

RESEARCH

Open Access



# BiMA-DTI: a bidirectional Mamba-Attention hybrid framework for enhanced drug-target interaction prediction

Youyuan Shui<sup>1</sup>, Xuewen Ge<sup>1</sup>, Chen Cao<sup>2</sup>, Junjie Wang<sup>1,3\*</sup>, Jie Hu<sup>1,3\*</sup> and Yun Liu<sup>1,3,4\*</sup>

## Abstract

**Background** Predicting drug-target interactions (DTIs) is essential for accelerating drug discovery, yet traditional experimental methods are time-consuming and costly. Computational approaches, especially those using machine learning and deep learning, offer a more efficient alternative.

**Results** This paper presents the BiMA-DTI framework, which integrates Mamba's State Space Model (SSM) with multi-head attention mechanisms. This combination maximizes Mamba's ability to process long sequences while taking advantage of the attention mechanism's strength in handling short sequences. We designed a hybrid Mamba-Attention Network (MAN) and a Graph Mamba Network (GMN) for processing multimodal inputs, including protein amino acid sequences, Simplified Molecular Input Line Entry System (SMILES) strings, and molecular graphs of drugs, enabling comprehensive feature extraction and fusion. To enhance the complementarity between features extracted from sequences and graphs, BiMA-DTI performs a two-step weighted fusion of sequence features of drugs and proteins with molecular graph features of drugs. Finally, these fused features are concatenated and passed through a fully connected network to predict DTIs.

**Conclusions** BiMA-DTI demonstrates its potential to discover new drugs, offering a powerful tool for drug discovery. Experimental results show that BiMA-DTI outperforms state-of-the-art competing methods on benchmark datasets. Additionally, ablation experiments validate the rationality of BiMA-DTI's architecture and its generalization ability. Visualization studies provide interpretability of biological mechanisms. Finally, case studies further confirm that BiMA-DTI is a reliable drug-target interaction prediction tool.

**Keywords** Drug discovery, Drug-target interaction prediction, Mamba, Multi-head attention, Multi-modal fusion

\*Correspondence:

Junjie Wang  
junjie2021@njmu.edu.cn  
Jie Hu  
hujie@njmu.edu.cn  
Yun Liu  
liuyun@njmu.edu.cn

<sup>1</sup> Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China

<sup>2</sup> Key Laboratory for Bio-Electromagnetic Environment and Advanced Medical Theranostics, School of Biomedical Engineering and Informatics, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China

<sup>3</sup> Institute of Medical Informatics and Management, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China

<sup>4</sup> Department of Information, The First Affiliated Hospital, Nanjing Medical University, No. 300 Guang Zhou Road, Nanjing 210029, Jiangsu, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Predicting drug-target interactions (DTIs) is a crucial step in drug discovery, especially for identifying and validating potential therapeutic candidates. Traditional experimental methods for determining DTIs are time-consuming and cost-prohibitive [1]. Moreover, exhaustively screening all possible drug-target pairs is infeasible due to the immense size of the search space. In recent years, computational approaches have emerged as a valuable alternative. By harnessing the power of machine learning and deep learning, these approaches offer a faster and more cost-effective way to accelerate drug development.

In the early stage of computational approaches for DTI task research, traditional machine learning, such as support vector machine (SVM) [2] and Random Forest (RF) [3], has been widely applied. These methods enhanced prediction accuracy by leveraging manually designed features and learning patterns from data, marking a shift from experience-driven to data-driven approaches in DTI prediction tasks [4–8]. However, despite their promising performance, traditional machine learning based methods still face challenges in processing complex and large-scale data.

Deep learning has demonstrated remarkable success across various domains, owing to its capability to discern intricate patterns within large datasets. Consequently, researchers have increasingly embraced deep learning methods, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), and Transformer, to enhance the prediction of DTIs. References [9–12] highlight the use of CNNs for extracting features from drug molecules and proteins. By excelling at capturing local features, CNNs are particularly effective in identifying key information from one-dimensional or two-dimensional data. In contrast, RNNs, as demonstrated in [13–16], are well-suited for handling sequential data, making them ideal for processing protein amino acid sequences and SMILES (Simplified Molecular Input Line Entry System) of drug. GNNs, discussed in [17–20], excel at modeling the spatial structures of drug molecules and proteins, as well as capturing the interactions between them. The Transformer-based models [21] have demonstrated exceptional performance in handling long-sequence processing tasks through the attention mechanism, providing new perspectives for addressing complex sequence data. For instance, TransformerCPI [22] employs a Transformer decoder to learn interaction features between proteins and compounds. Multi-TransDTI [23] adopts a multi-view strategy with a Transformer to extract key local residues of

proteins for enhanced representation learning. DTITR [24] demonstrates that incorporating a Cross-Attention Transformer-Encoder improves the discriminative power of robust aggregate representations for proteins and compounds. Additionally, GraphormerDTI [25] and DTI-GTN [26] utilize graph Transformers to extract representations of drugs and target proteins. The attention mechanism of Transformer is well-suited for learning drug-target interaction tasks due to its ability to capture complex relationships. However, the quadratic computational complexity of the attention mechanism [27], which scales with sequence length, makes Transformers resource-intensive to train and deploy. Recently, Mamba introduced a novel State Space Model (SSM) that achieves linear time complexity while matching or surpassing Transformer performance in various language modeling and vision tasks [28, 29]. This advancement offers a more efficient alternative for handling large-scale sequence data.

This study presents a novel framework, hybrid Bidirectional Mamba-Attention for Drug-Target Interaction prediction (BiMA-DTI), which integrates the strengths of Mamba and multi-head attention mechanisms. Mamba is adept at capturing long-range dependencies within sequences, while multi-head attention mechanism effectively focuses on dependencies and global attention within shorter sequences. By combining these complementary capabilities, BiMA-DTI aims to improve the accuracy and efficiency of DTI prediction. BiMA-DTI supports multiple modals of input data: the protein amino acid sequence, the SMILES of the drug, and the molecular graph of the drug. For processing the protein amino acid sequences and SMILES, we have designed a hybrid Mamba-Attention Network (MAN). Additionally, to analyze the molecular graph of the drug, we have developed a Graph Mamba Network (GMN). The extracted multimodal features from drugs and proteins are then fused using a multi-modal fusion network, enabling the prediction of DTIs.

To evaluate the performance of the BiMA-DTI, we conducted experiments on four medium-scale datasets and one large-scale dataset constructed by ourselves. Beside, we performed ablation experiments to validate the effectiveness of the core modules and multimodal inputs. Visualization study provides excellent interpretability of the biological mechanism for BiMA-DTI. A case study was also conducted to demonstrate BiMA-DTI's applicability in real-world scenarios. In summary, BiMA-DTI not only has the best performance, but also has the potential to discover new drugs and can provide new tools for drug discovery.

## Results

### Experimental setup

To simulate real DTI prediction application scenarios and assess the generalization ability and reliability of the models, we compared BiMA-DTI with other baseline methods under four different experimental settings.

- E1: the data sets were randomly split into training, validation, and test sets in a ratio of 7:1:2.
- E2: if a drug  $d$  in the test set is present in the training set, it is removed from the training set. Additionally, if a protein  $p$  in the test set is not present in the training set, it is removed from the test set.
- E3: if a protein  $p$  in the test set is present in the training set, it is removed from the training set. Similarly, if a drug  $d$  in the test set is not present in the training set, it is removed from the test set.
- E4: if both a drug  $d$  and a protein  $p$  in the test set are present in the training set, they are removed from the training set.

For the E2, E3, and E4 settings, 20% of the dataset was randomly selected as the test set, while the remaining 80% was used for training. Subsequently, drugs or proteins were removed according to the specific criteria of Each setting. After this process, the remaining test samples were further divided into a validation set and a test set at a 1:2 ratio.

For Each setup, we conducted 10 independent runs using different random seeds for data partitioning, except for the Therapeutic Target Database (TTD) dataset [30–32] where a single run was performed due to its very large scale. The model achieving the highest area under the receiver operating characteristic (ROC) curve (AUROC) on the validation set was selected as the optimal model and evaluated on the test set.

To better demonstrate the BiMA-DTI performance, we employed the AUROC, the area under the precision-recall (PR) curve (AUPRC), accuracy, F1-score, and Matthews' correlation coefficient (MCC) as Evaluation metrics. The Equation of the evaluation metrics can be found in Additional file 1: Eqs. S1–S5.

### Baseline methods

To evaluate our model, we compared it to following baseline models. The performance of all baseline models was re-evaluated in the same experimental environment in reference [33]. Their hyperparameters were explicitly set to their default values. For BiMA-DTI, we employed the Optuna framework [34] to perform automated hyperparameter optimization. Specifically, we defined the search space for each module, including

the learning rate, number of Mamba layers, number of attention heads, graph transformer layers, hidden dimensions, fusion expansion factor, and loss function. The Tree-structured Parzen Estimator (TPE) was used as the optimization algorithm, and the average AUROC and AUPRC on the validation set were used as the selection criteria. We perform hyperparameter tuning under the random experimental setting (E1) on the Human dataset and adopt the same hyperparameters for Each setting on other datasets. Additional file 1: Table S1 summarizes all considered hyperparameters, their respective search ranges, and the optimal values obtained.

All experiments were conducted on a server equipped with an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz and an NVIDIA GeForce RTX 3090 GPU. We measured the average time to train and infer on a single data instance for BiMA-DTI and some baseline models. The results are summarized in Additional file 1: Table S2.

- MGNDTI [33]: MGNDTI uses RetNet to extract features from drug SMILES sequences and protein amino acid sequences, and graph convolutional networks (GCNs) for extracting features from drug molecular graphs. It builds a multimodal gating network for feature filtering and fusion, and finally uses a fully connected network (FCN) for prediction.
- MolTrans [35]: MolTrans encodes drug and protein features using Transformer and employs CNN and FCN for prediction.
- TransformerCPI [22]: TransformerCPI uses CNN to extract protein features from amino acid sequences and GNN to extract drug features from SMILES sequences. Then processes these features using a Transformer decoder and utilizes a FCN for prediction.
- CPI-GNN [36]: CPI-GNN uses GNN and one-dimensional CNN to encode features of drugs and proteins. It analyzes the significance of protein subsequences in relation to a drug using a one-sided attention mechanism.
- CPGL [37]: CPGL extracted protein features through long short-term memory (LSTM) neural network and drug features through graph attention network.
- BACPI [38]: BACPI uses a bidirectional attention neural network to combine the features of drugs and proteins and utilizes a classifier for prediction.
- GIFDTI [39]: GIFDTI learns drug features from SMILES sequences and protein features from amino acid sequences using CNN and Transformer. Then uses the global features of drugs and proteins for classification prediction.

- FOTF-CPI [40]: FOTF-CPI incorporates an optimal transport-based fragmentation model with a fused attention mechanism for prediction.
- DO-GMA [41]: DO-GMA combines a depthwise overparameterized convolutional neural network and a graph convolutional network to handle the SMILES sequences and graphs of drugs simultaneously, and combines the gating mechanism and the attention mechanism to fuse the features of drugs and targets.
- LAM-DTI [42]: LAM-DTI solves the problem of sequence length differences between drugs and targets by applying the aconnectionist temporal classification module to generate normalized feature sequences.

### Comparison of BiMA-DTI with baselines

To comprehensively evaluate the performance of BiMA-DTI, we conducted comparative experiments on datasets of different scales, and the experiments on each dataset were carried out under four different experimental settings. On the medium-scale datasets (Human, *C.elegans*, BioSNAP, and BindingDB) [35, 43–49], comprehensive comparisons were performed against all baseline methods. For the large-scale dataset, considering its substantial size and computational demands, we adopted a more focused evaluation strategy. Specifically, we selected for comparison two recently published state-of-the-art models (2025), along with the top three performing models from our medium-scale benchmark analysis. This tiered evaluation approach allows for both broad benchmarking on standard datasets and meaningful performance assessment on challenging large-scale data.

Furthermore, in order to verify the practical applicability of BiMA-DTI, we constructed a test set with balanced positive and negative samples based on the TTD dataset, in which all the positive samples were the US Food and Drug Administration (FDA) approved drug target pairs. We once again tested the model performance under the double cold start condition on the TTD dataset using this test set and compared it with the baseline model, and conducted a reliability analysis. These are helpful for validate the model's utility in real-world settings.

