



Article

A Length-Adaptive LLM-Enhanced Method for Drug–Target Interaction Prediction

Hengli Zhao ^{1,†}, Yongyi Zhang ^{1,†}, Xinyu Tian ¹, Yilin Chen ¹ and Wei Song ^{1,2,*}

¹ School of Computer Science and Artificial Intelligence, Zhengzhou University, No. 100, Science Avenue, Zhengzhou 450001, China; hengli0650@stu.zzu.edu.cn (H.Z.); zyy_yongyi@stu.zzu.edu.cn (Y.Z.); xinyutian@gs.zzu.edu.cn (X.T.); ylchen@stu.zzu.edu.cn (Y.C.)

² National Supercomputing Center in Zhengzhou, Zhengzhou 450001, China

* Correspondence: iewsong@zzu.edu.cn

† These authors contributed equally to this work.

Abstract

Drug–target interaction (DTI) prediction is central to drug repurposing. Although existing methods integrate multi-dimensional features and have achieved certain progress, they rely heavily on molecular structure or graph structure information. This makes it difficult to capture complex biomedical semantic associations, leading to insufficient semantic information. To address this issue, this paper proposes BioLA-DTI, a large language model (LLM)-enhanced method for drug–target interaction prediction. The method leverages LLMs to provide semantic information for drugs and proteins. Furthermore, existing methods apply uniform weighting strategies to proteins with significantly different lengths. This introduces noise for short proteins and under-represents long ones. To tackle the feature imbalance problem caused by protein-length variation, this work designs a length-adaptive fusion module. This module dynamically weights the LLM-derived semantic features against structural features based on the actual protein-sequence length. Experiments on two benchmark datasets show that BioLA-DTI outperforms existing baseline methods across multiple evaluation metrics, particularly in cold-start scenarios. Case studies demonstrate that this model can provide valuable references for drug discovery. Ablation studies further confirm the critical roles of LLM enhancement and length adaptation modules.

Keywords: drug–target interaction; length-adaptive fusion; large language model; multi-head attention; deep learning

1. Introduction

Drug discovery is a lengthy, costly, and high-risk process [1]. The identification of drug–target interactions is a core bottleneck in this process. Against this background, computational drug–target interaction (DTI) prediction methods have emerged and gained increasing attention. These methods replace certain wet-lab experiments with in silico simulations. They can rapidly screen potential candidate molecules in the early stages of drug discovery. This effectively controls the cost and scale of subsequent experiments [2].

In recent years, deep learning technology has been widely applied in the field of drug discovery. Particularly in drug–target interaction (DTI) prediction, deep neural network-based methods have significantly improved the efficiency of drug repositioning [3–5]. In addition, approaches based on heterogeneous network integration provide complementary strategies for DTI prediction [6]. DTI prediction generally requires the integration



Academic Editor: Bruno Rizzuti

Received: 27 May 2026

Revised: 23 June 2026

Accepted: 24 June 2026

Published: 12 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of multi-source information, including molecular drug sequences, molecular graph structures, and protein amino acid sequences. Traditional methods rely on manual feature engineering. In contrast, deep learning models can automatically extract hierarchical representations from raw data. This advantage has made them the current mainstream technical approaches [7,8].

Deep learning has achieved substantial progress in DTI prediction [9]. However, most existing models adopt vector concatenation or simple summation. They use these operations to fuse multi-modal features of drugs and proteins [10,11]. Such linear fusion strategies struggle to capture the nonlinear correspondence between local chemical substructures of drugs and functional regions of proteins. This limitation hinders the model's ability to learn complex binding patterns [12,13].

Meanwhile, protein sequences exhibit inherent length heterogeneity [14]. Protein length is not merely a measure of molecular size. It is a key physical parameter that determines interaction patterns, functional space, and regulatory complexity [15,16]. In drug design, systematic analysis of protein length helps distinguish differential inhibition strategies between normal and aberrant pathways. This offers a new avenue for improving intervention precision [15]. Protein-length heterogeneity requires predictive models to possess multi-scale representation capabilities. Ignoring this information may compromise the effective capture of cooperative relationships [16]. Therefore, adopting protein length as the physical basis for subsequent weighting and analysis is biologically and computationally well justified. To reconcile data-driven threshold selection with prior knowledge from the literature, the length distribution of proteins in the experimental dataset was first analyzed, as shown in Figures 1b and 2b. Inflection points were observed at approximately 300 and 800 residues. Previous studies have indicated that 300 residues serve as an important statistical boundary for distinguishing short from long proteins in translational regulation [17]. The 800-residue threshold corresponds to the upper limit of the domain-size distribution; proteins exceeding this length often consist of multiple tandem domains [18]. Based on these findings, 300 and 800 residues were adopted as the cutoffs for classifying short and long proteins. Specifically, short proteins (e.g., fewer than 300 residues) typically consist of a single domain or a few compactly arranged domains, with functional sites (e.g., binding pockets and catalytic centers) concentrated along the sequence. For such proteins, local sequence motifs and co-evolutionary signals are often sufficient to encode the key features required for function [19]. In contrast, long proteins (e.g., more than 800 residues) are often composed of multiple domains in tandem, with key residues sparsely distributed among large, non-functional regions. Relying solely on raw sequences makes it difficult to capture long-range residue cooperativity and cross-domain functional coupling [20,21]. Consequently, long proteins require supplementary prior knowledge, such as function-enriched descriptions generated by large language models [22].

Current methods apply uniform feature processing and fusion pipelines to proteins of all lengths. For short proteins, this introduces irrelevant information and causes performance degradation. For long proteins, the lack of external knowledge leads to insufficient representation. The pre-training of large language models (LLMs) on biomedical texts provides new opportunities to introduce external prior knowledge [23]. Generative models such as BioGPT [24] can produce protein descriptions with rich biomedical semantics. Encoder models such as PubMedBERT [25] can provide high-quality semantic representations for protein-related biomedical texts. Both types of models hold promise in terms of filling the knowledge gaps of traditional sequence features. To date, few studies have reported the dynamic adjustment of the fusion weights between external LLM knowledge and underlying sequence or structural features according to protein length.

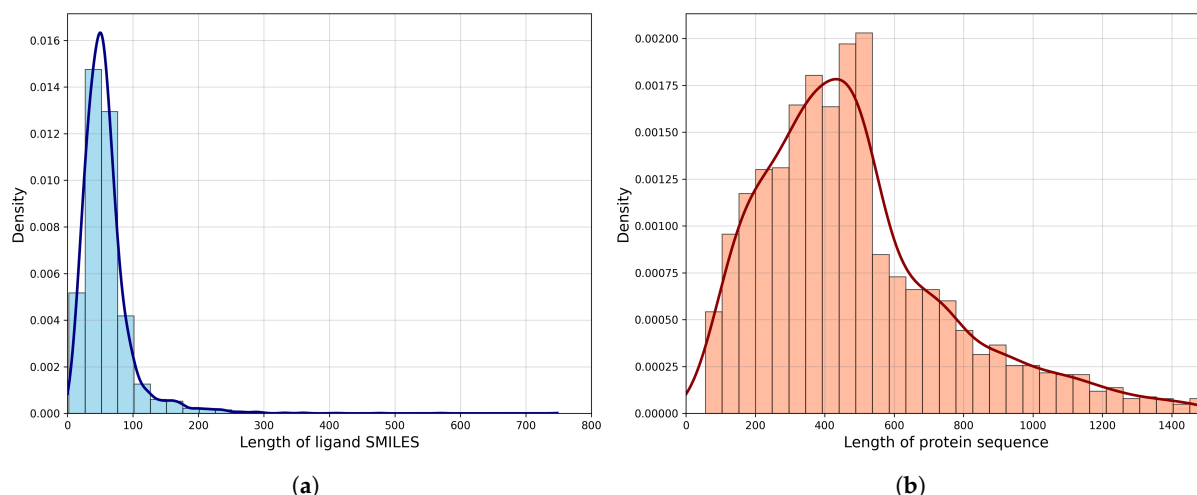


Figure 1. Distributions of ligand SMILES lengths and protein sequence lengths in the BioSNAP dataset. (a) Density distribution of ligand SMILES lengths. (b) Density distribution of protein sequence lengths.

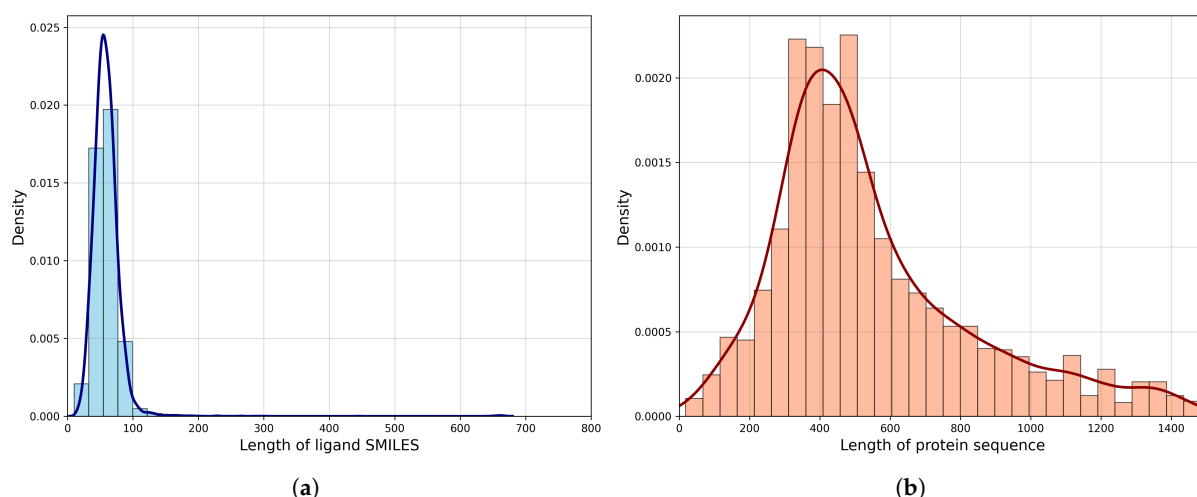


Figure 2. Distributions of ligand SMILES lengths and protein sequence lengths in the BindingDB dataset. (a) Density distribution of ligand SMILES lengths. (b) Density distribution of protein sequence lengths.

