JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

# MMSG-DTA: A Multimodal, Multiscale Model Based on Sequence and Graph Modalities for Drug-Target Affinity Prediction

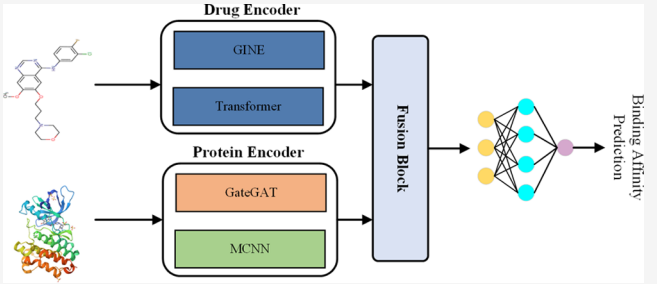
Jiahao Xu, Lei Ci, Bo Zhu, Guanhua Zhang, Linhua Jiang, Shixin Ye-Lehmann, and Wei Long*

Cite This: J. Chem. Inf. Model. 2025, 65, 981−996 Read Online

ACCESS | Metrics & More | Article Recommendations

Downloaded via INDIAN INST OF TECH MADRAS on November 11, 2025 at 03:28:09 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

**ABSTRACT:** Drug-Target Affinity (DTA) prediction is a cornerstone of drug discovery and development, providing critical insights into the intricate interactions between candidate drugs and their biological targets. Despite its importance, existing methodologies often face significant limitations in capturing comprehensive global features from molecular graphs, which are essential for accurately characterizing drug properties. Furthermore, protein feature extraction is predominantly restricted to 1D amino acid sequences, which fail to adequately represent the spatial structures and complex functional regions of proteins. These shortcomings impede the development of models capable of fully elucidating the mechanisms underlying drug-target interactions. To overcome these challenges, we propose a multimodal, multiscale model based on Sequence and Graph Modalities for Drug-Target Affinity (MMSG-DTA) Prediction. The model combines graph neural networks with Transformers to effectively capture both local node-level features and global structural features of molecular graphs. Additionally, a graph-based modality is employed to improve the extraction of protein features from amino acid sequences. To further enhance the model's performance, an attention-based feature fusion module is incorporated to integrate diverse feature types, thereby strengthening its representation capacity and robustness. We evaluated MMSG-DTA on three public benchmark data sets—Davis, KIBA, and Metz—and the experimental results demonstrate that the proposed model outperforms several state-of-the-art methods in DTA prediction. These findings highlight the effectiveness of MMSG-DTA in advancing the accuracy and robustness of drug-target interaction modeling.




## 1. INTRODUCTION

The development of new drugs offers more treatment options for patients, particularly those with rare or refractory diseases.[1] However, the traditional drug development process is complex and time-consuming, typically involving stages such as basic research, drug design, in vitro testing, animal studies, and clinical trials before submission for regulatory approval. This process can take 10 to 17 years, or even longer, with a cost of up to $2.6 billion.[2] More critically, approximately 90% of the drugs eventually fail during clinical trials and do not obtain approval for commercialization.[3]

Drugs exert their therapeutic effects by temporarily binding to target molecules, influencing their functions, and inhibiting specific catalytic reactions.[4] Therefore, the calculation of drug-target interactions (DTI) plays a crucial role in drug discovery. Identifying DTIs is essential for both new drug development and evaluation of potential side effects. By quantifying drug-target binding affinity (DTA), molecules with high affinity can be identified and further optimized as potential drug candidates.[5] Moreover, affinity quantification not only reveals the existence of binary interactions but also provides detailed information on the drug-target binding strength, offering stronger guidance for drug design.[6]

In drug discovery, there are two main computational approaches used to predict the interactions between drugs and specific targets: molecular docking-based methods and other computational methods.[7,8] However, molecular docking has a certain limitations. First, it is computationally expensive and time-consuming. Additionally, when researchers face novel proteins with unknown structures, docking methods may encounter difficulties due to their reliance on a detailed understanding of protein 3D structures.[9] Given these limitations, alternative computational methods have emerged as effective replacements for docking. These methods predict the drug-target binding affinity (DTA) by analyzing the features of drugs and proteins. They are primarily categorized into two types: machine learning-based approaches and deep learning-based approaches.[10]

Received: October 7, 2024
Revised: December 15, 2024
Accepted: December 30, 2024
Published: January 7, 2025

ACS Publications © 2025 American Chemical Society

Machine learning, as an efficient computational approach, has been widely employed for predicting drug-target binding affinities. By constructing features from large annotated data sets, machine learning models are trained to autonomously discern patterns within the data, allowing for predictions on new, unannotated samples. Compared with molecular docking methods, which rely on three-dimensional molecular structures and require extensive computational resources, machine learning offers greater efficiency, enabling large-scale predictions and the identification of a larger pool of potential drug candidates for subsequent experimental validation. The application of machine learning in predicting drug-target binding affinities primarily involves feature-based methods and similarity-based approaches.

Feature-based methods represent compound-protein pairs as descriptor vectors that encode various attributes of compounds and proteins. These feature vectors are then input into models such as the Random Forest (RF) and Support Vector Machine (SVM) to predict novel compound-protein interactions. For instance, Li et al.[11] demonstrated that RF can effectively utilize a larger number of structural features and training samples, achieving superior predictive performance compared to multiple linear regression. Similarly, Shar et al.[12] compared the performance of RF and SVM using the Pred-binding method and found that both models provided robust affinity predictions while avoiding overfitting.

Similarity-based methods rely on the hypothesis that compounds with biological, topological, and chemical similarities exhibit similar functions and biological activities and consequently target similar proteins. In the chemical space, similarity is computed through substructure and isomorphic searches based on molecular representations, whereas in the protein space, it is primarily determined through sequence alignment. These methods construct similarity matrices for compounds or target proteins. For example, the KronRLS model proposed by Pahikkala et al.[13] utilizes similarity metrics of compounds and proteins to construct a Kronecker kernel and applies regularized least-squares (RLS) regression to predict affinities. Additionally, the SimBoost model proposed by He et al.[14] uses the similarity information on compounds and proteins as input for the gradient boosting machine to predict compound-protein affinity.

In the field of DTA prediction, deep learning methods have demonstrated exceptional performance. Early models, such as DeepDTA,[15] utilized the SMILES strings of drugs and the amino acid sequences of proteins as inputs, incorporating two independent one-dimensional convolutional neural network (CNN) modules. Building upon this foundation, WideDTA[16] introduced the ligand's largest common substructure (LMCS) and protein motifs and domains (PDMs), significantly improving model performance. Furthermore, DeepCDA[17] combined long short-term memory networks (LSTMs) and convolutional neural networks (CNNs) to propose a cross-domain compound−protein affinity prediction method. DeepCDA captures long-range dependencies in drug and protein sequences by using LSTMs, while CNNs are used to extract local features, thereby improving both the accuracy and reliability of the predictions. Researchers have proposed utilizing molecular graph representations of drugs as an alternative strategy for creating enriched feature representations. For instance, DGraphDTA[18] is a model that integrates graph neural networks (GNNs) with deep learning, dynamically adjusting the contribution of different layers to enhance the expressive power of drug molecular structures, thereby improving prediction performance. Similarly, GraphDTA[19] analyzes drugs using atomic features and employs four GNN layers to capture complex graph representations, advancing the modeling of topological information compared to previous studies. Meanwhile, MGraphDTA[20] aims to improve prediction accuracy by leveraging multiscale learning to effectively capture both local and global structural features of drug molecules.

Despite significant progress in drug-target affinity prediction (DTA), which has improved accuracy and yielded satisfactory results, several challenges remain. Current methods face difficulties in fully capturing the global characteristics of molecular graphs in drug molecule feature extraction. Many approaches focus primarily on local features, and multilayer graph neural networks (GNNs) often encounter issues of oversmoothing and gradient vanishing, limiting their ability to effectively capture the global structure of molecular graphs. Similarly, protein feature extraction mainly relies on amino acid sequence information. However, linear amino acid sequences are insufficient to fully represent the spatial structure and complex functional domains of proteins, thus limiting the robustness of the feature representations. Moreover, sequence-based models struggle with capturing long-range interactions between residues, impairing their ability to model the global topological structure of proteins. Additionally, most deep learning models for DTA prediction simulate drug-target interactions by simply concatenating the feature vectors of compounds and proteins. This simplified approach lacks interpretability and fails to account for the intermolecular interactions between atoms in compounds and amino acids in proteins, which are crucial for accurately predicting affinity.

In light of these challenges, recent advancements, such as Graphormer,[21] have been widely used for tasks like molecular property prediction, protein structure analysis, and drug design, addressing some of the limitations of traditional methods. By utilizing self-attention mechanisms, Graphormer can better capture both local and global dependencies within molecular graphs and protein structures. Such approaches have inspired and informed our work in DTA prediction as they demonstrate the potential for enhanced interpretability and more accurate simulations of intermolecular interactions, which are critical for affinity prediction.

To overcome the limitations highlighted above and build on the strengths of these recent advancements, we propose a novel framework, MMSG-DTA, which incorporates several innovative modules:

**GraphTrans Module.** To solve the oversmoothing issue in GNNs, we designed the GraphTrans module, which effectively captures both local and global features of molecular graphs, significantly enhancing the representation of molecular structures.

**Enhanced Protein Feature Representation.** Our Model goes beyond traditional approaches that rely solely on sequence information by combining graph-based and sequence-based modalities. This integrated approach allows us to capture complex spatial structures and functional regions of proteins more effectively.

**Attention-Based Feature Fusion.** To optimize feature contribution, we introduced an attention mechanism that dynamically adjusts the weight of each modality. This dynamic fusion improves prediction robustness and accuracy, ensuring that the most relevant features are prioritized.