### Performances on medium-scale datasets

Through the comprehensive evaluation of four medium-sized benchmark datasets shown in Tables 1 and 2, the consistency advantage of BiMA-DTI in the DTI prediction task was revealed. On the Human dataset, BiMA-DTI achieved full leadership in the cases of drug cold start (under the E2 setting) and target cold start (under the E3 setting), respectively. All metrics under the E1 setting achieved suboptimal results, demonstrating a strong

precision-recall balance. The results of *C.elegans* are even more astonishing. Our model achieved an AUROC of 0.9933 and an AUPRC of 0.9940 at the E1 setting, approaching the theoretical maximum of this benchmark. This is attributed to the synergistic combination of Mamba's remote dependency modeling and the local interaction focus of the attention mechanism.

As shown in Fig. 1, similar advantages extend to the BindingDB and BioSNAP datasets. Under the E1 setting, BiMA-DTI achieves the optimal AUROC and AUPRC, respectively. The outstanding performance of this model on BindingDB, which includes over 49,000 interactions with significant drug advantages (5.58:1 drug-target ratio), highlights its ability to handle a variety of chemical compounds. It is worth noting that in the protein cold start scenario, BiMA-DTI maintained leading performance in all datasets, with an AUPRC of 0.5284 on BindingDB and 0.6826 on BioSNAP, which was 2.51–3.20% higher than the closest competitor. On the BindingDB and BioSNAP datasets, the performance of most models dropped by 50–30% in cold start settings. The double cold start evaluation (under the E4 setting) further demonstrated the generalization ability of BiMA-DTI, and BiMA-DTI achieved the optimal AUROC and AUPRC on all three datasets.

In summary, the proposed BiMA-DTI model demonstrates superior performance compared to ten baseline models across four datasets under four distinct experimental settings, showcasing exceptional generalization capability in DTI prediction tasks. While some baseline models (e.g., DO-GMA and LAM-DTI) show competitive performance on specific datasets or settings, none match BiMA-DTI's consistent cross-dataset reliability. The strong performance of BiMA-DTI can be attributed to two key factors. First, the integration of the Mamba module and multi-head attention effectively captures long-range dependencies while efficiently focusing on short-span interactions. Second, the MAN incorporates a Bi-Mamba module, further enhancing its capacity to extract global sequence information. This advanced attention strategy significantly improves the model's understanding of the intricate local DTI processes. The comparison of BiMA-DTI to the baseline model on the BindingDB and BioSNAP datasets is detailed in Additional file 1: Tables S3 and S4.

### Performances on large-scale datasets

To rigorously evaluate scalability in real-world drug discovery scenarios, we conducted extensive experiments on the TTD dataset. This dataset was chosen for its unique combination of characteristics: it provides quantitative binding affinity measurements (such as  $K_i$ ,  $K_d$ , and  $IC_{50}$ ), covers a wide spectrum of

**Table 1** Comparison results of BiMA-DTI and baselines on Human (10 random runs)

Method	AUROC	AUPRC	Accuracy	F1-score	MCC
E1					
MGNDTI	0.9855 ± 0.0031	0.9820 ± 0.0059	0.9481 ± 0.0056	0.9485 ± 0.0057	0.8951 ± 0.0111
MolTrans	0.9799 ± 0.0028	0.9785 ± 0.0044	0.9418 ± 0.0099	0.9207 ± 0.0291	0.8823 ± 0.0196
TransformerCPI	0.9795 ± 0.0036	0.9745 ± 0.0052	0.9316 ± 0.0071	0.9223 ± 0.0092	0.8613 ± 0.0147
CPI-GNN	0.9329 ± 0.0085	0.9174 ± 0.0164	0.8899 ± 0.0084	0.8858 ± 0.0098	0.7798 ± 0.0170
BACPI	0.9670 ± 0.0058	0.9608 ± 0.0087	0.9181 ± 0.0110	0.9070 ± 0.0130	0.8341 ± 0.0225
CPGL	0.9674 ± 0.0052	0.9673 ± 0.0079	0.9092 ± 0.0123	0.9055 ± 0.0141	0.8191 ± 0.0238
GIFDTI	0.9690 ± 0.0047	0.9645 ± 0.0084	0.9091 ± 0.0099	0.8967 ± 0.0120	0.8161 ± 0.0200
FOTF-CPI	0.9834 ± 0.0024	0.9803 ± 0.0035	0.9413 ± 0.0074	0.9326 ± 0.0090	0.8811 ± 0.0149
DO-GMA	<b>0.9891 ± 0.0029</b>	<b>0.9870 ± 0.0068</b>	<b>0.9541 ± 0.0052</b>	<b>0.9544 ± 0.0047</b>	<b>0.9098 ± 0.0101</b>
LAM-DTI	0.9860 ± 0.0044	0.9854 ± 0.0024	0.9502 ± 0.0057	0.9497 ± 0.0056	0.9007 ± 0.0114
BiMA-DTI	0.9860 ± 0.0039	0.9858 ± 0.0024	0.9523 ± 0.0053	0.9522 ± 0.0049	0.9043 ± 0.0101
E2					
MGNDTI	0.9162 ± 0.0188	0.9073 ± 0.0187	0.8361 ± 0.0183	0.8481 ± 0.0196	0.6797 ± 0.0383
MolTrans	0.8856 ± 0.0133	0.8783 ± 0.0152	0.8219 ± 0.0170	0.7769 ± 0.0263	0.6438 ± 0.0334
TransformerCPI	0.8689 ± 0.0140	0.8920 ± 0.0131	0.8390 ± 0.0114	0.8110 ± 0.0166	0.6766 ± 0.0207
CPI-GNN	0.7642 ± 0.0496	0.7765 ± 0.0444	0.7231 ± 0.0440	0.6967 ± 0.0540	0.4519 ± 0.0874
BACPI	0.8286 ± 0.0225	0.8152 ± 0.0309	0.7518 ± 0.0250	0.6957 ± 0.0373	0.4992 ± 0.0529
CPGL	0.8952 ± 0.0119	0.9068 ± 0.0087	0.8180 ± 0.0187	0.8061 ± 0.0292	0.6431 ± 0.0286
GIFDTI	0.8698 ± 0.0246	0.8749 ± 0.0181	0.8010 ± 0.0232	0.7586 ± 0.0344	0.6022 ± 0.0429
FOTF-CPI	0.9016 ± 0.0087	0.8995 ± 0.0086	0.8201 ± 0.0224	0.7739 ± 0.0434	0.6461 ± 0.0370
DO-GMA	0.9277 ± 0.0099	0.9420 ± 0.0123	0.8614 ± 0.0121	0.8606 ± 0.0122	0.7257 ± 0.0238
LAM-DTI	0.9290 ± 0.0001	0.9410 ± 0.0001	0.8746 ± 0.0005	0.8734 ± 0.0009	0.7584 ± 0.0014
BiMA-DTI	<b>0.9298 ± 0.0100</b>	<b>0.9480 ± 0.0067</b>	<b>0.8782 ± 0.0120</b>	<b>0.8770 ± 0.0125</b>	<b>0.7587 ± 0.0243</b>
E3					
MGNDTI	0.9795 ± 0.0079	0.9785 ± 0.0061	0.9401 ± 0.0178	0.9395 ± 0.0151	0.8789 ± 0.0335
MolTrans	0.9738 ± 0.0054	0.9699 ± 0.0080	0.9448 ± 0.0115	0.9199 ± 0.0211	0.8874 ± 0.0233
TransformerCPI	0.9569 ± 0.0069	0.9495 ± 0.0063	0.8934 ± 0.0140	0.8661 ± 0.0218	0.7836 ± 0.0249
CPI-GNN	0.9674 ± 0.0103	0.9656 ± 0.0165	0.9364 ± 0.0166	0.9305 ± 0.0179	0.8737 ± 0.0320
BACPI	0.9790 ± 0.0090	0.9790 ± 0.0077	0.9410 ± 0.0158	0.9279 ± 0.0192	0.8800 ± 0.0307
CPGL	0.9240 ± 0.0166	0.9326 ± 0.0139	0.8478 ± 0.0284	0.8239 ± 0.0471	0.7023 ± 0.0467
GIFDTI	0.9413 ± 0.0103	0.9357 ± 0.0127	0.8711 ± 0.0170	0.8469 ± 0.0213	0.7379 ± 0.0345
FOTF-CPI	0.9734 ± 0.0070	0.9728 ± 0.0059	0.9221 ± 0.0099	0.9014 ± 0.0150	0.8426 ± 0.0196
DO-GMA	0.9854 ± 0.0055	0.9647 ± 0.0229	0.9644 ± 0.0099	0.9638 ± 0.0101	0.9298 ± 0.0192
LAM-DTI	0.9897 ± 0.0042	0.9908 ± 0.0046	0.9403 ± 0.0324	0.9316 ± 0.0415	0.8850 ± 0.0584
BiMA-DTI	<b>0.9900 ± 0.0067</b>	<b>0.9913 ± 0.0049</b>	<b>0.9689 ± 0.0083</b>	<b>0.9677 ± 0.0091</b>	<b>0.9380 ± 0.0170</b>
E4					
MGNDTI	<b>0.8106 ± 0.0228</b>	<b>0.7780 ± 0.0278</b>	<b>0.7180 ± 0.0446</b>	<b>0.7708 ± 0.0219</b>	<b>0.4862 ± 0.0618</b>
MolTrans	0.6969 ± 0.0187	0.6377 ± 0.0198	0.5891 ± 0.0426	0.3258 ± 0.1520	0.2760 ± 0.0680
TransformerCPI	0.7387 ± 0.0201	0.7475 ± 0.0212	0.7095 ± 0.0209	0.5835 ± 0.0350	0.4261 ± 0.0370
CPI-GNN	0.6929 ± 0.0562	0.6771 ± 0.0450	0.6265 ± 0.0351	0.4985 ± 0.0551	0.2800 ± 0.0691
BACPI	0.7682 ± 0.0283	0.7186 ± 0.0327	0.6365 ± 0.0442	0.3516 ± 0.1836	0.2599 ± 0.1331
CPGL	0.7658 ± 0.0245	0.7390 ± 0.0267	0.5725 ± 0.0659	0.2447 ± 0.2166	0.1784 ± 0.1534
GIFDTI	0.7475 ± 0.0298	0.7406 ± 0.0241	0.6918 ± 0.0180	0.5545 ± 0.0471	0.3856 ± 0.0345
FOTF-CPI	0.7250 ± 0.0315	0.6766 ± 0.0353	0.6222 ± 0.0294	0.3301 ± 0.1123	0.2496 ± 0.0525
DO-GMA	0.7431 ± 0.0219	0.7397 ± 0.0241	0.6549 ± 0.0376	0.7119 ± 0.0197	0.4544 ± 0.0369
LAM-DTI	0.7809 ± 0.0233	0.7910 ± 0.0257	0.5700 ± 0.0839	0.2441 ± 0.2491	0.2063 ± 0.1720
BiMA-DTI	0.7681 ± 0.0204	0.7632 ± 0.0268	0.6942 ± 0.0244	0.7331 ± 0.0154	0.4078 ± 0.0406

