

# DeepCoVDR: deep transfer learning with graph transformer and cross-attention for predicting COVID-19 drug response

Zhijian Huang<sup>1,\*</sup>, Pan Zhang<sup>2,\*</sup>, Lei Deng <sup>1,\*</sup>

<sup>1</sup>School of Computer Science and Engineering, Central South University, Changsha 410083, China

<sup>2</sup>Hunan Provincial Key Laboratory of Clinical Epidemiology, Xiangya School of Public Health, Central South University, Changsha 410083, China

\*Corresponding author. School of Computer Science and Engineering, Central South University, Changsha 410083, China. E-mail: leidend@csu.edu.cn

<sup>†</sup>Equal contribution.

## Abstract

**Motivation:** The coronavirus disease 2019 (COVID-19) remains a global public health emergency. Although people, especially those with underlying health conditions, could benefit from several approved COVID-19 therapeutics, the development of effective antiviral COVID-19 drugs is still a very urgent problem. Accurate and robust drug response prediction to a new chemical compound is critical for discovering safe and effective COVID-19 therapeutics.

**Results:** In this study, we propose DeepCoVDR, a novel COVID-19 drug response prediction method based on deep transfer learning with graph transformer and cross-attention. First, we adopt a graph transformer and feed-forward neural network to mine the drug and cell line information. Then, we use a cross-attention module that calculates the interaction between the drug and cell line. After that, DeepCoVDR combines drug and cell line representation and their interaction features to predict drug response. To solve the problem of SARS-CoV-2 data scarcity, we apply transfer learning and use the SARS-CoV-2 dataset to fine-tune the model pretrained on the cancer dataset. The experiments of regression and classification show that DeepCoVDR outperforms baseline methods. We also evaluate DeepCoVDR on the cancer dataset, and the results indicate that our approach has high performance compared with other state-of-the-art methods. Moreover, we use DeepCoVDR to predict COVID-19 drugs from FDA-approved drugs and demonstrate the effectiveness of DeepCoVDR in identifying novel COVID-19 drugs.

**Availability and implementation:** <https://github.com/Hhhzj-7/DeepCoVDR>.

## 1 Introduction

With the outbreak of COVID-19, the lives and health of people worldwide have been seriously threatened. The pathogen causing the disease has been named SARS-CoV-2 (Guan et al. 2020), and it is the most pathogenic human coronavirus ever discovered (Cha et al. 2018). In order to prevent COVID-19 from further causing people's health crisis, it is urgent to find SARS-CoV-2 antiviral drugs. Several drugs and vaccines have been approved, but there is still a lack of specific therapeutics to block severe illness and mortality due to the emergence of some new SARS-CoV-2 strains (Spinner et al. 2020).

Recently, certain drugs targeting the viral protein SARS-CoV-2, such as remdesivir (Spinner et al. 2020; Yin et al. 2020), paxlovid (Owen et al. 2021; Hammond et al. 2022), and molnupiravir (Wang et al. 2021b; Jayk Bernal et al. 2022), have been approved by the FDA, but their clinical efficacy remains controversial. Therefore, targeting SARS-CoV-2 host receptors is emerging as another potential solution in the search for specific COVID-19 drugs, in addition to the two entry receptors ACE2 and TMPRSS2, Gordon et al. (2020) identified 332 host–virus protein interactions between SARS-CoV-2 and human proteins through affinity purification mass spectrometry. Moreover, transcriptome profiling also provides an opportunity to predict antiviral drug response by detecting quantitative changes in host gene activity and gene regulation, and changes in the level of gene expression always

contribute to altered pathways, which play a crucial role in determining the phenotypes of cells (Sharifi-Noghabi et al. 2019; Riva et al. 2020). Sometimes similar phenotypes can be observed in different cells due to changes in the same gene activity; For example, recent studies have shown increased expression of immune checkpoint receptors, including PD-1 and CTLA-4, in lung tissue from COVID-19 patients, both of which are associated with cancer development (Sharif-Askari et al. 2021). The research on COVID-19 drug prediction mainly focuses on drug repurposing. Traditional drug repurposing methods rely on expensive and time-consuming wet-lab experiments (Avorn 2015), and computational methods in drug repurposing can effectively alleviate these problems.

