BMC Biology

METHODOLOGY

Open Access

DrugDAGT: a dual-attention graph transformer with contrastive learning improves drug-drug interaction prediction

Yaojia Chen[1,5,6], Jiacheng Wang[2], Quan Zou[1,2], Mengting Niu[3,4], Yijie Ding[2], Jiangning Song[5,6*] and Yansu Wang[1*]

Abstract

Background Drug-drug interactions (DDIs) can result in unexpected pharmacological outcomes, including adverse drug events, which are crucial for drug discovery. Graph neural networks have substantially advanced our ability to model molecular representations; however, the precise identification of key local structures and the capture of long-distance structural correlations for better DDI prediction and interpretation remain significant challenges.

Results Here, we present DrugDAGT, a dual-attention graph transformer framework with contrastive learning for predicting multiple DDI types. The dual-attention graph transformer incorporates attention mechanisms at both the bond and atomic levels, thereby enabling the integration of short and long-range dependencies within drug molecules to pinpoint key local structures essential for DDI discovery. Moreover, DrugDAGT further implements graph contrastive learning to maximize the similarity of representations across different views for better discrimination of molecular structures. Experiments in both warm-start and cold-start scenarios demonstrate that DrugDAGT outperforms state-of-the-art baseline models, achieving superior overall performance. Furthermore, visualization of the learned representations of drug pairs and the attention map provides interpretable insights instead of black-box results.

Conclusions DrugDAGT provides an effective tool for accurately predicting multiple DDI types by identifying key local chemical structures, offering valuable insights for prescribing medications, and guiding drug development. All data and code of our DrugDAGT can be found at https://github.com/codejiajia/DrugDAGT.

Keywords Drug-drug interactions, Graph transformer, Attention, Interpretation

*Correspondence:
Jiangning Song
Jiangning.Song@monash.edu
Yansu Wang
wangyansu@uestc.edu.cn
Full list of author information is available at the end of the article

Background

Complex diseases are often treated with combinations of drugs to leverage their synergistic benefits, but unexpected drug-drug interactions (DDIs) can lead to adverse drug events (ADEs) [1, 2]. As the demand for multi-drug treatments grows, identifying drug interactions to minimize unforeseen ADEs becomes increasingly crucial. Traditionally, the clinical examination of all possible DDIs is time-consuming and costly [3, 4]. Consequently, computational methods, particularly machine learning, efficiently predict potential DDIs by analyzing established patterns [5–9].

Existing computational approaches can generally be categorized into network-based and chemical structure-based methods. Network-based methods integrate multiple sources to construct large-scale heterogeneous biological networks, encoding the chemical relationships between drugs into graphs or networks [10–17]. Advanced techniques such as graph embedding and knowledge graphs are then employed to predict new DDIs [18]. Graph embedding methods typically use network structures as input, employing random walks [19–21], matrix factorization [22–25], and neural networks [26–30] to learn node representations. Knowledge graph approaches integrate various entities (e.g., drugs, targets, side effects) and their relationships to construct structured knowledge, using models like graph convolutional and attention networks to extract higher-level semantic features for improved DDI prediction [31–33]. However, network-based methods often rely heavily on historical interactions or require additional biomedical knowledge, making them less suitable for drugs in early developmental stages that only have chemical structures available.

Conversely, chemical structure-based methods treat drugs as independent entities, predicting DDIs solely from drug pairs. DDIs are fundamentally determined by the interactions of important molecular substructures within the compounds. However, many models learn global representations with encoders without explicitly learning local interactions. For example, Molormer [34] first learns the global representation of a drug's entire structure, with mutual information only implicitly learned within a black-box decoding module. This reliance on global views restricts modeling precision and interpretability of predictions. Unlike earlier approaches, recent studies utilize various graph neural network (GNN) variants to extract features from drug chemical substructures effectively [35–38]. Specifically, SA-DDI and DGNN-DDI account for the varying sizes and shapes of crucial substructures within molecules, integrating a topology-based attention mechanism to enhance representation learning. However, a major limitation of these models is their difficulty in capturing long-range dependencies. They are constrained by the number of GNN layers and rely solely on aggregating information from neighboring nodes to learn drug representations. As a result, they struggle to fully understand the inherent complexity of drug molecular graphs, ultimately affecting their performance.

To address these limitations, we propose the interpretable dual-attention graph transformer-based model (DrugDAGT) for DDI prediction. Specifically, this model employs a graph transformer that utilizes bond attention to capture short-distance dependencies and atom attention for long-distance dependencies, thereby providing a comprehensive representation of local structures. Then, the encoded local representations are fed into the interaction-specific module, which extracts representations for explicitly learning about the local interactions between drug pairs. To enhance the model's ability to distinguish representations, we apply graph contrastive learning by introducing noise to generate different views of the drug and maximizing their similarity. Finally, our model employs a two-layer feed-forward network (FFN) to predict multiple types of DDIs. We perform an extensive performance evaluation of our method against baseline DDI methods in both warm-start and cold-start scenarios. Our findings indicate that our approach not only outperforms these state-of-the-art methods in terms of overall performance but also offers interpretable insights into the prediction outcomes.

Results and discussion

Problem formulation

In DDI prediction, the objective is to discover potential interactions between drug pairs, which may result in ADEs. Employing the simplified molecular-input line-entry system (SMILES) as input for the drug entails chemical atom and bond element information within the drug into a 1D sequence. However, 1D SMILES sequences inadequately capture molecular structures, potentially reducing model performance. Our model converts SMILES to 2D molecular graphs where atoms are nodes and bonds are edges. Such transformation enables GNNs to effectively capture key molecular features such as atomic hybridization and the number of covalent bonds, and as such, it can improve the comprehensive representation of drug properties. Given a pair of drug sequences d_i and d_j with an interaction type $r \in [0,86]$, DDI prediction endeavors to train a model M, which is capable of mapping the combined feature representation space $d_i \times d_j$ to an adverse drug event interaction probability score $p \in [0,1]$.

DrugDAGT framework

Our proposed DrugDAGT framework is illustrated in Fig. 1. Starting with the SMILES input for drug-drug pairs, DrugDAGT first utilizes a dual-attention graph transformer to integrate molecular graph representations by capturing both short- and long-range dependencies within the local structure. Following this initial feature extraction, the model explicitly learns the local interactions between molecules. To optimize the model, we employ graph contrastive learning in a regularized manner to boost the similarity across different views of molecular representations. Finally, a two-layer FFN module is established to predict the interaction probabilities of multiple DDIs in both warm-start and cold-start scenarios. Overall, the entire framework of DrugDAGT is underpinned by the labeled training data and end-to-end supervised learning strategy, thereby ensuring precise adaptability and efficacy in predicting DDIs.