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

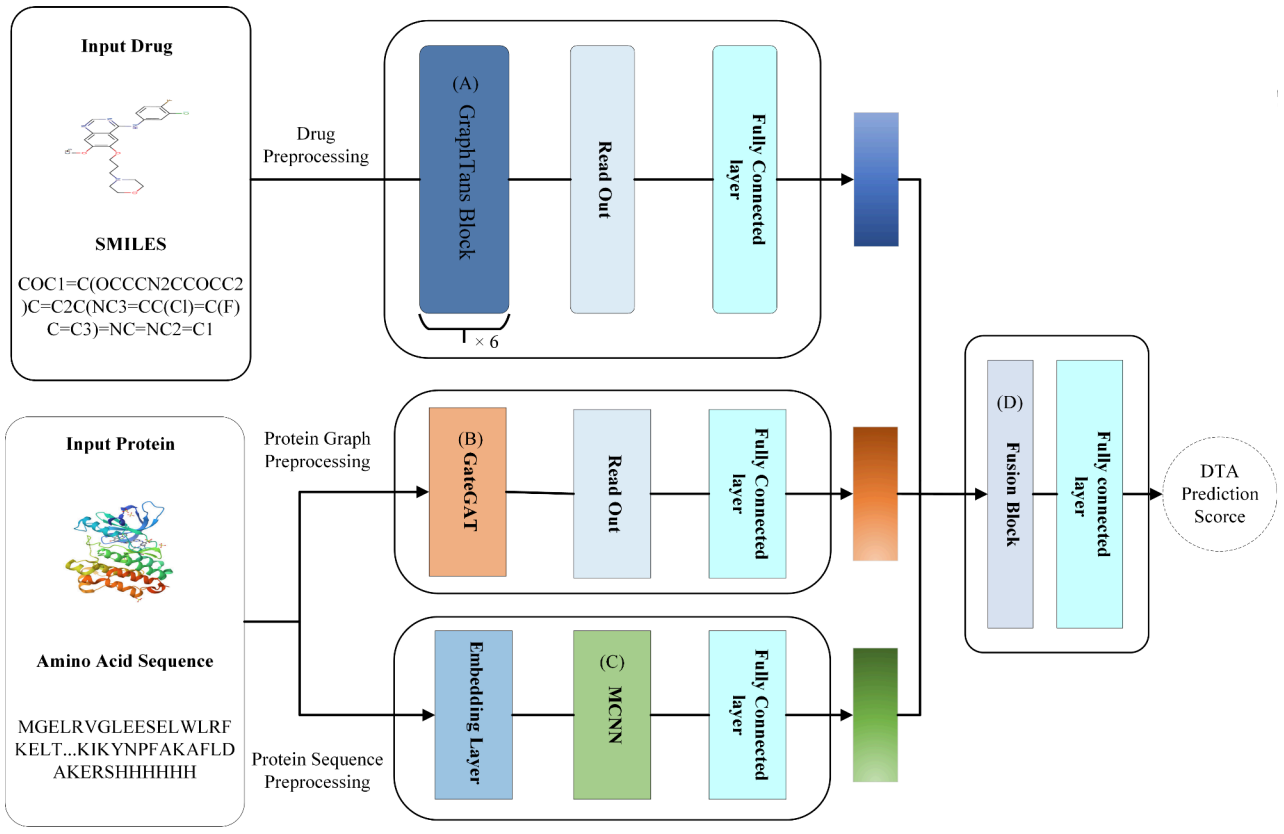



**Figure 1. Overall Architecture of MMSG-DTA.** (A) Structure details of GraphTrans. This architecture is responsible for processing features of the graph modality of small molecules. (B) Structure details of GateGAT. This architecture is responsible for processing features of the protein graph modality. (C) Structure details of MCNN. This architecture is responsible for processing features of the protein sequence modality. (D) The fusion module combines the extracted features using an attention mechanism. The fused features are then fed into a fully connected layer for prediction.

**Comprehensive Benchmark Evaluation.** We evaluated MMSG-DTA on three benchmark data sets (Davis, KIBA, and Metz), and the results demonstrated that our framework outperforms existing state-of-the-art methods, showcasing its potential for real-world applications. Additionally, we conducted cold-start evaluations on the Davis and KIBA data sets, effectively mitigating issues such as data leakage and further enhancing the reliability and generalizability of our model's performance.

## 2. THE PROPOSED METHOD

The proposed method for multimodal multiscale drug-target affinity (DTA) prediction combines both graph and sequence modalities, as depicted in Figure 1. The model consists of three key components: the data preprocessing module, the feature extraction module, and the DTA prediction network. First, the data preprocessing module transforms drug and protein target data into appropriate formats for further processing. Next, the feature extraction module extracts meaningful features from both drug molecules and protein targets. Finally, the DTA prediction network integrates these extracted features and passes them through a fully connected layer to predict the binding affinity.

This section provides a detailed overview of the key components that contribute to the enhanced performance of the proposed DTA prediction model. These components include learning about drug molecule characteristics, learning about protein target characteristics, fusion of these characteristics, and the affinity prediction network.

### 2.1. Feature Learning of Drug Molecules.

The drug feature extraction module is designed to capture the essential structural characteristics of drug molecules that are crucial for understanding their binding affinities with proteins. By incorporation of local topological features through the GINE layer and global attention mechanisms via the Attention layer, the model effectively identifies key structural motifs that influence drug−protein interactions. These features provide valuable insights into how molecular shape, flexibility, and specific functional groups contribute to the binding process.

The molecular input for the model is encoded using the Simplified Molecular Input Line Entry System (SMILES), which employs concise ASCII strings to represent the chemical structures of the compounds. These SMILES strings are processed using RDKit,[22] converting them into graph representations with node features and adjacency matrices. In this representation, the nodes correspond to atoms, while the edges represent chemical bonds. The detailed node features are summarized in Table 1, where each node is characterized by an 88-dimensional feature vector.

Inspired by Graph Neural Networks (GNNs)[23] and the Transformer[24] architecture, we introduce the GraphTrans module. This novel approach fuses local and global feature extraction to improve drug molecule graph representations, as illustrated in Figure 2.

In the first part of the GraphTrans module, local features are extracted using a variant of the Graph Isomorphism Network (GIN) called GINE.[25] The node feature matrix of the drug molecular graph, $\boldsymbol{H}_{\text{input}}^{(l)}$, is updated by aggregating features from


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

**Table 1. Node Features for a Drug Graph**

| Feature Name | Dimension |
|---|---|
| Atom Symbols | 44 |
| Formal Charge | 1 |
| Explicit Valence | 1 |
| Number of Atomic | 1 |
| Hybridization Type | 3 |
| Whether the atom is Donor | 1 |
| Degree of the atom (one-hot) | 11 |
| Number of Radical Electrons | 1 |
| Whether the atom is Acceptor | 1 |
| Whether the atom is Aromatic | 1 |
| Number of Explicit Hydrogens | 1 |
| Total number of H atoms bound to the heavy atom (one-hot) | 11 |
| Number of implicit H atoms bound to the heavy atom (one-hot) | 11 |
| Total | 88 |

neighboring nodes. This aggregation is modulated by a learnable parameter $\epsilon^{(l)}$, allowing for adaptive weighting between the node's own features and those of its neighbors. The update rule for the node feature $h_i'$ at layer $l$ is given by

$$h_i' = F_\Theta^{(l)}\left((1+\epsilon^{(l)})\cdot h_i^{(l)} + \sum_{j\in N(i)} \mathrm{ReLU}(h_j^{(l)} + e_{j,i})\right) \quad (1)$$

Here, $F_\Theta^{(l)}$ is a neural network layer that processes the aggregated features. This step allows the model to capture important local patterns within the drug molecule's graph structure.

To capture the global dependencies within the drug molecule, the GraphTrans module employs a Transformer mechanism. Specifically, the node feature matrix $\boldsymbol{H}_{\text{input}}^{(l)}$ at the $l$-th layer is used to compute the query $(\boldsymbol{Q})$, key $(\boldsymbol{K})$, and value $(\boldsymbol{V})$ matrices via learned weight matrices $\boldsymbol{W}_Q$, $\boldsymbol{W}_K$, and $\boldsymbol{W}_V \in \mathbb{R}^{d\times d_k}$. The attention scores are computed as

$$\boldsymbol{A} = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right) \quad (2)$$

The output $\boldsymbol{H}'$ for each attention head is weighted by the attention scores and computed as

$$\boldsymbol{H}' = \boldsymbol{A}\cdot\boldsymbol{V} \quad (3)$$

These outputs are concatenated and passed through a final weight matrix $\boldsymbol{W}_o$ to produce the multihead attention output $\boldsymbol{H}_{\text{att}}$:

$$\boldsymbol{H}_{\text{att}} = \mathrm{Concat}(\boldsymbol{H}_1', \boldsymbol{H}_2', \ldots, \boldsymbol{H}_h') \quad (4)$$

The global feature representation for the next layer is then computed by adding the multihead attention output to the original input features, followed by Dropout and Layer Normalization:

$$\boldsymbol{H}_{\text{global}} = \mathrm{LayerNorm}[\mathrm{Dropout}(\boldsymbol{H}_{\text{att}} + \boldsymbol{H}_{\text{input}}^{(l)})] \quad (5)$$

After extraction of both local and global features, these two sets of features are combined. In this study, we use the summation operation to merge the local features $\boldsymbol{H}_{\text{local}}$ and global features $\boldsymbol{H}_{\text{global}}$, resulting in the final feature matrix $\boldsymbol{H}_{\text{out}}$:

$$\boldsymbol{H}_{\text{out}} = \boldsymbol{H}_{\text{global}} + \boldsymbol{H}_{\text{local}} \quad (6)$$

To extract the final output vector $\boldsymbol{D}_G$ from the graph, the average of the vertex embeddings from the final layer is computed:

$$\boldsymbol{D}_G = \frac{1}{|V|}\sum_{v\in V} \boldsymbol{H}_{\text{out}}^{L}(v) \quad (7)$$

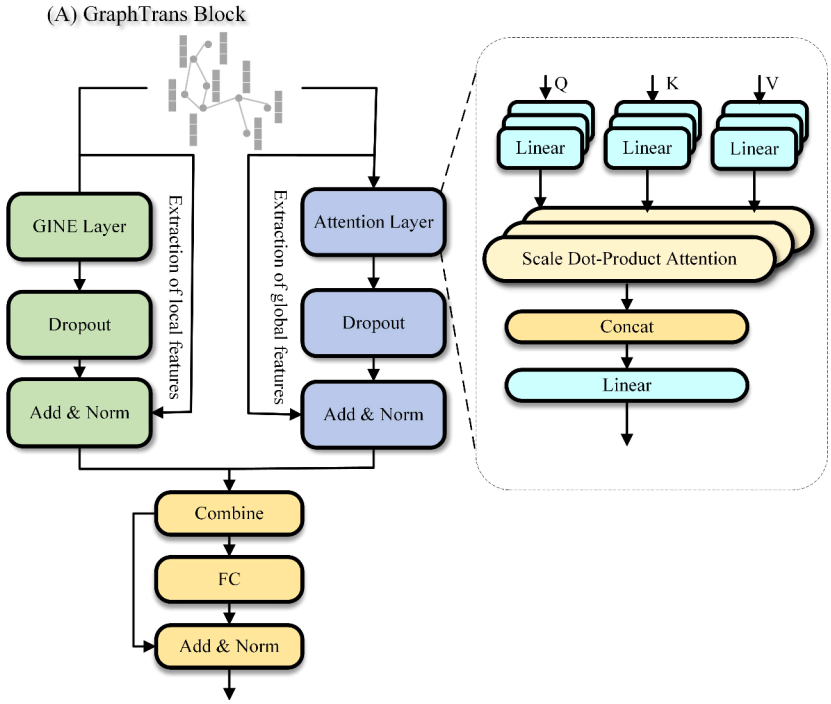



**Figure 2.** GraphTrans module. The green portion represents the extraction of local features, while the purple portion represents the extraction of global features. Finally, these two types of features are fused together.



https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

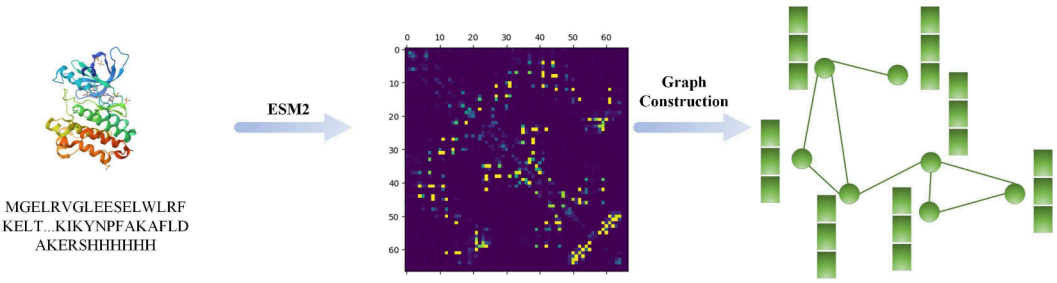



**Figure 3.** Process of Constructing the Protein Graph. Protein sequences are used to construct the protein graph. Due to the complexity of the amino acid sequence of the protein, we have illustrated only a portion of the entire molecular graph for clarity and ease of demonstration.