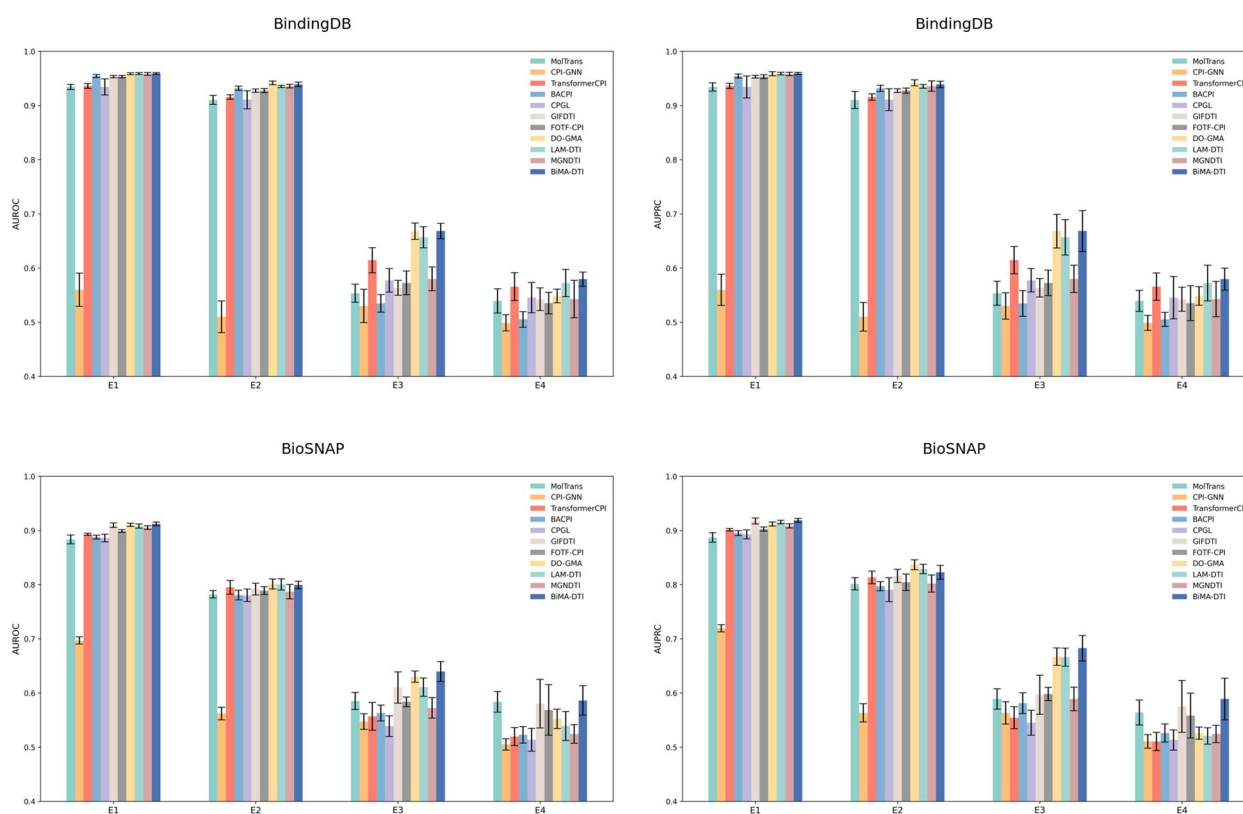
Note: Values in bold indicate the best performance across all models

**Table 2** Comparison results of BiMA-DTI and baselines on *C.elegans* (10 random runs)

Method	AUROC	AUPRC	Accuracy	F1-score	MCC
E1					
MGNDTI	0.9912 ± 0.0022	0.9916 ± 0.0020	0.9682 ± 0.0051	0.9682 ± 0.0050	0.9365 ± 0.0102
MolTrans	0.9918 ± 0.0024	0.9920 ± 0.0028	0.9670 ± 0.0032	0.9626 ± 0.0047	0.9342 ± 0.0064
TransformerCPI	0.9919 ± 0.0015	0.9916 ± 0.0018	0.9608 ± 0.0045	0.9608 ± 0.0044	0.9217 ± 0.0089
CPI-GNN	0.9536 ± 0.0079	0.9422 ± 0.0121	0.9146 ± 0.0100	0.9185 ± 0.0088	0.8292 ± 0.0201
BACPI	0.9864 ± 0.0044	0.9869 ± 0.0039	0.9489 ± 0.0114	0.9487 ± 0.0117	0.8981 ± 0.0227
CPGL	0.9757 ± 0.0034	0.9797 ± 0.0029	0.9265 ± 0.0117	0.9288 ± 0.0103	0.8538 ± 0.0224
GIFDTI	0.9827 ± 0.0068	0.9843 ± 0.0056	0.9435 ± 0.0110	0.9435 ± 0.0116	0.8873 ± 0.0221
FOTF-CPI	0.9919 ± 0.0030	0.9909 ± 0.0054	0.9663 ± 0.0051	0.9661 ± 0.0052	0.9328 ± 0.0101
DO-GMA	0.9926 ± 0.0026	0.9905 ± 0.0048	0.9716 ± 0.0033	0.9715 ± 0.0034	<b>0.9645 ± 0.0056</b>
LAM-DTI	0.9929 ± 0.0013	0.9914 ± 0.0013	0.9690 ± 0.0049	0.9690 ± 0.0051	0.9382 ± 0.0098
BiMA-DTI	<b>0.9933 ± 0.0021</b>	<b>0.9940 ± 0.0020</b>	<b>0.9724 ± 0.0041</b>	<b>0.9725 ± 0.0040</b>	0.9448 ± 0.0079
E2					
MGNDTI	0.8985 ± 0.0161	0.9134 ± 0.0084	0.8188 ± 0.0232	0.8215 ± 0.0229	0.6410 ± 0.0481
MolTrans	0.8259 ± 0.0290	0.8595 ± 0.0259	0.7465 ± 0.0298	0.7202 ± 0.0314	0.5014 ± 0.0616
TransformerCPI	0.8076 ± 0.0151	0.8657 ± 0.0122	0.7623 ± 0.0244	0.7299 ± 0.0453	0.5551 ± 0.0285
CPI-GNN	0.6830 ± 0.0649	0.7278 ± 0.0391	0.6320 ± 0.0584	0.6038 ± 0.0763	0.2844 ± 0.1126
BACPI	0.8083 ± 0.0280	0.8315 ± 0.0294	0.7079 ± 0.0272	0.6653 ± 0.0427	0.4435 ± 0.0526
CPGL	0.8625 ± 0.0231	0.8897 ± 0.0130	0.7546 ± 0.0323	0.7335 ± 0.0587	0.5464 ± 0.0500
GIFDTI	0.8516 ± 0.0244	0.8819 ± 0.0143	0.7656 ± 0.0174	0.7399 ± 0.0219	0.5549 ± 0.0281
FOTF-CPI	0.8712 ± 0.0213	0.8913 ± 0.0188	0.7861 ± 0.0228	0.7605 ± 0.0415	0.5993 ± 0.0334
DO-GMA	0.9141 ± 0.0188	0.9442 ± 0.0120	0.8440 ± 0.0168	0.8443 ± 0.0178	0.7709 ± 0.0305
LAM-DTI	<b>0.9393 ± 0.0101</b>	<b>0.9582 ± 0.0069</b>	0.8426 ± 0.0201	0.8483 ± 0.0221	<b>0.7067 ± 0.0298</b>
BiMA-DTI	0.9301 ± 0.0186	0.9539 ± 0.0110	<b>0.8728 ± 0.0195</b>	<b>0.8733 ± 0.0199</b>	0.7462 ± 0.0358
E3					
MGNDTI	0.9636 ± 0.0096	0.9684 ± 0.0055	0.9193 ± 0.0111	0.9181 ± 0.0107	0.8388 ± 0.0222
MolTrans	0.9457 ± 0.0166	0.9542 ± 0.0116	0.9034 ± 0.0164	0.8330 ± 0.0508	0.8069 ± 0.0324
TransformerCPI	0.9470 ± 0.0141	0.9486 ± 0.0159	0.8523 ± 0.0439	0.8262 ± 0.0660	0.7204 ± 0.0728
CPI-GNN	0.8998 ± 0.0330	0.9007 ± 0.0419	0.8351 ± 0.0409	0.8199 ± 0.0531	0.6797 ± 0.0810
BACPI	0.9438 ± 0.0166	0.9528 ± 0.0106	0.8657 ± 0.0183	0.8422 ± 0.0232	0.7494 ± 0.0292
CPGL	0.8641 ± 0.0231	0.8868 ± 0.0234	0.7644 ± 0.0456	0.7133 ± 0.0743	0.5666 ± 0.0759
GIFDTI	0.9408 ± 0.0122	0.9444 ± 0.0126	0.8523 ± 0.0314	0.8273 ± 0.0441	0.7192 ± 0.0529
FOTF-CPI	0.9438 ± 0.0096	0.9514 ± 0.0082	0.8497 ± 0.0211	0.8181 ± 0.0306	0.7238 ± 0.0349
DO-GMA	0.9832 ± 0.0044	0.9816 ± 0.0070	0.9633 ± 0.0062	0.9600 ± 0.0082	0.9280 ± 0.0120
LAM-DTI	0.9875 ± 0.0039	0.9866 ± 0.0047	0.9508 ± 0.0163	0.9370 ± 0.0224	0.9002 ± 0.0325
BiMA-DTI	<b>0.9878 ± 0.0064</b>	<b>0.9878 ± 0.0065</b>	<b>0.9691 ± 0.0101</b>	<b>0.9660 ± 0.0112</b>	<b>0.9365 ± 0.0213</b>
E4					
MGNDTI	0.7301 ± 0.0427	0.7252 ± 0.0382	<b>0.6555 ± 0.0514</b>	<b>0.7239 ± 0.0238</b>	0.3601 ± 0.0826
MolTrans	0.5983 ± 0.0279	0.6174 ± 0.0347	0.5089 ± 0.0150	0.2489 ± 0.1162	0.0779 ± 0.0440
TransformerCPI	0.6262 ± 0.0608	0.6213 ± 0.0643	0.5537 ± 0.0358	0.2706 ± 0.1324	0.1414 ± 0.0973
CPI-GNN	0.5842 ± 0.0588	0.5717 ± 0.0515	0.5166 ± 0.0423	0.3096 ± 0.1012	0.0625 ± 0.1052
BACPI	0.6662 ± 0.0405	0.6583 ± 0.0342	0.5621 ± 0.0333	0.2974 ± 0.1069	0.1715 ± 0.0724
CPGL	0.6827 ± 0.0608	0.6942 ± 0.0759	0.5028 ± 0.0389	0.0852 ± 0.1722	0.0385 ± 0.0873
GIFDTI	0.7172 ± 0.0342	0.7146 ± 0.0391	0.6221 ± 0.0495	0.4623 ± 0.1736	0.2856 ± 0.0930
FOTF-CPI	0.6561 ± 0.0355	0.6430 ± 0.0423	0.5577 ± 0.0254	0.2914 ± 0.1338	0.1653 ± 0.0635
DO-GMA	0.6556 ± 0.0188	0.6620 ± 0.0274	0.5339 ± 0.0215	0.6767 ± 0.0053	<b>0.4517 ± 0.0216</b>
LAM-DTI	0.7293 ± 0.0507	0.7093 ± 0.0433	0.5534 ± 0.0262	0.2423 ± 0.0890	0.1839 ± 0.0324
BiMA-DTI	<b>0.7437 ± 0.0503</b>	<b>0.7527 ± 0.0437</b>	0.6477 ± 0.0698	0.7169 ± 0.0284	0.3348 ± 0.1165

Note: Values in bold indicate the best performance across all models



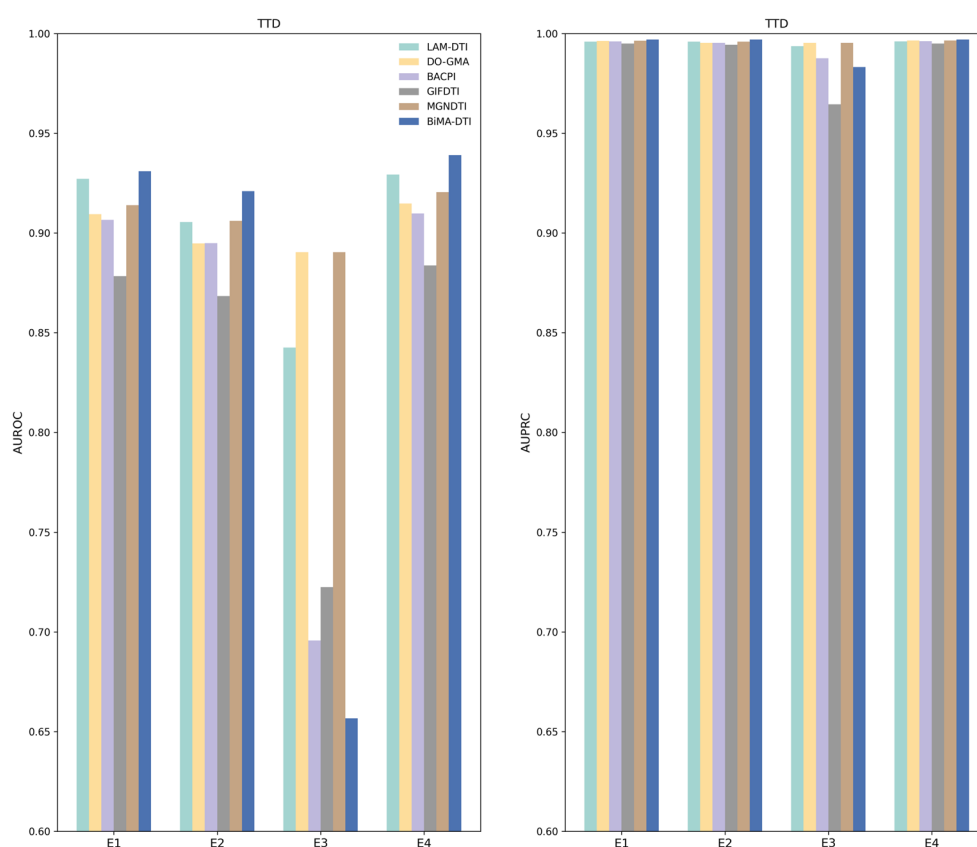


**Fig. 1** AUROC and AUPRC of BiMA-DTI and baselines on BioSNAP and BindingDB. All the data are the average of 10 repetitions. Raw data values are provided in Additional file 1: Tables S3 and S4 and Additional file 2: Sheet 1

FDA-approved drugs and clinical-stage compounds, and presents a pronounced class imbalance with a positive-to-negative sample ratio of approximately 32.7:1. These properties make TTD a realistic and challenging benchmark for testing model scalability and generalization beyond well-curated datasets. Given the large scale of the task, we selected five representative baseline methods for comparison. These were chosen based on their overall performance on a medium-scale dataset and include two of the most recent models published this year.