Experimental setting

Dataset and metrics

We evaluate the model performance on the public dataset DrugBank [39], a comprehensive bioinformatics and cheminformatics resource that aggregates comprehensive drug data. It is sourced from FDA and Health Canada drug labels, covering 1706 drugs and 191,808 DDIs. These interactions are classified into 86 types, each detailing how one drug influences the metabolism

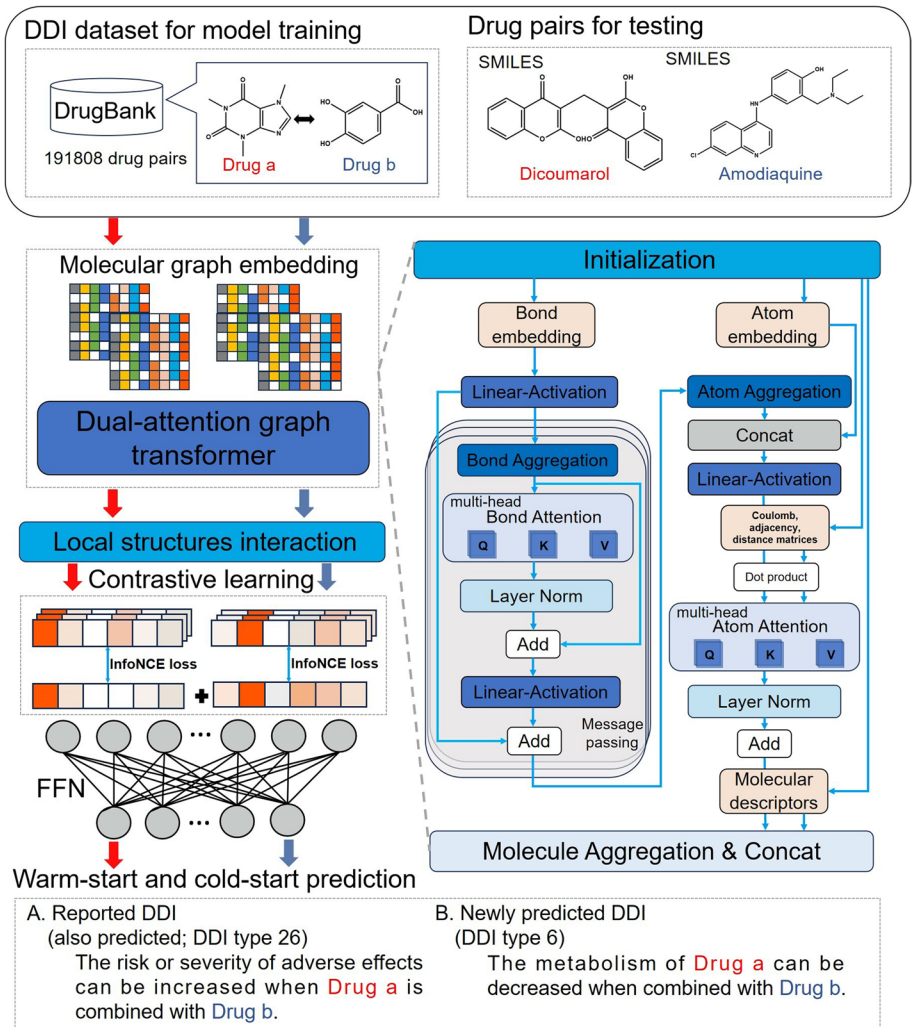

Fig. 1 Overview of the proposed DrugDAGT methodology. DrugDAGT first encodes drug pairs from both training and testing datasets into molecular graph embeddings using a dual-attention graph transformer to capture local structural features. It then processes these features to learn local interactions, thereby enhancing the drug representations via graph contrastive learning. Finally, the FFN module decodes these enhanced representations to predict DDI probabilities in both warm-start and cold-start scenarios

of another, with each pair linked to only one interaction type. Drugs are represented by SMILES, a text notation that describes chemical structures. Given the large size of our dataset, we employed the hold-out strategy to train our model, which helped enhance computational efficiency and ensure that the test set is never exposed to the training process, thus preserving its independence. We implement two dataset split strategies tailored for warm-start and cold-start scenarios. In the warm-start setting, the dataset is randomly divided with each drug in the test set potentially appearing in the training set as well. In contrast, the cold-start setting provides a more stringent and realistic evaluation, ensuring that all test drug pairs will not be observed during training. Both scenarios adhere to an 8:1:1 division ratio for training, validation, and testing sets. Moreover, for each positive DDI tuple (d_i, d_j, r), we generate a corresponding negative sample by altering either d_i or d_j according to the strategy proposed by Wang et al. [40].

Besides, the performance metrics included the area under the precision and recall curve (AUPR), F1-score (F1), precision (PRE), recall (REC), accuracy (ACC), and area under the curve (AUC).

Implementation

Our model is developed using Python 3.8 and Pytorch 1.13.0, incorporating functionalities from torch-geometric 1.6.3 [41], Numpy 1.23.0 [42], Pandas 2.0.3 [43], Scikit-learn 1.2.2 [44], and RDkit 2023.3.3 [45]. All experiments were run on Ubuntu OS with a NAVID GeForce TRX 3090 GPU. The Adam optimizer, with a learning rate of $1e-4$, is employed with a batch size of 200. We run the model for a maximum of 20 epochs across all datasets. The top-performing model, identified at the epoch with the highest AUPR score on the validation set, is then used for evaluating the final performance on the test set. The hyperparameter settings are shown in Table 1. Figure 2B illustrates the impact of different hyperparameter choices, including message passing steps (T), hidden feature dimension (D), and dropout probability (P), on metric scores. The analysis is conducted on 10 subsets of the DrugBank test set, with the dataset evenly divided into parts based on 86 categories, and results representing the averages across these subsets. We found that optimal results were achieved with $T=5$, $D=900$, and P was 0.05.

Table 1 A list of model hyperparameters and their respective values

Hyperparameter	Module	Value
Message passing iterations (T)	Message passing in the dual-attention graph transformer	5
Initial hidden dimension (D)	Initialization phase of the dual-attention graph transformer	900
Dropout probability (P)	Dual-attention graph transformer and FFN network classification	0.05

Baselines and variants

We compare DrugDAGT's performance against four other models in predicting DDIs. First, GMPNN-CS [35] utilizes a gated Message Passing Neural Network (MPNN) to extract diverse chemical substructures from molecular graph representations, accommodating variations in size and shape. Second, Molormer (34) considers DDI prediction as the identification of pairwise molecular graph interactions using spatial information and lightweight-based attention mechanism. Third, SA-DDI (36) combines a directed MPNN with an attention mechanism to acquire substructure features effectively. Fourth, DGNN-DDI (37), a dual GNN-based model, extracts molecular structure and interaction information.

To investigate the impact of atomic attention, bond attention, and contrastive learning on model performance, we consider four variants of our model. The first three variants, named No_Atom, No_Bond, and No_Atom_Bond, remove atomic attention, bond attention, and both, respectively. The final variant, named No_Contrastive, eliminates contrastive learning. For the first four models mentioned above, we adopt the hyperparameter settings recommended in the original paper, while for the latter four variants, we maintain consistency with DrugDAGT.

Performance comparison for each DDI type

To assess the performance of our model across different DDI types, we independently evaluated each by calculating metric scores based on the predicted scores and ground-truth labels. Predicted scores reflect the model's predicted probabilities for 86 interaction types and the non-interaction scenario. For each drug pair, we identified the interaction type by selecting the highest probability among these 87 options. The ground-truth label indicates the actual interaction type, with zero indicating no interaction.

Figure 2A demonstrates that our dataset displays an imbalanced distribution across 86 DDI types, and the model's predicted AUPR values do not correlate directly with the number of samples per category. This indicates that the model, utilizing dual-attention mechanisms, effectively captures short- and long-range dependencies within drug molecular structures, thereby enhancing its understanding of molecular complexity and extracting valuable information from limited data. As shown in Fig. 2C, the DrugDAGT model demonstrates superior