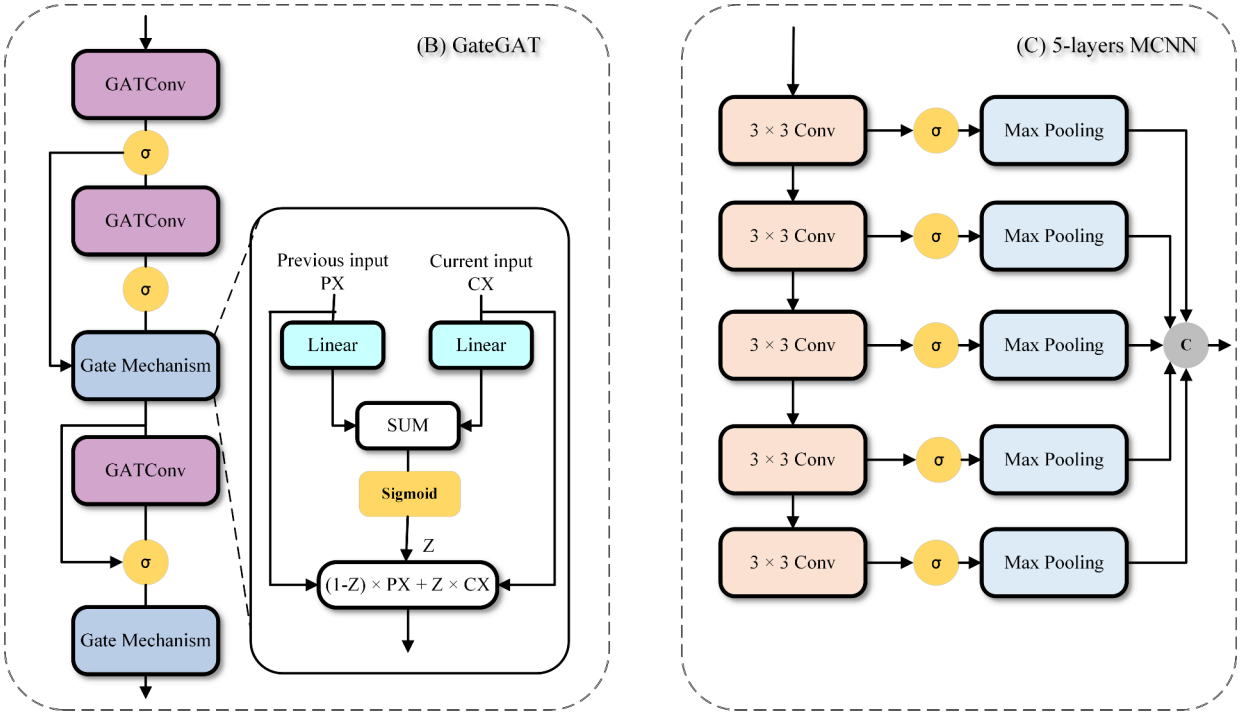



**Figure 4.** Left side illustrates the GateGAT module, which incorporates multiple GATConv layers with a gating mechanism. The right side depicts the MCNN module, designed to generate the feature representation of protein sequences.

In this approach, $D_G$ represents the comprehensive feature vector for the drug molecule, capturing both local and global graph information.

**2.2. Feature Learning for Protein Targets.** For proteins, our model integrates both sequence- and structure-based features. The GATConv module captures the physical and chemical interactions between amino acid residues, such as hydrogen bonding and hydrophobic interactions, which are fundamental to protein−ligand binding. By explicitly modeling these interactions, we can gain a deeper understanding of the key residues involved in drug binding and how protein dynamics may influence the binding process. Additionally, the MCNN module helps capture conserved regions of the protein sequence that may correspond to functional sites, providing further biological insights into the protein's functional architecture.

To enhance our understanding of protein structures and improve the quality of input features, we employed Meta's ESM-2[26] model to predict protein contact maps. ESM-2 is based on the Transformer architecture and leverages the self-aware mechanism to capture long-range dependencies within protein sequences, allowing it to understand and generate complex sequence patterns. The model is trained on a large corpus of protein sequence data from public databases, including UniProt, which encompass millions of protein sequences and provide rich evolutionary information. We chose ESM-2 due to its ability to accurately predict protein contact maps without relying on multiple sequence alignments (MSAs), significantly enhancing the prediction efficiency.

The contact map generated by the ESM-2 model is represented as a probability matrix, where each element corresponds to the interaction probability between different residues, with values ranging from [0,1].

As depicted in Figure 3, the protein graph construction begins by transforming the protein sequence into a weighted graph. Here, residues function as nodes, and interactions between them are represented as edges with edge features derived from probability values. To manage protein sequences longer than 1200 residues, we adopted a segmentation and stitching approach for building the contact map. Specifically, the sequence is divided into subsequences of length $L$,with an overlap stride of $\frac{L}{2}$. Each subsequence's contact map is independently predicted using the ESM-2 model. The resulting



https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

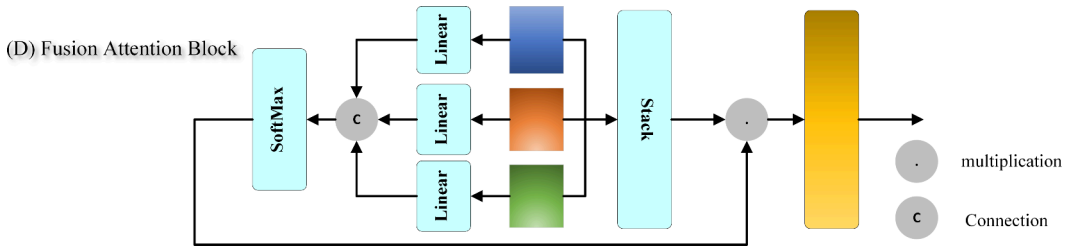



**Figure 5.** Feature fusion module. This process allows for the effective integration of multimodal features, enhancing the model's ability to predict drug−protein interactions.

probability matrices for each segment are then merged with overlapping regions averaged. Furthermore, each residue is represented by an initial 33-dimensional feature vector comprising attributes such as residue type, polarity, hydrophobicity, molecular weight, and group dissociation constant.

For feature learning in protein graphs, we employed the Graph Attention Network (GAT)[27] with a gated skip connection mechanism, as depicted on the right side of Figure 4. GAT dynamically adjusts the weights based on the importance of neighboring nodes, thereby capturing more meaningful neighborhood information. This is particularly crucial in protein graphs as different amino acid residues may contribute differently to the target task. Moreover, protein graphs typically exhibit heterogeneity, where nodes (amino acid residues) and edges (interactions) have different attributes.

In our implementation, each node $i$'s feature vector $X_i$ is first projected into a new space using a learnable weight matrix $W$. The attention coefficient $e_{i,j}$ between node $i$ and its neighboring node $j$ is then calculated as follows:

$$e_{i,j} = \alpha(W \cdot X_i \| W \cdot X_j) \tag{8}$$

Here, the symbol $\|$ represents the concatenation operation, and $\alpha$ is a learnable weight function used to compute the attention coefficient. Subsequently, the softmax function is applied to normalize over all neighboring nodes $j \in N(i)$, yielding the normalized attention coefficient $a_{i,j}$:

$$a_{ij} = \mathrm{softmax}(e_{ij}) = \frac{\exp(\mathrm{LeakyReLU}(e_{ij}))}{\sum_{k \in N(i)} \exp(\mathrm{LeakyReLU}(e_{ik}))} \tag{9}$$

Next, the hidden layer $P_i$ is obtained by the weighted sum of the features of node $i$'s neighboring nodes:

$$P_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \cdot W \cdot X_j\right) \tag{10}$$

To further address the issues of gradient vanishing and node feature degradation when stacking multiple layers of GNNs, we introduce a gating mechanism in the hidden layers. This mechanism integrates feature representations from different layers by controlling the retention and update ratios of the information. Specifically, the node representations in layer $l$ and layer $l + 1$ are updated using the gated skip connection mechanism as follows

$$Z_i = \mathrm{sigmoid}(W_1 \cdot P_i^{(l+1)} + W_2 \cdot P_i^{(l)} + b) \tag{11}$$

where $Z_i$ controls the ratio of update and retention, $W_1$ and $W_2$ are trainable weight matrices, and $b$ is the bias term. The feature of node $i$ is updated as follows:

$$P_i^{(l+1)} = Z_i \cdot P_i^{(l+1)} + (1 - Z_i) \cdot P_i^{(l)} \tag{12}$$

This mechanism adjusts the ratio coefficient $Z_i$, allowing the model to aggregate information from distant nodes while preserving its own features and thus preventing information loss in deep networks.

Unlike the feature extraction module for molecular graphs, in the protein graph feature extraction process, using Max Pooling during the Readout phase yields better results. This is because certain important amino acids may play a dominant role in determining the protein's behavior. Subsequent experiments will demonstrate this point, as well. Therefore, the final feature representation of the protein graph $\boldsymbol{P_G}$ is as follows:

$$\boldsymbol{P_G} = \mathrm{Max}(P_1, P_2, P_3, ..., P_i) \tag{13}$$

Here, $i$ represents the number of nodes in the protein graph.

For feature learning of protein sequences, we first establish a vocabulary, using one-hot encoding to map each amino acid to an integer (for example, alanine (A) is 1, cysteine (C) is 2, aspartic acid (D) is 3, and so on). This gives us an integer representation of the protein sequence.

After obtaining the integer representation of the protein sequence, we passed it through an embedding layer. The function of the embedding layer is to map the integer corresponding to each amino acid to a learnable 128-dimensional vector.

The embedding matrix of the protein sequence is denoted as $S^{(L_p \times E)}$, where $L_P$ represents the length of the protein sequence, and $E$ is the dimensionality of the embedding layer. As depicted on the right side of Figure 5, the multilayer convolutional neural network (MCNN) architecture consists of multiple convolutional and max-pooling layers, designed to extract hierarchical features from the input data.