As shown in Fig. 2, it is obvious that the AUPRC of all models under each experimental setting is quite high and very close, especially considering the extreme class imbalance in the TTD dataset. Nevertheless, BiMA-DTI can still achieve a subtle lead in AUPRC and a significant AUROC. This indicates that BiMA-DTI can maintain reliable accuracy in a huge and unbalanced interaction space, which is crucial for virtual screening applications. In virtual screening applications, even a small increase in false positives can lead to significant downstream costs. This robustness under large-scale and unbalanced conditions emphasizes the practical applicability of the model to the real-world drug discovery pipeline.

In the drug cold start scenario (under the E2 setting), BiMA-DTI achieves an AUROC of 0.921 and again attains an AUPRC of 0.997, reflecting its strong generalization capability to previously unseen drug compounds. The target cold start setting (under the E3 setting), however, poses a greater challenge. Due to the nature of the setting, only a small subset of the TTD dataset remains (a total of 700 interactions, including 264 in the training set) after ensuring that all test proteins are unseen during training. Unlike the other experimental settings, the extremely limited sample size under the E3 setting prevents preservation of the original positive-to-negative ratio (3% negative), leading to an even more skewed distribution. This data constraint significantly impacts AUROC, which is known to be sensitive to both class balance and sample size. Accordingly, BiMA-DTI's AUROC drops to 0.657, lower than several competing methods. Nevertheless, the model maintains a high AUPRC of 0.983, suggesting that despite reduced sensitivity in this low-data regime, BiMA-DTI retains strong precision and continues to make reliable predictions. This robustness is particularly valuable when only sparse interaction data is available for novel protein targets. In the most demanding double cold start scenario (under the E4 setting),



**Fig. 2** AUROC and AUPRC of BiMA-DTI and baselines on TTD. The data of E3 is the average of 10 repetitions, while the others were only run once. Raw data values are provided in Additional file 1: Table S5 and Additional file 2: Sheet 1

BiMA-DTI regains leading performance with an AUROC of 0.939 and sustains the AUPRC benchmark of 0.997, confirming its ability to scale under extreme generalization requirements.

These results demonstrate the effectiveness of BiMA-DTI's architecture in handling large-scale, imbalanced datasets. The consistently high AUPRC across all settings underscores its robustness in real-world applications where precision is paramount. Although AUROC values naturally decline under cold-start conditions, the preserved AUPRC performance highlights the model's potential for reliable interaction verification in large-scale drug discovery pipelines.

The comparison of BiMA-DTI to the baseline models on the TTD dataset is detailed in Additional file 1: Table S5.

### Reliability analysis

The reliability of predicted interaction scores is important for applying DTI models in real-world biomedical scenarios, especially when predictions are used for ranking or threshold-based decisions. A well-calibrated model means that the predicted scores closely reflect

the actual likelihood of interaction, making the model's output more interpretable and trustworthy. While evaluation metrics such as AUROC and AUPRC can show how well a model ranks interactions, they do not indicate whether the predicted probabilities themselves are reliable. Therefore, we conducted a reliability analysis to examine how well the predicted scores of BiMA-DTI and baseline models match true outcome frequencies [50].

To perform this analysis, we constructed an external test set based on the TTD, where all positive samples are FDA-approved drug–target pairs with known clinical relevance. For comparison and simulate the real new DTI prediction environment as much as possible, we randomly extract the same number of negative pairs that do not exist in the training set under the E4 partition of the TTD dataset. Although real-world datasets are often imbalanced, using a balanced test set makes it easier to evaluate calibration behavior and avoids biased results that may arise from class imbalance [51].

On this clinically grounded evaluation set, BiMA-DTI achieved the best overall performance among all compared methods as shown in Table 3. These results indicate that BiMA-DTI retains high discriminative power



**Table 3** Comparison results of BiMA-DTI and baselines on external test set

Method	AUROC	AUPRC	Accuracy	F1-score	MCC
BACPI	0.7148	0.7200	0.5486	0.6796	0.1689
GIFDTI	0.7053	0.7119	0.5753	0.6937	0.2375
MGNDTI	0.7840	0.7748	0.6866	0.7271	0.4030
LAM-DTI	0.7892	0.7741	0.6674	0.7290	<b>0.4160</b>
DO-GMA	0.7700	0.7713	0.6940	0.7207	0.3954
BiMA	<b>0.7903</b>	<b>0.7787</b>	<b>0.6959</b>	<b>0.7320</b>	0.4069

Note: Values in bold indicate the best performance across all models

**Table 4** Calibration evaluation of BiMA-DTI and baseline models

Method	ECE	Brier
BACPI	0.0327	0.3331
GIFDTI	0.0334	0.3491
MGNDTI	0.0299	0.2998
LAM-DTI	0.0287	0.2715
DO-GMA	0.0312	0.3179
BiMA-DTI	<b>0.0253</b>	<b>0.2701</b>

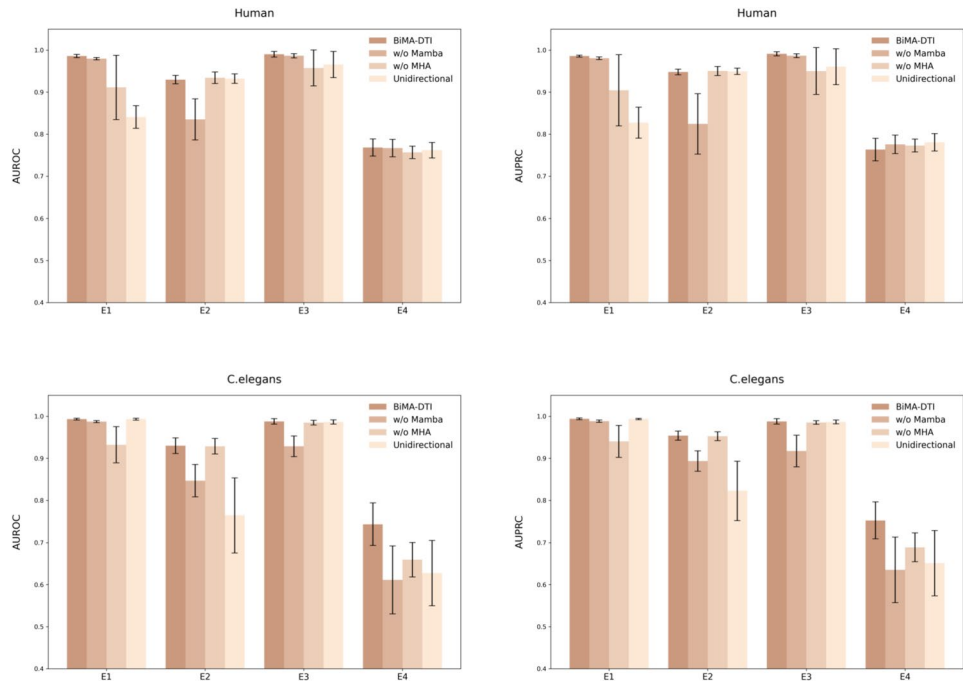
Note: Values in bold indicate the best performance across all models

even on high-confidence clinical data. MGNDTI, LAM-DTI, and DO-GMA also showed competitive performance, while BACPI and GIFDTI lagged behind. Then, we assessed model calibration through two widely

adopted metrics: expected calibration error (ECE) and Brier score. The ECE measures the average discrepancy between predicted probabilities and observed frequencies across multiple bins, while the Brier score captures the mean squared Error between predicted probabilities and ground-truth labels. The equations of ECE and Brier scores are shown as Additional file 1: Eqs. S6 and S7. As shown in Table 4, BiMA-DTI achieved the lowest ECE (0.0253) and Brier score (0.2701) among all compared models, indicating superior calibration. Other models, such as LAM-DTI and MGNDTI, also showed relatively good calibration, but with higher ECE and Brier values, suggesting room for improvement. These results demonstrate that BiMA-DTI not only excels in predictive accuracy but also provides reliable probabilistic estimates, showing its potential for real-world deployment.

**Ablation experiment**

To evaluate the impact of MAN and multi-modal input, we conducted ablation experiments on four experimental settings of the Human and *C.elegans* datasets. The full BiMA-DTI model served as the baseline. We assessed the contributions of Mamba and multi-head attention by removing each key module from MAN and comparing the resulting performance. Additionally, we developed two variants to evaluate the effectiveness of multi-modal input: SPMDTI (drug SMILES sequence + protein amino



**Fig. 3** AUROC and AUPRC of MAN ablation experiment on Human and *C.elegans*. All the data are the average of 10 repetitions. Raw data values are provided in Additional file 1: Tables S6 and S7 and Additional file 2: Sheet 2

acid sequence) and GPMDTI (drug molecular graph + protein amino acid sequence).

As shown in Fig. 3, the removal of the Mamba module and the multi-head attention mechanism led to varying degrees of performance degradation in BiMA-DTI, detailed data are shown in Additional file 1: Tables S6 and S7. The model lacking these components also demonstrated unstable predictive capabilities. This underscores the ability of MAN to not only integrate the strengths of Mamba and the multi-head attention mechanism but also to enable their complementary interaction, thereby enhancing the model's stability. Furthermore, we observed that processing unidirectional sequences not only resulted in reduced performance but also increased error rates. In conclusion, the MAN architecture's design is validated, as it significantly boosts the model's predictive accuracy while substantially improving its robustness and generalization capabilities.

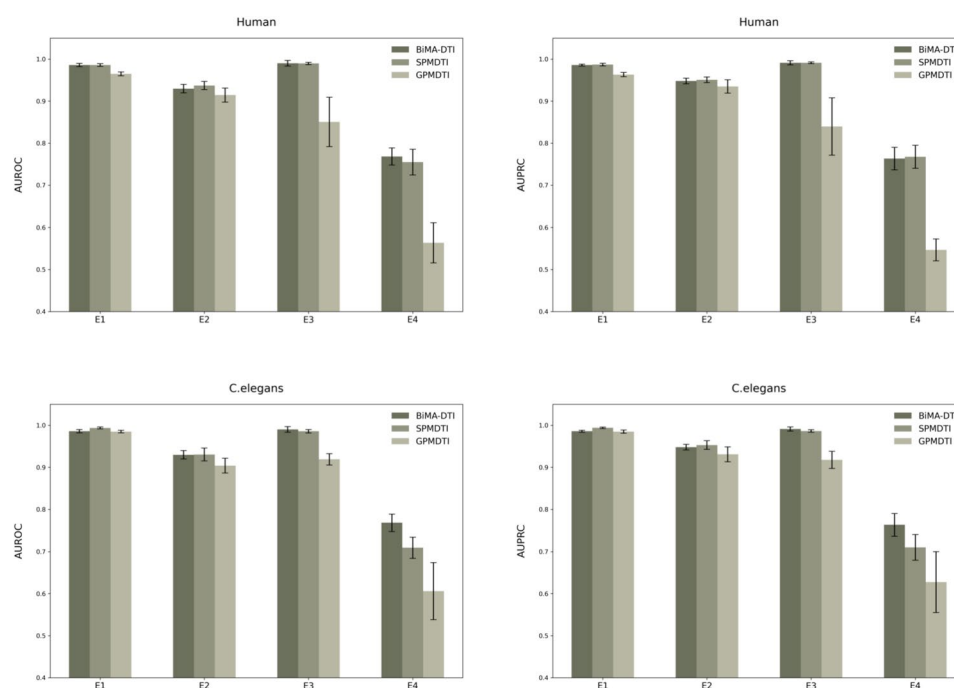
As shown in Fig. 4, the performance of BiMA-DTI does not consistently surpass all variants. Notably, SPMDTI outperforms BiMA-DTI under the E2 setting of the Human dataset and even achieves the best performance to varying degrees under the E1 setting on the *C.elegans* dataset. Since SPMDTI does not utilize graph inputs, it focuses more intensively on feature extraction from sequences, particularly the SMILES sequences of drugs. This is further corroborated by SPMDTI's strong performance under the E2 setting on both datasets. However,

SPMDTI's inability to enhance its performance in other experimental settings suggests that the structural information of drugs not only boosts the model's performance but also its generalization capabilities. In contrast, GPMDTI consistently underperforms compared to both BiMA-DTI and SPMDTI, suggesting that drug sequences play a more critical role than graphic information for BiMA-DTI. Detailed data are shown in Additional file 1: Tables S8 and S9.