To solve these problems systematically, BioLA-DTI, a length-adaptive, LLM-enhanced method, is proposed. As shown in the protein-length density distributions of BindingDB and BioSNAP (Figures 1b and 2b), protein sequences follow a long-tailed distribution with obvious length differences. Most proteins have 200–600 residues, while many are shorter than 300 or longer than 800 residues. Based on this distribution, the model dynamically adjusts fusion weights between semantic and structural LLM features according to protein length. Proteins are divided into three groups: <300, 300–800, and >800 residues, with learnable parameters of ω_1 , ω_2 , and ω_3 for each group. This grouping reflects biological heterogeneity in protein-domain organization and functional site density. Short proteins rely on their own sequence–structural features to avoid noise. Long proteins utilize LLM prior knowledge to compensate for sparse intrinsic information. Moreover, a gating mechanism is adopted for feature enhancement, and learnable multi-head attention is adopted for deep cross-modal drug–protein feature interaction.

A semantically enhanced representation method with length-adaptive fusion is proposed. Comprehensive advantages and strong cold-start performance are validated via the BioSNAP and BindingDB datasets.

2. Results

2.1. Baseline

To evaluate the performance of the proposed method, it is compared with several state-of-the-art baseline models on both the BioSNAP and BindingDB datasets. The compared baseline models include DeepConv-DTI [26], MolTrans [3], TransformerCPI [27], GraphDTI [28], MCANet [29], DrugBAN [30], MGNDTI [31], and MGMA-DTI [32].

2.2. Analysis of Performance

Table 1 presents the performance of all methods on the BioSNAP and BindingDB datasets, where bold numbers denote the optimal results. On the BindingDB dataset, the model achieves the best results in AUROC, recall and F1 score, demonstrating superior prediction accuracy and classification reliability. Although the AUPRC and accuracy of the model are slightly lower than those of MGMA-DTI [32], the performance gap is negligible. More importantly, the cold-start experiment better reflects practical application value and generalization ability (as shown in Figure 3). In this setting, the model significantly outperforms MGMA-DTI [32]. These results verify the effectiveness and practicability of the proposed strategy.

Table 1. Performance comparison of various methods on the BioSNAP and BindingDB datasets.

Dataset	Method	AUROC	AUPRC	Recall	F1 Score	Accuracy
BindingDB	DeepConv-DTI	0.945 ± 0.002	0.925 ± 0.005	0.873 ± 0.006	0.884 ± 0.018	0.882 ± 0.009
	MolTrans	0.952 ± 0.002	0.936 ± 0.001	0.877 ± 0.006	0.886 ± 0.016	0.887 ± 0.009
	TransformerCPI	0.941 ± 0.003	0.925 ± 0.005	0.876 ± 0.007	0.854 ± 0.008	0.875 ± 0.006
	GraphDTI	0.951 ± 0.002	0.934 ± 0.003	0.882 ± 0.012	0.888 ± 0.003	0.888 ± 0.005
	MCANet	0.958 ± 0.001	0.827 ± 0.003	0.889 ± 0.011	0.884 ± 0.002	0.903 ± 0.003
	DrugBAN	0.956 ± 0.001	0.943 ± 0.002	0.892 ± 0.003	0.901 ± 0.004	0.897 ± 0.002
	MGNDTI	0.958 ± 0.002	0.943 ± 0.003	0.899 ± 0.003	0.902 ± 0.004	0.899 ± 0.004
	MGMA-DTI	0.963 ± 0.003	0.951 ± 0.005	0.913 ± 0.011	0.898 ± 0.004	0.914 ± 0.006
	This work	0.964 ± 0.001	0.948 ± 0.002	0.915 ± 0.003	0.910 ± 0.002	0.907 ± 0.001
BioSNAP	DeepConv-DTI	0.886 ± 0.006	0.891 ± 0.006	0.761 ± 0.029	0.782 ± 0.013	0.805 ± 0.009
	MolTrans	0.887 ± 0.004	0.891 ± 0.005	0.755 ± 0.031	0.764 ± 0.023	0.802 ± 0.011
	TransformerCPI	0.897 ± 0.006	0.903 ± 0.009	0.793 ± 0.003	0.802 ± 0.007	0.821 ± 0.007
	GraphDTI	0.891 ± 0.003	0.898 ± 0.004	0.767 ± 0.010	0.802 ± 0.006	0.811 ± 0.006
	MCANet	0.903 ± 0.006	0.773 ± 0.005	0.839 ± 0.006	0.832 ± 0.004	0.829 ± 0.003
	DrugBAN	0.896 ± 0.005	0.901 ± 0.011	0.828 ± 0.011	0.834 ± 0.003	0.834 ± 0.004
	MGNDTI	0.905 ± 0.003	0.908 ± 0.004	0.852 ± 0.007	0.833 ± 0.005	0.829 ± 0.006
	MGMA-DTI	0.909 ± 0.011	0.911 ± 0.008	0.858 ± 0.006	0.857 ± 0.008	0.845 ± 0.003
	This work	0.924 ± 0.002	0.929 ± 0.003	0.876 ± 0.014	0.853 ± 0.004	0.846 ± 0.004

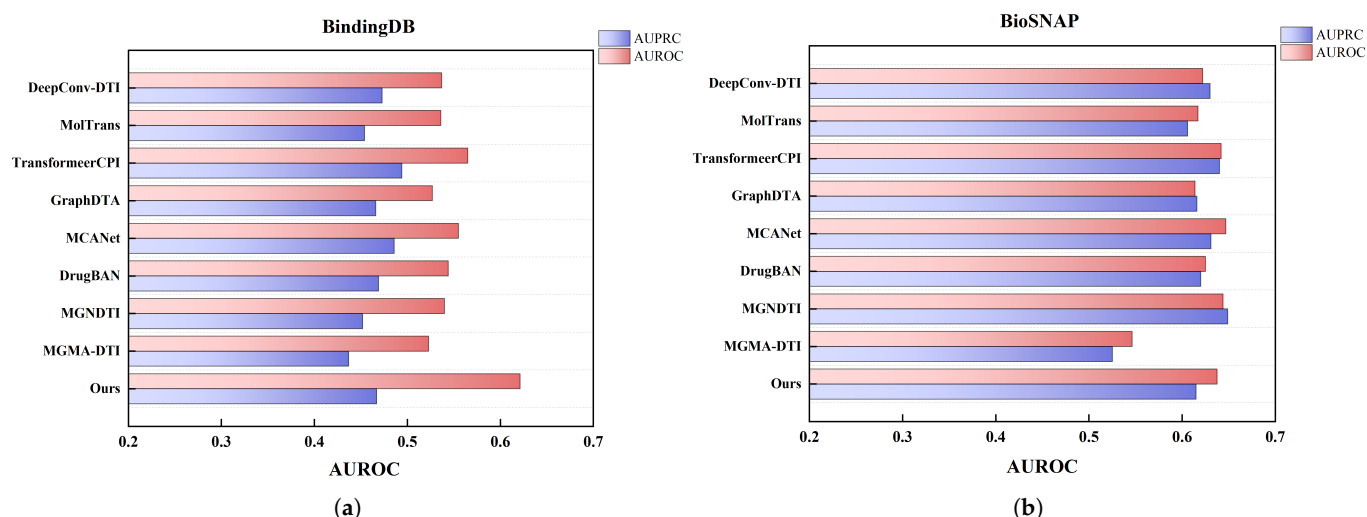


Figure 3. Performance comparison under cold-start scenarios. (a) Results on BindingDB. (b) Results on BioSNAP.

On the BioSNAP dataset, the method outperforms all baseline models across four key evaluation metrics. In particular, the AUROC and AUPRC are improved by more than 1.5 percentage points, showing stronger overall discrimination and prediction capability. Despite a slight gap in the F1 score compared with MGMA-DTI [32], the substantial improvement in comprehensive performance and core evaluation metrics fully demonstrates the superiority of this method.

It should be noted that the performance differences among baseline methods reflect their distinct design priorities. Specifically, MGMA-DTI [32] achieves marginal improvements in AUPRC and accuracy on the BindingDB dataset. This advantage is attributed to its specially designed graph aggregation module, which effectively extracts local substructure features from molecular drug graphs. However, this model does not incorporate pre-trained language models for semantic protein enhancement and lacks external prior semantic knowledge. Consequently, its generalization capability is constrained under cold-start scenarios.

