ISSN: 2688-8645

Drug-Target Affinity Prediction Based on Graph Representation and Attention Fusion Mechanism

Linhua Jiang[1, 2], Wenbiao Ye[1, 2], Wei Long[1], Wenbo Guo[1] and Lingxi Hu[1, *]

[1]School of Information Engineering, Huzhou University, Huzhou 313000, China

[2]Artificial Intelligence Laboratory of Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China

*03078@zjhu.edu.cn

Abstract

Predicting drug-target affinity is crucial in the field of drug discovery. To further improve the accuracy of predictions, this paper proposes a drug-target affinity prediction model, GRAM-DTA, based on graph representation and attention fusion mechanisms. The model represents the input features of drugs and targets as graph data and utilizes deep graph isomorphism networks and graph neural network modules combining graph convolutional networks and graph attention networks to process the feature information of drugs and targets, respectively. In the feature fusion stage, an attention mechanism is introduced to simulate the interactions between drug molecules and amino acids, dynamically adjust the importance of features, and capture the interaction patterns between drugs and targets. Experiments were conducted on the Davis and KIBA benchmark datasets, and the model was compared with current state-of-the-art models.The experimental results show that our model achieved a 3.1% and 3.4% improvement in the r_m^2 value compared to the best-performing baseline model, significantly outperforming other traditional methods and baseline models.

Keywords

Drug-target Affinity; Feature Extraction; Attention Mechanism.

1. Introduction

In the field of drug discovery, predicting drug-target affinity (DTA) plays a crucial role, with the core objective of accurately assessing the strength of interactions between drugs (ligands) and targets (usually proteins)[1]. Traditional computational methods, such as molecular docking, although capable of precisely simulating the binding mechanisms of drugs and targets, often face limitations in computational efficiency and accuracy when dealing with structurally complex and variable biomolecules0. With the rapid advancement of information technology, Computer Aided Drug Design (CADD)[3] has been able to develop swiftly. By leveraging advanced computational algorithms, CADD can efficiently identify and simulate the interaction patterns between drugs and proteins[4], providing a chemical basis for the structural optimization of small-molecule compounds, thereby reducing the trial-and-error costs in the drug development process and accelerating the pace of new drug discovery[5].

In the initial stage of drug development, identifying chemical molecules that bind to targets with high affinity is a crucial first step, which lays the foundation for the subsequent optimization of these molecules into lead compounds[6]. In the early stages of research, experimental screening of lead compounds, such as high-throughput screening techniques[7], often requires a significant amount of time and cost. With the development of CADD, many important techniques have emerged, among which molecular docking[8] stands out. Nowadays, there are various molecular docking tools available, such as DOCK and FlexX[9]. Although current

ISSN: 2688-8645

molecular docking techniques can clearly present the three-dimensional conformations of drug-target binding, they still require substantial computational resources and time since each docking process has to start from scratch[10]. Over the past decade, researchers have been exploring and applying machine learning-based predictive models. For example, Pahikkala et al.[11] used the Smith-Waterman algorithm and PubChem structure clustering tools to construct similarity matrices for proteins and drugs to predict DTA. In addition, the SimBoost model[12] extracts features from drugs, targets, and drug-target pairs through gradient boosting. In terms of performance, machine learning-based methods have achieved optimization and improvement. However, in practical applications, these methods rely on complex feature engineering, which often requires specialized domain knowledge to complete. In recent years, deep learning and graph neural networks have made significant progress in the field of drug-target affinity prediction. The DeepDTA[13] model uses the SMILES sequences of drugs and the amino acid sequences of proteins as feature inputs, employs two convolutional neural networks (CNNs) to extract local sequence information, and then predicts DTA through fully connected layers. Subsequently, the authors further incorporated Ligand Maximum Common Substructure (LMCS) and Protein Domain and Motif (PDM) into the WideDTA[14] model, introducing attention mechanisms to enhance the interpretability of DTA prediction. Nguyen et al.[16] proposed GraphDTA, which encodes drugs as undirected graphs represented by feature matrices and adjacency matrices, further improving the performance of DTA prediction. MGraphDTA[19] is a hyper-deep graph neural network with 27 layers that captures multi-scale features while cleverly avoiding the problem of vanishing gradients. DGraphDTA[20] utilizes the structural information of molecules and proteins to construct graph features of drug molecules and proteins, introducing graph neural networks to obtain their feature representations. In the study of drug-target interactions, three-dimensional structural data representation has also been applied. For example, the AtomNet[21] model processes the three-dimensional structures of drug-target complexes using three-dimensional convolutional neural networks. The three-dimensional structures of drugs and targets theoretically more accurately elucidate their interaction mechanisms. However, currently, for most drug-target pairs with binding sites, there is still a lack of three-dimensional structural datasets available for experiments, which leads to high computational costs and limits large-scale applications.

Accurate prediction of interactions between drugs and targets holds profound significance for drug development. Therefore, this paper proposes an innovative drug-target affinity prediction model, GRAM-DTA, based on graph representation and attention fusion mechanisms. Considering the complexity and diversity of drug molecular structures, this method transforms drug molecules into graph form to fully retain the structural information of the molecules and employs multi-layer Graph Isomorphism Networks (GIN) to extract drug features. To effectively represent the structural and functional information of targets, this method leverages the ESM-1b model proposed by Rao et al.[26] to infer end-to-end residue contact information of proteins. The method combines Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) to extract feature representations of targets. An attention mechanism is introduced in the model to simulate interactions between drug molecules and amino acids, thereby achieving feature fusion. To evaluate the performance of the model, this paper conducts experimental validation on the Davis and KIBA benchmark datasets. The overall framework of the model is shown in Figure 1. The main contributions of this method are as follows:

(1) Both drug molecules and amino acids are represented as graph data types. Deep GIN networks and multi-layer GCN-GAT modules are employed to process the feature information of drugs and targets, fully retaining the structural information of the molecules while effectively extracting the structural and functional features of the targets.

(2) An attention fusion mechanism is introduced in the feature fusion stage to simulate the interactions between drug molecules and amino acids, dynamically adjust the importance of

ISSN: 2688-8645

features, and capture the interaction patterns between drugs and targets, thereby more accurately predicting DTA.

