ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

A Multibranch Neural Network for Drug-Target Affinity Prediction Using Similarity Information

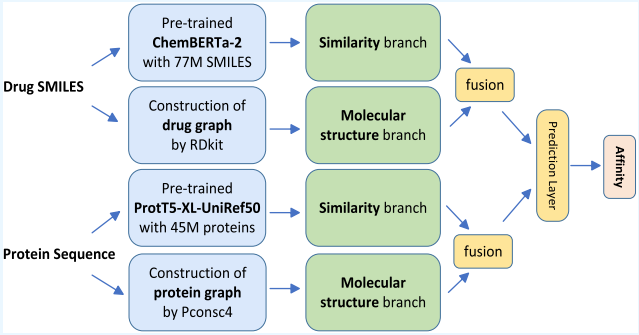
Jing Chen,* Xiaolin Yang, and Haoyu Wu

Cite This: *ACS Omega* 2024, 9, 35978−35989

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Predicting drug-target affinity (DTA) is beneficial for accelerating drug discovery. In recent years, graph structure-based deep learning models have garnered significant attention in this field. However, these models typically handle drug or target protein in isolation and only extract the molecular structure information on the drug or protein itself. To address this limitation, existing network-based models represent drug-target interactions or affinities as a knowledge graph to capture the interaction information. In this study, we propose a novel solution. Specifically, we introduce drug similarity information and protein similarity information into the field of DTA prediction. Moreover, we propose a network framework that autonomously extracts similarity information, avoiding reliance on knowledge graphs. Based on this framework, we design a multibranch neural network called GASI-DTA. This network integrates similarity information, sequence information, and molecular structure information. Comprehensive experimental results conducted on two benchmark data sets and three cold-start scenarios demonstrate that our model outperforms state-of-the-art graph structure-based methods in nearly all metrics. Furthermore, it exhibits significant advantages over existing network-based models, outperforming the best of them in the majority of metrics. Our study's code and data are openly accessible at http://github.com/XiaoLin-Yang-S/GASI-DTA.



INTRODUCTION

Drug discovery plays a crucial role in the development of new medications, aiming to identify compounds that interact with specific biological targets, especially the target proteins.[1] However, this process is time-consuming and requires substantial funding, making it challenging.[2] Predicting drug-target affinity provides valuable information about the strength of interaction between drugs and their targets, as measured by the dissociation constant (K_d), inhibition constant (K_i), or maximum inhibition concentration (IC_{50}), etc.[3] By using computational methods to predict affinity instead of relying solely on rigorous experiments, researchers can efficiently screen and prioritize drug candidates with promising research value. This accelerates the drug development process, providing timely and effective assistance in improving treatment outcomes for various diseases.[4]

The existing computational methods are mainly divided into three categories: traditional physics-based methods, machine learning-based methods, and deep learning-based methods. Traditional physics-based methods, such as molecular docking[5,6] and molecular dynamics simulations,[7,8] have achieved excellent predictive results by utilizing the three-dimensional (3D) structure of drug molecules and proteins. However, these methods face the challenge of computing resource consumption and scoring function design.[9] Machine learning-based approaches have had a significant impact in this field. Researchers use various methods to predict drug-target affinity, such as support vector machines, logistic regression, random forests, and least-squares.[10,11] However, due to their inherent limitations, such as reliance on complex feature engineering and expert domain knowledge, it is challenging to achieve adequate generalization in this field.[12,13]

In recent years, inspired by successful applications in various research fields, deep learning methods have also been widely utilized in bioinformatics and cheminformatics.[14−16] In the field of drug-target affinity (DTA) prediction, the advantage of deep learning-based methods is their ability to uncover hidden interactions between drugs and targets.[17] These methods can be categorized into structure-based methods and nonstructure-based methods. Structure-based methods focus on utilizing the three-dimensional structure of drug and target. These methods typically employ fully connected neural network (FCNN),[18]

Received: June 15, 2024
Revised: August 3, 2024
Accepted: August 6, 2024
Published: August 12, 2024

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

three-dimensional (3D) convolutional neural network (CNN),[19,20] graph neural network (GNN),[21,22] and so on for feature extraction. They are capable of capturing spatial relationships and geometric features between drugs and targets, thereby offering more accurate prediction results.[23] However, these methods have a notable limitation: they do not work when the 3D structure of the drug and target is unknown, and currently, obtaining the exact 3D structure of the drug and target remains a very challenging task. Nonstructure-based methods typically have a dual-encoder architecture, which separately learns representations of drugs and targets. These models can be divided into the following categories: sequence-based methods, graph structure-based methods, and network-based methods.

Sequence-based models take the sequence information on drugs and proteins directly as input for the model. For example, DeepDTA[24] used two convolutional neural networks to extract features from drug simplified molecular input line entry system (SMILES) strings and protein sequences. Subsequently, the learned features of drugs and proteins were combined and fed into a multilayer perceptron for DTA prediction. AttentionDTA[25] utilizes attention mechanisms to merge the obtained drug and protein representations, resulting in improved performance. ELECTRA-DTA[26] applied the ELECTRA[27] pretrained model, which extracts feature representations from a large amount of original sequence data, to the model. Utilizing the pretrained model alleviated the issue of insufficient data in DTA prediction tasks to some extent and achieved competitive performance at that time. MFR-DTA[28] found that most methods ignore the individual information on sequence elements, leading to inadequate sequence feature representation. Therefore, they proposed a new biological sequence feature extraction block, which includes a global feature extractor and an individual feature extractor to effectively extract global and individual features. Moreover, they use a spatial attention block to capture the local relationship among the adjacent elements, further enriching the extracted individual features. Although sequence-based approaches have made some progress, they overlook information regarding the molecular structure of drugs and proteins. This oversight can compromise the predictive power of models and their ability to learn functional correlations in underlying feature spaces.[3]

Graph structure-based models usually process the drug SMILES string or protein sequences into molecular graphs to extract molecular structure information. For instance, GraphDTA[29] constructs a drug molecule graph with atoms as nodes and bonds as edges. In this model, a GNN is used to extract features from the drug molecule graph, which enhances the prediction performance of DTA. DGraphDTA[30] further constructs the protein graph using Pconsc4.[31] This allows the model to acquire protein structure information simultaneously. WGNN-DTA[32] optimizes the generation of protein graph by introducing evolutionary scale modeling (ESM), which yields superior results. MGraphDTA[3] believes that GNN models with few layers are insufficient to capture the overall structure of compounds. Therefore, a deep graph convolutional network (GCN)[33] model is proposed to extract structure features. GLGN-DTA[34] integrates a graph learning module into the existing graph architecture. This module is designed to learn a soft adjacency matrix and extract more structure information than the traditional fixed adjacency matrix method.

Graph structure-based methods generally outperform sequence-based models because they incorporate molecular structure information. However, these models typically handle drug or protein in isolation and only extract the molecular structure information on the drug or protein itself. To solve this problem, network-based models have begun to emerge.

Network-based models typically represent drug-target interactions or affinities as a knowledge graph to extract the interaction information.[35] For example, BERT-DTA[36] represents drug-target interactions as a graph and utilizes GCN for feature extraction. HGRL-DTA[37] optimizes the modeling of interaction information by representing drug-target affinities as an affinity graph. In this graph, drugs and proteins serve as nodes, while drug-target affinities as the edges. HGRL-DTA inputs the affinity graph into the network along with the molecular graphs. It learns fine-level representations from molecular graphs and coarse-level representations from affinity graph. Finally, a coarse-to-fine information fusion method is used to further enhance the fine-level representation. Because of the aforementioned design, HGRL-DTA achieved the best current performance on both benchmark data sets.

Network-based models achieve better results compared with graph structure-based model. This offers a new research direction for DTA prediction tasks. However, these methods also have certain limitations. In cold start scenarios, new drugs or proteins are unseen in the training set, and we also do not know any interaction relationships between these new drugs/proteins and known proteins/drugs. In this case, it is impossible to construct a knowledge graph containing these new drugs or proteins during testing, making network-based methods unusable. One solution is a similarity-based representation inference method. During testing, the model only utilizes the graph node features obtained from the training. For nodes that are not included in the training set, it utilizes the similarity function to calculate their features by existing similar node features. However, this approach is limited by the manually designed similarity functions, which hinders the development of end-to-end deep learning methods in DTA prediction. Meanwhile, existing network-based methods typically employ simple, manually designed initial features. They do not leverage the features that contain rich semantic information provided by the rapidly evolving pretrained models.