DeepConv-DTI [26] employs local convolution and global max pooling to capture protein binding-site features. MolTrans [3] explicitly models fine-grained associations between drug substructures and protein subsequences via a Co-Attention mechanism, achieving significantly deeper interaction modeling than conventional methods. TransformerCPI [27] captures inter-sequence dependencies based on the Transformer decoder. Although these three methods achieve reasonable benchmark performance across various datasets, none of them introduces pre-trained semantic knowledge. Therefore, their performance upper bound is limited in scenarios where semantic information is scarce.

DrugBAN [30] introduces a domain adaptation mechanism, while MGNDTI [31] adopts a multi-modal gated fusion mechanism. Both methods achieve competitive results across multiple metrics, confirming the beneficial impact of multi-modal fusion on DTI prediction. However, neither of them incorporates pre-trained language models, nor do they explicitly design adaptive fusion strategies targeting protein sequence-length heterogeneity.

The above analysis indicates that existing methods have room for improvement in terms of both semantic enhancement and length adaptation, which are exactly the core motivations of the proposed method. Combining the results on both datasets, the proposed method introduces LLM semantic enhancement with length-adaptive fusion, which can more effectively capture the deep interaction patterns between drugs and targets and exhibits prominent overall superiority.

2.3. Ablation Studies

2.3.1. General Ablation Experiments

Ablation experiments are conducted on the BioSNAP dataset by gradually removing core modules. M1 denotes the model without the gating module, M2 represents the model removing the enhanced LLM component, M3 refers to the model that replaces the multi-head attention in the feature fusion module with direct concatenation, and Full stands for the complete proposed model.

AUROC, AUPRC, recall, F1 score and accuracy are adopted as evaluation metrics, and the experimental results are illustrated in Figure 4a.

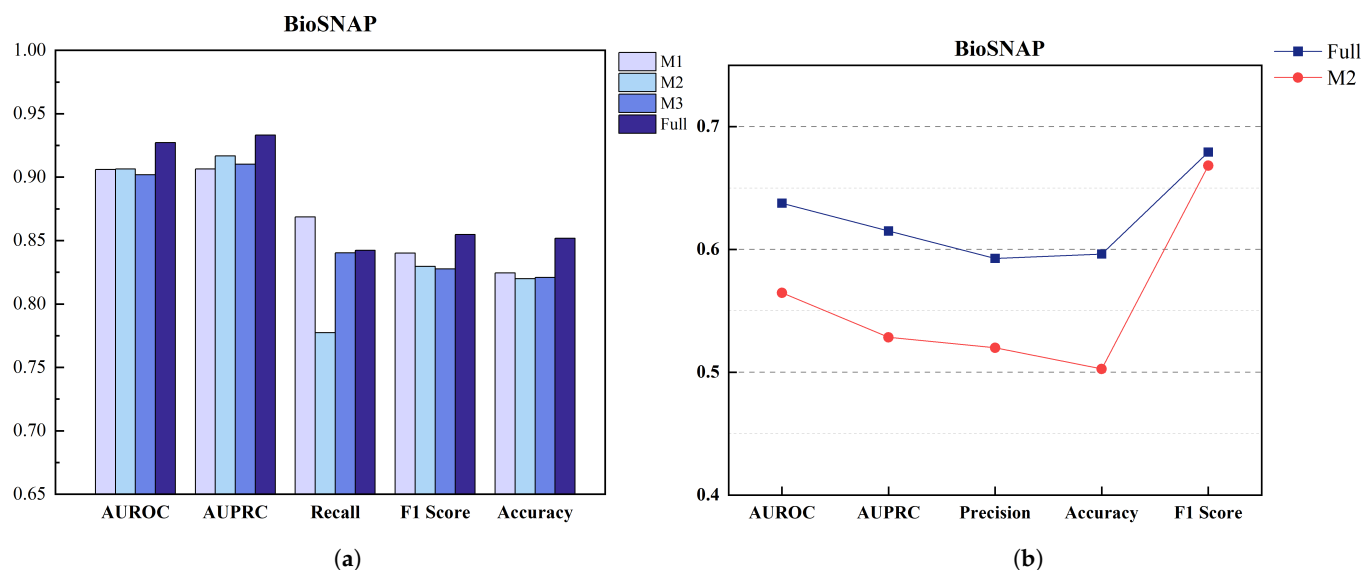


Figure 4. Experimental results on the BioSNAP dataset. (a) Experimental ablation results. (b) Comparison of metrics in cold-start settings.

The ablation results are analyzed as follows:

When Module 1 (M1) is removed, the overall model performance declines slightly, but there remains a clear gap relative to the Full model in core metrics such as AUROC and AUPRC. This indicates that the gating mechanism can effectively enhance feature representation capability and provides steady performance gains for the model.

When Module 2 (M2) is removed, the recall drops most significantly by approximately 0.06 compared with the Full model, accompanied by a noticeable decline in AUPRC, the F1 score and other metrics. This demonstrates that the biomedical semantic enhancement module built based on BioGPT and PubMedBERT is critical for the model's ability to identify positive samples and serves as a key factor in improving the sensitivity of interaction prediction.

Model 3 (M3) is set as an ablation group. The cross-modal multi-head attention fusion module is removed from the full model to evaluate its individual contribution. After removing M3, all model metrics drop moderately. This proves the module can capture implicit drug–protein interactions and improve prediction performance. The performance drop as a result of removing M3 is slighter than that resulting from the removal of the LLM semantic enhancement module. Hence, this fusion module is beneficial but secondary in the model. Previous drug–target interaction studies confirm that cross-attention outperforms simple fusion strategies such as concatenation and element-wise addition [5]. The results support this viewpoint. Most existing works regard cross-modal fusion as the core for interaction modeling. In the framework with LLM semantic pre-enhancement, this attention module only serves as an auxiliary unit for interaction representation learning.

2.3.2. Experimental Validation of Length-Adaptive Weighting

The length-adaptive weighting is compared with a uniform weighting baseline. Uniform weight assigns a fixed value for the fusion of structural and semantic features of proteins and drug molecules. Length-adaptive weight assigns different values based on protein-length intervals. Experiments are conducted on the BindingDB and BioSNAP datasets, with results reported in Table 2 and Table 3, respectively.

Table 2. Ablation study of length-adaptive weighting on BindingDB.

Metrics	Length-Adaptive	Uniform	Improvement (%)
AUROC	0.9649	0.9594	+0.55%
AUPRC	0.9500	0.9446	+0.54%
F1-Score	0.9085	0.8966	+1.19%
Accuracy	0.9051	0.8950	+1.01%
Precision	0.8682	0.8626	+0.56%

Table 3. Ablation study of length-adaptive weighting on BioSNAP.

Metrics	Length-Adaptive	Uniform	Improvement (%)
AUROC	0.9272	0.9251	+0.21%
AUPRC	0.9331	0.9313	+0.18%
F1-Score	0.8548	0.8547	+0.01%
Accuracy	0.8518	0.8405	+1.13%
Precision	0.8471	0.8064	+4.07%

On BindingDB, consistent gains are observed. The F1 score improves by 1.19%, followed by accuracy (+1.01%). These results demonstrate that protein-length-based weight allocation effectively enhances feature fusion. On BioSNAP, length-adaptive weighting achieves better performance across all metrics. Notable improvements include precision (+4.07%) and accuracy (+1.13%).

2.3.3. Cold-Start Ablation Experiments

To further verify the importance of the enhanced LLM module, experiments are conducted on the Full model and the M2 model under the cold-start scenario of the BioSNAP dataset. AUROC, AUPRC, recall, F1 score and accuracy are adopted as evaluation metrics, and the results are presented in Figure 4b.

In particular, the Full model outperforms the M2 model by approximately 8 to 10 percentage points in AUROC and AUPRC, with the gap in accuracy reaching nearly 10 percentage points. This fully demonstrates that the enhanced LLM module can provide effective biomedical semantic information for the model, enabling it to maintain stable and reliable prediction performance when facing unknown drug–protein pairs.

Cold-start experiments show that BioLA-DTI can effectively predict proteins and drugs not present in the training set. In practice, many new targets (e.g., orphan receptors) and new compounds lack known interaction records [33]. However, their sequences, annotations, and molecular structures can be obtained from existing databases. Relying solely on this information for novel drug–target pair prediction has limited reliability [34]. To address this, BioLA-DTI combines textual protein features generated by BioGPT, protein sequence features, and structural drug features. This assists in predicting potential interactions for low-resource drug–target pairs. Compared with traditional methods, the model reduces dependence on known pairs in the training set. It is suitable for initial screening of new targets and new drugs, as well as drug repurposing for rare diseases.

2.4. Cold-Start Experiment

The performance comparison of all baseline methods under the cold-start scenario on the BindingDB dataset is shown in Figure 3a, with AUROC and AUPRC adopted as evaluation metrics. In the BindingDB cold-start setting, the proposed method significantly outperforms all comparison models in terms of AUROC, indicating that the model possesses an outstanding discriminative ability to distinguish interactive from non-interactive unknown drug–target pairs.

Meanwhile, it also achieves competitive AUPRC performance, which further verifies the model's capability of accurately identifying positive samples.