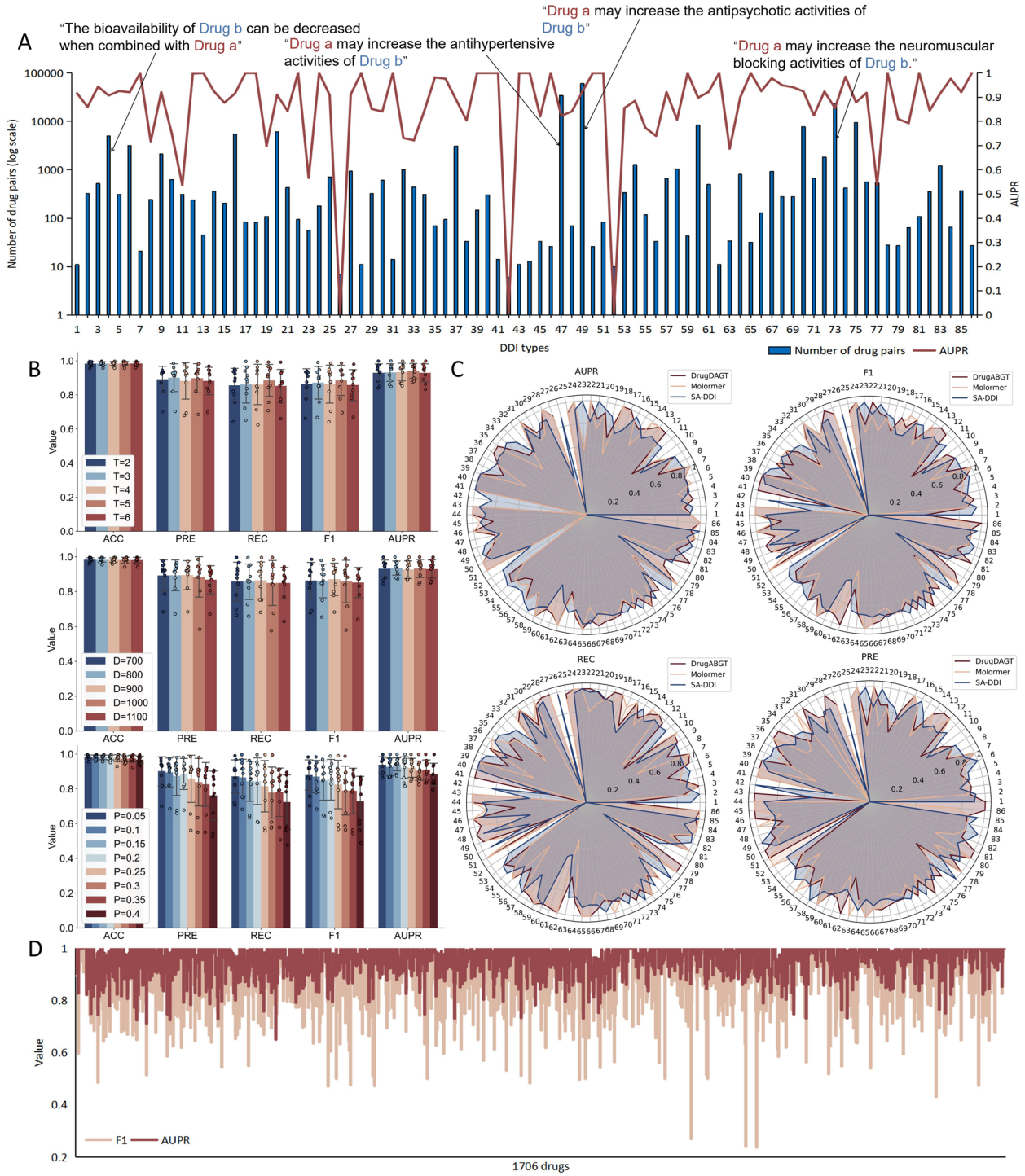

Fig. 2 Performance comparison across 86 DDI types in the DrugBank dataset. **A** Distribution of drug pair counts and predicted AUPR values for 86 DDI types in the DrugBank dataset. The left vertical axis corresponds to the blue bars representing the range of drug pair counts, displayed using logarithmic scaling for a balanced visualization across all categories. The right vertical axis corresponds to the red line graph showing the range of AUPR values. Due to space constraints, only odd-numbered category labels are displayed on the horizontal axis to prevent clutter. **B** Influence of hyperparameter message passing steps (*T*), hidden feature dimension (*D*), and dropout probability (*P*) on the performance metrics: accuracy (ACC), precision (PRE), recall (REC), F1-score (F1), and area under the precision-recall curve (AUPR). The scatter plots display results from ten experiments, and the bar graphs show group averages. **C** Performance comparison of our model against the suboptimal models SA-DDI and Molomer across 86 DDI types. **D** Comprehensive analysis of F1 and AUPR scores for 1706 drugs in the DrugBank dataset, with light red lines representing F1 and dark red lines indicating AUPR, respectively

performance in most categories, outperforming both Molormer and SA-DDI in over half of the DDI types. Further analysis segmented by drug types in DrugBank, as illustrated in Fig. 2D, uses line graphs to display the average F1 and AUPR performance across 1706 drug groups. This line graph illustrates the average F1 and AUPR performance across 1706 drug groups, with the majority nearing or achieving a score of 1, reflecting highly precise DDI predictions for most drugs. Integrating the insights from both Fig. 2A, C, and D, our model demonstrates consistently high and stable performance.

Performance evaluation under warm-start and cold-start scenarios

We conducted assessments under both warm-start and cold-start scenarios to thoroughly evaluate the model performance in practical applications. In the warm-start scenario, although both drugs are known, their interactions remain unidentified, making this setting suitable for detecting missing DDIs among known drugs.

Here, we compared DrugDAGT against eight baselines and variants: GMPNN-CS, Molormer, SA-DDI, DGNN-DDI, No_Atom, No_Bond, No_Atom_Bond, and No_Contrastive. Table 2 shows that DrugDAGT consistently outperforms these models in AUPR, F1, and PRE metrics and remains competitive in REC, ACC, and AUC, demonstrating its effectiveness in DDI prediction. DrugDAGT significantly outperforms No_Contrastive, underscoring the efficacy of graph contrastive learning. Although DrugDAGT shows less improvement with atom and bond attention, its dual attention mechanism effectively captures both short-range and long-range structural dependencies, enhancing the model's interpretability. Additionally, Fig. 3A presents the overall performance of each method across all DDI types through boxplots. Our proposed model consistently outperforms the others, showcasing the highest median values and tightest interquartile ranges across most performance metrics. In view of realistic conditions where the real-world datasets typically exhibit a higher number of negative samples than the positives, we extended our evaluation to more challenging settings. In particular, we evaluated DrugDAGT against the suboptimal SA-DDI method on the DrugBank dataset with a 1:5 positive to negative sample ratio, as shown in Fig. 3C. The results demonstrate that DrugDAGT outperforms SA-DDI, even under significant data imbalance, highlighting its robustness and effectiveness for drug-drug interaction prediction in practical scenarios.

The warm-start scenario can produce overly optimistic results due to data bias. To conduct a more challenging evaluation of the model, we assessed DrugDAGT in the cold-start scenario, where the predictions on test data cannot rely solely on the features of known drugs. This scenario is well-suited for predicting DDIs among emerging drugs in real-world applications. The experimental settings remain consistent with those of the warm start scenario. Figure 3B indicates that all models have a significant degradation in performance from the warm-start to cold-start scenario. This phenomenon may be attributed to most drugs in the DrugBank dataset having distinct structures, resulting in test and training sets that are mostly different but share a few common structures in cold-start scenarios. Despite a decrease in performance from warm-start to cold-start scenarios, DrugDAGT still outperforms GMPNN-CS, Molormer, SA-DDI, and DGNN-DDI, indicating its ability to generalize learned chemical substructure information to different drugs with similar substructures.