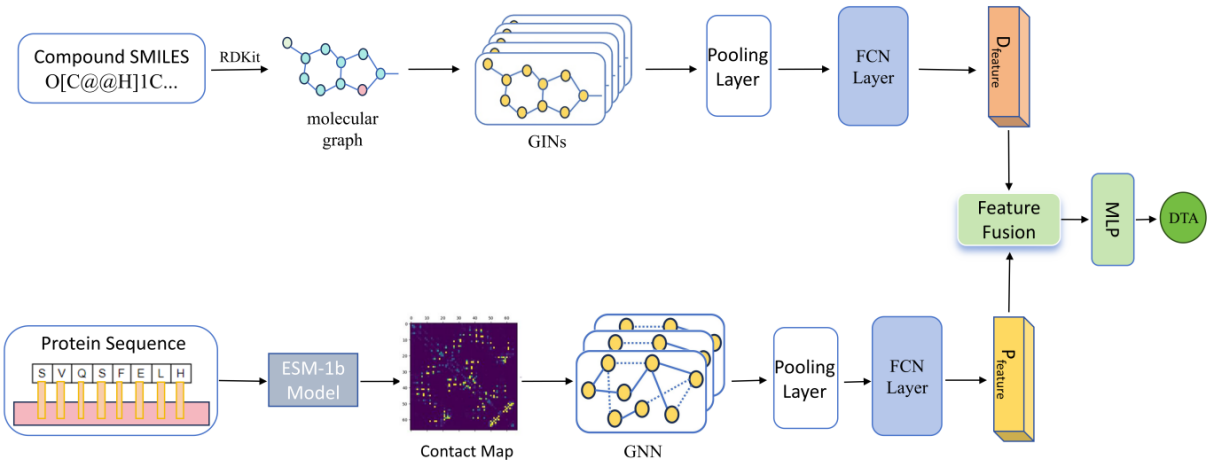

Figure 1. Overall Framework of the GRAM-DTA Model

2. Data and Methods

2.1. Datasets

This paper employs the Davis[23] and KIBA[24] benchmark datasets to evaluate the research on drug-target affinity prediction. The Davis dataset originates from experiments measuring the binding affinity between protein kinases and compounds, covering 442 targets and 68 drug molecules, forming a total of 30,056 drug-target interaction pairs. The affinity values in this dataset are represented by the dissociation constant (K_d). Given the sparse distribution of K_d values, it is common to transform them into logarithmic space to obtain more concentrated values pK_d. The transformed range is [5.0, 10.8], with higher values indicating stronger binding affinity. The specific transformation formula is as follows:

$$pK_d = -\log_{10}\left(\frac{K_d}{10^9}\right) \tag{1}$$

The KIBA dataset was compiled by Tang et al.[24] by integrating drug-target interaction data from various sources. The initial KIBA dataset contained a large amount of sparse interaction information. To improve data quality and the reliability of experiments, this study used the filtered KIBA dataset. The dataset now includes 2,111 drug molecules and 229 targets, totaling 118,254 drug-target interaction pairs. The affinity values in the KIBA dataset, known as KIBA scores, integrate multiple statistical information, including half-inhibition concentration (IC_{50}), inhibition constant (K_i), and dissociation constant (K_d), with a value range of [0.0, 17.2]. Table 1 summarizes the overall statistical information of these two benchmark datasets.

Table 1. Statistical Information of Benchmark Datasets

Datasets	Proteins	Drugs	Pairs
Davis	442	68	30,056
KIBA	229	2,111	118,254

2.2. Graph Representation of Drugs

The structure of drug molecules is complex and diverse. Transforming drug molecules into graph form can completely retain the structural information of the molecules. In the graph

ISSN: 2688-8645

representation, each atom of the drug molecule is regarded as a node, and the chemical bonds between atoms are represented as edges. This process begins with converting the SMILES string of the drug into a molecular graph. SMILES strings are a method of describing molecular structures using ASCII characters, detailing key information such as the types of atoms, bond types and directions, and ring structures within the molecule. To describe the nodes in the graph, we utilized the atomic feature set designed in DeepChem[25], as shown in Table 2, including atomic symbol, number of neighboring atoms, number of neighboring hydrogen atoms, atomic valence, and whether it is an aromatic structure in these five key categories. With the help of the open-source cheminformatics tool RDKit, the SMILES string is converted into the corresponding molecular graph G=(V, E), where V represents the set of nodes and E represents the set of edges. Each node is ultimately assigned a 78-bit binary feature vector, providing rich details for the graph representation of the drug molecule.

Table 2. Atomic Features and Dimensions

Feature	Dimension
One-hot encoding of the atom element	44
One-hot encoding of the degree of the atom in the molecule	11
One-hot encoding of the total number of H bound to the atom	11
One-hot encoding of the number of implicit H bound to the atom	11
Whether the atom is aromatic	1
All	78

Graph Neural Network (GNN) models are specifically designed for graph-structured data and can directly process nodes and edges in the graph representation. Traditional GNN models, however, may lose some important structural information when processing graph-structured data, resulting in insufficient feature representations. Based on the experiments by Nguyen et al.[16], we chose multi-layer Graph Isomorphism Networks (GIN) to extract the graph feature representation of drugs. The feature aggregation mechanism of GIN has unique advantages, and its node feature update process is shown in Figure 2.