The performance of all comparison methods in the cold-start scenario of the BioSNAP dataset is shown in Figure 3b. AUROC and AUPRC are used as evaluation metrics.

In the BioSNAP cold-start setting, the proposed method achieves relatively good performance in both AUROC and AUPRC, ranking among the top performers. It maintains stable and competitive discriminative performance and exhibits better robustness and adaptability. These results suggest that BioLA-DTI can serve as an effective preliminary screening tool for low-resource targets, reducing the cost of experimental validation in drug repurposing pipelines.

2.5. Visualization of Predicted Probability Distribution

The predicted probability distributions of positive and negative samples of the proposed method on the BindingDB and BioSNAP test sets are shown in Figure 5a and Figure 5b, respectively. The y-axis shows probability density, normalized separately for each class (positive/negative), not over the entire dataset. Thus, each class histogram integrates to 1. This within-class normalization highlights distributional shapes despite class imbalance. Sample sizes are given in the legend. The predicted probabilities of negative samples are highly concentrated near 0, while those of positive samples cluster closely around 1. The two types of samples are almost completely separated in the probability space, with no obvious sample distribution in the transitional interval.

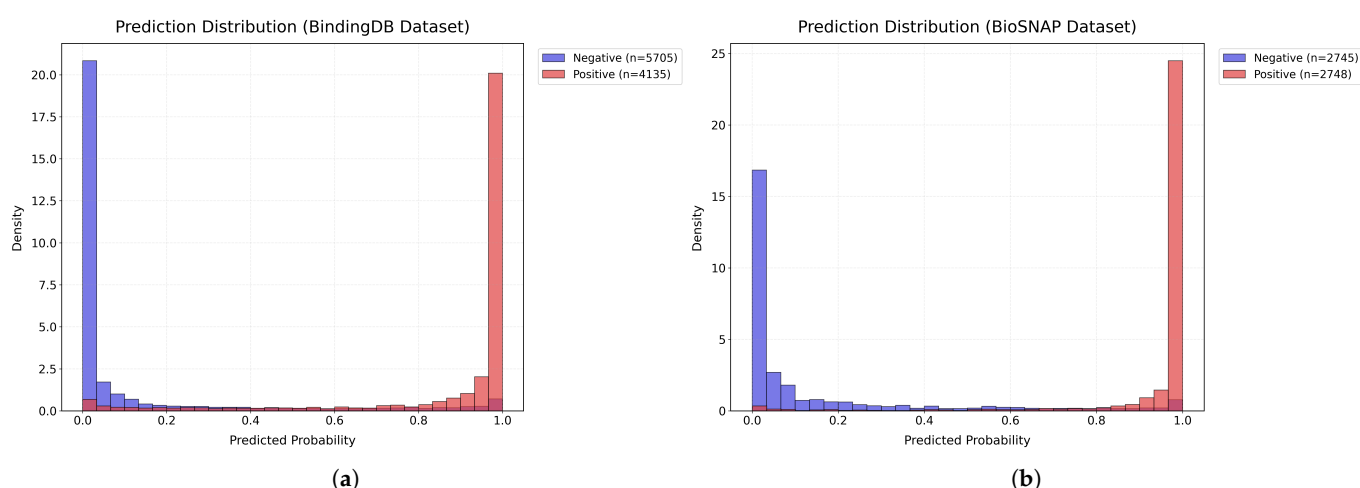


Figure 5. Prediction distribution for positive/negative samples. (a) Results on BindingDB. (b) Results on BioSNAP.

The results demonstrate that the DTI prediction model proposed in this work possesses an extremely strong ability to distinguish positive from negative samples on both the BindingDB and BioSNAP datasets. It achieves high confidence in prediction results, with a clear decision boundary and superior overall model performance.

3. Discussion

Predicting whether drug molecules can bind effectively with specific proteins is a core task in computer-aided drug design (CADD). In recent years, deep learning has made remarkable progress in this field due to its powerful feature extraction capability. TC-DTA [10], HyperAttentionDTI [35] and DeepPurpose [36] treat drug SMILES strings and protein amino acid sequences as 1D sequences, representing typical unimodal methods. They adopt 1D convolutional neural networks or recurrent neural networks for encoding. These methods extract local patterns and sequential dependencies to predict binding affinity. Sequence-based approaches have simple inputs and no need for complex structural computation. They are also easy to implement. Nevertheless, 1D sequences inevitably lose the spatial topology of drug molecules and the relative positions of functional groups. This is an intrinsic property of such representations because SMILES strings and amino acid sequences are inherently linear projections of three-dimensional molecular structures. They also have limited capacity to capture long-range dependencies among distant amino acid residues in proteins. Such dependencies are critical for protein folding and binding-site formation.

To compensate for the loss of structural information in unimodal methods, researchers have introduced graph neural networks. Such methods have found extensive applications in both biomedical and logical reasoning tasks [37]. They model drug molecules as topological graphs while retaining protein-sequence representations. This marks an initial step toward multi-modal modeling. Specifically, MGraphDTA [11], GUNE-DTA [38] and GraphCPI [39] treat atoms as nodes and chemical bonds as edges. They adopt graph convolutional networks to extract spatial features, significantly enhancing the perception of stereochemical drug properties. However, these early bimodal methods still adopt relatively simple fusion strategies. They typically concatenate drug graph features with protein-sequence features before feeding them into the prediction layer, lacking explicit modeling of deep interactions between the two modalities. Moreover, these methods still require the explicit construction of graph structures from protein sequences. The scale of such graphs grows quadratically with sequence length, leading to high computational complexity and limiting their applicability to long-chain proteins.

To overcome the limitation of simple concatenation in bimodal methods, researchers have introduced attention mechanisms. These mechanisms enable more fine-grained cross-modal interaction. Methods such as DrugBAN [30], iNGNN-DTI [40] and MIFAM-DTI [41] retain drug graph features and protein-sequence features. They employ bidirectional attention or cross-attention modules. These modules explicitly model local interactions between drugs and proteins. Compared to simple concatenation approaches, these methods achieve superior performance on multiple public benchmarks. However, existing fusion strategies all follow a static fusion paradigm. This holds true for both early simple concatenation and fine-grained cross-attention. Under this paradigm, identical fusion parameters are used for all protein samples. The impact of protein sequence-length heterogeneity is not considered. In fact, short-chain proteins have concentrated functional sites. Their sequences already contain sufficient discriminative information. In contrast, long-chain proteins contain multiple complex structural domains. Their key functional sites are sparsely distributed. Thus, they require rich contextual knowledge. Static fusion cannot adapt to these two distinct scenarios [42]. It may introduce redundant noise for short proteins. It also suffers from information insufficiency for long proteins. Meanwhile, large language models (e.g., BioGPT and PubMedBERT) can generate protein descriptions rich in biomedical semantics, providing a new avenue for the supplementation of missing knowledge for long-chain proteins. However, existing work simply concatenates such semantic features with struc-

tural features, lacking a mechanism to dynamically adjust the fusion intensity of LLM knowledge according to protein length.

Existing methods are capable of extracting structural features and association information from drugs and proteins, yet they suffer from two notable deficiencies. Primarily, most models fail to account for the heterogeneity in protein sequence lengths. Moreover, these models encounter difficulties in adaptively integrating knowledge from large language models. Such limitations consequently impair the performance improvement afforded by multi-modal fusion in prediction tasks. To address the aforementioned common shortcomings, a length-adaptive large language model-enhanced framework for drug–target interaction (DTI) classification is proposed. Using the BioSNAP and BindingDB datasets, the proposed method is evaluated via quantitative comparisons, ablation studies, and probability distribution visualizations. All quantitative metrics consistently demonstrate the effectiveness of the designed dynamic fusion strategy. To validate the model’s positive predictions biologically, this study designs a visual case analysis based on molecular conformations. The next section selects two groups of drug–target pairs with high-confidence positive predictions as research cases for interpretable analysis via molecular docking visualization.

Case Studies

To further evaluate the interpretability and practical drug discovery value of BioLA-DTI, two drug–target pairs with high positive prediction confidence are selected for case analysis: mefenamic acid (DB00784) and its target, Q04828, as well as CID 155903259 and its target, UniProt ID P0DTD1 (Figures 6 and 7). Among them, red represents the target protein, cyan represents the drug molecule, and blue indicates hydrogen bond interactions. In Figure 6, mefenamic acid forms multiple hydrogen bonds with several polar residues inside the pocket of target protein Q04828. In Figure 7, Nirmatrelvir establishes hydrogen-bond interactions with residues within the active region of the main SARS-CoV-2 protease. AutoDock Vina (1.5.7) is adopted to independently calculate the 3D binding conformations between small molecules and target proteins. Notably, the presented 3D binding diagrams serve only as visual aids to illustrate potential interaction patterns between ligands and target proteins.

