BMC Genomics

RESEARCH

Open Access

GS-DTA: integrating graph and sequence models for predicting drug-target binding affinity

Junwei Luo[1], Ziguang Zhu[1], Zhenhan Xu[1], Chuanle Xiao[3], Jingjing Wei[2] and Jiquan Shen[1,2*]

Abstract

Background Drug-target binding affinity (DTA) prediction is vital in drug discovery and repositioning, more and more researchers are beginning to focus on this. Many effective methods have been proposed. However, some current methods have certain shortcomings in focusing on important nodes in drug molecular graphs and dealing with complex structural molecules. In particular, when considering important nodes and complex substructures in molecules, they may not be able to fully explore the potential relationships between different parts. In addition, when dealing with protein structures, some methods ignore the connections between amino acid fragments that are far apart in sequence but may work synergistically in function.

Results In this paper, we propose a new method, called GS-DTA, for predicting DTA based on graph and sequence models. GS-DTA takes simplified molecular input line input system (SMILES) of the drug and the protein amino acid sequence as input. First, each drug is modeled as a graph, in which a vertex is an atom and an edge represents interaction between atoms. Then GATv2-GCN and the three-layer GCN networks are used to extract the features of the drug. GATv2-GCN enhances the model's ability to focus on important nodes by assigning dynamic attention scores, which improves the learning of the graph structure's intricate patterns. Besides, The three-layer GCN can captures hierarchical features of the drug through deeper propagation and feature transformation. Meanwhile, for each protein, a framework combining CNN, Bi-LSTM, and Transformer is used to extract the contextual and structural information of the protein amino acid sequences, and this combination can help to understand a comprehensive and detailed features of the protein. Finally, the obtained drug and protein feature vectors are combined to predict DTA through the fully connected layer. The source code can be downloaded from https://github.com/zhuziguang/GS-DTA.

Conclusions The results show that GS-DTA achieves good performance in terms of MSE, CI, and r^2_m on the Davis and KIBA datasets, improving the accuracy of DTA prediction.

Keywords Drug-target binding affinity, Graph neural networks, Transformer

*Correspondence:
Jiquan Shen
sjq@hpu.edu.cn
[1] School of Software, Henan Polytechnic University, Jiaozuo 454000, China
[2] College of Chemical and Environmental Engineering, Anyang Institute of Technology, Anyang 455000, China
[3] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou 510060, China

Background

Drug-target interaction (DTI) prediction is vital in drug discovery and repositioning [1–3] because only drugs and targets with similar molecular structures are compatible. Unlike DTI prediction methods based on binary classification tasks [4, 5], Drug-target binding affinity (DTA) refers to the binding strength between drug and their target proteins, a regression task [6, 7]. DTA

prediction provides important information and data support for DTI prediction [8, 9], helping to determine possible interactions between drugs and potential targets, thereby guiding the prediction and research of drug-target interactions, so accurate prediction and measurement of DTA is crucial in the drug discovery and development process. For this reason, more and more researchers are focusing on the research of DTA prediction.

Recently, multiple computational approaches for predicting DTA have been proposed. SimBoost [10] is a technique that uses gradient-enhanced supervised learning methods to predict DTA. In the KronRLS [11] model, the Kronecker products of drug and target pairs are generated to compute the kernel K for these drug and protein pairs, which is then used in a regularized least-squares regression model (RLS) to predict binding affinity.

With the development of deep learning [12], some deep learning-based methods for predicting DTA have been proposed. Some prediction methods are summarized in Table 1. DeepDTA [7] introduces a new method to predict drug-target binding affinity using deep learning, which only requires sequence information of proteins and drugs. Through convolutional neural networks (CNN), DeepDTA extracts features from the raw sequence data of proteins and drugs and uses fully connected layers to predict DTA. WideDTA [13] is designed based on DeepDTA and used the same architecture as DeepDTA, But the difference is that this method combines four different text information: Ligand SMILES, Ligand Max Common Substructure, Protein Sequence, and Protein Motifs and Domains. The results show that the added textual information helps the model achieve better results. DeepCDA [14] combines CNN and LSTM [15] to achieve better representation of drug and protein features. It also proposes a bidirectional attention mechanism to encode the binding strength between each protein structure and drug substructure. Although sequence-based methods have achieved good results in DTA prediction, there is a risk of losing molecular structure information. CSatDTA [16] uses a convolution-based self-attention mechanism to analyze drug and target sequences. With the continuous development of graph neural networks (GNN) [17, 18], many researchers adopt GNN models to predict DTA in order to retain as much molecular structure information as possible. Utilizing multiple graph convolutional layers (MGNN) and a multiscale convolutional neural network (MCNN), MGraphDTA [19] captures multiscale features of drug and target. DeepGLSTM [20] uses three GCN blocks to learn the topological information of drug. DGraphDTA [21] first converts protein sequences into contact graphs and then processes them into protein graphs. After this, the protein graph and the drug molecule graph are sent to the GNN to predict DTA. Compared to traditional methods based on one-dimensional protein sequences, 2D graph structures can provide more structural information about proteins and demonstrate significant advantages in DTA prediction. GraphDTA [6] uses various GNN variant models to extract drug feature representations, improving the prediction effectiveness of DTA. OdinDTA [22] integrates a mutual attention mechanism and a pre-trained protein model with a conventional three-channel prediction framework. DGDTA [23] proposes a method that combines a dynamic graph attention network with a bidirectional long short-term memory network to predict DTA. GDilatedDTA [24] proposes a DTA prediction model based on graph dilation convolution strategy, which uses a weight matrix to enhance the interpretability of the binding process