So far, some computational drug repurposing methods for screening SARS-CoV-2 antiviral drugs have been proposed, which can be mainly divided into two categories. The first category is the network-driven method, which analyzes the information extracted from the similarity of drugs and other nodes (such as viruses, proteins, etc.). Zhou et al. (2020) presented an integrative antiviral drug repurposing methodology that searches for potential SARS-CoV-2 antiviral drugs by computing the network proximity among the drug's target proteins and HCoV-related proteins. Meng et al. (2021) proposed the SCPMF method which integrated the known drug–virus interaction, drug–drug similarity, and virus–virus similarity network to construct a heterogeneous network, and

used the similarity constraint probability matrix decomposition method to search for COVID-19 therapeutics. Peng *et al.* (2021) constructed a new heterogeneous virus–drug network and developed a novel random walk with restart method (VDA-RWR) for identifying possible virus–drug association related to SARS-CoV-2. The second category is based on deep learning methods. Deepthi *et al.* (2021) proposed DLEVDA, which inputs features incorporating the pairwise similarities of drug chemical structures and virus genome sequences into a convolutional neural network, and finally infers promising candidates against SARS-CoV-2 infection through an XGBoost classifier. Wang *et al.* (2021a) proposed the COVIDVS1, COVIDVS-2, and COVIDVS-3. These three models base on a directed message passing neural network. COVIDVS1 first constructed a broad-spectrum anti-beta-coronavirus compound prediction model, and then COVIDVS-2 and COVIDVS-3 were obtained by applying transfer learning to fine-tune the model using the specific anti-SARS-CoV-2 compound datasets.

Although the prior studies have led to significant progress, there is still room for improvement.

- Researchers have been exploring new potential drugs by testing their antiviral activity against SARS-CoV-2 (Zaliani *et al.* 2022); therefore, the accurate prediction of COVID-19 drug response is of great significance for guiding the design of SARS-CoV-2 antiviral drugs and repurposing existing drugs in the treatment of SARS-CoV-2. However, previous methods are not able to predict COVID-19 drug response.
- The conventional drug feature lacks complete drug chemistry information. For example, the fingerprint-based drug features and similarity network features are insufficient in representing the structure information of drug molecules.
- Pharmacogenomics (Daly 2017) is a subject of exploring the influence of genetic variation of genes on drug treatment effects from the perspective of the genome. For example, transcriptome profiling provides an opportunity to predict antiviral drug response. However, the pharmacogenomics profiles have yet to be well utilized in the study of anti-SARS-CoV-2 drug prediction.
- Due to the limited and sparse publicly available SARS-CoV-2 data, it is challenging to apply deep learning methods to improve the predictive performance of drug response.

In this study, we propose a transfer learning method with a graph transformer for predicting COVID-19 drug response, namely DeepCoVDR. First, we construct a graph transformer that can simultaneously capture the information of neighbor nodes and long-distance node information in drug molecules to extract drug information. Second, we use a feed-forward neural network (DNN) to learn the representation of transcriptomics. Third, a cross-attention module is employed to produce new embeddings of drugs and cell lines which can supplement the relationship between each node in the drug molecule and the corresponding cell line. To solve the problem that the amount of COVID-19 drug response data is very limited, we use cancer cell lines, drugs, and their responses to pretrain our model and then use Vero E6, which is highly permissive to SARS-CoV-2 infection, drugs and their responses to fine-tune the model. DeepCoVDR can not only predict the half-maximal inhibitory concentration (IC<sub>50</sub>) sensitivity value in the regression task but

also classify a drug as sensitive or resistant in the classification task. We have designed many experiments to test the performance of DeepCoVDR. In the classification and regression settings, our model achieves state-of-the-art performance. We also apply our model to screen the FDA-approved drugs from Skuta *et al.* (2017), and the drug candidates also confirm the accuracy of DeepCoVDR.

