



# DeepMGT-DTI: Transformer network incorporating multilayer graph information for Drug–Target interaction prediction



Peiliang Zhang <sup>a</sup>, Ziqi Wei <sup>b</sup>, Chao Che <sup>a,\*</sup>, Bo Jin <sup>c</sup>

<sup>a</sup> Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education, Dalian, 116622, China

<sup>b</sup> School of Software, Tsinghua University, Beijing, 100084, China

<sup>c</sup> School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, 116024, China

## ARTICLE INFO

### Keywords:

DTI

Multilayer graph information

Transformer networks

COVID-19

Drug repositioning

## ABSTRACT

Drug–target interaction (DTI) prediction reduces the cost and time of drug development, and plays a vital role in drug discovery. However, most of research does not fully explore the molecular structures of drug compounds in DTI prediction. To this end, we propose a deep learning model to capture the molecular structure information of drug compounds for DTI prediction. This model utilizes a transformer network incorporating multilayer graph information, which captures the features of a drug's molecular structure so that the interactions between atoms of drug compounds can be explored more deeply. At the same time, a convolutional neural network is employed to capture the local residue information in the target sequence, and effectively extract the feature information of the target. The experiments on the DrugBank dataset showed that the proposed model outperformed previous models based on the structure of target sequences. The results indicate that the improved transformer network fuses the feature information between layers in the graph convolutional neural network and extracts the interaction data for the molecular structure. The drug repositioning experiment on COVID-19 and Alzheimer's disease demonstrated the proposed model's ability to find therapeutic drugs in drug discovery. The code of our model is available at <https://github.com/zhangpl109/DeepMGT-DTI>.

## 1. Introduction

The development of new drugs is an expensive and time-consuming process. A new drug's total development cost ranges from US \$200 million to US \$3 billion, and the development time is usually 13–15 years [34]. Therefore, drug repositioning methods have become a hot research topic in the field of drug development. An increasing number of drug repositioning studies show that drug–target interactions (DTIs) are crucial, but it usually takes 2–3 years to validate the accuracy of a DTI method through costly large-scale biochemical experiments [21]. However, computational DTI prediction methods can shorten DTI validation time and reduce research costs significantly. The rapid development of internet technology has led to the rapid accumulation of data about drug compounds, targets, and interactions for computational DTI research methods [15,17]. The internet further promotes the development of drug repositioning.

Computational DTI research methods include machine learning, deep learning, and graph neural networks. DTI research using machine learning can be traced to early work on pharmacological DTI prediction

[2]. The similarity/distance-based approach used to be a popular approach for DTI studies; similarity and distance functions were based on the pharmacological and genomic similarities of drugs and existing topological networks of drug–target relationships [4,12,35]. Zhang et al. [47] proposed the MultiviewDTI approach and improved DTI prediction by integrating drug data, target chemistry data, and known DTI data. Deep-learning methods are increasingly applied to DTI prediction [13, 16]. These models are trained by integrating multiple types of data, and the extracted feature information is correlated with molecular fingerprints for DTI prediction [41,42].

The rapid development of graph neural networks provides a powerful technique for DTI prediction. Wan et al. [38] proposed NeoDTI based on heterogeneous network data to perform DTI prediction. The results show that NeoDTI can achieve better prediction results. Lee et al. [23] represented SMILES sequences as drug molecule strings as Morgan fingerprints and used fully connected neural networks for drug feature extraction. This method extracted feature information more effectively than the Deep DTA [30] method. Although deep learning and graph neural network approaches have contributed significantly to the

\* Corresponding author.

E-mail addresses: [zhangpl109@163.com](mailto:zhangpl109@163.com) (P. Zhang), [weizq@tsinghua.edu.cn](mailto:weizq@tsinghua.edu.cn) (Z. Wei), [chechao@gmail.com](mailto:chechao@gmail.com) (C. Che), [jinbo@dlut.edu.cn](mailto:jinbo@dlut.edu.cn) (B. Jin).

development of DTI research, non-trivial problems remain. The graph neural network approach focuses on capturing node information and lacks feature learning of edges in the network. This leads to a lack of information capturing between nodes.

1. It is difficult to create a string representation that captures the relationship information in a drug molecule. It results in the loss of information about the interactions between drug atoms.
2. Moreover, the sequence representation in the form of a string is fragile, and a change in a string is likely to cause the molecular structure to change, which is not conducive to storage and research.

In response to these problems in DTI research, we propose a transformer network incorporating multilayer graph information (DeepMGT-DTI) to capture the molecular structure of the drug compounds involved in DTI prediction. In DeepMGT-DTI, the SMILES string of a drug is represented as a drug molecule graph and captures the feature information of the molecular structure of the drug compound using a graph convolutional neural network. The information in different hidden layers of the graph convolutional neural network is also fed into the transformer network for an improved multi-headed attention mechanism to mine the mutual information between different atoms in the molecular structure of the drug molecule at a deeper level. For the sequence structure of the target, we use a convolutional neural network to capture the local residue information in the target sequence and effectively extract the feature information of the target. This paper presents the principal contributions of this study, as summarized below.

1. The transformer network captures information from different layers in the graph convolutional neural network by an improved multi-headed attention mechanism. The original information about the interactions between atoms in the chemical structure of a drug is preserved. The problem of a lack of learning of edge features by the graph convolutional neural network is overcome.
2. The improved graph convolutional neural network is used to learn and capture the features of the molecular structure map of the drug and extract the atomic information in the molecular structure of the drug compound to the maximum extent.
3. The SMILES sequence string is represented as a molecular structure map, which overcomes the disadvantage of the fragility of the SMILES sequence. Also, the interconnections between atoms can be better represented.

The rest of this study is organized as follows. In section 2, related work is summarized. The proposed methodology is discussed in section 3. The experimental settings are discussed in section 4, and results are presented. Finally, conclusions are drawn in section 5.

## 2. Related work

In recent years, drug-target relationship prediction has become more closely linked to drug repositioning and has become a popular research component in drug development and biomedicine. The three research methods for DTI prediction are based on deep learning, graph neural networks, and transformer networks.

### 2.1. DTI prediction method based on deep learning

Deep learning methods can mine the information in a drug and target sequence structure to capture the characteristic information of the drug and the target. Yang et al. [45] proposed a mutual learning mechanism based on the multi-headed attention mechanism and location awareness. The model performance was significantly higher than that of the baseline when performing orphan-target and orphan-drug prediction, indicating that their proposed method improves the generalization and interpretation of DTI modeling. Bahi et al. [1] proposed two new

methods, SCA-DTIs and SCA-DTA, to predict convolutional neural networks (CNNs) and stacked-autoencoder (SAE) DTIs and drug-target binding affinity, respectively. The test results on different datasets demonstrated the potential of their proposed method for DTI and DTA prediction. Huang et al. [18] proposed DeepPurpose, a deep-learning library for DTI prediction. They trained DTI prediction models by implementing chemical structures of multiple compounds and more than 50 neural networks, which performed well on several datasets. Zhang et al. [46] used the PubMedBERT method to extract semantic elements focusing on COVID-19 and construct a knowledge graph from PubMed and other research literature. The TransE method was used for knowledge representation. Their method was shown to be feasible for drug candidate screening and generation and could be extended to potential drug discovery for other diseases. Lee et al. [23] used the DeepConv-DTI method to obtain information about local residues in protein sequences and used the obtained information for drug-target relationship prediction. The AUC of their model reached 0.852. The experimental results supported the aim of the model extracting the residue information in protein sequences. The above methods showed good prediction results of the models. However, it is challenging to learn the complex relationships between different entities for complex histological data because there is a lack of practical guidance for drug reuse.