Table 1 Predcition methods

Method	Published year	Model	Summary
SimBoost	2016	Gradient boosting regression trees	The binding affinity problem of compounds and proteins is considered as a continuous value prediction problem rather than a binary classification problem
DeepDTA	2018	CNN	Using Convolutional Neural Networks to Process 1D Representations of Protein and Drug Sequences
WideDTA	2019	CNN	This method integrates four distinct textual information sources pertaining to proteins and ligands
DeepCDA	2020	CNN + LSTM	Convolutional networks and LSTM layers are combined into a unified framework to effectively encode local and global temporal patterns
GraphDTA	2021	GCN + CNN	Using a variety of graph neural networks
DeepGLSTM	2022	GCN + BiLSTM	Three GCN blocks are used to learn the topological information of drug molecules, and BiLSTM is used to learn the representation of protein sequences
GdilatedDTA	2024	McGEN + MLRCN + BiLSTM	Predicting drug affinity using a graph-based dilated convolution strategy
AttentionMGT-DTA	2024	GraphTransformer + Cross-Attention	Two attention mechanisms are adopted to integrate and interact information between different protein patterns and drug target pairs

between drugs and targets. AttentionMGT-DTA [25] constructs a protein pocket graph based on the AlphaFold database to represent the target protein, and introduces a joint-attention mechanism to generate affinity results for the matrix interacting with atom-residue.

However, there are still some problems that affect accurate DTA prediction. (1) Some current methods may focus more on local information and relatively less on global information in drug feature extraction. This strategy, although effective, may miss some important features in the global context when dealing with the complex structure of drugs. (2) Some current methods may fail to fully capture the complex structural and functional information of proteins.

To solve the above problems, this paper proposes a method called GS-DTA. In GS-DTA, we designed and implemented a hybrid graph neural network model combining GATv2, GCN, and three-layer GCN to comprehensively extract the features of drug. By using multi-layer GCN, nodes can obtain more extensive neighborhood information, thereby extracting different levels of information in the molecular structure, which helps the model identify complex molecular features. The GATv2 network introduces an attention mechanism, which enables the model to dynamically focus on important nodes and edges in the drug molecule graph, making feature extraction more selective and targeted. GCN focuses on acquiring global information in molecular structure, while GATv2 can supplement the limitations of GCN. The combination of the two enhances the robustness of the model, making it more adaptable and generalizable when facing molecules with different structures. In addition, Proteins are composed of amino acids linked together by peptide bonds, forming a polypeptide chain whose sequence determines the protein's function. Within the protein sequence, there are crucial features such as ligand-binding sites and the chemical properties of amino acids, which tend to be concentrated in localized regions. Convolutional neural networks (CNNs) are particularly effective at identifying these local features, allowing the model to quickly capture important information in the sequence. However, the functionality of some amino acids depends not only on their immediate surroundings but also on the residues before and after them in the sequence. Since CNNs do not inherently account for the sequence order or dependencies between adjacent amino acids, Bidirectional Long Short-Term Memory (BiLSTM) comes in to model these contextual dependencies, offering a deeper understanding of how adjacent amino acids influence each other's function. Moreover, in protein structures, distant amino acid residues can sometimes be spatially close, interacting synergistically and impacting the drug-binding affinity. Because CNN and BiLSTM are limited in capturing these long-range dependencies, Transformer uses a attention mechanism that allows it to incorporate global information, enabling the model to understand how distant positions in the sequence may contribute to the protein's functional regions. By combining these three networks, the model is able to extract features from protein sequences at multiple levels, including local, contextual, and global, resulting in a more comprehensive representation. GS-DTA combines multiple network models to complement the weaknesses of a single network, reduce the deviation caused by a single model, and improve the robustness and generalization ability of the overall model.

