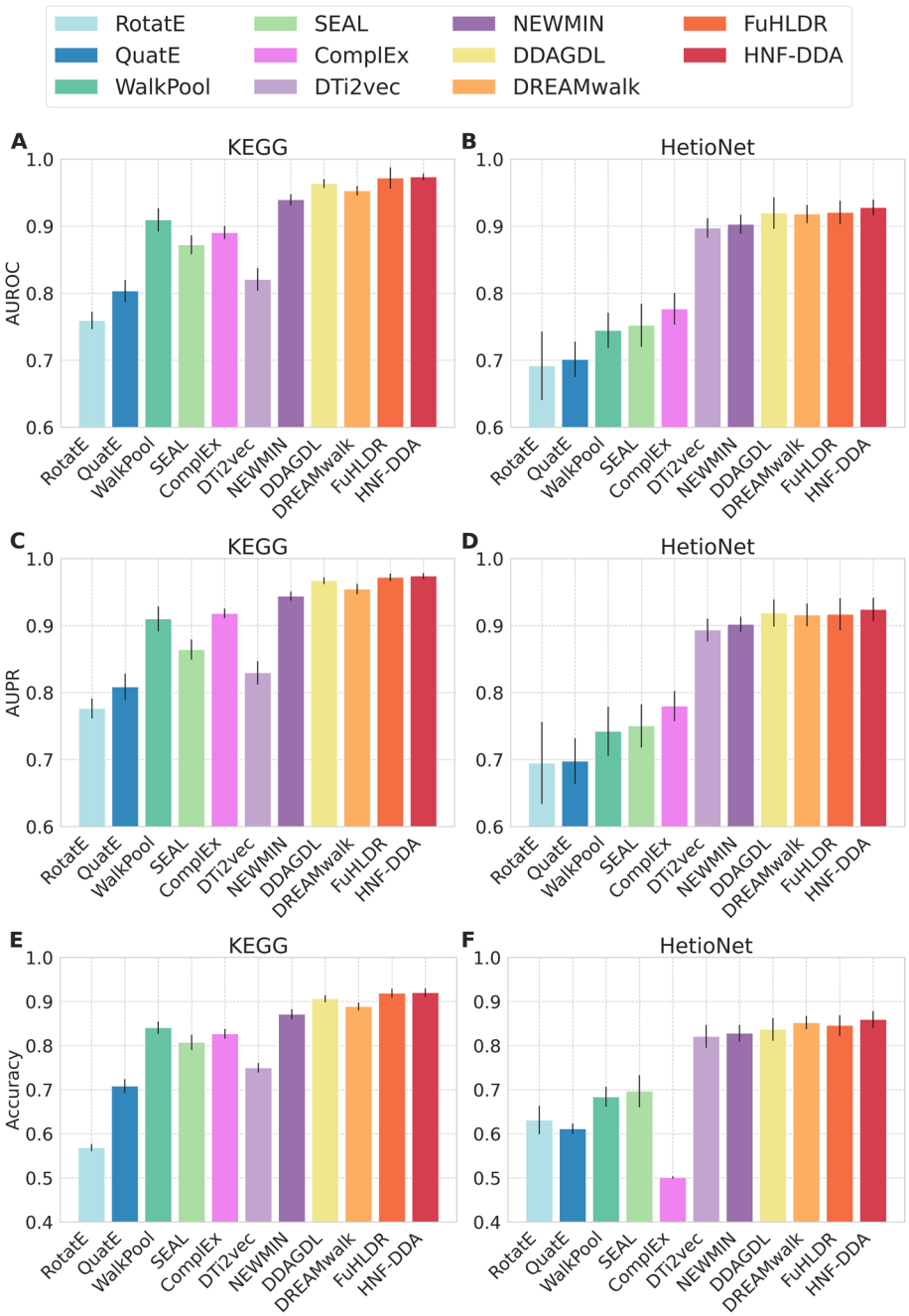

Fig. 2 The drug–disease association prediction performances of each model on the KEGG and HetioNet. **A** DDA prediction performance on KEGG dataset. **B** DDA prediction performance on HetioNet dataset. All methods are repeated 10 times, and the black short lines in the figures represent error bars. The specific experimental data sets can be found in Additional file 2

DDAs in the HetioNet dataset, which makes it difficult to perform category-based split experiments, no such experiments are conducted.

Ablation experiments

To comprehensively validate the predictive performance of the HNF-DDA model, we conducted ablation experiments. We created the following variants targeting the heterogeneous network encoder and learning objective modules of the HNF-DDA model:

- *w/o link*: Remove the edge-level learning objective.
- *w/o sub*: Remove the subgraph-level learning objective.

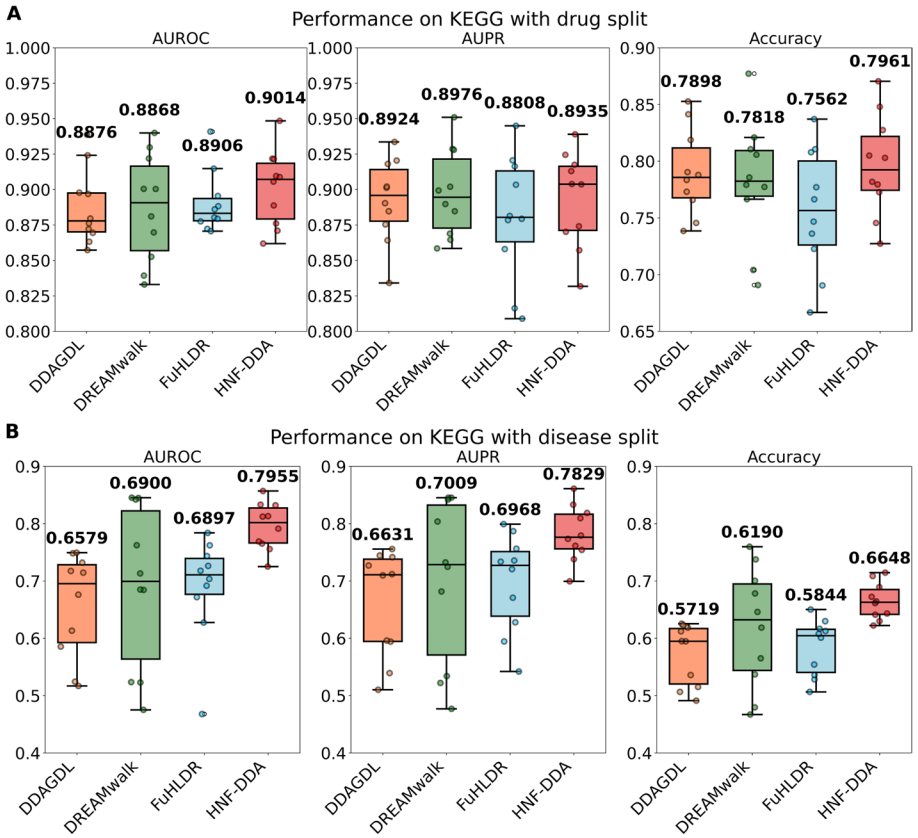

Fig. 3 Performance of DDAGDL, DREAMwalk, FuHLDR, and HNF-DDA based on split experiments in the dataset KEGG. **A** Performance of drug split experiment. **B** Performance of disease split experiment. Note: The bold numbers in the figure represent the average results of 10 experiments. The specific experimental data sets can be found in Additional file 2

- *w/o class*: Remove the node-level learning objective.
- *w/o init_feat*: Replace the initial embeddings learned by the large language model with one-hot encoding.
- *GAT*: Replace the all-pair message passing encoder with a GAT.
- *GCN*: Replace the all-pair message passing encoder with a GCN.