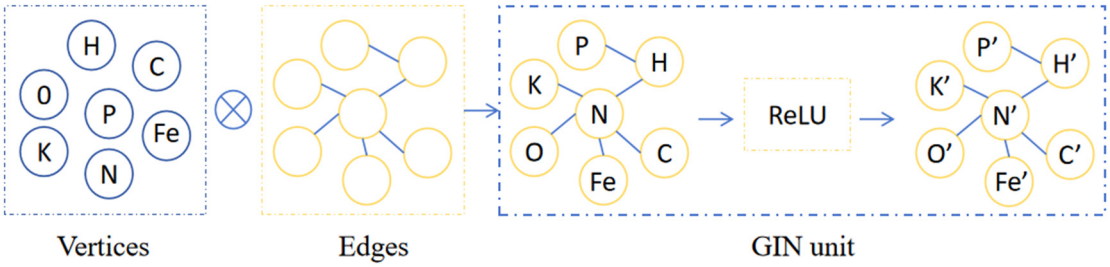

Figure 2. Updating Node Features

Firstly, it employs a summation method to integrate the features of neighboring nodes, which preserves the complete information of the neighboring node features and avoids losing important feature details during the aggregation process. Secondly, GIN introduces a multi-layer perceptron (MLP) and learnable parameters ε, enabling the model to flexibly adjust the contributions of both the node's own features and its neighbors' features. This mechanism allows GIN to capture complex interactions between nodes, thereby better learning the feature representation of the graph structure. Therefore, the feature update of each node in the GIN network using an MLP can be represented as follows:

ISSN: 2688-8645

$$d_i^k = MLP^k((1+\varepsilon^k)d_i^{k-1} + \sum_{j\in N(i)} d_j^{k-1}) \tag{2}$$

In the formula, d_i^k represents the feature representation of node i at the k-th layer, N(i) is the set of nodes adjacent to node i, and ε is a learnable parameter. In the model architecture of this study, we constructed a multi-layer graph isomorphism network composed of 5 GIN layers, with a batch normalization layer following each GIN layer. Within each graph feature embedding learning block, the output of the last GIN layer is passed to a global max pooling layer, which can effectively ignore differences in the node set and node count, thereby generating a graph-level embedding representation D:

$$D = gmp(BN(\{d_i^k | i \in G\})) \tag{3}$$

2.3. Graph Representation of Targets

In the task of drug-target affinity prediction, the feature representation of targets is equally crucial. To effectively represent the structural and functional information of targets, we chose the ESM-1b model proposed by Rao et al.[26] to infer end-to-end residue contact information of proteins. This model can obtain relatively accurate information without sequence alignment. After obtaining the contact graph, we constructed a weighted residue contact graph. Specifically, we treated each residue as a node in the graph and used a threshold of 0.5 to determine whether two residues are connected. When the contact probability between residues exceeds 0.5, it is considered a stable contact, and this relationship is established as a connecting edge; if it is below this threshold, it is considered no contact. We used the contact probability as the weight of the connecting edge to reflect the relative strength of different contacts, thereby comprehensively capturing the conformational information of the protein. The ESM-1b model requires that the input sequence length does not exceed 1024. In our model, for sequences longer than 1000, we processed them by truncation. To mine potential features, we used Graph Neural Networks， a type of deep learning model that has gradually emerged in recent years and can handle non-Euclidean structured data. Among them, Graph Convolutional Networks (GCN) is a commonly used type of GNN, with each convolutional operation as follows:

$$H^{(k+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(k)}w^{(k)}) \tag{4}$$

In the formula, $H^{(k)}$ represents the node feature matrix of the k-th layer, $\tilde{A}$ is the adjacency matrix A of the graph plus the identity matrix, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $w^{(k)}$ is the weight matrix of the k-th layer, and σ is the nonlinear activation function. The entire process can be understood as follows: First, the graph is normalized through $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, then the node feature matrix $H^{(k-1)}$ of the (k−1)-th layer is multiplied with the normalized graph matrix, followed by multiplication with the weight matrix $w^{(k)}$ of the k-th layer, and finally, the node feature matrix σ of the k-th layer is obtained through the nonlinear activation function $H^{(k)}$. Graph Attention Networks (GAT) is another type of graph neural network that uses attention mechanisms to aggregate feature information from neighboring nodes, thereby achieving dynamic weighted fusion of node features. The calculation formula for node features in GAT can be specifically expressed as:

$$h_v^{(k)} = \sigma(\sum_{u\in N(v)} t_{vu}^{(k)} W^{(k)} h_u^{(k-1)}) \tag{5}$$

Here, $h_v^{(k)}$ is the feature vector of node v in the k-th layer, $N(v)$ is the set of neighboring nodes of node v, and $W^{(k)}$ is the weight matrix of the k-th layer. $t_{vu}^{(k)}$ is the attention coefficient between node v and node u, which can be calculated using the following formula:

ISSN: 2688-8645

$$t_{vu}^{(k)} = \text{soft}\max_u(e_{vu}^{(k)}) \tag{6}$$

$$e_{vu}^{(k)} = \text{Leaky ReLU}(a^{(k)}(W^{(k)}h_v^{(k-1)}) \oplus a^{(k)}(W^{(k)}h_u^{(k-1)})) \tag{7}$$

In the formula, $a^{(k)}$ is the attention vector of the k-th layer, and $\oplus$ denotes the concatenation operation. This process can be understood as follows: The node feature vector $h_v^{(k-1)}$ of the (k−1)-th layer is transformed into a new feature vector through the weight matrix $W^{(k)}$. Then, the unnormalized attention coefficient $t_{vu}^{(k)}$ is calculated using the attention vector $a^{(k)}$ and the LeakyReLU function. Subsequently, the attention coefficient $t_{vu}^{(k)}$ is normalized using the softmax function. Finally, the node feature vector $h_v^{(k)}$ of the k-th layer is obtained through weighted summation and a nonlinear activation function σ. Therefore, in the model architecture of this study, we employed a strategy combining multi-layer GCN with GAT, as shown in Figure 3, to construct a conformation learning module specifically designed for aggregating node and edge features. To fully capture the evolutionary information in the sequence, we first embedded individual residues using the ESM-1b model and then obtained sequence-level embedding representations by averaging these residue embedding. In the feature embedding layer, we further utilized a multi-layer perceptron network to learn the evolutionary information of the sequence-level embedding, thereby providing a richer feature basis for subsequent analysis and prediction tasks.