### 2.2. DTI prediction method based on graph neural network

The graph neural network can mine for potential associations in the histological data and obtain more complex topological structure information between entities. To improve the practical use of DTI prediction, Nguyen et al. [28] represented the drug as a graph structure and used GraphDTA model for drug-target affinity prediction. The experimental results showed that graph neural networks further improve the prediction of drug-target affinity. Chu et al. [9] used DTI-CDF to learn multiple similarity features between drugs and similarity features between target proteins extracted from heterogeneous graphs for drug-target relationship prediction. The DTI-CDF method outperformed integrated learning methods, deep neural networks, and DDR [29] methods. In 2021, Cheng et al. [8] used graph embedding to obtain node information and substructure information in heterogeneous graphs. An autoencoder was used to accomplish drug-target linkage prediction, and good experimental results were obtained. The development of graph neural networks has also facilitated the study of molecular structure graphs in medicine. Li et al. [24] proposed a SIGN method based on graph neural networks that increase the prediction performance between nodes by improving the graph attention layer to integrate distance and angle information between nodes. Wang et al. [40] proposed a multi-view comparative representation learning approach (MIRACLE) to predict drug interactions. The method views the DDI network as a multi-view. Each node in the DDI network is itself a drug molecule graph. The MIRACLE method learns structural information of drug molecules by using GCN and bond-aware message-passing networks. The DeepCDR method [25] performs cancer drug response prediction by automatically learning potential representations between atoms and bonds in the chemical structure of a drug. The experimental results showed the potential value of DeepCDR in both predictive power and as a method to guide disease-specific drug design.

### 2.3. DTI prediction method based on transformer network

After Google proposed the transformer network [37], it was rapidly applied to natural language processing [11], computer vision [31], and medical artificial intelligence [6]. It was also widely used in drug molecular structure representation. Huang et al. [19] proposed the molecular interaction transformer (MolTrans) to address the shortcomings of current DTI methods that ignore the nature of DTI substructures and the value of unlabeled data. This method extracts the relationships between substructures from unlabeled biomedical data. Maziarka et al.

[27] proposed the molecule attention transformer (MAT). This method uses the molecular structure of a drug as the object of study and enhances the multi-headed attention mechanism in the transformer network using interatomic distances and molecular graph structures. Experiments show that MAT is competitive on some molecular prediction tasks.

The above study shows that a method based on graph neural networks can effectively extract topological information from the interaction network and that the drug molecule structure graph contains rich feature information, which has garnered considerable interest in the research community. The transformer network extracts some of the

feature information in the drug molecule structure, promoting DTI research and development.

### 3. Methods

DeepMGT-DTI performs DTI prediction by integrating the structural information of drug molecules and the sequence features of targets. DeepMGT-DTI has three main parts: data representation, feature extraction, and model prediction, as shown in Fig. 1. We first process and integrate the numbers from DrugBank and embed the data representation. Then, drug molecule structure information, target sequence

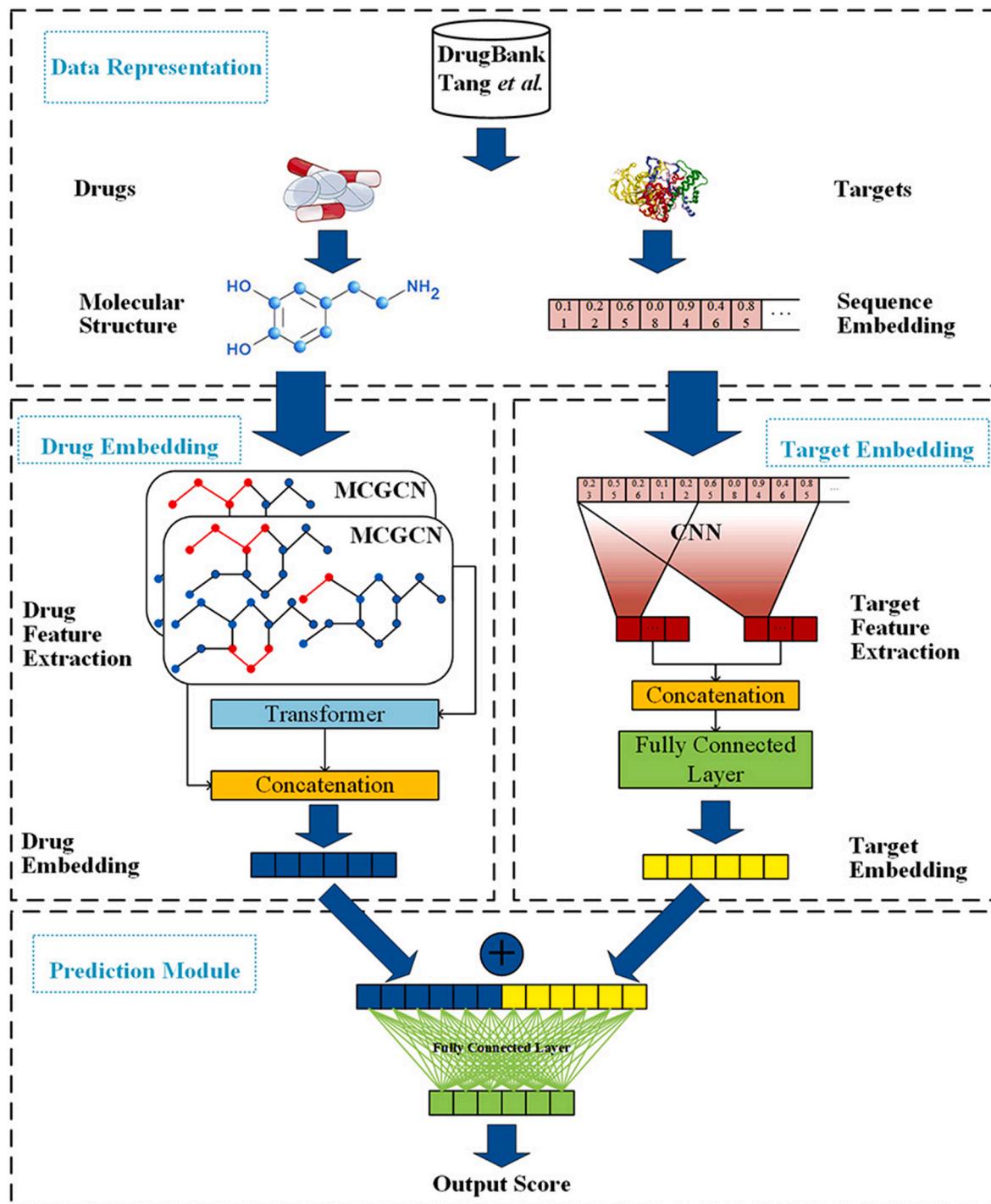


Fig. 1. Overview of the DeepMGT-DTI model.

information, and drug-target action relationships are input. The drug and target feature information is extracted using the transformer and convolutional neural networks that fuse the graph information, respectively. Finally, we connect the drug and target features and use a fully connected neural network for DTI prediction.