In this study, We introduce drug similarity information and protein similarity information into the field of DTA prediction. Specifically, different from traditional drug/protein sequence processing, which treats atoms/residues as tokens, each drug/protein is considered a sequence. We input each drug/protein as a token, and all or partial drugs/proteins as a sequence into a bidirectional long short-term memory (BiLSTM) network. In the BiLSTM network, each token can learn not only its own sequence information but also useful information contained in other similar tokens. In this way, we do not need to manually construct a knowledge graph, the model can autonomously extract similarity information. Then based on this method, we design a multibranch neural network called GASI-DTA, which integrates similarity information, sequence information, and molecular structure information. To be specific, we utilize two branches to extract information for both drugs and proteins. One branch is the molecular structure branch. We use the methods described in DGraphDTA[30] to construct molecular graphs for drugs and proteins, and utilize simple yet effective GCN to extract molecular structure information from these

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

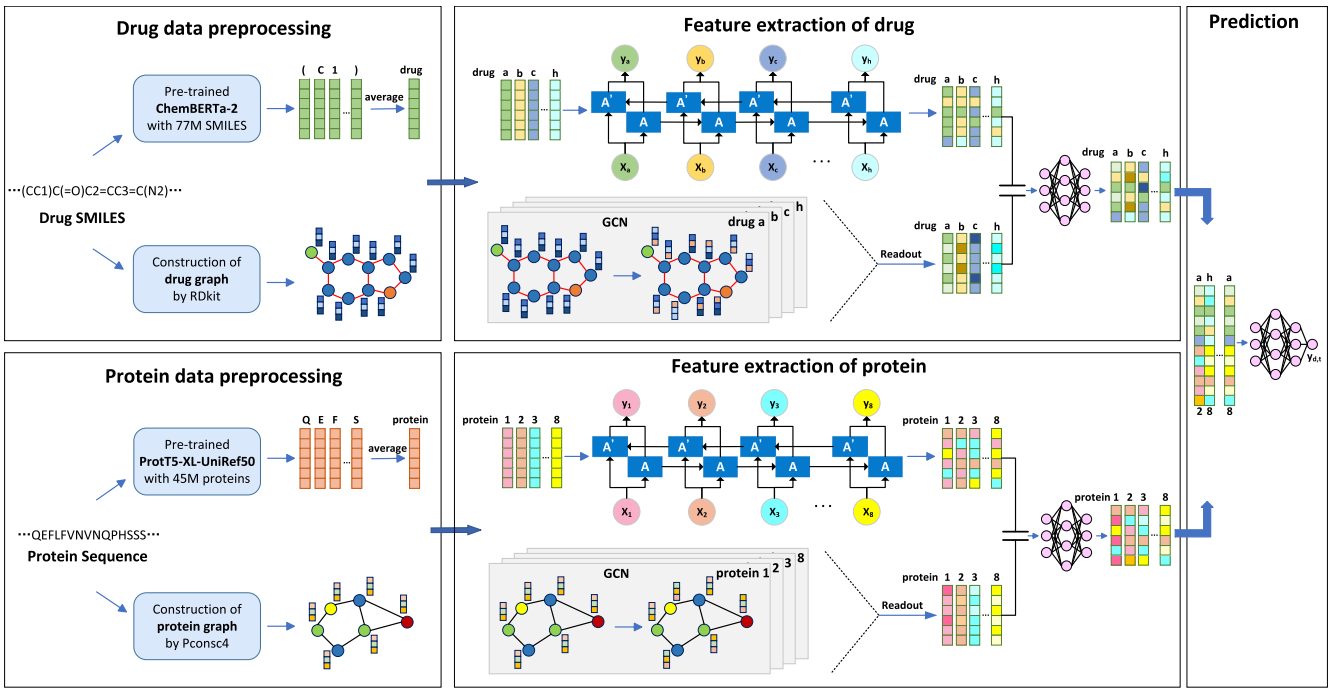


Figure 1. Framework of the proposed GASI-DTA model. Entire framework is divided into three modules: data preprocessing, feature extraction, and prediction. In the data preprocessing module, drug SMILE strings and protein sequences are preprocessed to obtain their respective sequence features and molecular graphs. In the feature extraction module, both drugs and proteins use two branches to exact information. The upper similarity branch utilizes Bi-LSTM to extract similarity information and sequence information from sequence features. The lower molecular structure branch utilizes GCN to extract molecular structure information from molecular graphs. The results from the two branches are fused to derive the ultimate features of drugs and proteins, respectively. In the prediction module, the features of drugs and proteins are concatenated and then input into fully connected layers to obtain affinity.

graphs. The other branch is similarity branch. We use pretraining models to obtain the sequence features of drugs and proteins that contain rich semantic information. These features are then input into the BiLSTM[38] to extract sequence and similarity information.

In cold start scenarios, our model has two advantages compared to network-based models. First, our model does not require the construction of a knowledge graph, enabling it to run directly in cold start scenarios. However, when the test set data differs significantly or even entirely from the training set data, the similarity information extracted by the training set is unusable and misleading in the test set. Hence, our model GASI-DTA would be subject to certain interference when directly utilizing all similarity information, which makes it challenging to compare its performance with the state-of-the-art models in these scenarios. Second, our model extracts drug similarity information and protein similarity information in a decoupled manner, ensuring that changes in one do not affect the extraction of similarity information for the other within the test set. Therefore, to address the aforementioned issues, we designed a combinatorial approach by leveraging this decoupling property to extract partial available similarity information and avoid interference from misleading information.

Our main contributions are summarized as follows:

- We utilize pretrained models to extract sequence features containing rich semantic information. This approach help alleviate the challenges of data scarcity and overfitting to some extent in the DTA prediction task.
- We introduce drug similarity information and protein similarity information into the field of DTA prediction. Moreover, we propose a network framework that autonomously learns similarity information, avoiding reliance on knowledge graphs.
- We design a multibranch neural network called GASI-DTA that integrates multiple information. Specifically, the model can autonomously extract similarity information and sequence information from sequence features using a similarity branch. Meanwhile, molecular structure information is extracted from molecular graphs using a molecular structure branch.
- Comprehensive experimental results conducted on two benchmark data sets and three cold-start scenarios demonstrate that our model outperforms state-of-the-art graph structure-based methods in nearly all metrics. Furthermore, it exhibits significant advantages over existing network-based models, outperforming the best of them in the majority of metrics.

METHODS

In this section, we introduce the details of our model GASI-DTA. The general framework of GASI-DTA is illustrated in Figure 1. Entire framework is divided into three modules: data preprocessing, feature extraction, and prediction. In the data preprocessing module, we preprocess the drug SMILES strings and protein sequences in the data set to derive the corresponding sequence features and molecular structure graphs. In particular, we utilize pretrained models to acquire sequence features that contain rich semantic information. In the feature extraction module, both drugs and proteins pass

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

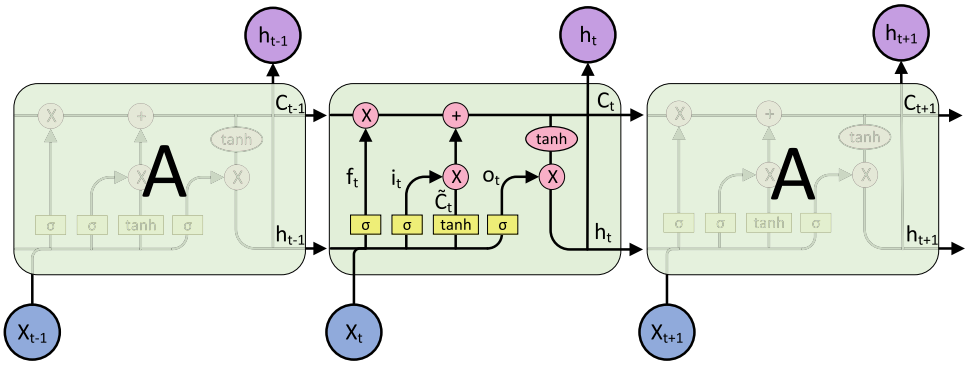


Figure 2. Framework of the LSTM model. Three tokens, X_{t-1}, X_t, and X_{t+1}, are input into the model, and the model processes them sequentially. For each input, the model passes through a Module A to obtain the cell state C and output h. In the Module A, yellow blocks represent neural network layers, while pink circles denote bitwise or element-wise operations.

through two network branches to exact information. The upper similarity branch utilizes Bi-LSTM to extract similarity information and sequence information. For each token in the sequence, Bi-LSTM generates the corresponding output using an Module A and an Module A'. The lower molecular structure branch employs GCN to extract molecular structure information. Subsequently, the results of the two branches are fused through multiple fully connected layers to obtain the final features of the drugs/proteins. Finally, the acquired drugs and proteins features are fed into the prediction network to obtain the affinity values of the corresponding drug-target pairs.