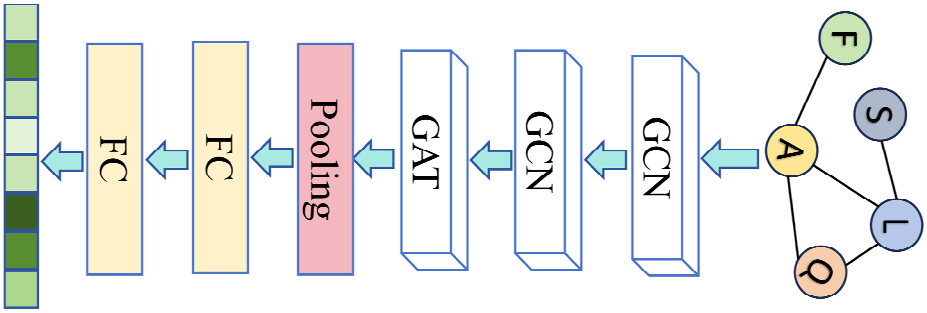

Figure 3. Graph Neural Network for Extracting Target Features

2.4. Attention Fusion Mechanism

To effectively simulate the interactions between drug molecules and amino acids, we introduced an attention mechanism to fuse features in the model design, as shown in Figure 4. Specifically, we dynamically adjusted the importance of each feature based on the input features using a nonlinear activation function and learnable parameterized mappings. This strategy can more sensitively capture the interaction patterns between drugs and targets, thereby improving the model's performance in DTA prediction tasks. First, we extracted the graph representation features D from the compounds and the structural features P from the proteins. Next, D and P were input into the attention network to calculate the corresponding attention coefficient weights, with the calculation formula as follows:

$$\alpha = \text{Sigmoid}(\text{attention}(D + P)) \tag{8}$$

$$\text{attention}(x) = w_2(\text{ReLU}(w_1 x + b_1) + b_2 \tag{9}$$

In the formula, w_1 and w_2 are trainable weights, while b_1 and b_2 are bias terms. Finally, using the calculated attention coefficient weights, the drug features D and target features P are

ISSN: 2688-8645

weighted and summed to obtain the feature embedding representation that integrates information from both the drug and target:

$$F = D \odot \alpha + P \odot (1 - \alpha) \tag{10}$$

In the formula, $\odot$ denotes the element-wise product.

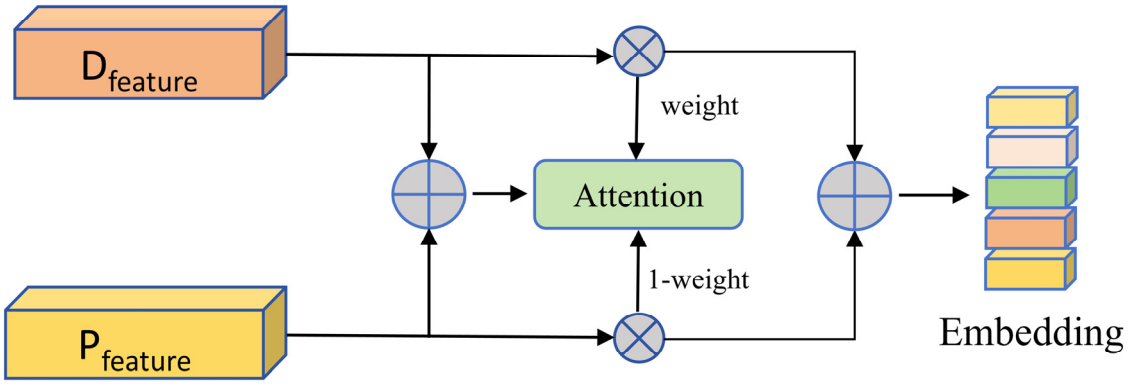

Figure 4. Attention Fusion Module

The attention module outputs the fused feature F of the drug and target, which is then input into a multi-layer perceptron (MLP) composed of three fully connected layers. The input layer of the MLP receives the output feature from the attention module, followed by two fully connected layers each with 1024 nodes. Each fully connected layer is succeeded by a Dropout layer to prevent model over-fitting. The Dropout mechanism regularizes the neural network by reducing co-adaptation among nodes. The third fully connected layer contains 512 nodes and is directly connected to the output layer without Dropout. Additionally, we use the ReLU activation function in the fully connected layers to facilitate the minimization of differences between label values and predicted values during training. During model training, we select Mean Squared Error (MSE) as the loss function, where y_i represents the actual label value of the i-th sample, $\hat{y}_i$ represents the model's predicted value for the i-th sample, and n represents the total number of samples. The calculation formula for Mean Squared Error is as follows:

$$loss = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

2.5. Evaluation Metrics

As a regression model, the task of drug-target affinity prediction employs three important evaluation metrics: Concordance Index (CI), Mean Squared Error (MSE), and Regression toward the Mean (r_m^2). The CI value is used to assess the consistency between predicted and actual values, the MSE value measures the difference between predicted and actual values, and the r_m^2 value evaluates the model's degree of fit. Through these three metrics, the predictive performance of the model can be comprehensively evaluated to ensure its reliability and accuracy in practical applications. The calculation formulas for these three evaluation metrics can be expressed as follows:

$$CI = \frac{1}{Z}\sum_{\delta_x > \delta_y} h(b_x - b_y) \tag{12}$$

ISSN: 2688-8645

$$h(x)=\begin{cases}0 & \text{if} \quad x<0\\ 0.5 & \text{if} \quad x=0\\ 1 & \text{if} \quad x>0\end{cases} \tag{13}$$

In the formula, δ_x, $\delta_y$represents the label value, b_x, $b_y$is the corresponding predicted value, Z is the normalization constant, and h(x) is the step function.

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{14}$$

In the formula,y_i represents the label value of the i-th sample, $\hat{y}_i$ represents the predicted value of the i-th sample, and N is the number of samples.

$$r_m^2=r^2\times(1-\sqrt{r^2-r_0^2}) \tag{15}$$

In the formula, r^2 and r_0^2 refer to the squared correlation coefficients between the label values and predicted values with and without intercept, respectively.

3. Experiments and Results

3.1. Baseline Methods

To further compare and analyze the performance of the model proposed in this paper, we conducted a comparative analysis with five other classic and excellent models, namely SimBoost[12], DeepDTA[13], GraphDTA[16], MGraphDTA[19], DGraphDTA[20] and IMAEN [29]. To ensure fairness and objectivity, all models were evaluated under the same conditions. Specifically, we used the same baseline datasets and followed a consistent dataset splitting strategy, with a training set to test set ratio of 4:1. Additionally, to enhance the reliability of the results, we performed five-fold cross-validation for each model and took the average of the experimental results as the final performance metric.

