

GraphomerDTI: A graph transformer-based approach for drug-target interaction prediction

Mengmeng Gao^a, Daokun Zhang^{b,***}, Yi Chen^{a,**}, Yiwen Zhang^c, Zhikang Wang^d, Xiaoyu Wang^d, Shanshan Li^c, Yuming Guo^c, Geoffrey I. Webb^b, Anh T.N. Nguyen^e, Lauren May^e, Jiangning Song^{d,*}

^a School of Biological Science and Medical Engineering, Southeast University, Nanjing, China

^b Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Monash University, Melbourne, Australia

^c Climate, Air Quality Research Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, 3004, Australia

^d Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia

^e Drug Discovery Biology Theme, Monash Institute of Pharmaceutical Sciences, Monash University, Melbourne, Australia

ARTICLE INFO

Keywords:

Deep learning
Drug-target interaction
Graph transformer
Attention mechanism

ABSTRACT

The application of Artificial Intelligence (AI) to screen drug molecules with potential therapeutic effects has revolutionized the drug discovery process, with significantly lower economic cost and time consumption than the traditional drug discovery pipeline. With the great power of AI, it is possible to rapidly search the vast chemical space for potential drug-target interactions (DTIs) between candidate drug molecules and disease protein targets. However, only a small proportion of molecules have labelled DTIs, consequently limiting the performance of AI-based drug screening. To solve this problem, a machine learning-based approach with great ability to generalize DTI prediction across molecules is desirable. Many existing machine learning approaches for DTI identification failed to exploit the full information with respect to the topological structures of candidate molecules. To develop a better approach for DTI prediction, we propose GraphomerDTI, which employs the powerful Graph Transformer neural network to model molecular structures. GraphomerDTI embeds molecular graphs into vector-format representations through iterative Transformer-based message passing, which encodes molecules' structural characteristics by node centrality encoding, node spatial encoding and edge encoding. With a strong structural inductive bias, the proposed GraphomerDTI approach can effectively infer informative representations for out-of-sample molecules and as such, it is capable of predicting DTIs across molecules with an exceptional performance. GraphomerDTI integrates the Graph Transformer neural network with a 1-dimensional Convolutional Neural Network (1D-CNN) to extract the drugs' and target proteins' representations and leverages an attention mechanism to model the interactions between them. To examine GraphomerDTI's performance for DTI prediction, we conduct experiments on three benchmark datasets, where GraphomerDTI achieves a superior performance than five state-of-the-art baselines for out-of-molecule DTI prediction, including GNN-CPI, GNN-PT, DeepEmbedding-DTI, MolTrans and HyperAttentionDTI, and is on a par with the best baseline for transductive DTI prediction. The source codes and datasets are publicly accessible at <https://github.com/mengmeng34/GraphomerDTI>.

1. Introduction

Given the effective treatments for many diseases (e.g., Alzheimer's and epilepsy) are limited, there is an ongoing need for new drug discovery. However, the emergence of new diseases and the raise of drug

resistance propose challenges in finding the optimal treatment. Screening the candidate drugs likely to interact with disease targets is the first step of drug discovery. Specifically, through interacting with disease targets, such as activating/inhibiting an enzyme, receptor, or ion channel, effective drugs can be identified after selection, followed by the

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: daokun.zhang@monash.edu (D. Zhang), yichen@seu.edu.cn (Y. Chen), jiangning.song@monash.edu (J. Song).

effectiveness verification based on clinical trials. Nevertheless, the traditional drug discovery pipeline is notoriously expensive, time-consuming, and more importantly, with a low success rate. It is reported on average 2.6 billion dollars and more than 10 years are required to develop a new drug, while the success rate at the first clinical trial phase is less than 10% [1].

As a useful complement to conventional wet-lab experiments aimed at identifying drug-target interactions, machine learning-based drug-target interaction (DTI) prediction [2] can enable high-throughput drug screening by significantly reducing the cost, time and human resources, as well as increasing the success rates of follow-up clinical trials [3,4]. Such approaches first train DTI prediction models based on the known interactions between drugs and targets. The trained models can then be applied to conduct drug screening [5–10], involving search through a vast candidate drug space to identify the drug molecules that are predicted to interact with target proteins. To ensure the effectiveness of DTI prediction-based drug screening, the DTI prediction models are required to have high prediction accuracies, particularly for the out-of-sample molecules that dominate the candidate drug space and have no known interaction records with any disease targets.

Machine learning-based DTI prediction methods require informative features/representations of drug molecules and target proteins as their input [11,12]. Various explorations have been made to construct drug features, including: 1) using the one-hot encodings spanned by the handcrafted molecular descriptors, e.g., substituent atoms, chemical bonds, structural fragments, and functional groups [13]; 2) transforming molecules into Simplified Molecular Input Entry System (SMILES) strings [14] and learning molecular representations through sequence learning models [15,16]; 3) modeling molecules as graphs and leveraging graph neural networks (GNNs) [17] to learn molecular graph representations [18–20]. On the other hand, sequence-based models are mainly used to learn protein features from protein primary chains, like 1D-CNN [21], LSTM [22], and Transformer [23], etc.

As an early-stage attempt of using AI to identify DTIs, DL-CPI [24] extracted handcrafted features to represent drug molecules and target proteins, based on which a fully connected neural network (FCNN) for DTI prediction was designed. The handcrafted features are usually agnostic to downstream DTI prediction tasks, limiting DTI prediction performance. To overcome this limitation, various end-to-end deep learning models have been proposed, which could learn task-relevant features more effectively. DeepConv-DTI [25] extracted the extended connectivity fingerprint (ECFP) features to represent drug molecules, and utilized a multi-scale 1D-CNN to learn protein representations. DeepDTA [26] used two 1-D CNNs to learn molecular and protein representations from SMILES strings and amino acid sequences respectively. Additionally, DrugVQA [27] used BiLSTM [28] to learn molecular representations from molecular SMILES strings, and a 2-dimensional convolutional neural network (2D-CNN) to learn protein representations from the 2-dimensional amino acid contact maps. GraphDTA [29] leveraged a GNN to extract molecular representations from molecular graphs and a 1D-CNN to learn protein representations from protein amino acid sequences.

Nevertheless, the DTI prediction methods mentioned above have not considered the varying contacting patterns between the atoms of drug molecules and amino acid segments of target proteins. In order to enhance DTI prediction performance, attention mechanisms [30] have been utilized to model the intricate interaction patterns between drug molecules and target proteins. Specifically, GNN-CPI [21] used a one-sided attention mechanism to characterize protein subsequences that are vital for DTI identification. Meanwhile, this attention mechanism was also adopted by DeepEmbedding-DTI [22], which utilized a local breadth-first search (BFS) to extract molecular subgraph features. E2E [31] employed a two-way attention mechanism to simultaneously account for the varying importance levels of drug molecule atoms and protein amino acid segments for DTIs. The Transformer-based self-attention mechanism [32] was utilized by the TransformerCPI to

characterize the interactions between drug molecule atoms and protein amino acids, which was further leveraged by the GNN-PT [23] to capture the complex interactions with a target-attention decoder [32]. MolTrans [33] employed the neighborhood interaction scheme to characterize DTIs by leveraging the correlations between neighboring drug atoms and the relatedness between neighboring protein amino acids. Notably, the above methods failed to capture the diverse types of non-covalent interactions between atoms and amino acids (e.g., hydrophobic interactions, hydrogen bonding, and π stacking). To address this issue, the HyperAttentionDTI model [34] assigned an attention vector to each atom-amino acid pair to model the varying interaction types.

However, existing molecular representations are limited in terms of their ability to capture the essential structural characteristics of drug molecules, regarding the relative importance and contribution of different atoms, as well as the structural distances and chemical bonding types between atoms, which are particularly important for describing molecular subgraph features. This limitation makes it difficult for current DTI prediction models to extract important molecular subgraph features, such as molecular functional groups, which are indicative of the interactions of drug molecules with their targets. With an ineffective molecular representation scheme, the DTI prediction for the out-of-sample molecules would have limited performance, as informative molecular subgraph features of the out-of-sample molecules could not be sufficiently explored for the DTI prediction.

To address this challenge, we proposed a novel DTI prediction model termed GraphomerDTI, which used the powerful Graph Transformer neural network to construct molecular representations from molecular graphs, with a stronger ability to generalize DTI prediction from the training molecules to novel out-of-sample molecules. With the Transformer [32] based message passing mechanism, Graph Transformer encodes the discriminative molecular subgraph features into molecular representations. Building upon the vanilla Graph Transformer architecture [35] and following the design principles proposed by Ying et al. [36], we augmented three key components to capture more informative structural features (i.e., molecular subgraph patterns): (1) the atom centrality encoding was utilized to measure the atom importance; (2) the shortest path distances between pairwise atoms were calculated and transformed into the atom spatial encodings to characterize the structural distances between different atoms; (3) the edge encoding was used to model the various chemical bond types that are informative for distinguishing different molecular subgraphs. Compared with the vanilla GNN, Graph Transformer provides a broader receptive field to capture more global structural relatedness between atoms. More importantly, it offers a greater flexibility to model the attention between different atoms with varying structural distances and chemical bond types. This can enable an automated selection of informative molecular subgraph features to be generalized to the out-of-sample molecules, thereby leading to improved DTI prediction. Furthermore, GraphomerDTI also employed a stacked 1D-CNN to learn target protein representations from the amino acid sequences. Then, an attention layer was established to model the complex interactions between molecular representations and protein representations.

To verify the efficacy of our proposed GraphomerDTI method, we conducted extensive experiments on three real-world DTI prediction benchmarks. The experimental results show that the proposed GraphomerDTI model significantly outperforms state-of-the-art baseline methods for predicting DTIs of novel drug molecules, and performs competitively for predicting DTIs of known drug molecules. In addition, we also illustrated the efficacy of DTI prediction through a case study, by identifying drug molecules that interact with the adrenergic receptors, which highlights that the proposed method can identify more interacting drug molecules than the state-of-the-art baseline methods.

2. Materials and methods

2.1. GraphomerDTI model

Fig. 1A illustrates the pipeline of the proposed GraphomerDTI model, which leverages the powerful Graph Transformer neural network to learn molecular representations. Specifically, GraphomerDTI consists of three key components: a drug representation learning component, a protein representation learning component, and a drug-protein interaction learning component. The inputs to the GraphomerDTI model include the molecular graph of the input drug and the amino acid sequence of the input target protein. A total of 12 stacked Graph Transformer layers [35] and three stacked 1D-CNN layers [34] were employed to extract the informative features of the input drug and protein, respectively. Then an attention layer [34] was applied to encode the associations between different drug and protein feature components into a decision vector. Finally, the obtained decision vector was used as the input to a fully connected neural network (FCNN) to predict the interaction between the input drug molecule and the target protein.

