MGNDTI: A Drug-Target Interaction Prediction Framework Based on Multimodal Representation Learning and the Gating Mechanism

Lihong Peng, Xin Liu, Min Chen,* Wen Liao, Jiale Mao, and Liqian Zhou*
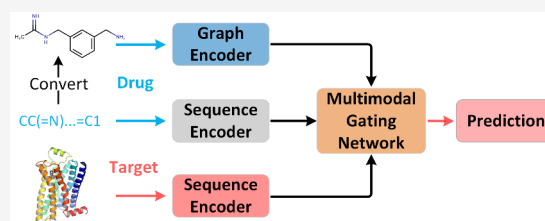
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Drug-Target Interaction (DTI) prediction facilitates acceleration of drug discovery and promotes drug repositioning. Most existing deep learning-based DTI prediction methods can better extract discriminative features for drugs and proteins, but they rarely consider multimodal features of drugs. Moreover, learning the interaction representations between drugs and targets needs further exploration. Here, we proposed a simple **M**ulti-modal **G**ating **N**etwork for **DTI** prediction, MGNDTI, based on multimodal representation learning and the gating mechanism. MGNDTI first learns the sequence representations of drugs and targets

using different retentive networks. Next, it extracts molecular graph features of drugs through a graph convolutional network. Subsequently, it devises a multimodal gating network to obtain the joint representations of drugs and targets. Finally, it builds a fully connected network for computing the interaction probability. MGNDTI was benchmarked against seven state-of-the-art DTI prediction models (CPI-GNN, TransformerCPI, MolTrans, BACPI, CPGL, GIFDTI, and FOTF-CPI) using four data sets (i.e., Human, *C. elegans*, BioSNAP, and BindingDB) under four different experimental settings. Through evaluation with AUROC, AUPRC, accuracy, F1 score, and MCC, MGNDTI significantly outperformed the above seven methods. MGNDTI is a powerful tool for DTI prediction, showcasing its superior robustness and generalization ability on diverse data sets and different experimental settings. It is freely available at https://github.com/plhhnu/MGNDTI.

1. INTRODUCTION

Accurately identifying Drug-Target Interactions (DTIs) benefits the discovery and repositioning of drugs. The wet-lab experiment-based DTI prediction methods are expensive, laborious, and time-consuming.[1−4] Consequently, computational methods have increasingly been devised to decipher new DTIs automatically. These methods mainly contain two categories: simulation-based and machine learning-based methods.[5]

Simulation-based methods typically contain molecular docking[6,7] and molecular dynamics simulations.[8] Simulation-based methods fully adopt 3D structures of proteins for inferring interaction sites and binding postures. Although they obtained remarkable progress, they rely heavily on known protein 3D structure information and require very high computational resources.

Machine learning-based methods work well in various association prediction tasks, for example, lncRNA-disease associations,[9,10] ligand—receptor interactions,[11−14] and drug-induced hepatotoxicity.[15] Specially, machine learning significantly promoted DTI inference. These methods contain nearest neighbor,[16] network embedding,[17,18] semisupervised learning,[19] and triple collaborative matrix factorization,[20] rotation forest,[21] and feature processing scheme.[22]

Recently, with the continuous accumulation of interacting drug-target pairs (DTPs), deep learning has obtained great success in DTI prediction, and many end-to-end deep learning

models have been devised to uncover DTI candidates.[23,24] These methods first separately encode drugs and targets based on their molecular graphs or 1D sequences from two different aspects, and the encoded features are concatenated as inputs of classifier.[25,26] For example, GraphDTA,[27] DeepConv-DTI,[28] MMDTA,[29] and GraphCL-DTA[30] adopted a 1D convolutional neural network (1D-CNN) to learn features of proteins from their amino acid sequences. GNN-PT,[31] iNGNN-DTI,[25] augmented graph neural network,[32] and DruBAN[33] used graph neural networks (GNNs) to obtain molecular representations. In addition, other models have been still employed to predict new DTIs and drug repositioning, for example, bilinear attention network,[34] graph convolutional network,[35] bidirectional encoder representation-based contrastive learning,[36] global-local perspective,[37] and the combination of BiLSTM and 2D-CNN.[38]

Deep learning has greatly improved the DTI prediction. However, most deep learning-based DTI identification models separately represent and process drugs and targets and

demonstrate extreme difficulty in deciphering their interacting context. The attention mechanism[39] can efficiently solve this problem.[40] HyperAttentionDTI[41] built an attention vector for each amino acid or atom. AttentionSiteDTI[42] employed the self-attention mechanism to uncover binding sites between a ligand and protein. CoaDTI[43] utilized a coattention mechanism for incorporating the drug and protein modalities. DrugBAN[33] analyzed pairwise local DTIs through a deep bilinear attention network with domain adaptation. MCANet[44] adopted the cross-attention mechanism to enhance the model feature representation ability.

Deep learning models have demonstrated promising progress in DTI prediction, but few of most existing deep learning-based DTI prediction models consider the aggregation of various features of drugs.[45] Moreover, the mutual influences between drugs and targets need further exploration.[40] The attention mechanism emphasizes important features by assigning different sizes of attention weights to the features.[46] But the gating mechanism[47,48] facilitates to retain useful features while filtering ineffective features for the learning task. Here, to solve the limitations, we propose a new DTI prediction model, called MGNDTI, by incorporating multimodal representation learning and the gating mechanism. The main contributions of this work are as follows:

- Two Retentive Networks (RetNets) are designed to extract drugs' features from their SMILES strings and proteins' features from their amino acid sequences.

- A simple multimodal representation learning model is devised to incorporate drug SMILES sequence features and molecular graph features.

- A multimodal gating network is developed to learn the representations of DTPs. To the best of our knowledge, it is the first model that incorporates multimodal representation learning and the gating mechanism for DTP representation learning.

2. MATERIALS AND METHODS

2.1. Data Sets. Four DTI data sets are used to train and assess the model's performance, namely, Human, *C.elegans*, BioSNAP, and BindingDB. The former two data sets were created by ref [49] and contain highly credible negative DTIs. The BioSNAP data set was built by refs [50] and [51] from the DrugBank database[52] and provide an equal number of validated positive DTIs and negative randomly selected DTIs from unseen DTPs. The BindingDB data set[53] focuses mainly on experimentally validated binding affinities between small molecules and proteins. In this study, BindingDB uses its low-bias version built by ref [54]. The description about the above four data sets is shown in Table 1. On the Human and BioSNAP data sets, Figure 1 shows the statistical histograms for protein length and compound atom number. The statistical histograms for the *C.elegans* and BindingDB data sets have been provided in Figure S1 in Supporting Information.

Table 1. Description of DTI datasets

Data set	Drug	Protein	Interaction	Positive	Negative
Human	2,726	2,001	5,997	2,633	3,364
C.elegans	1,767	1,876	7,785	3,893	3,892
BioSNAP	4,505	2,181	27,464	13,830	13,634
BindingDB	14,643	2,623	49,199	20,674	28,525

2.2. MGNDTI. To predict new DTIs, as illustrated in Figure 2, MGNDTI first separately extracts drugs' features from their SMILES sequences and proteins' features from their amino acid sequences to obtain their sequence representations using different RetNets. That is, the two RetNets use different parameters. Moreover, drugs' molecular graph features are extracted to obtain drugs' structural representation through GCN. Subsequently, a multimodal gating network is developed to obtain the feature representation of each DTP. Finally, a four-layer Fully Connected Network (FCN) is constructed for classifying each DTP.

2.2.1. Problem Formulation. Suppose that a drug is denoted as $\mathcal{D} = (\mathcal{S}, \mathcal{G})$, where $\mathcal{S} = (s_1, s_2, s_3, ..., s_n)$ denotes its SMILES sequence with n SMILES symbols, and $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denotes its molecular graph with the atom set $\mathcal{V}$ and the set of undirected edges between atoms $\mathcal{E}$. Each protein is denoted as $\mathcal{P} = (a_1, a_2, a_3, ..., a_m)$ with m amino acids. 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 DTP.

2.2.2. Encoding Drug and Protein Sequence Features with Retentive Network. RetNet[55] is a strong architecture of large language models. It obtains training parallelism, low-cost prediction, and better performance. Here, we adopt RetNet for acquiring feature representations of drugs and proteins based on their sequence information.

The SMILES string of a drug is composed of 64 different characters, and a protein consists of 23 different amino acids. MGNDTI begins with transforming each SMILES symbol s_i and amino acid a_j to an embedding vector based on two embedding layers, respectively. As a result, the embedding matrix with d embedding size $\mathbf{F}_S \in \mathbb{R}^{n \times d}$ for a drug and $\mathbf{F}_{\mathcal{P}} \in \mathbb{R}^{m \times d}$ for a protein is obtained.

Next, drug and protein features are encoded through RetNet. An L-layer RetNet block is built by stacking multiscale retention (MSR) and feed-forward network (FFN). It takes the feature matrix $\mathcal{M}_S^{(0)} = \mathbf{F}_S$ as input and outputs drug latent feature matrix $\mathcal{M}_S = \mathcal{M}_S^{(L)}$ by Eq. 1:

$$\begin{cases} \mathbf{Y}_S^{(l)} = \text{MSR}(\text{LN}(\mathcal{M}_S^{(l)})) + \mathcal{M}_S^{(l)} \\ \mathcal{M}_S^{(l+1)} = \text{FNN}(\text{LN}(\mathbf{Y}_S^{(l)})) + \mathbf{Y}_S^{(l)} \end{cases} \quad (1)$$

where $\text{LN}(\cdot)$ and $\text{MSR}(\cdot)$ denote LayerNorm[56] and multiscale retention,[55] respectively. $\mathcal{M}_S^{(l)} \in \mathbb{R}^{n \times d}$, $\mathbf{Y}_S^{(l)} \in \mathbb{R}^{n \times d}$, and $\text{FNN}(\mathbf{X}) = \text{gelu}(\mathbf{X}\mathbf{W}_1)\mathbf{W}_2$ $(\mathbf{X} \in \mathbb{R}^{n \times d})$ with parameter matrices $\mathbf{W}_1 \in \mathbb{R}^{d \times d/2}$ and $\mathbf{W}_2 \in \mathbb{R}^{d/2 \times d}$.

Similarly, it takes the feature matrix $\mathcal{M}_{\mathcal{P}}^{(0)} = \mathbf{F}_{\mathcal{P}}$ as input and outputs protein latent feature matrix $\mathcal{M}_{\mathcal{P}} = \mathcal{M}_{\mathcal{P}}^{(L)}$ by Eq. 2:

$$\begin{cases} \mathbf{Y}_{\mathcal{P}}^{(l)} = \text{MSR}(\text{LN}(\mathcal{M}_{\mathcal{P}}^{(l)})) + \mathcal{M}_{\mathcal{P}}^{(l)} \\ \mathcal{M}_{\mathcal{P}}^{(l+1)} = \text{FNN}(\text{LN}(\mathbf{Y}_{\mathcal{P}}^{(l)})) + \mathbf{Y}_{\mathcal{P}}^{(l)} \end{cases} \quad (2)$$

where $\mathcal{M}_{\mathcal{P}}^{(l)} \in \mathbb{R}^{m \times d}$ and $\mathbf{Y}_{\mathcal{P}}^{(l)} \in \mathbb{R}^{m \times d}$.