### 3.1. Data representation

We process and represent the drug and target data before model training. The molecular chemical structure specific to each drug is represented as a graph, whose vertices and edges represent the drug's atoms and chemical bonds, respectively. Each drug molecule is represented using a feature matrix and an adjacency matrix. Each row of the feature matrix corresponds to a property of one atom. Each drug is denoted as  $\{g_i = (X_i, A_i)_{i=1}^N\}$ , where  $N$  is the type of drug,  $X_i \in \mathbb{R}^{D_i \times C}$  denotes the feature matrix of the drug,  $A_i \in \mathbb{R}^{D_i \times D_i}$  is the adjacency matrix of the drug,  $D_i$  denotes the number of atoms of the  $i$ th drug, and  $C$  is the number of characteristic channels of the atoms.

We perform the embedding operation on the sequence structure of the target points before the model training. The embedding process is shown in Fig. 2. We randomly initialize a lookup table corresponding to all occurrences of amino acids in the target point sequence of size  $26 \times 20$ . Because the embedding vector is trainable, the relevant information in the lookup table changes as the model is optimized. Based on the lookup table, we can correspond the amino acids in the target sequence to construct the embedding matrix of the target sequence. The length of the embedding matrix is the maximum length in the target point sequence, which we set to 2,500. The width is the width of the lookup table.

### 3.2. Drug embedding

Considering the specificity of drug molecules and the limitation of the graph convolutional neural network, we designed a transformer network that fuses multilayer graph information that is used to extract the drug features. The transformer network with fused multilayer graph information consists of the molecular complementary graph

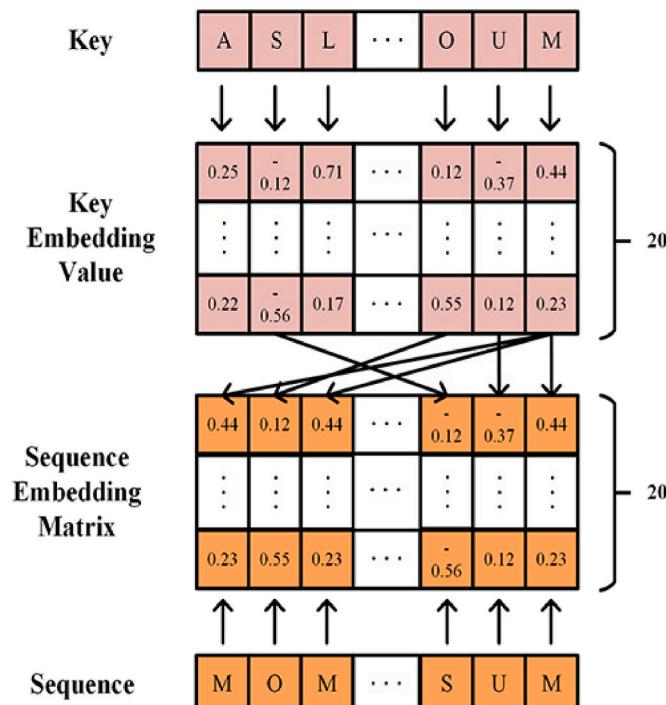


Fig. 2. Target sequence embedding representation.

convolutional neural network (MCGCN) and modified transformer network. The model structure is shown in Fig. 3. The network takes the molecular structure of the drug as input. The MCGCN extracts the drug molecular graph information, and the features of each of its layers are also input into the transformer network. Different multi-headed attention mechanisms are used in the transformer network for further feature extraction. By further extracting the features of different hidden layers in the MCGCN, the information of the edges between atoms in the drug molecule graph is preserved, which enriches the features of drug molecules and overcomes the shortcoming of the GCN, which ignores the relationship between nodes.

#### 3.2.1. Molecular complementary graph convolutional neural network

Because of the uniqueness of the molecular structure of a drug, different drugs have different molecular structures; thus, the molecular graph of each drug is unique. The original graph convolutional neural network [22] aims to classify nodes for a single graph. To adapt the convolutional neural network to our task, we used an MCGCN to ensure

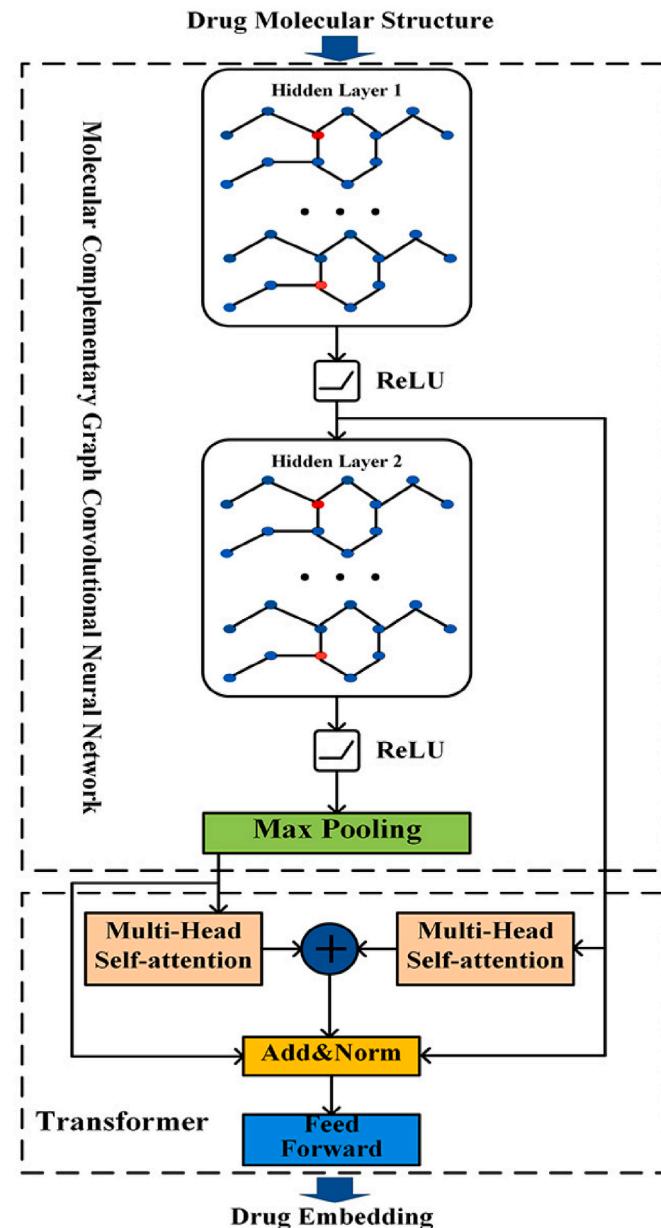


Fig. 3. Drug embedding structure.

that each drug molecule's adjacency matrix and feature matrix are the same size by adding a complementary graph to the original drug molecular graph where the original and the complementary graphs are independent. The MCGCN has two hidden layers, each followed by a ReLU activation function, and uses maximum pooling at the end of the MCGCN to reduce the dimensionality of the data.