2.1.1. Drug representation learning

Despite the broad application to drug representation, SMILES sequences fail to capture enough structural relatedness between different atoms due to its limited encoding scheme, where the three-dimensional molecular structures have to be converted into one-dimensional strings. For example, it fails to provide insights into the atom connection structures in a manner readily exploitable by a deep learning model. To rectify this problem, we used molecular graphs to describe drug molecules. The molecular graph modelling scheme is intuitive, where atoms are modelled as nodes and chemical bonds between atoms are represented by edges, describing all connections between atoms. Formally, we defined a molecular graph as $G = (V, E)$ with a set of nodes (atoms) V with the size n and a set of edges (chemical bonds) E . Each node $v_i (v_i \in V)$ is described by a d -dimensional feature vector $x_{v_i} (x_{v_i} \in \mathbb{R}^d)$, a learnable embedding vector corresponding to its atom type (e.g., C, H, O or N). Each edge $e_i (e_i \in E)$ is also characterized by a d -dimensional feature vector $x_{e_i} (x_{e_i} \in \mathbb{R}^d)$, and a learnable embedding vector

determined by its chemical bond type.

We used 12 Graph Transformer layers to construct drug molecular representations. As is shown in Fig. 1B, the Graph Transformer layer is implemented through the multi-head self-attention scheme. Expressively, the multi-head self-attention between atoms provides an elegant message-passing mechanism to capture informative molecular structural features, where node centrality encoding, node spatial encoding, and edge encoding are used to encode the importance of atoms, the structural relatedness between atoms, and chemical bond semantics, respectively.

2.1.1.1. Graph transformer layer. Following the GNN-based message passing scheme [17], Graph Transformer layer updates atom representations by aggregating the representations of neighboring atoms with an attention mechanism. By stacking multiple Graph Transformer layers, atoms' rich neighborhood structure within a large radius can be effectively encoded into final atom representations.

In the l -th layer and k -th head, for the atom $v_i \in V$, the input representation $h_{v_i}^l (h_{v_i}^l \in \mathbb{R}^{d_l})$ is updated to $h_{v_i}^{l+1,k} (h_{v_i}^{l+1,k} \in \mathbb{R}^{d_{l+1,k}})$ as

$$h_{v_i}^{l+1,k} = \sum_{j=1}^n w_{ij}^{k,l} V^{k,l} h_{v_j}^l \quad (1)$$

where $V^{k,l} (V^{k,l} \in \mathbb{R}^{d_l \times d_{l+1,k}})$ is the transformation matrix, and $w_{ij}^{k,l}$ is the attention score of node v_j received by the node v_i in the k -th head. The attention score $w_{ij}^{k,l}$ is obtained by summing up the attention scores in $d_{l+1,k}$ channels $\hat{w}_{ij}^{k,l} (\hat{w}_{ij}^{k,l} \in \mathbb{R}^{d_{l+1,k}})$, followed by the Softmax activation:

$$w_{ij}^{k,l} = \text{softmax}(\hat{w}_{ij}^{k,l} \bullet \mathbf{1}) \quad (2)$$

where \bullet denotes the dot product between two vectors, $\mathbf{1}$ is a $d_{l+1,k}$ -dimensional vector with every element being 1, and $\hat{w}_{ij}^{k,l} (\hat{w}_{ij}^{k,l} \in \mathbb{R}^{d_{l+1,k}})$ is defined as

$$\hat{w}_{ij}^{k,l} = \frac{(Q^{k,l} h_{v_i}^l) \odot (K^{k,l} h_{v_j}^l)}{\sqrt{d_{l+1,k}}} \quad (3)$$

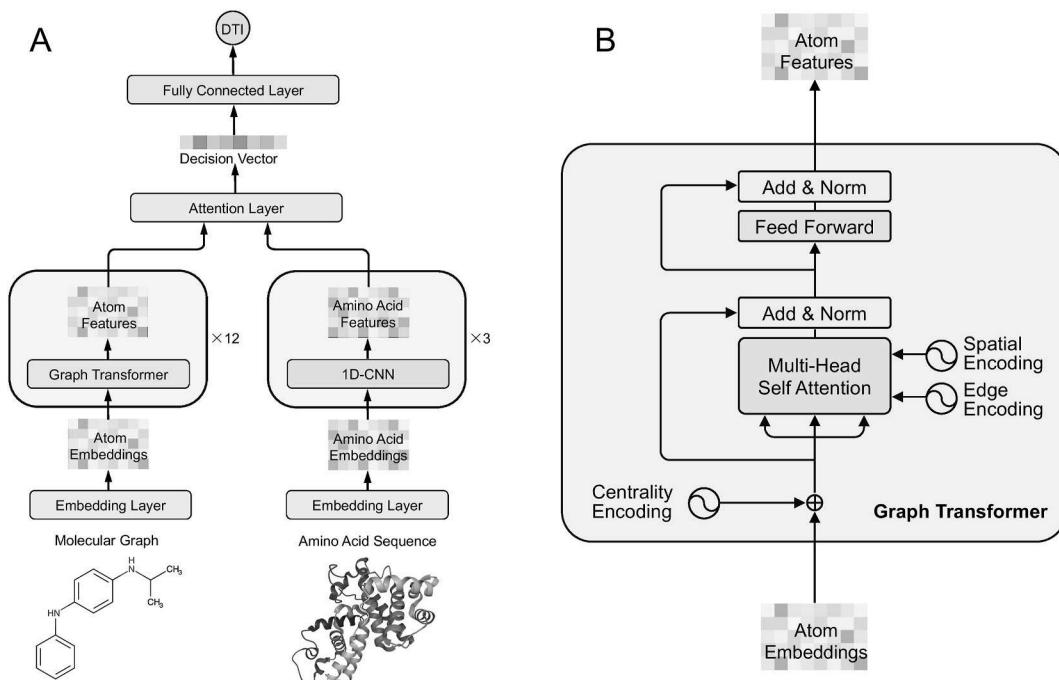


Fig. 1. The architecture of the proposed GraphomerDTI approach. (A) The overall framework of GraphomerDTI. (B) The architecture of the Graph Transformer layer.

where \odot represents the element-wise product between two vectors, and $Q^{k,l}, K^{k,l}, V^{k,l} \in \mathbb{R}^{d_l \times d_{l+1,k}}$ are the transformation matrices.

The output representation for atom $v_i \in V$ in l -th layer is obtained by concatenating the updated representations in different heads, $h_{v_i}^{l+1,k} (h_{v_i}^{l+1,k} \in \mathbb{R}^{d_l \times d_{l+1,k}})$, followed by a linear transformation:

$$h_{v_i}^{l+1} = O_h^l \parallel h_{v_i}^{l+1,k} \quad k=1 \quad (4)$$

where \parallel represents the concatenation operation, H denotes the number of heads, and O_h^l refers to the transformation matrix.

Based on the above Graph Transformer layer, following the work by Ying et al. [36], we added three important structural encoding components, i.e., node centrality encoding, node spatial encoding, and edge encoding, to fully capture the structural semantics in molecular graphs. In the upgraded Graph Transformer layer, the multi-channel attention score in Eq. (3) is expressed as

$$\hat{w}_{ij}^{k,l} = \begin{cases} \frac{(Q^{k,l}h_{v_i}^l) \odot (K^{k,l}h_{v_j}^l)}{\sqrt{d_{l+1,k}}} + b_{ij}^{k,l} + E^{k,l}h_{(v_i,v_j)}^l & \text{if } \varphi(v_i, v_j) \leq N \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where the node centrality encodings are used to augment the input atom representations $h_{v_i}^0$, $\varphi(v_i, v_j)$ denotes the shortest path distance between the atoms v_i and v_j with N being the pre-determined distance threshold, $b_{ij}^{k,l} \in \mathbb{R}^{d_{l+1,k}}$ is the spatial encoding for modelling the relatedness between the atoms v_i and v_j , $E^{k,l}$ is the transformation matrix, and $h_{(v_i,v_j)}^l$ is the l -th layer representation for the edge connecting atoms v_i and v_j . A detailed description of the three structural encoding components is provided in the Supplementary file.

Based on these, through the multiple stacked Graph Transformer layers, the informative structural characteristics of molecular graphs can be effectively encoded into atom representations. As illustrated in Fig. 2, the molecular subgraphs surrounding the highlighted carbon atoms of Theobromine (DrugBank ID: DB01412) in Fig. 2A can be transformed into the discriminative atom representations in Fig. 2B. The learned discriminative atom representations will then be utilized for the downstream DTI prediction.

2.1.2. Protein representation learning

We used amino acid sequences to describe target proteins. For a target protein, we represented it as amino acid sequence $P = (p_1, p_2, \dots, p_m)$, where p_i is the i -th amino acid and m is the number of amino acids in the protein.

We used 1D-CNN to encode amino acid subsequence patterns into amino acid representations. Like the drug representation learning component, the protein representation learning component also starts

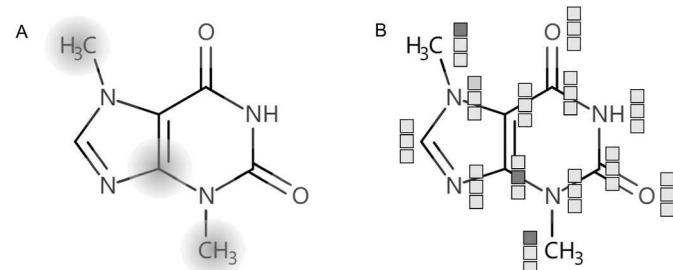


Fig. 2. Graphical illustration showing the molecular subgraph features and their correspondence to the discriminative atom representations. (A) Atoms centred by molecular subgraphs. (B) Discriminative atom representations.

from an embedding layer. Since there are 20 different amino acids in proteins, we constructed a learnable embedding table for the 20 amino acids, based on which each amino acid is encoded into a learnable embedding vector. In this way, the input protein is first represented as a collection of amino acid embedding vectors, which are then fed into the CNN-based protein representation learning component. The CNN-based protein representation learning module contains three consecutive 1D-CNN layers, which can effectively extract local subsequence patterns along the whole input amino acid sequence.

The 1D-CNN layer was implemented with a fixed-size convolution kernel sliding along the input sequence, which can capture the amino acid co-occurrence patterns within a predefined window size, informatively representing the protein. Consequently, the output is a sequence of updated amino acid feature vectors. Through stacking multiple 1D-CNN layers, more complex amino acid co-occurrence patterns can be encoded into amino acid representations, playing a critical role in accurately predicting DTIs.

2.1.3. Interaction learning

As the output of the drug representation learning component, the drug molecule is encoded into a feature matrix denoted as $D \in \mathbb{R}^{n \times d_f}$, where n represents the number of atoms of the molecule, d_f is the dimension of the final atom representations, and the i -th column of D , $d_i (d_i \in \mathbb{R}^{d_f})$, is the final representation of the i -th atom v_i . Similarly, with the protein representation learning component, the target protein is encoded into another feature matrix $P \in \mathbb{R}^{m \times d_f}$, where m denotes the number of amino acids in the protein, d_f is the dimension of the final amino acid representations and the j -th column of P , $p_j \in \mathbb{R}^{d_f}$, is the final representation of the j -th amino acid.

To capture the interactions between molecular atoms and protein segments, we employ the association attention algorithm [34] to transform the initial drug and protein representations D and P into the attention-aware representations D_a ($D_a \in \mathbb{R}^{n \times d_f}$) and D_p ($D_p \in \mathbb{R}^{m \times d_f}$), respectively, which capture the important interacting molecular atoms and protein segments. A detailed description of the drug-target interaction learning workflow is provided in the supplementary file.