2.2.3. Encoding Molecular Graph Features of Drugs with GCN. To encode drug molecular graph features, inspired by the previous works,[3,33] we transform its SMILES string to a 2D molecular graph $\mathcal{G}$. In $\mathcal{G}$, each atom node is initialized through the DGL-LifeSci[57] package based on its chemical properties.
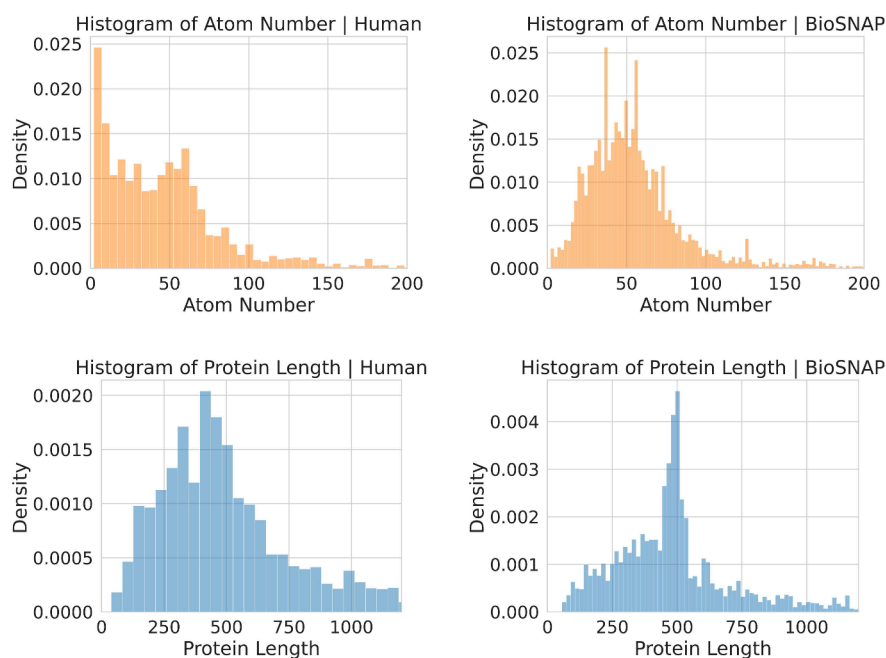
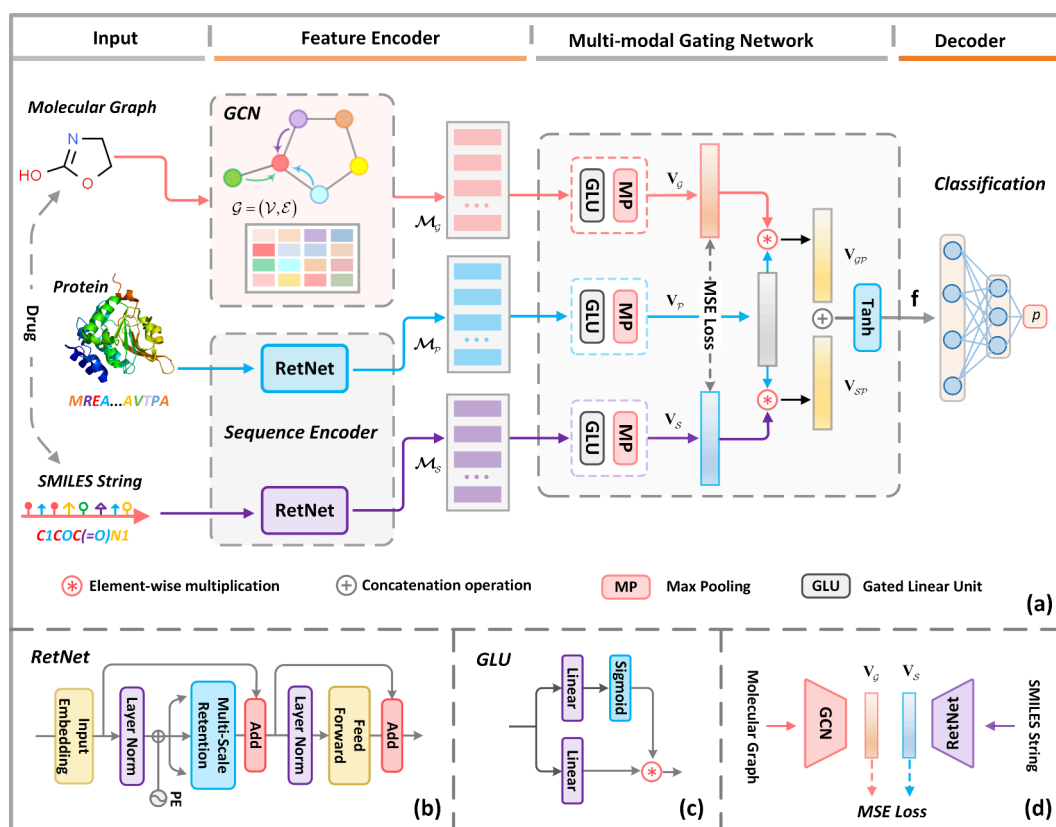**Figure 1.** Histogram of proteins and compounds on Human and BioSNAP.

Figure 2. Illustration of MGNDTI. (a) The architecture of MGNDTI mainly includes drug and protein feature encoding based on RetNets and GCN, DTP feature representation learning based on multimodal gating network, and DTP classification based on an FCN. (b) Architecture of retentive network. (c) Architecture of gated linear unit. (d) Architecture of multimodal representation learning.

Consequently, the molecular graph of each drug is represented as $\mathbf{M}_d \in \mathbb{R}^{n \times 74}$. Moreover, a linear transformation $\mathbf{M}_\mathcal{G} = \mathbf{M}_d \mathbf{W}_d^T$ ($\mathbf{W}_d \in \mathbb{R}^{d \times 74}$) is used to compute a real-valued dense matrix $\mathbf{M}_\mathcal{G} \in \mathbb{R}^{n \times d}$ as the input of GCN.

Next, a K-layer GCN is devised to initialize $\mathbf{H}^{(0)} = \mathbf{M}_\mathcal{G}$ and encode the molecular graph feature of a drug by Eq. 3:

$$\mathbf{H}^{(k+1)} = f(\mathbf{H}^{(k)}, \mathbf{A}) = \mathrm{ReLU}(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(k)} \mathbf{W}_\mathcal{G}^{(k)})$$

(3)

here, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{I} \in \mathbb{R}^{n \times n}$ denote an adjacency matrix and identity matrix, respectively. $\tilde{\mathbf{D}} \in \mathbb{R}^{n \times n}$ denotes the degree matrix corresponding to $\tilde{\mathbf{A}}$. $\mathbf{W}_{\mathcal{G}}^{(k)} \in \mathbb{R}^{d \times d}$ and $\mathbf{H}^{(k)} \in \mathbb{R}^{n \times d}$ denote the weight and input of the k-th layer, respectively.

Finally, the molecular graph of a drug can be depicted as $\mathcal{M}_{\mathcal{G}} = \mathbf{H}^{(K)}$.

2.2.4. Learning DTP Representation with Multimodal Gating Network. To better characterize each drug, as shown in Figure 2 (d), we learn drug multimodal representation by combining its SMILES sequence and molecular graph. Furthermore, we develop a multimodal gating network for learning DTP features by incorporating multimodal representation learning and the gating mechanism.

As shown in Figure 2 (c), we adopt a simple gated linear unit (GLU)[48] to filter nonimportant features and use the max pooling to down-sample features by Eq. 4:

$$\begin{cases} \mathbf{V}_{\mathcal{G}} = \mathrm{MP}((\mathcal{M}_{\mathcal{G}}\mathbf{W}_{\mathcal{G}} + \mathbf{b}_{\mathcal{G}}) \odot \sigma(\mathcal{M}_{\mathcal{G}}\mathbf{V}_{\mathcal{G}} + \mathbf{c}_{\mathcal{G}})) \\ \mathbf{V}_{S} = \mathrm{MP}((\mathcal{M}_{S}\mathbf{W}_{S} + \mathbf{b}_{S}) \odot \sigma(\mathcal{M}_{S}\mathbf{V}_{S} + \mathbf{c}_{S})) \\ \mathbf{V}_{\mathcal{P}} = \mathrm{MP}((\mathcal{M}_{\mathcal{P}}\mathbf{W}_{\mathcal{P}} + \mathbf{b}_{\mathcal{P}}) \odot \sigma(\mathcal{M}_{\mathcal{P}}\mathbf{V}_{\mathcal{P}} + \mathbf{c}_{\mathcal{P}})) \end{cases} \quad (4)$$

where $\mathbf{V}_{S} \in \mathbb{R}^d$ and $\mathbf{V}_{\mathcal{G}} \in \mathbb{R}^d$ indicate drug feature vectors obtained by its SMILES sequence and molecular graph, respectively. $\mathbf{V}_{\mathcal{P}} \in \mathbb{R}^d$ indicates protein feature vector based on its sequence. $\mathbf{W}_{.} \in \mathbb{R}^{d \times d}$, $\mathbf{b}_{.} \in \mathbb{R}^d$, $\mathbf{V}_{.} \in \mathbb{R}^{d \times d}$, and $\mathbf{c}_{.} \in \mathbb{R}^d$ are learnable parameters, and $\sigma(\cdot)$, $\odot$, and $\mathrm{MP}(\cdot)$ denote the sigmoid function, element-wise product, and the max pooling, respectively.

We acquire sequence features of a drug by combining RetNet and GLU with the maximum pooling and its molecular graph features by incorporating GCN and GLU with the max pooling. During training, we learn drug multimodal representation through Mean Square Error (MSE) loss. We can better preserve unique features of a drug from its SMILES sequence and molecular graph when obtaining its common representation through multimodal representation learning by Eq. 5:

$$\mathcal{L}_{mse} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{V}_{S}^{i} - \mathbf{V}_{\mathcal{G}}^{i})^2 \quad (5)$$

where $\mathbf{V}_{S}^{i}$ and $\mathbf{V}_{\mathcal{G}}^{i}$ denote the i-th feature of $\mathbf{V}_{S}$ and $\mathbf{V}_{\mathcal{G}}$, respectively.