As it is well-known that one-dimensional convolutional neural networks (1D-CNNs) struggle to capture global information, we designed a network with multiple branches to extract features from protein sequences. Specifically, through experiments, we found that a multiscale convolutional neural network with five branches yielded the best results. The operation is as follows: in the first branch, we use a 3 × 3 one-dimensional convolutional layer followed by a max-pooling layer to capture the local information on the protein sequence. In each subsequent branch, an additional 3× 3 convolutional layer is added to the previous structure to extract further feature information. Finally, the outputs from each branch are concatenated, allowing the network to capture as much of the complete information from the protein sequence as possible. The following formula provides a detailed explanation of the above process.


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

$$P_S = W_s \cdot \left( \sum_{f=1}^{5} Maxpooling(C_f(S)) \right) \tag{14}$$

Here, $C_f$ represents the $f$-th branch, with each branch containing $f$ 3× 3 convolutional layers. The input $S$ is mapped into a matrix $C \in \mathbb{R}^{1200\times h}$, where each convolutional layer is followed by a ReLU activation function. The MaxPooling operation then maps matrix $C$ into a $h$-dimensional vector. $W_S \in \mathbb{R}^{d\times fh}$ is a learnable matrix.

**2.3. Feature Fusion.** Through the aforementioned feature extraction modules, we obtained three different features: drug molecular graph features $D_G$, protein graph features $P_G$, and protein sequence features $P_S$. Next, we focus on how to integrate these multimodal features through the feature fusion module to enhance the accuracy and robustness of the model in predicting drug−protein interactions. As shown in Figure 5, the feature fusion module illustrates the specific process of integrating these diverse features into a unified representation for the prediction task. This process not only deepens the model's understanding of drug−protein interactions but also enables the model to combine complementary information from both drug and protein features, providing a more comprehensive perspective on the crucial relationships between drug properties and protein regions. By doing so, the fused features help pinpoint the most critical drug properties and protein regions for affinity, offering more profound insights for drug-target interaction predictions.

To effectively integrate these multimodal features, we propose an attention mechanism that dynamically assigns weights based on the contribution of each feature type, allowing the model to focus on the most relevant information and improve the accuracy of the drug-target affinity prediction.

For each feature modality, a projection network is applied to map the input features into a common latent space. For input features $h \in \mathbb{R}^d$, the projection network consists of two linear transformation layers and a nonlinear activation layer (Tanh). The formula is as follows

$$F(h) = W_2 \cdot \tanh(W_1 \cdot h + b_1) + b_2 \tag{15}$$

where $h$ represents the input feature vector (either $D_G$, $P_G$, or $P_S$). $W_1 \in \mathbb{R}^{d_{in}\times d_{hidden}}$ and $W_2 \in \mathbb{R}^{d_{hidden}\times 1}$ are two learnable linear transformation weight matrices, and $b_1 \in \mathbb{R}^{d_{hidden}}$ and $b_2 \in \mathbb{R}$ are the bias terms for the two linear transformations. The output of the projection network is a scalar $F(h)$, which represents the importance of the modality feature in the attention mechanism.

The attention mechanism includes three projection networks: $F(D_G)$, $F(P_G)$, and $F(P_S)$, which compute the attention scores for the drug molecular graph, protein graph, and protein sequence features, respectively. After concatenation of the outputs of these networks, the attention weights are generated by applying the softmax function:

$$a = \text{Softmax}([F(D_G), F(P_G), F(P_S)]) \tag{16}$$

Specifically expanded as

$$a_{i\in(D_G,P_G,P_S)} = \frac{\exp(F(i))}{\exp(F(D_G)) + \exp(F(P_G)) + \exp(F(P_S))} \tag{17}$$

These attention weights are used to scale the corresponding feature representations, and then, they are combined. Specifically, we use the attention weights $\alpha$ to perform an element-wise weighted sum of the drug molecular graph, protein graph, and protein sequence features:

$$F_{\text{emb}} = \sum_{i\in\{D_G,P_G,P_S\}} a_i \cdot F(i) \tag{18}$$

The fused feature embeddings are sequentially passed through two fully connected layers combined with a nonlinear activation function and a Dropout layer to produce the final output. This fusion strategy allows the model to adaptively adjust the prioritization of different modality features based on their contribution, ensuring that the most informative features are retained, thereby improving the prediction of drug-target interactions.

The attention-based feature fusion not only enhances the model's flexibility in handling diverse data modalities but also improves the robustness of predictions by focusing on the most relevant information from each input source.

## 3. EXPERIMENT

**3.1. Data Preparation.** In this study, to comprehensively evaluate the performance of our proposed method MMSG-DTA, we used three publicly available drug-target affinity benchmark data sets: the Davis data set,[28] the KIBA data set,[29] and the Metz data set.[30]

The Davis data set uses the negative logarithm of the dissociation constant ($K_d$) as the affinity value, and the logarithmic transformation process is represented in Equation 21. A higher $pK_D$ value indicates a stronger binding affinity. This data set contains 30,056 affinity values between 68 drugs and 442 targets, with an affinity range from 5.0 to 10.8.

$$pK_d = -\log_{10}\left(\frac{K_d}{10^9}\right) \tag{19}$$

The Metz data set also uses $pK_d$ to score binding affinity. It contains 35,259 binding affinity scores between 170 drugs and 1,423 targets, with a range from 4 to 11.1.

The KIBA data set integrates information from inhibition constant ($K_i$), dissociation constant ($K_d$), and half-maximal inhibitory concentration ($IC_{50}$) to score drug-target affinity. Following the preprocessing method described by He et al.,[14] the data set contains 118,254 binding affinity scores between 2,111 drugs and 229 targets, with values ranging from 0.0 to 17.2. Table 2 summarizes the statistics of all data sets.

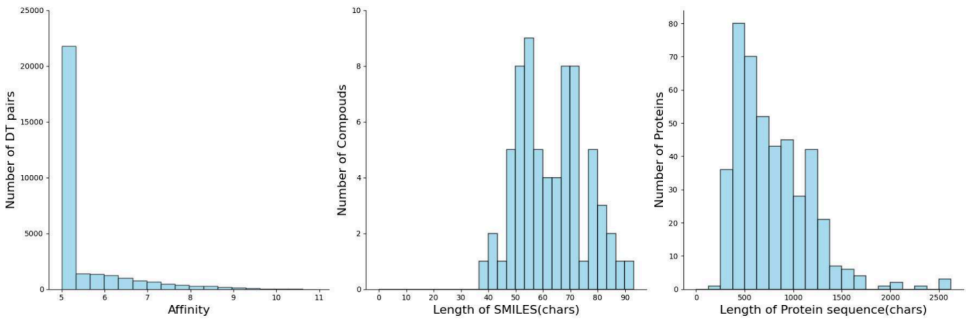
Figure 6 illustrates the distribution ranges of affinity values, drug SMILES lengths, and protein amino acid sequence lengths across the three benchmark data sets. The experimental

**Table 2. Summary of the Data Sets**

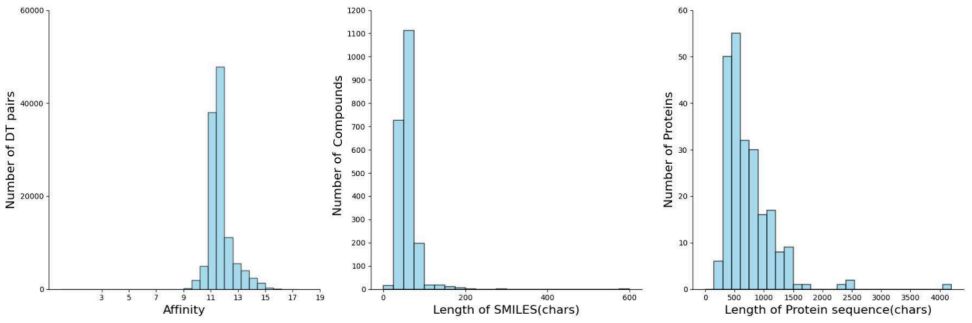
| | Davis | KIBA | Metz |
|---|---|---|---|
| No. of drugs | 68 | 2111 | 170 |
| No. of proteins | 442 | 229 | 1423 |
| No. of binding affinities | 30,056 | 118,254 | 35,259 |
| Maximum length of drugs | 103 | 590 | 112 |
| Maximum length of proteins | 2549 | 4128 | 2527 |
| The average length of drugs | 64 | 58 | 46 |
| The average length of proteins | 788 | 728 | 745 |
| Affinity Measures | $pK_d$ | KIBA score | $pK_d$ |
| Range of affinities | 5.0−10.8 | 0.0−17.2 | 4.0−11.1 |



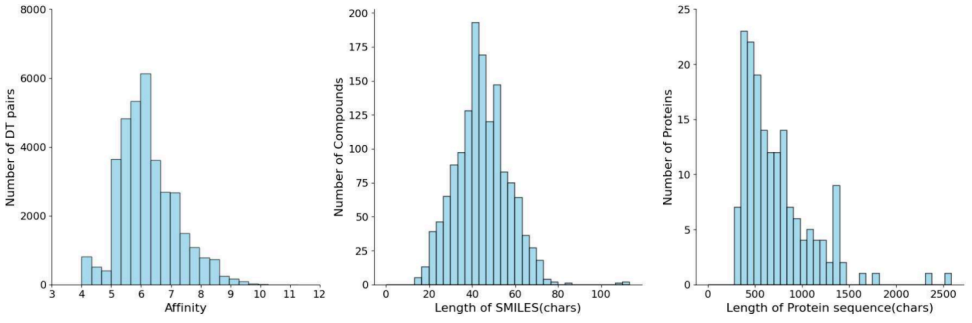
https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996




(a) Davis




(b) KIBA




(c) Metz

**Figure 6.** Distribution of binding affinity values, distribution of SMILES string lengths, and distribution of protein sequence lengths for the data sets. The top represents the Davis data set, the middle represents the KIBA data set, and the bottom represents the Metz data set.

results show that, for most drugs in these data sets, the SMILES lengths are less than 100, and the amino acid sequence lengths of the protein targets are less than 1,500. In the Davis data set, the majority of affinity values is concentrated around 5, indicating the very low affinity of this data set. The affinity scores in the KIBA and Metz data sets are centered in the middle, following a normal distribution.

**3.2. Evaluation Metrics.** Drug-target affinity (DTA) prediction is a regression task, and it uses the mean squared error (MSE), a commonly used loss function in regression tasks, as the loss function for this task. MSE is calculated based on the error between the true values and the predicted values. A smaller MSE indicates that the predicted values are closer to the true values. The MSE is defined as follows

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2 \tag{20}$$

where $y_i$ represents the true value of the $i$-th sample, and $p_i$ is the predicted value of the $i$-th sample.