By applying a global max pooling operation on D_a and P_a respectively, we then obtained the final representation of input drug $x_d (x_d \in \mathbb{R}^{d_f})$ and final representation of input protein $x_p (x_p \in \mathbb{R}^{d_f})$. Finally, x_d and x_p are concatenated together ($x_d \parallel x_p$) and fed into a multi-layer fully connected neural network (FCNN) to predict whether there exists an interaction between the input drug molecule and input target protein.

The overall DTI prediction model is trained by minimizing the following cross-entropy loss function:

$$\mathcal{L} = -[y \log(\hat{y}) + (1-y)\log(1-(\hat{y}))], \quad (7)$$

where y is the ground-truth interaction label, with “1” denoting the input drug has an interaction with the input target protein and “0” meaning there is no interaction between them, respectively, and \hat{y} represents the predicted likelihood of the interaction’s existence.

Table 1 provides a detailed description of the algorithm for training the proposed GraphomerDTI model: First, the model parameters are initialized with random numbers; Next, they are updated with the stochastic gradient descent by iteratively sampling a batch of training DTIs; Finally, the trained GraphomerDTI model is returned for DTI prediction.

2.2. Experiments

In this section, we conducted extensive experiments on three real-world benchmark datasets to verify the efficacy of the proposed GraphomerDTI model. We implemented the GraphomerDTI model using the Deep Graph Library (DGL) [37].

Table 1
Algorithm description for training GraphomerDTI.

Input:	Training DTIs as a set of DTI tuples (including drugs, targets and interaction labels).
Output:	The trained GraphomerDTI model.
1	Initialize model parameters with random numbers;
2	Repeat
3	Sample a batch of training DTIs;
4	Generate drug representations with Graph Transformer by Eq. (1);
5	Generate target representations with 1D-CNN;
6	Upgrade drug and target representations by interaction learning;
7	Update model parameters by minimizing the loss in Eq. (7) with gradient descent;
8	Until the model converges or a number of iterations is reached;
9	Return the trained GraphomerDTI model.

2.2.1. Benchmark datasets

We benchmarked performance on three commonly used datasets: DrugBank [38], Davis [39] and KIBA [40], each of which includes numerous drug molecules, target proteins, and more notably, their interactions and non-interactions. The statistics of the DrugBank, Davis and KIBA benchmarks are summarized in Table 2. More details on construction of the benchmark datasets are provided in the supplementary file.

2.2.2. Baseline methods

To verify the efficacy of the proposed GraphomerDTI model, we compared it with the following six state-of-the-art baseline methods that also used the advanced molecular graph and SMILES schemes to represent drug molecules:

- **GNN-CPI** [21] uses a GNN and a 1D-CNN to respectively learn molecular and protein representations from molecular graphs and protein sequences. The learned molecular and protein representations are then concatenated together to predict the interactions between molecules and proteins.
- **GNN-PT** [23] utilizes a GNN to learn drug representations and a Transformer to learn protein representations, which takes advantage of the self-attention mechanism to capture long-distance dependencies between amino acid residues.
- **DeepEmbedding-DTI** [22] first uses BERT [41] to learn protein subsequence features from protein sequences and then uses a bidirectional LSTM (BiLSTM) to learn protein representations. A local breadth-first search (BFS) based GNN is used to learn molecular representations from molecular graphs.
- **MolTrans** [33] transforms molecular SMILES strings and protein sequences into sequences of components, with each component being a subsequence occurring frequently, then uses Transformers to learn molecular and protein representations from the transformed sequences.
- **HyperAttentionDTI** [34] first uses two 1D-CNNs to learn molecular and protein representations from molecular SMILES strings and protein sequences. An attention mechanism is then employed to capture complex interactions between molecular and protein representations.
- **AttentionSiteDTI** [42] constructs molecular and protein representations with the Topology Adaptive GCN (TAGCN) [43] and uses the

self-attention mechanism [32] to capture the interactions between the molecular and protein representations.

All the six baseline methods were re-trained on the same training dataset of DTIs with the default parameter configurations. A detailed description of their implementation is provided in the supplementary file.

2.2.3. Experimental settings

To evaluate the DTI prediction performance of different models, we split each of the benchmark datasets into a training set and a test set, then trained the DTI prediction models on the training set and evaluated their performance on the test set. Specifically, by considering various drug screening cases, we applied the following three training/test set split settings:

• **Transductive setting:** The training and test DTIs have overlapping drugs and targets. For each test DTI, both its drug and target have at least one other DIT record in the training set.

• **Drug inductive setting:** The training and test DTIs have only overlapping targets but have no overlapping drugs. For each test DTI, its drug is new (i.e., unseen) to the training DTIs.

• **Drug-target inductive setting:** The training and test DTIs do not have any overlap for both drugs and targets. For each test DTI, its drug and target are both new (i.e., unseen) to the training DTIs.

Fig. 3 illustrates the differences between the three settings, which have different overlapping status of drugs and targets between the training and test sets. In the supplementary file, we provide a detailed description of how the three training/test set splittings were performed using the three benchmark datasets.

Under the three settings, the random training/test set splitting is repeated for 5 times on each benchmark dataset. As such, the averaged evaluation metrics and standard deviations are reported as the final results.

2.3. Performance evaluation metrics

In this study, the DTI prediction is actually a binary classification task. To evaluate its performance, we used the following four frequently used evaluation metrics, where TP, FP, TN and FN denote the numbers of true positive, false positive, true negative and false negative samples, respectively.

- **F1-Score** is the harmonic mean of precision and recall (i.e., $F1\text{-Score} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$), where $\text{precision} = \text{TP} / (\text{TP} + \text{FP})$ and $\text{recall} = \text{TP} / (\text{TP} + \text{FN})$.
- **AUC** is the area under the Receiver Operating Characteristic (ROC) curve. It measures the probability that a machine learning model ranks a random positive sample higher than a random negative sample.
- **AUPR** is the area under the precision-recall curve. It evaluates a machine learning model's capability to retrieve all positive samples (perfect recall) yet avoid predicting any negative samples as positive samples (perfect precision).
- **MCC** is Matthew's correlation coefficient. It evaluates the consistency between the predicted labels and ground-truth labels:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

For all of the four performance metrics, a higher score indicates that the predicted DTI is more consistent with the ground-truth DTI, i.e., the model with higher evaluation scores is more reliable for real-world drug screening.

Table 2

Statistical summary of the benchmark datasets used in this study.

Benchmark dataset	No. Proteins	No. Drugs	No. Interactions	No. Non-interactions
DrugBank	4293	6561	17,291	17,291
Davis	379	67	7119	18,274
KIBA	225	2058	22,137	94,082

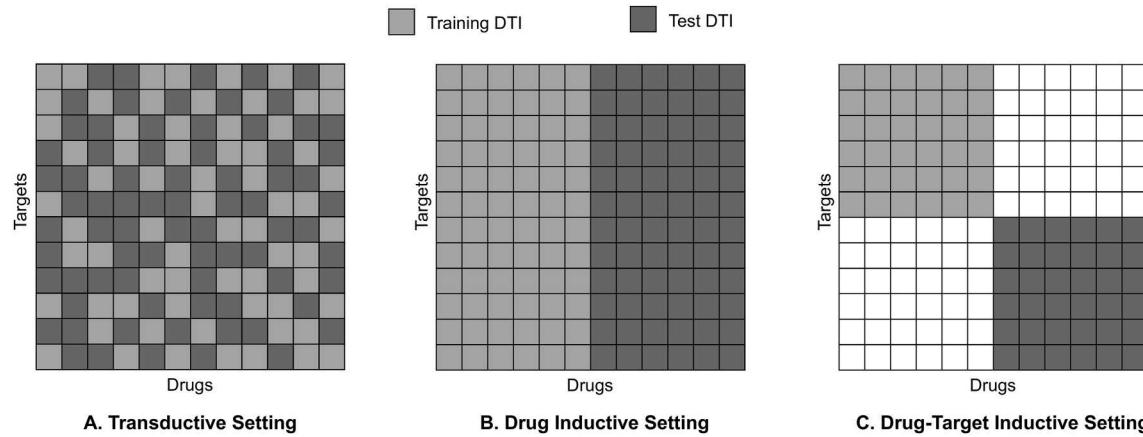


Fig. 3. Graphical illustration of the three different settings for DTI prediction. (A) Transductive setting. (B) Drug inductive setting. (C) Drug-target inductive setting.

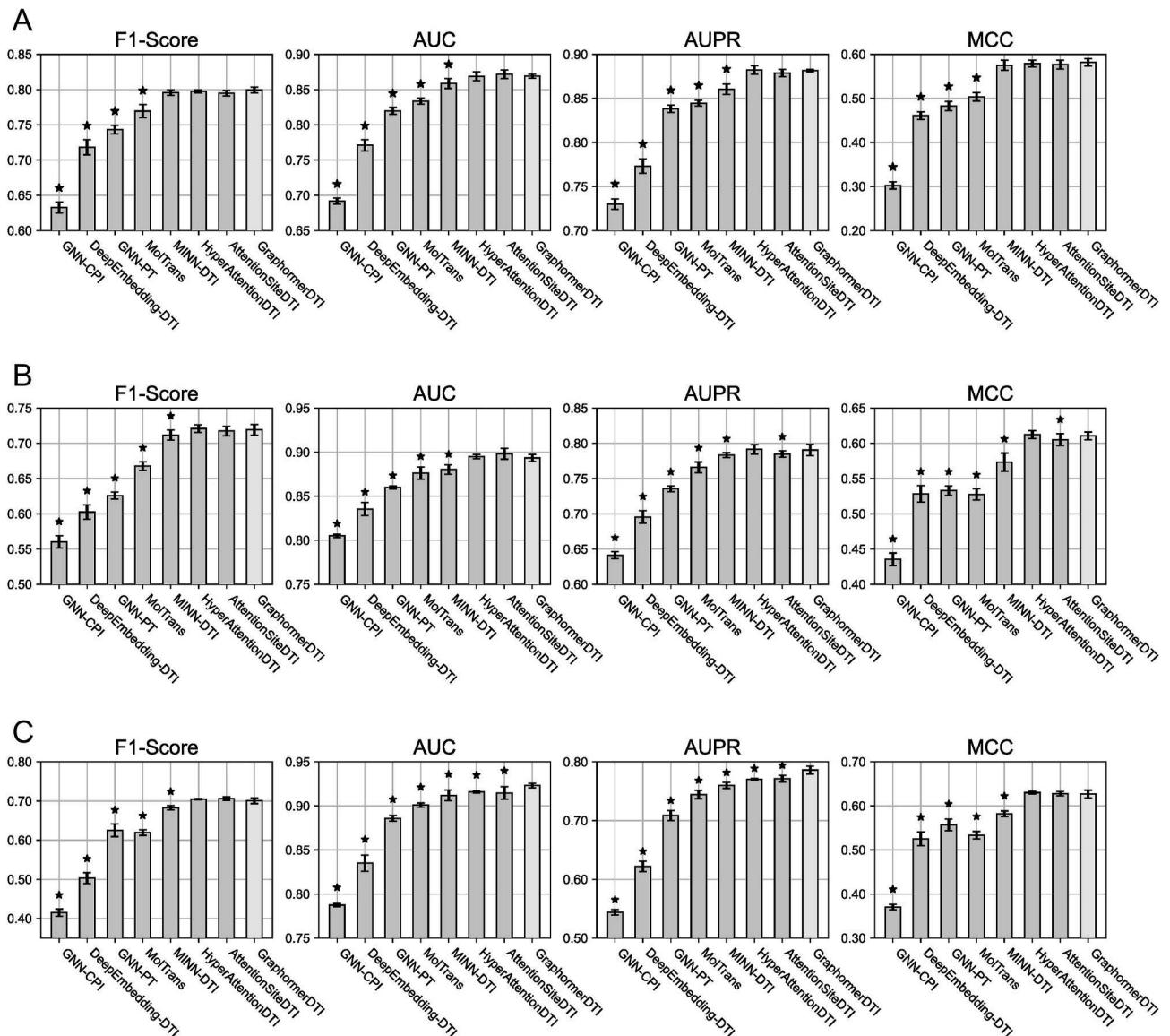


Fig. 4. Performance comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the transductive setting. (A) The results on the DrugBank dataset. (B) The results on the Davis dataset. (C) The results on the KIBA dataset.

