Escaping the Drug-Bias Trap: Using Debiasing Design to Improve Interpretability and Generalization of Drug-Target Interaction Prediction

Peidong Zhang, Jianzhu Ma, and Ting Chen

***Abstract*—Considering the high cost associated with determining reaction affinities through in vitro experiments, virtual screening of potential drugs bound to specific protein pockets from vast compounds is critical in AI-assisted drug discovery. Deep-learning approaches have been proposed for predicting Drug-Target Interactions (DTIs). However, they have shown an overestimated accuracy due to the drug-bias trap, a challenge where traditional multimodal models overly rely on the drug branch while underutilizing protein information. This raises doubts about the interpretability and generalizability of existing DTI models. Therefore, we introduce UdanDTI, an innovative deep-learning architecture explicitly designed for predicting drug-protein interactions. UdanDTI applies an unbalanced dual-branch system and an attentive aggregation module to enhance interpretability from a biological perspective. Across various public datasets, UdanDTI demonstrates outstanding performance, outperforming state-of-the-art models under in-domain, cross-domain, and structural interpretability settings. Notably, it demonstrates exceptional accuracy in predicting drug responses of two crucial subgroups of Epidermal Growth Factor Receptor (EGFR) mutations associated with non-small cell lung cancer, consistent with experimental results. Meanwhile, UdanDTI could complement the advanced molecular docking software DiffDock.**

***Index Terms*—Bioinformatics, machine learning, data mining.**

I. Introduction

IDENTIFYING drug-target interactions (DTI) represents a pivotal quest in drug discovery to uncover bioactive compound candidates and their potential interactions with target proteins. While traditional in vitro experimentation remains the gold standard for evaluating drug-target binding affinities [1], its utility is often limited by its exorbitant costs, time-consuming nature, and inefficiencies [2]. Consequently, computer-based virtual screening has emerged as a promising alternative for the rapid prediction of Drug-Target Interactions (DTIs) and elucidation of binding patterns.

Molecular docking is a physics-based virtual screening technique; however, the expansion of molecular space renders docking impractical for every potential drug-target pair. For instance, using commercial software to dock 10 billion candidate drug molecules would require approximately 3000 years [3]. Conversely, machine learning enables DTI models to quickly predict potential affinity from structure-independent raw features [4], [5], [6], [7], [8], [9], [10], [11]. The most popular dual-branch network model encodes proteins and drugs separately using two branches and then merges multimodal information for interaction predictions. In the dual-branch deep learning-based architectures, researchers have explored various encoding modules like Convolutional Neural Networks (CNNs) [12], [13], [14], [15], [16], Graph Neural Networks (GNNs) [17], [18], [19], [20], [21], [22], Deep Neural Networks (DNNs) [23], [24], and Transformers [25], [26], [27], [28], utilizing 1D protein sequences and 2D drug representations to decipher binding patterns. Additionally, incorporating multi-modal information has also improved performance [29], [30]. Recently, Large Language Models (LLMs) in bioinformatics have been introduced to support DTI prediction by encoding molecular properties [31], [32], [33].

Despite these advancements of deep learning methods, generalizability and interpretability remain two significant challenges [15], [34]. Potential series bias in data distribution is the main reason. Earlier work has highlighted this series bias using methods such as cross-domain splitting [15], modeling interactions [25], [28], focusing ligand [26], and introducing void protein experiments [22].

Beyond data issues, we think that the model architecture itself contributes to bias. Multimodal models often face challenges in learning fused knowledge [22], [26], [35]; instead, they tend to predict each modality-specific score independently [36]. Specifically, in multimodal DTI models, there is a tendency to over-rely on the drug modality while underutilizing information from the protein modality, primarily because drug structures are relatively more straightforward and easier to learn. We define this phenomenon as the "drug-bias trap". Such shortcut learning

Received 10 September 2024; revised 28 May 2025; accepted 29 May 2025. Date of publication 4 June 2025; date of current version 8 August 2025. This work was supported in part by the National Key R&D Program of China under Grant 2024YFF1207100, Grant 2024YFF1207103, Grant 2022YFC2703100, and Grant 2022YFC2703105, in part by the Guoqiang Institute of Tsinghua University, and in part by the Beijing National Research Center for Information Science and Technology (BNRist). *(Corresponding authors: Jianzhu Ma; Ting Chen.)*

Peidong Zhang and Ting Chen are with the Institute for Artificial Intelligence and Department of Computer Science and Technology and BNRist, Tsinghua University, Beijing 100190, China (e-mail: zpd24@mails.tsinghua.edu.cn; tingchen@tsinghua.edu.cn).

Jianzhu Ma is with the Institute for AI Industry Research and Department of Electronic Engineering, Tsinghua University, Beijing 100190, China (e-mail: majianzhu@air.tsinghua.edu.cn).

The codes and datasets of UdanDTI are available at https://github.com/CQ-zhang-2016/UdanDTI.

This article has supplementary downloadable material available at https://doi.org/10.1109/TCBBIO.2025.3576488, provided by the authors.