Another evaluation metric is the Concordance Index (CI), which measures whether the predicted values of two randomly selected drug-target pairs have the same relative order as the true values in the data set. A higher CI indicates a better predictive performance of the model. The definition is shown in Equation 21

$$\mathrm{CI} = \frac{1}{Z}\sum_{y_i > y_j} h(p_i - p_j) \tag{21}$$

where $p_i$ is the predicted value corresponding to the true affinity value $y_i$, which is greater, and $p_j$ is the predicted value corresponding to the true affinity value $y_j$, which is smaller. $h(x)$ is the step function. $Z$ is the normalization constant that maps the values to the range [0,1]. The definition of the step function is shown in eq 22.

$$h(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases} \tag{22}$$



https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

Regression toward the mean ($r_m^2$) is used to evaluate the external predictive potential of quantitative structure−activity relationship (QSAR) models. It indicates the degree to which the predictions approach the mean in subsequent measurements. A model with a higher $r_m^2$ value for the test set is considered acceptable. The calculation of $r_m^2$ is defined as shown in eq 23

$$r_m^2 = r^2 \times (1 - \sqrt{r^2 - r_0^2}) \quad (23)$$

where $r$ is the correlation coefficient with an intercept, and $r_0$ is the correlation coefficient without an intercept.

**3.3. Experimental Setup.** Our MMSG-DTA model is implemented using the open-source deep learning framework PyTorch, with the GNN component built using PyTorch Geometric (PyG) and the GraphGPS layers[31] for drug-target interaction prediction. We evaluated the performance of the proposed model on the Davis, KIBA, and Metz benchmark data sets. Table 3 presents the experimental hyperparameter settings used in our study, including the number of epochs, batch size, optimizer settings, model architecture (such as the number of layers and units in the GraphTrans, GateGAT, and MCNN components), and learning rate.

**Table 3. Experimental Hyperparameter Settings**[a]

| Parameters | Setting |
|---|---|
| Epoch | 2000 |
| Batch Size | 512 |
| Optimizer | Adam |
| Learning Rate | 5e-4 |
| Drought Rate | 0.2 |
| GraphTrans Layers | 6 |
| Dimension of the GraphTrans | N, 128, 128, 128, 128, 128, 128 |
| GateGAT Layers | 3 |
| Dimension of the GateGAT | N, 4N, 4N |
| MCNN Layers | 5 |
| Dimension of the MCNN | N, 2N, 3N, 4N, 5N |
| Fully connected layer hidden unit | 1024, 512 |

[a]N represents the dimension of the initial features.

To ensure fair and accurate hyperparameter tuning, we employed a dedicated validation set for selecting the optimal hyperparameters. The values listed in Table 3 were selected after extensive experimentation with the validation set, ensuring that the model performs optimally without overfitting or introducing data leakage. Importantly, no reference to the test set was made during the hyperparameter tuning process, thus preventing any data leakage.

The hyperparameters were tuned using cross-validation, where the training set was used to train the model, and the validation set was used solely for hyperparameter selection. This approach ensured that the test set remained untouched throughout the entire training and tuning process, providing an unbiased evaluation of the model's generalization ability. The final hyperparameter values were chosen based on the best performance on the validation set and were subsequently used for evaluation on the test set.

**3.4. Performance Comparison with Benchmark Models.** In the experiments, a 5-fold cross-validation (CV) strategy was employed to evaluate the performance of various models. For each fold, all methods shared the same training, validation, and test sets. The entire data set was randomly split into six parts, with five subsets used for the cross-validation process, and the remaining subset serving as the independent test data set.

For the Davis and KIBA data sets, we compared our model with widely used DTA prediction benchmark methods, including DeepDTA,[15] DeepCDA,[17] GraphDTA,[19] MGraphDTA,[20] TDGraphDTA,[32] AttentionMGT-DTA,[33] CCL-DTI,[34] TransVAEDTA,[35] MDCT-DTA,[36] and MDCT-DTA.[37] The Metz data set has received less attention in the academic community compared to other more commonly used DTA data sets, such as Davis and KIBA, resulting in fewer benchmark methods and related studies. Therefore, for the Metz data set, we have included ML-DTI[38] and GPCNDTA[39] in addition to the original benchmark methods for comparison. To ensure a fair comparison, we used the same training, validation, and test sets along with the same performance metrics for evaluation. Tables 4 to 6 summarize the results of our model alongside those reported in the original publications of the baseline methods.

The experimental results demonstrate that our proposed MMSG-DTA model achieves excellent performance across all three data sets, namely, Davis, KIBA, and Metz, confirming its robustness and generalization ability. In nearly all cases, MMSG-DTA outperformed the other baseline models, showcasing its superior predictive capabilities in drug-target interaction (DTA) prediction.

On the Davis data set, MMSG-DTA achieved the best performance with the lowest Mean Squared Error (MSE) of 0.193 (0.002), indicating superior prediction accuracy compared to the other models. The Concordance Index

**Table 4. Comparison of the Performance of MMSG-DTA and Previous Classic Models on the Davis Data Set**[a]

| Model | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|
| DeepDTA(2018) | 0.261 (0.007) | 0.878 (0.002) | 0.63 (0.015) |
| DeepCDA(2020) | 0.248 (0.002) | 0.889 (0.002) | 0.682 (0.008) |
| GraphDTA(2021) | 0.204 (0.005) | 0.883 (0.002) | 0.726 (0.016) |
| MGraphDTA(2022) | 0.207 (0.001) | 0.900 (0.004) | 0.714(0.005) |
| TDGraphDTA(2023) | 0.199 (0.005) | 0.906 (0.001) | 0.722(0.004) |
| AttentionMGT-DTA(2024) | 0.194 (0.004) | 0.891 (0.006) | 0.699 (0.027) |
| CCL-DTI(2024) | | 0.874 (0.003) | |
| TransVAEDTA(2024) | 0.332 (0.003) | 0.869 (0.008) | 0.571 (0.001) |
| GRA-DTA(2024) | 0.225 (0.005) | 0.897 (0.011) | 0.715 (0.004) |
| MDCT-DTA(2024) | 0.197 (0.006) | 0.908 (0.005) | 0.723 (0.008) |
| MMSG-DTA(Ours) | 0.193 (0.002) | 0.914 (0.007) | 0.763 (0.003) |

[a]The best results are highlighted in bold.



https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

Table 5. Comparison of the Performance of MMSG-DTA and Previous Classic Models on the KIBA Data Set[a]

| Model | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|
| DeepDTA(2018) | 0.194 (0.008) | 0.863 (0.005) | 0.673 (0.019) |
| DeepCDA(2020) | 0.176 (0.006) | 0.889 (0.002) | 0.649 (0.009) |
| GraphDTA(2021) | 0.139 (0.008) | 0.891 (0.001) | 0.725 (0.018) |
| MGraphDTA(2022) | 0.128 (0.001) | 0.902 (0.001) | 0.801 (0.001) |
| TDGraphDTA(2023) | 0.121 (0.006) | 0.899 (0.003) | 0.807 (0.003) |
| AttentionMGT-DTA(2023) | 0.140 (0.005) | 0.893 (0.001) | 0.786 (0.018) |
| CCL-DTI(2024) | | 0.882 (0.005) | |
| TransVAEDTA(2024) | 0.253 (0.003) | 0.822 (0.002) | 0.632 (0.001) |
| GRA-DTA(2024) | 0.142 (0.005) | 0.890 (0.011) | 0.784 (0.004) |
| MDCT-DTA(2024) | 0.130 (0.006) | 0.902 (0.005) | 0.792 (0.008) |
| MMSG-DTA(Ours) | 0.123 (0.002) | 0.911 (0.005) | 0.818 (0.003) |

[a]The best results are highlighted in bold.

Table 6. Comparison of the Performance of MMSG-DTA and Previous Classic Models on the Metz Data Set[a]

| Model | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|
| DeepDTA(2018) | 0.286 (0.004) | 0.815 (0.001) | 0.668 (0.003) |
| GraphDTA(2021) | 0.282 (0.005) | 0.815 (0.002) | 0.669 (0.008) |
| ML-DTI(2021) | 0.322 (0.003) | 0.799 (0.002) | 0.613 (0.015) |
| MGraphDTA(2022) | 0.265 (0.002) | 0.822 (0.001) | 0.701 (0.001) |
| TDGraphDTA(2023) | 0.264 (0.005) | 0.824 (0.005) | 0.708 (0.003) |
| GPCNDTA(2023) | 0.248 (0.005) | 0.834 (0.001) | 0.686 (0.001) |
| MDCT-DTA(2024) | 0.278 (0.003) | 0.824 (0.005) | 0.701 (0.008) |
| MMSG-DTA(Ours) | 0.254 (0.002) | 0.826 (0.004) | 0.717 (0.003) |

[a]The best results are highlighted in bold.

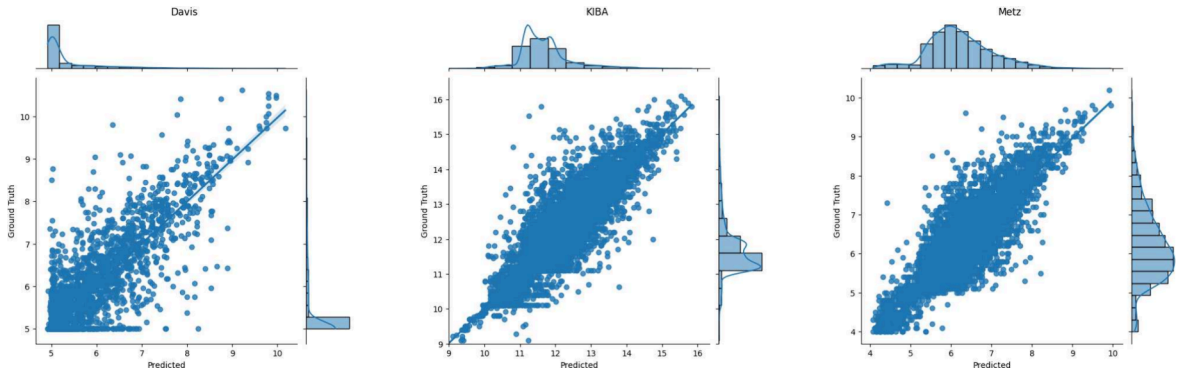



**Figure 7.** Scatter plots of true values versus predicted values on the Davis data set (left), KIBA data set (middle), and Metz data set (right). The horizontal axis represents the predicted binding affinity, and the vertical axis represents the true binding affinity. The bar charts on the top and right show the distribution of sample counts.

(CI) was 0.911 (0.005), which is the highest among all methods, underscoring the model's ability to accurately rank drug-target pairs. Additionally, MMSG-DTA achieved a robust $r_m^2$ of 0.763 (0.003), reflecting the model's effectiveness in capturing the underlying relationships between drugs and targets, which is critical for high-quality DTA prediction.