3.2. Comparison with Baseline Methods

The comparison results of the model with the baseline methods are shown in Tables 3 and 4, where the best results in each column are indicated in bold, and "-" indicates that the corresponding results were not provided in the original text. From the experimental results on the Davis dataset, our model achieved a 0.1% improvement in CI value and a 3.1% increase in r_m^2 value compared to the best-performing baseline model. Although our model's MSE value was 0.6% higher than the best-performing baseline model, the latter's CI/r_m^2 values were significantly lower than those of our model. On the KIBA dataset, compared to the best-performing baseline model, our model achieved a 0.3% decrease in MSE value and a 3.4% increase in r_m^2 value, while maintaining a CI value comparable to the best-performing baseline model.

The SimBoost model is an earlier machine learning-based model that uses traditional similarity calculation methods to extract features. Its high MSE indicates a large prediction error, and its overall performance is at a relatively low level among all baseline models. Our model significantly outperforms SimBoost in both CI and MSE values, especially in reducing prediction errors. DeepDTA and GraphDTA are deep learning-based models that use convolutional neural networks to extract protein features. Although there is a slight improvement in CI values, the MSE values remain relatively high. DGraphDTA employs graph convolutional networks to extract features from compounds and proteins, significantly improving CI values and reducing MSE values, which demonstrates the advantages of graph neural networks in handling graph-structured data. MGraphDTA shows performance close to DGraphDTA in CI and MSE, but its r_m^2

ISSN: 2688-8645

value is higher than that of DGraphDTA, indicating that optimizations in feature extraction and model design have led to performance improvements. IMAEN uses graph convolutional networks to extract features from compounds and proteins separately and achieves the highest CI values on six benchmark datasets. Despite this, our model also slightly outperforms the IMAEN model in CI values, indicating that our model performs best in terms of consistency in the prediction task.

Overall, with the advancement of technology, the performance of models has gradually improved. Traditional methods like SimBoost, although achieving good results in the early stages, have been significantly outperformed by the application of deep learning techniques, such as the DeepDTA and GraphDTA models. Particularly, the DGraphDTA model has demonstrated the advantages of Graph Neural Networks in handling graph-structured data. Our proposed model employs an improved GNN module to extract features from drugs and targets, achieving a high CI value while further reducing the MSE to 0.123. This indicates that the optimizations in feature extraction and model design have led to enhanced performance.

Table 3. Performance Comparison with Baseline Methods on the Davis Dataset

Method	Compounds	Proteins	CI	MSE	r_m^2
SimBoost	Pubchem-Sim	Smith-Waterman	0.872	0.282	0.644
DeepDTA	CNN	CNN	0.878	0.261	0.630
DGraphDTA	GCN	GCN	0.904	**0.202**	0.700
GraphDTA	GIN	CNN	0.893	0.229	0.660
MGraphDTA	MGNN	MCNN	0.900	0.207	0.710
IMAEN	GCN	GCN	0.905	0.211	0.705
Our model	MGIN	GCN	**0.906**	0.208	**0.741**

Table 4. Performance Comparison with Baseline Methods on the KIBA Dataset

Method	Compounds	Proteins	CI	MSE	r_m^2
SimBoost	Pubchem-Sim	Smith-Waterman	0.872	0.282	0.644
DeepDTA	CNN	CNN	0.878	0.261	0.630
DGraphDTA	GCN	GCN	0.904	0.126	0.786
GraphDTA	GAT&GCN	CNN	0.891	0.139	0.780
MGraphDTA	MGNN	MCNN	0.901	0.207	0.710
IMAEN	GCN	GCN	0.898	0.130	0.775
Our model	MGIN	GCN	**0.904**	**0.123**	**0.820**

In terms of feature representation, compared to the DeepDTA model that uses drug and target sequence data or the GraphDTA model that only employs graph representations for drug features, our model shows significant improvements in key evaluation metrics. Of course, compared to the MGraphDTA and DGraphDTA models, which also use graph feature inputs for drugs and targets, our model does not achieve a noticeable improvement in CI values. However, it obtains the best r_m^2 values in the experimental results on both the Davis and KIBA datasets. Moreover, as shown in the experimental results in Table 3, our model does not achieve the best MSE value on the Davis dataset. However, from the MSE results in Table 4, our model achieves the best MSE value, which is significantly better than the experimental results on the Davis dataset. The reason is that the KIBA dataset has a much larger volume and more appropriate

ISSN: 2688-8645

sample distribution compared to the Davis dataset. The above analysis of the results demonstrates that our model has achieved further improvements and can more effectively predict the binding affinity between drugs and targets.

3.3. Ablation Study

To thoroughly investigate the specific roles of different modules in the model, we conducted ablation studies on the KIBA dataset by modifying parts of the model to assess the contributions of each module. First, we analyzed the impact of the depth of Graph Isomorphism Networks (GIN) on model performance by increasing or reducing the number of GIN layers. In the second modification, we replaced the target feature extraction module with a one-dimensional convolutional neural network structure to extract sequence features of targets, thereby evaluating the contribution of the target graph representation to model performance. In the third modification, we employed a simple concatenation fusion method at the feature fusion stage of drugs and targets, comparing the impact of the attention fusion module on model performance. The results of the experiments are shown in Table 5, with the best results in each column highlighted in bold.

Table 5. Ablation Study on the KIBA Dataset

Different method		CI	MSE	r_m^2
Impact of the layers of the graph isomorphism network	3	0.866	0.153	0.771
	4	0.871	0.147	0.775
	5	0.904	0.123	0.820
	6	0.867	0.156	0.771
	7	0.869	0.152	0.783
Without protein graph		0.883	0.136	0.776
Without attention fusion mechanism		0.886	0.142	0.775
Our model		**0.904**	**0.123**	**0.820**

By comparing the experimental results, it can be clearly seen how the three modifications affect the model's performance. As shown in the line chart in Figure 5, the CI/MSE values do not maintain a positive correlation with the increase in the number of GIN layers. Specifically, as the number of GIN layers increases, the CI value no longer rises, while the MSE value or r_m^2 value may increase. Conversely, when the number of GIN layers is reduced, it also leads to a decrease or increase in different evaluation metrics. This indicates that the rational setting of the number of GIN layers is crucial for improving model performance.