Data Preprocessing. *Pretrained Model.* To encode drug SMILES sequences and proteins sequences, existing deep learning models such as DeepDTA[24] and DeepCDA[39] utilize simple label/one-hot encoding to represent each symbol in the sequences. These encoding methods can only extract simple features and cannot encode advanced semantic information within the sequence, which is often the most crucial. Using pretrained models to encode sequences can effectively address this issue. Through pretraining, the model can learn some general feature representations that capture advanced semantic information on the data more effectively. In addition, the pretrained model offers several advantages. First, it typically leverages a substantial amount of unlabeled data for training, enabling it to fully utilize data resources. This is particularly advantageous in the biological field, where data scarcity is common. For example, the data set used by the drug pretrained models chemBERTa-2[40] is over one million times larger than the DTA benchmark data set, Davis. Second, employing pretrained strategies can significantly enhance the model's generalization capability, leading to improved performance on unseen data. In conclusion, the feature encoding based on pretrained models can significantly provide additional and indispensable assistance to specific downstream research areas, especially in data-limited biological fields. Therefore, our study utilizes this feature encoding to acquire drug and protein sequence features.

Pretrained Representation of Drug SMILES. For drug SMILES strings, we selected the current best molecule pretrained model, chemBERTa-2,[40] to generate the corresponding string features. The authors of chemBERTa-2 collected 77 million strings from PubChem[41] to construct an unsupervised pretraining data set, which is one of the largest molecular pretrained data sets available to date. The pretrained model has demonstrated excellent performance in various downstream tasks, confirming its reliability. As illustrated in Figure 1, we first obtain features for each drug string symbol using this pretrained model. The features were averaged to derive embedding for the entire drug. Then all drug embeddings will be fed into the drug similarity branch.

Pretrained Representation of Protein Sequences. For protein sequences, we selected the classical ProtT5-XL-UniRef50[42] pretrained model, which has been validated and extensively utilized in numerous downstream tasks, to extract protein sequence features. This model was pretrained on UniRef50, which comprises 45 million protein sequences. We utilized ProtT5-XL-UniRef50 to obtain features for each protein residue. Then, we computed the average features of all residues for each protein to generate the overall protein sequence features. Subsequently, these features were input into the protein similarity branch.

Drug Graph Representation. For drug molecules, following the previous study GraphDTA,[29] we use RDKit[43] to convert the drug SMILES strings into molecular graphs, where nodes represent atoms and edges represent bonds. In this study, we select the same set of atomic features as GraphDTA to initialize the features of the graph nodes. Specifically, we concatenate several different one-hot encodings along with a determination of whether the atom is aromatic, forming a 78-dimensional vector. This vector serves as the initial feature input for the drug graph nodes in our model.

Protein Graph Representation. For protein molecular graphs, following the previous study DGraphDTA,[30] We first preprocess the protein sequence, which includes sequence alignment and filtering. Then, the results were fed into Pconsc4[31] to convert the sequence into a contact map. Finally, we apply the same threshold filtering to obtain the adjacency matrix of the protein graphs. Through this process, the protein sequence is transformed into graph form, where nodes represent protein residues. We utilize the same set of residue features as DGraphDTA to initialize the features of the graph nodes. Specifically, we concatenate the one-hot encoding of the residue, position-specific scoring matrix (PSSM), and the biological features of the residue to form a 54-dimensional vector as the input feature for the protein graph nodes.

Feature Extraction. *Similarity Branch.* In the existing sequence-based DTA prediction networks, each atom/residue within the drug/protein sequence is treated as a token, and the entire sequence is input into the network as a whole. In this setup, the network can only learn the intrinsic sequence information on the drug/protein itself. In our study, we regard

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

each drug/protein as a token, and input either all or a portion of the drugs/proteins as a sequence into the similarity branch to acquire its own sequence information and similarity information. The similarity branch employs a BiLSTM[38] network. BiLSTM consists of two long short-term memory (LSTM) networks in opposite directions, and the output of each token is determined collectively by these two LSTM networks.

The processing procedure of LSTM is illustrated in Figure 2. Three tokens, X_{t-1}, X_t, and X_{t+1}, are input into the model, and the model processes them by the Module A sequentially. After processing X_{t-1}, the model obtains the current cell state, C_{t-1}, and output, h_{t-1}. The cell state, C_{t-1}, is crucial in LSTM as it preserves the state information at time $t-1$, which is then propagated to the next time through the upper channel in Figure 2. Since the state information transmitted through this channel undergoes minimal linear operations, it only induces minor changes, enabling tokens further down the sequence to comprehend the information contained in earlier tokens. Meanwhile, h_{t-1} interacts with the input of token X_t through the lower channel to generate C_t and h_t at time t. The above process continues sequentially until the end of the sequence. It can be seen from the above that LSTM depends on the input order of the sequence, and a single LSTM can only predict the output of the next time based on the previous temporal information. Therefore, our study adopts BiLSTM to obtain similarity information. In BiLSTM, data will be input into two independent LSTM networks in the opposite order of $X_{t-1}X_tX_{t+1}$ and $X_{t+1}X_tX_{t-1}$. This arrangement enables each token to access all preceding and succeeding temporal information. In this way, each drug/protein can fully acquire useful information from other similar drugs/proteins.

Molecular Structure Branch. As demonstrated in previous studies,[3,30] graph structures-based models can effectively learn molecular structure information that is missing in sequence-based models, thereby achieving better predictive performance. In our study, we also acknowledge the importance of molecular structure information. Hence, we introduce a molecular structure branch to extract the molecular structure information on drugs and proteins. We employ a multilayered GCN as the backbone of the molecular structure branch. The calculation formula of GCN is shown in formula 1

$$H^{l+1} = f(H^l, A) = \sigma(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}H^lW^{l+1}) \quad (1)$$

where H^l is the output of the l-th layer and also serves as the input to the $(l+1)$-th layer, $A \in R^{n\times n}$ is the adjacency matrix of the molecular graph, where n is the number of nodes in the graph. $\hat{D}$ is the diagonal node degree matrix calculated from A, and with the same dimension as A. $\hat{A}$ is the sum of A and the identity matrix $I \in R^{n\times n}$. W^{l+1} is the learnable weight parameter of the $(l+1)$-th layer.

Feature Fusion. Through the two aforementioned branches, we obtain the sequence and similarity features of drugs/proteins as well as the molecular structure features. Then, we concatenate the obtained features and then allow the model to autonomously learn useful information from them through two fully connected layers, aiming to achieve better fusion effectiveness. During this process, we employ dropout to prevent overfitting and achieve better predictive performance.

Drug-Target Affinity Prediction. After the feature extraction module described above, we obtain feature vectors for both drugs and proteins. We concatenate these two feature vectors and further process the result through three fully connected layers to obtain the final drug-target affinity. During this process, we also employ dropout to prevent overfitting and achieve better predictive performance.

Prediction for Cold-Start Scenarios. Different from current network-based models, GASI-DTA autonomously extracts drug similarity information and protein similarity information in a decoupled manner. Therefore, in cold-start scenarios, when partial similarity information is unavailable, our model can still learn from other useful similarity information. Leveraging this advantage, we devise a combinatorial approach to extract only the partial available similarity information and circumvent interference from the misleading information in cold-start scenarios.