3. Results

In this study, we compared the DTI prediction performance of different algorithms based on the DrugBank, Davis and KIBA benchmarks under the three different settings. Figs. 4–6 respectively show the experimental results under the three settings. For each metric, we also conducted the paired *t*-test between the best performer and its competitors. The performers significantly inferior to the best performer at 0.05 significance level are marked with “★”. Based on the results, we concluded that the proposed GraphomerDTI approach achieved the best overall performance. To further prove our conclusion, Tables 3–5 provide specific values, with the highest values in bold for each condition.

3.1. Performance comparison under the transductive setting

Fig. 4 provides the DTI prediction performance of the GraphomerDTI model and baseline methods under the transductive setting on the three benchmark datasets. As is shown in Fig. 4, the proposed GraphomerDTI model significantly outperforms most of the baseline

methods and is on a par with the best baseline method termed HyperAttentionDTI. The detailed performance comparison results in terms of F1-score, AUC, AUPR and MCC are provided in Table 3. From the Tables, we can see that GraphomerDTI outperformed HyperAttentionDTI and AttentionSiteDTI in predicting drug molecules with unknown interactions. However, its performance was worse when it came to predicting known interactions. There might exist two possible reasons: The first reason might be that HyperAttentionDTI and AttentionSiteDTI with simple network architectures tended to be easily overfitting to the data of known interactions during the training process, potentially resulting in favourable performance under the transductive setting but inferior performance under the inductive setting. The second reason might be that there existed a limited number of known interactions in the training data, leading to inferior optimization of graph neural networks. Overall, we conclude that Graph Transformer has a remarkable ability to extract informative molecular characteristics from molecular graphs for accurate transductive DTI prediction. With the referred drugs and targets in training set, DTI prediction under the transductive setting is an easy task, making it hard to distinguish GraphomerDTI and HyperAttentionDTI. Hence, we also conducted performance comparison under

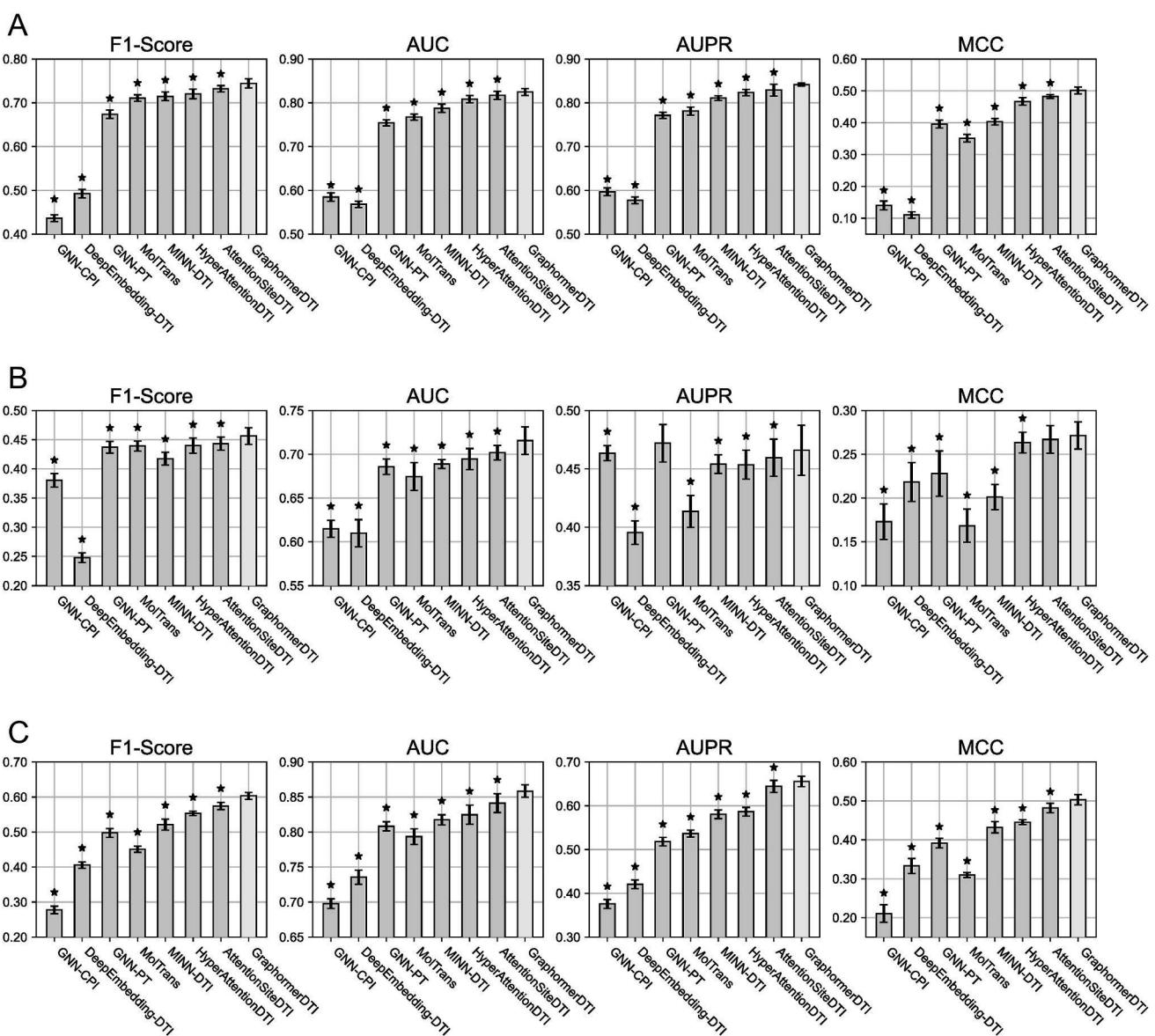


Fig. 5. Performance comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the drug inductive setting. (A) The results on the DrugBank dataset. (B) The results on the Davis dataset. (C) The results on the KIBA dataset.

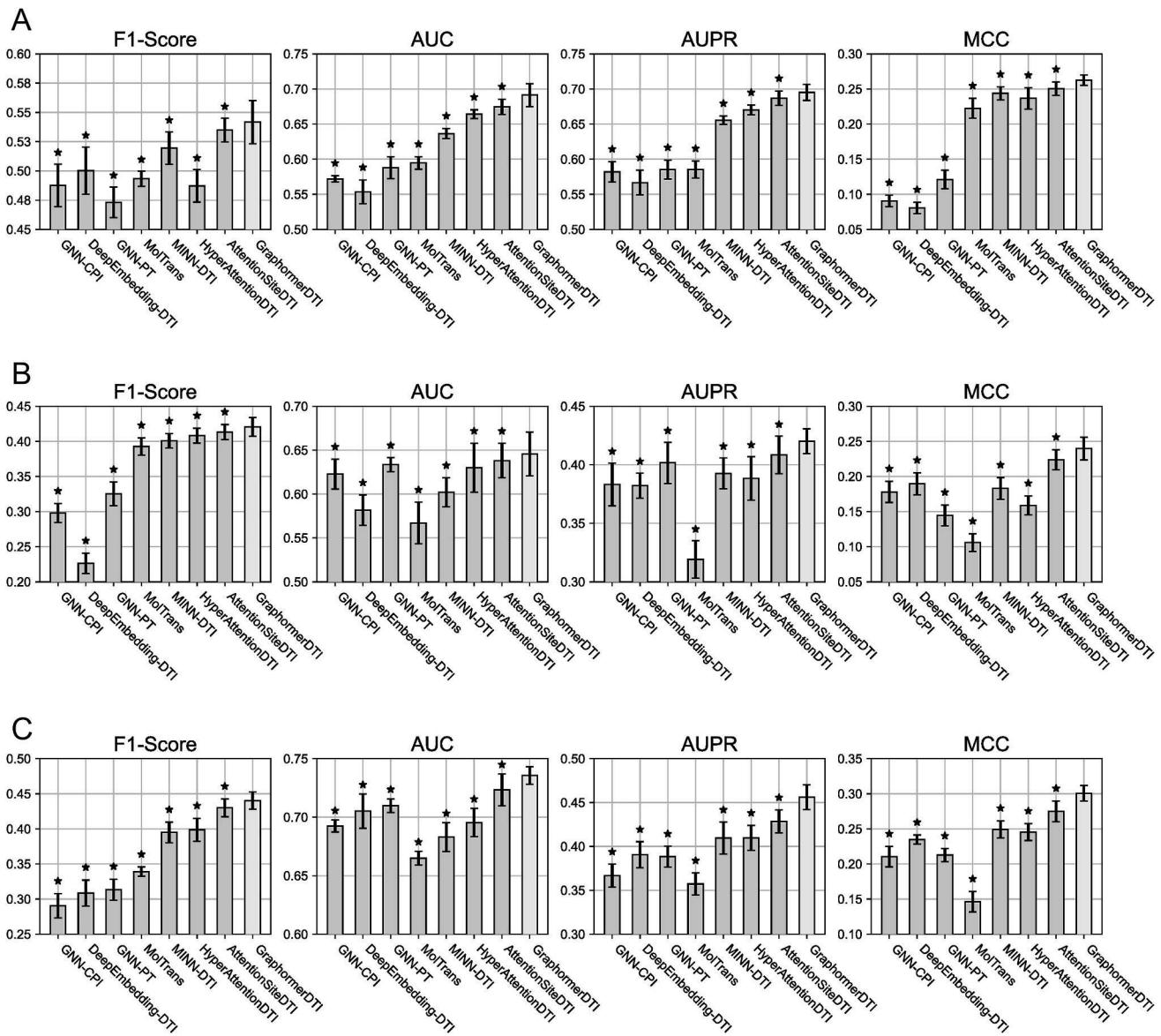


Fig. 6. Performance comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the drug-target inductive setting. (A) The results on the DrugBank dataset. (B) The results on the Davis dataset. (C) The results on the KIBA dataset.

the more challenging inductive settings.