Finally, a DTP can be denoted as $\mathbf{V}_{S\mathcal{P}} \in \mathbb{R}^d$ based on drug sequence features and protein sequence features and as $\mathbf{V}_{\mathcal{G}\mathcal{P}} \in \mathbb{R}^d$ of drug molecular graph features and protein sequence features by a simple element-by-element product operation. Consequently, the final representation $\mathbf{f} \in \mathbb{R}^{2d}$ of a DTP can be obtained by concatenating $\mathbf{V}_{S\mathcal{P}}$ and $\mathbf{V}_{\mathcal{G}\mathcal{P}}$ by Eq. 6:

$$\begin{cases} \mathbf{V}_{S\mathcal{P}} = \mathbf{V}_{S} \odot \mathbf{V}_{\mathcal{P}} \\ \mathbf{V}_{\mathcal{G}\mathcal{P}} = \mathbf{V}_{\mathcal{G}} \odot \mathbf{V}_{\mathcal{P}} \\ \mathbf{f} = \tanh(\mathrm{Concat}(\mathbf{V}_{S\mathcal{P}}, \mathbf{V}_{\mathcal{G}\mathcal{P}})) \end{cases} \quad (6)$$

2.2.5. Inference and Training. For computing the interaction probability p of each DTP, its feature vector $\mathbf{f}$ is fed to an FCN by Eq. 7:

$$p = \sigma(\delta(\delta(\delta(\mathbf{f}\mathbf{W}_{F1} + \mathbf{b}_1)\mathbf{W}_{F2} + \mathbf{b}_2)\mathbf{W}_{F3} + \mathbf{b}_3)\mathbf{W}_{F4} + b_4) \quad (7)$$

where $\mathbf{W}_{F1} \in \mathbb{R}^{2d \times 4d}$, $\mathbf{W}_{F2} \in \mathbb{R}^{4d \times 4d}$, $\mathbf{W}_{F3} \in \mathbb{R}^{4d \times d}$, $\mathbf{W}_{F4} \in \mathbb{R}^{d \times 1}$, $\mathbf{b}_1 \in \mathbb{R}^{4d}$, $\mathbf{b}_2 \in \mathbb{R}^{4d}$, $\mathbf{b}_3 \in \mathbb{R}^{d}$, and $b_4 \in \mathbb{R}$ denote learnable parameters. $\sigma(\cdot)$ and $\delta(\cdot)$ denote the sigmoid function and ReLU function, respectively. Moreover, we use Binary Cross Entropy (BCE) loss to train the model by Eq. 8:

$$\mathcal{L}_{bce} = \frac{1}{M} \sum_{i} -(y_i \log(p_i) \\ + (1 - y_i)\log(1 - p_i)) + \lambda \|\mathbf{\Theta}\|_2^2 \quad (8)$$

where y_i and p_i indicate the true and predicted labels, respectively. $\mathbf{\Theta}$, M, and λ indicate a set of all learnable parameters, the number of DTPs used to train the model, and the L_2 regularization coefficient, respectively.

In summary, the loss function of the MGNDTI model is represented as Eq. 9:

$$\mathcal{L}_{total} = \mathcal{L}_{bce} + \frac{1}{M} \sum_{j=1}^{M} \mathcal{L}_{mse}^{j} \quad (9)$$

Finally, we minimize $\mathcal{L}_{total}$ to train MGNDTI.

3. RESULTS

3.1. Experimental Setup. To better demonstrate the MGNDTI performance, we employed the Area Under the Receiver Operation Characteristics (ROC) curve (AUROC), the Area Under the Precision-Recall (PR) Curve (AUPRC), accuracy, F1-score, and Matthews' Correlation Coefficient (MCC) as evaluation metrics.

Given the drug set $\mathbf{D}_{train}$ and the protein set $\mathbf{P}_{train}$ in the training set, to infer whether there is an interaction between a drug d and a protein p in the testing set, we conducted comprehensive comparison experiments through the following four different settings:

- E1. A DTI data set is randomly partitioned into training set, validation set, and test set with a ratio 7:1:2.
- E2. During training, drug d was not used to train the model while protein p was used to train the model: $d \notin \mathbf{D}_{train}$ and $p \in \mathbf{P}_{train}$.
- E3. During training, drug d was used to train the model while protein p was not used to train the model: $d \in \mathbf{D}_{train}$ and $p \notin \mathbf{P}_{train}$.
- E4. During training, neither d nor p were used to train the model: $d \notin \mathbf{D}_{train}$ and $p \notin \mathbf{P}_{train}$.

For E2, E3, and E4, 20% DTPs were first randomly selected for test and the remaining were selected for training. Next, according to the three different settings, drugs or proteins were deleted, respectively. Finally, the remaining testing samples after deletion were partitioned into the validation set and test set at a ratio of 1:2.

During each setting, we conducted 10 independent runs which used different random seeds when partitioning a data set. And the optimal performing model was selected based on the highest AUROC from the validation set and was evaluated on the test set.

For MGNDTI, we set epoch_size = 50, batch_size = 64, learning_rate = 1e-4, learning_decay = 0.5, decay_interval = 20, weight_decay = 1e-5, atom_dim = 128, attention_heads = 8, GCN_layers = 3, RetNet_layers = 1, amino_acid_dim =

Table 2. Comparison results of MGNDTI and seven baselines on Human (10 random runs)[a]

Method	AUROC	AUPRC	Accuracy	F1-score	MCC
			E1		
CPI-GNN	0.9329 ± 0.0085	0.9174 ± 0.0164	0.8899 ± 0.0084	0.8858 ± 0.0098	0.7798 ± 0.0170
TransformerCPI	0.9795 ± 0.0036	0.9745 ± 0.0052	0.9316 ± 0.0071	0.9223 ± 0.0092	0.8613 ± 0.0147
MolTrans	0.9799 ± 0.0028	0.9785 ± 0.0044	0.9418 ± 0.0099	0.9207 ± 0.0291	0.8823 ± 0.0196
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
MGNDTI	**0.9855 ± 0.0031**	**0.9820 ± 0.0059**	**0.9481 ± 0.0056**	**0.9485 ± 0.0057**	**0.8951 ± 0.0111**
			E2		
CPI-GNN	0.7642 ± 0.0496	0.7765 ± 0.0444	0.7231 ± 0.0440	0.6967 ± 0.0540	0.4519 ± 0.0874
TransformerCPI	0.8689 ± 0.0140	0.8920 ± 0.0131	**0.8390 ± 0.0114**	0.8110 ± 0.0166	0.6766 ± 0.0207
MolTrans	0.8856 ± 0.0133	0.8783 ± 0.0152	0.8219 ± 0.0170	0.7769 ± 0.0263	0.6438 ± 0.0334
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
MGNDTI	**0.9162 ± 0.0188**	**0.9073 ± 0.0187**	0.8361 ± 0.0183	**0.8481 ± 0.0196**	**0.6797 ± 0.0383**
			E3		
CPI-GNN	0.9674 ± 0.0103	0.9656 ± 0.0165	0.9364 ± 0.0166	0.9305 ± 0.0179	0.8737 ± 0.0320
TransformerCPI	0.9569 ± 0.0069	0.9495 ± 0.0063	0.8934 ± 0.0140	0.8661 ± 0.0218	0.7836 ± 0.0249
MolTrans	0.9738 ± 0.0054	0.9699 ± 0.0080	**0.9448 ± 0.0115**	0.9199 ± 0.0211	**0.8874 ± 0.0233**
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
MGNDTI	**0.9795 ± 0.0079**	0.9785 ± 0.0061	0.9401 ± 0.0178	**0.9395 ± 0.0151**	0.8789 ± 0.0335
			E4		
CPI-GNN	0.6929 ± 0.0562	0.6771 ± 0.0450	0.6265 ± 0.0351	0.4985 ± 0.0551	0.2800 ± 0.0691
TransformerCPI	0.7387 ± 0.0201	0.7475 ± 0.0212	0.7095 ± 0.0209	0.5835 ± 0.0350	0.4261 ± 0.0370
MolTrans	0.6969 ± 0.0187	0.6377 ± 0.0198	0.5891 ± 0.0426	0.3258 ± 0.1520	0.2760 ± 0.0680
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
MGNDTI	**0.8106 ± 0.0228**	**0.7780 ± 0.0278**	**0.7180 ± 0.0446**	**0.7708 ± 0.0219**	**0.4862 ± 0.0618**

[a]Note: The best results are marked in bold, and the second-best results are marked in underline.

128, and FCN_hidden_size = 512. In addition, we selected a threshold under the best F1-score as the classification threshold.

3.2. Baseline Methods. To evaluate the MGNDTI performance, we compared it to seven baselines. Their source codes were downloaded from corresponding repositories, and the performances were re-evaluated in the same experimental environment. Their hyperparameters were explicitly set with corresponding defaults.

- CPI-GNN:[58] CPI-GNN first encoded features of drugs and proteins using GNN and 1D-CNNs, respectively, and investigated the importance of subsequences in a protein to a drug with the one-side attention mechanism, finally predicted DTIs by an FCN.
- TransformerCPI:[59] TransformerCPI adopted CNN and GCN to extract protein features from their amino acid sequences and drug features from their SMILES sequences, respectively. And then it utilized a Transformer decoder to convert the learned features into an FCN for predictions.
- MolTrans:[50] MolTrans encoded drug and protein features based on their sequences by Transformer,

constructed an interaction matrix based on the encoded features, and finally made predictions through CNNs and FCNs.
- BACPI:[60] BACPI incorporated drug and protein representations by bidirectional attention neural network and implemented predictions with a classifier.
- CPGL:[61] CPGL optimized drug feature representations through a graph attention network and protein feature representations through a long short-term memory neural network for enhancing the robustness and generalization ability of the model.
- GIFDTI:[62] GIFDTI learned drug features from their SMILES sequences and protein features from their amino acid sequences by CNN and Transformer, respectively. And then it extracted the global features and intermolecular interaction features for DTP classification.
- FOTF-CPI:[63] FOTF-CPI merged an optimal transport-based fragmentation model and a fused attention mechanism for predictions.