As illustrated in Figure 3, we design a sequence branch to only learn sequence information. The pretrained sequence features of drugs or proteins are used as inputs. These features are processed by four ResBlocks, each consisting of two fully connected layers. Leveraging pretrained features including rich semantic information, the network can effectively extract sequence information. In cold-start scenarios, we use this branch to substitute the similarity branch including unavailable similarity information, while keeping the input-output feature dimensions consistent for both branches. We designate the resulting model as GAS-DTA. Specifically, when only the drug is unseen in the training, we merely substitute the drug's

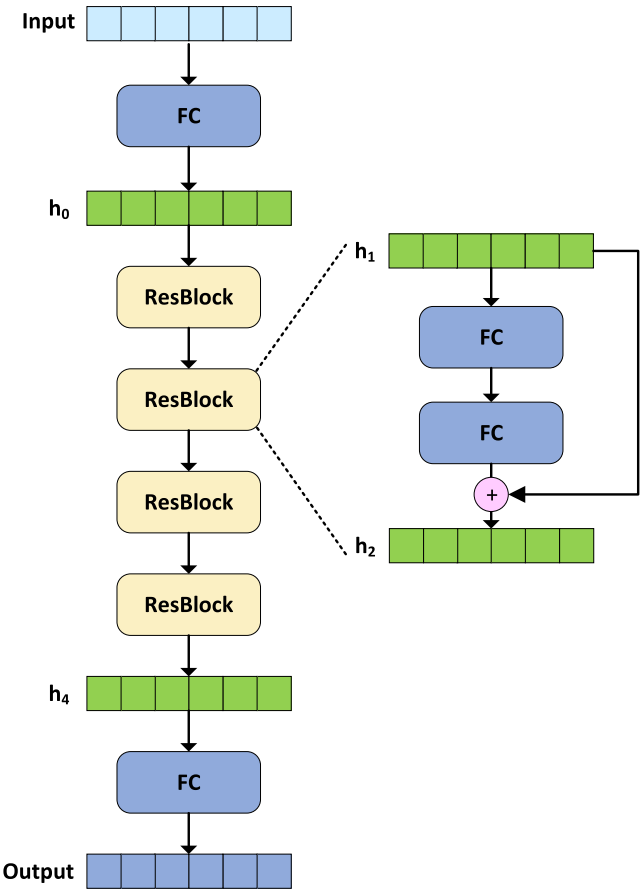


Figure 3. Sequence branch. FC represents Fully Connected layer, and ResBlock denotes residual block. Sequence information is extracted from sequence features by the multilayer residual blocks.

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

similarity branch, allowing GAS-DTA to still extract partial similarity information from the protein data.

RESULTS AND DISCUSSION

Data Sets. To ensure a comprehensive comparison with existing methods, we choose to assess all models on two widely used and publicly available DTA data sets, named Davis[44] and KIBA.[45] Based on these two data sets, we utilized four different experimental setups, including the baseline data set and three cold start scenarios. Detailed information about these two data sets is provided in Table 1.

Table 1. Summary of the DTA Datasets

data set	drugs	proteins	binding entries
Davis	68	442	30,056
KIBA	2111	229	118,254

In the Davis data set, the average length of drug SMILES strings is 64, while the average length of protein sequences is 788. The data set is curated by selecting specific kinase proteins and their inhibitors, where the affinity is represented by the dissociation constant (K_d). Higher K_d values indicate weaker binding strength for drug-target pairs. Following prior research,[24] affinity values in this data set are processed by formula 2

$$pK_d = -\log_{10}\frac{K_d}{10^9} \tag{2}$$

In the KIBA data set, the average length of drug SMILES strings is 58, while the average length of protein sequences is 728. The data set compiles the bioactivities of kinase inhibitors from multiple sources, including inhibition constants (K_i), dissociation constants (K_d), and half-maximal inhibitory concentrations (IC_{50}). The KIBA score is utilized to denote the binding affinity between drugs and proteins, where lower KIBA scores indicate stronger binding strength.

To better evaluate our model, we adopt the same experimental setup as HGRL-DTA. We divided each data set into training and test sets in a 5:1 ratio. Meanwhile, we conducted 5-fold cross-validation on the training set to select the best hyperparameters for our model.

Evaluation Metrics. In our study, we employ three widely used metrics to assess our model's performance: mean squared error (MSE), concordance Index (CI) and r_m^2.

MSE is a common metric in regression tasks used to quantify the deviation between predicted values and ground truths. A smaller MSE signifies that the model's predictions are closer to the actual values of the samples. MSE is defined by formula 3

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{3}$$

where p_i denotes the predicted value of the i-th data sample, y_i represents the true value of the i-th data sample, and n signifies the number of drug-target pairs in the data set.

CI measures the consistency in order between predicted and true values. This metric ranges from 0 to 1, where higher values indicate better results. The calculation of CI is described by formula 4

$$\mathrm{CI} = \frac{1}{Z}\sum_{d_x > d_y} h(b_x - b_y),\ h(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0.5, & \text{if } x = 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{4}$$

where b_x denotes the predicted value corresponding to the larger true value d_x, while b_y represents the predicted value corresponding to the smaller true value d_y. $h(x)$ stands for a commonly used step function, and Z is a normalization constant.

The r_m^2 involved in DeepDTA[24] is used to evaluate the model's external prediction capability. We calculate the r_m^2 metric by formula 5

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{5}$$

where r^2 and r_0^2 are the squared correlation coefficients with and without intercept respectively.

Experimental Environment and Settings. We implemented our method using PyTorch 2.0.0 and Torch-Geometric 2.3.1. All experiments were conducted on a NVIDIA GeForce RTX 3080 GPU. The hyperparameters of our model are presented in Table 2.

Table 2. Hyperparameter Settings of GASI-DTA

hyperparameter	setting
learning rate	0.0005
batch size	512
Epoch	2000
dropout rate	0.2
GCN layers	3
optimizer	Adam
loss function	MSE

Comparison with Existing Methods on Two Benchmark Data Sets. To thoroughly evaluate the effectiveness of our proposed method GASI-DTA in the DTA prediction task, we compare it with various existing methods across two benchmark data sets. These methods include sequence-based models such as DeepDTA,[24] AttentionDTA,[25] ELECTRA-DTA,[27] graph structure-based models like GraphDTA,[29] DGraphDTA,[30] MGraphDTA,[3] GSAML-DTA,[46] GLCN-DTA,[34] as well as network-based models BERT-DTA[36] and HGRL-DTA.[37]

Table 3 presents the comparative results. It can be observed that our model, GASI-DTA, still outperforms the state-of-the-art model HGRL-DTA on all metrics in the Davis data set, demonstrating the superiority of our model. Additionally, on the KIBA data set, our model slightly lags behind HGRL-DTA in the most important metric, MSE. But it also approaches the current best level, matching or surpassing other graph structure-based models. This is because in the Davis data set, where the data volume is relatively small, our method of autonomously extracting similarity and sequence information using pretrained sequence features enables the model to acquire more effective information, leading to the best experimental outcomes. However, in the KIBA data set, where the data volume is larger and information is more abundant, the performance differences between all models are generally smaller. Network-based models such as HGRL-DTA set a threshold when creating interaction knowledge graphs, which artificially filters the information to a certain extent, thereby avoiding information interference. In contrast, our

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

Table 3. Prediction Performance on the Two Benchmark Datasets[a]

	Davis			KIBA		
model	MSE	CI	r_m^2	MSE	CI	r_m^2
DeepDTA	0.245	0.888	0.665	0.181	0.868	0.711
AttentionDTA	0.233	0.889	0.676	0.150	0.883	0.760
ELECTRA-DTA	0.238	0.897	0.671	0.162	0.889	0.727
GraphDTA	0.229	0.893	0.685	0.139	0.891	0.730
DGraphDTA	0.216	0.900	0.686	0.132	0.902	0.800
GSAML-DTA	0.201	0.896	0.718	0.132	0.900	0.800
MGraphDTA	0.207	0.900	0.710	0.128	0.902	**0.801**
GLCN-DTA	0.215	0.903	0.720	0.127	0.899	0.792
BERT-GCN	0.199	0.896	0.741	0.149	0.888	0.761
HGRL-DTA	0.166	**0.911**	0.751	**0.125**	**0.906**	0.789
GASI-DTA(ours)	**0.157**	**0.911**	**0.778**	0.127	0.905	0.791

[a]The best score in each column is in bold and the second best score is underlined.