Digital Object Identifier 10.1109/TCBBIO.2025.3576488

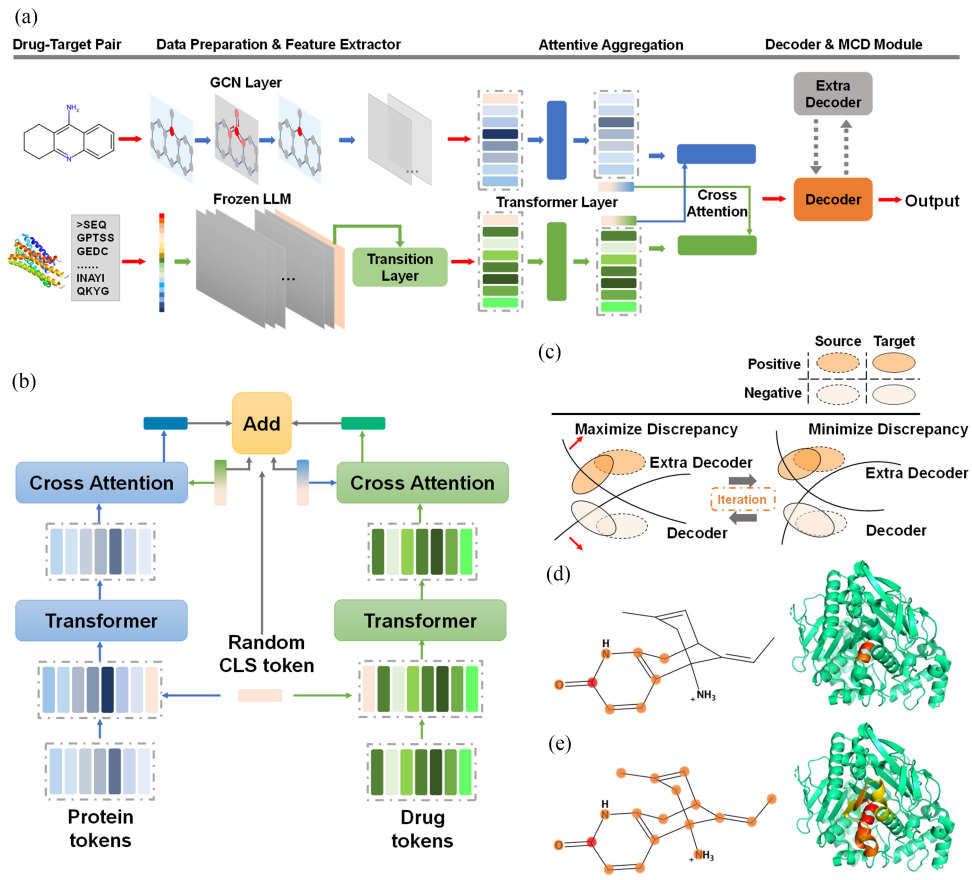

Fig. 1. (a) The overall architecture of UdanDTI. A sampled drug-target pair, represented as a drug 3-D structure and a protein 1-D sequence, sequentially passes through four modules: Data Preparation, Feature Extractor, Attentive Aggregation, and Decoder. If the feature distributions of the test and training data differ significantly, an MCD module is activated with an additional decoder for unsupervised domain adaptation. The final output is a natural number ranging from 0 to 1, indicating the binding potential between the pair. (b) Details of the Attentive Aggregation module. It involves two parallel branches containing a concatenated transformer-cross attention layer. A randomly generated token, incorporated as a head, is added to the representations of the protein and the drug for transformer training. Then, the head tokens from both branches are exchanged to guide the cross-attention training. Five tokens are weighted and summed to smooth the prediction. (c) Concept of the MCD module[43]. Two decoders (hyperplanes for decision-making) are trained to increase discrepancy in target domain data. Subsequently, the front-end network is trained to diminish discrepancy. This adversarial approach pulls target domain samples near the decision boundary into the correct class space. (d) and (e) are illustrations of an unbalanced dual-branch for protein Acetylcholinesterase and drug Huperzine A. (d) A 3-layer network's receptive field adequately captures the crucial functional group pyridine-carbonyl (interaction functional group in Complex 1GPK). However, it only covers a small portion of the acetylcholinesterase helix. (e) A 24-layer network (the averaged depth of LLMs) effectively focuses on the spatial characteristics of the protein pocket but obscures the crucial nodes of the drug molecule.

allows models to achieve superficially high performance on training-like distributions but fails when generalizing, especially when protein information is critical. The prevalent issue of the drug-bias trap is observed in models that overly rely on information from the training set or fall into overfitting due to high protein homology [15], [16], [25], [37]. Further investigations into the drug-bias trap are detailed in Section IV of the Supplementary Information, following the methodology outlined in [22].

To address these enduring challenges, we propose UdanDTI, an unbalanced dual-branch neural network, as a solution to enhance interpretability and generalization in DTI predictions (as illustrated in Fig. 1(a)) for virtual screening. Distinct from previous balanced networks, UdanDTI considers the disparate biochemical differences between proteins and drugs to prioritize unbalanced depths, ensuring equitable attention to both branches. Specifically, UdanDTI uses protein sequences and Simplified Molecular Input Line Entry System (SMILES) strings as input. UdanDTI first captures a high-dimensional representation of the protein using large language models (LLMs) like ESM [38] or ProtBert [39], which are trained with frozen weights and encode a broader range of contextual knowledge from the amino acid sequence. Meanwhile, the drug molecule is represented as a graph, with atoms as nodes and atomic bonds as edges, and then a GCN-based module is used to encode the local features of the drug's spatial structure. Our novel aggregation module, depicted in Fig. 1(b), is specifically designed to capture mutual information effectively. Firstly, two parallel transformer layers are used to assimilate self-information of the drug and the protein, respectively. Two cross-attention layers consider the different modeling scales and facilitate the detection of local interaction patterns through cross-attention weights. The aggregation module ensures that features in every channel of the final output come from both branches, explicitly disrupting the independence of the branches and forcing the model to learn from the fused features. In the end, UdanDTI is augmented with

a model called Maximum Classifier Discrepancy [38] (MCD, as shown in Fig. 1(c)) to facilitate cross-domain adaptation, enhancing robustness and partly alleviating cross-domain interpretability issues.

Comparative evaluations against current state-of-the-art (SOTA) methods across in-domain, cross-domain, and interpretability metrics on independent public datasets confirm UdanDTI's superiority. Moreover, UdanDTI performs well in two main subgroups (on the basis of sensitivity and structural changes) of epidermal growth factor receptor (EGFR) mutations. Additionally, experiments suggest that UdanDTI can also serve as a reliable auxiliary tool in molecular docking tasks.

II. Evaluation Strategies and Metrics

This study applies the proposed model to three public datasets —BindingDB [41], BioSNAP [42], and Human [43] —to conduct a comparative analysis against seven state-of-the-art models. Additionally, drug responses to EGFR mutations [39] are also used as a test set to evaluate the performance of UdanDTI. Detailed descriptions of baseline models and datasets can be found in the Supplementary Information. We apply two types of dataset splitting settings: in-domain and cross-domain. In-domain setting includes random splitting and cold-pair splitting strategies. The random splitting strategy divides the dataset into the training, validation, and testing sets at a ratio of 7:2:1. The cold-pair method reserves 5% and 10% of DTI pairs for the validation and test sets, respectively, and removes all relevant drugs and proteins from the training set.

In cross-domain experiments, we apply a stricter DT-cluster-based splitting strategy to split latent feature clusters of both proteins and drugs in the testing set from those in the training set [15]. Besides, we use three other cluster-based splitting methods: with unseen drugs, where the drugs in the testing set are not present in the training set; one with unseen proteins, where the proteins in the testing set are not present in the training set; and one using sphere exclusion clustering [40], which partitions drug data into clusters. Detailed methodologies of these splitting strategies are elaborated in Section VI of the Supplementary Information.

Additionally, to evaluate interpretability on protein sequences, we first extracted 1125 experimentally determined drug–protein complex structures from the PDB (https://www.rcsb.org/) that also appear in the BindingDB dataset. Because a number of these complexes involve the same protein bound to multiple ligands, they collectively correspond to 1862 distinct drug–protein pairs. We reserved these 1862 pairs as our test set, and randomly split the remaining BindingDB pairs into training and validation sets.

Typically, binary classification problems employ the AUROC (Area Under the Receiver Operating Characteristic Curve) and AUPRC (Area Under the Precision-Recall Curve) as gold-standard metrics for model performance evaluation. Additionally, accuracy, sensitivity, and specificity at the threshold of the best F1 score are also used for comparison purposes. Moreover, we introduce the Top-10-JS metric to evaluate the DTI model's capability to identify actual interactive protein fragments. By defining amino acids within 6 Å of the nearest ligand molecule as protein pockets, this metric calculates the Jaccard coefficient between the model's top-10 most attended residues and related protein pocket residues. A higher Top-10-JS implies that the DTI model is more adept at learning the actual interactive protein fragments in the given sample.

In the end, we conduct five independent runs with different random seeds to ensure stability and robustness. Five-fold cross-validation results and relevant statistical results are also provided in the Supplementary Information.

III Methods

A. Data Preparation

We advocate for the combination of protein embedding from a deep LLM and a drug embedding through shallower networks, thereby establishing an unbalanced dual-branch network. The rationale behind this design lies in the contrasting capabilities of deep and shallow networks. Given that amino acids bound in topological space may not align sequentially, deep LLM offers a broad receptive field capable of encompassing multiple amino acids constituting local spatial substructures. In contrast, the graph structure of a drug molecule explicitly delineates spatial properties, and crucial functional groups could be covered by a relatively small receptive field. Employing shallow networks could prevent overfitting. Illustratively, in Fig. 1(d), the inadequacy of a 3-layer network in extracting topologically bounded amino acids in Acetylcholinesterase is evident, despite its success in capturing the pyridine-carbonyl (the interactive functional group in complex 1GPK) of the Huperzine A drug. Conversely, a 24-layer network (usually the average depth of LLMs) effectively focuses on the spatial properties of proteins but obscures critical nodes of drug molecules (Fig. 1(e)).