3.2. Performance comparison under the drug inductive setting

Fig. 5 shows the DTI performance comparison under the drug inductive setting on the three datasets. From Fig. 5, we can see that the proposed GraphomerDTI outperforms all other baseline methods. Table 4 provides the detailed performance comparison results. The results in Table 4 show that GraphomerDTI achieved a better performance compared with other methods in terms of F1-score, AUC, AUPR, and MCC. The results suggest that GraphomerDTI has a strong capability of processing drug molecular data with unknown interactions and as such, it may have more suitable applications in drug design and discovery. Featured by the well-designed node centrality encoding, node spatial encoding, and edge encoding components, GraphomerDTI can effectively capture the essential graph inductive bias. This enables its trained model to better extrapolate representations for the out-of-sample drug molecules. The informative molecular representations then contribute to the superior DTI prediction performance of GraphomerDTI over baseline methods.

3.3. Performance comparison under the drug-target inductive setting

Fig. 6 provides a comparison of the DTI prediction performance of different models under the drug-target inductive setting on the three datasets and Table 5 provides the detailed performance comparison results. Under this setting, the proposed GraphomerDTI method also achieved the best overall performance in terms of F1-Score, AUC, AUPR and MCC. This further highlights the effectiveness of the molecular representations learned by the GraphomerDTI model, with the improved DTI prediction for the out-of-sample drug molecules with regard to both in-sample and out-of-sample target proteins. The results suggest that GraphomerDTI can be a useful computational drug screening tool for emerging diseases with few or no effective drug treatments. Table 6 shows the importance of the three structural coding components. The absence of any of the structural coding components resulted in reduced performance.

3.4. Normalized confusion matrix visualization

To study the DTI prediction performance of the proposed

Table 3

Statistical comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the transductive setting on the three benchmark datasets.

Benchmark dataset	Methods	F1-Score (std)	AUC (std)	AUPR (std)	MCC (std)
DrugBank	GNN-CPI	0.633 (0.008)	0.692 (0.004)	0.730 (0.006)	0.302 (0.008)
	DeepEmbedding-DTI	0.718 (0.011)	0.771 (0.008)	0.773 (0.008)	0.461 (0.009)
	GNN-PT	0.743 (0.006)	0.820 (0.005)	0.838 (0.004)	0.483 (0.010)
	MolTrans	0.770 (0.009)	0.834 (0.004)	0.845 (0.003)	0.504 (0.010)
	MINN-DTI	0.796 (0.004)	0.859 (0.007)	0.861 (0.006)	0.575 (0.011)
	HyperAttentionDTI	0.797 (0.002)	0.869 (0.006)	0.882 (0.005)	0.579 (0.008)
	AttentionSiteDTI	0.795 (0.004)	0.872 (0.006)	0.879 (0.004)	0.577 (0.010)
	GraphomerDTI	0.799 (0.004)	0.869 (0.003)	0.882 (0.001)	0.582 (0.008)
Davis	GNN-CPI	0.560 (0.009)	0.805 (0.002)	0.641 (0.005)	0.435 (0.009)
	DeepEmbedding-DTI	0.603 (0.010)	0.835 (0.007)	0.696 (0.009)	0.528 (0.012)
	GNN-PT	0.626 (0.005)	0.860 (0.002)	0.736 (0.004)	0.533 (0.007)
	MolTrans	0.668 (0.006)	0.876 (0.007)	0.766 (0.008)	0.528 (0.008)
	MINN-DTI	0.712 (0.007)	0.880 (0.005)	0.783 (0.003)	0.573 (0.013)
	HyperAttentionDTI	0.721 (0.005)	0.895 (0.003)	0.792 (0.007)	0.613 (0.006)
	AttentionSiteDTI	0.718 (0.007)	0.898 (0.006)	0.785 (0.005)	0.605 (0.009)
	GraphomerDTI	0.719 (0.008)	0.893 (0.004)	0.791 (0.008)	0.611 (0.006)
KIBA	GNN-CPI	0.415 (0.009)	0.787 (0.002)	0.544 (0.005)	0.370 (0.006)
	DeepEmbedding-DTI	0.503 (0.014)	0.835 (0.009)	0.622 (0.009)	0.525 (0.015)
	GNN-PT	0.625 (0.016)	0.886 (0.003)	0.709 (0.009)	0.557 (0.013)
	MolTrans	0.620 (0.007)	0.901 (0.002)	0.744 (0.007)	0.534 (0.009)
	MINN-DTI	0.683 (0.006)	0.912 (0.006)	0.760 (0.005)	0.582 (0.007)
	HyperAttentionDTI	0.705 (0.002)	0.916 (0.001)	0.770 (0.002)	0.630 (0.003)
	AttentionSiteDTI	0.706 (0.004)	0.915 (0.007)	0.771 (0.006)	0.628 (0.005)
	GraphomerDTI	0.701 (0.007)	0.923 (0.003)	0.786 (0.007)	0.627 (0.009)

GraphomerDTI model in more details, we visualized normalized confusion matrices of GraphomerDTI predictions. Fig. 7 presents the plot of the normalized confusion matrices on DrugBank, Davis and KIBA for one training/test split under the transductive and drug inductive settings. From Fig. 7, we can see that the proposed GraphomerDTI yields reasonably good prediction performance, except for the imbalanced Davis dataset under the drug inductive setting. As predicting DTIs for novel out-of-sample molecules is a challenging task, we can observe an obvious performance drop by comparing Fig. 7B with Fig. 7A. Nevertheless, GraphomerDTI still achieved satisfactory performance on predicting DTIs for novel molecules on both DrugBank and KIBA.

To demonstrate the performance for individual targets, Fig. 8 shows the normalized confusion matrices for individual targets under the transductive and drug inductive settings. Taking the target P11217 in DrugBank, target EPHA4 in Davis, and target Q9HAZ1 in KIBA as examples, most molecules could still be accurately predicted despite the significant performance decrease under the drug inductive setting in Fig. 8B compared with the transductive setting in Fig. 8A. In conclusion,

Table 4

Statistical comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the drug inductive setting on the three benchmark datasets.

Benchmark dataset	Methods	F1-Score (std)	AUC (std)	AUPR (std)	MCC (std)
DrugBank	GNN-CPI	0.437 (0.008)<	0.585 (0.010)	0.597 (0.009)	0.140 (0.014)
	DeepEmbedding-DTI	0.493 (0.010)	0.568 (0.007)	0.577 (0.008)	0.110 (0.010)
	GNN-PT	0.674 (0.009)	0.754 (0.007)	0.771 (0.007)	0.396 (0.012)
	MolTrans	0.711 (0.008)	0.767 (0.007)	0.781 (0.009)	0.351 (0.012)
	MINN-DTI	0.715 (0.010)	0.788 (0.010)	0.811 (0.005)	0.403 (0.010)
	HyperAttentionDTI	0.720 (0.011)	0.809 (0.008)	0.823 (0.007)	0.467 (0.012)
	AttentionSiteDTI	0.732 (0.007)	0.817 (0.009)	0.829 (0.014)	0.482 (0.006)
	GraphomerDTI	0.745 (0.010)	0.825 (0.008)	0.841 (0.003)	0.501 (0.011)
	Davis	GNN-CPI	0.380 (0.012)	0.615 (0.010)	0.464 (0.006)
	DeepEmbedding-DTI	0.248 (0.009)	0.610 (0.016)	0.395 (0.010)	0.218 (0.022)
Davis	GNN-PT	0.437 (0.010)	0.686 (0.009)	0.472 (0.016)	0.228 (0.026)
	MolTrans	0.439 (0.009)	0.675 (0.016)	0.414 (0.014)	0.168 (0.019)
	MINN-DTI	0.417 (0.011)	0.689 (0.005)	0.454 (0.008)	0.201 (0.014)
	HyperAttentionDTI	0.440 (0.013)	0.695 (0.012)	0.454 (0.012)	0.263 (0.012)
	AttentionSiteDTI	0.443 (0.011)	0.702 (0.008)	0.460 (0.016)	0.267 (0.016)
	GraphomerDTI	0.456 (0.014)	0.716 (0.016)	0.466 (0.022)	0.271 (0.016)
	KIBA	GNN-CPI	0.278 (0.011)	0.698 (0.007)	0.376 (0.010)
	DeepEmbedding-DTI	0.406 (0.009)	0.736 (0.010)	0.421 (0.010)	0.333 (0.019)
	GNN-PT	0.498 (0.012)	0.808 (0.007)	0.518 (0.010)	0.391 (0.013)
	MolTrans	0.451 (0.009)	0.794 (0.011)	0.537 (0.008)	0.309 (0.007)
KIBA	MINN-DTI	0.521 (0.016)	0.818 (0.007)	0.581 (0.010)	0.432 (0.015)
	HyperAttentionDTI	0.553 (0.006)	0.825 (0.014)	0.586 (0.010)	0.445 (0.006)
	AttentionSiteDTI	0.574 (0.010)	0.841 (0.014)	0.644 (0.014)	0.482 (0.012)
	GraphomerDTI	0.603 (0.010)	0.858 (0.009)	0.656 (0.012)	0.503 (0.013)

GraphomerDTI performs well in predicting DTIs for novel drug molecules.

3.5. Ablation studies

To verify the importance of the three structural encoding components of GraphomerDTI, we compared the performance of the full GraphomerDTI model with that of its three ablated variants, i.e., without node centrality encoding (“w/o centrality encoding”), without node spatial encoding (“w/o spatial encoding”) and without edge encoding (“w/o edge encoding”). We also compared the GraphomerDTI model with the ablated version where the Graph Transformer component was replaced by the vanilla GCN model [17]. Fig. 9 shows the DTI prediction performance of the GraphomerDTI model and its ablated versions on the DrugBank dataset, under the transductive and drug inductive settings. For each metric, we also conducted the paired *t*-test between the best performer and its competitors. The performers significantly worse than the best performer at 0.05 significance level are

Table 5

Statistical comparison of GraphomerDTI and other baseline methods in terms of F1-score, AUC, AUPR and MCC under the drug-target inductive setting on the three benchmark datasets.