From the results in Fig. 4, HNF-DDA performs the best on the KEGG and HetioNet datasets. From the overall trend shown in Fig. 4, the changes in KEGG are relatively minor. As seen in Table 1, this is because the KEGG dataset contains a larger number of known DDA samples for training, which allows the model to fully learn the drug–disease association patterns. Therefore, the effects of different experimental conditions are minimal, and the performance metrics are higher than those of HetioNet.

Comparing the results of *w/o link*, *w/o sub*, *w/o class*, and *w/o init_feat*, it can be observed that the results of *w/o sub* and *w/o init_feat* are relatively worse. The *w/o sub* results indicate that the subgraph capture module proposed by HNF-DDA effectively mines the potential associations between drugs and diseases. The *w/o init_feat* results suggest that HNF-DDA effectively integrates semantic information of biological entities and heterogeneous network structure, and replacing the semantic features learned by the large language model may degrade the predictive performance. The comparison of *GAT* and *GCN* with HNF-DDA indicates that the all-pair message passing encoder used by HNF-DDA can effectively capture signals between any pair of nodes in the heterogeneous network, integrating multiple sources of heterogeneous information comprehensively and enhance the prediction performance. These experimental results demonstrate the effectiveness of the all-pair message passing and subgraph capture modules proposed by HNF-DDA.

Visualization of embeddings

We visualize the learned heterogeneous network node embeddings using T-sne [52]. Figure 5A and B show the visualization results of node embeddings on KEGG and HetioNet, respectively. Figure 5C and D show the visualization results of node embeddings on KEGG and HetioNet after removing the subgraph capture module. In

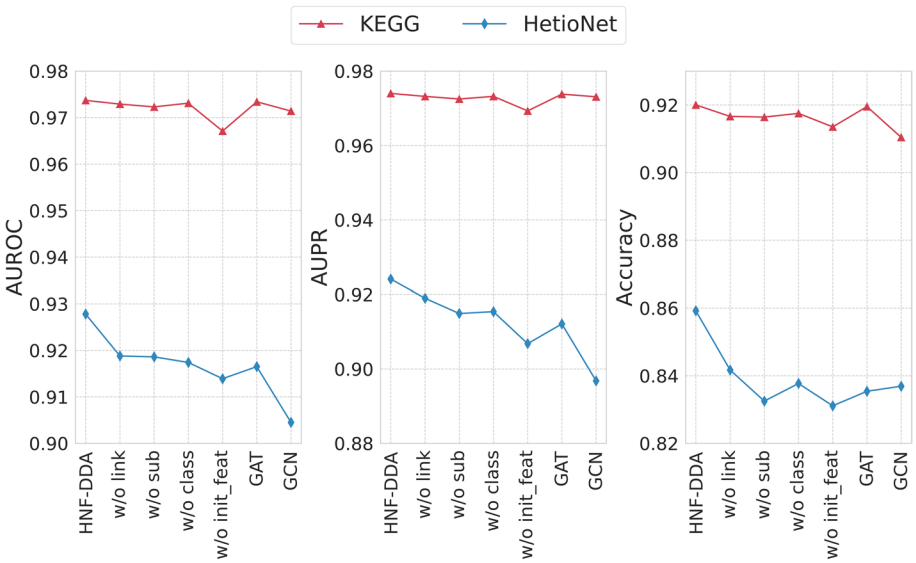

Fig. 4 Performance of HNF-DDA and different variants on KEGG and HetioNet datasets. The specific experimental data sets can be found in Additional file 2

Fig. 5A, drug clusters (red) are relatively close to disease clusters (blue) and relatively far from pathway clusters (yellow), in Fig. 5C, after removing the subgraph module, the pathway clusters (yellow) are situated between the drug clusters (red) and disease clusters (blue). Compared to Fig. 5B, D removing the subgraph module, drug clusters (red) are relatively close to pathway clusters (yellow) and relatively far from disease clusters (blue). These results indicate that the subgraph capture module can uncover potential associations between drugs and diseases, bringing them closer in the embedding space, which benefits the improvement of downstream DDA prediction performance.

Case study

To further validate the reliability of DDA predictions by HNF-DDA in drug repositioning, we conduct literature verification on candidate drugs for breast cancer and prostate cancer from the KEGG dataset. Firstly, we average the prediction scores of all DDAs obtained from tenfold cross-validation repeated 10 times, ensuring different data splits for each tenfold cross-validation. Then, we sort the predicted scores of all unknown DDAs. Finally, we select the top 10 candidate drugs with the highest predicted scores for breast cancer and prostate cancer for literature validation analysis. Table 2 lists candidate drugs and the corresponding literature reports.

As shown in Table 2, among the top 10 candidate drugs for breast cancer predicted by HNF-DDA, 9 have supporting literature evidence. For prostate cancer, 8 out of the top 10 candidate drugs have supporting literature evidence. Among the candidate drugs for breast cancer and prostate cancer, seven drugs overlap (including drugs for which there is no literature evidence). These drugs are either chemotherapy drugs that can treat various types of cancer (such as Etoposide and Vincristine sulfate) or can play a supportive role in managing cancer, or the symptoms and side effects related to its treatment. For example, Ephedrine is mainly used as a bronchodilator and decongestant and can sometimes be used in supportive care to manage low blood pressure during surgery or chemotherapy; Desmopressin is primarily used to treat diabetes insipidus and certain bleeding disorders but can also be used to manage bleeding complications in cancer patients; Prednisone, a corticosteroid, is used to treat various diseases, including inflammation, autoimmune diseases, and as part of certain chemotherapy regimens. The other three non-overlapping drugs also have specific related literature reports. These results further demonstrating the reliability of HNF-DDA in practical disease applications.

Discussion

HNF-DDA shows multiple improvements over existing SOTA models. It outperforms other models in prediction accuracy in different scenarios of both datasets, including predictions in new drug or disease categories. This study highlights the importance of using large language models and capturing both global and local structures of heterogeneous networks for DDA prediction. Although contrastive learning methods has shown creativity, the quality of negative samples limits the prediction performance.