3.3. Comparison of MGNDTI with Seven Baselines. In this section, we compared MGNDTI with seven baselines on

Table 3. Comparison results of MGNDTI and seven baselines on *C.elegans* (10 random runs)[a]

Method	AUROC	AUPRC	Accuracy	F1-score	MCC
			E1		
CPI-GNN	0.9536 ± 0.0079	0.9422 ± 0.0121	0.9146 ± 0.0100	0.9185 ± 0.0088	0.8292 ± 0.0201
TransformerCPI	**0.9919 ± 0.0015**	0.9916 ± 0.0018	0.9608 ± 0.0045	0.9608 ± 0.0044	0.9217 ± 0.0089
MolTrans	0.9918 ± 0.0024	**0.9920 ± 0.0028**	0.9670 ± 0.0032	0.9626 ± 0.0047	0.9342 ± 0.0064
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
MGNDTI	0.9912 ± 0.0022	0.9916 ± 0.0020	**0.9682 ± 0.0051**	**0.9682 ± 0.0050**	**0.9365 ± 0.0102**
			E2		
CPI-GNN	0.6830 ± 0.0649	0.7278 ± 0.0391	0.6320 ± 0.0584	0.6038 ± 0.0763	0.2844 ± 0.1126
TransformerCPI	0.8076 ± 0.0151	0.8657 ± 0.0122	0.7623 ± 0.0244	0.7299 ± 0.0453	0.5551 ± 0.0285
MolTrans	0.8259 ± 0.0290	0.8595 ± 0.0259	0.7465 ± 0.0298	0.7202 ± 0.0314	0.5014 ± 0.0616
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
MGNDTI	**0.8985 ± 0.0161**	**0.9134 ± 0.0084**	**0.8188 ± 0.0232**	**0.8215 ± 0.0229**	**0.6410 ± 0.0481**
			E3		
CPI-GNN	0.8998 ± 0.0330	0.9007 ± 0.0419	0.8351 ± 0.0409	0.8199 ± 0.0531	0.6797 ± 0.0810
TransformerCPI	0.9470 ± 0.0141	0.9486 ± 0.0159	0.8523 ± 0.0439	0.8262 ± 0.0660	0.7204 ± 0.0728
MolTrans	0.9457 ± 0.0166	0.9542 ± 0.0116	0.9034 ± 0.0164	0.8330 ± 0.0508	0.8069 ± 0.0324
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
MGNDTI	**0.9636 ± 0.0096**	**0.9684 ± 0.0055**	**0.9193 ± 0.0111**	**0.9181 ± 0.0107**	**0.8388 ± 0.0222**
			E4		
CPI-GNN	0.5842 ± 0.0588	0.5717 ± 0.0515	0.5166 ± 0.0423	0.3096 ± 0.1012	0.0625 ± 0.1052
TransformerCPI	0.6262 ± 0.0608	0.6213 ± 0.0643	0.5537 ± 0.0358	0.2706 ± 0.1324	0.1414 ± 0.0973
MolTrans	0.5983 ± 0.0279	0.6174 ± 0.0347	0.5089 ± 0.0150	0.2489 ± 0.1162	0.0779 ± 0.0440
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
MGNDTI	**0.7301 ± 0.0427**	**0.7252 ± 0.0382**	**0.6555 ± 0.0514**	**0.7239 ± 0.0238**	**0.3601 ± 0.0826**

[a]Note: The best results are marked in bold, and the second best results are marked in underline.
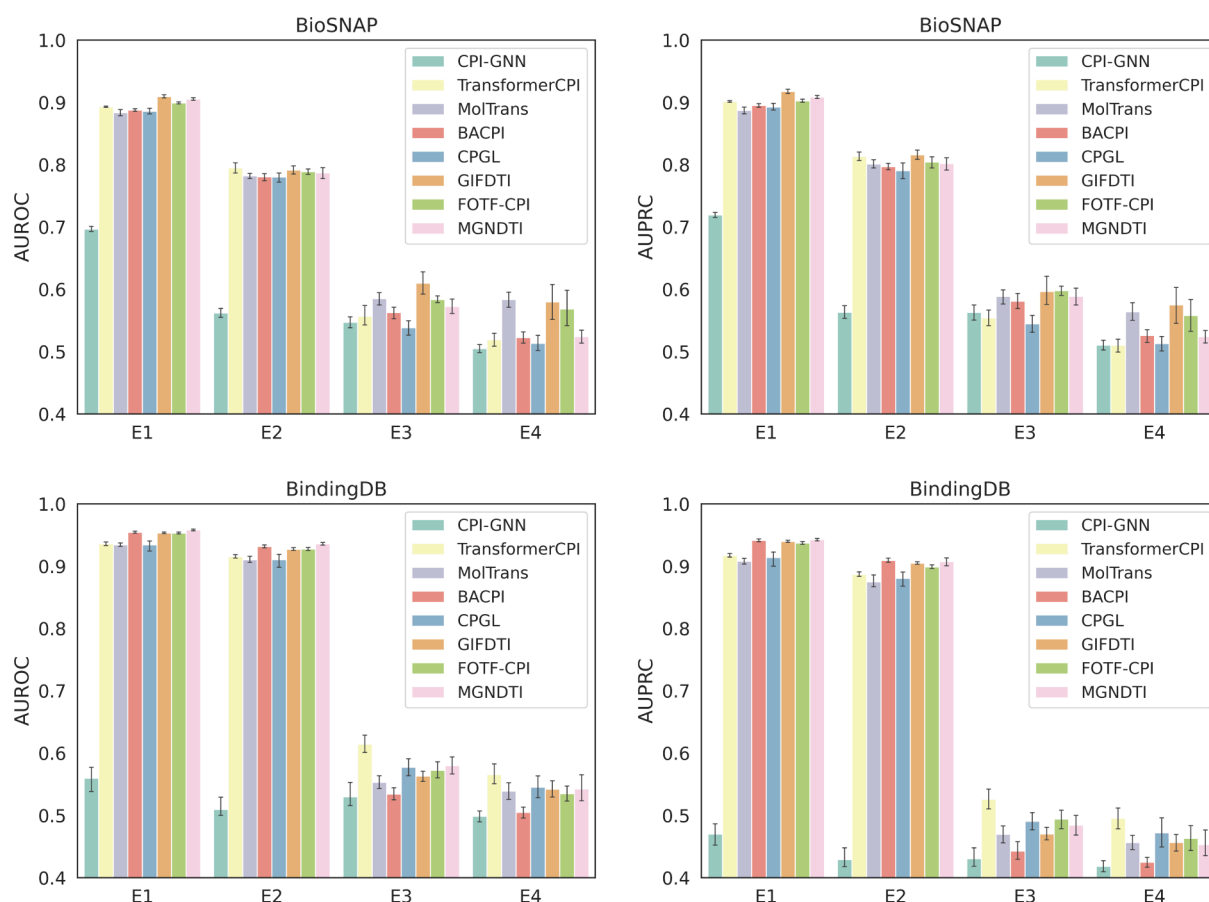
the Human, *C.elegans*, BioSNAP, and BindingDB benchmarking data sets under four different experimental settings. As shown in Tables 2 and 3, our proposed MGNDTI method performed better than seven baselines based on AUROC, AUPRC, accuracy, F1-score, and MCC on Human and *C.elegans* under four different experimental settings. Particularly, the MGNDTI performance was significantly improved on Human under E4. Its computed AUROC improved by 4.24% (from 0.7682 to 0.8106), AUPRC by 3.05% (from 0.7475 to 0.7780), accuracy by 0.85% (from 0.7095 to 0.7180), F1-score by 18.73% (from 0.5835 to 0.7708), and MCC by 6.01% (from 0.4261 to 0.4862).

On *C.elegans*, MGNDTI surpassed seven baselines on all evaluation metrics under E2, E3, and E4. Under E1, although MGNDTI slightly underperformed on AUROC and AUPRC, its decrease was very little, that is, 0.07% (from 0.9919 to 0.9912) and 0.04% (from 0.9920 to 0.9916). More importantly, it improved accuracy, F1-score, and MCC, with accuracy improvement by 0.12% (from 0.9670 to 0.9682), F1-score by 0.21% (from 0.9661 to 0.9682), and MCC by 0.23% (from 0.9342 to 0.9365).

Figure 3 depicted AUROC and AUPRC of MGNDTI and seven baselines on BioSNAP and BindingDB under four experimental settings. Under E1 and E2 settings, MGNDTI significantly outperformed seven baseline methods on the BindingDB data set, and its performance was comparable to the best baseline method on the BioSNAP data set. Under E3 and E4 settings, MGNDTI reached the baseline level on the two data sets. Detailed values for AUROC, AUPRC, accuracy, F1-score, and MCC for MGNDTI and baseline are listed in Tables S1 and S2 of the Supporting Information and at https://github.com/plhhnu/MGNDTI/tree/main/results

In summary, our proposed MGNDTI model elucidated outstanding performance compared to seven baselines on four data sets under four different experimental settings, demonstrating powerful robustness and generalization ability when making DTI predictions.

3.4. Ablation Experiments. To assess the effect of feature encoders (i.e., GCN and RetNet), multimodal representation of drugs, and the multimodal gating network on DTI prediction, we performed the following three ablation experiments:

Figure 3. AUROC and AUPRC of MGNDTI and seven baselines on BioSNAP and BindingDB (10 random runs).

(1) To validate the contribution of the feature extractors GCN and RetNet in the model, we investigated three variants of MGNDTI: TransDTI (i.e., replacing RetNet with Transformer), SRetDTI (i.e., two RetNets with shared weights), and GATDTI (i.e., replacing GCN with GAT).

As shown in Figure 4, the AUROC and AUPRC values of MGNDTI were higher than those of TransDTI and SRetDTI under all four experimental setups. The two RetNet feature extractors without shared parameters were more effective than the two RetNet with shared parameters when extracting drug and protein features. Moreover, RetNet surpassed TransDTI and Transformer when extracting drug and protein features. In addition, MGNDTI outperformed GATDTI on Human and BindingDB data sets. On *C.elegans* and BioSNAP, the performance of MGNDTI and GATDTI was essentially equal. Thus, we extracted drug sequence features based on RetNet and GLU with the max pooling and encoded drug molecular graph features based on GCN and GLU with the max pooling.

(2) To compare the role of drug unimodality and multimodality on DTI prediction, we investigated two MGNDTI variants based on drug unimodality and multimodality: SGNDTI (drug SMILES sequence + protein sequence) and GGNDTI (drug molecular graph + protein sequence).
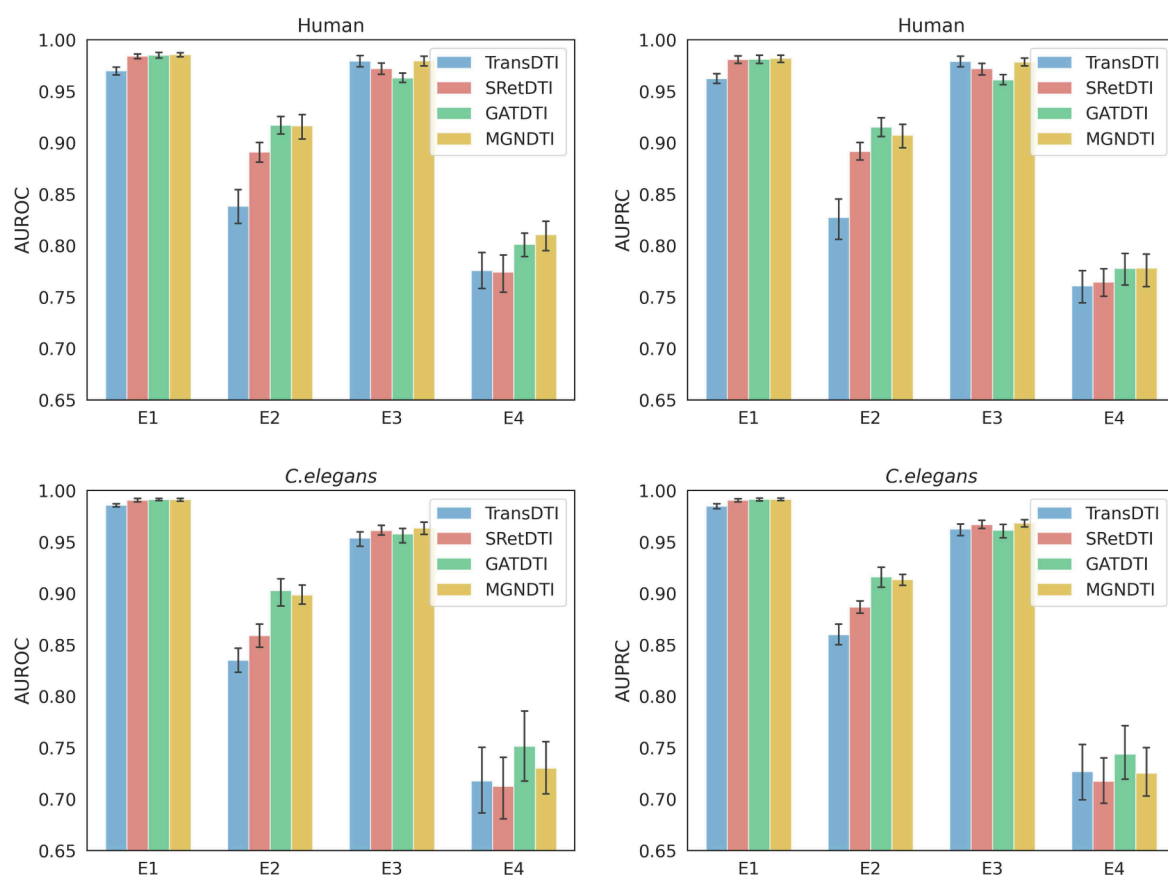
Figure 5 demonstrates the performance of MGNDTI under drug unimodality and multimodality. As shown in Figure 5, MGNDTI significantly outperformed the other

two unimodality variants under all four experimental settings. Hence, multimodality facilitates to boost the DTI prediction accuracy.