Benchmark dataset	Methods	F1-Score (std)	AUC (std)	AUPR (std)	MCC (std)
DrugBank	GNN-CPI	0.488 (0.018)	0.572 (0.005)	0.582 (0.015)	0.091 (0.008)
	DeepEmbedding-DTI	0.500 (0.020)	0.554 (0.017)	0.567 (0.018)	0.081 (0.008)
	GNN-PT	0.473 (0.013)	0.588 (0.016)	0.585 (0.014)	0.121 (0.013)
	MolTrans	0.493 (0.006)	0.595 (0.009)	0.585 (0.012)	0.223 (0.014)
	MINN-DTI	0.520 (0.014)	0.637 (0.007)	0.656 (0.006)	0.244 (0.009)
	HyperAttentionDTI	0.487 (0.014)	0.664 (0.006)	0.670 (0.007)	0.237 (0.015)
	AttentionSiteDTI	0.535 (0.010)	0.675 (0.011)	0.687 (0.010)	0.251 (0.010)
	GraphomerDTI	0.542 (0.019)	0.691 (0.016)	0.695 (0.011)	0.263 (0.007)
	GNN-CPI	0.298 (0.014)	0.623 (0.017)	0.383 (0.018)	0.178 (0.015)
	DeepEmbedding-DTI	0.226 (0.014)	0.582 (0.017)	0.382 (0.011)	0.190 (0.016)
Davis	GNN-PT	0.325 (0.017)	0.634 (0.008)	0.402 (0.018)	0.145 (0.015)
	MolTrans	0.393 (0.012)	0.567 (0.024)	0.319 (0.016)	0.106 (0.013)
	MINN-DTI	0.401 (0.010)	0.602 (0.017)	0.393 (0.013)	0.183 (0.016)
	HyperAttentionDTI	0.408 (0.011)	0.630 (0.028)	0.388 (0.019)	0.159 (0.014)
	AttentionSiteDTI	0.413 (0.011)	0.638 (0.020)	0.408 (0.016)	0.224 (0.014)
	GraphomerDTI	0.421 (0.014)	0.646 (0.025)	0.420 (0.011)	0.240 (0.016)
	GNN-CPI	0.290 (0.017)	0.693 (0.005)	0.367 (0.013)	0.210 (0.015)
	DeepEmbedding-DTI	0.309 (0.019)	0.705 (0.015)	0.391 (0.015)	0.235 (0.007)
	GNN-PT	0.313 (0.015)	0.710 (0.006)	0.388 (0.012)	0.213 (0.010)
	MolTrans	0.339 (0.007)	0.665 (0.006)	0.357 (0.013)	0.146 (0.015)
KIBA	MINN-DTI	0.395 (0.015)	0.683 (0.012)	0.410 (0.018)	0.249 (0.012)
	HyperAttentionDTI	0.399 (0.016)	0.695 (0.012)	0.410 (0.014)	0.246 (0.012)
	AttentionSiteDTI	0.430 (0.013)	0.723 (0.014)	0.429 (0.013)	0.275 (0.015)
	GraphomerDTI	0.440 (0.012)	0.736 (0.008)	0.456 (0.014)	0.301 (0.011)

marked with “★”. At the same time, the detailed performance results in terms of F1-score, AUC, AUPR and MCC values are also provided in Tables 6, in which the maximum values for each metrics are highlighted in bold. From Fig. 9, we can see that the GraphomerDTI model performed best, while the ablated version consistently went through a significant performance drop under the two settings. These results demonstrate the advantages of the Graph Transformer over the vanilla GCN in capturing molecular structural characteristics. On the other hand, the results also show that all the three encoding components adopted by Graph Transformer are vital for the learning of informative molecular representations for accurate DTI prediction. Ablating any one of them would result in uninformative molecular representations and inferior DTI prediction performance.

3.6. Interaction visualization

To illustrate the effectiveness of the interaction learning component for DTI prediction, we visualized the important atoms of Theobromine

Table 6

The ablation study results under the transductive setting and the drug inductive setting on the DrugBank dataset.

Setting	Methods	F1-Score (std)	AUC (std)	AUPR (std)	MCC (std)
Transductive setting	GCN	0.621 (0.008)	0.821 (0.007)	0.795 (0.006)	0.526 (0.010)
	w/o centrality	0.786 (0.004)	0.854 (0.006)	0.863 (0.006)	0.548 (0.013)
	encoding	0.792 (0.005)	0.860 (0.004)	0.870 (0.005)	0.565 (0.010)
	w/o spatial	0.798 (0.003)	0.864 (0.002)	0.874 (0.004)	0.580 (0.004)
	encoding	0.799 (0.004)	0.869 (0.003)	0.882 (0.001)	0.582 (0.008)
	GraphomerDTI	0.799 (0.004)	0.869 (0.003)	0.882 (0.001)	0.582 (0.008)
	GCN	0.601 (0.013)	0.801 (0.009)	0.735 (0.008)	0.412 (0.012)
	w/o centrality	0.735 (0.011)	0.817 (0.004)	0.834 (0.004)	0.496 (0.009)
	encoding	0.744 (0.010)	0.820 (0.009)	0.827 (0.006)	0.500 (0.015)
	w/o spatial	0.743 (0.008)	0.821 (0.007)	0.836 (0.006)	0.498 (0.006)
Drug inductive setting	GraphomerDTI	0.745 (0.010)	0.825 (0.008)	0.841 (0.003)	0.501 (0.011)

(DrugBank ID: DB01412) and residue subsequences of Adenosine A₁ receptor (UniProt ID: P30542) that have high attention weights for predicting their interaction, as well as the binding residues of the Adenosine A₁ receptor predicted by the P2Rank toolbox [44] (a powerful protein binding pocket prediction toolbox) in Fig. 10. The GraphomerDTI model for predicting the interactions between Theobromine and Adenosine A₁ receptor was trained under the transductive setting. The attention weights were retrieved from the interaction learning component by feeding the drug and target into the trained GraphomerDTI model. From Fig. 10A and B, we can observe that three carbon atoms of Theobromine and residue subsequences (centred at the residue 15: ILE and residue 60: ALA) of Adenosine A₁ receptor are particularly important for predicting their interaction. By comparing Fig. 10B and C, we can find that the binding residues predicted by the GraphomerDTI model exhibited some consistency with the predictions produced by the P2Rank toolbox [43], showing that the proposed GraphomerDTI model could effectively capture the important interacting patterns between Theobromine and Adenosine A₁ receptor.

3.7. Case study

To further validate the reliability of the proposed GraphomerDTI model, we conducted a case study on DTI prediction regarding to the adrenergic receptors (adrenoceptors). We predicted drug molecules that interact with adrenergic receptors, and then validated with the ground-truth drug molecules that have been experimentally identified to interact with adrenergic receptors in the DrugBank database [38]. Adrenergic receptors are targets of norepinephrine and epinephrine, and a variety of drugs can bind to them. According to their different responses to norepinephrine, they are divided into alpha receptors and beta receptors [45,46]. There are two alpha receptor subtypes, each of which has three subclasses, with a total of six target proteins [45]. The beta receptor has three subtypes and three target proteins [46].

We compared the GraphomerDTI model with the HyperAttentionDTI model [17] for predicting drugs having interactions with the adrenergic receptors. The case study was operated on the DrugBank dataset. The drug molecules that interact with the target proteins (adrenergic receptors) in DrugBank were divided into the training and test set with the ratio of 4:1. We trained the two models using training molecules' DTIs and predicted interactions regarding to the nine adrenergic receptors for test molecules. The test drug molecules were

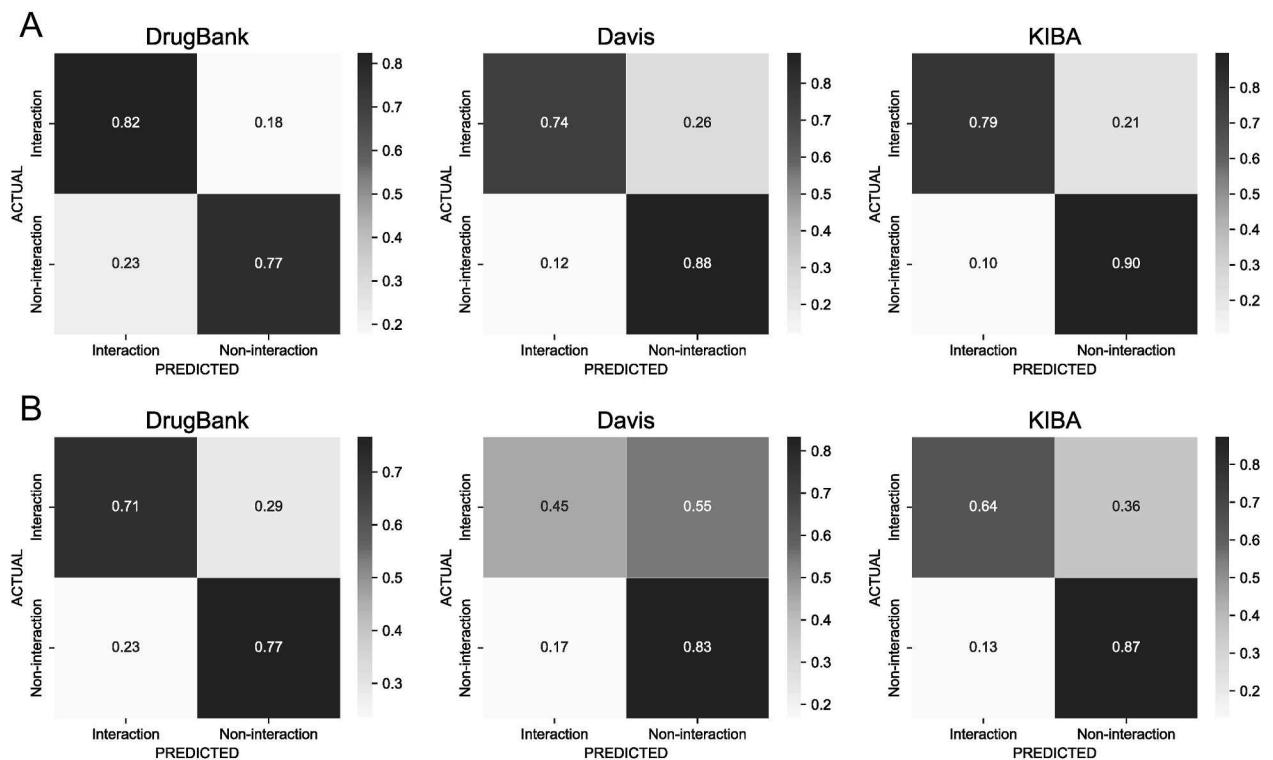


Fig. 7. Normalized confusion matrices of GraphomerDTI. (A) The normalized confusion matrices under the transductive setting. (B) The normalized confusion matrices under the drug inductive setting.