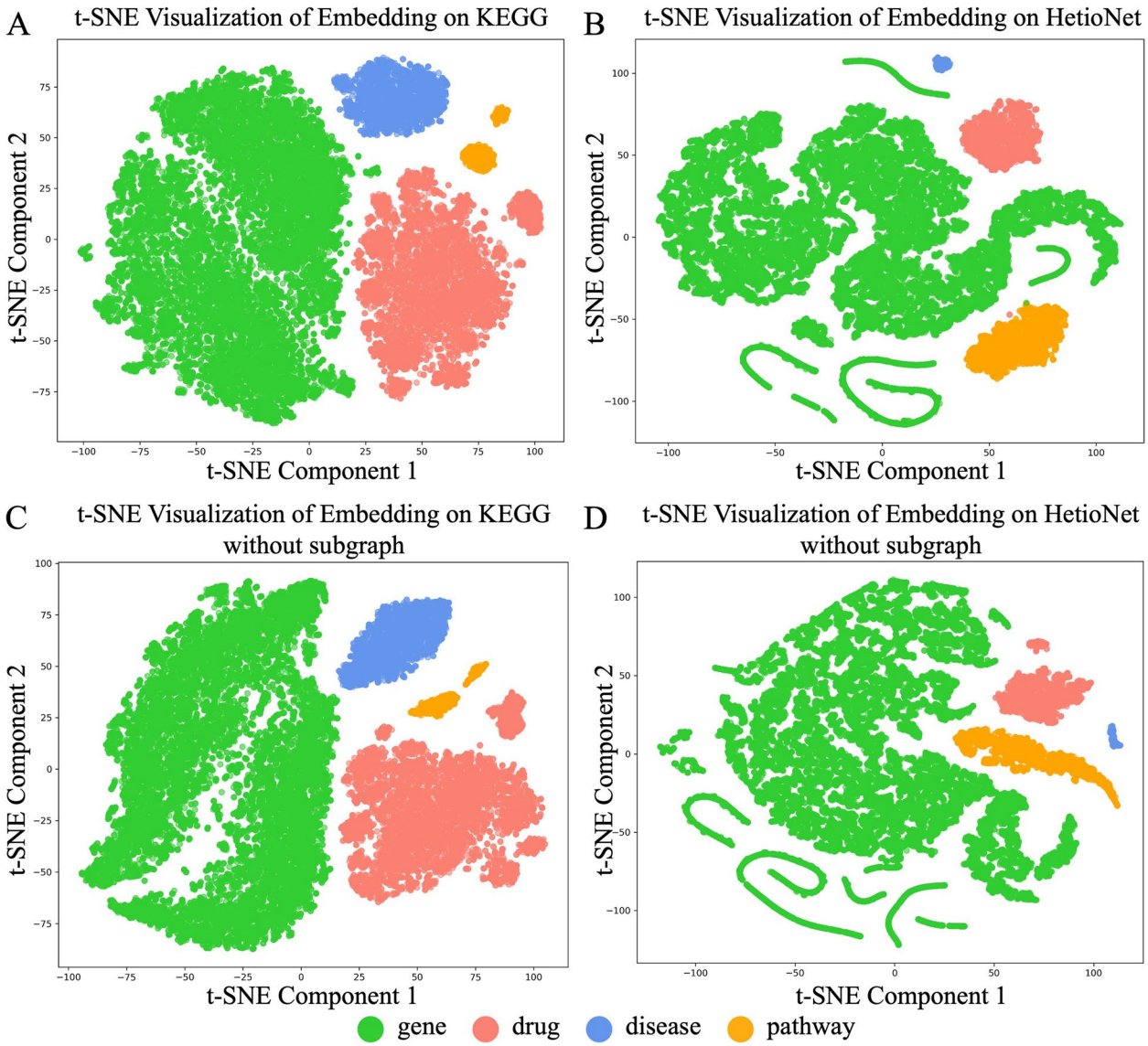

Fig. 5 HNF-DDA embedding visualization experiment. **A** Visualization on KEGG. **B** Visualization on HetioNet. **C** Visualization on KEGG without subgraph module. **D** Visualization on HetioNet without subgraph module

Although we use a subgraph capture strategy to preserve the local structure of nodes and learn the high-level semantic information of nodes, the subgraph negative samples obtained through the random replacement strategy have relatively low interference and discriminability. Additionally, the classifier used in HNF-DDA is also trained on drug-disease negative pairs generated by random sampling, which may result in false negative pairs. Therefore, in our future work, we plan to investigate sampling strategies for negative samples to obtain more realistic and reliable negative samples for more accurate DDA prediction.

Conclusions

We propose HNF-DDA, a subgraph contrast-driven transFormer-based Heterogeneous Network embedding model for predicting drug–disease associations (DDAs). HNF-DDA utilizes an all-pair message passing strategy to preserve the global information of heterogeneous network nodes, enabling the integration of multi-omics data. It also proposes a subgraph capture module to retain the local structure of drug-disease subgraphs, learning the multiple high-level semantic information. Experimental results on two benchmark datasets demonstrate that HNF-DDA outperforms 10 state-of-the-art methods. Dataset split experiments

Table 2 Top 10 candidates of HNF-DDA for breast cancer and prostate cancer

Datasets	Breast cancer		Prostate cancer	
Rank	Drug	Evidences	Drug	Evidences
1	Ephedrine	[53–55]	Somatropin	[56, 57]
2	Etoposide	[58–60]	Prednisone	[61–63]
3	Desmopressin	[64–66]	Doxorubicin hydrochloride	[67, 68]
4	Mupirocin calcium	[69]	Budesonide	
5	Vincristine sulfate	[70–72]	Desmopressin	[73–75]
6	Somatropin		Vinblastine sulfate	[76]
7	Budesonide	[77]	Ephedrine	
8	Cortisone acetate	[78]	Etoposide	[79–81]
9	Pirbuterol	[82]	Paclitaxel	[83–85]
10	Prednisone	[86–88]	Fluorouracil	[89–91]

reveal HNF-DDA's potential in predicting DDAs for novel drug or disease categories. Ablation and visualization experiments indicate that the proposed all-pairs message passing and subgraph capturing strategies effectively reveal latent associations between drugs and diseases, enhancing DDA prediction performance. Finally, a literature validation analysis of the top 10 candidate drugs for breast and prostate cancer confirms the reliability of HNF-DDA in identifying candidate drugs. In summary, our model, HNF-DDA, offers a powerful tool for drug-disease prediction.

Methods

The main objective of this paper is to predict associations between drugs and diseases. We propose a subgraph contrastive-driven transFormer-style Heterogeneous Network embedding model, HNF-DDA, as illustrated in Fig. 1. First, we construct a biomedical heterogeneous network and utilize a biological language model to learn the initial embeddings of the heterogeneous network nodes. Next, we learn the embeddings of drugs and diseases using all-pairs message passing and subgraph capture strategies. Finally, we employ an XGBoost classifier to predict the association probabilities between drugs and diseases. Table 3 is a summary of all notations used in the "Methods" section.

Biomedical heterogeneous network

In this study, we construct a biomedical heterogeneous network using the interactions between biological entities. The nodes represent drugs, diseases, proteins, and other biological entities, while the edges represent the relationships between these entities. A schematic diagram of this network is shown in Fig. 1.