In the second modification, after replacing the graph neural network for extracting target features with a one-dimensional convolutional neural network, the model's predictive performance significantly decreased, with a 2.1% drop in CI value, a 1.3% drop in MSE value, and a 4.6% drop in r_m^2. This result indicates that, for the extraction of target features, graph neural networks have a clear advantage over one-dimensional convolutional neural networks, being able to more effectively capture the structural and semantic information of targets, thereby positively impacting the model's predictive performance. Similarly, the model lacking attention fusion in the feature fusion stage also experienced a substantial decline in predictive performance. This demonstrates that the attention mechanism plays a crucial role in the feature fusion process, dynamically adjusting the importance of different features to enhance the model's focus on key information and improve its predictive capability.

These ablation study results indicate that the rational setting of the number of GIN layers and the selection of appropriate feature representation methods play a crucial role in enhancing

ISSN: 2688-8645

model performance. Even after removing or modifying certain key modules, the model still maintains comparable levels of key evaluation metrics compared to the baseline model, demonstrating its robustness and adaptability. The synergistic effect of these modules effectively captures and utilizes the feature information of drugs and targets, thereby improving the overall performance of the model. Through these in-depth analyses and experiments, a better understanding of the role of each component in the model can be achieved, and targeted directions for future model optimization are also provided.

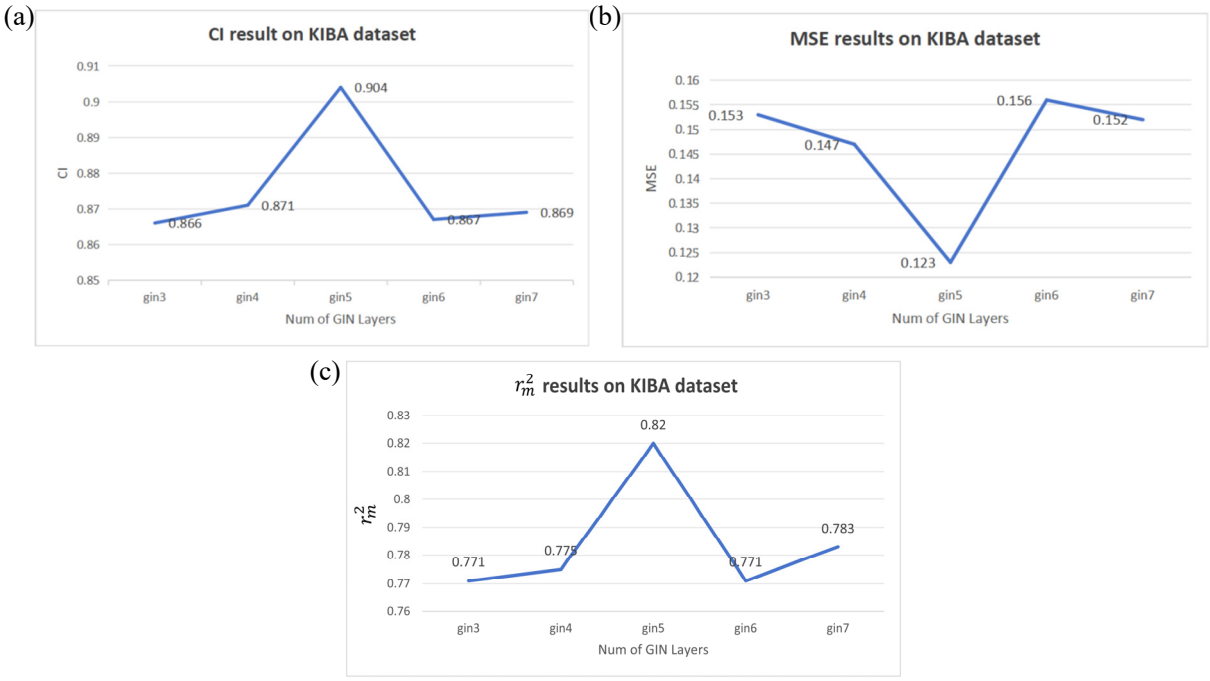

Figure 5. Impact of GIN Layers on the Performance of GRAM-DTA

3.4. Interpretability of the Model

Deep learning models are often regarded as "black boxes" due to their lack of interpretability, as it is difficult to trace which features have a key impact on the results. This lack of interpretability limits the application of deep learning methods. Inspired by Grad-AAM[32], we utilize gradient information to describe the influence of each atom or node on the prediction results. Through this method, the regions of the graph structure that contribute the most to the prediction results are presented in the form of heatmaps, thereby enhancing the interpretability of the model.

In the GRAM-DTA model, the last layer of GIN contains rich high-level semantic information. Therefore, we choose to visualize the drug graph representation to generate heatmaps that depict the atoms and functional groups that contribute most significantly to the prediction results. Specifically, we represent the feature map of the last graph convolutional layer in the model as m. To generate the probability map P_d for each atom node v in the drug molecule, we calculate the gradient of the predicted affinity score DTA of atom node v with respect to the c-th channel in the feature map m.

$$U_c = \frac{1}{|V|}\sum_{v\in V}\frac{\partial DTA}{\partial m_v^c} \tag{16}$$

Subsequently, each channel of the feature map m is weighted and summed, and then processed through the ReLU activation function.

$$P_d = ReLU(\sum_c U_c m^c) \tag{17}$$

ISSN: 2688-8645

Finally, the gradient weights are adjusted to the range [0,1] using the min-max normalization method, thereby generating a probability map P_d of the weighted distribution of the drug molecule, which is further visualized as a heatmap, as shown in Figure 6.