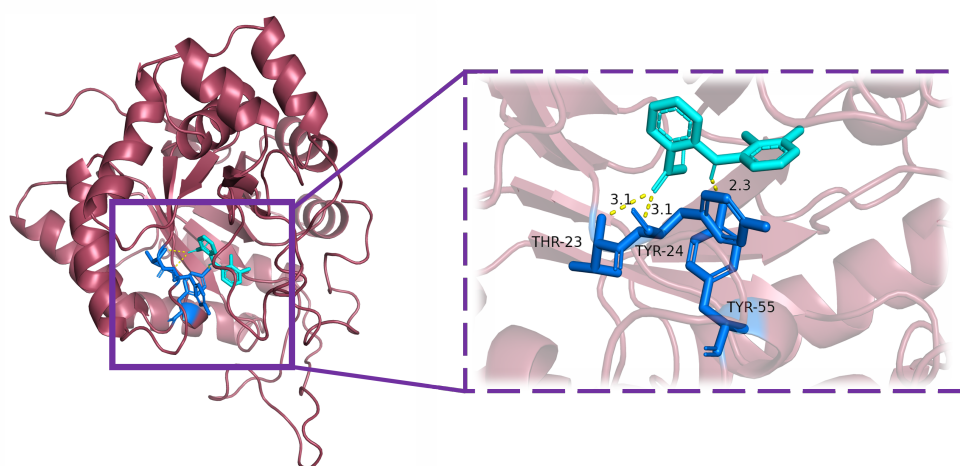


Figure 6. 3D complex structure obtained by molecular docking of Mefenamic acid with Q04828.

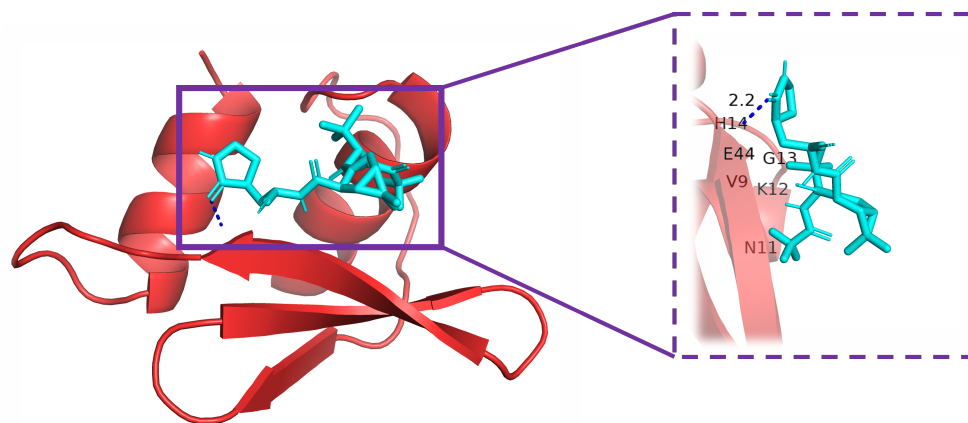


Figure 7. 3D complex structure obtained by molecular docking of Nirmatrelvir with the main SARS-CoV-2 protease.

The two high-confidence positive prediction cases, combined with the docking-visualized binding modes, demonstrate that BioLA-DTI can serve as an effective front-end filtering tool in virtual screening pipelines. It can rapidly screen out high-probability positive candidate molecules from large compound libraries. This significantly reduces the search space for subsequent computationally expensive molecular docking or molecular dynamics simulations. Ultimately, it improves the overall efficiency of the virtual drug screening workflow.

4. Materials and Methods

4.1. Datasets

Widely adopted public biomedical benchmark datasets are utilized for model validation to guarantee reliable experimental results and impartial evaluation. Detailed information about the datasets is presented as follows.

BioSNAP [43] is derived from the Stanford BioSNAP biomedical network and constructed based on the DrugBank database. It contains experimentally validated interaction pairs between approved drugs and target proteins. This dataset is fully balanced with a positive-to-negative sample ratio of 1:1. As listed in Table 4, it includes 27,464 interaction samples, covering 4510 drugs and 2181 proteins. Figure 1a,b show the density distributions of ligand SMILES lengths and protein sequence lengths in the BioSNAP dataset. In this dataset, drug SMILES strings exhibit a unimodal distribution, with most lengths concentrated between 20 and 100 characters. Protein sequences follow a long-tailed distribution, with most lengths ranging from 200 to 600 residues and a notable number exceeding 1000 residues. This significant variation in protein length further motivates the length-adaptive fusion module.

Table 4. Statistics of datasets.

Dataset	Drug	Protein	Interaction	Positive	Negative
BindingDB	14,643	2623	49,199	20,674	28,528
BioSNAP	4510	2181	27,464	13,830	13,634

BindingDB [44] collects compound-target binding affinity data validated by literature and patents. This study uses its standard low-bias preprocessed version. The samples are labeled according to the binding affinity: samples with $K_d \leq 30$ nM are defined as positive, and those with $K_d > 30$ nM are defined as negative. The dataset is slightly imbalanced,

with a positive–negative ratio of approximately 5:7. It contains 49,199 interaction samples involving 14,643 drugs and 2623 proteins, as listed in Table 4. Figure 2a,b show the density distributions of ligand SMILES lengths and protein sequence lengths in the BindingDB dataset. In this dataset, most drug SMILES strings range from 20 to 80 characters, while protein sequences exhibit a long-tailed distribution with large variations in length. This variation motivates our introduction of the length-adaptive fusion module.

4.2. Evaluation Metrics and Implementation

4.2.1. Evaluation Metrics

This paper adopts multiple mainstream metrics to comprehensively evaluate the model's drug–target interaction prediction performance. These metrics include AUROC, AUPRC, accuracy, precision, recall, and F1 score. Among them, AUROC measures the model's overall ability to distinguish between positive and negative samples across different thresholds. Values closer to 1 indicate stronger discriminative ability. AUPRC reflects the trade-off between precision and recall and is more sensitive to the identification performance of the minority class in imbalanced data scenarios. Accuracy is the proportion of correctly predicted samples among all samples, reflecting the overall correctness of the model's predictions. Precision is the proportion of true-positive samples among all samples predicted as positive, measuring the exactness of the predictions. Recall is the proportion of actual positive samples correctly identified by the model, measuring the completeness of the model in capturing positive samples. The F1-score is the harmonic mean of precision and recall, providing a comprehensive reflection of both metrics, and is commonly used for holistic evaluation in imbalanced data scenarios [31,45]. The definitions of each metric are presented as follows.

$$\text{AUROC} = \int_0^1 \text{TPR}(k) d(\text{FPR}(k)) \quad (1)$$

$$\text{AUPRC} = \int_0^1 \text{Precision}(k) d(\text{Recall}(k)) \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Specifically, *TP* (True Positives) is the number of true-positive samples. *TN* (True Negatives) is the number of true-negative samples. *FP* (False Positives) is the number of false-positive samples, and *FN* (False Negatives) is the number of false-negative samples. All metrics range from 0 to 1, and higher values indicate better model performance.

4.2.2. Implementation

To ensure experimental reproducibility and the reliability of results, standardized experimental parameters and training strategies are adopted in this work. The detailed settings are presented as follows:

- Training setup: 5-fold cross-validation is employed with a maximum of 100 training epochs. Early stopping is adopted with patience set to 15, monitoring the AUROC on the validation set. The weight decay is set to 1×10^{-4} , and the ReduceLROnPlateau learning rate scheduler is used with a decay factor of 0.5 and patience of 10.

- LLM configuration: BioGPT (microsoft/BioGPT) is deployed locally. PubMedBERT is used to encode descriptive text, and the projection layers are set to be trainable.
- Cold-start setting: An independent test set is strictly constructed with no overlapping drugs or proteins between the training and test sets. The test set contains drugs and proteins that are completely unseen during training. This rigorous setting enables a fair evaluation of the model's generalization ability for unknown drugs and proteins. AUROC and AUPRC are adopted as evaluation metrics.

4.3. Method

The overall framework of the proposed method is shown in Figure 8. It consists of a feature extraction module, a multi-modal feature fusion module and a classification module. The feature extraction module constructs multi-modal and multi-level representations for drugs and proteins. It includes three complementary perspectives: molecular graph, sequence, and LLM semantic information. The multi-modal feature fusion module adopts a length-adaptive weight mechanism. It achieves adaptive fusion of features for individual entities. Meanwhile, it employs multi-head attention to model deep interactions between drug and protein features. The adaptive weight mechanism is illustrated in Figure 9. The classification module takes the fused interactive features as input. It uses a multi-layer perceptron to finally predict the probability of drug–target interaction.