Similarly, on the KIBA data set, MMSG-DTA once again led the performance metrics with an impressive MSE of 0.123 (0.002), substantially lower than that of the second-best model, MGraphDTA (2022), at 0.128 (0.001). The model also demonstrated a superior CI of 0.914 (0.004), indicating an enhanced ability to distinguish between active and inactive drug-target interactions. Furthermore, MMSG-DTA achieved the highest $r_m^2$ value of 0.818 (0.003), surpassing other models like MGraphDTA (2022), which had an $r_m^2$ of 0.801 (0.001), thus confirming its robustness in capturing complex relationships in DTA.

On the Metz data set, MMSG-DTA achieved a competitive performance, with an MSE of 0.254 (0.002), which, although not the absolute lowest, still places it ahead of several methods such as ML-DTI (2021) with 0.322 (0.003). The CI score of 0.831 (0.004) and the $r_m^2$ of 0.717 (0.003) also demonstrate that MMSG-DTA performs well in terms of distinguishing between different drug-target interactions and capturing the underlying predictive patterns. While GPCNDTA (2023) outperforms MMSG-DTA slightly in CI and $r_m^2$, the overall performance of MMSG-DTA on the Metz data set remains highly competitive.

We present the results in the form of a scatter plot, as shown in Figure 7, where the *X*-axis represents the predicted values from the model and the *Y*-axis represents the true affinity values. The antidiagonal indicates the line where the predicted values exactly match the true values. Data points closer to this line indicate higher prediction accuracy.


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

Table 7. Performance Evaluation on More Realistic Settings of Davis Data Sets[a]

| Scenario | Method | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|---|
| Drug | GraphDTA(2021) | 0.920 (0.029) | 0.678 (0.036) | 0.160 (0.019) |
| | MGraphDTA(2022) | 0.563 (0.065) | 0.729 (0.022) | 0.192 (0.021) |
| | NHGNN-DTA(2023) | 0.554 (0.091) | 0.752 (0.017) | 0.207 (0.030) |
| | MT-DTA(2023) | 0.433 (0.018) | 0.743 (0.010) | 0.158 (0.015) |
| | GRA-DTA(2024) | 0.578 (0.005) | 0.725 (0.015) | 0.121 (0.024) |
| | MMSG-DTA(Ours) | 0.448 (0.087) | 0.765 (0.017) | 0.319 (0.032) |
| Target | GraphDTA(2021) | 0.510 (0.086) | 0.729 (0.012) | 0.154 (0.014) |
| | MGraphDTA(2022) | 0.411 (0.006) | 0.831 (0.012) | 0.436 (0.028) |
| | NHGNN-DTA(2023) | 0.344 (0.029) | 0.855 (0.016) | 0.479 (0.021) |
| | MT-DTA(2023) | 0.384 (0.005) | 0.844 (0.003) | 0.452 (0.024) |
| | GRA-DTA(2024) | 0.376 (0.006) | 0.827 (0.015) | 0.453 (0.024) |
| | MMSG-DTA(Ours) | 0.316 (0.011) | 0.868 (0.018) | 0.533 (0.031) |
| ALL | GraphDTA(2021) | 0.968 (0.096) | 0.579 (0.017) | 0.026 (0.016) |
| | MGraphDTA(2022) | 0.874 (0.090) | 0.636 (0.021) | 0.071 (0.051) |
| | NHGNN-DT(2023) | 0.857 (0.096) | 0.665 (0.038) | 0.087 (0.051) |
| | GRA-DTA(2024) | 0.560 (0.007) | 0.690 (0.012) | 0.101 (0.024) |
| | MMSG-DTA(Ours) | 0.499 (0.076) | 0.661 (0.016) | 0.171 (0.033) |

[a]The best results are highlighted in bold.

Table 8. Performance Evaluation on More Realistic Settings of KIBA Data Sets[a]

| Scenario | Method | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|---|
| Drug | GraphDTA(2021) | 0.442 (0.012) | 0.728 (0.008) | 0.355 (0.022) |
| | MGraphDTA(2022) | 0.421 (0.009) | 0.746 (0.002) | 0.366 (0.016) |
| | NHGNN-DTA(2023) | 0.385 (0.029) | 0.756 (0.007) | 0.400 (0.015) |
| | MT-DTA(2023) | 0.408 (0.007) | 0.752 (0.008) | 0.381 (0.024) |
| | GRA-DTA(2024) | 0.337 (0.007) | 0.765 (0.012) | 0.466 (0.024) |
| | MMSG-DTA(Ours) | 0.336 (0.015) | 0.780(0.004) | 0.454(0.022) |
| Target | GraphDTA(2021) | 0.519 (0.045) | 0.658 (0.050) | 0.314 (0.057) |
| | MGraphDTA(2022) | 0.435 (0.055) | 0.674 (0.028) | 0.382 (0.047) |
| | NHGNN-DTA(2023) | 0.382 (0.071) | 0.732(0.041) | 0.452 (0.054) |
| | MT-DTA(2024) | 0.390 (0.013) | 0.751 (0.009) | 0.406 (0.024) |
| | GRA-DTA(2024) | 0.435 (0.007) | 0.692 (0.012) | 0.355 (0.024) |
| | MMSG(Ours) | 0.382 (0.041) | 0.752 (0.022) | 0.441 (0.045) |
| ALL | GraphDTA(2021) | 0.676 (0.113) | 0.601 (0.030) | 0.149 (0.067) |
| | MGraphDTA(2022) | 0.590 (0.094) | 0.626 (0.028) | 0.182 (0.012) |
| | NHGNN-DTA(2023) | 0.565 (0.094) | 0.649 (0.037) | 0.218 (0.047) |
| | GRA-DTA(2024) | 0.441 (0.007) | 0.657 (0.012) | 0.228 (0.024) |
| | MMSG-DTA(Ours) | 0.425 (0.072) | 0.667 (0.021) | 0.248 (0.014) |

[a]The best results are highlighted in bold.

In summary, MMSG-DTA exhibits outstanding generalization ability across different data sets. It consistently outperforms the majority of baseline models and is comparable to the best methods, surpassing all others in multiple evaluation metrics such as MSE, CI, and $r_m^2$. These results highlight the model's effectiveness and robustness in predicting drug-target interactions, making it a promising tool for DTA prediction tasks.

**3.5. Performance Evaluation on More Realistic Settings.** Previously, experiments typically relied on random splits to divide the data set into training, validation, and test sets. However, such random splitting methods can lead to overly optimiztic results, as they may cause leakage of drug and protein information into the test set.[40] For the purpose of drug discovery, models must generalize to drug−protein (DP) pairs that have never been encountered during training. Therefore, we propose a novel data splitting strategy to more accurately evaluate the performance of the DTA models. In our approach, we first partition the data set based on distinct drugs, ensuring that drugs in the test set do not appear in either the training or validation sets, and the training and test sets are mutually exclusive. Similarly, we perform data splitting according to different proteins. Finally, we introduce an extreme case in which neither the drugs nor the proteins in the test set appear in the training or validation sets. These three distinct data splitting strategies provide a better reflection of the generalizability of our DTA model, aligning with the practical requirements of novel drug development and effectively mitigating data leakage.

We evaluate the performance of MMSG-DTA and several previous state-of-the-art (SOTA) methods on the Davis and KIBA data sets, under scenarios designed to avoid data leakage. One such method is MT-DTA,[41] which is a drug-target interaction prediction model based on an associative learning mechanism. This method aims to enhance prediction accuracy and robustness by capturing the associations between drugs and targets. Another method is NHGNN-DTA,[42] a drug-target interaction prediction model based on a heterogeneous graph


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

Table 9. Comparison of Ablation Experiments on the Davis and KIBA Data Sets[a]d

| Data set | Variant | MSE ↓ | CI ↑ | $r_m^2$ ↑ |
|---|---|---|---|---|
| Davis | MMSG_no_att | 0.207 (0.018) | 0.903 (0.013) | 0.728 (0.014) |
| | MMSG_no_protein_Graph | 0.203 (0.021) | 0.905 (0.010) | 0.722 (0.017) |
| | MMSG_no_protein_Sequence | 0.205 (0.019) | 0.903 (0.015) | 0.737 (0.012) |
| | MMSG_no_drug_local | 0.197 (0.015) | 0.904 (0.008) | 0.710 (0.010) |
| | MMSG_no_drug_global | 0.212 (0.023) | 0.892 (0.018) | 0.705 (0.016) |
| | MMSG-DTA | 0.193 (0.002) | 0.914 (0.009) | 0.763 (0.003) |
| KIBA | MMSG_no_att | 0.127 (0.010) | 0.905 (0.009) | 0.783 (0.015) |
| | MMSG_no_protein_Graph | 0.124 (0.009) | 0.901 (0.011) | 0.793 (0.013) |
| | MMSG_no_protein_Sequence | 0.130 (0.014) | 0.903 (0.012) | 0.809 (0.009) |
| | MMSG_no_drug_local | 0.128 (0.011) | 0.884 (0.017) | 0.772 (0.020) |
| | MMSG_no_drug_global | 0.134 (0.012) | 0.891 (0.014) | 0.796 (0.016) |
| | MMSG-DTA | 0.123 (0.002) | 0.911 (0.005) | 0.818 (0.003) |

[a]The best results are highlighted in bol.

neural network that integrates multimodal information from both drugs and targets to enhance prediction accuracy and robustness. To ensure a fair comparison, all methods are evaluated using the same data splits. The experimental results for the cold-start scenarios are presented in Tables 7 and 8.

In the Davis data set, the proposed MMSG-DTA model demonstrates notable performance improvements across all three cold-start settings, achieving 8% improvement in MSE, 6% improvement in CI, and 4% improvement in $r_m^2$. Across all three cold-start settings, MMSG-DTA consistently delivers superior performance, improving all evaluation metrics by significant margins (P ⟨0.05).

Similarly, on the KIBA data set, MMSG-DTA demonstrates impressive performance enhancements across the cold-start scenarios, with an average improvement of 12% in MSE, 7% in CI, and 8% in $r_m^2$. In the ALL scenario, MMSG-DTA achieves the best overall performance with an MSE of 0.425 (0.072), CI of 0.667 (0.021), and $r_m^2$ of 0.248 (0.014), outperforming GRA-DTA(2024) with an MSE of 0.441 (0.007) and $r_m^2$ of 0.228 (0.024).

In both the Davis and KIBA data sets, MMSG-DTA demonstrates significant improvements in all evaluation metrics across the three cold-start settings (P ⟨0.05). The model's ability to effectively capture shared features of both drugs and proteins—by jointly learning local and global features of drugs, as well as the amino acid sequences and graph structural features of proteins—plays a key role in its superior performance. This innovative feature learning framework is likely the main contributing factor to MMSG-DTA's enhanced generalization ability, making it particularly effective in cold-start scenarios. This performance advantage highlights the model's robustness and its ability to handle the complexities of drug-target interaction prediction.