(3) To analyze the role of multimodal gating network and its GLU with MSE loss during drug multimodality learning, we considered three other variants of MGNDTI: CatDTI (i.e., MGNDTI without the multimodal gating network), MSEDTI (i.e, MGNDTI without GLU), and GLUDTI (i.e, MGNDTI without multimodal representation learning).

Similarly, as shown in Figure 6, MGNDTI with multimodal gating network (GLU + MSE loss) acquired the best prediction against the other three variants, demonstrating the strong learning ability of the multimodal gating network. The ablation experiments on BioSNAP and BindingDB are listed in Figures S2, S3, and S4 in Supporting Information and at https://github.com/plhhnu/MGNDTI/tree/main/ablation.

3.5. Visualization and Model Analysis. To investigate the distributions of multimodal features (i.e., drug features from their SMILES strings, drug features from their molecular graphs, and protein features from their sequences) in MGNDTI with MSE loss or not (i.e., MGNDTI with multimodal representation learning or not), we performed the feature visualization operations on four data sets: we first trained MGNDTI on the training set and then randomly selected 1,200 samples and extracted their multimodal features using the pretrained MGNDTI on the test set, finally reduced the dimensionality of these features through UMAP and carried out visualization analysis.

Figure 4. Ablation experiments on feature encoders (GCN and RetNet) on Human and *C.elegans* under four experimental settings (10 random runs).

Figure 7 shows the distributions of the three types of features in MGNDTI with MSE loss or not on Human and *C.elegans*. On Human, the distributions of the obtained features were relatively sparse for MGNDTI without MSE loss, and the distribution was extremely sparse for drug features from their SMILES sequences. However, these distributions became relatively dense and formed three clusters in MGNDTI with a MSE loss. On *C.elegans*, drug feature distribution from their SMILES sequences was always sparse regardless of MGNDTI with MSE loss or not, but the other two distributions were always dense. The distributions of feature visualization on BioSNAP and BindingDB are listed in Figure S5 in Supporting Information and at https://github.com/plhhnu/MGNDTI/tree/main/feature.

Furthermore, we analyzed the MSE Loss, BCE Loss, and Total Loss (Total Loss = MSE Loss + BCE Loss) when the number of epochs changed during training. As shown in Figure 8, the MSE Loss sharply decreased and then quickly tended toward a stable state while the BCE Loss decreases relatively slowly. Furthermore, the Total Loss tended to a steady state at Epoch = 20 on two small data sets (i.e., Human and *C.elegans*) and at Epoch = 40 on two larger data sets (BioSNAP and BindingDB). More importantly, the BCE Loss and Total Loss significantly decreased at Epoch 20 and Epoch 40 on BioSNAP and BindingDB, respectively. Additionally, the learning rate decay with decay_interval = 20 could speed up the model training.

Finally, we analyzed GPU memory consumption and model complexity of MGNDTI and baseline methods. GPU memory consumption is measured by the GPU memory used by the models during training. Model complexity is evaluated by the number of parameters and floating-point operations (FLOPs).

Table 4 shows the comparison between MGNDTI and the seven baseline methods in terms of FLOPs, number of parameters, and GPU memory usage: (1) FLOPs: FLOPs of MGNDTI were 6.542G, which was much less than MolTrans, BACPI, GIFDTI, and FOTF-CPI. (2) Number of Parameters: the number of parameters of MGNDTI was 0.655M, which was less than that of MolTrans, GIFDTI, and FOTF-CPI. (3) GPU memory usage: the GPU memory usage of MGNDTI was 4.038GB, which was much lower than that of Trans-formerCPI, MolTrans, BACPI, GIFDTI, and FOTF-CPI. In summary, MGNDTI was better than most of the baselines in terms of computational efficiency and resource usage, and the number of parameters was moderate.

3.6. Case Study. To further verify the MGNDTI effectiveness, we performed case studies from the following aspects: predicting potential interacting targets for two drugs, olanzapine (DrugBank ID: DB00334) and valproic acid (DrugBank ID: DB00313), and inferring potential drugs for two targets CYP2D6 (UniProt ID: P10635) and CYP2C8 (UniProt ID: P10632) based on the pretrained MGNDTI model. And the predicted interactions were further validated through the DrugBank database.

Olanzapine is an antipsychotic drug and is used to treat schizophrenia and bipolar 1 disorder.[64,65] Valproic acid is an anticonvulsant and is applied to control complex partial seizures along with both simple and complex absence seizures.[66,67]
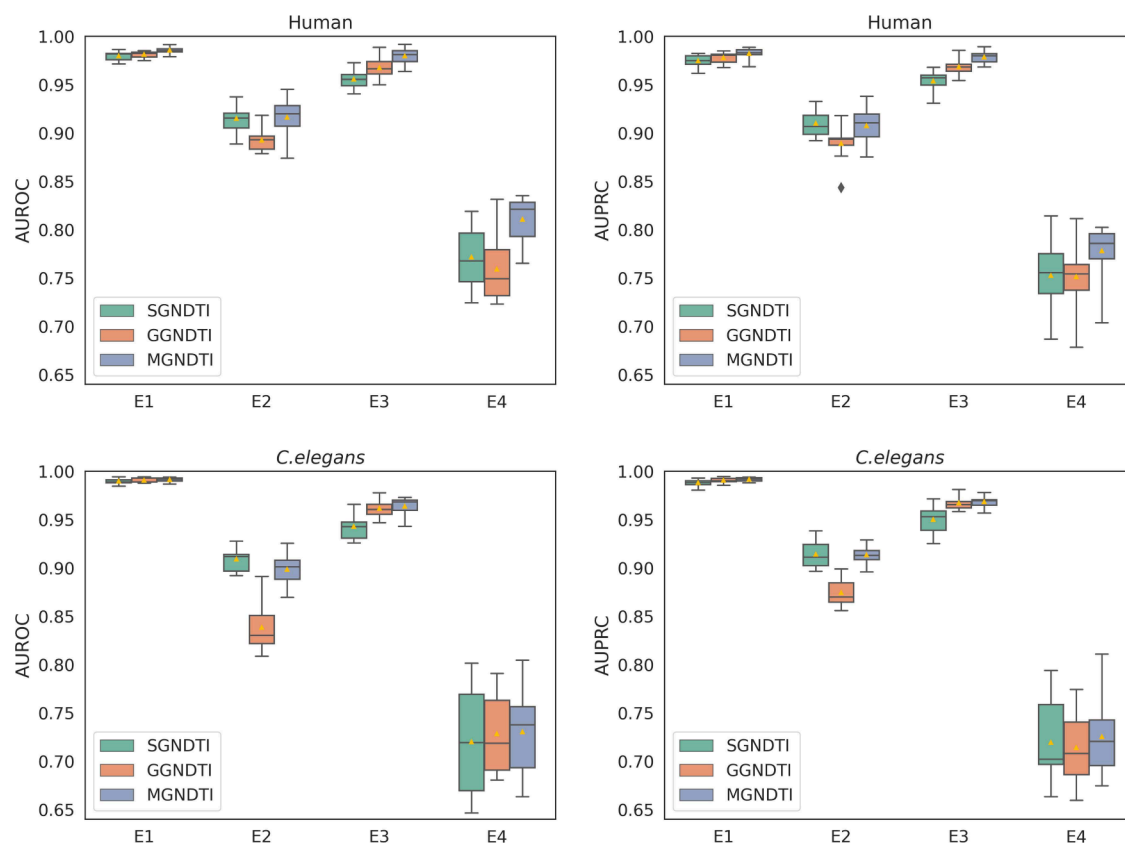
Figure 5. Ablation experiments on drug multimodality on Human and *C.elegans* under four experimental settings (10 random runs).
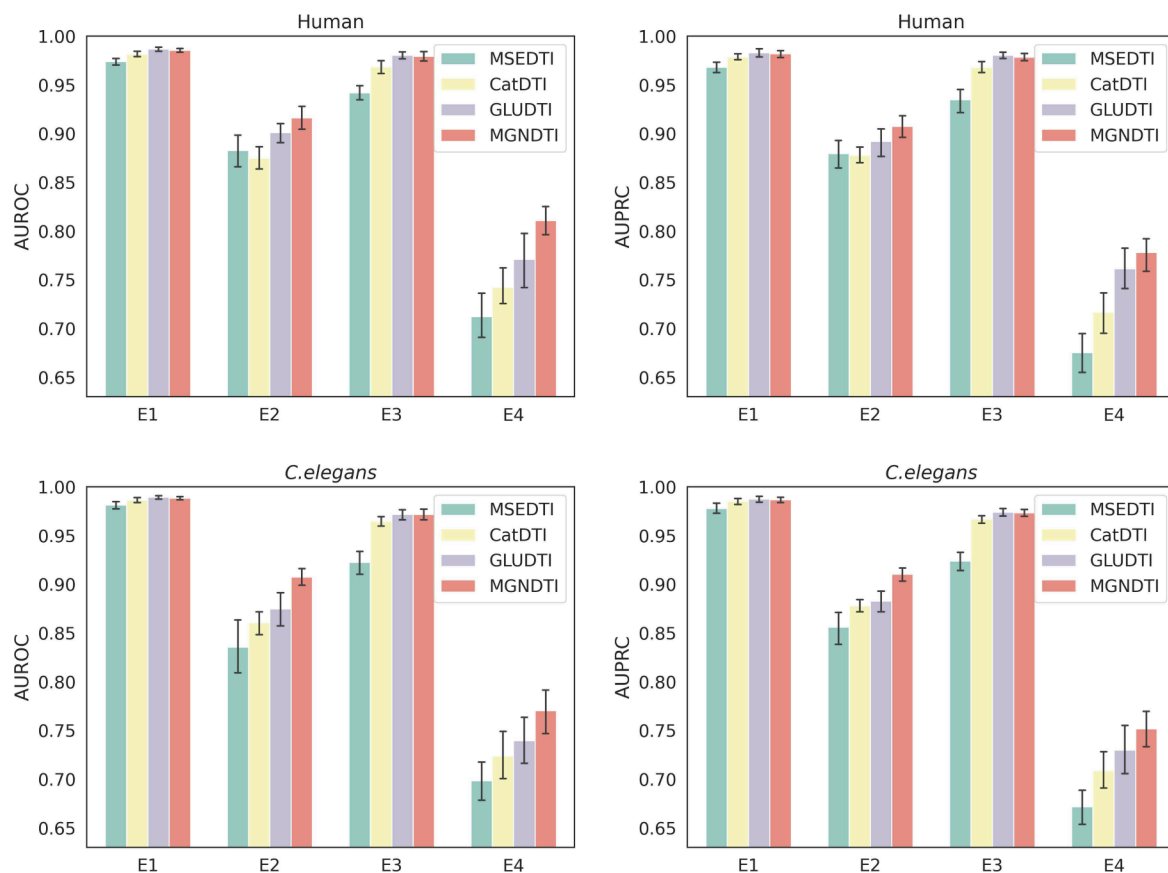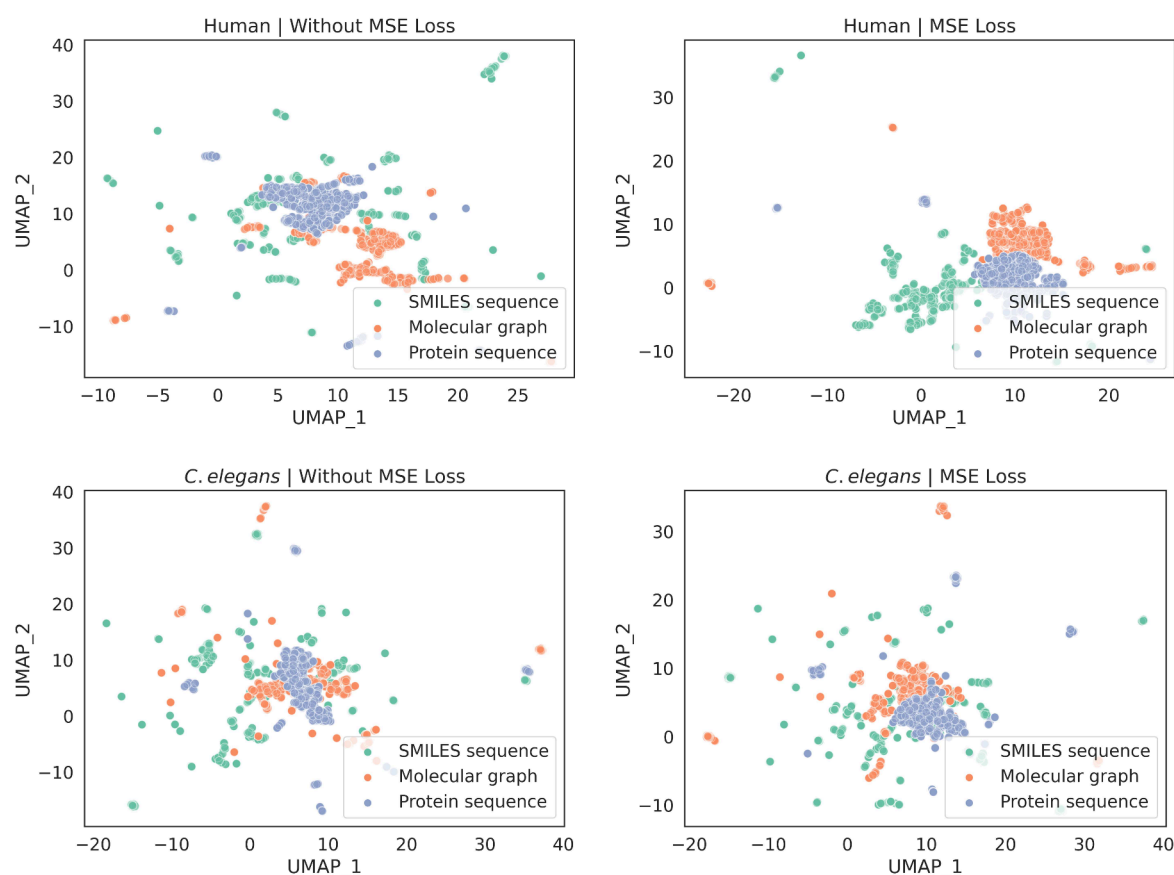