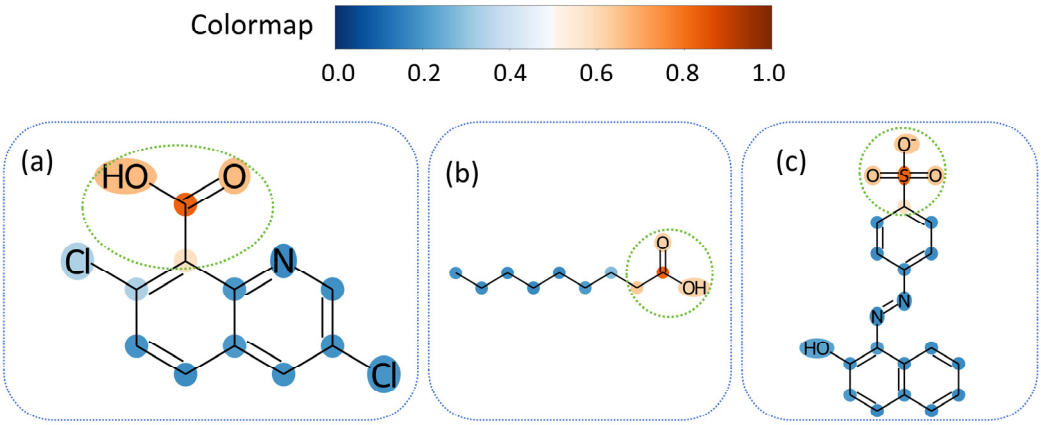

Figure 6. Heatmap of Atom and Functional Group Contributions to the Active Structure of Drug Molecules

The figure displays heatmaps of atomic and functional group contributions, intuitively reflecting the extent of each atom's contribution to the prediction results. Specifically, green circles highlight the fatty acids in the different molecular structures of (a) and (b), as well as the importance of the sulfonic acid group atoms in (c). According to previous studies, these structures have been confirmed as common binding sites for targets. Representing drug molecules as graph structures and learning their topological patterns through Graph Neural Networks can accurately distinguish the active structures of drug molecules. Based on this, and in conjunction with the aforementioned experimental results, we can observe that selecting an appropriate depth for GIN can further enhance the model's ability to capture the structural features of drug molecules.

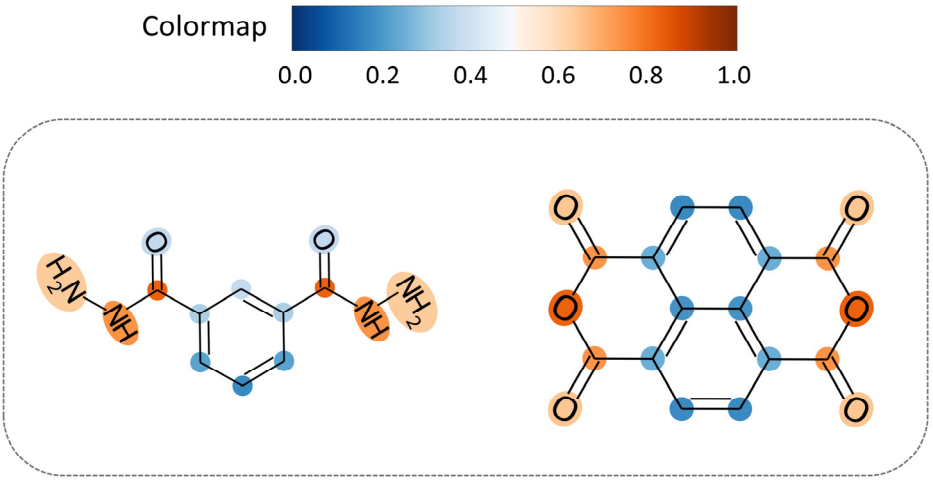

Figure 7. Atomic Heatmap Distribution of Symmetric Compounds

Figure 7 shows that the distribution of atomic heatmaps in compounds with symmetrical structures also exhibits symmetry. This indicates that representing compounds as graph structures and using GNNs to extract their feature patterns can effectively preserve the original structural information of the compounds. Overall, Grad-AAM provides biological

ISSN: 2688-8645

interpretability for DTA prediction models, helping us to gain a deeper understanding of the model's decision-making process.

4. Summary and Discussion

This paper proposes a drug-target affinity prediction model, GRAM-DTA, based on graph representation and attention fusion mechanisms. The model represents drug molecules and target proteins as graph data types and utilizes deep Graph Isomorphism Networks, multi-layer Graph Convolutional Networks, and Graph Attention Networks to process the feature information of drugs and targets. In the feature fusion stage, an attention mechanism is introduced to simulate the interactions between drug molecules and target proteins, dynamically adjust the importance of features, and capture the interaction patterns between drugs and targets. Experiments were conducted on the Davis and KIBA benchmark datasets, and the model was compared with current state-of-the-art models. The results show that the model performs excellently in the drug-target affinity prediction task, outperforming other traditional methods and existing models. Specifically, on the Davis dataset, our model achieved a 0.1% improvement in CI value and a 3.1% increase in r_m^2 value compared to the best-performing baseline model. On the KIBA dataset, compared to the best-performing baseline model, our model achieved a 0.3% decrease in MSE value and a 3.4% increase in r_m^2 value, while maintaining a CI value comparable to the best-performing baseline model. The experimental results demonstrate the effectiveness and practicality of the model, providing a powerful tool for drug discovery and development.

Despite the significant performance improvements achieved by the GRAM-DTA model in the DTA prediction task, there are still some limitations. For example, when constructing the contact graph of targets, sequences longer than 1000 are simply truncated without using more efficient methods. Although this approach simplifies the computation to some extent, it may lose important structural information, thereby affecting the model's prediction accuracy. Future work can explore more efficient methods to handle long target sequences. Additionally, with the continuous development of deep learning and graph neural network technologies, future research can further explore more advanced model architectures and algorithms to improve the accuracy and efficiency of drug-target affinity prediction. For example, integrating multimodal data to construct more comprehensive feature representations can further enhance the model's performance and applicability.

Acknowledgments

The research was partly supported by the National Natural Science Foundation of China (No. 62175037) and the funding of Zhejiang-French Digital Monitoring Lab for Aquatic Resources and Environment, Department of Science and Technology of Zhejiang Province. Thanks to Professor Sihua Peng for guiding the study of GNN knowledge and introducing GNN research.