To generate protein target features, we utilize a pretrained protein LLM to produce embeddings $E_P \in \mathbb{R}^{m \times d_p}$, where m is the length of a protein and d_p is the preset dimension of the hidden layer. ESM [41] and ProtBert [42] (separately used in distinct downstream tasks) are selected to process the original sequence data. Our input drug molecules are represented using SMILES. We preprocess SMILES into graphs, incorporating node features and adjacency matrices using RDKit. Each node's feature is expressed through one-hot encoding, encompassing the atom type, degree, formal charge, number of radical electrons, hybridization, aromaticity, and the total number of hydrogen atoms for the atom. Consequently, this processing yields a drug ligand embedding, $E_D \in \mathbb{R}^{n \times d_d}$, where n is the length of the drug and d_d is 74. During the training process, the maximum lengths of protein and drug are set to 1200 and 290, respectively. Additionally, we explored the utilization of other protein and drug LLMs; all ablation experiments on both protein and drug LLMs are presented in the Supplementary Information.

B. Feature Extractor

GCNs are potent neural architectures explicitly tailored to process graphs, leveraging their structural information effectively. For the drug compound, we fed the molecular graph E_D

into a three-layer GCN module to capture the information on non-hydrogen heavy atoms and functional groups. The message propagation mechanism of GCN ensures weighted aggregation of information concerning topologically bonded atoms (chemical bonds). The deeper layers employ superposition to broaden the receptive field, accurately extracting and characterizing substructures within small drug molecules. The formulation of the GCN extractor utilizes a convolution operation described as follows:

$$F_D^{i+1} = GCN\ \left(F_D^i,\ A\right) = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}F_D^i W_D^{i+1} + b_D^{i+1}\right) \tag{1}$$

Where $A \in \mathbb{R}^{n\times n}$ is the adjacency matrix of the drug graph (n is the number of the nodes in the graph), I is the identity matrix, and $\hat{A} = A + I$. $\hat{D}$ is the diagonal node degree matrix calculated from A. W_D^{i+1} and b_D^{i+1} are the learnable weight matrix and bias vector of the layer i+1, respectively, and the output is $F_D^{i+1} \in \mathbb{R}^{n\times d}$. It should be noted that $F_D^0 = E_D$, the initial embedding vector of the drug.

Besides, we apply a linear layer to adjust the dimension of protein embedding E_P to $F_P \in \mathbb{R}^{m\times d}$.

C. Attentive Aggregation

The attention mechanism, well-known in Transformer architectures [43] for machine translation, has raised considerable research interest. The mechanism typically comprises three vectors, query vector (q), key vector (k), and value vector (v), to compute interaction values (e.g., dot products or correlation coefficients) between the query and key vectors to weight the summation of the value vectors [44], [45], [46]. A classic multi-head attention mechanism can be represented by (2)–(4).

$$\begin{aligned} &MultiHead\ (Q,\ K,\ V) \\ &\quad = Concat\left(head_1, \ldots,\ head_{num}\right) W^O \end{aligned} \tag{2}$$

$$head_i = Attention\left(QW_i^Q,\ KW_i^K,\ VW_i^V\right) \tag{3}$$

$$Attention\ (q,\ k,\ v) = softmax\left(\frac{qk^T}{\sqrt{d_k}}\right) v \tag{4}$$

Where d_k is the dimension of vectors q and k, num is the preset number of attention heads, W^O, W_i^Q, W_i^K,and W_i^V are trainable parameters.

We have designed a novel attentive aggregation architecture (shown in Fig. 1(b)) specifically aimed at capturing the pairwise interactions between protein and drug substructures. As the feature representations of proteins $F_P$and drugs F_D share the same dimension d, they could be viewed as sentences (token sets) of varying lengths. We first randomly initialize a token $h \in \mathbb{R}^{1\times d}$ as the head of F_P and F_D, respectively, and then define $\hat{F}_P = concat(h,\ F_P) \in \mathbb{R}^{(m+1)\times d}$ and $\hat{F}_D = concat(h,\ F_D) \in \mathbb{R}^{(n+1)\times d}$. In the two parallel branches, $\hat{F}_P$ and $\hat{F}_D$ each is calculated by a distinct 1-layer Transformer. The updated heads are retrieved to serve as the query vectors for calculating cross-attention in the opposite branch. Outputs from four positions are summed with pre-set weights to the original token h. We retain the original head h to ensure the smoothness of model predictions. The mathematical representation of this process is as follows:

Transformer step:

$$\hat{F}_P' = MultiHead\left(\hat{F}_P,\ \hat{F}_P, \hat{F}_P\right) \tag{5}$$

$$\hat{F}_P'' = FFN\ \left(\hat{F}_P'\right) = \sigma\left(\hat{F}_P' W_{P_1} + b_{P_1}\right) W_{P_2} + b_{P_2} \tag{6}$$

$$\hat{F}_D' = MultiHead\left(\hat{F}_D,\ \hat{F}_D, \hat{F}_D\right) \tag{7}$$

$$\hat{F}_D'' = FFN\ \left(\hat{F}_D'\right) = \sigma\left(\hat{F}_D' W_{D_1} + b_{D_1}\right) W_{D_2} + b_{D_2} \tag{8}$$

Where $W_{P\_1}, W_{P\_2}, b_{P\_1}, b_{P\_2}, W_{D\_1}, W_{D\_2}, b_{D\_1}$,and W_{D_2} are trainable parameters.

Cross-attention step:

$$h_{P-2} = MultiHead\left(h_{P-1},\ \hat{F}_{D-h},\ \hat{F}_{D-h}\right) \tag{9}$$

$$h_{D-2} = MultiHead\left(h_{D-1},\ \hat{F}_{P-h}, \hat{F}_{P-h}\right) \tag{10}$$

Where h_{P-1} and $\hat{F}_{P-h}$ are the head token and the rest of the tokens of $\hat{F}_P''$, respectively, and h_{D-1} and $\hat{F}_{D-h}$ are the head token and the rest of the tokens of $\hat{F}_D''$, respectively. Thus we could calculate the final output p^0.

$$\begin{aligned} p^0 = {} & 0.04 * h + 0.08 * h_{P-1} + 0.08 * h_{D-1} \\ & + 0.4 * h_{P-2} + 0.4 * h_{D-2} \end{aligned} \tag{11}$$

D. Decoder and Loss Function

For the calculated $output_{token} \in \mathbb{R}^{1\times d}$, we fed it into a 3-layer multilayer perceptron to compute the interaction probability. The calculation process in each layer can be represented by the following formula.

$$p^{i+1} = \sigma\left(p^i W_o^{i+1} + b_o^{i+1}\right) \tag{12}$$

Where the predictive score $p = sigmoid(p^3)$.

Finally, the cross-entropy loss function is used to minimize the distance between labels and predictions.