Datasets

In our study, we utilized two widely recognized benchmark datasets, Davis [26] and KIBA [27], to both train and evaluate the performance of GS-DTA. These datasets, which are publicly accessible, serve as established standards for DTA prediction. The Davis dataset contains 442 proteins and 72 drugs, measured by K_d values, and a total of 30,056 interactions between drugs and targets. The KIBA dataset comprises 229 proteins, 2116 drugs, and 118,254 drug-target interactions [6].

In the benchmark setting, each dataset is divided into several parts, where one part is reserved for testing, and other parts are used for training. This paper conducts tests on these two data sets to comprehensively evaluate the predictive power of GS-DTA. The details of these two datasets are shown in Table 2.

Methods

GS-DTA uses the SMILES of drug and protein sequences to predict DTA. The architecture of GS-DTA is shown in Fig. 1. The overall framework consists of three main steps: (1) Extracting drug features; The SMILES of drug are converted into a drug molecular graph, and a hybrid graph neural network is used to extract drug features. (2) Extracting protein features; By using a convolutional neural network (CNN), a Bidirectional long short-term memory network (Bi-LSTM), and a Transformer [28] module, we consider the local and global information between amino acid sequences to extract important features. (3) Performing DTA prediction; the drug and protein features are connected and input into the fully connected layer for DTA prediction.

Table 2 Datasets

Dataset	Proteins	Compounds	Interactions	Train	Test
Davis	442	72	30,056	25,046	5010
KIBA	229	2116	118,254	98,545	19,709