A biomedical heterogeneous network is defined as an undirected network $G = (V, E, \mathcal{A}, \mathcal{R})$ and $N = |V|$ represent number of nodes. $V = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes in the network, where $v_i \in V$ represents the i node in the network. $E = \{(v_i, v_j)|v_i, v_j \in V\} \subseteq V \times V$ is the set of edges in the network, and each edge represents

Table 3 Summary of all notations

Symbol	Description	Symbol	Description
G	Biomedical heterogeneous network	H	Feature Embedding Matrix
V	Heterogeneous network node set	t	Node type
E	Heterogeneous network edge set	W	Weight matrix
$\mathcal{A}$	Node attribute set	B	Bias matrix
$\mathcal{R}$	Node type set	d	Embedding feature dimensions
N	Number of nodes	z	Node embedding feature vector
v_i	The i node of network	g	Sampled from Gumbel distribution
a_i	Attribute feature of node v_i	$\mathcal{T}$	Temperature coefficient
$\tau(\bullet)$	Node type mapping function	b	Learnable weight parameter
$\sigma(\bullet)$	Activation function	$\phi(\bullet)$	Positive Random Features (PRF)
$\widehat{Y}$	Predicted probability of the label	C	The intermediate node set
s	The ending node (disease/drug)	M	Subgraph node set

the interaction or association that exists between two nodes, where (v_i, v_j) represents the connection between nodes i and j. $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ is the set of attributes of a node, including the SMILES structure of drugs, protein sequences, and biological text descriptions of diseases and other biological entities, where $a_i \in \mathcal{A}$ represents the attribute feature associated with node v_i. $\mathcal{R}$ is the set of type of a node, we describe the type of each node by a mapping function $\tau : V \rightarrow R$, namely: $\tau(v_i) \in \{Drug, Protein, Disease, Others\}$, where $\tau(v_i)$ represents the type of node v_i. Additionally, we removed all DDAs from the biomedical heterogeneous network, allowing the integration of network structure and biological entity semantic information during the heterogeneous network embedding process without relying on drug-disease treatment information. These DDAs will serve as supervisory information for the DDA prediction task using the XGBoost classifier.

Computing initial embeddings

The biomedical heterogeneous network contains various types of biological entities. We employ specific model for each type of entity based on their attributes to extract semantic information, which serves as the initial embeddings for the nodes in the heterogeneous network. This approach integrates external knowledge with the structure of the heterogeneous network. These models utilized to compute the initial embeddings are outlined below:

- We utilize the SMILES as the attribute information for drugs. SMILES encodes the structure of a molecule into a string of characters, with each character representing information about atoms, bonds, and rings [92, 93]. This encoding provides a comprehensive description of the molecular structure, including the connections between atoms, ring structures, and stereochemistry. We employ a language model for drug molecules, MolFormer [94], to obtain embeddings from the drug's SMILES. MolFormer employs masked language modeling and combines linear attention Transformers with rotary embeddings.
- We utilize amino acid sequence data as the attribute information for proteins. This data comprises a sequence of characters that represent the specific amino acids constituting a protein and their sequential arrangement. Each amino acid is represented by a letter, and the sequence can reflect the protein's structure, function, and activity. We employ a pre-trained protein model, ProtBert [95], to obtain initial embeddings from the protein sequence data. ProtBert is based on the BERT [96] architecture and encodes amino acid sequences into token-level or sentence-level representations, which can be used for downstream protein tasks, such as contact prediction.
- We utilize biological text descriptions as attribute information for diseases and other biological entities. We employ a biomedical text language model, BiomedBERT [97], to obtain initial embeddings. BiomedBERT is based on the BERT architecture and is pre-trained from scratch using text abstracts from PubMed and full-text articles from PubMed Central as its corpus.

Finally, we obtain the initial embeddings $H^{init} = \{H_t^{init} \in \mathbb{R}^{|V_t| \times d_t}\}$, where $t \in \{Drug, Protein, Disease, Others\}$. Details of these models are in the Additional file 1.

Heterogeneous network embedding

This section introduces the heterogeneous network embedding (HNFormer) module of HNF-DDA. In this module, we employ a Transformer-style graph embedding method and design a subgraph capture strategy to learn the embeddings of heterogeneous network nodes.

All-pair message passing encoder

The mechanisms of drug action and disease pathology involve various types of biomolecules and signaling pathways. Additionally, the known edges in the biological heterogeneous network are incomplete, and many potential associations exist between nodes. Therefore, signal transmission in a heterogeneous network should not be limited to entities of the same type or local entity relationships. Inspired by NodeFormer [40], we employ an all-pairs message passing encoder to enable signal transmission between any pair of entities in the heterogeneous network, ensuring the full integration of multi-source heterogeneous information.

First, we utilize multiple MLPs to map the initial embeddings H^{init} of different type node into the same space:

$$H_t^0 = \sigma\left(W_t^{init} H_t^{init} + B_t^{init}\right), \tag{5}$$

where $H_t^0 \in \mathbb{R}^{|V_t| \times d}$, $t \in \{Drug, Protein, Disease, Others\}$, and $\sigma(\bullet)$, W_t^{init}, B_t^{init} represent Exponential Linear Units activation function, weight, bias parameter, respectively. We concatenate the embeddings of different types to form the complete node embeddings $H^0 \in \mathbb{R}^{N \times d}$:

$$H^0 = \begin{bmatrix} H_{Drug}^0 \\ H_{Protein}^0 \\ H_{Disease}^0 \\ H_{Others}^0 \end{bmatrix}, \tag{6}$$

where $z_u^0 \in H^0$ represents the u th node representation vector in the 0 layer, and 0 represents initial representation vector of heterogeneous network.

Next, for any node u, we use $z_u^{(l)}$ to represent its corresponding representation vector at layer l. Thus, the update for the next layer $z_u^{(l+1)}$ is:

$$z_u^{(l+1)} = \sum_{s=1}^{N} \frac{\exp((q_u)^T k_s)}{\sum_{w=1}^{N} \exp((q_u)^T k_w)} \bullet v_s, \tag{7}$$

where $k_u = W_K^{(l)} z_u^{(l)}$, $q_u = W_Q^{(l)} z_u^{(l)}$, $v_u = W_V^{(l)} z_u^{(l)}$ are obtained from the feature transformation of the l th layer, and $W_K^{(l)}$, $W_Q^{(l)}$, and $W_V^{(l)}$ are learnable parameters in l th layer. Equation (7) can be viewed as a graph attention mechanism defined on a fully connected graph where all nodes are pairwise connected.

Because for any node, it is necessary to calculate the attention of the other N nodes separately. Therefore, using a kernel method approximate the exponential-then-dot operation, which is $\exp(a^T b) = \kappa(a,b) \approx \phi(a)^T \phi(b)$, where $\phi : \mathbb{R}^d \to \mathbb{R}^m$ is a low-dimensional feature map (RF). For example, the commonly used Positive Random Feature (PRF) [98] can be defined as:

$$\phi(x) = \frac{\exp(\frac{-\|x\|_2^2}{2})}{\sqrt{m}} \left[\exp\left(w_1^T x\right), \ldots, \exp\left(w_m^T x\right)\right], \tag{8}$$

This enables us to rewrite Eq. (7):

$$z_u^{(l+1)} = \sum_{s=1}^{N} \frac{\phi(q_u)^T \phi(k_s)}{\sum_{w=1}^{N} \phi(q_u)^T \phi(k_w)} \bullet v_s = \frac{\phi(q_u)^T \sum_{s=1}^{N} \phi(k_s) v_s^T}{(q_u)^T \sum_{w=1}^{N} \phi(k_w)} \tag{9}$$

In this way, only one computation is needed, the total complexity is kept within $O(N)$.