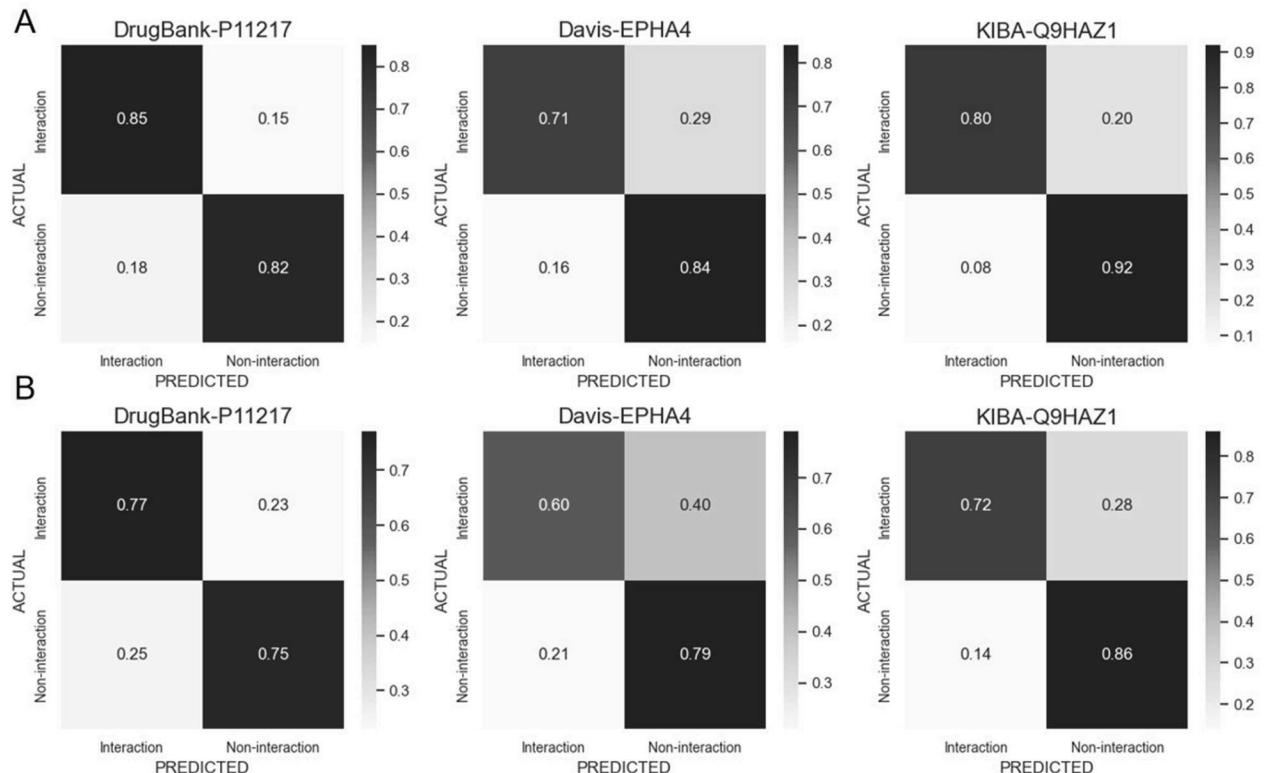


Fig. 8. Normalized confusion matrices for individual targets. (A) The normalized confusion matrices under the transductive setting. (B) The normalized confusion matrices under the drug inductive setting.

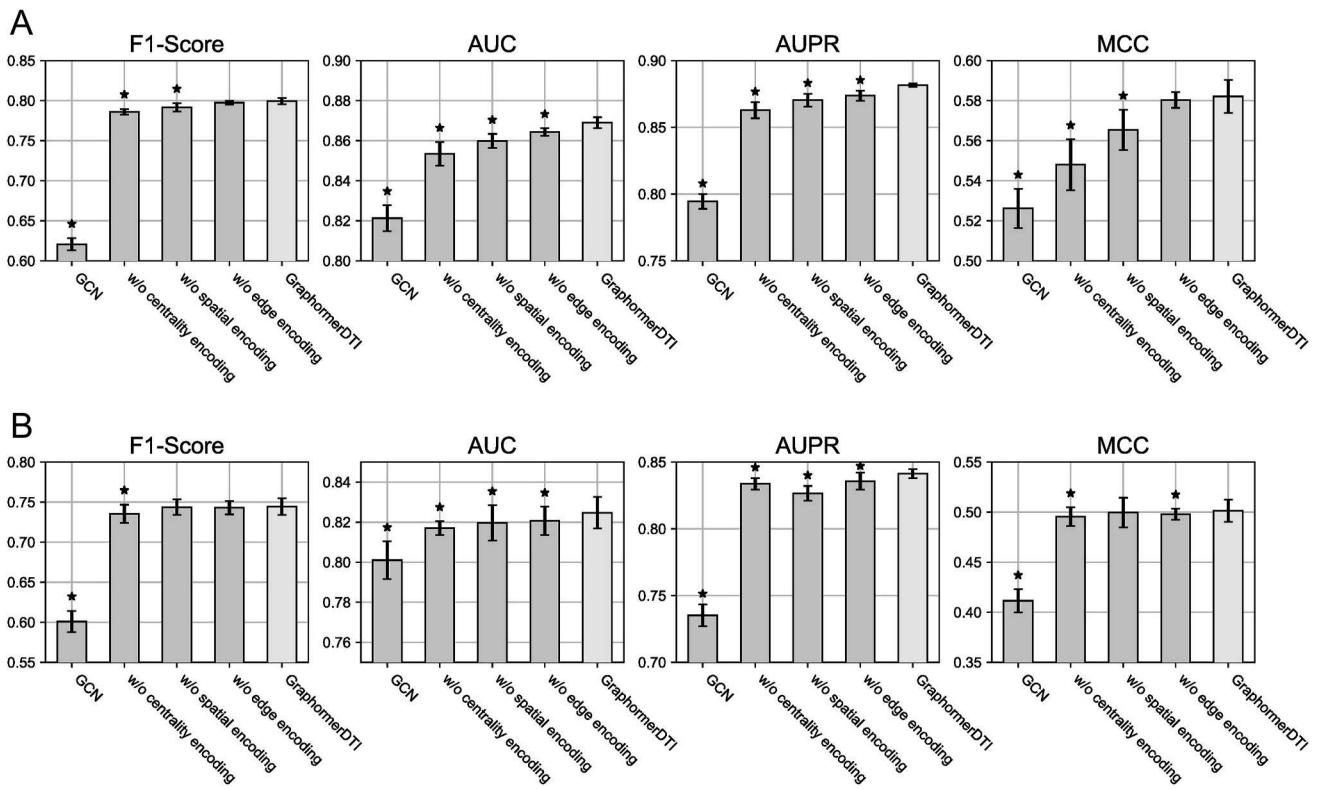


Fig. 9. The ablation study results on the DrugBank dataset. (A) The ablation study results under the transductive setting. (B) The ablation study results under the drug inductive setting.

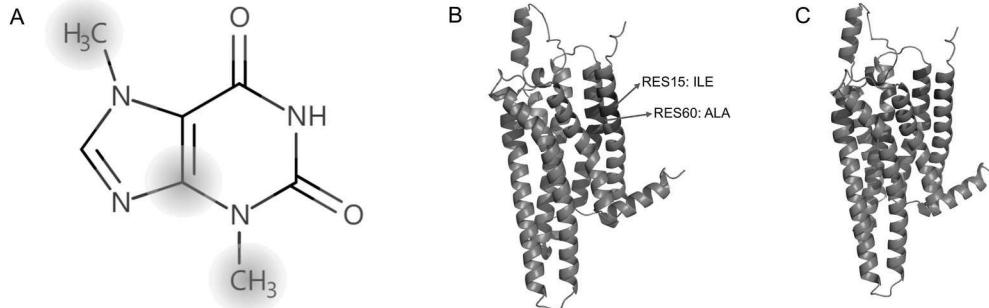


Fig. 10. Illustration of the important atoms and amino acids for predicting the interaction between Theobromine and Adenosine A₁ receptor. (A) Important atoms of Theobromine. (B) Important amino acids of Adenosine A₁ receptor. (C) Important binding residues identified by P2Rank.

then ranked according to their predicted interaction likelihoods. Regarding different adrenergic receptors, the comparison between GraphomerDTI and HyperAttentionDTI in terms of the number of hit molecules (that interact with the receptors) in the top 20 predictions is illustrated in Fig. 11A. In Fig. 11B, we also plotted the top 20 GraphomerDTI predictions for the molecules that interact with the Alpha-2B adrenergic receptor. As can be seen, GraphomerDTI correctly predicted more ground-truth drugs interacting with adrenergic receptors than HyperAttentionDTI, highlighting the greater predictive power of the GraphomerDTI on identifying effective drug molecules. This proves that GraphomerDTI has the promising potential to contribute to the related disease treatment.

4. Conclusion

In this paper, we proposed the GraphomerDTI model to predict the interactions between drug molecules and target proteins. GraphomerDTI used the Graph Transformer neural network to learn

informative molecular representations through encoding the essential structural characteristics of drug molecules, i.e., the importance of atoms, the structural distance between atoms, and motif subgraph patterns. In addition, 1D-CNN is used to learn informative protein representations, and an attention operation is leveraged to model the complex interactions between molecular and protein representations. The informative molecular representations learned by the proposed model contribute to the advanced DTI performance than the other five state-of-the-art methods. Finally, the real-world case study highlighted the exceptional predictive power of the proposed GraphomerDTI for the real-world out-of-molecule DTI prediction, providing the insights that the GraphomerDTI model is highly applicable for the real-world drug virtual screening and of a substantial practical value for precise medicine including that targeting adrenergic receptors related diseases.

GraphomerDTI can effectively learn informative features of target proteins from amino acid sequences. Furthermore, existing protein feature extraction methods and tools, such as BioSeq-BLM [47], BioSeq-Diabolo [48] and iFeatureOmega [49], hold great potential for

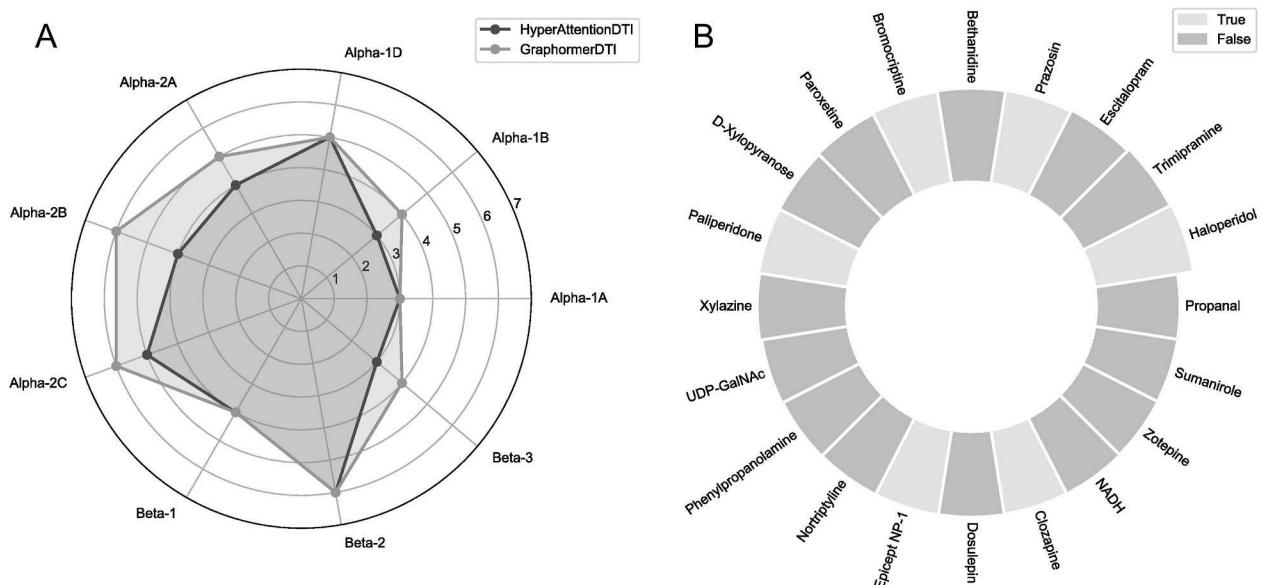


Fig. 11. The case study results. (A) The number of hit molecules with the top 20 molecules predicted by HyperAttentionDTI and GraphomerDTI for 9 adrenergic receptors. (B) The top 20 molecules predicted by GraphomerDTI for the Alpha-2B adrenergic receptor.

being integrated and enhanced in conjunction with GraphomerDTI. For instance, protein features extracted by BioSeq-BLM or BioSeq-Diabolo can serve as inputs to GraphomerDTI, generating a protein graph for subsequent DTI prediction. This strategy can leverage the advantages of protein feature extraction tools and introduce more potentially useful protein information into GraphomerDTI models, thereby improving the prediction performance. By combining protein features with the characterization capabilities of graph neural networks, protein structure and function information can be better captured, further improving the accuracy and reliability of DTI predictions. In addition, these protein feature extraction methods and tools can also be engaged in pre-processing steps to better prepare input data, such as reducing the dimensions of input data and enabling a more meaningful representation. We plan to employ these advanced methods in the future to achieve a more comprehensive characterization of protein data in order to further improve the accuracy of predictions and provide useful guidance for data-driven drug design and discovery.