Table 4. Prediction Performance in Cold-Start Scenarios[a]

		Davis			KIBA		
scenarios	model	MSE	CI	r_m^2	MSE	CI	r_m^2
S1	DeepDTA	0.985	0.548	0.027	0.494	0.747	0.377
	AttentionDTA	0.869	0.642	0.079	0.506	0.744	0.298
	GraphDTA	0.801	0.659	0.160	0.475	0.753	0.380
	DGraphDTA	0.818	0.646	0.114	0.458	0.754	0.298
	MGraphDTA	0.907	0.599	0.082	0.469	0.752	0.366
	HGRL-DTA	0.776	**0.684**	0.163	0.434	0.757	0.370
	GAS-DTA(ours)	**0.765**	0.674	**0.164**	**0.388**	**0.775**	**0.443**
S2	DeepDTA	0.552	0.729	0.258	0.732	0.676	0.273
	AttentionDTA	0.436	0.787	0.304	0.529	0.693	0.254
	GraphDTA	0.860	0.666	0.134	0.469	0.710	0.388
	DGraphDTA	0.445	0.788	0.289	0.364	0.718	0.429
	MGraphDTA	**0.359**	0.813	**0.415**	0.483	0.674	0.342
	HGRL-DTA	0.383	0.816	0.375	**0.322**	**0.741**	**0.502**
	GAS-DTA(ours)	0.386	**0.817**	0.349	0.340	0.735	0.489
S3	DeepDTA	0.767	0.508	0.009	0.700	0.627	0.140
	AttentionDTA	0.679	0.554	0.005	0.609	0.629	0.143
	GraphDTA	0.988	0.569	0.020	0.676	0.641	0.149
	DGraphDTA	0.658	0.569	0.020	0.676	0.641	0.148
	MGraphDTA	0.764	0.507	0.001	0.660	0.627	0.152
	HGRL-DTA	0.642	0.602	**0.044**	0.532	0.642	0.207
	GAS-DTA(ours)	**0.616**	**0.610**	0.041	**0.493**	**0.679**	**0.245**

[a]The best score in each column is in bold and the second best score is underlined.

method of autonomously extracting similarity information did not conduct additional filtering, which affected the results to some extent. In our future work, we will further study this issue to avoid the mutual influence of large amounts of information and improve prediction performance.

Among all baselines, sequence-based methods such as DeepDTA, AttentionDTA, and ELECTRA-DTA simply represent drugs and proteins as one-dimensional sequences, which are insufficient to capture molecular structure information. Hence, their performance is relatively poor. On the other hand, graph structure-based models represent drugs and proteins as graphs, leading to better predictive performance. However, these models only extract molecular structure information and handle drug and protein in isolation. This is why they generally perform lower than our model in most cases. Compared to network-based methods, our model GASI-DTA achieved better performance on the Davis data set, indicating the advantages of our approach.

Comparison with Existing Methods in Cold-Start Scenarios. In the previous experiment, we described the experimental results on two benchmark data sets (i.e., where the training and test sets share the same drugs and proteins but contain different drug-target pairs). However, in more practical applications, we still encounter cold-start scenarios where new drugs or proteins are completely unseen in the training set. In these cold-start scenarios, only partial similarity information is available, and other similarity information may be misleading when inferring new drugs or proteins. To address this issue, we modified our model to adapt to such scenarios. The modified model is named GAS-DTA. In this section, to provide a more comprehensive evaluation of our model, GAS-DTA, we conduct experiments using three different cold-start scenarios.

- **S1:** each drug in the test set is unseen in the training set, but all protein is seen.
- **S2:** each protein in the test set is unseen in the training set, but all drug is seen.

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

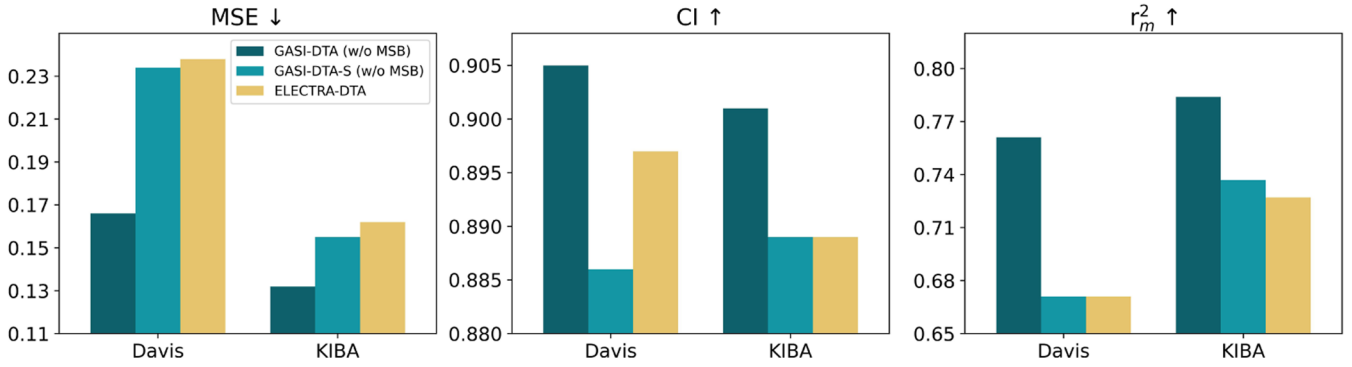


Figure 4. Performance evaluation of models with or without similarity information.

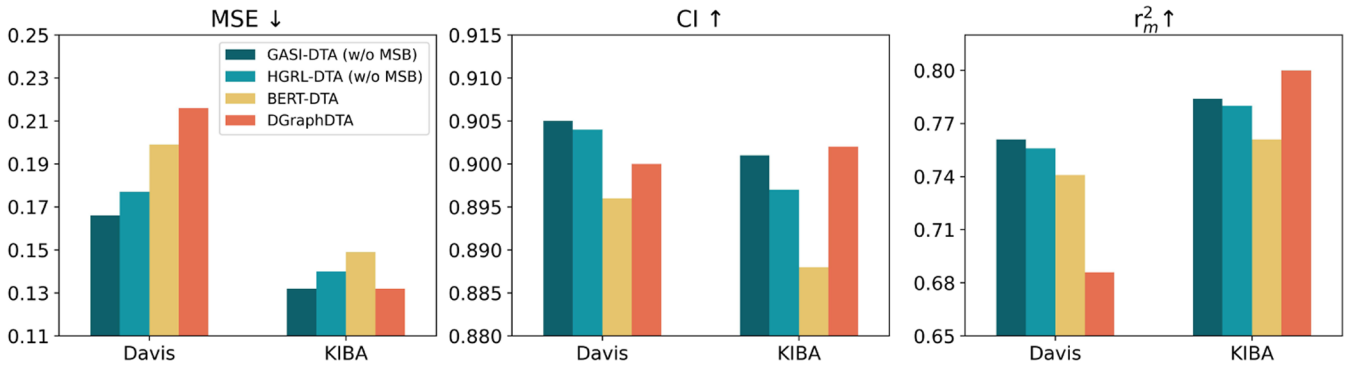


Figure 5. Performance evaluation of GASI-DTA (w/o MSB) compared to network-based models and graph structure-based model.

- **S3:** both the drug and protein in the test set are unseen in the training set.

In the cold-start scenarios, we compared our GAS-DTA model with the models introduced in the previous experiment. It is worth noting that some models, like GLCN-DTA and GSAML-DTA, their papers did not include cold-start scenarios results. Therefore, we did not compare them with our method. All cold-start scenario experimental results are presented in Table 4, where our model outperforms the current best models in almost all metrics for experimental settings S1 and S3, effectively validating the generalization and effectiveness of our proposed GAS-DTA model for cold-start scenarios. In experimental setting S2, our model exhibits a slight performance gap compared to the current best model HGRL-DTA. One possible reason is that in both data sets, the similarity information on proteins is crucial. In this cold-start scenario, our model, GAS-DTA, only extracts drug similarity information, leading to a performance gap. At the same time, MGraphDTA achieved the best results on the MSE metric of the Davis data set in the S2 experiment. One possible reason is that MGraphDTA utilizes a deep graph network, while other models typically use a shallow graph network. The overall structure of the graph can be better obtained through the deep graph networks.