Assuming that the original graph of the drug  $\{g_i = (X_i, A_i)\}_{i=1}^N$  is given as representation  $N$ , the complementary graph of the drug can be represented as  $\{g_i^c = (X_i^c, A_i^c)\}_{i=1}^N$ , where the feature and adjacency matrices are  $X_i^c \in \mathbb{R}^{(D-D_i) \times C}$ ,  $A_i^c \in \mathbb{R}^{(D-D_i) \times (D-D_i)}$ , and  $D$  denotes the number of atoms after the complementation and is set to 100. After the complementation operation, the molecular graph of the drug can be represented as

$$A_i' = \begin{bmatrix} A_i & B_i \\ B_i^T & A_i^c \end{bmatrix}, X_i' = \begin{bmatrix} X_i \\ X_i^c \end{bmatrix}, \quad (1)$$

where  $B_i \in \mathbb{R}^{D_i \times (D-D_i)}$  denotes the connection matrix between the original and the complemented graph of the  $i$ th drug, and  $A_i' \in \mathbb{R}^{D \times D}$ ,  $X_i' \in \mathbb{R}^{D \times C}$  are the adjacency matrix and the feature matrix after the complementation. The drug is denoted as  $f(A_i', X_i')$  in the MCGCN. Moreover, each layer of the MCGCN is denoted as

$$H_i^{(l+1)} = \sigma\left(\tilde{M}_i^{-\frac{1}{2}} \tilde{A}_i' \tilde{M}_i'^{-\frac{1}{2}} H_i^{(l)} \Theta^{(l)}\right), \quad (2)$$

where  $\tilde{A}_i' = A_i' + I_N$  is the adjacency matrix with self-attention added;  $\tilde{M}_i'$  is the weight matrix of  $\tilde{A}_i'$ , where  $\tilde{M}_i'[j,j] = \sum_k \tilde{A}_i'[j,k]$ .  $H_i^{(l)}$  and  $\Theta^{(l)}$  are the convolution signals and filter parameters of layer  $l$ ; and  $\sigma(\bullet)$  is the activation function, which is set to  $\text{ReLU}(\bullet) = \max(0, \bullet)$ . We further denote the first  $D_i$  rows of  $H_i^{(l)}$  as  $H_i^{(l,\alpha)}$  and the remaining  $(D - D_i)$  rows as  $H_i^{(l,\beta)}$  to obtain the propagation function of MCGCN between the layers:

$$H_i^{(l+1,\alpha)} = \sigma\left(\left((\tilde{M}_i + M_i^B)^{-\frac{1}{2}} \tilde{A}_i (\tilde{M}_i + M_i^B)^{-\frac{1}{2}} H_i^{(l,\alpha)} + (\tilde{M}_i + M_i^B)^{-\frac{1}{2}} \tilde{A}_i (\tilde{M}_i^c + M_i^{B^T})^{-\frac{1}{2}} H_i^{(l,\beta)}\right) \Theta^{(l)}\right). \quad (3)$$

### 3.2.2. Transformer network

The transformer network [37] is based on an attention mechanism and contains both an encoder and a decoder. In DTI prediction, the transformer network extracts additional features. Therefore, only the encoder part of the transformer network is used, as shown in Fig. 3. We have improved the multi-headed attention mechanism in the encoder. The drug feature information from different hidden layers of the MCGCN is processed using different multi-headed attention mechanisms. In the transformer network, we use two multi-headed attention modules with different heads, 4 and 6. The feature vectors processed by the multi-headed attention mechanism are connected by the concat operation and fed to the layer normalization part, then the fully connected forward neural network.

In the transformer network, the attention functions  $Q$ ,  $K$  and  $V$ , represent the query, key, and value, respectively. The multi-headed attention mechanism projects  $Q$ ,  $K$  and  $V$ , undergo different linear transformations, obtain the representation results within each attention module by splicing different attention results and then obtain the final result by splicing different attention module representation results. The formula is expressed as follows:

$$\text{MultiHead}_i(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_i), \quad (4)$$

$$\text{AllMul} = \text{Concat}(\text{MultiHead}_1, \dots, \text{MultiHead}_j), \quad (5)$$

where  $\text{AllMul}$  represents the concatenation of vectors from different

attention modules,  $i$  denotes the number of heads in each attention module, and  $j$  represents the number of multi-headed attention modules, set to 2.

In addition to the multi-headed attention module, each encoder layer contains a fully connected feedforward network and a two-layer linear transform. The two-layer linear transform uses the ReLU activation function.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2. \quad (6)$$

### 3.3. Target embedding

We use CNN capture to extract the residue information from the target point sequence and the feature information from the target point.

The structure of the target point feature extraction module is shown in Fig. 4. The target point sequence goes through the embedding layer to get the embedding matrix representation of the sequence. The length of the target point sequence is filled with empty labels if it is smaller than the length of the embedding matrix. We use convolution kernels of different sizes to convolve the target sequence embedding matrix. The size of the convolution kernel is 10, 15, and 20, respectively, and the step size is 1. From the  $j$ th amino acid to the  $(j + WS)$  amino acid, it can be defined as

$$(x * w)_j = \sum_{a=1}^{ES} \sum_{b=0}^{WS-1} w_{a,b} x_{a,j+b}, \quad (7)$$

where  $WS$  is the size of the convolutional kernel. After each convolutional layer, we use the ELU activation function.

$$\sigma(x, x) = \begin{cases} \sigma(e^x - 1) & x < 0 \\ x & x \geq 0 \end{cases}. \quad (8)$$

Finally, to extract the most important local features, we perform a global maximum pooling of the results after each convolution, defined as

$$\text{MaxPooling}(E_{Pk}) = \max((x * w)_j), \quad (9)$$

where  $j$  contains all the convolution results of the embedding matrix of the target sequence  $Pk$ . After maximum pooling, we extract the features of the interacting amino acids in the sequence. Finally, the result after maximum pooling is fed to the fully connected layer to obtain the embedding representation of the target sequence.

### 3.4. Prediction module

We use a concatenation operation to merge the embedded representation of the drug molecule structure and the target sequence. The merged vector representation is then fed into the constructed fully concatenated layer along with the original label to predict drug-target interaction.