### Visualization study

To further improve the interpretability of BiMA-DTI and provide insights into the model's decision-making process, we conducted a series of visual analyses targeting both the input and internal mechanisms. Interpretability is a key requirement in biomedical applications, where understanding which molecular features contribute to a model's prediction can aid in rational drug design and hypothesis generation. We analyzed the contribution of specific drug substructures and protein residues to the model outputs and explored whether these features align with known biochemical interactions or binding patterns.

First, we performed an attribution analysis on the molecular graph inputs using the Saliency method provided by the Captum interpretability library [52]. This method computes the gradient of the model's output with respect to Each input node feature, Effectively measuring the sensitivity of the prediction to perturbations in

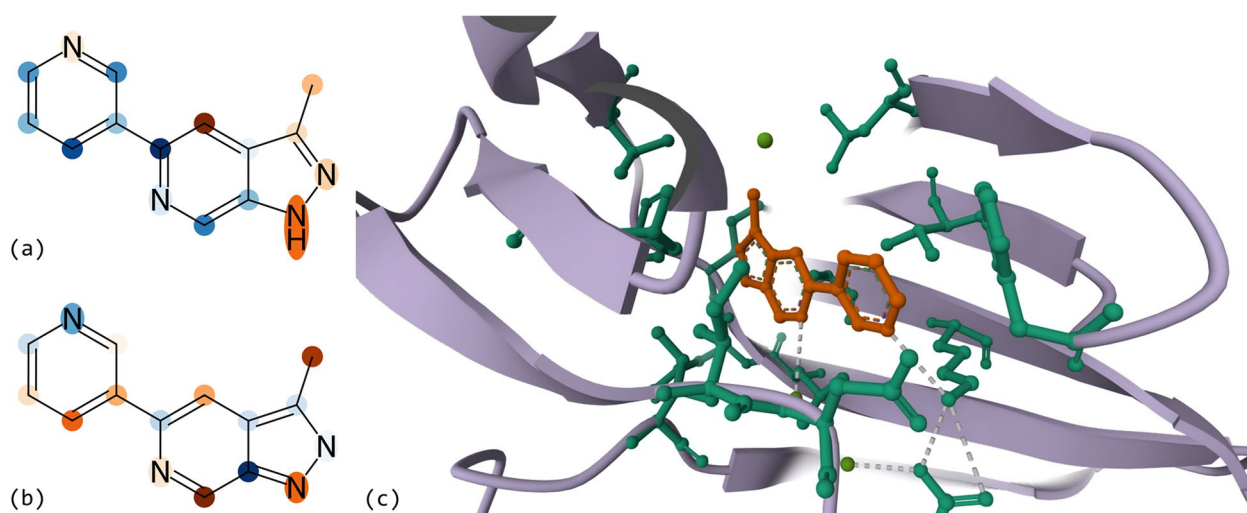


**Fig. 4** AUROC and AUPRC of drug multimodal input ablation experiment on Human and *C.elegans*. All the data are the average of 10 repetitions. Raw data values are provided in Additional file 1: Tables S8 and S9 and Additional file 2: Sheet 2

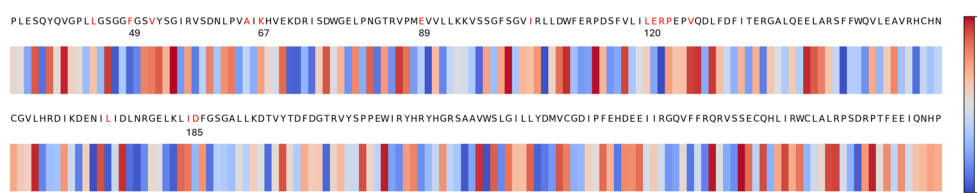
atom level representations. We visualized the resulting saliency scores on 2D molecular structures, where atoms with higher attribution scores were highlighted in deeper orange, indicating greater importance in driving the predicted DTI. This analysis helps to reveal which substructures within the molecule are most influential in the final decision. In addition to input-level attribution, we also visualized the Mamba-Attention weights produced by the model's MAN to examine how the model attends to different components of the drug and protein during interaction prediction. Specifically, we mapped the embedding values back to the drug's 2D molecular graph and the protein's amino acid sequence, the latter visualized as a chart. These offer an interpretable view into how the model integrates information from both modalities, and allow for potential comparison with known active sites or binding motifs in future analyses. For each analyzed drug target pair, we used the three dimensional (3D) binding sites provided in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) as the real reference. We selected two protein ligand pairs and analyzed the small molecule ligand in them: The Proviral Insertion site for Moloney murine lymphoma virus (Pim) kinase family (PDB ID:5DHJ)

involved in the discovery of Pim kinase inhibitors for the treatment of tumor diseases [53]. Another is orally available Factor7a (PDB ID:2BZ6) inhibitor which includes the clotting factor VIIa and is a target for antithrombotic therapy [54].

For the drug, as shown in Fig. 5a and b, the two 2D molecular diagrams of the attribution analysis and Mamba-Attention Weights both label the key nitrogen atoms involved in the binding of the 5DHJ drug, and Fig. 5c also shows that these nitrogen atoms are very close to the molecules at the binding sites. Additional file 1: Fig. S1(a) and (b) respectively highlight the oxygen, nitrogen, and fluorine atoms involved in the binding of the 2BZ6 drug molecule, and Additional file 1: Fig. S1(c) also proves that these atoms are involved in the binding with the protein. For the protein, as shown in Fig. 6, the Weight of the 5DHJ protein amino acid after MAN reaches the maximum within approximately the three intervals of 40–60, 100–120, and 180–200, which coincides with the actual binding amino acid positions in the upper part of Fig. 6. Similarly, MAN focused on the amino acids within the 30–50, 80–100, and 170–190 intervals of the 2BZ6 protein, and the significance of



**Fig. 5** The 2D molecular graph of PDB ID:5DHJ drug and the 3D binding site structure from RCSB PDB. **a** Visualization of the importance of the Saliency method. **b** Visualization of Mamba-Attention weights. **c** The position of the amino acid sequence bound to the drug and the 3D binding site structure from RCSB PDB. The 2D molecular graph of PDB ID:5DHJ drug and the 3D binding site structure from RCSB PDB



**Fig. 6** The importance weight of PDB ID:5DHJ protein and the position of the amino acids involved in the combination

these intervals was also confirmed in Additional file 1: Fig. S2.

In addition, we can notice that the 2D molecular diagrams of drugs drawn by the two visualization methods are complementary, that is, the key atoms not covered by one method will be covered by the other method. This further proves the validity of the input of the BiMA-DTI molecular map. These outstanding results indicate that BiMA-DTI can capture key information to a considerable extent and has good biological interpretability. However, we cannot ignore the atoms that have been wrongly focused on, which also indicates that BiMA-DTI has limited perception of structural information, or it may potentially reveal previously unidentified local interaction sites.

Case study

To further evaluate the predictive capability and biological interpretability of BiMA-DTI, we conducted case studies from both drug and target perspectives using the DrugBank dataset [55]. Two drugs (DB00143 and DB09068) and two protein targets (P54284 and P34969) were selected, each with more than five known positive or negative interactions. These molecules were excluded from the training set, and BiMA-DTI was used to predict their interactions with all potential partners. The detailed prediction outcomes, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), are presented in Table 5. For the prediction results of Each drug and target, only 10 are displayed in the order of prediction scores (5 of the highest scores for positive cases and 5 of the lowest scores for negative cases), as shown in Tables 6 and 7.

From the drug perspective, BiMA-DTI achieved highly accurate predictions. For DB00143, a redox-active tripeptide involved in detoxification processes [56], the model correctly predicted 40 out of 42 true interaction cases, with only 6 false positives and 2 false negatives. Similarly, DB09068 (Vortioxetine), a multimodal serotonin modulator [57], showed 4 true positives and 7 true negatives, with only one FP and three FNs. These results indicate that BiMA-DTI can robustly capture molecular

**Table 5** Prediction results for selected drugs and protein targets in the case study

DrugBankID/UniProtID	TP	TN	FP	FN
DB09068 (Drug)	4	7	1	3
DB00143 (Drug)	40	4	6	2
P54284 (Target)	3	8	4	12
P34969 (Target)	3	10	1	34

**Table 6** Prediction results of drug DB00143 and DB09068

DrugBankID	UniProtID	True label	Predict label
DB00143	P18283	True	True
	P28161	True	True
	P09488	True	True
	P35754	True	True
	P22352	True	True
	Q9FBC5	False	False
	P13827	False	False
	P43235	False	False
	Q9NQX3	False	False
	P24310	False	True
DB09068	P08588	True	True
	P31645	True	True
	P08908	True	True
	P28222	True	True
	P46098	True	False
	P13929	False	False
	Q13423	False	False
	Q06528	False	False
	Q16659	False	False
	P53985	False	False

**Table 7** Prediction results of target protein P54284 and P34969

UniProtID	DrugBankID	True label	Predict label
P54284	DB00273	True	True
	DB04855	True	True
	DB01388	True	True
	DB09231	True	False
	DB00825	True	False
	DB00261	False	False
	DB02521	False	False
	DB06515	False	False
	DB08256	False	False
	DB13868	False	False
P34969	DB00216	True	True
	DB01224	True	True
	DB13345	True	True
	DB08815	True	False
	DB00248	True	False
	DB07726	False	False
	DB02192	False	False
	DB00992	False	False
	DB00119	False	False
	DB07958	False	False

interaction patterns, particularly for well-annotated small molecules.

From the target perspective, predictions for P54284 and P34969 were more challenging. For P54284, a kinase associated with hematological malignancies [58], the model achieved 3 true positives and 8 true negatives, but misclassified 4 false positives and 12 false negatives. Many of the missed drugs, including DB09231 (Pomalidomide) and DB00825 (Tamoxifen), are known to engage in indirect or condition-specific interactions, which may not be fully captured through sequence-level features. Likewise, for P34969, a coagulation-related protease [59], BiMA-DTI correctly identified 3 interacting drugs and 10 non-interacting ones, while generating 1 FP and 34 FNs. This large number of false negatives can be attributed to the strong reliance of these anticoagulants on 3D conformational binding, which our model currently does not explicitly encode.

In summary, while BiMA-DTI performs reliably on the drug side with relatively few misclassifications, performance on certain targets is more sensitive to context-dependent and structure-based factors. These case studies reinforce the model's strengths in capturing global interaction trends and suggest its potential applicability in drug repurposing, where existing compounds are evaluated for novel targets. At the same time, the results highlight areas where incorporating structural priors or binding site information could further enhance prediction fidelity in future work.