The above process assumes that each edge has a continuous attention weight. To further consider the "discretization" of edges, for any node u, need to find an "optimal" set of neighbors in each layer for information passing. Therefore, treating the attention weights generated by N nodes as a categorical distribution and then sample the neighbor set from it. Specifically, replacing the Softmax in Eq. (7) with Gumbel-Softmax:

$$z_u^{(l+1)} = \sum_{s=1}^{N} \frac{\exp((q_u^T k_u + g_s)/\mathcal{T})}{\sum_{w=1}^{N} \exp((q_u^T k_w + g_w)/\mathcal{T})} \bullet v_s, g_u \sim Gumbel(0,1), \tag{10}$$

Then, following the before approximate using RF with linear complexity:

$$z_u^{(l+1)} \approx \sum_{s=1}^{N} \frac{\phi\left(q_u/\sqrt{\mathcal{T}}\right)^T \phi\left(k_s/\sqrt{\mathcal{T}}\right)^{e g_s/\mathcal{T}}}{\sum_{w=1}^{N} \phi\left(q_u/\sqrt{\mathcal{T}}\right)^T \phi\left(k_w/\sqrt{\mathcal{T}}\right)^{e g_w/\mathcal{T}}} \bullet v_s = \frac{\phi\left(q_u/\sqrt{\mathcal{T}}\right)^T \sum_{s=1}^{N} e^{g_s/\mathcal{T}} \phi(k_s/\sqrt{\mathcal{T}}) v_s^T}{\left(q_u/\sqrt{\mathcal{T}}\right)^T \sum_{w=1}^{N} e^{g_w/\mathcal{T}} \phi(k_w/\sqrt{\mathcal{T}})} \tag{11}$$

In addition to considering the message passing between all node pairs in the network, the topology of the heterogeneous network itself contains a lot of useful information. During each layer of message passing, to strengthen the weights of the observed edges. Therefore, assigning a shared learnable weight to each edge, referred to as relational bias, and update the formula for each layer as follows:

$$z_u^{(l+1)} \leftarrow z_u^{(l+1)} + \sum_{s, a_{us}=1} \sigma(b^{(l)}) \bullet v_s, \tag{12}$$

where $b^{(l)}$ is the learnable weight parameter corresponding to layer l, $a_{us=1}$ indicates that there is an association between nodes u and s. We can obtain the last layer of node u embeddings $z_u \in H$ based on all-pair message passing.

Finally, we employ an MLP to predict the labels of the nodes:

$$\widehat{Y} = MLP(H) \tag{13}$$

where $\widehat{Y} \in \mathbb{R}^{N \times |\mathcal{R}|}$ represent the predicted probability of the label.

Subgraph structure capture

Drugs act on multiple target proteins and participate in various functional pathways, working together to treat diseases. Therefore, in a heterogeneous network, the relationships between drugs, diseases, and other biological entities collectively form higher-order subgraph structures. In addition to considering signal transmission between nodes in the heterogeneous network, it is crucial to preserve the contextual semantic information contained in the high-order structures of the heterogeneous network. While existing methods often rely on meta-path approaches to explore high-order structures, they may fall short in capturing rich semantics and extracting high-order patterns [99]: (1) Meta paths often focus on single relationships, ignoring the multiple associations between different entities; (2) Starting from a source node, the number of nodes that a meta path can reach is too large, resulting in the extracted structure lacking restrictions and containing insufficient semantic information.

Inspired by CPT-HG [99], we recognize that the mechanisms of drug action and disease pathology involve

various types of biomolecules and signaling pathways. Consequently, drugs, diseases, and other biological entities collectively form subgraph with intricate high-order structures. To address this, we construct positive and negative subgraphs and leverage contrastive learning to capture the intricate subgraph structures and subtle contextual semantic information within the heterogeneous network.

Specifically, given a drug (disease) node as the starting node u and a disease (drug) node as the ending node s, we take the common first-order neighbors between the drug and the disease as the intermediate node set C. We construct subgraphs using only the first-order common neighbors of drugs and diseases as intermediate nodes, capturing the strong associations between drugs and diseases, and enhancing the structural constraints of the subgraph to include rich high-level semantic information. Therefore, the positive subgraph corresponding to node u is:

$$M_u^+ = \{s\} \cup C \tag{14}$$

Then, we randomly replace half of the elements in the intermediate node set C to obtain a new set C^-, thus we can obtain the negative subgraph for node u:

$$M_u^- = \{s\} \cup C^- \tag{15}$$

Finally, we apply the concept of contrastive learning to ensure that node u is closer to its positive subgraph embedding and farther from its negative subgraph embedding. The subgraph-level loss objective is defined as:

$$\mathcal{L}_{sub} = -\frac{1}{|V^{dd}|} \sum_{u \in V^{dd}} \frac{\exp\left(H_u f\left(M_u^+\right)\right)}{\exp\left(H_u f\left(M_u^+\right)\right) + \exp\left(H_u f\left(M_u^-\right)\right)}, \tag{16}$$

where V^{dd} is the set of drug and disease nodes, and $f(\bullet)$ denotes the pooling function (e.g., average pooling) that gets the subgraph embeddings.

Learning objectives

Given the node labels Y and the predicted labels $\widehat{Y}$, the node-level supervised loss is defined as:

$$\mathcal{L}_n = -\frac{1}{N} \sum_{v \in V} \sum_{r=1}^{|\mathcal{R}|} \mathbb{I}[y_v = r] log \widehat{y_{v,r}}, \tag{17}$$

where $\mathbb{I}[\bullet]$ is an indicator function. $\widehat{y_{v,r}}$ represent the probability that the v th node belongs to the class r.

Treating the attention estimates of each layer in the model as a categorical distribution, with the observed edges as samples. Thus, we define an edge-level loss objective using maximum likelihood estimation:

$$\begin{aligned} \mathcal{L}_e &= -\frac{1}{NL} \sum_{l=1}^{L} \sum_{(u,s) \in E} \frac{1}{d_u} log \pi_{us}^{(l)} \\ \pi_{us}^{(l)} &= \frac{\phi(q_u)^T \phi(k_s)}{\phi(q_u)^T \sum_{w=1}^{N} \phi(k_w)} \end{aligned}, \tag{18}$$

where d_u represents the in-degree of node u, and $\pi_{us}^{(l)}$ represents the predicted probability for edge (u, s) at the l th layer of the model.

The final objective is the sum of the node-level, edge-level, and subgraph-level loss:

$$\mathcal{L} = \mathcal{L}_n + \alpha \mathcal{L}_e + \beta \mathcal{L}_{sub}, \tag{19}$$

where α and β are weight parameters.

Drug–disease association prediction

After obtaining the embeddings of the heterogeneous network nodes, we utilize the known DDAs as supervision information and predict DDA scores based on drug and disease embeddings using the XGBoost model. To enhance the stability of the prediction results, we conduct multiple independent training sessions with XGBoost and average the resulting prediction scores.

XGBoost (eXtreme Gradient Boosting) is a powerful and widely used machine learning algorithm, primarily designed for supervised learning tasks such as classification and regression [100]. XGBoost is an implementation of gradient boosting machines (GBM), which are a type of ensemble learning method. Ensemble learning methods combine multiple base learners (in this case, decision trees) to improve overall performance. Gradient boosting, specifically, builds models sequentially, where each new model attempts to correct the errors made by the previous models. XGBoost has gained popularity due to its high performance, speed, and scalability. Details of the training procedure and the parameters of XGBoost are in the Additional file 1.

Abbreviations

AUROC	The area under the receiver operating characteristics curve
AUPR	The area under the precision and recall curve
DDA	Drug–disease association
GAT	Graph attention networks
GCN	Graph convolutional networks
XGBoost	EXtreme Gradient Boosting

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02206-x.