Interpretability and visualization

To examine the evolution of drug pair representations during training, we utilized t-distributed stochastic neighbor embedding (t-SNE) [46] to visualize the learned representations of randomly selected 8 DDI types, as

Table 2 Performance comparison of different methods in the warm-start scenario

Method	AUPR	F1	PRE	REC	ACC	AUC
GMPNN-CS	0.8347	0.7854	0.799	0.7915	0.9968	0.9928
Molormer	0.8634	0.8113	0.8157	0.833	0.9968	0.9941
SA-DDI	0.8693	0.8257	0.8408	0.8297	**0.9976**	0.9959
DGNN-DDI	0.7375	0.8377	0.8335	0.8547	0.9972	0.9881
No_Atom	*0.8859*	*0.8743*	*0.8723*	*0.8873*	0.9971	*0.9982*
No_Bond	0.8856	0.8714	0.8718	0.8901	0.9973	**0.9983**
No_Atom_Bond	0.8826	0.8687	0.8577	**0.8923**	0.9971	0.9977
No_Contrastive	0.8753	0.8543	0.8628	0.8683	0.9968	0.9968
DrugDAGT	**0.8959**	**0.8807**	**0.8941**	0.8857	*0.9974*	0.9975

bold indicates optimal performance and italics denote sub-optimal performance

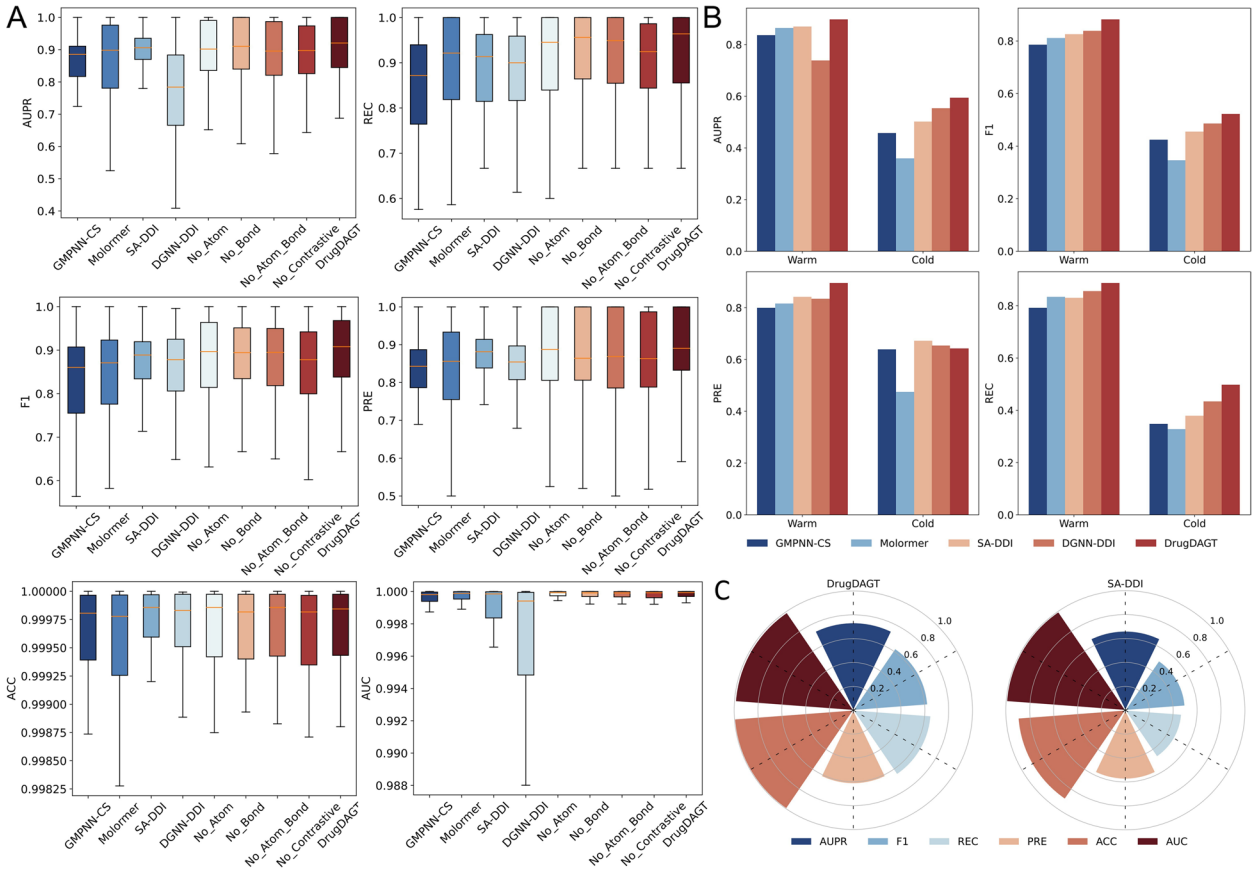

Fig. 3 Performance evaluation under multiple scenarios. **A** Performance comparison of the proposed DrugDAGT with eight baselines and variants under the warm-start scenario. **B** Performance comparison between the warm-start and cold-start scenarios. **C** Performance comparison of DrugDAGT with the suboptimal method SA-DDI on the DrugBank dataset with a positive-to-negative sample ratio of 1:5

shown in Fig. 4A. The visualization initially exhibits a degree of chaos, then gradually clusters more distinctly according to DDI types during the learning process, particularly evident in types 32 and 72. This implies that the drug pair representations learned by DrugDAGT can effectively discriminate between different types of DDIs. Additionally, based on the t-SNE visualization, we used the NMI (Normalized Mutual Information Score) and ARI (Adjusted Rand Index) metrics to assess the consistency and similarity between the clustering results and the actual DDI types. NMI measures to what extent the information the clustering shares with the actual labels, normalized between 0 and 1, with higher values indicating more effective clustering. ARI assesses the clustering performance by comparing the proportion of correctly and incorrectly clustered data, ranging from −1 (worst) to 1 (best). These quantitative results provide an objective measure of model performance. With increasing epochs, the rising NMI and ARI scores demonstrate the improved capacity of the model to learn and adapt to the data structure.

A further strength of DrugDAGT is that it provides critical molecular-level insights and interpretations for drug design efforts. Here, we employ the similarity map [47] implemented in RDKit, which utilizes atomic attention weights to visualize the contribution of each local structure to the final DDI prediction. We investigated interactions between drugs ketoconazole and loxoprofen with five additional drugs not in the training set, validated using the DrugBank dataset, with the visualized results depicted in Fig. 4B.

It is noteworthy that green areas often highlight halogens, chalcogens, or pnictogens, while the carbon atoms in these drugs typically show minimal attention values. This highlights the crucial role of non-carbon atoms in drug activity, as halogens form halogen bonds with receptors, enhancing interactions and significantly boosting both membrane permeability and metabolic stability [48, 49]. Additionally, the DrugDAGT's attention mechanism enhances the learning of molecular functional group representations. For instance, in interactions between Ketoconazole and other drugs, our model prioritizes the