## 2 Materials

Deep learning methods usually utilize a large amount of data that provide a good basis for solving various problems. Unfortunately, applying deep learning to predict COVID-19 drug response is a major challenge. SARS-CoV-2, as a newly discovered virus, has very limited data available. This will significantly affect the performance of a model in the potential data space. According to the principle of similarity (Willett 2014; Kubinyi 1998), similar molecule structures may work on the same genes. That shows the feasibility of adopting transfer learning on this issue. A sufficient amount of data helps the model to learn enough knowledge of drugs and cell line during pretraining, so we choose cancer dataset which is large enough for pretraining and easy to collect.

For the cancer dataset, we extract the dataset from Jiang *et al.* (2022). This dataset mainly uses the new version of the public database GDSC (Yang *et al.* 2013), GDSC2, containing 135, 242 IC<sub>50</sub> values (natural log-transformed) across 809 cell lines and 198 compounds. It selects 17 777 gene expression values from GDSC2 for each cell line and gets the SMILES strings of all compounds from PubChem (Kim *et al.* 2019).

For the SARS-CoV-2 dataset, we collected a set of IC<sub>50</sub> values of 318 drugs on Vero E6 from Zaliani *et al.* (2022). For the omics data of Vero E6, we use the data from Riva *et al.* (2020), which contain 16 122 gene expression values. To ensure the consistency of feature dimension, we filter the genes that are different from the cancer dataset. Finally, only 11 793 gene expression values in both the cancer dataset and the SARS-CoV-2 dataset are retained.