Additional file 1. Biological large language model details, algorithm S1, supplementary details, Tables S1, and Figures S1. Biological large language model details: Details of the biological large language model used to extract initial features of biological entities. Algorithm S1: Algorithm training procedure and complexity analysis. Supplementary details: Compared Methods Detail. Fig. S1: 1st-order vs 2nd-order prediction performance. Table S1: Parameter Settings of XBGoost in HetioNet and KEGG datasets, respectively.

Additional file 2. Figures experiments data. This file contains the specific data of Figs. 2, 3 and 4.

Acknowledgements
We are very much indebted to the anonymous reviewers, whose constructive comments are very helpful for this paper.

Authors' contributions
Yifan Shang and Zixu Wang design experiments as well as writing manuscripts. Yangyang Chen, Xinyu Yang and Zhonghao Ren draw fgures and analyse the results. Xiangxiang Zeng, Lei Xu revise the manuscript. Lei Xu and Yifan Shang provide fnancial help with experiments as well as revising papers. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (Nos. 62422113, 62271329, 62402166); Shenzhen Science and Technology Program (20231129091450002); Key Field of Department of Education of Guangdong Province (2022ZDZX2082); Natural Science Foundation of Hunan Province (2024JJ6158).

Data availability
All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The Python and Torch implementation of the HNF-DDA model is accessible at https://doi.org/https://doi.org/10.5281/zenodo.15117258 or https://github.com/ShangCS/HNF-DDA. The sources of all analyzed datasets are as follows: KEGG dataset can be downloaded from https://www.kegg.jp/kegg/rest/keggapi.html, HetioNet dataset can be downloaded from https://github.com/hetio/hetionet. Additionally, the dataset used in this study is available at Zenodo: https://doi.org/https://doi.org/10.5281/zenodo.15117258. The specific data of Figs. 2, 3 and 4 can be found in Additional file 2.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China. [2]Department of Computer Science, University of Tsukuba, Tsukuba 305-8577, Japan. [3]School of Electronic and Communication Engineering, Shenzhen Polytechnic University, Shenzhen 518055, China.

Received: 15 November 2024 Accepted: 3 April 2025
Published online: 16 April 2025

References

1. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. Nature. 2023;616(7958):673–85.
2. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022;12(7):3049–62.
3. Mullard A. 2021 FDA approvals. Nat Rev Drug Discov. 2022;21(2):83–8.
4. Qi R, Zou Q. Trends and potential of machine learning and deep learning in drug study at single-cell level. Research. 2023;6:0050.
5. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58.
6. Jourdan J-P, Bureau R, Rochais C, Dallemagne P. Drug repositioning: a brief overview. J Pharm Pharmacol. 2020;72(9):1145–51.
7. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(8):673–83.
8. Ru X, Ye X, Sakurai T, Zou Q. NerLTR-DTA: drug–target binding affinity prediction based on neighbor relationship and learning to rank. Bioinformatics. 2022;38(7):1964–71.
9. Li H, Liu B. BioSeq-Diabolo: Biological sequence similarity analysis using Diabolo. PLoS Comput Biol. 2023;19(6):e1011214.
10. Ai C, Yang H, Ding Y, Tang J, Guo F. Low rank matrix factorization algorithm based on multi-graph regularization for detecting drug-disease association. IEEE/ACM Trans Comput Biol Bioinform. 2023;20(5):3033–43.
11. Zhao BW, Su XR, Hu PW, Huang YA, You ZH, Hu L. iGRLDTI: an improved graph representation learning method for predicting drug–target interactions over heterogeneous biological information network. Bioinformatics. 2023;39(8):btad451.
12. von Delft A, Hall MD, Kwong AD, Purcell LA, Saikatendu KS, Schmitz U, et al. Accelerating antiviral drug discovery: lessons from COVID-19. Nat Rev Drug Discov. 2023;22(7):585–603.
13. Ballard C, Aarsland D, Cummings J, O'Brien J, Mills R, Molinuevo JL, et al. Drug repositioning and repurposing for Alzheimer disease. Nat Rev Neurol. 2020;16(12):661–73.
14. Liu T, Qiao H, Wang Z, Yang X, Pan X, Yang Y, et al. CodLncScape provides a self-enriching framework for the systematic collection and exploration of coding LncRNAs. Adv Sci. 2024;11:2400009.
15. Ru X, Zou Q, Lin C. Optimization of drug–target affinity prediction methods through feature processing schemes. Bioinformatics. 2023;39(11):btad615.
16. Li H, Pang Y, Liu B. BioSeq-BLM: a platform for analyzing DNA, RNA, and protein sequences based on biological language models. Nucleic Acids Res. 2021;49(22):e129.
17. Li X, Ma S, Xu J, Tang J, He S, Guo F. TranSiam: Aggregating multi-modal visual features with locality for medical image segmentation. Expert Syst Appl. 2024;237:121574.
18. Guo X, Huang Z, Ju F, Zhao C, Yu L. Highly accurate estimation of cell type abundance in bulk tissues based on single-cell reference and domain adaptive matching. Adv Sci. 2024;11(7):2306329.
19. Su R, Wu H, Xu B, Liu X, Wei L. Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data. IEEE/ACM Trans Comput Biol Bioinform. 2019;16(4):1231–9.
20. Su R, Liu X, Wei L, Zou Q. Deep-resp-forest: a deep forest model to predict anti-cancer drug response. Methods. 2019;166:91–102.
21. Luo H, Li M, Yang M, Wu F-X, Li Y, Wang J. Biomedical data and computational models for drug repositioning: a comprehensive review. Brief Bioinform. 2021;22(2):1604–19.
22. Pang C, Qiao J, Zeng X, Zou Q, Wei L. Deep generative models in de novo drug molecule generation. J Chem Inf Model. 2023;64(7):2174–94.
23. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-Pro: support bio-sequence machine for proteins. Sci China Inform Sci. 2024;67(11):212106.
24. Zhao BW, Su XR, Hu PW, Ma YP, Zhou X, Hu L. A geometric deep learning framework for drug repositioning over heterogeneous information networks. Brief Bioinform. 2022;23(6):bbac384.
25. Zhao B-W, Wang L, Hu P-W, Wong L, Su X-R, Wang B-Q, et al. Fusing higher and lower-order biological information for drug repositioning via graph representation learning. IEEE Trans Emerg Top Comput. 2023;12(1):163–76.
26. Yang X, Niu Z, Liu Y, Song B, Lu W, Zeng L, et al. Modality-DTA: multimodality fusion strategy for drug–target affinity prediction. IEEE/ACM Trans Comput Biol Bioinform. 2022;20(2):1200–10.
27. Zhang P, Che C, Jin B, Yuan J, Li R, Zhu Y. NCH-DDA: Neighborhood contrastive learning heterogeneous network for drug–disease association prediction. Expert Syst Appl. 2024;238:121855.
28. Meng Y, Wang Y, Xu J, Lu C, Tang X, Peng T, et al. Drug repositioning based on weighted local information augmented graph neural network. Brief Bioinform. 2024;25(1):bbad431.
29 Yang K, Yang Y, Fan S, Xia J, Zheng Q, Dong X, et al. DRONet: effectiveness-driven drug repositioning framework using network embedding and ranking learning. Brief Bioinform. 2023;24(1):bbac518.
30. Gao Z, Ma H, Zhang X, Wang Y, Wu Z. Similarity measures-based graph co-contrastive learning for drug–disease association prediction. Bioinformatics. 2023;39(6):btad357.