E. Cross-Domain Module

As shown in Fig. 1(c), the MCD module introduces an additional decoder to implement adversarial learning while maintaining the core architecture. MCD achieves domain alignment from the target to source domains at a low computational cost, minimizing the impact on model interpretability.

Given a source domain $S_S = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ of $N_s$labeled drug–target pairs and a target domain of $S_T = \{(x_j^t)\}_{j=1}^{N_t}$ of N_t unlabeled drug–target pairs, UdanDTI could be decomposed as two decoders with different initialization methods, F_1 and F_2, and the main architecture G. The MCD module repeats the following three steps iteratively until convergence. In Step 1, G, F_1, and F_2 are trained simultaneously on S_S to ensure the classification accuracy of the model in the source domain. In Step 2, G remains fixed while two diverse classifiers are trained on S_T to widen the discrepancy. Notably, Step 1 is simultaneously

TABLE I
IN-DOMAIN COMPARISON BETWEEN UDANDTI AND SEVEN OTHER ADVANCED DTI MODELS ON THE BINDINGDB AND BIOSNAP DATASETS

Dataset	Method	AUROC	AUPRC	Accuracy	Sensitivity	Specificity
BindingDB	RF	0.942±0.011	0.921±0.016	0.880±0.012	0.875±0.023	0.892±0.020
	DeepConv-DTI	0.945±0.002	0.925±0.005	0.882±0.007	0.873±0.018	0.894±0.009
	GraphDTA	0.951±0.002	0.934±0.002	0.888±0.005	0.882±0.012	0.897±0.008
	MolTrans	0.952±0.002	0.936±0.001	0.887±0.006	0.877±0.016	0.902±0.009
	MGraphDTA	0.951±0.002	0.936±0.002	0.888±0.006	0.881±0.017	0.899±0.011
	MCANet	0.956±0.001	0.940±0.003	0.892±0.003	0.887±0.011	0.904±0.004
	DrugBAN	0.960±0.001	0.948±0.002	0.904±0.004	0.900±0.008	**0.908±0.004**
	UdanDTI	**0.965±0.001**	**0.955±0.001**	**0.911±0.004**	**0.920±0.007**	0.900±0.007
BioSNAP	RF	0.860±0.005	0.886±0.005	0.804±0.005	0.823±0.032	0.786±0.025
	DeepConv-DTI	0.886±0.006	0.890±0.006	0.805±0.009	0.760±0.029	0.851±0.013
	GraphDTA	0.887±0.008	0.890±0.007	0.800±0.007	0.745±0.032	0.854±0.025
	MolTrans	0.895±0.004	0.897±0.005	0.825±0.010	0.818±0.031	0.831±0.013
	MGraphDTA	0.896±0.006	0.899±0.004	0.811±0.012	0.785±0.035	0.856±0.018
	MCANet	0.901±0.006	0.899±0.005	0.827±0.009	0.810±0.035	0.827±0.011
	DrugBAN	0.903±0.005	0.902±0.004	0.834±0.008	0.820±0.021	0.847±0.010
	UdanDTI	**0.941±0.003**	**0.942±0.004**	**0.876±0.006**	**0.876±0.016**	**0.877±0.009**

included to restrict the direction of the decoder hyperplane changes. In Step 3, parameters in classifiers are kept constant and G is trained on S_T to reduce divergence (samples with low confidence) between them by extracting improved features. Through this alternating process of divergence and compromise, UdanDTI$_{\mathrm{MCD}}$ enables G to learn domain-invariant features that are robust to different decision boundaries. Finally, we chose the converged model obtained after 200 epochs.

IV. RESULTS

A. In-Domain Experiment

In the in-domain experimental setup, we conduct a comparative analysis against seven state-of-the-art models. We obtained results for Random Forest (RF), DeepConv-DTI, GraphDTA, MolTrans, and DrugBAN from cited references [15]. Additionally, MGraphDTA and MCANet were replicated and deployed utilizing publicly available codes.

Table I summarizes the comparative results on the BindingDB and BioSNAP datasets. Across all evaluation metrics—AUROC, AUPRC, accuracy, sensitivity, and specificity—UdanDTI consistently demonstrates superior performance over the baseline models. These results suggest that UdanDTI captures crucial information from fused dual-branch features more effectively than learning from data or descriptors based on statistical learning. We also designed ablation experiments on BioSNAP to evaluate the individual contributions of the unbalanced dual-branch and attentive aggregation module (which can be found in the Supplementary Information).

The results on the Human dataset, depicted in Fig. 2(a), indicate that most models exhibit promising performance under random splitting, as measured by both AUROC and AUPRC. However, a previous study [26] has highlighted the risk of overfitting in the Human dataset. To mitigate this risk, we adopted a cold-pair splitting strategy for further evaluation. Fig. 2(a) illustrates a notable decline in the AUROC and AUPRC performance of all models when transitioning from random segmentation to cold-pair splitting. Under this strategy, UdanDTI demonstrates a clearer competitive edge than the other models, consolidating its performance advantage.

B. Cross-Domain Experiment

The in-domain setting may not authentically represent the actual performance of DTI predictive models in prospective forecasting. As previously highlighted, a drug-bias trap in the DTI dataset may lead to accurate predictions solely based on drug characteristics, overlooking actual interaction patterns. The apparent high accuracy may stem from biases toward drugs and overfitting, grounded in protein homology, rather than a model's genuine predictive performance.

In real-world scenarios, the vast chemical space often contains drug-target pairs that are Out-of-Distribution (OOD) with the training data. Employing the same source and target domain setup as in DrugBAN, latent features of proteins and drugs in the test set are explicitly distanced from the feature distribution in the training set through clustering methods. To confront the cross-domain challenges, we activated the MCD module of UdanDTI. Table II presents the performance evaluation of the BindingDB and BioSNAP datasets under cluster-based pair segmentation. All DTI models showed a notable decrease in performance due to the minimized information overlap between the training and test sets. Under the cross-domain setting, vanilla UdanDTI demonstrated robustness, exhibiting AUROC values 6% and 10% higher than DrugBAN on the BioSNAP and BindingDB datasets, respectively. Random Forest (RF) exhibited commendable performance in the cross-domain settings, surpassing some deep learning baselines (DeepConv, GraphDTA, and MolTrans),