## 3 Methods

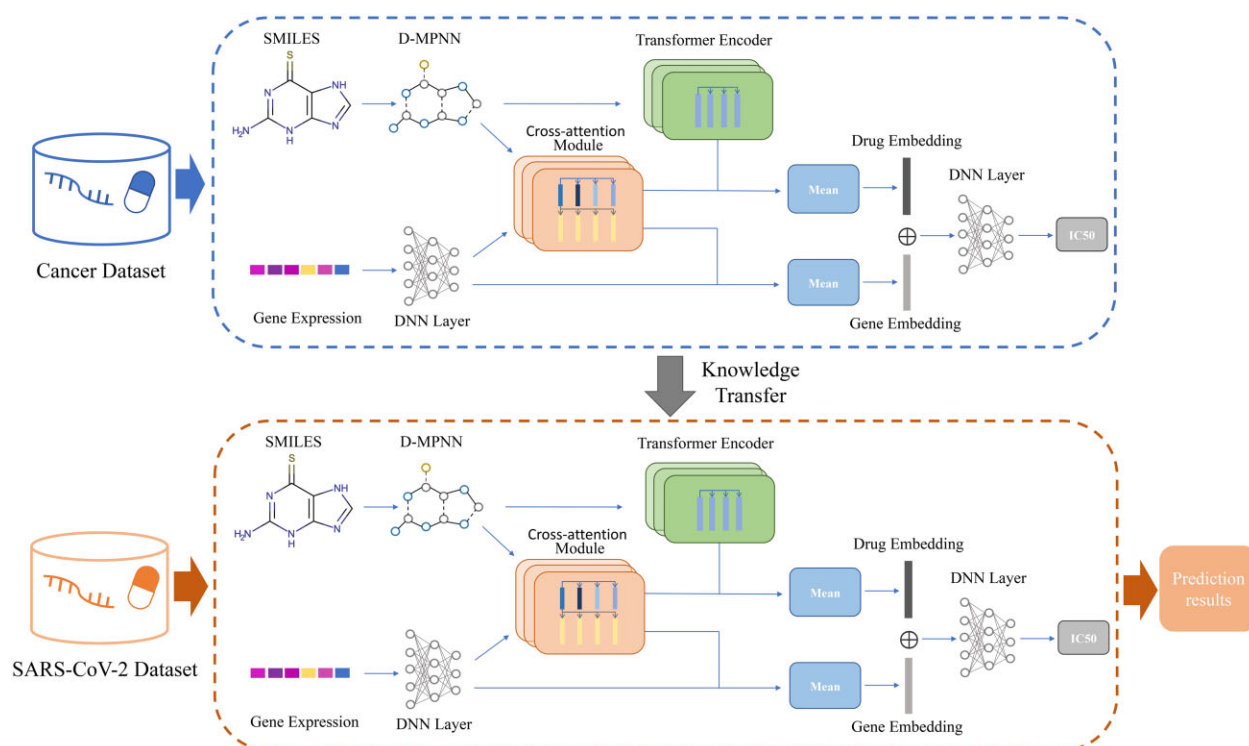
### 3.1 Overview of DeepCoVDR

DeepCoVDR is a deep learning method based on graph transformer, cross-attention, and transfer learning for COVID-19 drug response prediction. The flowchart of DeepCoVDR is shown in Fig. 1. DeepCoVDR has three parts: features mining, cross-attention module, and drug response prediction.

In the first part, we extract drug representation and cell line representation from the molecular structure of drugs and gene expression values. For the drug, we first use D-MPNN (Yang *et al.* 2019) to aggregate information from neighboring atoms and bonds, and then the node embeddings are fed into a transformer encoder to obtain the final representation of the drug. For the cell line, we get the representation through a five-layer DNN network.

In the cross-attention module, we stack the drug node feature from D-MPNN and the broadcasting cell line feature. Under the attention mechanism of the transformer, we get another drug representation and another cell line representation. Both of them contain fusion information.

Finally, we average two drug and two cell line representations separately and concatenate result features. After extraction of a three-layer DNN, we can obtain the predicted result.



**Figure 1.** Flowchart of DeepCoVDR. We first use the cancer dataset to pre-train the model and then fine-tune the model on the SARS-CoV-2 dataset. The model consists of three parts: feature representation extraction, cross-attention module, and drug response prediction. The drug's chemical structure as a graph is input into a graph transformer-based network and transformed into a high-level representation. And the gene expression data is used to extract cell line feature representation by a feed-forward neural network. Then, we employ a cross-attention module to calculate the interaction between the drug and cell line and obtain the representation of fusion information. In drug response prediction, we generate the final embeddings of the drug and cell line and take them as input to predict drug response.

We take mean square error as the loss function for the regression task. For the classification task, we add an additional sigmoid layer and use the cross-entropy as the loss function. The optimizer in the back-propagation process is Adam. To reduce the impact of data scarcity on model performance, we adopt transfer learning in the model. Transfer learning can address data scarcity by leveraging existing knowledge from source tasks to low-data target tasks (Pan and Yang 2010). We first use the cancer dataset to pretrain the model. Then we fine-tune the pretrained model with the SARS-CoV-2 dataset and obtain the target model.

### 3.2 Drug information extraction

Drugs can be compiled into molecular graphs. Using the open-source package RDKit (<https://rdkit.org/docs/index.html>), the SMILES string of each drug can be converted into a chemical structure with atom features and bond features. To fully use drug structure information, we construct a graph transformer framework that not only uses D-MPNN (Yang et al. 2019) to learn the local representation of the immediate adjacent structure of nodes but also uses transformer encoder (Vaswani et al. 2017) as the global inference module. As an extension of MPNN (Gilmer et al. 2017), D-MPNN uses messages associated with a directed edge to make use of key attributes and avoid unnecessary loops in the message delivery path (Yang et al. 2019). Taking a drug  $D$  of  $n$  atoms as an example, we first initialized edge hidden states:

$$h_{ab}^0 = f([n_a, e_{ab}]W^i) \quad (1)$$

where  $W^i \in \mathbb{R}^{h_c \times h}$  is a trainable parameter,  $n_a$  is atom features for atom  $a$  and  $e_{ab}$  is bond features for bond  $ab$ , and  $f$  is the ReLU activation function (Nair and Hinton 2010). Then, the message-passing process includes  $t_f$  steps. The messages and the hidden states of step  $t+1$  are respectively  $m_{ab}^{t+1}$  and  $h_{ab}^{t+1}$ :