31. Yang R, Fu Y, Zhang Q, Zhang L. GCNGAT: Drug-disease association prediction based on graph convolution neural network and graph attention network. Artif Intell Med. 2024:102805.
32. Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun. 2023;14(1):3570.
33. Yu L, Xia M, An Q. A network embedding framework based on integrating multiplex network for drug combination prediction. Brief Bioinform. 2022;23(1):bbab364.
34. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife. 2017;6:e26726.
35. Sun X, Wang B, Zhang J, Li M. Partner-specific drug repositioning approach based on graph convolutional network. IEEE J Biomed Health Inform. 2022;26(11):5757–65.
36. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug–disease associations through layer attention graph convolutional network. Brief Bioinform. 2021;22(4):bbaa243.
37. Wang Z, Zhou M, Arnold C. Toward heterogeneous information fusion: bipartite graph convolutional networks for in silico drug repurposing. Bioinformatics. 2020;36(Supplement_1):i525–33.
38. Yan K, Lv H, Guo Y, Peng W, Liu B. sAMPpred-GAT: prediction of antimicrobial peptide by graph attention network and predicted peptide structure. Bioinformatics. 2023;39(1):btac715.
39. Chen Y, Wang J, Wang C, Zou Q. AutoEdge-CCP: a novel approach for predicting cancer-associated CircRNAs and drugs based on automated edge embedding. PLoS Comput Biol. 2024;20(1):e1011851.
40. Wu Q, Zhao W, Li Z, Wipf DP, Yan J. Nodeformer: a scalable graph structure learning transformer for node classification. Adv Neural Inf Process Syst. 2022;35:27387–401.
41. Sun Z, Deng Z-H, Nie J-Y, Tang J. Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv preprint arXiv:190210197. 2019.
42. Yang J, Li Z, Wu WKK, Yu S, Xu Z, Chu Q, et al. Deep learning identifies explainable reasoning paths of mechanism of action for drug repurposing from multilayer biological network. Brief Bioinform. 2022;23(6):bbac469.
43. Su X, Hu L, You Z, Hu P, Zhao B. Attention-based knowledge graph representation learning for predicting drug-drug interactions. Brief Bioinform. 2022;23(3):bbac140.
44. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 04-2. 2015;1(4):1–4.
45. Zhu H, Hao H, Yu L. Identifying disease-related microbes based on multi-scale variational graph autoencoder embedding Wasserstein distance. BMC Biol. 2023;21(1):294.
46. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
47. Zhang S, Tay Y, Yao L, Liu Q. Quaternion knowledge graph embeddings. Adv Neural Inf Process Syst. 2019;32.
48. Pan L, Shi C, Dokmanić I. Neural link prediction with walk pooling. arXiv preprint arXiv:211004375. 2021.
49. Zhang Z, Tang J, Guo F. Complex detection in PPI network using genes expression information. Curr Proteomics. 2018;15(2):119–27.
50. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G, editors. Complex embeddings for simple link prediction. PMLR. 2016;2071–2080.
51. Thafar MA, Olayan RS, Albaradei S, Bajic VB, Gojobori T, Essack M, et al. DTi2Vec: drug–target interaction prediction using network embedding and ensemble learning. J Cheminform. 2021;13:1–18.
52. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11):339–51.
53. Chen D, Ma F, Liu XH, Cao R, Wu XZ. Anti-tumor effects of ephedrine and Anisodamine on Skbr3 human breast cancer cell line. Afr J Tradit Complement Altern Med. 2016;13(1):25–32.
54. Sioud F, Amor S, Toumia IB, Lahmar A, Aires V, Chekir-Ghedira L, et al. A new highlight of ephedra alata decne properties as potential adjuvant in combination with cisplatin to induce cell death of 4T1 breast cancer cells in vitro and in vivo. Cells. 2020;9(2):362.
55. Mohammed L, Mohammed R. Cytotoxic activity of ephedra alata extracts on human lymphocytes and breast cancer cell line in vitro. Iraqi J Sci. 2023;30:4210–22.
56. Fuhrman B, Barba M, Schünemann HJ, Hurd T, Quattrin T, Cartagena R, et al. Basal growth hormone concentrations in blood and the risk for prostate cancer: a case-control study. Prostate. 2005;64(2):109–15.
57. Stangelberger A, Schally AV, Djavan B. New treatment approaches for prostate cancer based on peptide analogues. Eur Urol. 2008;53(5):890–900.
58. Sledge GW Jr. Etoposide in the management of metastatic breast cancer. Cancer. 1991;67(S1):266–70.
59. Alpsoy A, Yasa S, Gündüz U. Etoposide resistance in MCF-7 breast cancer cell line is marked by multiple mechanisms. Biomed Pharmacother. 2014;68(3):351–5.
60. Atienza DM, Vogel CL, Trock B, Swain SM. Phase II study of oral etoposide for patients with advanced breast cancer. Cancer. 1995;76(12):2485–90.
61. Auchus RJ, Yu MK, Nguyen S, Mundle SD. Use of prednisone with abiraterone acetate in metastatic castration-resistant prostate cancer. Oncologist. 2014;19(12):1231–40.
62. Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone plus prednisone in metastatic, castration-sensitive prostate cancer. N Engl J Med. 2017;377(4):352–60.
63. Sartor O, Weinberger M, Moore A, Li A, Figg WD. Effect of prednisone on prostate-specific antigen in patients with hormone-refractory prostate cancer. Urology. 1998;52(2):252–6.
64. Ripoll GV, Garona J, Pifano M, Farina HG, Gomez DE, Alonso DF. Reduction of tumor angiogenesis induced by desmopressin in a breast cancer model. Breast Cancer Res Treat. 2013;142:9–18.
65. Garona J, Pifano M, Orlando UD, PAsTRIAN MB, Iannucci NB, Ortega HH, et al. The novel desmopressin analogue [V4Q5] dDAVP inhibits angiogenesis, tumour growth and metastases in vasopressin type 2 receptor-expressing breast cancer models. Int J Oncol. 2015;46(6):2335–45.
66. Weinberg RS, Grecco MO, Ferro GS, Seigelshifer DJ, Perroni NV, Terrier FJ, et al. A phase II dose-escalation trial of perioperative desmopressin (1-desamino-8-d-arginine vasopressin) in breast cancer patients. Springerplus. 2015;4:1–8.
67. David-Beabes GL, Overman MJ, Petrofski JA, Campbell PA, de Marzo AM, Nelson WG. Doxorubicin-resistant variants of human prostate cancer cell lines DU 145, PC-3, PPC-1, and TSU-PR1: characterization of biochemical determinants of antineoplastic drug sensitivity. Int J Oncol. 2000;17(6):1077–163.
68. Newling D. The use of adriamycin and its derivatives in the treatment of prostatic cancer. Cancer Chemother Pharmacol. 1992;30:S90–4.
69. Dagsuyu E, Yanardag R. Purification and characterization of thioredoxin reductase enzyme from commercial Spirulina platensis tablets by affinity chromatography and investigation of the effects of some chemicals and drugs on enzyme activity. Biotechnol Appl Biochem. 2024;71(1):176–92.
70. Ghosh S, Lalani R, Maiti K, Banerjee S, Bhatt H, Bobde YS, et al. Synergistic co-loading of vincristine improved chemotherapeutic potential of pegylated liposomal doxorubicin against triple negative breast cancer and non-small cell lung cancer. Nanomedicine. 2021;31:102320.
71. Katsumata K, Tomioka H, Kusama M, Aoki T, Koyanagi Y. Clinical effects of combination therapy with mitoxantrone, vincristine, and prednisolone in breast cancer. Cancer Chemother Pharmacol. 2003;52:86–8.
72. Chen J, Li S, Shen Q, He H, Zhang Y. Enhanced cellular uptake of folic acid–conjugated PLGA–PEG nanoparticles loaded with vincristine sulfate in human breast cancer. Drug Dev Ind Pharm. 2011;37(11):1339–46.
73. Sasaki H, Klotz LH, Sugar LM, Kiss A, Venkateswaran V. A combination of desmopressin and docetaxel inhibit cell proliferation and invasion mediated by urokinase-type plasminogen activator (uPA) in human prostate cancer cells. Biochem Biophys Res Commun. 2015;464(3):848–54.
74. Hoffman A, Sasaki H, Roberto D, Mayer MJ, Klotz LH, Venkateswaran V. Effect of combination therapy of desmopressin and docetaxel on prostate cancer cell du145 proliferation, migration and growth: MP83-17. J Urol. 2017;197(4):e1112–3.
75. Bass R, Roberto D, Wang DZ, Cantu FP, Mohamadi RM, Kelley SO, et al. Combining desmopressin and docetaxel for the treatment of castration-resistant prostate cancer in an orthotopic model. Anticancer Res. 2019;39(1):113–8.
76. Brady SF, Pawluczyk JM, Lumma PK, Feng D-M, Wai JM, Jones R, et al. Design and synthesis of a pro-drug of vinblastine targeted at treatment