**3.6. Ablation Experiments.** To better understand the underlying factors that contribute to the performance of our proposed MMSG-DTA model in drug-target affinity (DTA) prediction, we performed a series of ablation experiments. The goal of these experiments is to systematically assess the role of each component in our model, isolating their contributions and identifying the critical mechanisms that drive the overall predictive accuracy. By modifying or removing specific parts of the architecture, we can gain insights into how different modalities and feature extraction techniques influence the performance of the model.

We evaluated several variants of the MMSG-DTA model on two widely used benchmark data sets, Davis and KIBA, to investigate the significance of individual components. These ablation variants allow us to study the contributions of key elements, such as the attention-based feature fusion module, the protein graph modality, and both local and global drug features. Each variant highlights a specific aspect of the model, providing a comprehensive understanding of its inner workings and revealing potential areas for optimization.

- **MMSG_no_att**: In this variant, the attention mechanism within the feature fusion module was removed, and the three learned features were directly concatenated to predict DTA. This experiment evaluates the significance of attention-based feature fusion in enhancing prediction accuracy by effectively weighting different features.
- **MMSG_no_protein_Graph**: In this variant, the protein graph modality, relying solely on the drug graph modality and the protein sequence modality for feature learning. It aims to assess the contribution of the protein graph representation in capturing spatial and structural information essential for DTA prediction.
- **MMSG_no_protein_Sequence**: In this variant, the protein sequence modality is removed, utilizing only the drug graph modality and the protein graph modality for feature extraction. This experiment evaluates the importance of sequence-based information in complementing the protein's graph structure to improve DTA prediction.
- **MMSG_no_drug_local**: In this variant, the learning of local features within the drug molecular graph modality. By removing the local structural information of the drug, we assess the role of local-level representation in the overall prediction performance of the model.
- **MMSG_no_drug_global**: In this variant, the global features of the drug molecular graph are excluded during feature learning. This experiment helps determine the contribution of global structural information of the drug in predicting DTA.

The results of the ablation experiments are listed in Table 9. Based on these results, we summarize the following:

The complete MMSG-DTA model demonstrates the best performance on both the Davis and KIBA data sets. After removal of the local or global features of the drug, the model's performance declines significantly, indicating that the multi-level representation of the drug is crucial for predicting drug−protein interactions. When the protein sequence or graph structure information is removed, the model's performance


https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

fluctuates slightly across different data sets, suggesting that the contribution of protein features varies between data sets. Additionally, the model's performance decreases when the attention mechanism is removed, highlighting the critical role of the attention mechanism in enhancing feature capture and information fusion.

**3.7. Case Study.** To further assess the generalization ability of our model, we conducted an experiment by randomly selecting a set of FDA-approved candidate drugs from the DrugBank[43] database, none of which are included in the KIBA data set. A total of 300 candidate drugs were retained. Subsequently, we focused on epidermal growth factor receptor (EGFR),[44] a key target in cancer therapy, as the protein of interest. Among the 300 drugs, 10 are known to interact with EGFR, while another 10 drugs do not exhibit any interaction with EGFR. The amino acid sequence of EGFR, obtained from UniProt, was combined with the 300 drugs to form drug-target pairs, which were then input into the MMSG-DTA model trained on the larger KIBA data set for prediction.

Table 10 presents the top 10 drugs as well as the bottom 5 drugs ranked based on their predicted EGFR affinity. Among the 10 drugs with the highest predicted affinity for EGFR according to the MMSG-DTA model, 7 are known to interact with EGFR. The remaining 3 drugs, while not confirmed as EGFR inhibitors, are tyrosine kinase inhibitors commonly used in cancer-targeted therapies. Given that EGFR belongs to the tyrosine kinase family, these drugs may potentially bind to EGFR and warrant further experimental validation.

**Table 10. Predicted KIBA Score Ranking of Drug Candidates with EGFR[a]**

| RANK | ID | Durg Name | Predict Score |
|---|---|---|---|
| 1 | DB05424 | Canertinib | 13.443 |
| 2 | DB08916 | **Afatinib** | 13.407 |
| 3 | DB01259 | **Lapatinib** | 13.395 |
| 4 | DB13164 | Olmutinib | 13.204 |
| 5 | DB00317 | **Gefitinib** | 13.111 |
| 6 | DB00530 | **Erlotinib** | 13.021 |
| 7 | DB05524 | **Pelitinib** | 12.784 |
| 8 | DB09053 | Ibrutinib | 12.460 |
| 9 | DB05294 | **Vandetanib** | 12.441 |
| 10 | DB12267 | **Brigatinib** | 12.366 |
| | | ... | |
| 296 | DB00945 | Aspirin | 6.442 |
| 297 | DB00196 | Fluconazole | 6.247 |
| 298 | DB00811 | Ribavirin | 5.928 |
| 299 | DB01050 | Ibuprofen | 5.733 |
| 300 | DB00608 | Chloroquine | 5.540 |

[a]Bold font in the table indicates drugs known to interact with EGFR.

Among the bottom 5 drugs predicted by the model, Fluconazole is used to treat fungal infections, particularly in immunocompromised patients. Ribavirin is employed in the treatment of hepatitis (especially in combination with interferon) and certain respiratory viral infections. Aspirin and Ibuprofen are commonly used for pain relief, fever reduction, and alleviation of inflammation in arthritis. Chloroquine is used to treat malaria, rheumatoid arthritis, and systemic lupus erythematosus, among other immune-related conditions. These drugs are not directly associated with EGFR-targeted cancer therapies.

Although the model predicts some drugs with no direct association with EGFR-targeted therapies, further validation of EGFR-drug interactions was carried out through molecular docking experiments. To further validate the accuracy of the MMSG-DTA model in predicting drug-target affinity, we downloaded the crystal structure of EGFR with PDB ID 6ZU8 (UniProt ID: P00533) from the Protein Data Bank (PDB). Subsequently, we performed molecular docking experiments using Autodock[45] software and identified the candidate binding sites of the specific ligand and receptor based on the lowest binding free energy values. To further analyze the interactions between the drug molecule and protein, we visualized the molecular docking results using PyMOL software, highlighting the hydrogen bonds formed between the drug molecule and the EGFR amino acid residues, as shown in Figure 8.

## 4. CONCLUSIONS

In this study, we proposed the MMSG-DTA method for predicting drug-target affinity (DTA), which integrates multiple extraction techniques for drug molecules and protein targets. Our model combines the GraphTrans module for drug molecular graphs, the GateGAT module for protein graphs, and the MCNN module for protein sequences. By leveraging an attention mechanism in the feature fusion module, we effectively integrated these different types of features to enhance the prediction accuracy.

The experimental results demonstrate that our proposed MMSG-DTA model significantly outperforms existing methods in predicting drug-target affinity (DTA), as evidenced by the substantial improvements in mean squared error (MSE) and concordance index (CI) scores on the benchmark data sets, such as Davis, KIBA, and Metz. These improvements are particularly pronounced in cold-start settings, where the model is evaluated on previously unseen drug-target pairs, simulating real-world scenarios in drug discovery and repurposing. In these settings, MMSG-DTA consistently showed a marked reduction in MSE (8% on Davis, 12% on KIBA), and improvements in CI (6% on Davis, 7% on KIBA), highlighting its robustness and generalizability.

In conclusion, the MMSG-DTA method demonstrates exceptional performance in DTA prediction, offering a robust and generalizable approach for both drug discovery and repurposing. The model's ability to integrate multimodal feature representations from both drug and protein structures, combined with a novel data partitioning strategy, ensures its reliability and effectiveness in real-world scenarios, particularly in the face of cold-start challenges. Future work will focus on incorporating additional ligand feature representations, exploring pretraining and fine-tuning strategies and expanding the model's applicability to broader drug discovery contexts.

However, while our model demonstrates promising results, several areas for improvement remain. In particular, the incorporation of additional compound features could further enhance its performance. These could include drug-specific features, such as both one-dimensional and two-dimensional drug structures and protein-specific features, including one-dimensional sequences, two-dimensional graph structures, and three-dimensional conformations of proteins. Furthermore, the integration of 3D protein−ligand complex structures could lead to a more accurate representation of the drug-target interaction.



https://doi.org/10.1021/acs.jcim.4c01828
J. Chem. Inf. Model. 2025, 65, 981−996

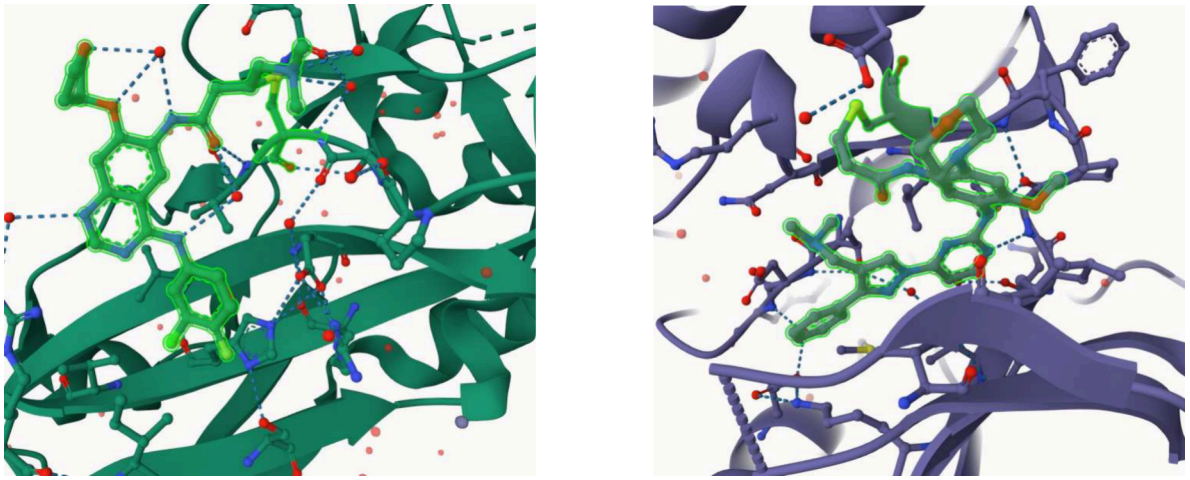


**Figure 8.** Molecular docking results of two different ligands with EGFR, visualized using PyMOL. The left image shows the docking of a ligand with EGFR, highlighting the hydrogen bonds (dashed blue lines) between the ligand and EGFR amino acid residues. The right image displays another ligand's interaction with EGFR, emphasizing similar hydrogen bonds formed between the ligand and protein residues. These visualizations demonstrate key interactions critical for drug binding.

Looking ahead, we identified several exciting directions for future work. First, we plan to integrate additional ligand features, such as ligand structure and the physicochemical properties of atoms, to enhance the model's predictive power. Furthermore, we intend to explore pretraining and fine-tuning techniques, as well as identify loss functions that are more suited to our model, moving beyond basic task loss functions. Lastly, one of our primary objectives is to broaden the application of the MMSG-DTA model in drug discovery. We aim to utilize the PLINDER[46] data set to improve the model's ability to generalize to previously unseen drug−protein pairs, a critical aspect for identifying potential drug repurposing candidates and predicting novel drug-target interactions.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The source codes are available at https://github.com/JiahaoXY/MMSG-DTA. All data sets used in this article are publicly available and can be accessed via the following links. The Davis data set is available at https://github.com/hkmztrk/DeepDTA/tree/master/data/davis. The KIBA data set is available at https://github.com/hkmztrk/DeepDTA/tree/master/data/kiba. The Metz data set is available at https://github.com/simonfqy/PADME/tree/master/metzdata.