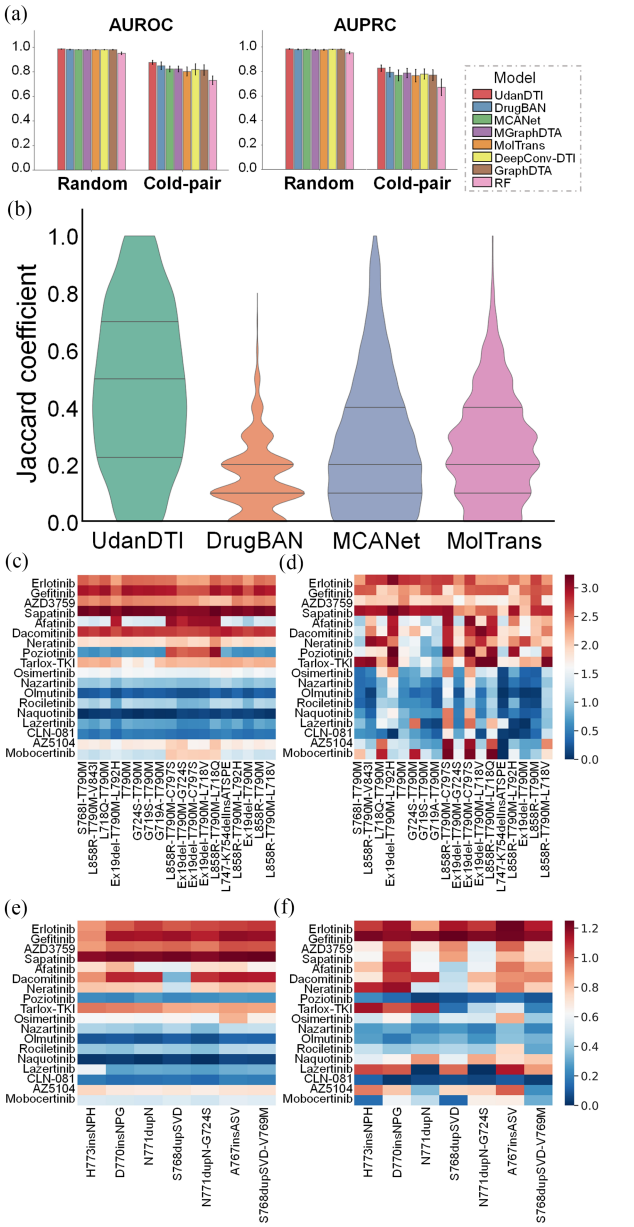

Fig. 2. (a) indicates the comparison between UdanDTI and seven other advanced DTI models on the Human dataset. AUROC and AUPRC performances under random splitting and cold-pair splitting are separately shown. (b) Performance comparison between UdanDTI and other interpretable models under the Top-10 JS metric. (c)–(f) show the drug-protein response values predicted by UdanDTI and measured by in-vitro experiments, including predicted results among T790 M-like subgroup (c) and Ex20ins-L subgroup (e), and in-vitro experimental results among T790 M-like subgroup (d) and Ex20ins-L subgroup (f). Heatmaps in the same row (from the same subgroup) follow the same temperature coefficient.

likely due to its reliance on statistics-based fingerprint features. Meanwhile, deep learning models emphasizing interactive learning (MCANet, DrugBAN, and UdanDTI) exhibited superior predictive performance. Further refinement and tuning of the MCD module improved UdanDTI's generalization capabilities, enhancing AUROC by 8% and 5% on the two datasets, respectively. These outcomes highlight UdanDTI's capability to address real-world challenges. Additionally, we provide an ablation experiment of various cross-domain modules and the attentive aggregation module in the Supplementary Information.

TABLE II
CROSS-DOMAIN COMPARISON BETWEEN UDANDTI AND EIGHT OTHER ADVANCED DTI MODELS ON THE BINDINGDB AND BIOSNAP DATASETS

Method	$\text{AUROC}_{\text{BindingDB}}$	$\text{AUPRC}_{\text{BindingDB}}$	$\text{AUROC}_{\text{BioSNAP}}$	$\text{AUPRC}_{\text{BioSNAP}}$
RF	0.564±0.018	0.503±0.035	0.614±0.027	0.604±0.036
DeepConv-DTI	0.539±0.029	0.474±0.051	0.627±0.021	0.632±0.027
GraphDTA	0.530±0.014	0.467±0.061	0.637±0.011	0.644±0.010
MolTrans	0.536±0.022	0.477±0.063	0.635±0.023	0.629±0.028
MGraphDTA	0.535±0.025	0.476±0.064	0.638±0.032	0.627±0.031
MCANet	0.557±0.020	0.488±0.059	0.651±0.019	0.632±0.026
DrugBAN	0.575±0.025	0.507±0.065	0.654±0.023	0.630±0.026
$\text{DrugBAN}_{\text{CDAN}}$	0.604±0.039	0.556±0.068	0.684±0.026	0.736±0.017
UdanDTI	0.636±0.021	0.571±0.027	0.756±0.013	0.773±0.021
$\text{UdanDTI}_{\text{MCD}}$	**0.713±0.017**	**0.671±0.019**	**0.805±0.011**	**0.825±0.008**

C. Quantitative Comparison of Interpretability on Protein

Addressing branch bias is a common challenge in multi-branch learning frameworks. While some black-box models directly concatenate protein and drug representations to achieve high accuracy, their practical significance is relatively limited. The pervasive drug bias trap has emerged as a substantial obstacle for ML-based DTI models [22], [35]. We highlighted that it is partly a trap caused by the dual-branch network paradigm and the relatively simple structure of drug molecules.

To evaluate the efficiency of models in learning interaction patterns, we utilized a Top-10-JS metric to assess model performance. Although some other methods [47], [48] have been used to measure the interpretability, we choose a more intuitive quantitative comparison: the overlap between the attentive scores predicted by the model and the actual domains of protein pockets. We reproduced models such as MCANet, MolTrans, and DrugBAN based on their respective paper details, trained them on the BindingDB training set, and then evaluated their Top-10-JS values on 1862 DTI pairs with actual structures. The experimental results, displayed in Fig. 2(b), unequivocally demonstrate UdanDTI's superior performance compared to other advanced models. We further compared the frequency of effective hits of different amino acids. The experimental results (presented in the Supplementary Information) demonstrate that UdanDTI is more effective in predicting the interactions between heavy atoms on the side chains of PHE, LEU, MET, and HIS amino acids and ligands. The specially designed attentive aggregation module facilitates an intuitive exploration of the influence exerted by the protein and drug substructures on the predicted results. By solely applying Large Language Models (LLM) to the protein branch, UdanDTI effectively balances both branches. These results suggest that UdanDTI relatively mitigates the drug-bias trap, and its interpretability is substantiated through quantitative experiments

D. Prediction on T790 M and EX20 Subgroup

EGFR mutations commonly occur in exons 18-21 and are primary driver mutations identified in non-small cell lung cancer

(NSCLC). Researchers categorized EGFR mutants into four subgroups based on sensitivity and structural changes [39]. Predicting interactions between protein mutants and drugs presents an immensely challenging task due to the tendency of EGFR gene mutations to occur predominantly on a single exon or even a single amino acid. Moreover, the public benchmark DTI datasets contain solely the wild EGFR gene and a limited set of early-developed drugs (Afatinib, Erlotinib, Gefitinib, and Osimertinib), further complicating predictions. To facilitate domain transfer learning, we converted the experimental mutant-drug affinities to binary values (0-1) and activated the MCD module, comparing its performance with that of DrugBANCDAN.

Results indicate that in T790 M-like and Ex20ins-L subgroups, where gene mutations significantly modify the protein-binding pocket's structure, resulting in resistance, UdanDTI demonstrated robust performance. UdanDTI achieved an AUROC of 0.83 and an AUPRC of 0.91 in the T790 M-like subgroup, outperforming DrugBANCDAN by 7% and 4%, respectively. In the Ex20ins-L subgroup, UdanDTI attained an AUROC of 0.79 and an AUPRC of 0.86, surpassing DrugBANCDAN by 8% and 9%, respectively.