## Discussion

In practical drug discovery pipelines, particularly during early-phase virtual screening, the ability to efficiently and accurately evaluate large chemical libraries is crucial. Although our evaluation primarily focused on benchmark datasets with controlled scenarios (e.g., balanced test sets and double cold start settings), BiMA-DTI is inherently compatible with large-scale screening tasks due to its End-to-end architecture and minimal reliance on structural prerequisites. Specifically, BiMA-DTI operates directly on raw molecular graphs and amino acid sequences, without requiring 3D coordinates or pre-defined binding pockets. This design choice makes the model particularly advantageous when structural data are incomplete, unavailable, or unreliable conditions that frequently occur in early-stage drug discovery campaigns. Once trained, BiMA-DTI can rapidly compute interaction scores for massive drug–target pairings in a single forward pass, providing scalable and cost-effective virtual screening capabilities.

At the same time, we acknowledge that this structural simplicity may also limit the model's capacity to capture

fine-grained spatial features that are Essential to molecular recognition. While BiMA-DTI achieves strong performance using only sequence and graph-based modalities, it may overlook critical geometric determinants of binding affinity, Such as 3D conformations, binding pocket orientations, or intermolecular steric effects. This limitation is particularly evident in our visualization study, where predicted important residues or substructures sometimes diverge from experimentally validated binding regions. These observations suggest that while the current design is efficient and generalizable, it may benefit from the inclusion of structural cues in cases where accuracy is prioritized over speed.

Incorporating 3D information into BiMA-DTI presents both opportunities and challenges. On one hand, integrating protein-ligand structural features could significantly improve binding site localization and affinity Estimation. On the other hand, 3D inputs dramatically increase the complexity of the model, both in terms of data dimensionality and computational cost. Atom level coordinates introduce additional input channels, such as spatial distances, angles, or torsions. That expand memory requirements and slow inference, which is problematic for large-scale applications. Furthermore, many DTI benchmark datasets do not consistently provide high-quality structural data, necessitating preprocessing steps like docking, homology modeling, or structure prediction. These procedures can introduce additional noise and uncertainty and may not scale efficiently [60, 61].

Nevertheless, future Extensions of BiMA-DTI could incorporate 3D information in a modular and selective fashion. For proteins, coupling with structure prediction models to identify putative binding pockets, followed by voxelization or point cloud construction, would Enable the use of 3D GNN [62–64]. For Ligands, 3D conformers can be generated [65], and distance or angle features embedded into molecular graphs, allowing geometric message passing. These structural embeddings could then be fused with our existing sequence and graph-based representations through multimodal integration. Such extensions offer a promising path to enhance accuracy while preserving the scalability of the current framework.

Finally, the multimodal fusion mechanism itself invites further refinement. Although our Multimodal Weight Fusion (MMWF) module effectively integrates sequence and graph features, real-world scenarios often involve conflicting or noisy information. For example, a drug's SMILES encoding may indicate strong binding motifs, while its graph representation lacks critical substructures. Future versions of BiMA-DTI could incorporate hierarchical attention mechanisms to dynamically resolve such conflicts, selectively prioritizing high confidence



signals and attenuating less reliable inputs. This would further enhance the interpretability and robustness of the model across diverse biomedical datasets and screening environments.

Conclusions

In this paper, we present a model called BiMA-DTI for DTI prediction, which is based on the Mamba model and multi-head attention mechanism. Under four different experimental setups of four medium-scale datasets and one large-scale dataset, BiMA-DTI achieved the best prediction performance. Additionally, we conducted ablation experiments to verify the effectiveness of the core modules and multimodal input, visualization study further proves that. Finally, we conduct a case study to further demonstrate the performance of the model.

In summary, BiMA-DTI has surpassed sota in performance and has good DTI prediction capabilities. However, we also noticed that in real-world prediction tasks, BiMA-DTI still faces the challenge of reduced accuracy. Therefore, in subsequent research, we will focus on solving more complex DTI prediction problems.

Methods

Dataset

Our study uses four medium-scale and one large-scale DTI datasets to train and evaluate the model's performance; detailed descriptions of these five datasets are provided in Table 8.

Medium-scale datasets

The *C.elegans* and Human datasets include balanced positive and negative samples, providing the comprehensive information on known and clinical kinase inhibitors across the entire kinase family [43]. It fully encompasses the human protein kinase family, with positive samples sourced from highly reliable biochemical databases, including the DrugBank and Matador databases [44, 45]. The BioSNAP dataset, constructed by reference [35, 46], is derived from DrugBank data and includes an equal number of validated positive DTIs and randomly selected negative DTIs from unseen drug-target pairs. The BindingDB dataset [47] focuses primarily on experimentally validated binding affinities

between small molecules and proteins. In this study, we utilize a low-bias version of BindingDB, which is constructed by reference [48].

Large-scale dataset

Additionally, we applied a large-scale dataset by collecting all drug-target compound data from TTD [30]. We carefully selected data with experimental binding affinity measurements (Ki/Kd/IC50/EC50/AC50/Potency), including outliers or null values will be deleted, and determined DTI based on these metrics. Following the study by reference [31], interactions were considered non-existent when these metrics Exceeded 100 μM. To ensure higher reliability, we adopted more stringent criteria (exceeded 10μM) for negative sample identification.

Problem formulation

In this section, we first introduce some definitions and notation used in our model and then formulate the problem. A drug is denoted as  $\mathcal{D} = (\mathcal{S}, \mathcal{G})$ , where  $\mathcal{S} = s_1, s_2, s_3, \dots, s_n$  is a set of SMILES sequences sets  $s_i$  with  $n$  symbols and  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is a molecular graph with the atom set  $\mathcal{V}$  and the set  $\mathcal{E}$  of undirected edges between atoms. A protein is denoted as  $\mathcal{P} = (a_1, a_2, a_3, \dots, a_m)$  with  $m$  amino acids sets  $a_j$ . Thus, the potential DTI identification task is to learn a function  $\mathcal{F} : \mathcal{D} \times \mathcal{P} \rightarrow [0, 1]$  for computing interaction probability of each drug-target pair. While proteins can also be represented as graphs, it is more difficult because the tertiary structure is not always available in a reliable form. Therefore, we abandon the graph representation of proteins and opted to use the widely adopted one-hot encoding method to represent proteins. The List of atom and Edge features is summarized in Additional file 1: Tables S10 and S11.

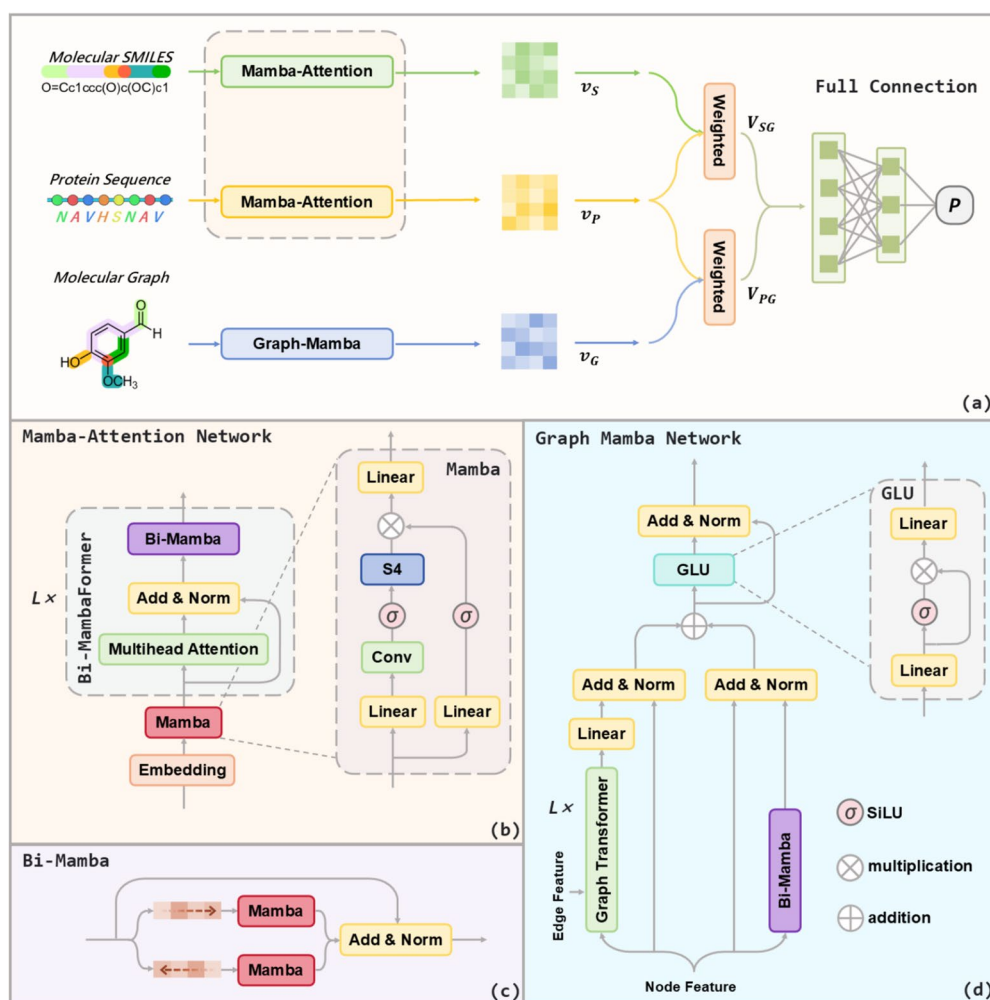
BiMA-DTI

As shown in Fig. 7a, BiMA-DTI mainly consists of four modules, namely MAN, GMN, feature fusion network, and predictor. The MAN module extracts sequence features from the drug's SMILES and the target protein's amino acid sequence, while the GMN module captures the structural representation of the drug. The MAN and GMN modules form the backbone of BiMA-DTI, leveraging the strengths of Attention and Mamba architectures to comprehensively capture the information for drugs and targets. The extracted sequence and structural features from both modules are then integrated within the feature fusion network. The fused features are then fed into a fully connected layer, which makes the prediction of an interaction occurring the drug and the target

Table 8 Description of DTI datasets

Dataset	Drug	Protein	Interaction	Positive	Negative
Human	2726	2001	5997	2633	3364
<i>C.elegans</i>	1767	1876	7785	3893	3892
BioSNAP	4505	2181	27464	13830	13634
BindingDB	14643	2623	49199	20674	28525
TTD	415616	1477	733489	711743	21746





**Fig. 7** **a** The whole process of BiMA-DTI. **b** Details of Mamba-Attention Network and Mamba module. **c** Details of Bi-Mamba. **d** Details of Graph-Mamba Network and GLU

protein. The following sections provide detailed explanations of these components.

### Mamba-Attention Network

We design a MAN module to extract information from the protein amino acid and drug molecular SMILES inputs. As shown in Fig. 7b, the MAN module includes an embedding layer and Bi-MambaFormer layers.

**Embedding layer** In our study, Each amino acid in the target protein amino acid sequence is represented by a unique letter, with Each of the 25 amino acids assigned a corresponding integer value (e.g., “A”: 1, “C”: 2, “D”: 3). Similarly, the SMILES notation of a drug is encoded using integer values, where each symbol is mapped to a specific integer (e.g., “C”: 1, “H”: 2, “N”: 3). The embedding layer consists of two components: a learnable word embedding layer and a positional embedding. The learnable word

embedding layer transforms the integer values of amino acids and drug SMILES symbols into high-dimensional vectors. Given the amino acid and SMILES as  $x_s$  and  $x_p$ , we can obtain the sequence representation of drug and target as  $E_s \in R^{n \times d}$  and  $E_p \in R^{m \times d}$ , respectively, via the learnable word embedding layer. Here,  $n$  and  $m$  represent the lengths of the protein amino acid sequence and SMILES, while  $d$  denotes the embedding size.