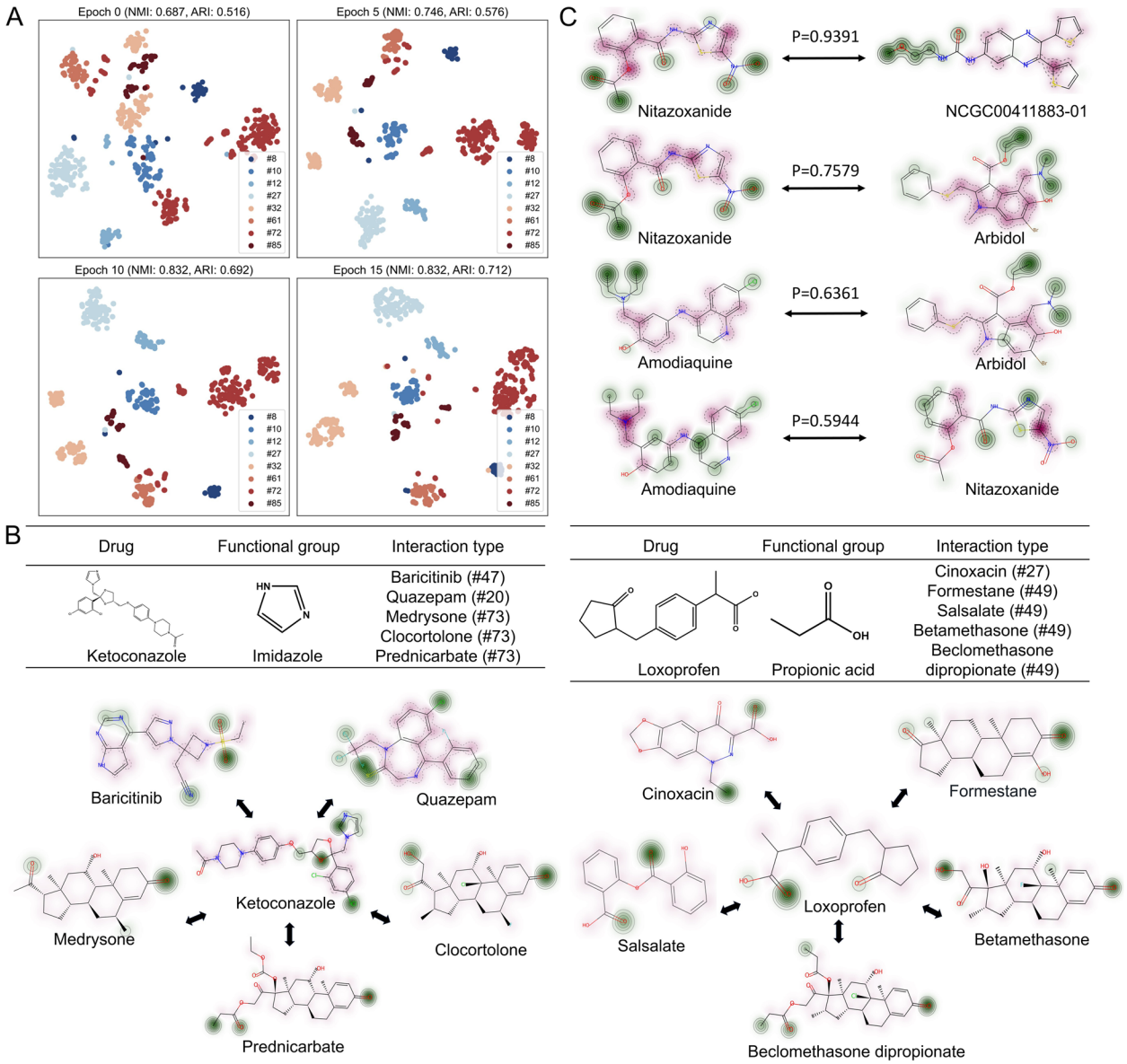

Fig. 4 Visualization of drug pairs representations and local structures. **A** t-SNE visualization of drug pair representations learned during training. NMI and ARI are used to evaluate clustering performance. **B** Visualization of the key local structures for ketoconazole and loxoprofen with five other drugs. In the attention maps, atoms with positive impacts are shown in green, while those with negative impacts are highlighted in red. The darker the color, the stronger the impact. The Tables above each map provide a detailed list of the functional groups and DDI types for ketoconazole and loxoprofen with five other drugs. **C** Visualization of the key local structures for the SARS-CoV-2 drug combinations. "*P*" represents the predicted DDI probability generated by DrugDAGT

imidazole functional group. This reflects Ketoconazole's mechanism as an imidazole antifungal that blocks ergosterol synthesis, increasing membrane fluidity and fungal growth [50]. Similarly, when interacting with other drugs like cinoxacin, quazepam, salsalate, betamethasone, and beclomethasaone dipropionate, the model prioritizes the propionic acid group in loxoprofen. The reason may be its role in inhibiting COX enzymes through this group to reduce inflammation and pain [51].

Our analysis extended to the model evaluation on a dataset formulated for SARS-CoV-2 treatment, comprising 73 interactions among 32 drugs, including 12 combinations that exhibited synergistic effects [37, 52]. Figure 4C displays the predicted probabilities and visualization of key local structures for four synergistic drug

combinations. We observed significant synergy against SARS-CoV-2 with nitazoxanide when combined with NCGC00411883-01, arbidol, and amodiaquine. Our results herein are in excellent alignment with findings from prior studies (52). Across these combinations, nitazoxanide consistently exhibited key substructures, primarily centered on the cresyl acetate functional group ('CC(=O)Oc1ccccc1C').

Conclusions

In this work, we present DrugDAGT, a dual attention deep learning framework for DDI prediction. We employ a graph transformer to integrate both short-range and long-range dependency information and map attention weights to the atomic and bond levels of molecules. This approach identifies key local structures and offers biological insights into the nature of interactions. We have also incorporated graph contrastive learning into our model to enhance its ability to distinguish between structures in a regularized fashion. Experimental results indicate that DrugDAGT consistently outperforms other state-of-the-art DDI models and variants of our own models in both warm-start and cold-start settings, achieving superior DDI prediction performance.

DrugDAGT has a few limitations. It is limited to the analysis of DDIs based on 2D molecular graphs and does not consider three-dimensional (3D) structural information. Since precise 3D structures are often unavailable, especially for new drugs, our method does not incorporate this type of structural detail yet. In future work, we plan to develop new methods that can incorporate realistic 3D structural information through generative AI models to enhance the performance and interpretability of the model. Furthermore, we plan to develop future versions of DrugDAGT that will extend beyond single-scale molecular structures to include molecular network scales, creating a multi-scale framework that can demonstrate better potential for broader applications in data-driven drug discovery. Finally, although the primary DDI databases focus on two-drug interactions, real-world scenarios often involve multiple drugs. As such, another interesting future direction will be to explore the mechanisms of action of multiple drugs by leveraging advanced techniques like text mining and large language models (LLMs).

Methods

Dual-attention graph transformer for molecular graph representation

For each drug, we utilize RDKit to transform each 1D SMILES sequence into its respective 2D graph structure, illustrated in Fig. 5A. Specifically, a drug molecular graph is defined as $G = (V, E)$, where V denotes nodes (atoms) and E denotes edges (chemical bonds). Each atom is characterized by a feature vector M_i encompassing eight attributes: atom type, degree, formal charge, chirality, number of H, hybridization, aromaticity, and atomic mass. Likewise, each bond is represented by a feature vector E_{ij} describing four pieces of information: bond type, conjugated, ring, and stereo. Detailed feature explanations for atoms and bonds are provided in Table 3.

The GNN naturally fits into modeling molecular structures, facilitating the representation of atoms and bonds within molecules for computational chemistry applications.

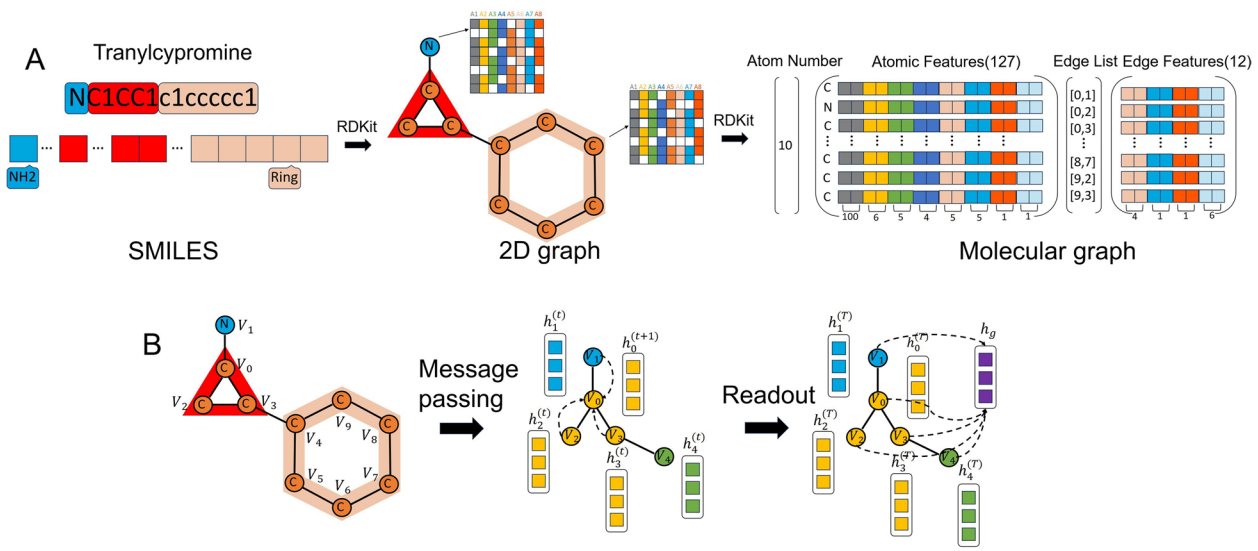

Fig. 5 Drug representation and graph embedding. **A** Tranylcypromine graph representation using RDKit. **B** The message passing and readout phases in graph embedding