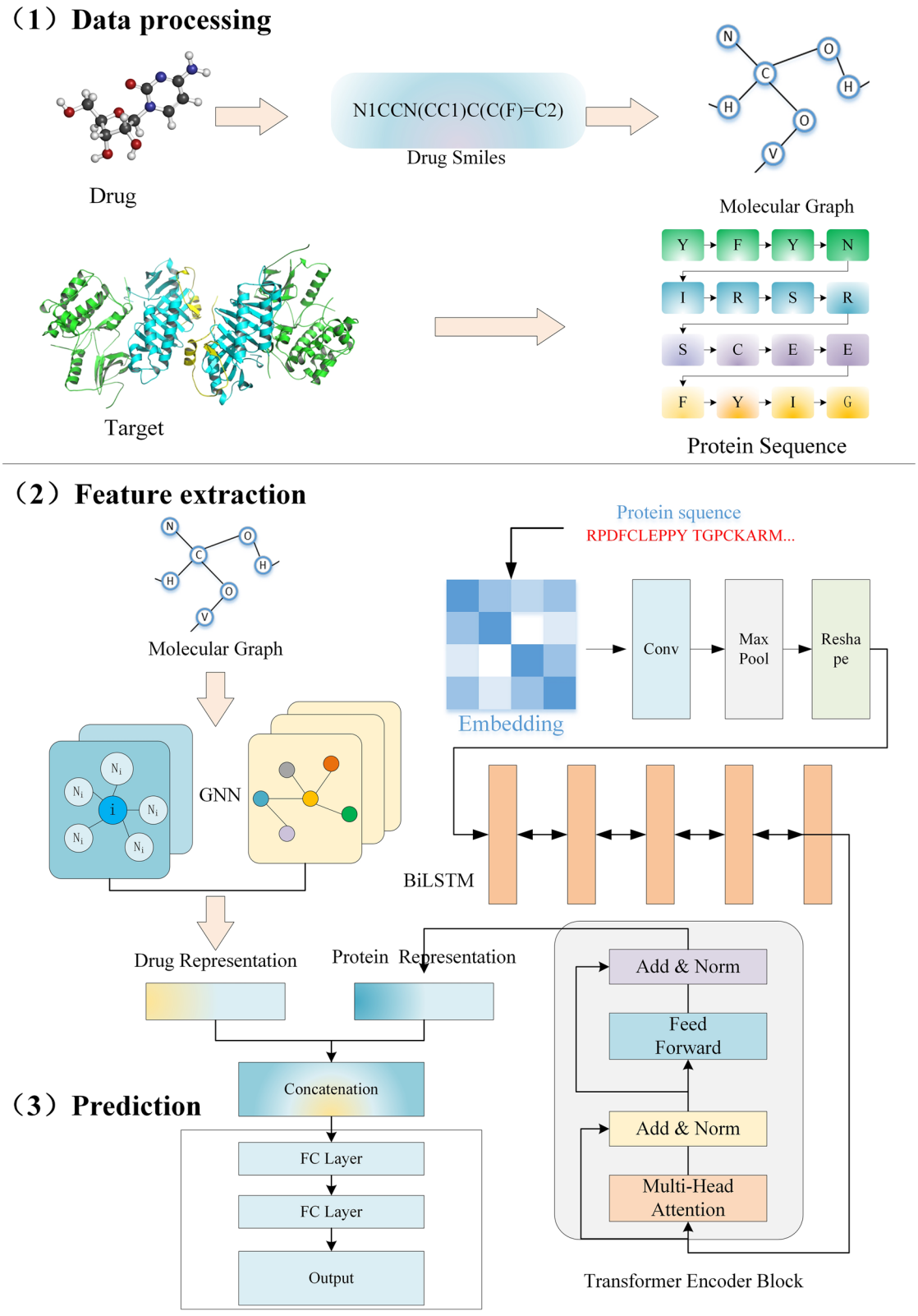

Fig. 1 The architecture diagram of GS-DTA

Extracting drug features

SMILES is a string representation used to depict molecular structures, it is a linear encoding method based on atoms and bonds that can concisely represent the chemical molecular structure, including atom types, connections between atoms, and the topology of the molecule. This representation applies to most biological macromolecules such as proteins, peptides, and nucleic acids.

SMILES can effectively represent important information about drug. Through the SMILES representation, the molecule's atoms, bonds, and topological structure can be described, thereby facilitating the storage, transmission, and processing of chemical information. However, using SMILES may cause the following problems: (1) SMILES is a linear string representation. Although it can represent the molecular structure, it easily ignores the topological information of the molecule. (2) For a long SMILES sequences, specific structural features may be difficult to interpret intuitively. To solve the above problems, this paper converts the SMILES of drug into graphs. Moreover, GNN can propagate information between nodes, capture long-range dependencies and local information within molecules, and enhance the richness and robustness of feature representation.

In this paper, GS-DTA uses the SMILES of drug as input, and then uses RDKit [28, 29] to build a graph for each drug. For each drug, we construct a graph *G(V, E)*, *V* is the set of vertices in the graph, each vertex refers to an atom. *E* is the set of edges in the graph, and each edge represents the chemical bond between atoms. The features of a vertex include a variety of properties, such as atomic symbols, atomic degrees, the number of hydrogen atoms around the atoms, atomic implicit valence, and atomic aromaticity. Table 3 lists the drug characteristics in detail.

First, a vertex is represented as $v=\{f_1, f_2, f_3, f_4, f_5\}$. Among them, f_1-f_5 represent the one-hot encoding vector of atom type, atom degree, atom implicit valence and atom aromatic respectively. Then the characteristics of each drug $H=\{v_1, v_2, v_3..., v_n\}$ are obtained.

$$h_i^{(0)} = one - hot(Featuer) \tag{1}$$

In the above formula, *Feature* represents the atom type, atomic degree, atomic implicit valence, and atomic aromaticity.

After initializing the node features, update the node features through the attention mechanism of GATv2:

$$h_i^{(1)} = \sigma\left(\sum_{j\in N(i)} \alpha_{ij}^{(1)} W^{(1)} h_j^{(0)}\right) \tag{2}$$

The attention coefficient $\alpha_{ij}^{(1)}$ is calculated in the following way:

$$e_{ij}^{(1)} = LeakyReLU\left(a^{(1)T}\left[W^{(1)}h_i^{(0)} || W^{(1)}h_j^{(0)}\right]\right) \tag{3}$$

$$\alpha_{ij}^{(1)} = \frac{\exp\left(e_{ij}^{(1)}\right)}{\sum_{k\in N(i)} \exp\left(e_{ik}^{(1)}\right)} \tag{4}$$

$W^{(1)}$ is the trainable weight matrix, $a^{(1)}$ is the parameter vector of the attention mechanism, and || represents the connection operation of the vector. According to the calculated attention coefficient, the above features are weighted and summed:

$$h_i^{(1)\prime} = \sigma\left(\sum_{j\in N_i} \alpha_{ij} W h_j\right) \tag{5}$$

$h_i^{(1)\prime}$ is the new feature output by GATv2 for each node *i*, and σ is the activation function. To further improve the expression ability of the model, the attention mechanism is expanded to a multi-head attention mechanism.

$$h_i^{(1)\prime\prime} = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i} \alpha_{ij}^{(k)} W^{(k)} h_j^{(0)}\right) \tag{6}$$

Table 3 The atom features for drug graph representation

Feature	Description	Dimension
Atom type	'C', 'N', 'O', 'S', 'F', 'Si', 'P', 'CL', 'Br', 'Mg', 'Na', 'Ca', 'Fe', 'As', 'Al', 'I', 'B', 'V', 'K', 'Tl', 'Yb', 'Sb', 'Sn', 'Ag', 'Pd', 'Co', 'Se', 'Ti', 'Zn', 'H', 'Li', 'Ge', 'Cu', 'Au', 'Ni', 'Cd', 'In', 'Mn', 'Zr', 'Cr', 'Pt', 'Hg', 'Pb', 'Unknown'	44
Atom degree	0,1,2,3,4,5,6,7,8,9,10	11
Atom total num H	0,1,2,3,4,5,6,7,8,9,10	11
Atom implicit valence	0,1,2,3,4,5,6,7,8,9,10	11
Atom is aromatic	0 or 1	1
total		78

where K is the number of attention heads, $\alpha_{ij}^{(k)}$ is the attention coefficient of the k-th head, and $W^{(k)}$ is the weight matrix of the k-th head.