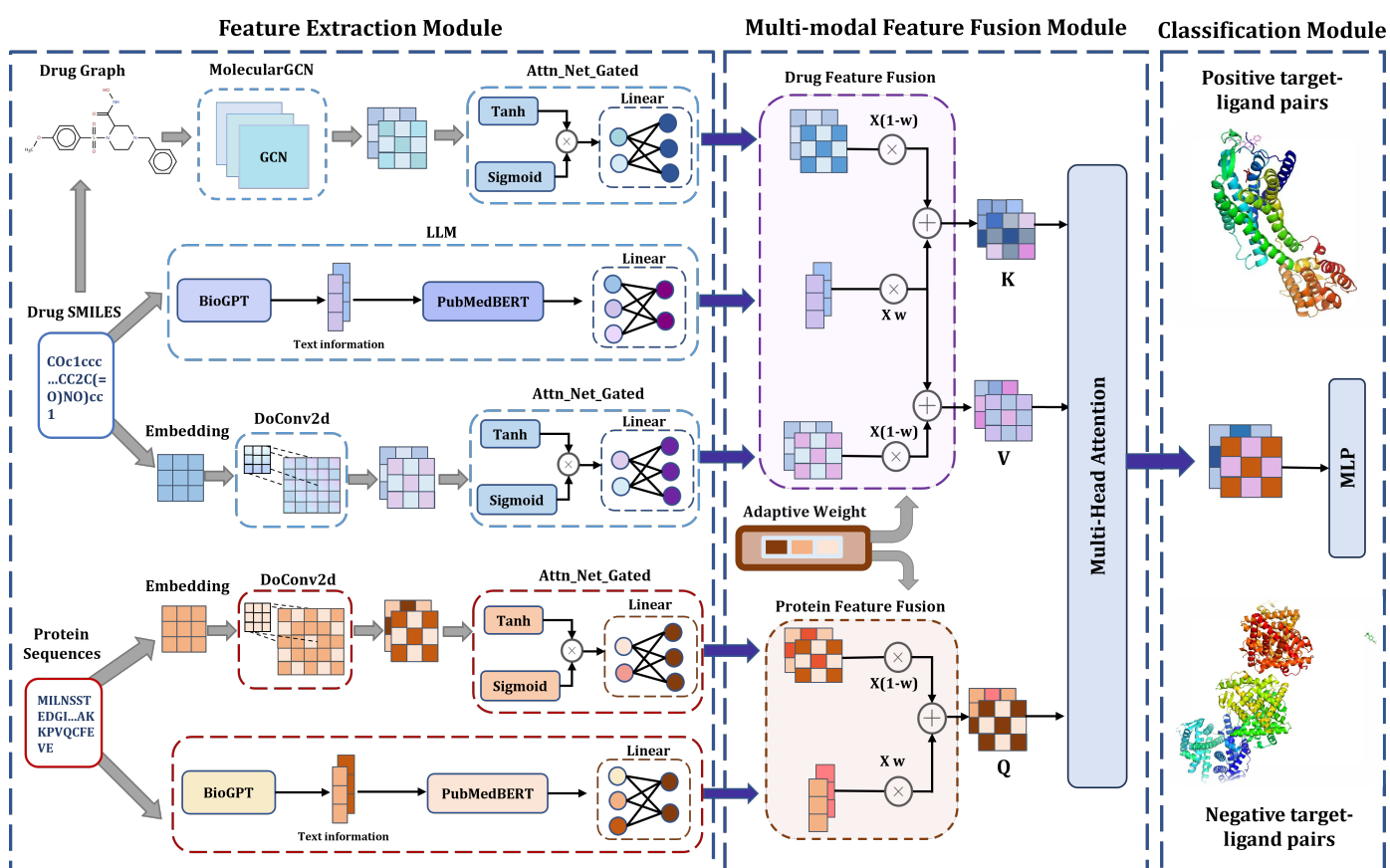


Figure 8. Overall architecture of BioLA-DTI, consisting of feature extraction, length-adaptive fusion, and classification modules. Drug and protein features are extracted via sequence encoders, GCN (Graph Convolutional Network), and biomedical LLMs, then dynamically weighted according to protein sequence length, adaptively fused, and fed into an attention-enhanced MLP for interaction prediction.

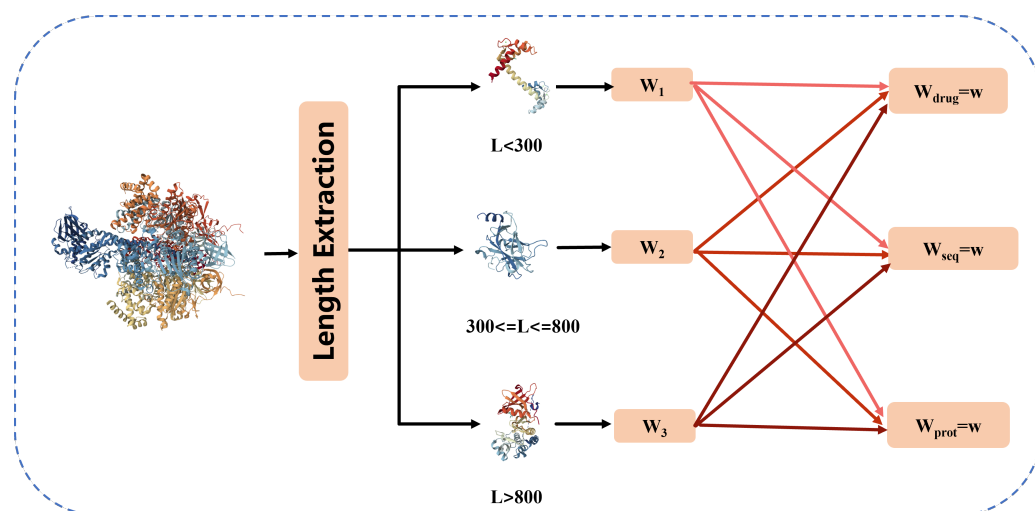


Figure 9. Weight allocation of LLM-derived features based on 300/800 residue boundaries (BindingDB & BioSNAP Distribution-Guided).

4.3.1. Feature Extraction Module

Step 1: Sequence Feature Extraction

The protein sequence is first integer-encoded and rearranged. It is flattened by DO-Conv2d to obtain the sequence feature.

Let P denote the input feature map with C_{in} input channels and C_{out} output channels. Let $K_h \times K_w$ be the kernel size and D_{mul} be the depth multiplier for over-parameterization. DO-Conv2d maintains two independent learnable kernels in training:

- Pointwise kernel: $W \in \mathbb{R}^{C_{out} \times D_{mul} \times C_{in}}$;
- Depthwise kernel: $D \in \mathbb{R}^{C_{in} \times (K_h \cdot K_w) \times D_{mul}}$.

In forward propagation, the two kernels are fused into a unified 4D composite kernel ($W' \in \mathbb{R}^{C_{out} \times C_{in} \times K_h \times K_w}$) via tensor contraction. The fusion formulation is written as

$$W'_{c_{out}, c_{in}, i, j} = \sum_{d=1}^{D_{mul}} W_{c_{out}, d, c_{in}} \cdot D_{c_{in}, (i-1)K_w + j, d}, \quad (7)$$

where $i \in [1, K_h]$ and $j \in [1, K_w]$ index spatial positions inside the kernel. Standard 2D convolution is then conducted with the fused kernel on input feature map P :

$$O = W' * P. \quad (8)$$

Following original DO-Conv settings, D adopts fixed diagonal initialization with d_{diag} . The initialized composite kernel approximates a vanilla convolution kernel. Stable convergence is guaranteed at early training epochs.

After sequence feature extraction, a gating module is introduced to refine representations. Convolution layers generate gate values, followed by Sigmoid normalization. The mathematical formulation is given as follows:

$$H_p^{gate} = H_p^{raw} \odot g \quad (9)$$

Then, features are obtain via global average pooling ($h_d^{struct} \in \mathbb{R}^{128}$).

The drug SMILES string is embedded first. DO-Conv2d is used to capture local sequence features. The feature map is then flattened and processed by a linear layer (h_d^{seq}). The gating mechanism follows the same principle as that for proteins.

Step 2: Molecular Graph Feature Extraction

The drug SMILES is first converted into a DGL graph with initial atom features. Node features are obtained through three GCN layers. Global average pooling is then applied to obtain $h_d^{\text{graph}} \in \mathbb{R}^{128}$.

Step 3: Semantic Feature Extraction

We introduce BioGPT to construct prompts for drug SMILES and protein sequences (e.g., “Describe the pharmacological properties of this molecule”). The generation temperature is set to 0.6, with a maximum of 120 new tokens, and the results are cached.

During encoding, PubMedBERT extracts the token vector, which is then projected via two linear layers to obtain $f_d^{\text{llm}} \in \mathbb{R}^{128}$.

4.3.2. Multi-Modal Feature Fusion Module

Step 1: Length-Adaptive Fusion

For proteins of varying lengths, we design a length-adaptive weighted fusion mechanism based on the actual sequence length. It dynamically adjusts the weights of original features and LLM pre-trained features to achieve balanced fusion. First, the input protein sequence is preprocessed by removing padding tokens, and the true amino acid number is counted as the actual length (L). Then, protein sequences are divided into short, medium and long groups according to length. Independent weights (w_1 , w_2 , and w_3) are defined for each group.

The actual protein length (L , after padding removal) is defined as follows:

$$\alpha(L) = \begin{cases} w_1, & L < 300 \\ w_2, & 300 \leq L \leq 800 \\ w_3, & L > 800 \end{cases} \quad (10)$$

Based on the above adaptive weight (α), weighted fusion is performed on three types of features and pretrained LLM features to realize complementary enhancement of multi-modal information.

$$\tilde{h}_d^{\text{seq}} = (1 - \alpha)h_d^{\text{seq}} + \alpha f_d^{\text{llm}} \quad (11)$$

$$\tilde{h}_d^{\text{graph}} = (1 - \alpha)h_d^{\text{graph}} + \alpha f_d^{\text{llm}} \quad (12)$$

$$\tilde{h}_p = (1 - \alpha)h_p^{\text{struct}} + \alpha f_p^{\text{llm}} \quad (13)$$

The fused vector is expanded to a sequence length of 1 for attention computation. This design adaptively adjusts the contribution of LLM features based on protein length, reducing their influence for short proteins and enhancing it for long-chain proteins.

Step 2: Learnable Multi-Head Attention Fusion

The query, key and value vectors of the multi-head attention module are constructed from length-adaptive LLM-weighted multi-modal fused features.