$$m_{ab}^{t+1} = \sum_{i \in \{N_a, b\}} h_{ia}^t \quad (2)$$

$$h_{ab}^{t+1} = f(h_{ab}^0 + m_{ab}^{t+1}W^m) \quad (3)$$

where  $N_a$  is the set of neighbors of  $a$  in graph  $D$ ,  $W^m \in \mathbb{R}^{h \times h}$  is a trainable parameter. And the final messages and representation of the atom  $a$  are  $m_a$  and  $h_a$ :

$$m_a = \sum_{i \in N_a} h_{ai}^{t_f} \quad (4)$$

$$h_a = f([n_a, m_a]W^o) \quad (5)$$

where  $W^o \in \mathbb{R}^{h \times h}$  is a trainable parameter.

Transformer is a deep learning method with attention mechanism for seq2seq tasks (Vaswani et al. 2017). By using a self-attention module, the transformer can extract syntax information effectively. Herein, we use a transformer encoder to mine pairwise message between atom encodings of drug molecules. After we obtain the final pennode encoding  $h_d \in$

$\mathbb{R}^{n \times h}$  of drug  $D$  from D-MPNN, we input them into a transformer encoder. Transformer mainly includes two parts, a multiattention layer and a feed-forward network. A multiattention layer consists of several self-attention layers which takes the keys, the queries and the values as input to calculate scaled dot-product attention. We first get query  $Q_d$ , key  $K_d$ , and value  $V_d$  by

$$Q_d = h_d W_d^Q, K_d = h_d W_d^K, V_d = h_d W_d^V \quad (6)$$

where  $W_d^Q, W_d^K, W_d^V \in \mathbb{R}^{h \times z}$  are learnable weight matrixes. To allow model consider different subspaces, we run  $u$  parallel self-attention layers and concatenate the output per-head encodings.

$$\begin{aligned} \text{head}_i &= \text{Attention}(Q_d, K_d, V_d) \\ &= \text{softmax}\left(\frac{Q_d K_d^T}{\sqrt{z}}\right) V_d \end{aligned} \quad (7)$$

$$\text{MultiHead}(h_d) = \text{concat}(\text{head}_1, \dots, \text{head}_u) W^T \quad (8)$$

where  $W^T \in \mathbb{R}^{uz \times z}$  is a weight parameter and  $\frac{1}{\sqrt{z}}$  is a scaled factor. To further adjust the representation, the output of the multiattention layer is input into the feed-forward layer:

$$E_1^d = f\left(\text{MultiHead}(h_d) W_1^d + b_1^d\right) W_2^d + b_2^d \quad (9)$$

where  $E_1^d$  is the first drug representation,  $W_1^d \in \mathbb{R}^{z \times o}$ ,  $W_2^d \in \mathbb{R}^{o \times o}$ , and  $b_1^d, b_2^d \in \mathbb{R}^o$  are weight parameters.

### 3.3 Cell line information extraction

For cell lines, we design a five-layer DNN to extract feature representation from gene expression data. Taking cell line  $C$  as an example, the first representation of Cell line  $E_1^c$  is the final output of DNN.

$$h_i^c = f(b_{i-1}^c W_i^c + b_i^c) \quad (10)$$

where  $i \in [1, 5]$  is the number of layers,  $h_i^c$  is the output vector of layer  $i$ ,  $W_i^c \in \mathbb{R}^{d_{i-1} \times d_i}$  is weight parameter.

### 3.4 Cross-attention module

Although we have obtained drug feature representation  $E_1^d$  and cell line feature representation  $E_1^c$ , which contain the information of their own characteristics, their mutual connection is ignored. Inspired by multimodality cross attention network for image and sentence matching (Wei et al. 2020), we present a cross-attention module based on a transformer encoder that allows us to calculate the effect of each atom in the molecule and the corresponding cell line, to generate  $E_2^d$  and  $E_2^c$  which are the second representation of drug and cell line containing fusion information. The core operations are as follows. We first use stacked features of drug and cell line to obtain the query  $Q_F$ , key  $K_F$  and value  $V_F$ :