Extracting protein features

Protein sequences are usually represented by a string of letters, each letter corresponding to a specific amino acid. Some previous studies (DeepDTA, DeepCDA, etc.) usually use protein sequences as input to the model. These studies used a 1D convolutional layer to extract valuable features from protein sequences.

In this paper, we convert each amino acid into a numerical value, representing the protein as a sequence of integers. We apply an embedding layer to this sequence, where each amino acid is represented by a 128-dimensional vector. For training, the sequences are standardized to a fixed length of 1000; shorter sequences are padded with zeros to reach this length. First, CNN is used to initially extract protein features, and pooling operations are used to improve the generalization ability of the model. Using P to represent the protein features extracted by the CNN network and a bidirectional long short-term memory network (Bi-LSTM) is used to further extract features:

$$\overrightarrow{h_t} = LSTM_{forward}(P_t, h_{t-1}) \tag{7}$$

$$\overleftarrow{h_t} = LSTM_{backward}(P_t, h_{t+1}) \tag{8}$$

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{9}$$

h_{t-1} and h_{t+1} represent the hidden state at time steps *t-1* and $t+1$ respectively. $H=[h_1, h_2, h_3, ..., h_n]$, H represents the vector composition of Bi-LSTM module output.

Based on the features extracted by Bi-LSTM, we adopt Transformer to dynamically assign different weights to each feature through its attention mechanism, allowing the model to automatically focus on the most informative parts of the sequence, thereby further extracting useful features:

$$Z = LayerNorm\big(H + MultiHeadSelfAttention(H) + FeedForward(H)\big) \tag{10}$$

$$MultiHeadSelfAttention(H) = Concat(head_1, head_2, ..., head_h) \cdot W^O \tag{11}$$

For each attention head i, the calculation process is as follows:

The Calculation of *Query*, *Key*, and *Value*:

$$Query^{(i)} = H \cdot W^Q, Key^{(i)} = H \cdot W^K, Value^{(i)} = H \cdot W^V \tag{12}$$

The Calculation of attention weight:

$$Attention^{(i)} = Softmax\left(\frac{Query^{(i)} \cdot \left(Key^{(i)}\right)^T}{\sqrt{d_k}}\right) \cdot Value^{(i)} \tag{13}$$

Merger of multiple heads' attention:

$$head_i = Attention^{(i)} \cdot W_i^O \tag{14}$$

Output features:

$$Z^{(l)} = LayerNorm\Big(Concat(head_1, head_2, ..., head_h) \cdot W^O + H\Big) \tag{15}$$

Among them, H is the output feature sequence of Bi-LSTM, $Z^{(l)}$ is the output feature sequence of Transformer, W^Q, W^K, W^V, W^O, is the weight matrix, d_k is the dimension of Key, and h is the number of attention heads.

Performing DTA prediction

The above-learned drug vector and protein vector are connected and sent to the fully connected layer to obtain the final DTA prediction value y:

$$y = W_{output}[D, P] + b_{output} \tag{16}$$

where W_{output} represents the weight matrix of the fully connected layer, b_{output} represent the bias of the fully connected layer.

Experiment and results

Experimental settings

The data sets used in the experiment include Davis dataset and KIBA dataset. GS-DTA takes drug SMILES and amino acid sequence as input. This article uses Python 3.7.12, PyTorch1.13 and PyG2.3.1 to implement dynamic GAT, LSTM and Transformer. In this paper, dropout is set to 0.2. Then, the proposed method was trained on the above dataset for 1000 epochs and used the Adam optimizer with a learning rate of 0.0005. The device used for the experiment was Intel(R) Xeon(R) Gold 6330 CPU @2.00 GHz and NVIDIA GeForce RTX 4090GPU.The best settings of hyperparameter optimization are presented in Table 4.

Evaluation metrics

We assess regression tasks using the three evaluation metrics: mean squared error (MSE), concordance index (CI), and r_m^2 index.

The mean-square error (MSE) is a measure that reflects the degree of difference between the model's predicted value and the true value. The smaller the value, the better. The calculation method is detailed in Eq. 17.