Positional encodings allow the model to exploit the order of a sequence by injecting some information about the relative or absolute position of tokens in the sequence. Common types of positional encoding include sinusoidal encoding [21], learnable encoding [66], relative positional encoding [67], and rotary positional encoding [68]. In this study, we employ the Mamba as positional embedding layer. Mamba can be regarded as a RNN where the hidden state at the current sequence position is updated based on the hidden state at the previous position. Such

recurrence mechanism to process tokens enables Mamba naturally consider order information of sequences [28]. As a result, the Mamba is leveraged to internally incorporate positional information into the embeddings. The SSM model processes input  $x[i]$  at sequence position  $i$  as Eq. 1:

$$\begin{aligned} h[i] &= Ah[i-1] + Bx[i] \\ SSM(x[i]) &= Ch[i] \end{aligned} \quad (1)$$

where  $h[i]$  is the hidden state at sequence position  $i$ .  $A$ ,  $B$ , and  $C$  are learnable matrices, and  $A$ ,  $B$  is discrete form. Therefore, a standard Mamba  $Mamba(e)$  can be presented as Eq. 2:

$$\begin{aligned} z &= Conv(Linear(e)) \\ E &= MLP(SSM(\sigma(z)) \cdot \sigma(z)) \end{aligned} \quad (2)$$

where  $Conv$  refers convolution.  $\sigma$  denotes SiLU operation [69] and  $\cdot$  denotes the multiplication.  $SSM$  represents the SSM model. So the drug and protein amino acid sequence vectors are processed by Mamba, as shown in Eq. 3:

$$\begin{aligned} H_{s1} &= Mamba(E_s) \\ H_{p1} &= Mamba(E_p) \end{aligned} \quad (3)$$

where  $H_{s1} \in R^{n \times d}$  and  $H_{p1} \in R^{m \times d}$ .

**Bi-MambaFormer layer** A single Mamba model executes the selection mechanism in only one direction, which can limit its ability to capture global dependencies [70]. To address this limitation, we propose Bi-MambaFormer layer to process both the original and reverse input sequences to improve performance, which integrates Bi-Mamba and Multi-head Attention mechanisms, as shown in Fig. 7b and c. The Multi-head Attention mechanism focuses on extracting local sequence information. Simultaneously, the Bi-Mamba component consists of two Mamba blocks that capture global sequence information in two directions. Additionally, residual connections are incorporated to enhance the overall performance and stability of the model. More specifically, given the embeddings of protein amino acid sequence and SMILES, the  $l$ th Bi-MambaFormer layer can be represented as Eq. 4:

$$\begin{aligned} H_{s2}^l &= MultiHead(H_{s1}^l) + H_{s1}^l \\ H_{p2}^l &= MultiHead(H_{p1}^l) + H_{p1}^l \\ H_{s1}^{l+1} &= Mamba(H_{s2}^l) + Mamba(H_{s2}^{l'}) + H_{s2}^l \\ H_{p1}^{l+1} &= Mamba(H_{p2}^l) + Mamba(H_{p2}^{l'}) + H_{p2}^l \end{aligned} \quad (4)$$

where  $H_{s2}^l \in R^{n \times d}$  and  $H_{p2}^l \in R^{m \times d}$  represent the intermediate states of the protein amino acid sequence and

SMILES, respectively, and  $H_{s2}^{l'}$  and  $H_{p2}^{l'}$  denote reverse sequences states. The Multi-head Attention function  $MultiHead(x)$  is defined as Eq. 5:

$$\alpha(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

$$MultiHead(x) = Concat(head_1, \dots, head_h)W^O$$

where  $head_i = \alpha(xW_i^Q, xW_i^K, xW_i^V)$ .

By stacking multiple Bi-MambaFormer layers, the model can progressively extract richer and more nuanced information from the input sequences. The final outputs,  $v_P \in R^d$  for the protein amino acid sequence and  $v_S \in R^d$  for the SMILES notation, encapsulate the comprehensive feature representations generated by the Bi-MambaFormer layers.

### Graph Mamba Network

We propose a hybrid Graph Transformer and Bi-Mamba to model the drug structure, as shown in Fig. 7d. The Graph Transformer employs the attention mechanism to assess the significance of each neighboring node relative to the current node, taking into account both node and edge features [71]. Meanwhile the Bi-Mamba network also learns node features, which leverages its selection mechanism to minimize the influence of less relevant nodes. This combination allows the model to effectively capture both local and global structural information, enhancing the representation of the drug's molecular structure [72].

**Graph Transformer layer** Given the drug structure graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , atom feature as  $H_i \in \mathcal{V}$  and edge feature as  $e_{ij} \in \mathcal{E}$ , we can obtain the  $i$ th node and edge  $(i, j)$  representation of  $\mathcal{G}$  as  $H_i$  and  $e_{ij}$ , respectively. The  $l$ th Graph Transformer layer is a non-linear transformation function that maps node  $i$ 's embedding  $H_i^{l-1} \in R^d$  to  $H_i^l \in R^d$  for  $i \in N$ ,  $N$  is the number of nodes in  $\mathcal{G}$  and  $d$  denotes the embedding size. When  $l = 0$ , the embedding  $H_i^0 = H_i \in R^d$  is just the node feature of  $i$ th node. In addition,  $e_{ij} \in R^{d_e}$ , where  $d_e$  is the dimension of input edge features.

Formally, in the  $l$ th Graph Transformer layer, the hidden representation  $H_i^l$  is updated by performing a message passing between node  $i$  and its neighbors, as Eq. 6:

$$H_i^l = W_1 H_i^{l-1} + \sum_{j \in \mathcal{N}(i)} \alpha_{ij} (W_2 H_j^{l-1} + W_3 e_{ij}) \quad (6)$$

where  $\mathcal{N}(i)$  is the set of neighbor nodes of node  $i$  in the graph,  $W_1 \in R^{d \times d}$ ,  $W_2 \in R^{d \times d}$ , and  $W_3 \in R^{d \times d}$  are

learnable parameters, and  $\alpha_{ij}$  is the attention weight used to aggregate messages. The weights  $\alpha_{ij}$  are computed using self-attention as Eq. 7:

$$\alpha_{ij} = \text{softmax}\left(\frac{(W_4 H_i^{l-1})^T (W_5 H_j^{l-1} + W_3 e_{ij})}{\sqrt{d}}\right) \quad (7)$$

where  $W_4 \in R^{d \times d}$  and  $W_5 \in R^{d \times d}$  are learnable parameters. After the final layer, we perform residual connection operations on all node features as the outputs:

$$H_{li} = H_i^L + H_i^0 \quad (8)$$

where  $H_{li} \in R^d$  is the final output of the  $l$ th atom feature after stacked multiple Graph Transformer layers.

**Bi-Mamba layer** We use the Bi-Mamba network to independently analyze node features in both forward and backward directions as shown in Fig. 7c, capturing global dependencies across the graph:

$$H_{gi} = \text{Mamba}(H_i) + \text{Mamba}(H'_i) + H_i \quad (9)$$

where  $H_{gi} \in R^d$  is the output of the  $i$ th atom feature after Bi-Mamba layer.

The output from the Graph Transformer is element-wise summed with the output from the Bi-Mamba. The resulting enriched representation is then passed through a global max pooling layer, which aggregates the information across all nodes to generate a global-level embedding  $v_G \in R^d$ .

### Multimodal feature fusion

We propose MMWF module to calibrate the learned features from protein amino acid sequence features, the SMILES and graph of drug. The MMWF employs a two-step weighted fusion process to combine these diverse features effectively.

Initially, a combined feature vector  $f_{input}$  is constructed by concatenating two input features  $f_1$  and  $f_2$ . This vector is processed through a feedforward neural network (FNN), followed by a sigmoid function to generate the first weight  $W_1$ . Using  $W_1$ , the inputs are weighted and concatenated again to form an intermediate feature vector  $f_{mid}$ . Next,  $f_{mid}$  is fed into another FNN, and its output is processed by a sigmoid function to produce the second weight  $W_2$ . Finally, the inputs are weighted once more using  $W_2$  and concatenated to produce the final fused output  $f_{out}$ .

This two-stage weighting mechanism ensures a refined and balanced fusion of  $f_1$  and  $f_2$ , enhancing the representation of their combined information. The

weighted feature fusion network  $MMWF(f_1, f_2)$  can be formulated as Eq. 10:

$$\begin{aligned} f_{input} &= \text{Concat}(f_1, f_2) \\ W_1 &= \theta(\text{FNN}(f_{input})) \\ f_{mid} &= \text{Concat}(f_1 W_1, f_2(1 - W_1)) \\ W_2 &= \theta(\text{FNN}(f_{mid})) \\ f_{out} &= \text{Concat}(f_1 W_2, f_2(1 - W_2)) \end{aligned} \quad (10)$$

where the  $f_1 \in R^d$  and  $f_2 \in R^d$  are inputs to be fused and  $\text{Concat}$  is a function to concatenate the two input features.  $\theta$  is the sigmoid function.  $\text{FNN}$  is a simple MLP maps the input dimension from  $2d$  to  $d$ . Therefore,  $f_{input} \in R^{2d}$ ,  $f_{mid} \in R^{2d}$ ,  $f_{out} \in R^{2d}$ ,  $W_1 \in R^d$ , and  $W_2 \in R^d$ .

Then, a fused feature can be denoted as vector  $V_{PS}$  for protein amino acid features and drug SMILES features and as  $V_{PG}$  for the fusion of protein features and drug molecular graph features. Finally, we concatenate  $V_{PS}$  and  $V_{PG}$  to obtain the final features  $V$ :

$$\begin{aligned} V_{PS} &= MMWF(v_P, v_S) \\ V_{PG} &= MMWF(v_P, v_G) \\ V &= \text{Concat}(V_{PS}, V_{PG}) \end{aligned} \quad (11)$$

where  $V_{PS} \in R^{2d}$ ,  $V_{PG} \in R^{2d}$ , and  $V \in R^{4d}$ .

### Prediction layer

Our prediction layer is a three fully connection network using GELU as the activation function. The total feature passes through the prediction layer to output the predicted interaction probability  $p \in R$  of drug molecular and protein.

Then, the binary cross entropy is adopted as the loss function to train the model, which is defined as Eq. 12:

$$L = \frac{1}{n} \sum_{i=1}^n -w_i [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (12)$$

where  $n$  represents the number of training set and  $y_i$  is the real label of a given drug protein pair.

### Abbreviations

AUROC	Area under the receiver operating characteristic curve
AUPRC	Area under the precision-recall curve
CNNs	Convolutional neural networks
DTIs	Drug-target interactions
ECE	Expected calibration error
FCN	Fully connected network
FDA	The US Food and Drug Administration
FN	False negatives
FP	False positives
GCN	Graph convolutional networks
GNNs	Graph neural networks
GMN	Graph Mamba Network
MAN	Mamba-Attention Network
MCC	Matthews' correlation coefficient

MMWF	Multimodal Weight Fusion
PDB	Protein Data Bank
Pim	Proviral Insertion site for Moloney murine lymphoma virus
PR	Precision-recall
RCSB	Research Collaboratory for Structural Bioinformatics
RF	Random Forest
RNNs	Recurrent neural networks
ROC	Receiver operating characteristic
SSM	State Space Model
SVM	Support vector machine
SMILES	Simplified Molecular Input Line Entry System
TN	True negatives
TP	True positives
TTD	Therapeutic Target Database
TPE	Tree-structured Parzen Estimator

## Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-025-02407-4>.

Additional file 1: Supplementary Information. This file contains the supplementary materials supporting the main text, including additional tables, equations, and figures. Tables S1-S2. The hardware facilities for model training, hyperparameter settings, and rates. Tables S3-S5. Comparison results of BiMA-DTI and baselines on BindingDB, BioSNAP and TTD. Tables S6-S9. Ablation experiments results about MAN and Multimodal on Human and *C.elegans*. Tables S10-S11. The list of atom and edge features. Equations S1-S7. The equations for evaluation metrics used to assess the performance of the model. Fig. S1. The 2D molecular graph of PDB ID:2BZ6 drug and the 3D binding site structure from RCSB PDB. Fig. S2. The importance weight of PDB ID:2BZ6 protein and the position of the amino acids involved in the combination.

Additional file 2: Supporting Data Values. This file contains all the experimental data of BiMA-DTI. Sheet 1. Comparison results of BiMA-DTI with baselines on five datasets, along with FDA-approved evaluation and calibration evaluation results. Sheet 2. Ablation experiments about MAN and multimodal on Human and *C.elegans*.

## Acknowledgements

We also appreciate the contributions of all authors involved in this study, whose collaboration made this research possible.

## Authors' contributions

LY, WJJ, and SYY conceived the study. SYY, GXW, and CC designed the algorithm and wrote and revised the paper. LY, WJJ, and HJ contributed the idea and revised the paper. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Key Research & Development Plan of Ministry of Science and Technology of China (Grant no. 2023YFC3605800, 2023YFC3605801).