Table 3 Atoms and bond features

Feature type	Attribute	Description	Size
Atom feature	Atom type	Chemical elements with atomic number ≤ 100	100
	Degree	Number of covalent bonds	6
	Formal charge	Electronic charge of the atom	5
	Chirality	Unspecified, tetrahedral CW/CCW, or other types of chirality	4
	Number of H	Number of bonded hydrogen atoms	5
	Hybridization	sp, sp^2, sp^3, sp^3d, sp^3d^2	5
	Aromaticity	Whether the atom is a component of an aromatic system	1
	Atomic mass	Mass of the atom (divided by 100)	1
Bond feature	Bond type	Single, double, triple, aromatic	4
	Conjugated	Whether the bond is conjugated	1
	Ring	Whether the bond is in a ring	1
	Stereo	Stereochemistry of bonds (none, any, E/Z, cis/trans)	6

As depicted in Fig. 5B, the GNN framework comprises two stages: (1) Message passing stage updates node features by aggregating their neighbor information. (2) Readout stage aggregates all node features to generate the overall molecular graph feature. However, GNN aggregates information at the node level, and tends to generate unnecessary loops and redundancies during the message passing stage when applied directly to DDI tasks. Moreover, the average impact during the readout phase causes every atom and bind to exert an equal influence on the predicted outcome, thus failing to highlight the critical substructures essential for DDI.

We use the atom-bond attention-based GNN [53] for the molecule graph representation. This framework utilizes a directed message-passing neural network (D-MPNN) [54] for message aggregating through directional bonds and advances it by incorporating transformer-like self-attention mechanisms [55] at both atomic and bond levels. Specifically, it comprises three stages: initialization phase, message passing phase, and readout phase.

Initialization phase

Considering that information is transmitted directionally, each bond is initialized with two feature vectors representing bond information in opposing directions. We denote the hidden feature of each bond $e_{i\to j}$ as h_{ij}^0, and its initialization process is as follows:

$$h_{ij}^0 = \sigma(W_0(M_i \parallel E_{ij})) \tag{1}$$

where M_i and E_{ij} denote the node and edge features generated when a smile is converted into a molecular graph, with $i \in V$, and $j \in N(i)$, where $N(i)$ represents the neighbors of node i in graph G. W_o is a learnable weight matrix. σ and ‖ respectively represent the ReLU activation function and the concatenate operation. Furthermore, molecular descriptors (h_f) and three inter-atomic matrices [53, 56, 57] (Coulomb, adjacency and distance) are produced. The summary of the molecular descriptors generated by RDKit is included in Additional File 1: Table S1. A brief description of the three inter-atomic matrices is as follows: the Coulomb matrix captures the interactions between atoms by representing the geometric and electronic properties of a molecule. The matrix elements include the nuclear charges along the diagonal and the Coulomb repulsion between atoms off the diagonal, expressed as follows:

$$M_{ij} = \begin{cases} 0.5Z_i^{2.4} (i = j) \\ \frac{Z_i Z_j}{|R_i - R_j|} (i \neq j) \end{cases} \tag{2}$$

where Z_i is the atomic number corresponding to the Cartesian coordinates R_i. The adjacency and distance matrices, on the other hand, are two graphical representations of molecules that capture the connectivity and distance information for each pair of atoms. In an adjacency matrix, elements are set to 1 if a chemical bond exists between the corresponding atoms, and to 0 if there is no bond, using a binary approach. Conversely, a distance matrix represents the topological distances between atoms, calculated from the 3D coordinates of each pair, thereby reflecting the spatial configuration of the molecule.

Message passing phase

To update the bond message r_{ij}^t in each iteration t, we aggregate all incoming neighboring hidden vectors h_{xi}^{t-1}

from the previous iteration, excluding those representing the inverse direction of the bond h_{ji}^{t-1}.

$$r_{ij}^{t} = \sum_{x \in \mathrm{N(i)}} h_{xi}^{t-1} - h_{ji}^{t-1} \tag{3}$$

Next, utilizing all bond messages as input, a bond-level transformer block is employed to generate bond attention, which is then merged with the input bond hidden features to produce updated features.

$$t_{ij}^{t} = \alpha_{bond}\left(r_{ij}^{t}\right) + r_{ij}^{t} \tag{4}$$

The bond attention mechanism updates the hidden state for N bonds within the molecular graph G during each iteration. This process begins with the input bond hidden feature matrix $R_b = [r_{b_1}^t, \ldots, r_{b_k}^t, \ldots, r_{b_{2N}}^t]$, where $R_b \in \mathbb{R}^{2N \times d}$ and d is the hidden dimension. Corresponding query $Q_b = [q_{b_1}, \ldots, q_{b_k}, \ldots, q_{b_{2N}}]$, key $K_b = [k_{b_1}, \ldots, k_{b_k}, \ldots, k_{b_{2N}}]$, and value $V_b = [v_{b_1}, \ldots, v_{b_k}, \ldots, v_{b_{2N}}]$ matrices, all derived from R_b, facilitate the subsequent attention operations. Bond attention steps include the following:

1) Global query formation: a global query vector q_b is computed by first calculating the additive attention weights α_{b_k} for each bond message, then summing the weighted bond query vectors.

$$\alpha_{b_k} = \frac{\exp(q_{b_k} w_q / \sqrt{d})}{\sum_{j=1}^{2N} \exp(q_{b_j} w_q / \sqrt{d})} \tag{5}$$

$$q_b = \sum_{k=1}^{2N} \alpha_{b_k} q_{b_k} \tag{6}$$

where q_{b_k} is a learnable weight matrix.

2) Key value interaction: each bond key vector k_{b_k} interacts with the global bond query to form a product vector p_{b_k}, which is then used to compute a global bond key k_b via the additive attention.

$$p_{b_k} = q_b \cdot k_{b_k} \tag{7}$$

$$\gamma_{b_k} = \frac{\exp(p_{b_k} w_k / \sqrt{d})}{\sum_{j=1}^{2N} \exp(p_{b_j} w_k / \sqrt{d})} \tag{8}$$

$$k_b = \sum_{i=1}^{2N} \gamma_{b_k} q_b \tag{9}$$

The global bond key is then used to transform the bond value vectors v_{b_k} via element-wise multiplication to obtain g_{b_k}.