Table 4 The setting of hyperparameters

Hyperparameters	Setting
Number of transformer layers	2
Number of attention heads	4
Dropout rate of transformer	0.2
Learning rate	0.0005
Batch size	512
Epoch	1000

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y_i} - y_i\right)^2 \tag{17}$$

The concordance index (CI) evaluates the discriminative performance of different models. The CI value closer to 1 indicates a better fit of the model. Equation 18 shows the calculation method.

$$CI = \frac{1}{Z}\sum_{\sigma_i > \sigma_j}\varphi\left(b_i - b_j\right) \tag{18}$$

The r_m^2 index is utilized in DeepGLSTM, serves as an evaluation metric for regression tasks. Equation 19 outlines the calculation procedure.

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{19}$$

Where the squares of the correlation coefficients with and without the intercept are denoted as r^2 and r_0^2, respectively.

Results

Comparison of the predicted and real values

In this section, we compare the predicted and real values of the Davis and KIBA datasets. As shown in Fig. 2, the results confirm that the data predicted by GS-DTA is very close to the real values of the Davis and KIBA datasets.

Performance comparison

In this section, Table 5 shows the experimental results of GS-DTA and other methods. The evaluation metric values of other methods are extracted from their published papers. To be consistent with the ablation experiments in 3.4.3, we use the same dataset and evaluation metrics. As shown in Table 5 on the Davis dataset and the Kiba dataset, GS-DTA is better than most other methods in CI, MSE, and r_m^2 indicators. GS-DTA ranks first in CI and r^2_m on the Davis dataset, and ranks first in CI and MSE on the KIBA dataset. Although it fails to rank first in MSE on the Davis dataset and r^2_m on the KIBA dataset, it ranks second. The experimental results prove the effectiveness of GS-DTA.

Ablation study

In our ablation experiments, we demonstrate the efficacy of GS-DTA. To ensure fairness, we selected identical training and test datasets, as well as evaluation metrics. GS-DTA uses GATV2 combined with GCN and three-layer GCN for drug processing, while combining CNN, Bi-LSTM and Transformer to extract features from protein amino acid sequences to improve model accuracy. Bi-LSTM and Transformer represent two widely used neural network architectures for sequence data analysis. Bi-LSTM effectively captures long-term dependencies within sequences, mitigating issues such as vanishing or exploding gradients encountered in traditional RNNs. Transformer, employs a attention mechanism, facilitating direct modeling of dependencies between any two positions in a sequence, regardless of sequence length.

This parallel processing capability enables efficient handling of long sequences, with the added benefit of