## Data availability

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The medium-scale datasets and source codes are available at <https://github.com/YouyuanShui/BiMA-DTI> and <https://doi.org/10.5281/zenodo.16935249>, and the large-scale dataset is available at <https://doi.org/10.6084/m9.figshare.29974717.v1>. All supporting data values underlying the figures are available in Additional file 1 and Additional file 2.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 22 February 2025 Accepted: 3 September 2025

Published online: 15 October 2025

## References

- Zhang Y, Hu Y, Han N, Yang A, Liu X, Cai H. A survey of drug-target interaction and affinity prediction methods via graph neural networks. *Comput Biol Med*. 2023;163:107136.
- Heikamp K, Bajorath J. Support vector machines for drug discovery. *Expert Opin Drug Discov*. 2014;9(1):93–104.
- Ezzat A, Wu M, Li XL, Kwok CK. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinformatics*. 2016;17(Suppl 19):S09.
- Jacob L, Vert JP. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*. 2008;24(19):2149–56.
- Shi W, Yang H, Xie L, Yin XX, Zhang Y. A review of machine learning-based methods for predicting drug–target interactions. *Health Inf Sci Syst*. 2024;12(1):30.
- Chen L, Jiang J, Dou B, Feng H, Liu J, Zhu Y, et al. Machine learning study of the extended drug–target interaction network informed by pain related voltage-gated sodium channels. *Pain*. 2024;165(4):908.
- Li Z, Huang R, Xia M, Patterson TA, Hong H. Fingerprinting interactions between proteins and ligands for facilitating machine learning in drug discovery. *Biomolecules*. 2024;14(1):72.
- He Z, Zhang J, Shi XH, Hu LL, Kong X, Cai YD, et al. Predicting drug-target interaction networks based on functional groups and biological features. *PLoS ONE*. 2010;5(3):e9603.
- Öztürk H, Özgür A, Ozkirimli E. Deepdta: deep drug-target binding affinity prediction. *Bioinformatics*. 2018;34(17):i821–9.
- Öztürk H, Ozkirimli E, Özgür A. WideDTA: prediction of drug-target binding affinity [Internet]. 2019. <http://arxiv.org/abs/1902.04166>. Cited 15 Jan 2025.
- 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(6):e1007129.
- Zhong KY, Wen ML, Meng FF, Li X, Jiang B, Zeng X, et al. MMDTA: a multimodal deep model for drug-target affinity with a hybrid fusion strategy. *J Chem Inf Model*. 2024;64(7):2878–88.
- Kavipriya G, Manjula D. Drug-target interaction prediction model using optimal recurrent neural network. *Intell Autom Soft Comput*. 2023;35(2):1675–89.
- Wang YB, You ZH, Yang S, Yi HC, Chen ZH, Zheng K. A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. *BMC Med Inform Decis Mak*. 2020;20(2):49.
- Elbasani E, Njimboum SN, Oh TJ, Kim EH, Lee H, Kim JD. GCRNN: graph convolutional recurrent neural network for compound–protein interaction prediction. *BMC Bioinformatics*. 2022;22(5):616.
- Karim MR, Cochez M, Jares JB, Uddin M, Beyan O, Decker S. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network. In: *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '19)*. New York: Association for Computing Machinery; 2019. pp. 113–23. <https://doi.org/10.1145/3307339.3342161>.
- Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions. *J Chem Inf Model*. 2019;59(10):4131–49.
- Jiang D, Hsieh CY, Wu Z, Kang Y, Wang J, Wang E, et al. InteractionGraphNet: a novel and efficient deep graph representation learning framework for accurate protein-ligand interaction predictions. *J Med Chem*. 2021;64(24):18209–32.
- Yang Z, Zhong W, Lv Q, Dong T, Yu-Chian CC. Geometric interaction graph neural network for predicting protein-ligand binding affinities from 3D structures (GIGN). *J Phys Chem Lett*. 2023;14(8):2020–33.



20. E Z, Qiao G, Wang G, Li Y. GSL-DTI: Graph structure learning network for drug-target interaction prediction. *Methods*. 2024;223:136–45.
21. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2023. <http://arxiv.org/abs/1706.03762>. Cited 15 Jan 2025.
22. Chen L, Tan X, Wang D, Zhong F, Liu X, Yang T, et al. TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*. 2020;36(16):4406–14.
23. Wang G, Zhang X, Pan Z, Rodríguez Patón A, Wang S, Song T, et al. Multi-Transdti: transformer for drug-target interaction prediction based on simple universal dictionaries with multi-view strategy. *Biomolecules*. 2022;12(5):644.
24. Monteiro NRC, Oliveira JL, Arrais JP. DTITR: End-to-end drug–target binding affinity prediction with transformers. *Comput Biol Med*. 2022;147:105772.
25. Gao M, Zhang D, Chen Y, Zhang Y, Wang Z, Wang X, et al. Graphormerdti: a graph transformer-based approach for drug-target interaction prediction. *Comput Biol Med*. 2024;173:108339.
26. Wang H, Guo F, Du M, Wang G, Cao C. A novel method for drug-target interaction prediction based on graph transformers model. *BMC Bioinformatics*. 2022;23(1):459.
27. Lin Z, Feng M, Santos CN, Yu M, Xiang B, Zhou B, et al. A structured self-attentive sentence embedding. 2017. <http://arxiv.org/abs/1703.03130>. Cited 15 Jan 2025.
28. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. 2024. <http://arxiv.org/abs/2312.00752>. Cited 15 Jan 2025.
29. Liu J, Liu M, Wang Z, An P, Li X, Zhou K, et al. RoboMamba: efficient vision-language-action model for robotic reasoning and manipulation. 2024. <https://arxiv.org/abs/2406.04339>. Cited 15 Jan 2025.
30. Zhou Y, Zhang Y, Zhao D, Yu X, Shen X, Zhou Y, et al. TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Res*. 2024;52(D1):D1465–77.
31. Tomašić T, Zubrienė A, Skok Ž, Martini R, Pajk S, Sosič I, et al. Selective DNA gyrase inhibitors: multi-target in silico profiling with 3D-pharmacophores. *Pharmaceuticals*. 2021;14(8):789.
32. Shui Y. Large-scale dataset for BiMA-DTI. 2025. <https://doi.org/10.6084/m9.figshare.2997471.v1>.
33. 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(16):6684–98.
34. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19)*. New York: Association for Computing Machinery; 2019. pp. 2623–31. <https://doi.org/10.1145/3292500.3330701>.
35. Huang K, Xiao C, Glass LM, Sun J. Moltrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics*. 2021;37(6):830–6.
36. Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*. 2019;35(2):309–18.
37. Zhao M, Yuan M, Yang Y, Xu SX. CPGL: prediction of compound-protein interaction by integrating graph attention network with long short-term memory neural network. *IEEE ACM Trans Comput Biol Bioinform*. 2023;20(3):1935–42.
38. Li M, Lu Z, Wu Y, Li Y. BACPI: a bi-directional attention neural network for compound-protein interaction and binding affinity prediction. *Bioinformatics*. 2022;38(7):1995–2002.
39. Zhao Q, Duan G, Zhao H, Zheng K, Li Y, Wang J. GIFDTI: prediction of drug-target interactions based on global molecular and intermolecular interaction representation learning. *IEEE ACM Trans Comput Biol Bioinform*. 2023;20(3):1943–52.
40. Yin Z, Chen Y, Hao Y, Pandiyan S, Shao J, Wang L. FOTF-CPI: a compound-protein interaction prediction transformer based on the fusion of optimal transport fragments. *iScience*. 2024;27(1):108756.
41. Peng L, Mao J, Huang G, Han G, Liu X, Liao W, et al. 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(3):1318–37.
42. Wei Z, Wang Z, Tang C. Dynamic prediction of drug-target interactions via cross-modal feature mapping with learnable association information. *J Chem Inf Model*. 2025;65(8):3915–27.
43. Liu H, Sun J, Guan J, Zheng J, Zhou S. Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics*. 2015;31(12):i221–9.
44. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008;36(suppl\_1):D901–6.
45. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res*. 2008;36(suppl\_1):D919–22.
46. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. 2018. <https://snap.stanford.edu/biodata/index.html>.
47. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res*. 2016;44(D1):D1045–53.
48. Bai P, Miljković F, Ge Y, Greene N, John B, Lu H. Hierarchical clustering split for low-bias evaluation of drug-target interaction prediction. In: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2021. pp. 641–44. <https://ieeexplore.ieee.org/document/9669515>. Cited 31 Oct 2024.
49. Shui Y. YouyuanShui/BiMA-DTI: the code and medium-scale datasets for BiMA-DTI. 2025. <https://doi.org/10.5281/zenodo.16935249>.
50. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. Sydney: JMLR.org; 2017. pp. 1321–30. (ICML'17).
51. Jiang H, Kim B, Guan MY, Gupta M. To trust or not to trust a classifier. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc.; 2018. pp. 5546–57. (NIPS'18).
52. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. 2014. <http://arxiv.org/abs/1312.6034>. Cited 15 Jan 2025.
53. Hu H, Wang X, Chan GKY, Chang JH, Do S, Drummond J, et al. Discovery of 3,5-substituted 6-azaindazoles as potent pan-Pim inhibitors. *Bioorg Med Chem Lett*. 2015;25(22):5258–64.
54. Groebke Zbinden K, Banner DW, Hilpert K, Himber J, Lavé T, Riederer MA, et al. Dose-dependent antithrombotic activity of an orally active tissue factor/factor VIIa inhibitor without concomitant enhancement of bleeding propensity. *Bioorg Med Chem*. 2006;14(15):5357–69.
55. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res*. 2024;52(D1):D1265–75.
56. Struzyńska L, Chalimoniuk M, Sulkowski G. The role of astroglia in Pb-exposed adult rat brain with respect to glutamate toxicity. *Toxicology*. 2005;212(2):185–94.
57. Bang-Andersen B, Ruhland T, Jørgensen M, Smith G, Frederiksen K, Jensen KG, et al. Discovery of 1-[2-(2,4-Dimethylphenyl)sulfonyl]phenyl]piperazine (Lu AA21004): a novel multimodal compound for the treatment of major depressive disorder. *J Med Chem*. 2011;54(9):3206–21.
58. Yamada Y, Masuda K, Li Q, Ihara Y, Kubota A, Miura T, et al. The structures of the human calcium channel  $\alpha 1$  subunit (CACNL1A2) and  $\beta$  subunit (CACNLB3) genes. *Genomics*. 1995;27(2):312–9.
59. Heidmann DEA, Metcalf MA, Kohlen R, Hamblin MW. Four 5-hydroxytryptamine7 (5-HT7) receptor isoforms in human and rat produced by alternative splicing: species differences due to altered intron-exon organization. *J Neurochem*. 1997;68(4):1372–81.
60. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9.
61. Morris GM, Huey R, Olson AJ. Using autodock for ligand-receptor docking. *Curr Protoc Bioinformatics*. 2008;24(1):8.14.1–8.14.40.
62. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871–6.
63. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. Quantum-chemical insights from deep tensor neural networks. *Nat Commun*. 2017;8(1):13890.

64. Gasteiger J, Groß J, Günnemann S. Directional message passing for molecular graphs. 2022. <http://arxiv.org/abs/2003.03123>. Cited 2 Jun 2025.
65. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*. 2006;25(2):247–60.
66. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. 2017. <http://arxiv.org/abs/1705.03122>. Cited 15 Jan 2025.
67. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. 2018. <http://arxiv.org/abs/1803.02155>. Cited 15 Jan 2025.
68. Su J, Lu Y, Pan S, Murtadha A, Wen B, Liu Y. RoFormer: enhanced transformer with rotary position embedding. 2023. <http://arxiv.org/abs/2104.09864>. Cited 15 Jan 2025.
69. Elfving S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw*. 2018;107:3–11.
70. Wang Z, Kong F, Feng S, Wang M, Yang X, Zhao H, et al. Is Mamba effective for time series forecasting?. 2024. <http://arxiv.org/abs/2403.11144>. Cited 2 Jan 2025.
71. Shi Y, Huang Z, Feng S, Zhong H, Wang W, Sun Y. Masked label prediction: unified message passing model for semi-supervised classification. 2021. <http://arxiv.org/abs/2009.03509>.
72. Rampásek L, Galkin M, Dwivedi VP, Luu AT, Wolf G, Beaini D. Recipe for a general, powerful, scalable graph transformer. *Adv Neural Inf Process Syst*. 2022;35:14501–515.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.