Fig. 2(c)–(f) presents a visualization comparison between UdanDTI's predictions and the experimental results in the original article [39]. To facilitate comparison, we sorted the response and predicted values of all protein-drug pairs and color-coded the corresponding cells accordingly. Visualization results show that the effect of the drug-bias trap was further reinforced due to the concealment of genetic mutations. For pairs of the same drug and different mutants, DTI models often result in approximate predictions. Notably, UdanDTI successfully predicted a series of exon-20 missing proteins in the T790 M subgroup and a series of S768dup proteins in the Ex20ins-L subgroup, partially escaping the amplified drug bias and achieving accurate predictions. However, challenges persisted in accurately predicting the interactions with drugs Naquotinib, Lazertinib, and AZ5104, primarily due to their complex structures that are located far from the training set in the feature space. This underscores the need for further enhancements in model design and feature representation to capture these complex interactions more effectively.

E. Interpretability With Attention Visualization

UdanDTI, equipped with an attentive aggregation module, explicitly learns the molecular substructure contributions of both proteins and drugs to the final predictions. Visualizing the attention weights provides insight into the molecular and amino acid levels. We categorized the co-crystalized ligands into three groups for analysis: Classical, Discrete fragments, and Complex structure, and visualized the top two predictions for each category.

Classical samples typically contain proteins of moderate length and drugs with simple structures, featuring fewer binding points and relatively loose pockets. For PDB structure 2B17 (DIF compound bound to Phospholipase A2 protein), UdanDTI correctly identified the carboxyl group and 22nd amino acid TYR involved in the binding patterns, and TYR also forms hydrogen bonds with secondary amines, which is considered an essential amino acid. The 30th GLY, which is involved in the arene-H reaction, is also considered necessary. Unfortunately, the 49th ASP, which reacts with carboxyl and two water molecules, has not been correctly identified. For PDB structure 3FY8 (XCF compound bound to Dihydrofolate reductase), UdanDTI accurately identified the amino group and unilateral carbon atoms involved in the reaction at the 2, 4-diaminopyrimidine site. UdanDTI recognizes the amino acids, 5th LEU and 92nd PHE, that react with an amino group on one side and the amino acid binding point at the 22nd TRP, which binds to carbon atoms. However, it ignores the 30th, 109th, and 111th amino acids that react with an amino group on the other side.

Discrete fragment samples refer to pockets of a target protein, often composed of non-sequentially adjacent amino acid fragments. This category challenges DTI models to discern the 3D spatial structure from 1D protein sequence data. We examine the results of UdanDTI in two high-confidence samples in the following. In the PDB structure 3G5K (a complex of Actinonin drug and peptide deformylases), UdanDTI correctly predicted almost all interaction patterns. The protein fragments that form the binding site (51st Val - 57th Gln, 85th Arg - 89th Pro, and 110th Phe - 115th Glu), and the carbonyl and hydroxyl groups involved in the binding reaction are all highlighted in red. Unfortunately, UdanDTI ignored the reaction of the 157th Glu with the amyl carbon atom, possibly because the attention mechanism amplifies those more significant interactions. Similarly, in the PDB structure 2DRC (a complex of Methotrexate drug and Dihydrofolate Reductase), UdanDTI identified multiple protein fragments surrounding the ligand molecule (5th Ile - 9th Ala, 19th Ala - 27th Asp, 44th Arg - 48th Glu, 52th Arg - 57th Arg, 93th Val - 97th Gly, 117th Ala - 121th Gly). It indicated that UdanDTI identified potential pockets based on drug molecules. For specific binding patterns, UdanDTI accurately identified the 52nd Arg and 57th Arg as sidechain donors of benzoyl and glutaric acid, and the 5th Ile and 94th Ile as the backbone acceptors of 2, 4-diaminopteridine. However, the model partially identified the triangular interactions between the 27th Asp, 24-diaminopteridine, and the 113th Thr, exposing the inadequacy of the cross-attention mechanism in the complex response.

Complex structure samples involve target proteins with intricate spatial configurations that may confound DTI models. As shown in Fig. 3, only one reaction between the 276th Leu and the benzoic acid is observed in the PDB structure 4P38 (complex of 21T and Corticosteroid 11-beta-dehydrogenase isozyme 1). UdanDTI highlighted multiple amino acids neighboring ligands, including 276Leu and the carboxyl carbon atom of benzoic acid. In the PDB structure 8F4W (a complex of VFX and N-acetyltransferase Eis), the ligand reacts primarily with amino acids on the 26th Asp-36th Trp protein fragments, which UdanDTI accurately captures. However, UdanDTI performs relatively poorly in the ligand molecules. It only extracts the binding patterns of the alcohol groups on dimethylamino and cyclohexane with their corresponding amino acids, missing the vital influence of methoxyphenyl groups in the interaction. Such polycyclic and benzene structures challenge the ability of GNN.

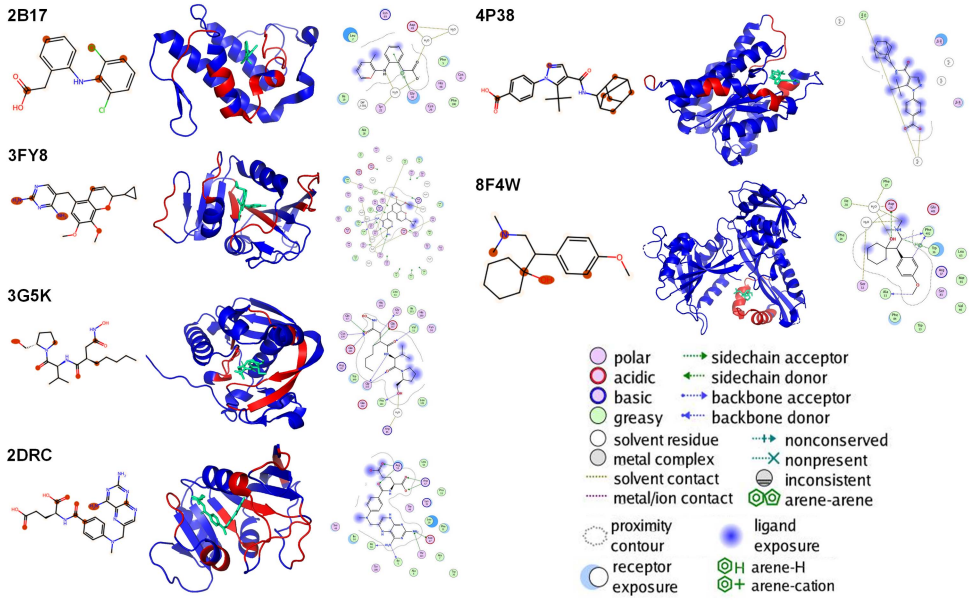

Fig. 3. Visualization results of UdanDTI. For each sample, we visualize the ligand molecule using RDKit and color the top 20% of the atoms that UdanDTI emphasizes in red (left sub-figure). Using PyMol, we illustrate the actual co-crystal structure obtained from the PDB library and highlight the top ten weighted amino acids with their preceding and following amino acids, all in red (middle sub-figure). Finally, we utilize Molecular Operating Environment (MOE) software to visualize the actual ligand molecular reaction diagram as a comparative standard (right sub-figure).

F. Complementary Tool for Docking

Molecular docking is a computational technique used to predict the preferred binding poses of ligands (drugs) to receptors (proteins) under specific conditions. DifDock [49] Innovatively adopts a diffusion model to simulate the uncertainty in molecular binding in real scenarios. It introduces a noise process by sampling positional changes, conformational flips, and bond twists (steps, turns, and twists) in ligand conformation and then trains a denoising model by feeding the altered ligand conformation and protein conformation.