3) Output formation: The attention output for each bond is then constructed by adding the projected value vector g_{b_k} to the original bond query vector q_{b_k}, followed by a layer normalization step to produce the normalized bond attention output.

$$c_{b_k} = g_{b_k} w_v + q_{b_k} \tag{10}$$

$$O_b = LayerNorm([c_{b_1}, \ldots, c_{2N}]) \tag{11}$$

Finally, the attention message t_{ij}^t is projected into a higher-dimensional space using a weight matrix W_t. It is concatenated with the original bond feature h_{ij}^0 and passed through the ReLU activation function σ to generate the bond-level representation in step t, represented as follows:

$$h_{ij}^{t} = \sigma(h_{ij}^{0} + W_t t_{ij}^{t}) \tag{12}$$

Each atom obtains its neighboring bond features through message-passing layers, concatenates them with atom features, and transforms them using a weight matrix and ReLU activation, resulting in the generation of atom-level hidden messages.

$$x_i = \sigma(W_T \parallel (M_i, \sum_{j \in N(i)} h_{ij}^T)) \tag{13}$$

Employing the multi-head self-attention mechanism along with three matrices designed for atoms assists in generating atomic-level attention, which is then combined with input atomic hidden features to produce the final atomic-level representation.

$$h_i = \alpha_{atom}(x_i, A_c, A_a, A_d) + x_i \tag{14}$$

The atom attention mechanism focuses on the interactions between atoms in a molecule. Unlike bond attention, which is built upon additive attention for efficiency, atom attention uses the scaled dot-product attention from the original Transformer network to capture a more comprehensive representation of molecular structures. We start by initializing the atom hidden matrix X_a for a molecule with V atoms, signifying the initial features of the atoms: $X_a = [x_a^1, \ldots, x_a^V] \in \mathbb{R}^{V \times d}$. In the atom attention mechanism, the same matrix X_a acts as the query Q_a, key K_a, and value V_a matrices. Atom attention steps include the following:

1) Attention matrix calculation: For each of the six attention heads, an attention matrix A_a is calculated by adding a bias term X_{graph} representing specific molecular features (like distance, adjacency or coulomb) to the scaled dot-product of Q_a and K_a:

$$A_a = Softmax(\frac{Q_a W_q (K_a W_k)^T}{\sqrt{d}} + X_{graph}) \tag{15}$$

where W_q and W_k are learnable weight matrices.

2) Output generation: The attention output is derived by aggregating the value vectors, with the weights being the attention scores from the previous step. Finally, this output is normalized using layer normalization:

$$O_a = Norm(A_a V_a W_a) \tag{16}$$

where the W_a is another learnable weight matrix.

Readout phase

The global representation for a molecule is derived by aggregating all learned atom representations as follows:

$$T = \sum\nolimits_{i \in V} h_i \tag{17}$$

$$p_i^y = softmax((W_x T_x) \odot (W_y h_i^y)) \tag{18}$$

Extracting interaction-specific local structures

To explicitly learn local interactions between drug pairs, we assign scores to each based on their probability of interacting with another drug [36]. In practice, considering a drug pair (d_x, d_y), we leverage the substructure information of d_x to detect critical substructures in d_y. Initially, we assess the interaction probability between d_x and each substructure in d_y:where T_x is the global representation of d_x, $\odot$ is dot product, p_i^y measures the importance of the substructure that is centered around the i-th atom in d_y. And W_x and W_y are weight matrices that transform features. Finally, the graphical representation of d_y is computed using the formula below:

$$H_y = \sum\nolimits_{i \in V} p_i^y \cdot h_i^y \cdot T_x + h_f \tag{19}$$

where $(\cdot)$ denotes element-wise multiplication. Similar processing is also applied to d_x to obtain its graphical representation H_x, h_f is the molecular descriptors.

Optimization with graph contrastive learning

We leverage graph contrastive learning to enhance drug representations, intending to improve the accuracy of DDI predictions. Given its exceptional efficacy in unsupervised learning for graph-based data, our model incorporates graph contrastive learning as a regularization strategy [58, 59]. This method enriches the model's discriminative ability by generating diverse views of each drug, achieved by infusing random noise into the nodes' representations within drug graphs and then contrasting these views.

For each drug molecular graph$G$, we randomly generate noise δ_i for each node i, maintaining with $\|\delta_i\| = \delta$, and perturb the features of each node in a consistent direction, as indicated by:

$$H' = H + |\delta_i| \cdot sign(H) \tag{20}$$

This procedure maintains the core structural features of drug molecules while introducing slight variations, creating different views for contrastive learning.

We enhance the discriminative power of drug representations by setting a contrastive learning objective that maximizes similarity between varied views of the same drug and contrasts them with others. The steps include the following: Randomly selecting a mini-batch of M drug molecular graphs, each yielding two distinct representations. Applying the InfoNCE loss function [60] for the mth drug molecule in the mini-batch:

$$L_{con} = -\frac{1}{M} \sum\nolimits_{m=1}^{M} log \frac{e^{f_m \diamond f'_m / \tau}}{\sum_{m'=1}^{M} e^{f_m \diamond f'_{m'} / \tau}} \tag{21}$$

where f_m and f'_m correspond to the representations of two different views of the mth drug. And the temperature parameter τ is set to 0.5 following [61], with $\diamond$ denoting the cosine similarity between vectors.

Drug-drug interaction prediction

To calculate probabilities for multiple DDI types, we feed the concatenated drug pair representation into two FFN network classification layers, followed by a softmax function:

$$p = Softmax(FFN(Concat(H'_x, H'_y))) \tag{22}$$

We then optimize all learnable parameters using backpropagation, aiming to minimize the combination of the cross-entropy loss and graph contrastive learning loss as follows:

$$L = -\frac{1}{Z} \sum\nolimits_{xy}^{Z} \left[y_{xy} \log(p_{xy}) + (1 - y_{xy}) \log(1 - p_{xy}) \right] + L_{con} \tag{23}$$

where y_{xy} is the ground-truth label of the xyth drug-drug pair among Z total pairs, and p_{xy} is the probability predicted by the model.

Abbreviations

DDIs	Drug-drug interactions
ADEs	Adverse drug events
GNN	Graph neural network
FFN	Feed-forward network
SMILES	Employing the simplified molecular-input line-entry system
AUPR	Area under the precision and recall curve
F1	F1-score
PRE	Precision
REC	Recall
ACC	Accuracy
AUC	Area under the curve

SNE	T-distributed stochastic neighbor embedding
D-MPNN	Directed message-passing neural network
LLMs	Large language models

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-024-02030-9.

Additional file 1: Table S1. The categories and examples of the molecular descriptors generated by RDKit.

Acknowledgements
The authors would like to thank the anonymous reviewers for their constructive comments.

Authors' contributions
Q.Z. and Y.C. conceived and designed the experiment. Y.C. and J.W. performed the experiment. Y.C., N.M. and Y.D analyzed the results. Y.C., J.S. and W.Y. wrote and revised the manuscript. J.S. and W.Y. provided funding and resources and project administration. All authors provided feedback on the manuscript. All authors read and approved the final manuscript.

Funding
The work was supported by the National Natural Science Foundation of China (No. 62102269, No.62373080, No. 62131004, No. 62302341, No. 62301369, No.62303328, No. 62172076, No. U22A2038), the Major and Seed Inter Disciplinary Research (IDR) projects awarded by Monash University, the National Key R&D Program of China (2022ZD0117700), the Shenzhen Polytechnic Research Fund (6024310027 K), the National funded postdoctoral researcher program of China (GZC20230382), the Municipal Government of Quzhou (No. 2023D036), the Tianfu Emei Plan and the Zhejiang Provincial Natural Science Foundation of China (No. LY23F020003).

Data availability
All code and data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories, which are available in the Zenodo repository (https://doi.org/10.5281/zenodo.13788384 [62]) and GitHub (https://github.com/codejiajia/DrugDAGT).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China. [2]Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China. [3]School of Electronic and Communication Engineering, Shenzhen Polytechnic University, Shenzhen, China. [4]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China. [5]Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC 3800, Australia. [6]Wenzhou Medical University-Monash Biomedicine Discovery Institute Alliance in Clinical and Experimental Biomedicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325035, China.