In the DeepMGT-DTI model, the loss is calculated using a binary cross-entropy function.

$$J(W, b) = -\frac{1}{n} \sum_i^n [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (10)$$

To prevent overfitting, we optimize the loss function with the  $L2$  parametrization.

$$J_{L2}(W, b) = J(W, b) + \lambda \sum_{l=1}^{L-1} W_l^2. \quad (11)$$

## 4. Experiments and results

### 4.1. Dataset

The data used in the experiments includes drug and target data, sequences representation of the targets and SMILES sequences of the drugs. The drug and target data with DTI relationship were extracted

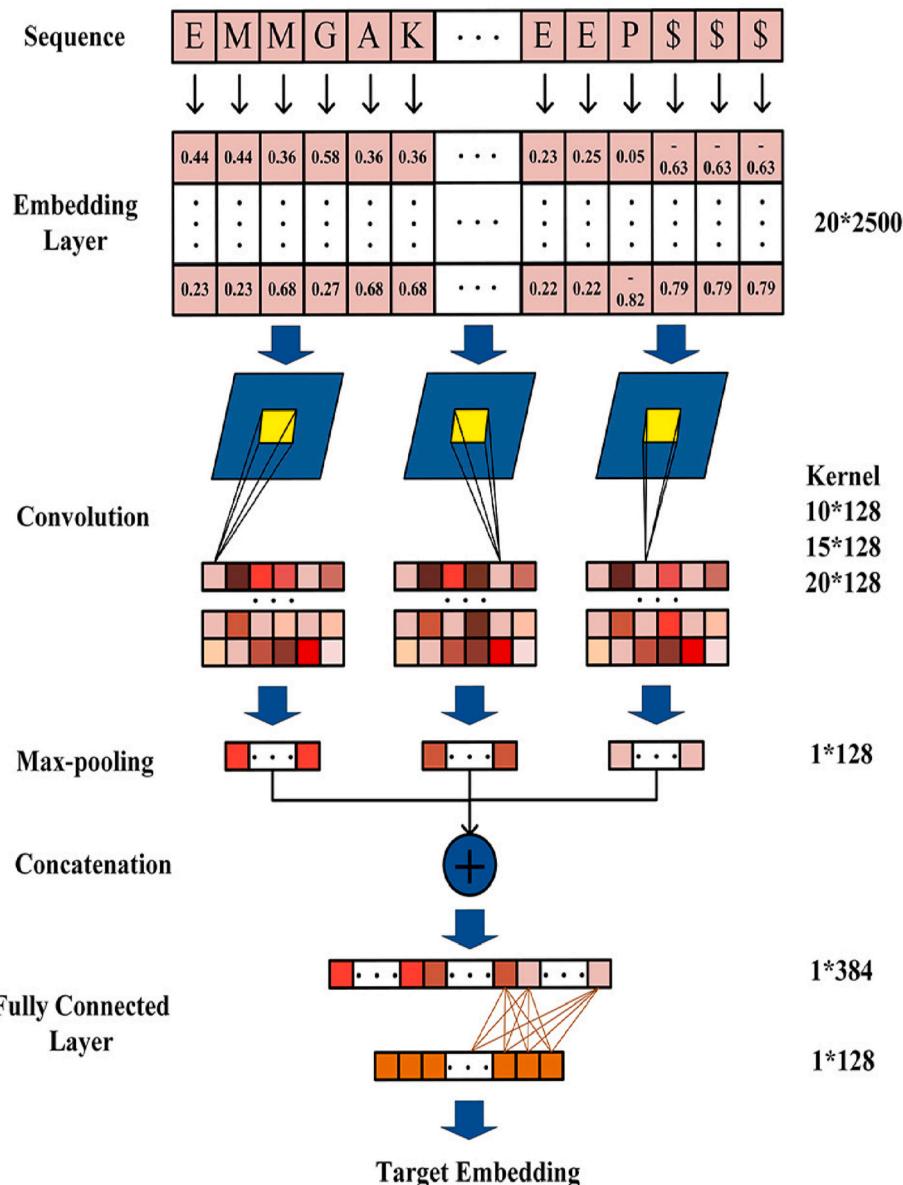


Fig. 4. Target embedding structure.

from Tang et al. [36] and the DrugBank database [44]. The sequences representation of the target and SMILES sequences of the drugs are obtained from KEGG database [20] and PubChem database [39], respectively. After we removing the DTI data with missing sequence representations, 12,496 drug molecular structures, 5,462 target sequences and 21,158 DTI data were obtained, as shown in Table 1. We randomly constructed negative examples in the ratio of 1:2 (positive examples to negative examples) and added them to the data. The training and test sets of the experiments in this paper were divided in an 8:2 ratio.

**Table 1**  
Statistics of the dataset.

Nodes Name	Number	DTI Type	Number
Drug	12,496	Positive (known)	21,158
Target	5,462	Negative (unknown)	42,316
<b>Total</b>	<b>17,958</b>	<b>Total</b>	<b>63,474</b>

#### 4.2. Experimental settings

We used DTI data related to Delta variants and Alzheimer's disease from the PubChem database in our case study. We used an NVIDIA GeForce RTX 3090 graphics card with 24 GB of RAM for training. The weights were updated using the Adam optimizer with an optimized loss

**Table 2**  
Model experimental parameters.

Experimental parameters	Size
epoch	150
Batch size	32
learning rate	0.000 1
Decay rate	0.000 1
dropout size	0.1
CNN input size	1,500
CNN filters	128
CNN windows size	10,15,20
GCN input size	75
Transformer MultiHead Attention Numbers	2

function. The experimental model parameters are shown in Table 2.

#### 4.3. Evaluation metrics

We chose four metrics to evaluate the effectiveness of the model comprehensively: the area under the ROC curve (AUC), area under the PR curve (AUPR), F1, and accuracy (ACC). We also used the sensitivity (Sen), specificity (Spe), and precision (Pre) metrics to evaluate the performance of the model more comprehensively. The F1, ACC, Sen, Spe and Pre are defined as follows,

$$F1 = \frac{2TP}{2TP + FP + FN}$$

$$ACC = \frac{TP + TN}{FP + FN + TP + TN}$$

$$Sen = \frac{TP}{P}$$

$$Spe = \frac{TN}{N}$$

$$Pre = \frac{TP}{TP + FP}$$

where TP and TN are the number of drug-target with interaction and drug-target without interaction that were successfully identified, respectively. FP(FN) represents the number of drug-target with (without) interaction examples that are incorrectly identified. P(N) represents the number of drug-target with (without) interaction examples.

#### 4.4. Experimental results

##### 4.4.1. Comparison with existing methods

To verify the performance of the DeepMGT-DTI model, we compared the proposed model to two machine learning models: Random Forest [3] and SVM [10], and five SOTA deep learning models: Deep DTA [30], Deep DTI [43], Deep Conv-DTI [23], TransformerCPI [7] and ML-DTI [45]. We applied these comparison methods to the dataset in the study and modified some parameters of the models. The obtained experimental results are shown in Table 3. The bold value indicates the best performance.

The experimental results show that the ML-DTI and DeepMGT-DTI models outperformed the Deep DTA, Deep DTI, Deep Conv-DTI and TransformerCPI models for each evaluation index. This is because the Deep DTA, Deep DTI, Deep Conv-DTI and TransformerCPI models focused on the sequence structure of the target. To a certain extent, the capture of drug molecular features was neglected. However, the ML-DTI and DeepMGT-DTI models focused more on mining drug molecule information while capturing the sequence features of the target. Thus, more feature information in the data was captured, and the performance of the models was improved. By comparing the results of the Deep DTA, Deep DTI, and Deep Conv-DTI models, it can be seen that the Deep Conv-DTI model achieved the best experimental results. This is because Deep

**Table 3**  
Comparison of DTI prediction results of DeepMGT-DTI and baseline models.

Model	AUC	AUPR	F1	ACC
Random Forest [3]	58.62	49.68	60.06	66.83
SVM [10]	60.62	53.36	62.82	67.93
Deep DTA [30]	75.96	63.71	71.95	79.81
Deep DTI [43]	79.39	70.04	72.86	80.33
Deep Conv-DTI [23]	86.10	68.35	72.36	80.68
TransformerCPI [7]	86.69	74.39	75.66	83.26
ML-DTI [45]	88.02	<b>77.63</b>	77.10	83.34
DeepMGT-DTI	<b>90.24</b>	77.11	<b>79.31</b>	<b>85.15</b>

DTI uses the physicochemical properties in the whole target sequence for characterization. The entire protein sequence contains many subsequences or structural domains that are not involved in the interaction with the compound. Deep DTA is designed and optimized for specific proteins. The performance is not suitable for target data with multiple classes of proteins.