In our experiments, we utilized the pre-trained DiffDock model on 1125 drug-protein pairs with experimentally validated complex structures (detailed in 3.2) to infer the most probable ligand poses (rank_1 provided by the confidence predictor of DiffDock). We further refined this approach by restricting the input to DiffDock to only include amino acids covered by the top 20% of UdanDTI's predictions. This focused input allowed DiffDock to achieve higher accuracy in ligand pose prediction, with the average RMSD improving from 6.55 Å to 5.24 Å.

As shown in Fig. 4, we have chosen two representative examples to illustrate. In complex 5YVC (Structure of CaMKK2 in complex with CKI-012, as shown in Fig. 4(a)), a narrow but precisely targeted perspective enabled DiffDock to better focus on the protein pocket, resulting in more precise ligand pose generation. Excluding the terminal propionic acid, the ligand generated by DiffDock with UdanDTI's assistance exhibited better alignment with the native small molecule. In complex 7X9D (DNMT3B in complex with harmine, as shown in Fig. 4(b)), the amino acid range provided by UdanDTI effectively avoided other potential misleading protein pockets. While the ligand pose was not pixel-perfect, it was accurately positioned within the correct pocket, illustrating UdanDTI's ability to mitigate some of DiffDock's limitations in positional changes.

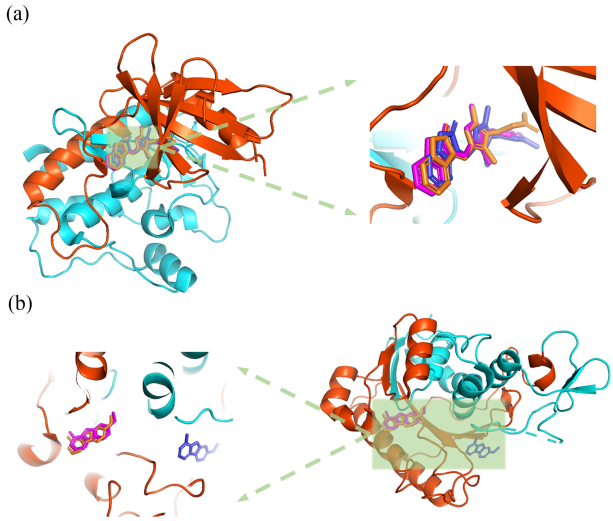

Fig. 4. Two representative examples of UdanDTI-assisted DiffDock docking. The range of amino acids selected by UdanDTI on the protein is highlighted in orange. The ligands inferred by UdanDTI-assisted DiffDock, directly predicted by DiffDock, and experimentally measured were colored in orange, blue, and magenta, respectively. (a) UdanDTI assists DiffDock in inferring a more precise binding pose. (b) UdanDTI assists DiffDock in avoiding other erroneous protein binding pockets.

These results indicate that UdanDTI can supplement docking models like DiffDock in molecular docking tasks. Furthermore, we include comparison experiments with DiffDock on time performance and virtual screening capability (assessed manually) in the Supplementary Information section.

V. Conclusion

This paper introduces UdanDTI, an advanced deep learning architecture designed for predicting DTIs. Through comprehensive experiments, UdanDTI has demonstrated exceptional performance on diverse public datasets, outperforming existing state-of-the-art DTI models across both in-domain and cross-domain evaluations. Furthermore, UdanDTI serves as a crucial auxiliary tool for molecular docking, enhancing the accuracy of ligand docking by identifying active protein fragments.

A significant aspect of our study is the identification and mitigation of the drug-bias trap caused by balanced dual-branch architectures. This trap leads to over-reliance on the drug branch, hindering the interpretability and widespread application of existing DTI models. UdanDTI's distinctive unbalanced dual-branch and attentive aggregation module mitigates the drug-bias trap, providing advanced interpretability from a biological standpoint. This architecture allows for the visualization of molecular interaction patterns, offering valuable insights into drug-protein interactions at both the atomic and amino acid levels. Our work represents the first quantitative comparison of DTI model interpretability on proteins, demonstrating UdanDTI's ability to capture the locally active protein fragments and potential protein pockets.

The introduction of the cross-domain module and LLM significantly enhances UdanDTI's generalization ability. UdanDTI$_{\mathrm{MCD}}$ achieved 11% and 12% higher AUROC than the best DTI model on two public datasets, BindingDB and BioSNAP, under a cross-domain setting, respectively. Additionally, UdanDTI excels in predicting drug responses among two crucial subgroups of EGFR mutants. The flexibility of the UdanDTI framework suggests its potential adaptability to a wide range of biological interaction challenges, including protein-protein interactions.

While our study did not incorporate highly accurate 3D structural protein data, given that only a few known protein sequences have detailed structural information. Recent advancements in protein 3D structure prediction, like AlphaFold [50], show promising strides. However, direct integration of predictive structures into the DTI framework may present challenges. Given the confidence score provided by AlphaFold, a composite framework considering the availability and reliability of predictive structures might offer a more viable solution to address the DTI problem effectively.

Acknowledgment

The funders had no role in the study design, data collection and analysis, the decision to publish, or manuscript preparation.

References

[1] H. Ozturk, A. Ozgur, and E. Ozkirimli, "DeepDTA: Deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.

[2] J. R. Broach and J. Thorner, "Education reforms struggle against Apartheid's legacy," *Nature*, vol. 384, no. 6604, pp. 14–16, 1996.

[3] A. A. Sadybekov et al., "Synthon-based ligand discovery in virtual libraries of over 11 billion compounds," *Nature*, vol. 601, no. 7893, pp. 452–459, 2022.

[4] J. L. Faulon et al., "Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor," *Bioinformatics*, vol. 24, no. 2, pp. 225–233, 2008.

[5] Y. Yamanishi et al., "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[6] L. Jacob and J. P. Vert, "Protein-ligand interaction prediction: An improved chemogenomics approach," *Bioinformatics*, vol. 24, no. 19, pp. 2149–2156, 2008.

[7] M. J. Keiser et al., "Predicting new molecular targets for known drugs," *Nature*, vol. 462, no. 7270, pp. 175–181, 2009.

[8] K. Bleakley and Y. J. B. Yamanishi, "Supervised prediction of drug–target interactions using bipartite local models," *Bioinformatics*, vol. 25, no. 18, pp. 2397–2403, 2009.

[9] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug-target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.

[10] W. Wang, S. Yang, and J. Li, "Drug target predictions based on heterogeneous graph inference," in *Proc. Pacific Symp. Biocomputing*, 2013, vol. 18, pp. 53–64.

[11] Y. Shi et al., "Protein-chemical interaction prediction via kernelized sparse learning SVM," in *Biocomputing 2013*. Singapore: World Scientific, 2013, pp. 41–52.

[12] F. Wan et al., "DeepCPI: A deep learning-based framework for large-scale in silico drug screening," *Genomic. Proteomic. Bioinf.*, vol. 17, no. 5, pp. 478–495, 2019.

[13] Q. Zhao et al., "HyperAttentionDTI: Improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism," *Bioinformatics*, vol. 38, no. 3, pp. 655–662, 2022.

[14] M. Tsubaki, K. Tomii, and J. Sese, "Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," *Bioinformatics*, vol. 35, no. 2, pp. 309–318, 2019.

[15] P. Bai et al., "Interpretable bilinear attention network with domain adaptation improves drug–target prediction," *Nat. Mach. Intell.*, vol. 5, no. 2, pp. 126–136, 2023.