Figure 6. Ablation experiments on multimodal gating networks on Human and *C.elegans* under four experimental settings (10 random runs).

Figure 7. Distribution of multimodal feature visualizations. UMAP visualization of multimodal features on Human and *C.elegans* in MGNDTI with MSE loss or not.

As shown in Table 5, MGNDTI successfully predicted the top 10 target candidates for the two drugs. All predicted top 10 targets for olanzapine have been validated by the DrugBank database. Most of the predicted top 10 targets for valproic acid have been confirmed by the DrugBank database.

We predicted that two proteins (i.e., P15509 and P04183) could interact with DB00313. To further validate the above results, as shown in Figure 9, we implemented molecular docking for the two proteins and DB00313 using the molecular docking tool CB-Dock2.[68] Both of their binding affinities were −5.0 kcal/mol, further elucidating their higher interacting probabilities with valproic acid.

The proteins CYP2D6 and CYP2C8 are important drug-metabolizing enzymes. CYP2D6 participates in the metabolism of more than 25% of commonly used drugs in clinic practice.[69] CYP2C8 metabolizes more than 100 drugs and has high expression in human liver.[70]

As shown in Table 6, MGNDTI inferred the top 10 drug candidates for CYP2D6 and CYP2C8. All predicted drug candidates were verified to interact with the two targets by the DrugBank database. In summary, the results from case studies elucidated that MGNDTI was capable of precisely identifying potential interacting candidates for unknown drugs and targets from a large number of samples, demonstrating its significance for drug repositioning.
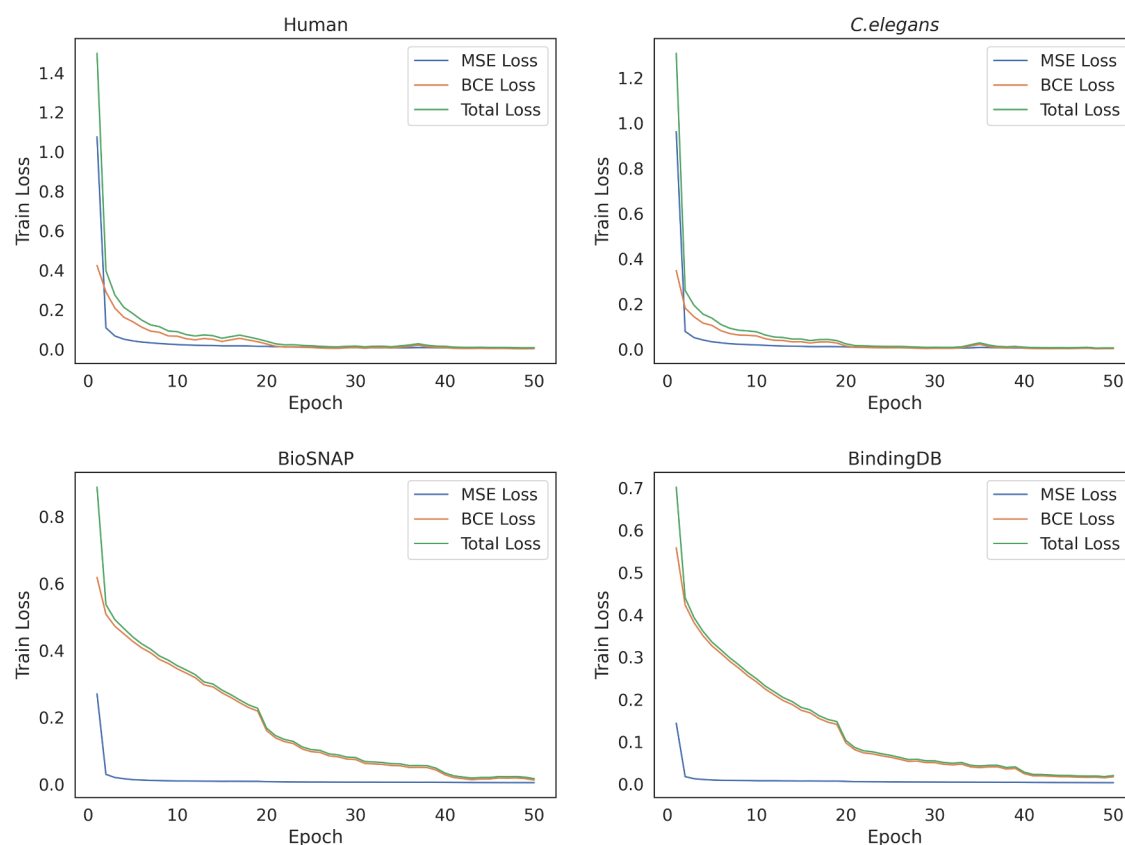
4. DISCUSSION

In this study, we proposed a new deep learning model for DTI prediction, MGNDTI, based on multimodal representation learning and the gating mechanism. MGNDTI first learned

drug and target sequence representations by using different RetNets. Next, it extracted drug molecular graph features through GCN. Subsequently, it devised a simple multimodal gating network for obtaining DTP feature embeddings. Finally, it built an FCN to classify DTPs.

We conducted an in-depth comparison experiment. First, MGNDTI was benchmarked against seven baseline DTI prediction models (CPI-GNN, TransformerCPI, MolTrans, BACPI, CPGL, GIFDTI, and FOTF-CPI) using four data sets (i.e., Human, *C.elegans*, BioSNAP, and BindingDB) under four different experimental settings. Through evaluation with AUROC, AUPRC, accuracy, F1 score, and MCC, MGNDTI outperformed the above seven baselines, demonstrating its strong robustness and generalization ability. Moreover, compared to four attention mechanism-based methods (i.e., CPI-GNN, TransformerCPI, BACPI, and FOTF-CPI), MGNDTI computed the best performance, verifying that the multimodal gating network facilitates learning the interaction representations between drugs and targets.

Additionally, we conducted three ablation experiments. The results demonstrated that RetNet and GCN could effectively extract drug or protein sequence and molecular graph features and drug multimodal features could further improve DTI prediction compared to their unimodal features. Moreover, MGNDTI with multimodal gating network could more efficiently learn DTP features compared to three other variants of MGNDTI: CatDTI, MSEDTI, and GLUDTI. Subsequently, drug multimodal features and protein sequence features were visualized. The visualization results clearly depict the distributions of the three types of features. Finally, the results

Figure 8. Variation of MSE loss, BCE loss, and total loss when the number of training Epoch changes on Human, *C.elegans*, BioSNAP, and BindingDB.