On the contrary, the Deep Conv-DTI model effectively extracts the features of target sequences by extracting the key information of residues in the target sequences and achieves better modeling results. By comparing the results of the ML-DTI and DeepMGT-DTI models, this can be clearly seen. The results of the AUC, F1, and ACC evaluation indexes of the DeepMGT-DTI model were better than those of the ML-DTI model. Based on multihead attention and position-aware attention, ML-DTI uses a mutual learning mechanism to bridge the gap between the drug encoder and target encoder. DeepMGT-DTI captures the molecular structure of the drug compounds by a transformer network incorporating multilayer graph information. In contrast, ML-DTI learns more the feature information of drugs and targets from a global perspective, and uses probability maps to effectively filter important features. Therefore, ML-DTI achieved the highest average AUPR (AUPR = 77.63%) of the DTI prediction experiments, which is 0.5% higher than that of DeepMGT-DTI.

##### 4.4.2. Ablation experiments

To verify the validity of each module of the model, we designed and conducted ablation experiments, and the results are summarized in Table 4. The bold value indicates the best performance.

**Effectiveness of MCGCN.** As shown in Table 4, MCGCN can further improve the performance by (1.2%, 2.3%, 2.4%, 0.2%) on AUC, AUPR, F1 and ACC. Based on MCGCN, our model learns the atomic information and the interaction information between atoms in the drug molecule structure map. It can be concluded that MCGCN effectively retain the information of drug molecules. This is why MCGCN can improve the prediction performance for DTI.

**Effectiveness of Transformer network.** As shown in Table 4, Transformer network contributes to a performance improvement of around (1.2%, 2.3%, 2.5%, 2.1%) on AUC, AUPR, F1 and ACC. This result shows that Transformer network extracts the feature information from MCGCN and improves the prediction performance of DeepMGT-DTI.

**Effectiveness of multi-head self-attention.** The multi-head self-attention module brings substantial improvements (1.7%, 4.2%, 2.0%, 2.2%) on AUC, AUPR, F1 and ACC. The improvement is especially significant on AUPR. This is consistent with our expectation that multi-head self-attention module not only preserves the feature information of MCGCN to the greatest extent, but also extracts the missing information in the hidden layer of MCGCN. It effectively overcome the deficiency of graph neural network for learning edge features and improve the prediction performance of the model.

**Robustness of imbalanced data.** To verify the robustness of the DeepMGT-DTI model on imbalanced data sets, we expand the test data in different ratios (positive sample: negative sample = 1:2, 1:5, 1:10 and 1:20) while maintaining the balance of the training data. The experimental results are shown in Table 5. The results shows that the comprehensive performance of DeepMGT-DTI is less affected by the class imbalance, while AUC and ACC show an increasing trend with the expansion of the sample ratio.

##### 4.5. Example of drug prediction

To validate the model's effectiveness for drug repositioning, the DeepMGT-DTI model was used to find therapeutic drugs for COVID-19. The Delta variant of COVID-19 currently covers 130 countries, is 55%–90% more infectious than previous COVID-19 variants, and 30%–100% more infectious than the Alpha variant in several countries, including the US. The Delta target is identified as the major target of COVID-19

**Table 4**

Results of ablation experiments on DTI prediction. The “w/o” indicates “without”.

Model	AUC	AUPR	F1	ACC	Sen	Spe	Pre
w/o MCGCN and Transformer	86.10	68.35	72.36	80.68	76.35	81.62	70.26
w/o Transformer	87.36	70.69	74.78	80.89	78.66	83.08	74.34
w/o Multi-Head Self-attention	88.53	72.97	77.26	82.96	81.50	84.90	74.88
DeepMGT-DTI	<b>90.24</b>	<b>77.11</b>	<b>79.31</b>	<b>85.15</b>	<b>82.17</b>	<b>86.73</b>	<b>76.65</b>

**Table 5**

Experimental results of DeepMGT-DTI on an imbalanced dataset.

Positive: Negative	AUC	ACC	Sen	Spe
1:2	90.24	85.15	82.17	86.73
1:5	91.88	87.79	80.35	89.32
1:10	92.82	88.43	83.98	88.88
1:20	92.95	88.98	83.53	89.29

by Bernal et al. [26]. Thus, the rapid discovery and identification of drugs that effectively inhibit the Delta variant are urgently needed to control the pandemic. In this study, information about Delta targets was extracted from the PubChem [39] database. The DeepMGT-DTI model was used for the prediction of potential therapeutic drugs. The results are shown in Table 6.

In the experimental results, four of the top five drugs, including Tramadol [14], were clinical treatments for COVID-19 or had literature support for inhibition of COVID-19. Tramadol [14], Amitriptyline [5], and Dextromethorphan [33] all had Delta targets with close interactions. Dexamethasone [32] and Dextromethorphan [33] are widely used in the clinical treatment of COVID-19 and have successfully alleviated the complications of COVID-19. Tramadol can increase antioxidant enzymes, superoxide dismutase, and glutathione peroxidase. It also reduces the effects of malondialdehyde, thus protecting COVID-19 patients from disease complications [14]. It has been shown that treating cells with different concentrations of Amitriptyline reduces the chance of cells being infected by 90%. This provides the basis for the use of Amitriptyline for COVID-19 treatment [5]. In order to demonstrate the distinguishing feature of DeepMGT-DTI, we used DeepConv-DTI [23] method to compare with our model in the COVID-19 potential

**Table 6**

The top 5 drugs related to COVID-19 recommended by DeepMGT-DTI.

No	Compound CID	Drug Name	Evidence
1	33 741	Tramadol	Tramadol could protect the COVID-19 patient from disease complications by increasing the antioxidant enzymes superoxide dismutase and glutathione peroxidase while diminishing malondialdehyde [14].
2	5311 356	Pholcodine	None.
3	5743	Dexamethasone	A recent clinical trial has revealed that dexamethasone and convalescent plasma treatment can reduce mortality in patients with severe COVID-19 [32].
4	2160	Amitriptyline	Treating volunteers with a low dose of amitriptyline prevents infection of freshly isolated nasal epithelial cells with a pp-VSV-SARS-CoV-2 spike [5].
5	5360 696	Dextromethorphan	In this scenario, we undertook a repurposing project of common cough and cold drugs—Dextromethorphan, Prednisolone, and Dexamethasone—to explore their anti-Covid property. Individual and sequential docking study with MD simulation and RMSD and RMSF analyses revealed that a combination of these three considered ligands may prove to be a successful therapy against COVID-19 [33].

therapeutic drugs prediction experiments. The results showed that the top 5 drugs predicted by DeepConv-DTI hit two COVID-19 therapeutic drugs, less than the four drugs of our method. This is because that DeepMGT-DTI represents SMILES sequences as a graph, effectively extracting the information of interactions among atoms in drug molecules by using MCGCN and Transformer networks.