- **Query Vector (Q):** The fused protein feature ($\tilde{h}_p$) serves as query vector (Q). It is calculated by adaptively weighting raw protein convolutional features and protein LLM features with a learnable coefficient (α). The model uses proteins as query carriers to extract interaction signals from multi-modal drug features.
- **Key Vector (K):** The fused drug graph feature ($\tilde{h}_{d_{\text{graph}}}$) acts as key vector (K). It adaptively fuses GCN-extracted topological drug features and drug LLM embeddings to encode 3D molecular structures and establish topological matching relationships with proteins.
- **Value Vector (V):** The fused drug SMILES sequence feature ($\tilde{h}_{d_{\text{seq}}}$) functions as value vector (V). It combines 1D convolution drug sequence features and drug LLM features

via weighted aggregation. This feature also works as matching keys for the sequence branch and retains semantic information of linear molecular strings.

The model adopts the three groups of Q , K , and V vectors. It produces three separate attention maps to reveal correlations between proteins and multi-modal drug features. Details are listed as follows:

- Drug graph–protein attention (A_1) characterizes the correlation between protein and drug molecular structural features;
- Drug sequence–protein attention (A_2) characterizes the correlation between protein and drug SMILES sequence features;
- Fused drug–protein attention (A_3) characterizes the correlation between protein and the fused features of the drug graph and sequence.

d denotes the feature dimension of single-head attention. A learnable weight vector ($\alpha = [\alpha_1, \alpha_2, \alpha_3]$) is introduced, whose weights (β_1 , β_2 , and β_3) are obtained by Softmax normalization (ensuring their sum is 1) to adaptively balance the contributions of the three attention maps. The final fused attention map (A) is the weighted sum of the three attention maps:

$$A_1 = \text{Softmax}\left(\frac{KQ^\top}{\sqrt{d_k}}\right) \quad (14)$$

$$A_2 = \text{Softmax}\left(\frac{VQ^\top}{\sqrt{d_k}}\right) \quad (15)$$

$$A_3 = \text{Softmax}\left(\frac{(V + K)Q^\top}{\sqrt{d_k}}\right) \quad (16)$$

$$[\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3] = \text{Softmax}([\alpha_1, \alpha_2, \alpha_3]) \quad (17)$$

$$A = \tilde{\beta}_1 A_1 + \tilde{\beta}_2 A_2 + \tilde{\beta}_3 A_3 \quad (18)$$

This design takes proteins as queries to retrieve complementary cues from sequential and structural drug modalities. It faithfully models the induced-fit effect underlying drug–target interactions. The trainable weight mechanism enables adaptive integration of heterogeneous multi-modal features. It yields biologically meaningful embeddings for subsequent binary prediction tasks.

4.3.3. Loss Function

In DTI prediction, some positive samples share high feature similarity. These samples are easily misclassified. Traditional cross-entropy loss hardly guides the model to focus on such hard samples. PolyLoss [46] extends cross-entropy loss by a polynomial expansion term. It imposes additional penalties on positive samples with large prediction errors. This forces the model to pay more attention to boundary samples. PolyLoss is employed as the optimization objective function to strengthen the model's discrimination performance on hard samples.

$$L_{\text{CE}} = -1/N \sum_i [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (19)$$

$$L_{\text{Poly}} = L_{\text{CE}} + \delta \cdot 1/N \sum_i (1 - \hat{y}_i) \quad (20)$$

where y_i denotes the true label of the i -th sample, N represents the number of samples in a batch, and δ is a hyperparameter that controls the strength of the polynomial correction term. The first term (L_{CE}) is the standard cross-entropy loss. The second term ($\delta \cdot 1/N \sum_i (1 - \hat{y}_i)$) imposes an additional penalty on the prediction error of positive samples. By doing so, this

loss alleviates the common class-imbalance issue in DTI tasks and improves the model's recognition performance on minority classes. Whereas Focal Loss [47] relies on exponential decay to reduce the loss from easy samples, PolyLoss controls only the first-order term weight in the loss expansion with a single lightweight hyperparameter. This achieves explicit regularization on hard positive samples without incurring the computational overhead of complex exponential operations. Subsequently, the asymmetric polynomial loss (APL) variant [48] further validated the effectiveness of polynomial-type loss designs in mitigating positive–negative sample imbalance in relation prediction tasks. Concurrently, a DTI-specific study published in the *Briefings in Bioinformatics* journal confirmed that PolyLoss can effectively address the issue of feature homogenization among interacting samples [29]. These findings collectively substantiate the adaptability and extensibility of this loss paradigm in bioinformatics interaction prediction scenarios.

5. Conclusions

This paper proposes a length-adaptive, LLM-enhanced method named BioLA-DTI for drug–target interaction prediction. Its core idea is to adjust the fusion weights between LLM-derived semantic features and conventional structural features according to protein sequence length. This addresses the problems of noise accumulation in short proteins and under-representation in long proteins. Extensive experiments on BioSNAP and BindingDB show that BioLA-DTI consistently outperforms state-of-the-art baseline models across multiple evaluation metrics. It performs particularly well under cold-start settings. Ablation experiments confirm that LLM semantic enhancement, the gating mechanism, and learnable multi-head attention fusion all contribute significantly to performance improvement. The above results show that the model can generate targeted candidate molecule lists for scarce targets. This supports drug repurposing studies. Unlike existing models that rely on a single modality, BioLA-DTI has a length-adaptive perception capability. This gives it a unique generalization advantage when predicting targets with extreme sequence lengths (e.g., short peptides or long, disordered proteins). It effectively overcomes the limitations of conventional methods, which accumulate noise on short sequences and lack adequate representation for long ones.

In terms of application, the framework can be directly used for virtual screening of orphan receptors with low sequence homology. This reduces the experimental costs of blind high-throughput screening. In addition, BioLA-DTI can serve as a front-end pre-filtering tool in large-scale virtual screening pipelines. The model enables rapid preliminary screening to select a subset of compounds with high potential binding activity from massive compound libraries. This substantially decreases the computational cost of subsequent high-precision molecular docking and molecular dynamics simulations. Ultimately, it improves the overall efficiency of the entire virtual drug screening workflow.

It should be noted that the length thresholds (300 and 800) used in the method are based on common protein length distributions. They may not be optimal for certain special families, such as very short antimicrobial peptides or very long cytoskeletal proteins [49]. Future work will explore more refined weight functions based on domain number or intrinsically disordered region proportions. The method will also be applied to practical drug repositioning projects, especially for rapid screening of orphan receptors and emerging pathogen targets.

Author Contributions: Conceptualization, H.Z. and Y.Z.; methodology, H.Z.; formal analysis, H.Z.; writing—original draft preparation, H.Z.; data curation, Y.Z.; visualization, Y.Z.; writing—review and editing, Y.Z.; validation, X.T.; software, H.Z., Y.Z. and Y.C.; supervision, W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China grant number 2023YFC2206400. The APC was funded by the same grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets supporting the findings of this study are publicly available on GitHub at the following link: <https://github.com/liLi0000L/BioLA-DTI.git> (accessed on 26 May 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Villoutreix, B.O.; Poyet, J.L.; Tsaïoun, K. Editorial: Drug discovery and development explained: Introductory notes for the general public. *Front. Drug Discov.* **2024**, *4*, 1528467. [\[CrossRef\]](#)
- Ru, X.; Xu, L.; Han, W.; Zou, Q. In silico methods for drug-target interaction prediction. *Cell Rep. Methods* **2025**, *5*, 10101184. [\[CrossRef\]](#)
- Huang, K.; Xiao, C.; Glass, L.M.; Sun, J. MolTrans: Molecular Interaction Transformer for drug-target interaction prediction. *Bioinformatics* **2021**, *37*, 830–836. [\[CrossRef\]](#)
- Ozturk, H.; Ozgur, A.; Ozkirimli, E. DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics* **2018**, *34*, 821–829. [\[CrossRef\]](#) [\[PubMed\]](#)
- Peng, L.; Mao, J.; Huang, G.; Han, G.; Liu, X.; Liao, W.; Tian, G.; Yang, J. DO-GMA: An End-to-End Drug-Target Interaction Identification Framework with a Depthwise Overparameterized Convolutional Network and the Gated Multihead Attention Mechanism. *J. Chem. Inf. Model.* **2025**, *65*, 1318–1337. [\[CrossRef\]](#) [\[PubMed\]](#)
- Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **2017**, *8*, 573. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, Y.; Luo, D.; Xue, W. Deep Learning for Drug-Target Interaction Prediction: A Comprehensive Review. *Chem. Biol. Drug Des.* **2025**, *106*, e70183. [\[CrossRef\]](#) [\[PubMed\]](#)
- Behrooznia, S.; Hooshmand, M. A Systematic Review of Drug-Related Interactions-Utilizing Deep Learning and LLMs for Prediction and Mitigation. *ACS Omega* **2025**, *10*, 61048–61075. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, Q.; Nie, G.; Li, X.; Xu, Y.; Tan, C.S.H. BioFusionDTI: Assimilating Graph and Sequence Modalities for Generalizable Drug-Target Interaction Prediction. *J. Chem. Inf. Model.* **2025**, *65*, 13187–13202. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tang, X.; Zhou, Y.; Yang, M.; Li, W. TC-DTA: Predicting Drug-Target Binding Affinity With Transformer and Convolutional Neural Networks. *IEEE Trans. Nanobiosci.* **2024**, *23*, 572–578. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Feng, Y.; Yang, X.; Guan, Y.; Zhang, J.; Yang, H.; Wang, Z.; Cheng, Q. KG-MACNF: A nonlinear cross-modal fusion model for predicting drug-target interactions via multi-relational embedding and fine-grained structure. *PLoS ONE* **2025**, *20*, e0331037. [\[CrossRef\]](#) [\[PubMed\]](#)
- Qian, Y.; Li, X.; Wu, J.; Zhang, Q. MCL-DTI: Using drug multimodal information and bi-directional cross-attention learning method for predicting drug-target interaction. *BMC Bioinform.* **2023**, *24*, 323. [\[CrossRef\]](#) [\[PubMed\]](#)
- Orengo, C.A.; Thornton, J.M. Protein Families and Their Evolution-A Structural Perspective. *Annu. Rev. Biochem.* **2005**, *74*, 867–900. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tan, T.; Frenkel, D.; Gupta, V.; Deem, M.W. Length, Protein-Protein Interactions, and Complexity. *Phys. A Stat. Mech. Its Appl.* **2005**, *350*, 52–62. [\[CrossRef\]](#)
- Wu, L.; Huang, Y.; Tan, C.; Gao, Z.; Hu, B.; Lin, H.; Liu, Z.; Li, S.Z. PSC-CPI: Multi-Scale Protein Sequence-Structure Contrasting for Efficient and Generalizable Compound-Protein Interaction Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*; Association for the Advancement of Artificial Intelligence: Washington, DC, USA, 2024; Volume 38, pp. 310–319. [\[CrossRef\]](#)
- Zhang, G.; Ignatova, Z. Generic Algorithm to Predict the Speed of Translational Elongation: Implications for Protein Biogenesis. *PLoS ONE* **2009**, *4*, e5036. [\[CrossRef\]](#) [\[PubMed\]](#)
- Traut, T. Multidomain Proteins. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014. [\[CrossRef\]](#)
- Giudice, J. Structure-Activity Relationships in the Endosomal Escape of Designed Miniature Proteins. In *ProQuest Dissertations and Theses Global*; University of California, Berkeley: Berkeley, CA, USA, 2025.

20. Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C.L.; Ma, J.; et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2016239118. [CrossRef] [PubMed]
21. Vazzana, G.; Manfredi, M.; Savojardo, C.; Martelli, P.L.; Casadio, R. Embedding-Based Alignments Capture Structural and Sequence Domains of Distantly Related Multifunctional Human Proteins. *Computation* **2026**, *14*, 25. [CrossRef]
22. Xu, M.; Yuan, X.; Miret, S.; Tang, J. ProtST: Multi-Modality Learning of Protein Sequences and Biomedical Texts. *arXiv* **2023**, arXiv:2301.12040.
23. Tang, W.; Zhao, Q.; Wang, J. LLMDTA: Improving Cold-Start Prediction in Drug-Target Affinity with Biological LLM. *IEEE Trans. Comput. Biol. Bioinform.* **2025**, *22*, 2398–2409. [CrossRef] [PubMed]
24. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.Y. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409. [CrossRef] [PubMed]
25. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthc.* **2022**, *3*, 1–23. [CrossRef]
26. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129. [CrossRef] [PubMed]
27. Chen, L.; Tan, X.; Wang, D.; Zhong, F.; Liu, X.; Yang, T.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. TransformerCPI: Improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* **2020**, *36*, 4406–4414. [CrossRef] [PubMed]
28. Liu, G.; Singha, M.; Pu, L.; Neupane, P.; Feinstein, J.; Wu, H.C.; Ramanujam, J.; Brylinski, M. GraphDTI: A robust deep learning predictor of drug-target interactions from multiple heterogeneous data. *J. Cheminform.* **2021**, *13*, 58. [CrossRef] [PubMed]
29. Bian, J.; Zhang, X.; Zhang, X.; Xu, D.; Wang, G. MCANet: Shared-weight-based MultiheadCrossAttention network for drug-target interaction prediction. *Brief. Bioinform.* **2023**, *24*, bbad082. [CrossRef] [PubMed]
30. Bai, P.; Miljkovic, F.; John, B.; Lu, H. Interpretable bilinear attention network with domain adaptation improves drug-target prediction. *Nat. Mach. Intell.* **2023**, *5*, 126–136. [CrossRef]
31. Peng, L.; Liu, X.; Chen, M.; Liao, W.; Mao, J.; Zhou, L. MGNDTI: A Drug-Target Interaction Prediction Framework Based on Multimodal Representation Learning and the Gating Mechanism. *J. Chem. Inf. Model.* **2024**, *64*, 6684–6698. [CrossRef] [PubMed]
32. Li, C.; Mi, J.; Wang, H.; Liu, Z.; Gao, J.; Wan, J. MGMA-DTI: Drug target interaction prediction using multi-order gated convolution and multi-attention fusion. *Comput. Biol. Chem.* **2025**, *118*, 108449. [CrossRef] [PubMed]
33. Scharf, M.M.; Humphrys, L.J.; Berndt, S.; Di Pizio, A.; Lehmann, J.; Liebscher, I.; Nicoli, A.; Niv, M.Y.; Peri, L.; Schihada, H.; et al. The dark sides of the GPCR tree—Research progress on understudied GPCRs. *Br. J. Pharmacol.* **2025**, *182*, 3109–3134. [CrossRef] [PubMed]
34. Torrisi, M.; de la Vega de León, A.; Climent, G.; Loos, R.; Panjkovich, A. Improving the Assessment of Deep Learning Models in the Context of Drug-Target Interaction Prediction. *bioRxiv* **2022**, preprint. [CrossRef]
35. Zhao, Q.; Zhao, H.; Zheng, K.; Wang, J. HyperAttentionDTI: Improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics* **2022**, *38*, 655–662. [CrossRef] [PubMed]
36. Huang, K.; Fu, T.; Glass, L.M.; Zitnik, M.; Xiao, C.; Sun, J. DeepPurpose: A deep learning library for drug-target interaction prediction. *Bioinformatics* **2020**, *36*, 5545–5547. [CrossRef] [PubMed]
37. Xu, T. Towards a Complete Logical Framework for GNN Expressiveness. In Proceedings of the 13th International Conference on Learning Representations (ICLR 2025), Singapore, 24–28 April 2025.
38. Xu, X.; Jiang, W.; Liu, W.; Liu, B.; Wang, Y. Graph Unseen Node Embedding Learning Network for Drug Target Affinity Prediction. In Proceedings of the 2024 IEEE 17th International Conference on Signal Processing (ICSP); IEEE: Piscataway, NJ, USA, 2024.
39. Quan, Z.; Guo, Y.; Lin, X.; Wang, Z.J.; Zeng, X. GraphCPI: Graph Neural Representation Learning for Compound-Protein Interaction. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; Yoo, I., Bi, J., Hu, X., Eds.; NSF, IEEE International Conference on Bioinformatics and Biomedicine-BIBM; IEEE: Piscataway, NJ, USA, 2019; pp. 717–722.
40. Sun, Y.; Li, Y.Y.; Leung, C.K.; Hu, P. iNGNN-DTI: Prediction of drug-target interaction with interpretable nested graph neural network and pretrained molecule models. *Bioinformatics* **2024**, *40*, btae135. [CrossRef] [PubMed]
41. Li, J.; Sun, L.; Liu, L.; Li, Z. MIFAM-DTI: A drug-target interactions predicting model based on multi-source information fusion and attention mechanism. *Front. Genet.* **2024**, *15*, 1381997. [CrossRef] [PubMed]
42. Cao, Z.; Xie, J.; Xu, J.; Li, B. Cross-Modal Interaction-Aware Progressive Fusion Network for Drug-Target Interaction Prediction. *J. Chem. Inf. Model.* **2025**, *65*, 10713–10726. [CrossRef] [PubMed]
43. Zitnik, M.; Sosic, R.; Leskovec, J. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. 2018; Volume 5. Available online: <http://snap.stanford.edu/biodata> (accessed on 5 March 2026).
44. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201. [CrossRef] [PubMed]

45. Abir, A.R.; Alif, M.N.; Zhang, W.; Ahmed, K.T.; Zhang, W. DCGAT-DTI: Dynamic cross-graph attention network for drug-target interaction prediction. *Bioinform. Adv.* **2026**, *6*, vbaf306. [[CrossRef](#)] [[PubMed](#)]
46. Leng, Z.; Tan, M.; Liu, C.; Cubuk, E.D.; Shi, X.; Cheng, S.; Anguelov, D. PolyLoss: A Polynomial Expansion Perspective of Classification Loss Functions. *arXiv* **2022**, arXiv:2204.12511.
47. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
48. Huang, Y.; Qi, J.; Wang, X.; Lin, Z. Asymmetric Polynomial Loss for Multi-Label Classification. In *Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2023; pp. 4288–4292. [[CrossRef](#)]
49. Nevers, Y.; Glover, N.; Dessimoz, C.; Lecompte, O. Protein length distribution is remarkably consistent across Life. *bioRxiv* **2021**, preprint, later published in *Genome Biology*. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.