References

[1] Zhao L, Zhu Y, Wang J, et al. A brief review of protein–ligand interaction prediction[J]. Computational and Structural Biotechnology Journal, 2022, 20: 2831-2838.

[2] Zhao L, Wang H, Shi S. PocketDTA: an advanced multimodal architecture for enhanced prediction of drug– target affinity from 3D structural data of target binding pockets[J]. Bioinformatics, 2024, 40(10): btae594.

[3] Yu W, MacKerell A D. Computer-aided drug design methods[J]. Antibiotics: methods and protocols, 2017: 85-106.

[4] Vemula D, Jayasurya P, Sushmitha V, et al. CADD, AI and ML in drug discovery: A comprehensive review[J]. European Journal of Pharmaceutical Sciences, 2023, 181: 106324.

[5] TANG Yue-wei LIU Zhi-ping. Drug-target Affinity Prediction Based on Deep Learning and Multi-layered Information Fusion[J]. China Biotechnology, 2021, 41(11): 40-47.

[6] LIU Xiaoguang, LI Mei. A survey of deep learning-based drug-target interaction prediction[J]. CAAI transactions on intelligent systems, 2024, 19(3): 494–524.

[7] Reuter J A, Spacek D V, Snyder M P. High-throughput sequencing technologies[J]. Molecular cell, 2015, 58(4): 586-597.

[8] Pagadala N S, Syed K, Tuszynski J. Software for molecular docking: a review[J]. Biophysical reviews, 2017, 9: 91-102.

[9] Zou Y, Wang R, Du M, et al. Identifying Protein–Ligand Interactions via a Novel Distance Self-Feedback Biomolecular Interaction Network[J]. The Journal of Physical Chemistry B, 2023, 127(4): 899-911.

[10] Ru X, Ye X, Sakurai T, et al. Current status and future prospects of drug–target interaction prediction[J]. Briefings in Functional Genomics, 2021, 20(5): 312-322.

[11] Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug–target interaction predictions[J]. Briefings in bioinformatics, 2015, 16(2): 325-337.

[12] He T, Heidemeyer M, Ban F, et al. SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines[J]. Journal of cheminformatics, 2017, 9: 1-14.

[13] Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction[J]. Bioinformatics, 2018, 34(17): i821-i829.

[14] Öztürk H, Ozkirimli E, Özgür A. WideDTA: prediction of drug-target binding affinity[J]. arXiv preprint arXiv:1902.04166, 2019.

[15] Zhao Q, Duan G, Yang M, et al. AttentionDTA: Drug–target binding affinity prediction by sequence-based deep learning with attention mechanism[J]. IEEE/ACM transactions on computational biology and bioinformatics, 2022, 20(2): 852-863.

[16] Nguyen T, Le H, Quinn T P, et al. GraphDTA: predicting drug–target binding affinity with graph neural networks[J]. Bioinformatics, 2021, 37(8): 1140-1147.

[17] Wang S, Song X, Zhang Y, et al. MSGNN-DTA: multi-scale topological feature fusion based on graph neural networks for drug–target binding affinity prediction[J]. International Journal of Molecular Sciences, 2023, 24(9): 8326.

[18] Qian Y, Ni W, Xianyu X, et al. DoubleSG-DTA: Deep Learning for Drug Discovery: Case Study on the Non-Small Cell Lung Cancer with EGFR T 790 M Mutation[J]. Pharmaceutics, 2023, 15(2): 675.

[19] Yang Z, Zhong W, Zhao L, et al. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction[J]. Chemical science, 2022, 13(3): 816-833.

[20] Jiang M, Li Z, Zhang S, et al. Drug–target affinity prediction using graph neural network and contact maps[J]. RSC advances, 2020, 10(35): 20701-20712.

[21] Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery[J]. arXiv preprint arXiv:1510.02855, 2015.

[22] Li Y, Rezaei M A, Li C, et al. DeepAtom: A framework for protein-ligand binding affinity prediction[C]//2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019: 303-310.

[23] Davis M I, Hunt J P, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity[J]. Nature biotechnology, 2011, 29(11): 1046-1051.

[24] Tang J, Szwajda A, Shakyawar S, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis[J]. Journal of Chemical Information and Modeling, 2014, 54(3): 735-743.

[25] Ramsundar B, Eastman P, Walters P, et al. Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy[J]. Drug Discovery, and More, 2019, 1.

ISSN: 2688-8645

[26] Rao R, Meier J, Sercu T, et al. Transformer protein language models are unsupervised structure learners[J]. Biorxiv, 2020: 2020.12. 15.422761.

[27] Bian J, Zhang X, Zhang X, et al. MCANet: shared-weight-based MultiheadCrossAttention network for drug–target interaction prediction[J]. Briefings in Bioinformatics, 2023, 24(2): bbad082.

[28] Li Z, Ren P, Yang H, et al. TEFDTA: a transformer encoder and fingerprint representation combined prediction method for bonded and non-bonded drug–target affinities[J]. Bioinformatics, 2024, 40(1): btad778.

[29] Zhang J, Liu Z, Pan Y, et al. IMAEN: An interpretable molecular augmentation model for drug–target interaction prediction[J]. Expert Systems with Applications, 2024, 238: 121882.

[30] Deng J, Zhang Y, Pan Y, et al. Multidta: drug-target binding affinity prediction via representation learning and graph convolutional neural networks[J]. International Journal of Machine Learning and Cybernetics, 2024: 1-10.

[31] Jiang M, Wang S, Zhang S, et al. Sequence-based drug-target affinity prediction using weighted graph neural networks[J]. BMC genomics, 2022, 23(1): 449.

[32] Zhu Z, Yao Z, Zheng X, et al. Drug–target affinity prediction method based on multi-scale information interaction and graph optimization[J]. Computers in Biology and Medicine, 2023, 167: 107621.

[33] Feng Y H, Zhang S W. Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs[J]. Molecules, 2022, 27(9): 3004.

[34] Jin Y, Lu J, Shi R, et al. Embeddti: enhancing the molecular representations via sequence embedding and graph convolutional network for the prediction of drug-target interaction[J]. Biomolecules, 2021, 11(12): 1783.