Table 4. Comparison of MGNDTI and seven baseline methods based on GPU memory usage, number of parameters, and FLOPs

Method	FlOPs	Params	GPU memory
CPI-GNN	1.649 G	0.003 M	0.459 GB
TransformerCPI	2.232 G	0.349 M	6.037 GB
MolTrans	137.373 G	47.163 M	14.024 GB
BACPI	15.305 G	0.195 M	10.877 GB
CPGL	0.728 G	0.012 M	2.564 GB
GIFDTI	619.523 G	20.759 M	24.370 GB
FOTF-CPI	135.493 G	36.128 M	11.326 GB
MGNDTI	6.542 G	0.655 M	4.038 GB

Table 5. Predicted top 10 targets interacting with olanzapine and valproic acid by MGNDTI

	Olanzapine (DB00334)		Valproic acid (DB00313)	
Rank	UniProt ID	Result	UniProt ID	Result
1	P11229	True	P15509	Unknown
2	P41595	True	P05177	True
3	P78334	True	Q92769	True
4	P05177	True	P08684	True
5	P10635	True	P23219	True
6	P28221	True	Q99250	True
7	P50406	True	P33260	True
8	Q9UN88	True	Q9HAW8	True
9	Q99928	True	P04183	Unknown
10	P14416	True	P22310	True

from case studies showed that all predicted top 10 targets for olanzapine, most of the predicted targets for valproic acid, and all top 10 predictions for CYP2D6 and CYP2C8 have been validated by the DrugBank database.

We still compared the FlOPs, number of model parameters, and GPU memory consumption of MGNDTI and seven baseline methods. The results elucidated that MGNDTI outperformed most of the baseline methods based on computational efficiency and resource utilization. Moreover, the MGNDTI had a moderate number of parameters. During training, the training loss changed with the number of epochs. Moreover, the learning rate decay sped up the model training. During the optimization process, MGNDTI was relatively easy to train, and its training loss decreased steadily, elucidating its robustness and reliability.

MGNDTI can precisely infer potential DTIs, and it may be attributed to the following two features. (i) Most existing drug feature extraction methods are based on the unimodal strategy and fail to fully learn drug abundant features. In contrast, we extracted drug multimodal features using RetNet and GCN and learned drug multimodal representations. (ii) Most DTI prediction methods used the attention mechanism to learn the interaction representations between drugs and targets, resulting in high computational burden. In this situation, we designed a simple multimodal network for reducing computational complexity and alleviating computational burden.

In summary, our proposed MGNDTI model accurately captured the DTI candidates. Considering that 3D structures of proteins contribute to their feature learning, in the future, we will fully utilize their 3D structures for DTI prediction through DeepMind's AlphaFold.[71] Moreover, noncoding
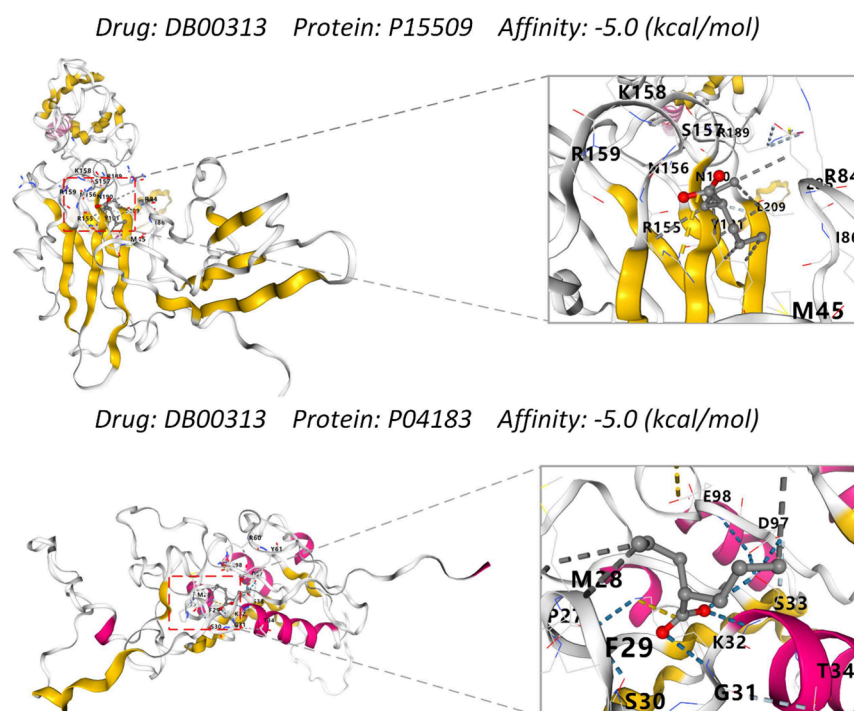
Drug: DB00313   Protein: P15509   Affinity: -5.0 (kcal/mol)

Drug: DB00313   Protein: P04183   Affinity: -5.0 (kcal/mol)

Figure 9. Molecular docking for two proteins (P15509 and P04183) and DB00313 using CB-Dock2.

Table 6. Predicted top 10 drug candidates interacting with CYP2D6 and CYP2C8 by MGNDTI

Rank	CYP2D6 (P10635)		CYP2C8 (P10632)	
	DrugBank ID	Result	DrugBank ID	Result
1	DB00521	True	DB01050	True
2	DB00627	True	DB00945	True
3	DB00316	True	DB01435	True
4	DB01182	True	DB01201	True
5	DB05109	True	DB00495	True
6	DB05271	True	DB00938	True
7	DB01151	True	DB08901	True
8	DB00281	True	DB08828	True
9	DB01698	True	DB00451	True
10	DB01191	True	DB00604	True

RNAs have been a class of important drug targets. Thus, we will further analyze the interactions between drugs and noncoding RNAs. Specifically, in the MGNDTI model, we will replace proteins with noncoding RNAs as inputs to the model and redesign RetNet to extract features of noncoding RNAs for predicting the probability of interaction between drugs and noncoding RNAs.

5. CONCLUSION

In this manuscript, we presented an end-to-end deep learning model called MGNDTI for DTI identification based on multimodal representation learning and the gating mechanism. MGNDTI obtained the best DTI prediction performance on four benchmark data sets under four different experimental settings. Moreover, it predicted that P15509 and P04183 could interact with valproic acid and require further validation. We anticipate that the proposed MGNDTI model can promote drug discovery and repurposing.

■ ASSOCIATED CONTENT

Data Availability Statement

All data are from publicly available resources. The BindingDB and BioSNAP data sets can be downloaded from https://github.com/peizhenbai/DrugBAN. The Human and *C.elegans* data sets can be downloaded from https://github.com/masashitsubaki/CPI_prediction. The Supporting Information and source codes are available at https://github.com/plhhnu/MGNDTI.

⑤ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00957.

(1) Statistical histograms for the *C.elegans* and BindingDB data sets. (2) AUROC, AUPRC, Accuracy, F1-score, and MCC computed by MGNDTI and seven baseline methods on BindingDB and BioSNAP under four experimental settings. (3) Results of ablation experiments of MGNDTI on BindingDB and BioSNAP. (4) UMAP visualization of feature distributions on BindingDB and BioSNAP (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Min Chen — *School of Computer Science and Engineering, Hunan Institute of Technology, Hengyang, Hunan 421002, China*; Email: chenmin@hnit.edu.cn

Liqian Zhou — *College of Life Science and Chemistry, Hunan University of Technology, Zhuzhou, Hunan 412007, China*; Email: zhoulq11@163.com

Authors

Lihong Peng — *College of Life Science and Chemistry, Hunan University of Technology, Zhuzhou, Hunan 412007, China*; ⓞ orcid.org/0000-0002-2321-3901

Xin Liu − *College of Life Science and Chemistry, Hunan University of Technology, Zhuzhou, Hunan 412007, China;* ⊚ orcid.org/0009-0008-8661-9288

Wen Liao − *School of Computer Science, Hunan University of Technology, Zhuzhou, Hunan 412007, China*

Jiale Mao − *School of Computer Science, Hunan University of Technology, Zhuzhou, Hunan 412007, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.4c00957

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Wang, L.; You, Z.-H.; Chen, X.; Yan, X.; Liu, G.; Zhang, W. Rfdt: A rotation forest-based predictor for predicting drug-target interactions using drug structure and protein sequence information. *Curr. Protein Pept. Sci.* **2018**, *19*, 445−454.

(2) Baruah, S.; Devi, D. A comparative analysis of Drug-Target Interaction detection approaches using Ensemble of Machine learning models. *Biochemistry and Bioinformatics* **2024**, 21−29.

(3) Peng, L.; Liu, X.; Yang, L.; Liu, L.; Bai, Z.; Chen, M.; Lu, X.; Nie, L. BINDTI: A bi-directional Intention network for drug-target interaction identification based on attention mechanisms. *IEEE Journal of Biomedical and Health Informatics* **2024**.

(4) Song, W.; Xu, L.; Han, C.; Tian, Z.; Zou, Q. Drug-target interaction predictions with multi-view similarity network fusion strategy and deep interactive attention mechanism. *Bioinformatics* **2024**, *40*, No. btae346, DOI: 10.1093/bioinformatics/btae346.

(5) Iliadis, D.; De Baets, B.; Pahikkala, T.; Waegeman, W. A comparison of embedding aggregation strategies in drug-target interaction prediction. *BMC Bioinf.* **2024**, *25*, 59.

(6) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(7) Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int. J. Mol. Sci.* **2019**, *20*, 4331.

(8) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129−1143.

(9) Peng, L.; Ren, M.; Huang, L.; Chen, M. GEnDDn: An lncRNA-Disease Association Identification Framework Based on Dual-Net Neural Architecture and Deep Neural Network. *Interdisciplinary Sciences: Computational Life Sciences* **2024**, *16*, 418.

(10) Peng, L.; Huang, L.; Su, Q.; Tian, G.; Chen, M.; Han, G. LDA-VGHB: identifying potential lncRNA−disease associations with singular value decomposition, variational graph auto-encoder and heterogeneous Newton boosting machine. *Briefings Bioinf.* **2023**, *25*, bbad466.

(11) Peng, L.; Tan, J.; Xiong, W.; Zhang, L.; Wang, Z.; Yuan, R.; Li, Z.; Chen, X. Deciphering ligand-receptor-mediated intercellular communication based on ensemble deep learning and the joint scoring strategy from single-cell transcriptomic data. *Comput. Biol. Med.* **2023**, *163*, 107137.

(12) Peng, L.; Yuan, R.; Han, C.; Han, G.; Tan, J.; Wang, Z.; Chen, M.; Chen, X. Cellenboost: a boosting-based ligand-receptor interaction identification model for cell-to-cell communication inference. *IEEE Trans. NanoBiosci.* **2023**, *22*, 705−715.

(13) Peng, L.; Xiong, W.; Han, C.; Li, Z.; Chen, X. CellDialog: A Computational Framework for Ligand-Receptor-Mediated Cell-Cell Communication Analysis. *IEEE Journal of Biomedical and Health Informatics* **2024**, *28*, 580−591.