[16] J. Bian et al., "MCANet: Shared-weight-based MultiheadCrossAttention network for drug–target interaction prediction," *Brief. Bioinform.*, vol. 24, no. 2, 2023, Art. no. bbad082.

[17] T. Zhao et al., "Identifying drug-target interactions based on graph convolutional network and deep neural network," *Brief. Bioinform.*, vol. 22, no. 2, pp. 2141–2150, 2021.

[18] Y. Li et al., "Drug-target interaction predication via multi-channel graph neural networks," *Brief. Bioinform.*, vol. 23, no. 1, 2022, Art. no. bbab346.

[19] Y. Wu et al., "BridgeDPI: A novel graph neural network for predicting drug–protein interactions," *Bioinformatics*, vol. 38, no. 9, pp. 2571–2578, 2022.

[20] Y. Hua et al., "CPInformer for efficient and robust compound-protein interaction prediction," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 1, pp. 285–296, 2023.

[21] T. Nguyen et al., "GraphDTA: Predicting drug-target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, 2021.

[22] J. Wang and N. V. Dokholyan, "Yuel: Improving the generalizability of structure-free compound-protein interaction prediction," *J. Chem. Inf. Model*, vol. 62, no. 3, pp. 463–471, 2022.

[23] T. Hinnerichs and R. Hoehndorf, "DTI-Voodoo: Machine learning over interaction networks and ontology-based background knowledge predicts drug-target interactions," *Bioinformatics*, vol. 37, no. 24, pp. 4835–4843, 2021.

[24] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences," *PLoS Comput. Biol.*, vol. 15, no. 6, 2019, Art. no. e1007129.

[25] K. Huang et al., "MolTrans: Molecular interaction transformer for drug-target interaction prediction," *Bioinformatics*, vol. 37, no. 6, pp. 830–836, 2021.

[26] L. Chen et al., "TransformerCPI: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," *Bioinformatics*, vol. 36, no. 16, pp. 4406–4414, 2020.

[27] B. Gao et al., "DrugCLIP: Contrasive protein-molecule representation learning for virtual screening," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 44595–44614.

[28] A. Schulman et al., "Attention-based approach to predict drug–target interactions across seven target superfamilies," *Bioinformatics*, vol. 40, no. 8, 2024, Art. no. btae496.

[29] A. Dehghan et al., "CCL-DTI: Contributing the contrastive loss in drug-target interaction prediction," *BMC Bioinf.*, vol. 25, no. 1, 2024, Art. no. 48.
[30] A. Dehghan et al., "TripletMultiDTI: Multimodal representation learning in drug-target interaction prediction with triplet loss function," *Expert Syst. Appl.*, vol. 232, 2023, Art. no. 120754.
[31] R. Singh et al., "Contrastive learning in protein language space predicts interactions between drugs and protein targets," *Proc. Nat. Acad. Sci. USA*, vol. 120, no. 24, 2023, Art. no. e2220778120.
[32] H. Kang et al., "Fine-tuning of BERT model to accurately predict drug-target interactions," *Pharmaceutics*, vol. 14, no. 8, 2022, Art. no. 1710.
[33] M. A. Thafar et al., "Affinity2Vec: Drug-target binding affinity prediction through representation learning, graph mining, and machine learning," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 4751.
[34] A. Chatterjee et al., "Improving the generalizability of protein-ligand binding predictions with AI-Bind," *Nat. Commun.*, vol. 14, no. 1, 2023, Art. no. 1989.
[35] J. Sieg, F. Flachsenberg, and M. Rarey, "In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening," *J. Chem. Inf. Model*, vol. 59, no. 3, pp. 947–961, 2019.
[36] F. -E. Eid et al., "Systematic auditing is essential to debiasing machine learning in biology," *Commun. Biol.*, vol. 4, no. 1, 2021, Art. no. 183.
[37] Z. Yang et al., "MGraphDTA: Deep multiscale graph neural network for explainable drug-target binding affinity prediction," *Chem. Sci.*, vol. 13, no. 3, pp. 816–833, 2022.
[38] K. Saito et al., "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3723–3732.
[39] J. P. Robichaux et al., "Structure-based classification predicts drug response in EGFR-mutant NSCLC," *Nature*, vol. 597, no. 7878, pp. 732–737, 2021.
[40] J. Simm et al., "Splitting chemical structure data sets for federated privacy-preserving machine learning," *J. Cheminform.*, vol. 13, no. 1, 2021, Art. no. 96.
[41] Z. Lin et al., "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, no. 6637, pp. 1123–1130, 2023.
[42] A. Elnaggar et al., "ProtTrans: Toward Understanding the language of life through self-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7112–7127, Oct. 2022.
[43] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[44] R. Hou et al., "Cross attention network for few-shot classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 4003–4014.
[45] H. Lin et al., "Cat: Cross attention in vision transformer," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2022, pp. 1–6.
[46] C. -F. R. Chen, Q. Fan, and R. Panda, "CrossVit: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 347–356.
[47] R. Rodriguez-Perez and J. Bajorath, "Interpretation of machine learning models using Shapley values: Application to compound potency and multi-target activity predictions," *J. Comput. Aided Mol. Des.*, vol. 34, no. 10, pp. 1013–1026, 2020.
[48] N. R. C. Monteiro et al., "Explainable deep drug-target representations for binding affinity prediction," *BMC Bioinf.*, vol. 23, no. 1, 2022, Art. no. 237.
[49] G. Corso et al., "Diffdock: Diffusion steps, twists, and turns for molecular docking," in *Proc. Int. Conf. Learn. Representations*, 2022.
[50] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

Peidong Zhang received the BSc degree in automation from Shanghai Jiao Tong University, China, in 2020, and the MSc degree in electronic information from Shanghai Jiao Tong University, in 2023. Currently, he is working toward the PhD degree in computer science and technology with Tsinghua University. His research interests include computer-aided drug discovery and deep learning.

Jianzhu Ma received the the PhD degree from Toyota Technological Institute at Chicago in 2016, followed by a position as a project scientist with the University of California, San Diego. He is currently an associate professor with the Institute for AI Industry Research, Tsinghua University, and a recipient of the Overseas Young Talent Program. From 2020 to 2021, he served as a Walther assistant professor with the Department of Computer Science and Biochemistry, Purdue University, and later as an associate professor with Peking University's AI Institute and School of Public Health. His research focuses on artificial intelligence, systems biology, biopharmaceuticals, and smart healthcare. He has received multiple awards, including the Best Paper Award at RECOMB, the Warren DeLano Award at ISMB, and Best Poster Award with the RNA and Protein Folding Conference. He has published in top journals like Nature Methods and Nature Machine Intelligence, with some of his work featured as cover articles.

Ting Chen received the bachelor's degree from Tsinghua University, in 1993, and the PhD degree from Stony Brook University in 1997. He is a professor with the Department of Computer Science and Technology, Tsinghua University. He was a lecturer with Harvard Medical School and later held faculty positions with the University of Southern California, where he directed the Computational Biology Division. His research focuses on Big Data algorithms and machine learning in genomics and proteomics. He has published more than 120 papers in top journals like Cell, Science, and Nature Communications, with more than 10000 citations, and has led projects funded by NSF, NIH, and NSFC. He received the Sloan Research Fellowship in 2004.