Comparative Experiment to Verify the Role of Similarity Information in Our Model. In the Method chapter, we analyzed how the similarity branch proposed in our study extracts similarity information in principle. However, this branch extracts both sequence information and similarity information, making it challenging to quantify the contribution of each part based solely in principle. Therefore, in this section, we conduct a comparative experiment to address this issue. The comparison models are as follows:

- **GASI-DTA (w/o MSB):** GASI-DTA without Molecular Structure Branch learns similarity information and sequence information from sequence features.
- **GASI-DTA-S (w/o MSB):** The model utilizes sequence branch, as shown in Figure 3, to replace the similarity branch in GASI-DTA (w/o MSB). Hence, it only learns sequence information.
- **ELECTRA-DTA:** This is a sequence-based DTA prediction model from recent years and only learns sequence information. This model has the same architecture as GASI-DTA-S (w/o MSB), both utilizing pretrained models to extract initial sequence features. To achieve more authoritative verification, we use the model for supplementary comparison.

The experimental results are shown in Figure 4, it can be seen that the performance of GASI-DTA-S (w/o MSB) and ELECTRA-DTA is not significantly different in most metrics across the two data sets. However, GASI-DTA (w/o MSB) outperforms other models in all metrics due to the influence of similarity information, especially in the most important performance metric MSE. Compared with GASI-DTA-S (w/o MSB), GASI-DTA (w/o MSB) increased the MSE from 0.234 to 0.166 and from 0.155 to 0.132 on two data sets, respectively. Therefore, this comparative experiment confirms the significance of similarity information in the field of DTA from an experimental perspective.

Comparative Experiment to Investigate the Advantages of Our Similarity Branch. Our model, GASI-DTA, consists of two branches: the molecular structure branch and the similarity branch. In this section, to comprehensively validate the advantages of our proposed similarity branch, we conduct comparative experiments using only the similarity branch against the classical models. These models include

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

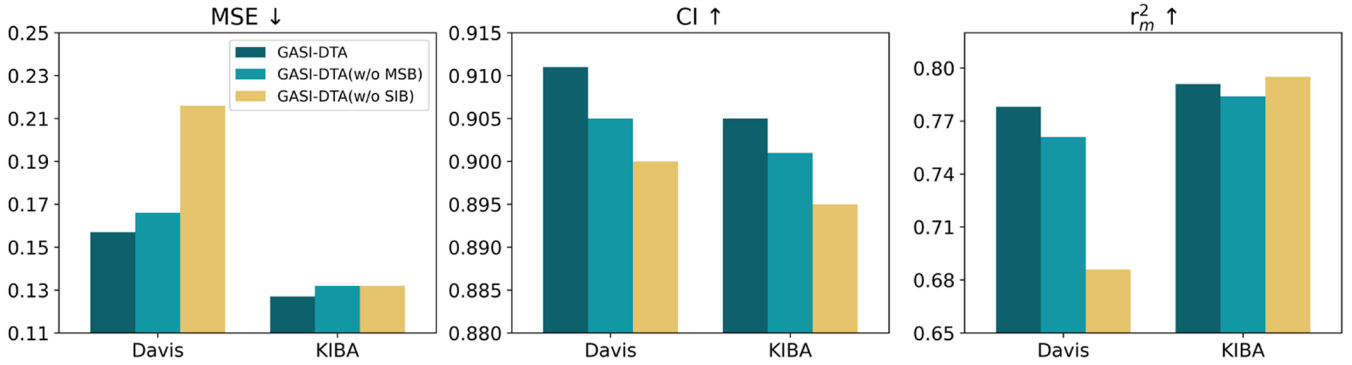


Figure 6. Results of the ablation study for network branch.

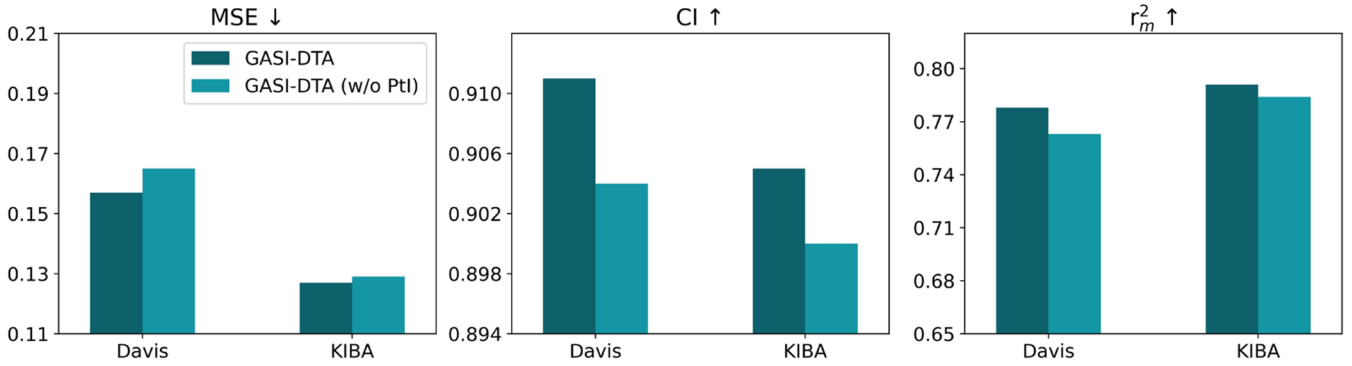


Figure 7. Results of the ablation study for pretrained information.

graph structure-based model DGraphDTA, network-based models BERT-DTA and HGRL-DTA without Molecular Structure Branch(w/o MSB).

As shown in Figure 5, our GASI-DTA (w/o MSB) model outperforms network-based models on almost all metrics across both data sets, effectively demonstrating the advantages of our proposed method. On the Davis data set, our model and network-based models exhibit markedly superior performance compared to the graph structure-based model DGraphDTA. However, on the KIBA data set, DGraphDTA's performance surpasses them. Only our proposed model, GASI-DTA (w/o MSB), achieves comparable performance across various metrics. The occurrence of this phenomenon illustrates a conclusion: graph structure-based models have a bigger dependency on data compared to network-based models and our model.

Ablation Study for Network Branch. To analyze the role of each branch in our model, we conducted an ablation study using the following GASI-DTA variants on two data sets:

- **GASI-DTA (w/o MSB):** GASI-DTA without molecular structure branch learns similarity information and sequence information from sequence features.
- **GASI-DTA (w/o SIB):** GASI-DTA without similarity branch only learns molecular structure information from molecular graphs.

Figure 6 illustrates the performance comparison between GASI-DTA and its variants on two data sets. GASI-DTA outperforms the other variants in nearly all metrics across both data sets. This clearly demonstrates the significant roles played by both the similarity branch and the molecular structure branch in our model. There is no straightforward substitution relationship between similarity-based method and graph structure-based method; instead, they exhibit a complementary relationship. This is because these models capture distinct information from the data set. Through mutual supplementation, better results can be achieved. This also explains the high performance of GASI-DTA across both data sets.

GASI-DTA and GASI-DTA(w/o SIB) exhibit the more significant performance difference in all metrics on the Davis data set, suggesting that similarity branch contributes more to our model on this data set. Conversely, on the KIBA data set, the two variants, GASI-DTA(w/o SIB) and GASI-DTA(w/o MSB), each have advantages in CI and r_m^2 metrics. In the MSE metrics, there is only a very small gap between the two variants, indicating that the contributions of both types of branch are equally important on this data set. It can be seen from the different results of the two data sets that our model GASI-DTA(w/o MSB) does not have high data requirements. Contrarily, GASI-DTA(w/o SIB) heavily relies on data sets. It does not perform well on the Davis data set.