Except for COVID-19, we also predicted potential therapeutic drugs for Alzheimer's disease. In the experiment, we selected apolipoprotein E closely related to Alzheimer's disease as targets, and searched for corresponding acting drugs in KEGG database and PubChem database to construct DTI data. Appropriate negative examples were selected for samples expansion. Among the top 5 predicted drugs, Rivastigmine tartrate and Memantine hydrochloride have been used in the clinical treatment of Alzheimer's disease.

## 5. Conclusion

This study proposed a transformer network with fused graph information for predicting DTIs. The DeepMGT-DTI model takes the molecular structure map of a drug and the sequence structure of a target as inputs. A transformer network extracts the features of the drug with fused graph information, and the prediction results are output. The experimental results indicated that the improved transformer network fused the feature information between different layers in the graph convolutional neural network, which compensates for the lack of learning of side features by the graph convolutional neural network. The DeepMGT-DTI model was also used to predict the therapeutic drugs for COVID-19, and four of the top five recommended drugs, including Tramadol, were found to have inhibitory effects on COVID-19. The results indicate the scalability of the model. In the future, we will further improve the model's generalization by incorporating more drug-target information and predicting potential therapeutic agents for more diseases (e.g., diabetes) to enhance the drug development process and reduce its cost.

## CRediT authorship contribution statement

**Peiliang Zhang:** Conceptualization of this study, Data processing, Methodology, Analysis of results, Paper writing and revision. **Ziqi Wei:** Data processing, Analysis of results, Paper revision. **Chao Che:** Conceptualization of this study, Methodology, Paper revision. **Bo Jin:** Methodology, Paper revision.

## Funding

This research was funded by the National Natural Science Foundation of China (No. 62 076 045, No. 62 102 058) and the Guidance Program of Liaoning Natural Science Foundation (No. 2019-ZD-0569). The funders did not play any role in the design of the study, the collection, analysis, and interpretation of data, or in writing of the manuscript.

## Ethical approval

This article does not contain any data, or other information from studies or experimentation, with the involvement of human or animal subjects.

## Availability of data and material

The data used in this study were obtained from the DrugBank database (<https://go.drugbank.com/>), the KEGG database (<https://www.genome.jp/kegg/>) and the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>).

## Declaration of competing interest

The authors declare no competing financial and non-financial interests.