of prostate cancer with enhanced efficacy and reduced systemic toxicity. J Med Chem. 2002;45(21):4706–15.
77. Collins D, Gaynor N, Conlon N, Gullo G, Eustace A, Crown J. Abstract P4–07–08: Budesonide and loperamide do not impact the cytotoxicity of neratinib or HER2-directed monoclonal antibodies in HER2+ breast cancer cell lines. Cancer Res. 2019;79(4_Supplement):P4–07–8-P4–8.
78. Lundgren S, Gundersen S, Klepp R, Lønning P, Lund E, Kvinnsland S. Megestrol acetate versus aminoglutethimide for metastatic breast cancer. Breast Cancer Res Treat. 1989;14:201–6.
79. Kamradt JM, Pienta KJ. Etoposide in prostate cancer. Expert Opin Pharmacother. 2000;1(2):271–5.
80. Pienta KJ, Lehr JE. Inhibition of prostate cancer growth by estramustine- and etoposide: evidence for interaction at the nuclear matrix. J Urology. 1993;149(6):1622–5.
81. Cattrini C, Capaia M, Boccardo F, Barboro P. Etoposide and topoisomerase II inhibition for aggressive prostate cancer: data from a translational study. Cancer Treat Res Commun. 2020;25:100221.
82. Carie A, Sebti S. A chemical biology approach identifies a beta-2 adrenergic receptor agonist that causes human tumor regression by blocking the Raf-1/Mek-1/Erk1/2 pathway. Oncogene. 2007;26(26):3777–88.
83. Obasaju C, Hudes GR. Paclitaxel and docetaxel in prostate cancer. Hematology/Oncol Clin. 2001;15(3):525–45.
84. Hua M-Y, Yang H-W, Chuang C-K, Tsai R-Y, Chen W-J, Chuang K-L, et al. Magnetic-nanoparticle-modified paclitaxel for targeted therapy for prostate cancer. Biomaterials. 2010;31(28):7355–63.
85. Kelly WK, Curley T, Slovin S, Heller G, McCaffrey J, Bajorin D, et al. Paclitaxel, estramustine phosphate, and carboplatin in patients with advanced prostate cancer. J Clin Oncol. 2001;19(1):44–53.
86. Bonnefoi H, Grellety T, Tredan O, Saghatchian M, Dalenc F, Mailliez A, et al. A phase II trial of abiraterone acetate plus prednisone in patients with triple-negative androgen receptor positive locally advanced or metastatic breast cancer (UCBG 12–1). Ann Oncol. 2016;27(5):812–8.
87. Marini G, Murray S, Goldhirsch A, Gelber R, Castiglione-Gertsch M, Price K, et al. The effect of adjuvant prednisone combined with CMF on patterns of relapse and occurrence of second malignancies in patients with breast cancer. Ann Oncol. 1996;7(3):245–50.
88. Wong NS, Buckman RA, Clemons M, Verma S, Dent S, Trudeau ME, et al. Phase I/II trial of metronomic chemotherapy with daily dalteparin and cyclophosphamide, twice-weekly methotrexate, and daily prednisone as therapy for metastatic breast cancer using vascular endothelial growth factor and soluble vascular endothelial growth factor receptor levels as markers of response. J Clin Oncol. 2010;28(5):723–30.
89. Atkins JN, Muss HB, Case LD, Frederick Richards I, Grote T, McFarland J. Leucovorin and high-dose fluorouracil in metastatic prostate cancer: a phase II trial of the Piedmont oncology association. Am J Clin Oncol. 1996;19(1):23–5.
90. Dewys WD, Begg CB, Brodovsky H, Creech R, Khandekar J. A comparative clinical trial of adriamycin and 5-fluorouracil in advanced prostatic cancer: prognostic factors and response. Prostate. 1983;4(1):1–11.
91. Swanson GP, Faulkner J, Smalley SR, Noble MJ, Stephens RL, O'Rourke TJ, et al. Locally advanced prostate cancer treated with concomitant radiation and 5-fluorouracil: Southwest oncology group study 9024. J UROLOGY. 2006;176(2):548–53.
92. Liu XW, Shi TY, Gao D, Ma CY, Lin H, Yan D, et al. iPADD: a computational tool for predicting potential antidiabetic drugs using machine learning algorithms. J Chem Inf Model. 2023;63(15):4960–9.
93. Yang Y, Gao D, Xie X, Qin J, Li J, Lin H, et al. DeepIDC: a prediction framework of injectable drug combination based on heterogeneous information and deep learning. Clin Pharmacokinet. 2022;61(12):1749–59.
94. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P. Large-scale chemical language representations capture molecular structure and properties. Nat Mach Intell. 2022;4(12):1256–64.
95. Ahmed E, Heinzinger M, Dallago C, Rihawi G, Wang Y, Jones L, et al. Prottrans: towards cracking the language of life's code through self-supervised learning. IEEE Trans Pattern Anal Mach Intell. 2021;44:7112–27.
96. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.
97. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthc. 2021;3(1):1–23.
98. Choromanski K, Likhosherstov V, Dohan D, Song X, Gane A, Sarlos T, et al. Rethinking attention with performers. arXiv preprint arXiv:200914794. 2020.
99. Jiang X, Lu Y, Fang Y, Shi C, editors. Contrastive pre-training of GNNs on heterogeneous graphs. CIKM. 2021;803–812.
100. Ma CY, Luo YM, Zhang TY, Hao YD, Xie XQ, Liu XW, et al. Predicting coronary heart disease in Chinese diabetics using machine learning. Comput Biol Med. 2024;169:107952.

Publisher's Note