Ablation Study for Pretrained Information. In this section, we conduct an ablation study for the pretrained information. Specifically, we compared the performance of GASI-DTA and GASI-DTA without Pretrained Information (w/o PtI) to investigate the impact of pretraining on our model. GASI-DTA (w/o PtI) utilizes the label encoding method employed by AttentionDTA[25] to acquire initialized sequence features. As shown in Figure 7, GASI-DTA, which utilizes pretrained models to obtain initialized sequence features, outperforms GASI-DTA (w/o PtI) across all metrics on both data sets, demonstrating the importance of pretrained information in our study.

Visualization Analysis. In this section, we designed an additional visualization experiment to explore the specific prediction performance of our model on different data

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

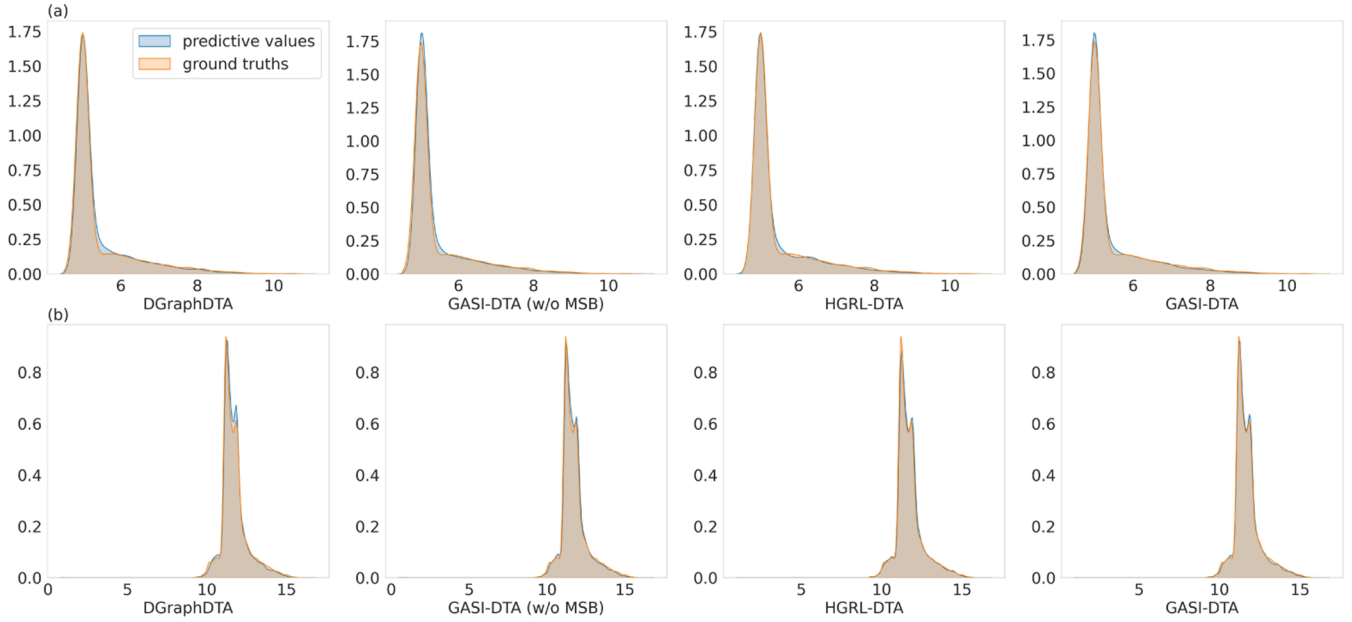


Figure 8. Kernel density estimate plots of affinities between predictive values and ground truths in (a) Davis and (b) KIBA data sets.

distributions and to analyze the characteristics of the model. As shown in Figure 8, we compared our models GASI-DTA and GASI-DTA (w/o MSB) with the network-based model HGRL-DTA[37] and the classical graph structure-based model DGraphDTA.[30] We utilize kernel density plots to visualize the distribution of model predictive values and ground truths. In the kernel density plots, the horizontal axis represents affinity, and the vertical axis indicates the data volume, with higher values representing larger data volumes. The more the two color graphics coincide, the closer the data distribution of predictive values and ground truths. In the Davis data set, there is a peak that represents a large subset of the data, while the rest can be represented as a small subset. It can be observed that in the Davis data set, DGraphDTA shows large deviations for smaller data subsets, while our GASI-DTA (w/o MSB) treats the data equally, exhibiting similar gaps in different distribution.

Because HGRL-DTA integrates interaction information into molecular structure information, enhancing the representation capacity of molecular structure features. Compared with DGraphDTA, it shows more similar distributions in larger data subsets. However, in the smaller data subset where DGraphDTA struggles to make accurate predictions, HGRL-DTA shows less improvement. In contrast, our model integrates the two types of model in a decoupled manner, combining the advantages of both methods, and overall shows more similar distributions in the Davis data set. The characteristics of each model can also be observed from the results of the KIBA data set. Our model GASI-DTA combines the advantages of both methods, while HGRL-DTA represents an optimization of DGraphDTA.

CONCLUSIONS AND FUTURE WORK

Graph structure-based models typically handle drugs or proteins in isolation and only extract the molecular structure information on the drug or protein itself. In order to solve this problem, We introduce drug similarity information and protein similarity information into the field of DTA prediction. Moreover, we propose a network framework that autonomously learns similarity information, avoiding reliance on knowledge graphs. Additionally, based on this framework, we design a new multibranch DTA prediction model, GASI-DTA. The model integrate multiple information. Specifically, it can autonomously extract similarity information and sequence information from pretrained sequence features using the similarity branch. Molecular structure information is extracted from molecular graphs using molecular structure branch.

We offer an interpretation of the model's behavior through theoretical analysis and substantial experimental conclusions, thereby enhancing the model's credibility and comprehensibility. From a theoretical perspective, we utilize BiLSTM as the backbone network for the similarity branch. In BiLSTM, each network unit receives two inputs: the sequence information on the drug/protein and the useful information on other similar drugs/proteins. This approach effectively extracts both sequence and similarity information. From the perspective of experimental conclusions, we designed a comparative experiment between the separate similarity branch and other classic model branches, and a comparison experiment between models with and without similarity information to quantify the contributions of similarity information and sequence information to the model. Additionally, we design visualization experiments to further study the impact of each branch on the model's specific predictive capability. This further explains the model's behavior and fully demonstrates the value of the proposed similarity branch for DTA prediction. Certainly, employing post hoc interpretability, local interpretability, intrinsic interpretability methods, and so on can provide more insight into the model's decision-making process, aiding in understanding the model's behavior on specific instances. The absence of this section is also one of the limitations of this paper. We will delve deeper into the model's interpretability in future research.

In cold-start scenarios, we introduce a combinatorial method to gain partial available similarity information and avoid interference from the misleading information. In most of the cold-start experiments, we have achieved better performance compared to state-of-the-art models. However, our method

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989

also has limitation as it requires separate designs when dealing with cold-start scenarios. In future research, we will continue to delve into the cold-start scenarios problem to design improved methods. This method should utilize partial effective similarity information and circumvent interference from misleading information without separate design. For instance, we are considering incorporating the voting method in ensemble learning into our model to achieve this purpose.

■ ASSOCIATED CONTENT

Data Availability Statement

The source data and code are available as open source at http://github.com/XiaoLin-Yang-S/GASI-DTA.

■ AUTHOR INFORMATION

Corresponding Author

Jing Chen − *School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computing Intelligence, Jiangnan University, Wuxi 214122, China*; Email: chenjing@jiangnan.edu.cn

Authors

Xiaolin Yang − *School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China*; orcid.org/0009-0000-0363-9334

Haoyu Wu − *School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c05607

Author Contributions

J.C.: ideas, supervision, guide and proofread manuscripts. X.Y.: ideas, creation of models, implementation of the computer code and supporting algorithms, writing of main manuscript text, preparation of all the figures and tables. H.W.: guide and proofread manuscripts. All authors read and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Schoolmate Hao Wang, XingDa Shang, WenChao Gu and MingHui Wu for stimulating and useful discussions.

■ REFERENCES

(1) Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239−1249.