## References

- [1] M. Bahi, M. Batouche, Convolutional neural network with stacked autoencoders for predicting drug-target interaction and binding affinity, *Int. J. Data Min. Model. Manag.* 13 (2021) 81–113, <https://doi.org/10.1504/IJDMMM.2021.112914>.
- [2] J.R. Bock, D.A. Gough, Virtual screen for ligands of orphan g protein-coupled receptors, *J. Chem. Inf. Model.* 45 (2005) 1402–1414, <https://doi.org/10.1021/ci050006d>.
- [3] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [4] B. Cao, X. Zhang, J. Wu, B. Wang, Q. Zhang, X. Wei, Minimum free energy coding for dna storage, *IEEE Trans. NanoBioscience* 20 (2021) 212–222, <https://doi.org/10.1109/TNB.2021.3056351>.
- [5] A. Carpinteiro, M.J. Edwards, M. Hoffmann, G. Kochs, B. Gripp, S. Weigang, C. Adams, E. Carpinteiro, A. Gulbins, S. Keitsch, et al., Pharmacological inhibition of acid sphingomyelinase prevents uptake of sars-cov-2 by epithelial cells, *Cell Reports Medicine* 1 (2020) 100142, <https://doi.org/10.11016/j.xcrm.2020.100142>.
- [6] C. Che, P. Zhang, M. Zhu, Y. Qu, B. Jin, Constrained transformer network for ecg signal processing and arrhythmia classification, *BMC Med. Inf. Decis. Making* 21 (2021) 1–13, <https://doi.org/10.1186/s12911-021-01546-2>.
- [7] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, M. Zheng, Transformercipi: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments, *Bioinformatics* 36 (2020) 4406–4414, <https://doi.org/10.1093/bioinformatics/btaa524>.
- [8] S. Cheng, L. Zhang, B. Jin, Q. Zhang, X. Lu, Drug Target Prediction Using Graph Representation Learning via Substructures Contrast, 2021, <https://doi.org/10.3390/app11073239>.
- [9] Y. Chu, A.C. Kaushik, X. Wang, W. Wang, Y. Zhang, X. Shan, D.R. Salahub, Y. Xiong, D.Q. Wei, Dti-cdf: a cascade deep forest model towards the prediction of drug–target interactions based on hybrid features, *Briefings Bioinf.* 22 (2021) 451–462, <https://doi.org/10.1093/bib/bbz152>.
- [10] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297, <https://doi.org/10.1007/BF00994018>.
- [11] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: *arXiv Preprint arXiv: 1810.04805*, 2018, <https://doi.org/10.18653/v1/N19-1423>.
- [12] M.S. Donaldson, J.M. Corrigan, L.T. Kohn, et al., *To Err Is Human: Building a Safer Health System*, 2000.
- [13] S. Ekins, The next era: deep learning in pharmaceutical research, *Pharmaceut. Res.* 33 (2016) 2594–2603, <https://doi.org/10.1007/s11095-016-2029-7>.
- [14] N.E. El-Ashmawy, A.H.A. Lashin, K.M. Okasha, A.M.A. Kamer, T.M. Mostafa, M. El-Aasr, A.E. Goda, Y.A. Haggag, H.O. Tawfik, M.A. Abo-Saif, The plausible mechanisms of tramadol for treatment of covid-19, *Med. Hypotheses* 146 (2021) 110468, <https://doi.org/10.1016/j.mehy.2020.110468>.
- [15] A. Ezzat, P. Zhao, M. Wu, X.L. Li, C.K. Kwok, Drug-target interaction prediction with graph regularized matrix factorization, *IEEE ACM Trans. Comput. Biol. Bioinf.* 14 (2016) 646–656, <https://doi.org/10.1109/TCBB.2016.2530062>.
- [16] E. Gawehn, J.A. Hiss, G. Schneider, Deep learning in drug discovery, *Molecular informatics* 35 (2016) 3–14, <https://doi.org/10.1002/minf.201501008>.
- [17] R. Gowthaman, S.A. Miller, S. Rogers, J. Khowsathit, L. Lan, N. Bai, D.K. Johnson, C. Liu, L. Xu, A. Anbanandam, et al., Darc: mapping surface topography by ray-casting for effective virtual screening at protein interaction sites, *J. Med. Chem.* 59 (2016) 4152–4170, <https://doi.org/10.1021/acs.jmedchem.5b00150>.
- [18] K. Huang, T. Fu, L.M. Glass, M. Zitnik, C. Xiao, J. Sun, Deepurpose: a deep learning library for drug–target interaction prediction, *Bioinformatics* 36 (2020) 5545–5547, <https://doi.org/10.1093/bioinformatics/btaa100>.
- [19] K. Huang, C. Xiao, L.M. Glass, J. Sun, Moltrans: molecular interaction transformer for drug–target interaction prediction, *Bioinformatics* 37 (2021) 830–836, <https://doi.org/10.1093/bioinformatics/btaa880>.
- [20] M. Kanehisa, S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, From genomics to chemical genomics: new developments in kegg, *Nucleic Acids Res.* 34 (2006) D354–D357, <https://doi.org/10.1093/nar/gkj102>.
- [21] I. Kapetanovic, Computer-aided drug discovery and development (cadd): in silico-chemico-biological approach, *Chem. Biol. Interact.* 171 (2008) 165–176, <https://doi.org/10.1016/j.cbi.2006.12.006>.
- [22] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016 *arXiv preprint arXiv:1609.02907* <https://doi.org/abs/1609.02907>.
- [23] I. Lee, J. Keum, H. Nam, Deepconv-dti: prediction of drug–target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.* 15 (2019), e1007129, <https://doi.org/10.1371/journal.pcbi.1007129>.
- [24] S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, H. Xiong, Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity, 2021 *arXiv preprint arXiv:2107.10670* <https://doi.org/abs/2107.10670>.
- [25] Q. Liu, Z. Hu, R. Jiang, M. Zhou, Deepcdr: a hybrid graph convolutional network for predicting cancer drug response, *Bioinformatics* 36 (2020) i911–i918, <https://doi.org/10.1093/bioinformatics/btaa822>.
- [26] J. Lopez-Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrerla, R. Myers, C.N. Campbell, G. Amirthalingam, M. Edmunds, M. Zambon, K.E. Brown, S. Hopkins, M. Chand, M. Ramsay, Effectiveness of covid-19 vaccines against the b.1.617.2 (delta) variant, *N. Engl. J. Med.* 385 (2021) 585–594, <https://doi.org/10.1056/NEJMoa2108891>.
- [27] L. Maziarka, T. Daniel, S. Mucha, K. Rataj, J. Tabor, S. Jastrzbski, Molecule attention transformer, in: *arXiv Preprint arXiv:2002.08264*, 2020 [arxiv.org/abs/2002.08264v1](https://arxiv.org/abs/2002.08264v1).
- [28] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, Graphdta: predicting drug–target binding affinity with graph neural networks, *Bioinformatics* 37 (2021) 1140–1147, <https://doi.org/10.1093/bioinformatics/btaa921>.
- [29] R.S. Olayan, H. Ashoor, V.B. Bajic, Ddr: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches, *Bioinformatics* 34 (2018) 1164–1173, <https://doi.org/10.1093/bioinformatics/btx731>.
- [30] H. ÖzTÜRK, A. Özgür, E. Ozkirimli, Deepdta: deep drug–target binding affinity prediction, *Bioinformatics* 34 (2018) i821–i829, <https://doi.org/10.1093/bioinformatics/bty593>.
- [31] A. Prakash, K. Chitta, A. Geiger, Multi-modal fusion transformer for end-to-end autonomous driving, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7077–7087. [arxiv.org/abs/2104.9224v1](https://arxiv.org/abs/2104.9224v1).
- [32] C.D. Raj, D.K. Kandaswamy, R.C.S.R. Danduga, R. Rajasabapathy, R.A. James, Covid-19: molecular pathophysiology, genetic evolution and prospective therapeutics—a review, *Arch. Microbiol.* (2021) 1–15, <https://doi.org/10.1007/s00203-021-02183-z>.
- [33] I. Sarkar, A. Sen, In silico screening predicts common cold drug dextromethorphan along with prednisolone and dexamethasone can be effective against novel coronavirus disease (covid-19), *J. Biomol. Struct. Dyn.* (2020) 1–5, <https://doi.org/10.1080/07391102.2020.1850528>.
- [34] J.W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical r&d efficiency, *Nat. Rev. Drug Discov.* 11 (2012) 191–200, <https://doi.org/10.1038/nrd3681>.
- [35] M. Takarabe, M. Kotera, Y. Nishimura, S. Goto, Y. Yamanishi, Drug target prediction using adverse event report systems: a pharmacogenomic approach, *Bioinformatics* 28 (2012) i611–i618, <https://doi.org/10.1093/bioinformatics/bts413>.
- [36] C. Tang, C. Zhong, D. Chen, J. Wang, Drug–target interactions prediction using marginalized denoising model on heterogeneous networks, *BMC Bioinf.* 21 (2020) 1–29, <https://doi.org/10.1186/s12859-020-03662-8>.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008, <https://doi.org/10.5555/3295222.3295349>.
- [38] F. Wan, L. Hong, A. Xiao, T. Jiang, J. Zeng, Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions, *Bioinformatics* 35 (2019) 104–111, <https://doi.org/10.1093/bioinformatics/bty543>.
- [39] Y. Wang, S.H. Bryant, T. Cheng, J. Wang, A. Gindulyte, B.A. Shoemaker, P. A. Thiessen, S. He, J. Zhang, Pubchem bioassay: 2017 update, *Nucleic Acids Res.* 45 (2017) D955–D963, <https://doi.org/10.1093/nar/gkw1118>.
- [40] Y. Wang, Y. Min, X. Chen, J. Wu, Multi-view graph contrastive representation learning for drug–drug interaction prediction, in: *Proceedings of the Web Conference 2021*, 2021, pp. 2921–2933, <https://doi.org/10.1145/3442381.3449786>.
- [41] Y.B. Wang, Z.H. You, S. Yang, H.C. Yi, Z.H. Chen, K. Zheng, A deep learning-based method for drug–target interaction prediction based on long short-term memory neural network, *BMC Med. Inf. Decis. Making* 20 (2020) 1–9, <https://doi.org/10.1186/s12911-020-1052-0>.
- [42] Y.C. Wang, C.H. Zhang, N.Y. Deng, Y. Wang, Kernel-based data fusion improves the drug–protein interaction prediction, *Comput. Biol. Chem.* 35 (2011) 353–362, <https://doi.org/10.1016/j.combiolchem.2011.10.003>.
- [43] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, H. Lu, Deep-learning-based drug–target interaction prediction, *J. Proteome Res.* 16 (2017) 1401–1409, <https://doi.org/10.1021/acs.jproteome.6b00618>.
- [44] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., Drugbank 5.0: a major update to the drugbank database for 2018, *Nucleic Acids Res.* 46 (2018) D1074–D1082, <https://doi.org/10.1093/nar/gkx1037>.

- [45] Z. Yang, W. Zhong, L. Zhao, C.Y.C. Chen, MI-dti: mutual learning mechanism for interpretable drug–target interaction prediction, *J. Phys. Chem. Lett.* 12 (2021) 4247–4261, <https://doi.org/10.1021/acs.jpclett.1c00867>.
- [46] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman, H. Kilicoglu, Drug repurposing for covid-19 via knowledge graph completion, *J. Biomed. Inf.* 115 (2021) 103696, <https://doi.org/10.1016/j.jbi.2021.103696>.
- [47] X. Zhang, L. Li, M.K. Ng, S. Zhang, Drug–target interaction prediction by integrating multiview network data, *Comput. Biol. Chem.* 69 (2017) 185–193, <https://doi.org/10.1016/j.combiolchem.2017.03.011>.