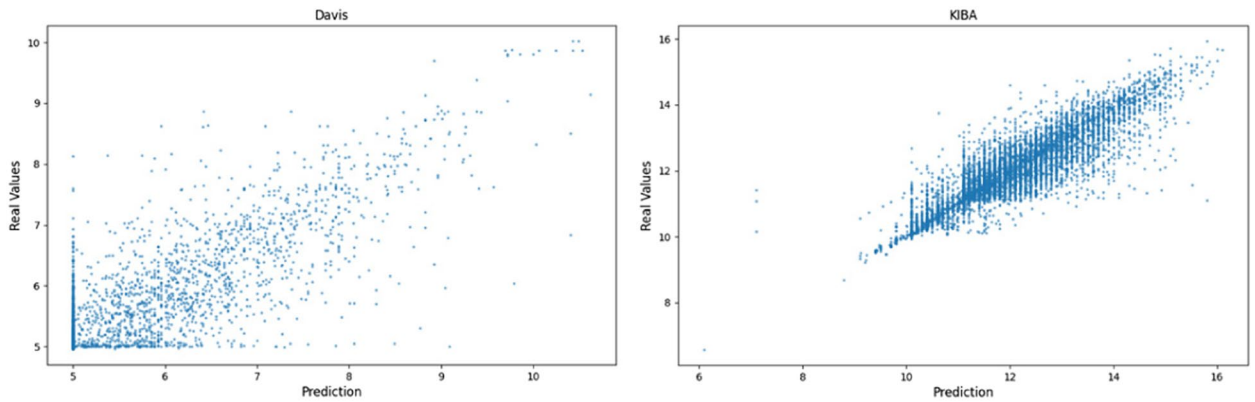

Fig. 2 Comparison of predicted values with true values

Table 5 Comparison result based on the Davis and KIBA dataset

Dataset	Methods	CI	MSE	r^2_m
Davis	KronRLS	0.871	0.379	0.407
	SimBoost	0.872	0.282	0.644
	DeepDTA	0.878	0.261	0.631
	DeepCDA	0.891	0.248	0.652
	GraphDTA(GAT)	0.892	0.232	0.689
	GraphDTA(GAT-GCN)	0.881	0.245	0.667
	DeepGLSTM	0.893	0.236	0.679
	GdilatedDTA	0.885	0.237	0.686
	AttentionMGT-DTA	0.891	**0.193**	0.699
	DGDTA-CL	0.889	0.237	0.672
	DGDTA-AL	0.899	0.225	0.707
	GS-DTA	**0.903**	0.213	**0.709**
KIBA	KronRLS	0.782	0.411	0.342
	SimBoost	0.836	0.222	0.629
	DeepDTA	0.863	0.194	0.673
	DeepCDA	0.889	0.176	0.682
	GraphDTA(GAT)	0.866	0.179	0.738
	GraphDTA(GAT-GCN)	0.891	0.139	0.789
	DeepGLSTM	0.890	0.143	0.789
	GdilatedDTA	0.876	0.156	0.775
	AttentionMGT-DTA	0.893	0.140	0.786
	DGDTA-CL	0.902	0.125	**0.809**
	DGDTA-AL	0.881	0.162	0.762
	GS-DTA	**0.905**	**0.124**	0.806

capturing various levels of feature representation through stacked Transformer layers. By combining LSTM and Transformer, we synergistically harness their respective strengths: LSTM excels at capturing long-term dependencies, while Transformer's parallel processing and self-attention mechanisms enable comprehensive modeling of positional dependencies. This integration ensures that both long-term dependencies and local intricacies are effectively addressed, thereby enhancing the performance and efficacy of sequence data processing. To verify the effectiveness of the model, we reduce the network one by one and conduct detailed ablation experiments, which are listed in Table 6

The ablation experiment results show that GS-DTA achieved good results on both the Davis dataset and the KIBA dataset. On the Davis dataset, the CI, MSE, and r^2_m of GS-DTA reached 0.903, 0.213, and 0.709; on the KIBA dataset, the CI, MSE, and r^2_m of GS-DTA reached 0.905, 0.124, and 0.806. Although the MSE on the Davis dataset is 0.3% lower than that of model 1, r^2_m is 0.1% lower than model 3 on the KIBA dataset. GS-DTA still has the best results from an overall perspective. The ablation experiment results prove the effectiveness of the model proposed in this paper.

Discussion

This study introduces GS-DTA, which integrates a hybrid graph neural network model to extract drug features and a network model that combines CNN, Bi-LSTM, and Transformer to extract protein features. Compared with other models, our model better considers both local and global information when extracting drug and protein features, and obtains richer protein and drug features. Besides, When extracting features, the model automatically identifies molecular features or protein domains that are highly correlated with binding affinity. These regions "noticed" by the model may play a key role in targeted drug design. In addition, We conducted ablation experiments on Davis and KIBA datasets and compared the proposed method with some other DTA models. The results demonstrate the effectiveness of GS-DTA.

In the context of drug development, GS-DTA's approach has the potential to significantly accelerate the drug discovery process by reducing the need for

Table 6 Ablation study on the Davis and KIBA dataset

Dataset	Methods	GATv2-GCN	3×GCN	CNN	BiLSTM	Transformer	CI	MSE	r^2_m
Davis	Model-1	√	-	√	√	√	0.900	**0.210**	0.708
	Model-2	√	√	√	√	-	0.892	0.224	0.674
	Model-3	√	√	√	-	√	0.901	0.215	0.670
	Model-4	√	√	-	√	√	0.900	0.212	0.708
	GS-DTA	√	√	√	√	√	**0.903**	0.213	**0.709**
KIBA	Model-1	√	-	√	√	√	0.899	0.125	0.793
	Model-2	√	√	√	√	-	0.901	0.126	0.796
	Model-3	√	√	√	-	√	0.904	0.124	**0.807**
	Model-4	√	√	-	√	√	0.896	0.137	0.797
	GS-DTA	√	√	√	√	√	**0.905**	**0.124**	0.806

costly and time-consuming experimental screenings. By enabling efficient predictions of drug-target binding affinity, this model aids in early-stage drug development by quickly identifying promising candidate drugs with favorable interaction profiles. In the future, we will also focus on other areas that can help drug development, such as drug synergy prediction [9], drug-drug interactions, etc.

Conclusions

Although GS-DTA has achieved good performance in predicting DTA by considering both global and local information of drug and protein features, it also has some limitations. First, the fusion of hybrid graph neural networks increases the model complexity and training overhead; Second, the structural information of proteins is important for predicting DTA, GS-DTA only consider one-dimensional protein sequences for extracting features. In the future, we will focus on other dimensions of protein information and further enhance the features by combining data from different dimensions.