(14) Peng, L.; Gao, P.; Xiong, W.; Li, Z.; Chen, X. Identifying potential ligand-receptor interactions based on gradient boosted neural network and interpretable boosting machine for intercellular communication analysis. *Comput. Biol. Med.* **2024**, *171*, 108110.

(15) Su, R.; Wu, H.; Liu, X.; Wei, L. Predicting drug-induced hepatotoxicity based on biological feature maps and diverse classification strategies. *Briefings Bioinf.* **2021**, *22*, 428−437.

(16) He, Z.; Zhang, J.; Shi, X.-H.; Hu, L.-L.; Kong, X.; Cai, Y.-D.; Chou, K.-C. Predicting drug-target interaction networks based on functional groups and biological features. *PLoS One* **2010**, *5*, e9603.

(17) Van Laarhoven, T.; Nabuurs, S. B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **2011**, *27*, 3036−3043.

(18) Ji, B.-Y.; You, Z.-H.; Jiang, H.-J.; Guo, Z.-H.; Zheng, K. Prediction of drug-target interactions from multi-molecular network based on LINE network representation method. *J. Transl. Med.* **2020**, *18*, 1−11.

(19) Peng, L.; Liao, B.; Zhu, W.; Li, Z.; Li, K. Predicting drug-target interactions with multi-information fusion. *IEEE Journal of Biomedical and Health Informatics* **2017**, *21*, 561−572.

(20) Ding, Y.; Tang, J.; Guo, F.; Zou, Q. Identification of drug-target interactions via multiple kernel-based triple collaborative matrix factorization. *Briefings Bioinf.* **2022**, *23*, bbab582.

(21) Li, Y.; Liu, X.-z.; You, Z.-H.; Li, L.-P.; Guo, J.-X.; Wang, Z. A computational approach for predicting drug-target interactions from protein sequence and drug substructure fingerprint information. *International Journal of Intelligent Systems* **2021**, *36*, 593−609.

(22) Ru, X.; Zou, Q.; Lin, C. Optimization of drug-target affinity prediction methods through feature processing schemes. *Bioinformatics* **2023**, *39*, btad615.

(23) Xie, Z.; Tu, S.; Xu, L. Multilevel Attention Network with Semi-supervised Domain Adaptation for Drug-Target Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* **2024**, *38*, 329−337.

(24) Xu, J.; Xu, J.; Meng, Y.; Lu, C.; Cai, L.; Zeng, X.; Nussinov, R.; Cheng, F. Graph embedding and Gaussian mixture variational autoencoder network for end-to-end analysis of single-cell RNA sequencing data. *Cell Reports Methods* **2023**, *3*, 100382.

(25) 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.

(26) 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. *Proceedings of the AAAI Conference on Artificial Intelligence* **2024**, *38*, 310−319.

(27) Nguyen, T.; Le, H.; Quinn, T. P.; Nguyen, T.; Le, T. D.; Venkatesh, S. GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37*, 1140−1147.

(28) 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.

(29) Zhong, K.-Y.; Wen, M.-L.; Meng, F.-F.; Li, X.; Jiang, B.; Zeng, X.; Li, Y. MMDTA: A Multimodal Deep Model for Drug-Target Affinity with a Hybrid Fusion Strategy. *J. Chem. Inf. Model.* **2024**, *64*, 2878−2888.

(30) Yang, X.; Yang, G.; Chu, J. GraphCL-DTA: A Graph Contrastive Learning With Molecular Semantics for Drug-Target Binding Affinity Prediction. *IEEE Journal of Biomedical and Health Informatics* **2024**, 1−10.

(31) Wang, J.; Li, X.; Zhang, H. GNN-PT: enhanced prediction of compound-protein interactions by integrating protein transformer. *arXiv* **2020**. DOI: 10.48550/arXiv.2009.00805

(32) Meng, Y.; Wang, Y.; Xu, J.; Lu, C.; Tang, X.; Peng, T.; Zhang, B.; Tian, G.; Yang, J. Drug repositioning based on weighted local information augmented graph neural network. *Briefings Bioinf.* **2023**, *25*, bbad431.

(33) Bai, P.; Miljković, F.; John, B.; Lu, H. Interpretable bilinear attention network with domain adaptation improves drug-target prediction. *Nature Machine Intelligence* **2023**, *5*, 126−136.

(34) Tang, X.; Zhou, C.; Lu, C.; Meng, Y.; Xu, J.; Hu, X.; Tian, G.; Yang, J. Enhancing drug repositioning through local interactive learning with bilinear attention networks. *IEEE Journal of Biomedical and Health Informatics*; **2023**.

(35) Zeng, P.; Zhang, B.; Liu, A.; Meng, Y.; Tang, X.; Yang, J.; Xu, J. Drug repositioning based on tripartite cross-network embedding and graph convolutional network. *Expert Systems with Applications* **2024**, *252*, 124152.

(36) Wang, R.; Jin, J.; Zou, Q.; Nakai, K.; Wei, L. Predicting protein-peptide binding residues via interpretable deep learning. *Bioinformatics* **2022**, *38*, 3351−3360.

(37) Zhou, Z.; Liao, Q.; Wei, J.; Zhuo, L.; Wu, X.; Fu, X.; Zou, Q. Revisiting drug-protein interaction prediction: a novel global-local perspective. *Bioinformatics* **2024**, *40*, btae271.

(38) Zheng, S.; Li, Y.; Chen, S.; Xu, J.; Yang, Y. Predicting drug-protein interaction using quasi-visual question answering system. *Nature Machine Intelligence* **2020**, *2*, 134−140.

(39) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, 2017; Vol. 30.

(40) Li, F.; Zhang, Z.; Guan, J.; Zhou, S. Effective drug-target interaction prediction with mutual interaction neural network. *Bioinformatics* **2022**, *38*, 3582−3589.

(41) 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.

(42) Yazdani-Jahromi, M.; Yousefi, N.; Tayebi, A.; Kolanthai, E.; Neal, C. J.; Seal, S.; Garibay, O. O. AttentionSiteDTI: an interpretable graph-based model for drug-target interaction prediction using NLP sentence-level relation classification. *Briefings Bioinf.* **2022**, *23*, bbac272.

(43) Huang, L.; Lin, J.; Liu, R.; Zheng, Z.; Meng, L.; Chen, X.; Li, X.; Wong, K.-C. CoaDTI: multi-modal co-attention based framework for drug-target interaction annotation. *Briefings Bioinf.* **2022**, *23*, bbac446.

(44) Bian, J.; Zhang, X.; Zhang, X.; Xu, D.; Wang, G. MCANet: shared-weight-based MultiheadCrossAttention network for drug-target interaction prediction. *Briefings Bioinf.* **2023**, *24*, bbad082.

(45) 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 Bioinf.* **2023**, *24*, 323.

(46) Hong, Z.; Zeng, X.; Wei, L.; Liu, X. Identifying enhancer-promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism. *Bioinformatics* **2020**, *36*, 1037−1043.

(47) De, S.; Smith, S. L.; Fernando, A.; Botev, A.; Cristian-Muraru, G.; Gu, A.; Haroun, R.; Berrada, L.; Chen, Y.; Srinivasan, S.; Desjardins, G.; Doucet, A.; Budden, D.; Teh, Y. W.; Pascanu, R.; Freitas, N. D.; Gulcehre, C. Griffin: Mixing Gated Linear Recurrences with Local Attention for Efficient Language Models. *arXiv* 2024.

(48) van den Oord, A.; Kalchbrenner, N.; Espeholt, L.; kavukcuoglu, k.; Vinyals, O.; Graves, A. Conditional Image Generation with PixelCNN Decoders. *Advances in Neural Information Processing Systems*; Curran Associates, 2016; Vol. 29.

(49) 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*, i221−i229.

(50) Huang, K.; Xiao, C.; Glass, L. M.; Sun, J. MolTrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics* **2021**, *37*, 830−836.

(51) Zitnik, M.; Sosic, R.; Leskovec, J. BioSNAP Datasets: Stanford biomedical network dataset collection. https://snap.stanford.edu/biodata/index.html 2018.

(52) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901−D906.

(53) Gilson, M. K.; 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*, D1045−D1053.

(54) Bai, P.; Miljković, F.; Ge, Y.; Greene, N.; John, B.; Lu, H. Hierarchical clustering split for low-bias evaluation of drug-target interaction prediction. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* **2021**, 641−644.

(55) Sun, Y.; Dong, L.; Huang, S.; Ma, S.; Xia, Y.; Xue, J.; Wang, J.; Wei, F. Retentive network: A successor to transformer for large language models. *arXiv* 2023. DOI: 10.48550/arXiv.2307.08621

(56) Ba, J. L.; Kiros, J. R.; Hinton, G. E. Layer normalization. *arXiv*. 2016. DOI: 10.48550/arXiv.1607.06450

(57) Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. Dgl-lifesci: An open-source toolkit for deep learning on graphs in life science. *ACS omega* **2021**, *6*, 27233−27238.

(58) 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*, 309−318.

(59) 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.

(60) 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*, 1995−2002.

(61) Zhao, M.; Yuan, M.; Yang, Y.; Xu, S. X. CPGL: Prediction of Compound-Protein Interaction by Integrating Graph Attention Network With Long Short-Term Memory Neural Network. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2023**, *20*, 1935−1942.

(62) 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. Bioinf.* **2023**, *20*, 1943−1952.

(63) 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*, 108756.

(64) Citrome, L.; McEvoy, J. P.; Todtenkopf, M. S.; McDonnell, D.; Weiden, P. J. A commentary on the efficacy of olanzapine for the treatment of schizophrenia: the past, present, and future. *Neuropsychiatr. Dis. Treat.* **2019**, *15*, 2559−2569.

(65) Komossa, K.; Rummel-Kluge, C.; Hunger, H.; Schmid, F.; Schwarz, S.; Duggan, L.; Kissling, W.; Leucht, S. Olanzapine versus other atypical antipsychotics for schizophrenia. *Cochrane Database of Systematic Reviews* 2010. DOI: 10.1002/14651858.CD006654.pub2

(66) Prasad, M.; Krishnan, P. R.; Sequeira, R.; Al-Roomi, K. Anticonvulsant therapy for status epilepticus. *Cochrane Database of Systematic Reviews* 2014. DOI: 10.1002/14651858.CD003723.pub3

(67) Rosenberg, G. The mechanisms of action of valproate in neuropsychiatric disorders: can we see the forest for the trees? *Cell. Mol. Life Sci.* **2007**, *64*, 2090−2103.

(68) Liu, Y.; Yang, X.; Gan, J.; Chen, S.; Xiao, Z.-X.; Cao, Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res.* **2022**, *50*, W159−W164.

(69) Owen, R. P.; Sangkuhl, K.; Klein, T. E.; Altman, R. B. Cytochrome P450 2D6. *Pharmacogenet. Genomics* **2009**, *19*, 559−562.

(70) Backman, J. T.; Filppula, A. M.; Niemi, M.; Neuvonen, P. J. Role of cytochrome P450 2C8 in drug metabolism and interactions. *Pharmacol. Rev.* **2016**, *68*, 168−241.

(71) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493.