Received: 13 June 2024 Accepted: 2 October 2024
Published online: 14 October 2024

References

1. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol. 2017;35(5):463–74.
2. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356(9237):1255–9.
3. Sun X, Vilar S, Tatonetti NP. High-throughput methods for combinatorial drug discovery. Scienc Translational Medicine. 2013;5(205):205rv1-205rv1.
4. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: In vitro safety pharmacology profiling: an essential tool for successful drug development. Drug Discovery Today. 2005;10(21):1421–33.
5. 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.
6. Ja Qin, Yang Y, Ai C, Ji Z, Chen W, Song Y, et al. Antibiotic combinations prediction based on machine learning to multicentre clinical data and drug interaction correlation. International Journal of Antimicrobial Agents. 2024;63(5):107122.
7. Wei L, He W, Malik A, Su R, Cui L, Manavalan B. Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. Briefings in Bioinformatics. 2021;22(4):bbaa275.
8. Shen X, Li Z, Liu Y, Song B, Zeng X. PEB-DDI: a task-specific dual-view substructural learning framework for drug-drug interaction prediction. IEEE J Biomed Health Inform. 2024;28(1):569–79.
9. Dou M, Tang J, Tiwari P, Ding Y, Guo F. Drug-Drug Interaction Relation Extraction Based on Deep Learning: A Review. ACM Comput Surv. 2024;56(6):1–33.
10. Cami A, Manzi S, Arnold A, Reis BY. Pharmacointeraction network models predict unknown drug-drug interactions. PLoS ONE. 2013;8(4): e61468.
11. Zhang P, Wang F, Hu J, Sorrentino R. Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects. Sci Rep. 2015;5(1): 12339.
12. Park K, Kim D, Ha S, Lee D. Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks. PLoS ONE. 2015;10(10): e0140816.
13. Zhang W, Jing K, Huang F, Chen Y, Li B, Li J, et al. SFLLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. Inf Sci. 2019;497:189–201.
14. Li H-L, Pang Y-H, 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.
15. Ren S, Chen L, Hao H, Yu L. Prediction of cancer drug combinations based on multidrug learning and cancer expression information injection. Futur Gener Comput Syst. 2024;160:798–807.
16. Pang C, Qiao J, Zeng X, Zou Q, Wei L. Deep Generative Models in De Novo Drug Molecule Generation. J Chem Inf Model. 2024;64(7):2174–94.
17. Ma T, Lin X, Song B, Yu PS, Zeng X. KG-MTL: Knowledge Graph Enhanced Multi-Task Learning for Molecular Interaction. IEEE Trans Knowl Data Eng. 2023;35(7):7068–81.
18. Lin X, Dai L, Zhou Y, Yu ZG, Zhang W, Shi JY, et al. Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction. Briefings in Bioinformatics. 2023;24((4):bbad235.
19. Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014. p. 701–710.
20. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 855–864.
21. Ribeiro LFR, Saverese PHP, Figueiredo DR. struc2vec: Learning node representations from structural identity. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017. p. 385–394.
22. Shi J-Y, Mao K-T, Yu H, Yiu S-M. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. Journal of Cheminformatics. 2019;11(1):28.
23. Cao S, Lu W, Xu Q. GraRep: learning graph representations with global structural information. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015. p. 891–900.

24. Ou M, Cui P, Pei J, Zhang Z, Zhu W. Asymmetric Transitivity Preserving Graph Embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016. p. 1105–14.
25. Zhang W, Liu X, Chen Y, Wu W, Wang W, Li X. Feature-derived graph regularized matrix factorization for predicting drug side effects. Neurocomputing. 2018;287:154–62.
26. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. LINE: Large-scale Information Network Embedding. Proceedings of the 24th International Conference on World Wide Web. 2015. p. 1067–1077.
27. Wang D, Cui P, Zhu W. Structural Deep Network Embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016. p. 1225–34.
28. Kipf TN, Welling M. Variational Graph Auto-Encoders. arXiv preprint arXiv:161107308. 2016.
29. Liu B, Gao X, Zhang H. BioSeq-Analysis2.0: an updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. Nucleic Acids Research. 2019;47(20): e127.
30. 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. Advanced Science. 2024;11(7): 2306329.
31. Lin X, Quan Z, Wang ZJ, Ma T, Zeng X. KGNN: Knowledge Graph Neural Network for drug-drug interaction prediction. In: IJCAI. 2020. p. 2739–2745.
32. Yu Y, Huang K, Zhang C, Glass LM, Sun J, Xiao C. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. Bioinformatics. 2021;37(18):2988–95.
33. Hong Y, Luo P, Jin S, Liu X. LaGAT: link-aware graph attention network for drug–drug interaction prediction. Bioinformatics. 2022;38(24):5406–12.
34. Zhang X, Wang G, Meng X, Wang S, Zhang Y, Rodriguez-Paton A, et al. Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug–drug interactions prediction. Brief Bioinform. 2022;23(5):bbac296.
35. Nyamabo AK, Yu H, Liu Z, Shi J-Y. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. Brief Bioinform. 2022;23(1):bbab441.
36. Yang Z, Zhong W, Lv Q, Yu-Chian CC. Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network. Chem Sci. 2022;13(29):8693–703.
37. Ma M, Lei X. A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. PLoS Comput Biol. 2023;19(1): e1010812.
38. Li Z, Zhu S, Shao B, Zeng X, Wang T, Liu T-Y. DSN-DDI: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. Brief Bioinform. 2023;24(1):bbac597.
39. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Research. 2014;42(D1):D1091–7.
40. Wang Z, Zhang J, Feng J, Chen Z. Knowledge Graph Embedding by Translating on Hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence. Québec City; 2014. p. 1112–9.
41. Fey M, Lenssen JE. Fast graph representation learning with PyTorch Geometric. arXiv preprint arXiv:190302428. 2019.
42. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62.
43. The pandas development team. pandas-dev/pandas: Pandas (v2.0.3). Zenodo; 2023. https://doi.org/10.5281/zenodo.8092754.
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
45. RDKit: open-source cheminformatics. https://www.rdkit.org.
46. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11):2579–605.
47. Riniker S, Landrum GA. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. Journal of Cheminformatics. 2013;5(1): 43.
48. Mendez L, Henriquez G, Sirimulla S, Narayan M. Looking back, looking forward at halogen bonding in drug discovery. Molecules. 2017;22(9): 1397.
49. Frontera A, Bauza A. On the importance of pnictogen and chalcogen bonding interactions in supramolecular catalysis. Int J Mol Sci. 2021;22(22): 12550.
50. Van Tyle JH. Ketoconazole; mechanism of action, spectrum of activity, pharmacokinetics, drug interactions, adverse reactions and therapeutic use. Pharmacotherapy. 1984;4(6):343–73.
51. Meek IL, Van de Laar MA, Vonkeman HE. Non-steroidal anti-inflammatory drugs: an overview of cardiovascular risks. Pharmaceuticals. 2010;3(7):2146–62.
52. Jitobaom K, Boonarkart C, Manopwisedjaroen S, Punyadee N, Borwornpinyo S, Thitithanyanont A, et al. Synergistic anti-SARS-CoV-2 activity of repurposed anti-parasitic drug combinations. BMC Pharmacol Toxicol. 2022;23(1):41.
53. Liu C, Sun Y, Davis R, Cardona ST, Hu P. ABT-MPNN: an atom-bond transformer-based message-passing neural network for molecular property prediction. J Cheminform. 2023;15(1):29.
54. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370–88.
55. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017. p. 6000–6010.
56. Rupp M, Tkatchenko A, Müller K-R, von Lilienfeld OA. Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett. 2012;108(5): 058301.
57. David L, Thakkar A, Mercado R, Engkvist O. Molecular representations in AI-driven drug discovery: a review and practical guide. Journal of Cheminformatics. 2020;12(1):56.
58. Zeng J, Xie P. Contrastive self-supervised learning for graph classification. Proc AAAI Conf Artif Intell. 2021;35(12):10824–32.
59. Yu J, Yin H, Xia X, Chen T, Cui L, Nguyen QVH. Are graph augmentations necessary? Simple graph contrastive learning for recommendation. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022. p. 1294–1303.
60. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 9729–9738.
61. You Y, Chen T, Sui Y, Chen T, Wang Z, Shen Y. Graph contrastive learning with augmentations. In: Proceedings of the 34st International Conference on Neural Information Processing Systems. 2020. p. 5812–5823.
62. Chen Y. DrugDAGT (v1.0.0). Zenodo. 2024. https://doi.org/10.5281/zenodo.13788384.

Publisher's Note