Abbreviations

GS-DTA	Graph and Sequence DTA
DTA	Drug–target binding affinity
GAT	Graph attention network
GCN	Graph convolutional network
Bi-LSTM	Bidirectional long short-term memory
SMILES	Simplified molecular input line entry system
CNN	Convolutional neural network
GATv2	Dynamic graph attention network

Authors' contributions

All authors contributed to the design of the method. JWL and ZGZ participated in the design of the study and the analysis of the experimental results. ZHX, CLX and ZGZ provided important support in data collection and preprocessing, and assisted in building the preliminary framework of the model. JJW and JQS provided technical support during model training, hyperparameter adjustment, and experiments. All authors have read and approved the final manuscript for publication.

Funding

This research was supported by the Henan Provincial Department of Science and Technology Research Project (Grant No. 242102210097), and Innovative Research Team of Henan Polytechnic University (Grant No. T2021-3).

Data availability

The Davis and KIBA data can be downloaded from https://github.com/thinng/GraphDTA/tree/master. The source code is available from GitHub at https://github.com/zhuziguang/GS-DTA.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 August 2024 Accepted: 10 January 2025
Published online: 04 February 2025

References

1. Hughes JP, Rees S, Kalindjian SB, et al. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239–49.
2. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016;12(1):2694–718.
3. Strittmatter SM. Overcoming drug development bottlenecks with repurposing: old drugs learn new tricks. Nat Med. 2014;20(6):590–1.
4. Chu Y, Shan X, Chen T, et al. DTI-MLCD: predicting drug-target interactions using multi-label learning with community detection method. Briefings in bioinformatics. 2021;22(3):bbaa205.
5. Weng Y, Lin C, Zeng X, et al. Drug target interaction prediction using multi-task learning and co-attention[C]//2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2019:528–533.
6. Nguyen T, Le H, Quinn TP, et al. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140–7.
7. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics. 2018;34(17):i821–9.
8. Dehghan A , Abbasi K , Razzaghi P ,et al.CCL-DTI: contributing the contrastive loss in drug–target interaction prediction. BMC Bioinformatics. 2024;25(1). https://doi.org/10.1186/s12859-024-05671-3.
9. Rafiei F, Zeraati H, Abbasi K, et al. CFSSynergy: combining feature-based and similarity-based methods for drug synergy prediction. J Chem Inf Model. 2024;64(7):2577–85.
10. He T, Heidemeyer M, Ban F, et al. SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. Journal of cheminformatics. 2017;9:1–14.
11. Cichonska A, Pahikkala T, Szedmak S, et al. Learning with multiple pairwise kernels for drug bioactivity prediction. Bioinformatics. 2018;34(13):i509–18.
12. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw. 2015;61:85–117.
13. Öztürk H, Ozkirimli E, Özgür A. WideDTA: prediction of drug-target binding affinity. arXiv preprint arXiv:1902.04166, 2019.
14. Abbasi K, Razzaghi P, Poso A, et al. DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. Bioinformatics. 2020;36(17):4633–42.
15. Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. 2019;31(7):1235–70.
16. Ghimire A, Tayara H, Xuan Z, et al. CSatDTA: prediction of drug–target binding affinity using convolution model with self-attention. Int J Mol Sci. 2022;23(15):8453.
17. Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications. AI open. 2020;1:57–81.
18. Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826, 2018.
19. Yang Z, Zhong W, Zhao L, et al. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. Chem Sci. 2022;13(3):816–33.
20. Mukherjee S, Ghosh M, Basuchowdhuri P. DeepGLSTM: deep graph convolutional network and LSTM based approach for predicting drug-target binding affinity[C]//Proceedings of the 2022 SIAM international conference on data mining (SDM). Soc Industr Appl Mathematics, 2022: 729–737.
21. Jiang M, Li Z, Zhang S, et al. Drug–target affinity prediction using graph neural network and contact maps. RSC Adv. 2020;10(35):20701–12.
22. Xu S, Wang R. OdinDTA: Combining Mutual Attention and Pre-training for Drug-target Affinity Prediction[C]//2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE. 2023: 680–687.
23. Zhai H, Hou H, Luo J, et al. DGDTA: dynamic graph attention network for predicting drug–target binding affinity. BMC Bioinformatics. 2023;24(1):367.

24. Zhang L, Zeng W, Chen J, et al. GDilatedDTA: Graph dilation convolution strategy for drug target binding affinity prediction. Biomed Signal Process Control. 2024;92: 106110.
25. Wu H, Liu J, Jiang T, et al. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. Neural Netw. 2024;169:623–36.
26. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011;29(11):1046–51.
27. Tang J, Szwajda A, Shakyawar S, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J Chem Inf Model. 2014;54(3):735–43.
28. Bento AP, Hersey A, Félix E, et al. An open source chemical structure curation pipeline using RDKit. Journal of Cheminformatics. 2020;12:1–16.
29. Landrum G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. Greg Landrum. 2013;8(31.10):5281.

Publisher's Note