## ■ AUTHOR INFORMATION


### Corresponding Author

**Wei Long** − *School of Information Engineering, Huzhou University, Huzhou 313000, China;* orcid.org/0000-0001-8162-0954; Email: lw@zjhu.edu.cn

### Authors

**Jiahao Xu** − *School of Information Engineering, Huzhou University, Huzhou 313000, China; Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China*

**Lei Ci** − *School of Information Engineering, Huzhou University, Huzhou 313000, China*

**Bo Zhu** − *School of Information Engineering, Huzhou University, Huzhou 313000, China*

**Guanhua Zhang** − *School of Information Engineering, Huzhou University, Huzhou 313000, China; Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China*

**Linhua Jiang** − *School of Information Engineering, Huzhou University, Huzhou 313000, China; Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China*

**Shixin Ye-Lehmann** − *Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China; Faculty of Medicine, University Paris-Saclay, Paris 94276, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c01828


### Author Contributions

The conceptualization and design of the study, including the formulation of specific ideas and methodologies, were jointly led by Jiahao Xu and Guanhua Zhang. The literature review was conducted by Bo Zhu. The development of the code was a collaborative effort between Jiahao Xu and Lei Ci. The drafting and revisions of the manuscript were undertaken by Jiahao Xu and Wei Long, with significant contributions from Shixin Ye-Lehmann during the revision process. The study was funded by Professor Linhua Jiang.

### Notes

The authors declare no competing financial interest.

## ■ 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.

## ■ REFERENCES

(1) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics* **2016**, *47*, 20−33.

(2) Hinkson, I. V.; Madej, B.; Stahlberg, E. A. Accelerating therapeutics for opportunities in medicine: a paradigm shift in drug discovery. *Frontiers in pharmacology* **2020**, *11*, 770.

(3) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve r&d

productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9* (3), 203−214.
(4) Pan, X.; Lin, X.; Cao, D.; Zeng, X.; Yu, P. S.; He, L.; Nussinov, R.; Cheng, F. Deep learning for drug repurposing: Methods, databases, and applications. *Wiley interdisciplinary reviews: Computational molecular science* **2022**, *12* (4), No. e1597.
(5) Gilson, M. K.; Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36* (1), 21−42.
(6) Chan, H. S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing drug discovery via artificial intelligence. *Trends in pharmacological sciences* **2019**, *40* (8), 592−604.
(7) D'Souza, S.; Prema, K.; Balaji, S. Machine learning models for drug−target interactions: current knowledge and future directions. *Drug Discovery Today* **2020**, *25* (4), 748−756.
(8) Ezzat, A.; Wu, M.; Li, X.-L.; Kwoh, C.-K. Computational prediction of drug−target interactions using chemogenomic approaches: an empirical survey. *Briefings in bioinformatics* **2019**, *20* (4), 1337−1357.
(9) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* **2011**, *7* (2), 146−157.
(10) Zhang, L.; Tan, J.; Han, D.; Zhu, H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today* **2017**, *22* (11), 1680−1685.
(11) Li, H.; Leung, K.-S.; Wong, M.-H.; Ballester, P. J. Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study. *BMC bioinformatics* **2014**, *15*, 291.
(12) Shar, P. A.; Tao, W.; Gao, S.; Huang, C.; Li, B.; Zhang, W.; Shahen, M.; Zheng, C.; Bai, Y.; Wang, Y. Pred-binding: large-scale protein−ligand binding affinity prediction. *Journal of enzyme inhibition and medicinal chemistry* **2016**, *31* (6), 1443−1450.
(13) Pahikkala, T.; Airola, A.; Pietilä, S.; Shakyawar, S.; Szwajda, A.; Tang, J.; Aittokallio, T. Toward more realistic drug−target interaction predictions. *Briefings in bioinformatics* **2015**, *16* (2), 325−337.
(14) He, T.; Heidemeyer, M.; Ban, F.; Cherkasov, A.; Ester, M. Simboost: a read-across approach for predicting drug−target binding affinities using gradient boosting machines. *Journal of cheminformatics* **2017**, *9*, 24.
(15) Öztürk, H.; Özgür, A.; Ozkirimli, E. Deepdta: deep drug−target binding affinity prediction. *Bioinformatics* **2018**, *34* (17), i821−i829.
(16) Öztürk, H.; Ozkirimli, E.; Özgür, A. Widedta: prediction of drug-target binding affinity, *arXiv preprint* arXiv:1902.04166 https://arxiv.org/abs/1902.04166.
(17) Abbasi, K.; Razzaghi, P.; Poso, A.; Amanlou, M.; Ghasemi, J. B.; Masoudi-Nejad, A. Deepcda: deep cross-domain compound−protein affinity prediction through lstm and convolutional neural networks. *Bioinformatics* **2020**, *36* (17), 4633−4642.
(18) Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; Wei, Z. Drug−target affinity prediction using graph neural network and contact maps. *RSC Adv.* **2020**, *10* (35), 20701−20712.
(19) Nguyen, T.; Le, H.; Quinn, T. P.; Nguyen, T.; Le, T. D.; Venkatesh, S. Graphdta: predicting drug−target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37* (8), 1140−1147.
(20) Yang, Z.; Zhong, W.; Zhao, L.; Chen, C. Y.-C. Mgraphdta: deep multiscale graph neural network for explainable drug−target binding affinity prediction. *Chemical science* **2022**, *13* (3), 816−833.
(21) Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; Liu, T.-Y. Do transformers really perform badly for graph representation? *Advances in neural information processing systems* **2021**, *34*, 28877−28888.
(22) Bento, A. P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An open source chemical structure curation pipeline using rdkit. *Journal of Cheminformatics* **2020**, *12*, 51.
(23) Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S. Y. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* **2021**, *32* (1), 4−24.
(24) Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems*.
(25) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. Strategies for pre-training graph neural networks, *arXiv preprint*. arXiv:1905.12265 https://arxiv.org/abs/1905.12265.
(26) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv* **2022**, *2022*, 500902.
(27) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint*. arXiv:1710.10903 https://arxiv.org/abs/1710.10903.
(28) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology* **2011**, *29* (11), 1046−1051.
(29) Tang, J.; Szwajda, A.; Shakyawar, S.; Xu, T.; Hintsanen, P.; Wennerberg, K.; Aittokallio, T. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J. Chem. Inf. Model.* **2014**, *54* (3), 735−743.
(30) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating the kinome. *Nat. Chem. Biol.* **2011**, *7* (4), 200−202.
(31) Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; Beaini, D. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems* **2022**, *35*, 14501−14515.
(32) Zhu, Z.; Yao, Z.; Zheng, X.; Qi, G.; Li, Y.; Mazur, N.; Gao, X.; Gong, Y.; Cong, B. Drug−target affinity prediction method based on multi-scale information interaction and graph optimization. *Computers in Biology and Medicine* **2023**, *167*, 107621.
(33) Wu, H.; Liu, J.; Jiang, T.; Zou, Q.; Qi, S.; Cui, Z.; Tiwari, P.; Ding, Y. Attentionmgt-dta: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks* **2024**, *169*, 623−636.
(34) Dehghan, A.; Abbasi, K.; Razzaghi, P.; Banadkuki, H.; Gharaghani, S. Ccl-dti: contributing the contrastive loss in drug−target interaction prediction. *BMC bioinformatics* **2024**, *25* (1), 48.
(35) Zhou, C.; Li, Z.; Song, J.; Xiang, W. Transvae-dta: Transformer and variational autoencoder network for drug-target binding affinity prediction. *Computer Methods and Programs in Biomedicine* **2024**, *244*, 108003.
(36) Tang, X.; Lei, X.; Zhang, Y. Prediction of drug-target affinity using attention neural network. *International Journal of Molecular Sciences* **2024**, *25* (10), 5126.
(37) Zhu, Z.; Zheng, X.; Qi, G.; Gong, Y.; Li, Y.; Mazur, N.; Cong, B.; Gao, X. Drug−target binding affinity prediction model based on multi-scale diffusion and interactive learning. *Expert Systems with Applications* **2024**, *255*, 124647.
(38) Yang, Z.; Zhong, W.; Zhao, L.; Chen, C. Y.-C. Ml-dti: mutual learning mechanism for interpretable drug−target interaction prediction. *J. Phys. Chem. Lett.* **2021**, *12* (17), 4247−4261.
(39) Zhang, L.; Wang, C.-C.; Zhang, Y.; Chen, X. Gpcndta: prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores. *Computers in Biology and Medicine* **2023**, *166*, 107512.
(40) Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J. K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on chembl. *Chemical science* **2018**, *9* (24), 5441−5451.
(41) Zhu, Z.; Yao, Z.; Qi, G.; Mazur, N.; Yang, P.; Cong, B. Associative learning mechanism for drug-target interaction prediction. *CAAI Transactions on Intelligence Technology* **2023**, *8* (4), 1558−1577.
(42) He, H.; Chen, G.; Chen, C. Y.-C. Nhgnn-dta: a node-adaptive hybrid graph neural network for interpretable drug−target binding affinity prediction. *Bioinformatics* **2023**, *39* (6), btad355.
(43) Knox, C.; Wilson, M.; Klinger, C. M.; Franklin, M.; Oler, E.; Wilson, A.; Pon, A.; Cox, J.; Chin, N. E.; Strawbridge, S. A.; et al.

Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic acids research* **2024**, *52* (D1), D1265−D1275.
(44) Peng, D.; Liang, P.; Zhong, C.; Xu, P.; He, Y.; Luo, Y.; Wang, X.; Liu, A.; Zeng, Z. Effect of egfr amplification on the prognosis of egfr-mutated advanced non−small-cell lung cancer patients: a prospective observational study. *BMC cancer* **2022**, *22* (1), 1323.
(45) Ding, J.; Tang, S.; Mei, Z.; Wang, L.; Huang, Q.; Hu, H.; Ling, M.; Wu, J. Vina-gpu 2.0: further accelerating autodock vina and its derivatives with graphics processing units. *J. Chem. Inf. Model.* **2023**, *63* (7), 1982−1998.
(46) Durairaj, J.; Adeshina, Y.; Cao, Z.; Zhang, X.; Oleinikovas, V.; Duignan, T.; McClure, Z.; Robin, X.; Kovtun, D.; Rossi, E.; et al. Plinder: The protein-ligand interactions dataset and evaluation resource. *bioRxiv* **2024**, 2024−07.




https://doi.org/10.1021/acs.jcim.4c01828
*J. Chem. Inf. Model.* 2025, 65, 981−996