(2) Roses, A. D. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat. Rev. Drug Discovery* **2008**, *7*, 807−817.

(3) Yang, Z.; Zhong, W.; Zhao, L.; Chen, C. Y.-C. MGraphDTA: deep multiscale graph neural network for explainable drug−target binding affinity prediction. *Chem. Sci.* **2022**, *13*, 816−833.

(4) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334−395.

(5) Saikia, S.; Bordoloi, M. Molecular docking: challenges, advances and its use in drug discovery perspective. *Curr. Drug Targets* **2019**, *20*, 501−521.

(6) Lee, A.; Lee, K.; Kim, D. Using reverse docking for target identification and its applications for drug discovery. *Expert Opin. Drug Discovery* **2016**, *11*, 707−715.

(7) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129−1143.

(8) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, No. 71.

(9) Wang, K.; Zhou, R.; Li, Y.; Li, M. DeepDTAF: a deep learning method to predict protein-ligand binding affinity. *Briefings Bioinf.* **2021**, *22*, No. bbab072.

(10) Perez-Lopez, C.; Molina, A.; Lozoya, E.; Segarra, V.; Municoy, M.; Guallar, V. Combining machine-learning and molecular-modeling methods for drug-target affinity predictions. *WIREs Comput. Mol. Sci.* **2023**, *13*, No. e1653.

(11) Thafar, M. A.; Alshahrani, M.; Albaradei, S.; Gojobori, T.; Essack, M.; Gao, X. Affinity2Vec: drug-target binding affinity prediction through representation learning, graph mining, and machine learning. *Sci. Rep.* **2022**, *12*, No. 4751.

(12) 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. *J. Cheminf.* **2017**, *9*, No. 24.

(13) 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. *J. Enzyme Inhib. Med. Chem.* **2016**, *31*, 1443−1450.

(14) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

(15) Cao, Y.; Geddes, T. A.; Yang, J. Y. H.; Yang, P. Ensemble deep learning in bioinformatics. *Nat. Mach. Intell.* **2020**, *2*, 500−508.

(16) Mater, A. C.; Coote, M. L. Deep learning in chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545−2559.

(17) Zhang, Y.; Li, S.; Xing, M.; Yuan, Q.; He, H.; Sun, S. Universal approach to de novo drug design for target proteins using deep reinforcement learning. *ACS Omega* **2023**, *8*, 5464−5474.

(18) Zhu, F.; Zhang, X.; Allen, J. E.; Jones, D.; Lightstone, F. C. Binding affinity prediction by pairwise function based on neural network. *J. Chem. Inf. Model.* **2020**, *60*, 2766−2772.

(19) Jones, D.; Kim, H.; Zhang, X.; Zemla, A.; Stevenson, G.; Bennett, W. D.; Kirshner, D.; Wong, S. E.; Lightstone, F. C.; Allen, J. E. Improved protein−ligand binding affinity prediction with structure-based deep fusion inference. *J. Chem. Inf. Model.* **2021**, *61*, 1583−1592.

(20) Wang, Z.; Zheng, L.; Liu, Y.; Qu, Y.; Li, Y.-Q.; Zhao, M.; Mu, Y.; Li, W. OnionNet-2: a convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells. *Front. Chem.* **2021**, *9*, No. 753002.

(21) Shen, C.; Zhang, X.; Deng, Y.; Gao, J.; Wang, D.; Xu, L.; Pan, P.; Hou, T.; Kang, Y. Boosting protein−ligand binding pose prediction and virtual screening based on residue−atom distance likelihood potential and graph transformer. *J. Med. Chem.* **2022**, *65*, 10691−10706.

(22) Jiang, D.; Hsieh, C.-Y.; Wu, Z.; Kang, Y.; Wang, J.; Wang, E.; Liao, B.; Shen, C.; Xu, L.; Wu, J.; et al. Interactiongraphnet: A novel and efficient deep graph representation learning framework for accurate protein−ligand interaction predictions. *J. Med. Chem.* **2021**, *64*, 18209−18232.

(23) Zhang, Y.; Li, S.; Meng, K.; Sun, S. Machine Learning for Sequence and Structure-Based Protein-Ligand Interaction Prediction. *J. Chem. Inf. Model.* **2024**, *64*, 1456−1472.

(24) Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: deep drug−target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821−i829.

(25) Zhao, Q.; Xiao, F.; Yang, M.; Li, Y.; Wang, J. In *AttentionDTA: Prediction of Drug-Target Binding Affinity Using Attention Model*, 2019 IEEE international conference on bioinformatics and biomedicine (BIBM); IEEE, 2019; pp 64−69.

(26) Wang, J.; Wen, N.; Wang, C.; Zhao, L.; Cheng, L. ELECTRA-DTA: a new compound-protein binding affinity prediction model based on the contextualized sequence encoding. *J. Cheminf.* **2022**, *14*, No. 14.

(27) Clark, K.; Luong, M.-T.; Le, Q. V.; Manning, C. D. Pre-Training Text Encoders as Discriminators Rather Than Generators. 2020, arXiv:2003.10555. arXiv.org e-Printarchive. https://arxiv.org/abs/2003.10555.
(28) Hua, Y.; Song, X.; Feng, Z.; Wu, X. MFR-DTA: a multi-functional and robust model for predicting drug-target binding affinity and region. *Bioinformatics* **2023**, *39*, No. btad056.
(29) 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*, 1140−1147.
(30) 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*, 20701−20712.
(31) Michel, M.; Hurtado, D. M.; Elofsson, A. PconsC4: fast, accurate and hassle-free contact predictions. *Bioinformatics* **2019**, *35*, 2677−2679.
(32) Jiang, M.; Wang, S.; Zhang, S.; Zhou, W.; Zhang, Y.; Li, Z. Sequence-based drug-target affinity prediction using weighted graph neural networks. *BMC Genomics* **2022**, *23*, No. 449.
(33) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. 2016, arXiv:1609.02907. arXiv.org e-Printarchive https://arxiv.org/abs/1609.02907.
(34) Qi, H.; Yu, T.; Yu, W.; Liu, C. Drug−target affinity prediction with extended graph learning-convolutional networks. *BMC Bioinf.* **2024**, *25*, No. 75.
(35) Peng, J.; Wang, Y.; Guan, J.; Li, J.; Han, R.; Hao, J.; Wei, Z.; Shang, X. An end-to-end heterogeneous graph representation learning-based framework for drug−target interaction prediction. *Briefings Bioinf.* **2021**, *22*, No. bbaa430.
(36) Lennox, M.; Robertson, N.; Devereux, B. In *Modelling Drug-Target Binding Affinity Using a BERT Based Graph Neural Network*, 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); IEEE, 2021; pp 4348−4353.
(37) Chu, Z.; Huang, F.; Fu, H.; Quan, Y.; Zhou, X.; Liu, S.; Zhang, W. Hierarchical graph representation learning for the prediction of drug-target binding affinity. *Inf. Sci.* **2022**, *613*, 507−523.
(38) Graves, A. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*, Studies in Computational Intelligence; Springer, 2012; Vol. *385*, pp 37−45.
(39) 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*, 4633−4642.
(40) Ahmad, W.; Simon, E.; Chithrananda, S.; Grand, G.; Ramsundar, B. Chemberta-2: Towards Chemical Foundation Models. 2022, arXiv:2209.01712. arXiv.org e-Printarchive. https://arxiv.org/abs/2209.01712.
(41) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated Platform of Small Molecules and Biological Activities. In *Annual Reports in Computational Chemistry*; Elsevier, 2008; Chapter 12, Vol. *4*, pp 217−241.
(42) Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7112−7127.
(43) Landrum, G. RDKit: Open-Source Cheminformatics Software. 2016.
(44) 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. *Nat. Biotechnol.* **2011**, *29*, 1046−1051.
(45) 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*, 735−743.
(46) Liao, J.; Chen, H.; Wei, L.; Wei, L. GSAML-DTA: an interpretable drug-target binding affinity prediction model based on graph neural networks with self-attention mechanism and mutual information. *Comput. Biol. Med.* **2022**, *150*, No. 106145.

https://doi.org/10.1021/acsomega.4c05607
ACS Omega 2024, 9, 35978−35989