5. Key points

- We propose a graph transformer-based deep learning framework, termed GraphomerDTI, for predicting the interactions between drug molecules and target proteins.
- Evaluated on the three different benchmark datasets, GraphomerDTI achieves a superior performance than the other five state-of-the-art methods under the inductive settings and is on a par with the best baseline under the transductive setting.
- The case study on real-world DTI predictions illustrates the exceptional predictive capability of GraphomerDTI for predicting DTIs of novel drug molecules.
- The source codes and datasets curated in this study are publicly accessible from GitHub at <https://github.com/mengmeng34/GraphomerDTI>.

Code and data availability

The source code and datasets curated in this study can be downloaded from GitHub at <https://github.com/mengmeng34/GraphomerDTI>.

CRediT authorship contribution statement

Mengmeng Gao: Writing – original draft, Project administration, Methodology. **Daokun Zhang:** Methodology. **Yi Chen:** Resources, Project administration, Methodology, Conceptualization. **Yiwen Zhang:** Writing – review & editing, Visualization. **Zhikang Wang:** Writing – review & editing, Validation. **Xiaoyu Wang:** Writing – review & editing, Visualization. **Shanshan Li:** Writing – review & editing, Resources, Formal analysis. **Yuming Guo:** Resources, Formal analysis. **Geoffrey I. Webb:** Project administration, Methodology. **Anh T.N. Nguyen:** Supervision, Data curation. **Lauren May:** Formal analysis, Data curation. **Jiangning Song:** Writing – original draft, Project administration, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the anonymous reviewers for their valuable suggestions. This work is supported in part by funds from NHMRC Ideas grant APP2013629 and the Major and Seed Inter-Disciplinary Research Projects awarded by Monash University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2024.108339>.

References

- [1] H. Dowden, J. Munro, Trends in clinical success rates and therapeutic focus, *Nat. Rev. Drug Discov.* 18 (2019) 495–496.
- [2] J. Vamathevan, D. Clark, P. Czodrowski, et al., Applications of machine learning in drug discovery and development, *Nat. Rev. Drug Discov.* 18 (2019) 463–477.
- [3] H.S. Chan, H. Shan, T. Dahoun, et al., Advancing drug discovery via artificial intelligence, *Trends Pharmacol. Sci.* 40 (2019) 592–604.
- [4] X. Pan, X. Lin, D. Cao, et al., Deep learning for drug repurposing: methods, databases, and applications, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* (2022) e1597.

- [5] X. Zeng, S. Zhu, X. Liu, et al., deepDR: a network-based deep learning approach to in silico drug repositioning, *Bioinformatics* 35 (2019) 5191–5198.
- [6] K. Huang, T. Fu, L.M. Glass, et al., DeepPurpose: a deep learning library for drug–target interaction prediction, *Bioinformatics* 36 (2020) 5545–5547.
- [7] M. Bagherian, E. Sabeti, K. Wang, et al., Machine learning approaches and databases for prediction of drug–target interaction: a survey paper, *Briefings Bioinf.* 22 (2021) 247–269.
- [8] W.X. Shen, X. Zeng, F. Zhu, et al., Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations, *Nat. Mach. Intell.* 3 (2021) 334–343.
- [9] V. Kanakaveti, A. Shammugam, C. Ramakrishnan, et al., Computational approaches for identifying potential inhibitors on targeting protein interactions in drug discovery, *Advances in protein chemistry and structural biology* 121 (2020) 25–47.
- [10] B. Liu, K. Pliakos, C. Vens, et al., Drug–target interaction prediction via an ensemble of weighted nearest neighbors with interaction recovery, *Appl. Intell.* 52 (2022) 3705–3727.
- [11] S. D’Souza, K. Prema, S. Balaji, Machine learning models for drug–target interactions: current knowledge and future directions, *Drug Discov. Today* 25 (2020) 748–756.
- [12] K. Sachdev, M.K. Gupta, A comprehensive review of feature based methods for drug target interaction prediction, *J. Biomed. Inf.* 93 (2019) 103159.
- [13] Y. Yang, D. Gao, X. Xie, et al., DeepIDC: a prediction framework of injectable drug combination based on heterogeneous information and deep learning, *Clin. Pharmacokinet.* 61 (2022) 1749–1759.
- [14] L. David, A. Thakkar, R. Mercado, et al., Molecular representations in AI-driven drug discovery: a review and practical guide, *J. Cheminf.* 12 (2020) 1–22.
- [15] S. Honda, S. Shi, H.R. Ueda, Smiles Transformer: Pre-trained Molecular Fingerprint for Low Data Drug Discovery, 2019 arXiv preprint arXiv:1911.04738.
- [16] N.R. Monteiro, B. Ribeiro, J.P. Arrais, Drug–target interaction prediction: end-to-end deep learning approach, *IEEE ACM Trans. Comput. Biol. Bioinf.* 18 (2020) 2364–2374.
- [17] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: ICLR, 2017.
- [18] H. Chen, Y. Lu, Y. Yang, et al., A drug combination prediction framework based on graph convolutional network and heterogeneous information, *IEEE ACM Trans. Comput. Biol. Bioinf.* 20 (2022) 1917–1925.
- [19] P. Wang, S. Zheng, Y. Jiang, et al., Structure-aware multimodal deep learning for drug–protein interaction prediction, *J. Chem. Inf. Model.* 62 (2022) 1308–1317.
- [20] T. Zhao, Y. Hu, L.R. Valsdottir, et al., Identifying drug–target interactions based on graph convolutional network and deep neural network, *Briefings Bioinf.* 22 (2021) 2141–2150.
- [21] M. Tsubaki, K. Tomii, J. Sese, Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, *Bioinformatics* 35 (2019) 309–318.
- [22] W. Chen, G. Chen, L. Zhao, et al., Predicting drug–target interactions with deep-embedding learning of graphs and sequences, *J. Phys. Chem.* 125 (2021) 5633–5642.
- [23] J. Wang, X. Li, H. Zhang, GNN-PT: Enhanced Prediction of Compound-Protein Interactions by Integrating Protein Transformer, 2020 arXiv preprint arXiv: 2009.00805.
- [24] K. Tian, M. Shao, Y. Wang, et al., Boosting compound-protein interaction prediction by deep learning, *Methods* 110 (2016) 64–72.
- [25] I. Lee, J. Keum, H. Nam, DeepConv-DTI: prediction of drug–target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.* 15 (2019) e1007129.
- [26] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug–target binding affinity prediction, *Bioinformatics* 34 (2018) i821–i829.
- [27] S. Zheng, Y. Li, S. Chen, et al., Predicting drug–protein interaction using quasi-visual question answering system, *Nat. Mach. Intell.* 2 (2020) 134–140.
- [28] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997) 1735–1780.
- [29] T. Nguyen, H. Le, T.P. Quinn, et al., GraphDTA: predicting drug–target binding affinity with graph neural networks, *Bioinformatics* 37 (2021) 1140–1147.
- [30] D. Bahdanau, K. Cho, Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, 2014 arXiv preprint arXiv:1409.0473.
- [31] K.Y. Gao, A. Fokoue, H. Luo, et al., Interpretable Drug Target Prediction Using Deep Neural Representation, *IJCAI*, 2018, pp. 3371–3377.
- [32] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [33] K. Huang, C. Xiao, L.M. Glass, et al., MolTrans: molecular Interaction Transformer for drug–target interaction prediction, *Bioinformatics* 37 (2021) 830–836.
- [34] Q. Zhao, H. Zhao, K. Zheng, et al., HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism, *Bioinformatics* 38 (2022) 655–662.
- [35] V.P. Dwivedi, X. Bresson, A Generalization of Transformer Networks to Graphs, 2020 arXiv preprint arXiv:2012.09699.
- [36] C. Ying, T. Cai, S. Luo, et al., Do transformers really perform badly for graph representation? *Adv. Neural Inf. Process. Syst.* 34 (2021).
- [37] M.Y. Wang, Deep graph library: towards efficient and scalable deep learning on graphs, in: ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019 Jan.
- [38] D.S. Wishart, C. Knox, A.C. Guo, et al., DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res.* 34 (2006) D668–D672.
- [39] M.I. Davis, J.P. Hunt, S. Herrgard, et al., Comprehensive analysis of kinase inhibitor selectivity, *Nat. Biotechnol.* 29 (2011) 1046–1051.
- [40] J. Tang, A. Szwajda, S. Shakyawar, et al., Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis, *J. Chem. Inf. Model.* 54 (2014) 735–743.
- [41] J. Devlin, M.-W. Chang, K. Lee, et al., Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018 arXiv preprint arXiv:1810.04805.
- [42] M. Yazdani-Jahromi, N. Yousefi, A. Tayebi, E. Kolanthai, C.J. Neal, S. Seal, O. O. Garibay, AttentionSiteDTI: an interpretable graph-based model for drug–target interaction prediction using NLP sentence-level relation classification, *Briefings Bioinf.* 23 (4) (2022 Jul 18) bbac272.
- [43] J. Du, S. Zhang, G. Wu, J.M. Moura, S. Kar, Topology Adaptive Graph Convolutional Networks, 2017 Oct 28 arXiv preprint arXiv:1710.10370.
- [44] R. Krivák, D. Hoksza, P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure, *J. Cheminf.* 10 (2018 Dec) 1–2.
- [45] B.N. Taylor, M. Cassagnol, Alpha Adrenergic Receptors, StatPearls Publishing, 2021. StatPearls [Internet].
- [46] P.J. Barnes, Beta-adrenergic receptors and their regulation, *Am. J. Respir. Crit. Care Med.* 152 (1995) 838–860.
- [47] H.L. Li, Y.H. Pang, B. Liu, BioSeq-BLM: a platform for analyzing DNA, RNA and protein sequences based on biological language models, *Nucleic Acids Res.* 49 (2021) e129.
- [48] H.L. Li, B. Liu, BioSeq-Diabolo: biological sequence similarity analysis using Diabolo, *PLoS Comput. Biol.* 19 (2023) e1011214.
- [49] Z. Chen, X. Liu, P. Zhao, et al., iFeatureOmega: an integrative platform for engineering, visualization and analysis of features from molecular sequences, structural and ligand data sets, *Nucleic Acids Res.* 50 (2022) W434–W447.