$$Q_F = \begin{pmatrix} Q_D \\ Q_C \end{pmatrix} = \begin{pmatrix} h_d W_f^Q \\ h_c W_f^Q \end{pmatrix} \quad (11)$$

$$K_F = \begin{pmatrix} K_D \\ K_C \end{pmatrix} = \begin{pmatrix} h_d W_f^K \\ h_c W_f^K \end{pmatrix} \quad (12)$$

$$V_F = \begin{pmatrix} V_D \\ V_C \end{pmatrix} = \begin{pmatrix} h_d W_f^V \\ h_c W_f^V \end{pmatrix} \quad (13)$$

where  $W_f^Q, W_f^K, W_f^V \in \mathbb{R}^{h \times f}$  are learnable weight matrices and  $h_c \in \mathbb{R}^{n \times h}$  is obtain from  $E_1^c$  by broadcasting. Then the output of head  $i$  is calculated as follows:

$$\begin{aligned} \text{head}_i^F &= \text{Attention}(Q_F, K_F, Q_F) \\ &= s\left(\begin{pmatrix} Q_D \\ Q_C \end{pmatrix} \begin{pmatrix} K_D^T & K_C^T \end{pmatrix}\right) \begin{pmatrix} V_D \\ V_C \end{pmatrix} \\ &= \begin{pmatrix} s(Q_D K_D^T) & s(Q_D K_C^T) \\ s(Q_C K_D^T) & s(Q_C K_C^T) \end{pmatrix} \begin{pmatrix} V_D \\ V_C \end{pmatrix} \\ &= \begin{pmatrix} s(Q_D K_D^T) V_D + s(Q_D K_C^T) V_C \\ s(Q_C K_D^T) V_D + s(Q_C K_C^T) V_C \end{pmatrix} \\ &= \begin{pmatrix} E_2^d \\ E_2^c \end{pmatrix} \end{aligned} \quad (14)$$

where  $s$  represent the softmax and scaled function,  $E_2^d$  and  $E_2^c$  are the hidden representation of drug and cell line. From  $E_2^d = s(Q_D K_D^T) V_D + s(Q_D K_C^T) V_C$  and  $E_2^c = s(Q_C K_D^T) V_D + s(Q_C K_C^T) V_C$ , the cross-attention module is able to consider the interaction of drug feature and cell line feature for the representation of drug and cell line. Similar to formulas 9 and 10, we can get the output of the transformer encoder based on multiattention module  $(E_2^d E_2^c)^T$ .  $E_2^d$  and  $E_2^c$  are the second representation of drug and cell line.

### 3.5 Baseline methods

To test the performance of DeepCoVDR, we compared our methods with six baselines: support vector machine (SVM) (Hearst et al. 1998), a random forest (RF) (Breiman 2001), XGBoost (Chen and Guestrin 2016), feed-forward neural network (DNN), graph convolutional network (GCN) Kipf and Welling (2016), and graph attention network (GAT) (Veličković et al. 2017). The input feature of SVM, RF, XGBoost, and DNN are the concatenation of Morgan fingerprints (Morgan 1965) and gene expression data. GCN and GAT can process graph structure data as graph neural network methods, so their inputs are molecular structures that contain atom features and bond features and gene expression data. The two methods use the same way as DeepCoVDR to process gene expression data. Their main difference lies in the treatment of drug structure. The parameters of these baselines are default or the best. The details are as follows. For SVM, RF and XGBoost, we choose the default hyperparameters from sklearn library (Pedregosa et al. 2011). And the hyperparameters of DNN, GCN and GAT are selected by experimental results. For DNN, we set three hidden layers and the dimension of hidden layers are [1024, 512, 128]. For GCN, we use three GCN layers and the dimension are [78, 156, 312]. For GAT, we set six attention heads and two GAT layers which dimension are [128, 128]. The optimizer of DNN, GCN and GAT is Adam.



## 4 Results

### 4.1 Experimental settings and model evaluation

We conduct five 5-fold cross-validations by five different random seeds to evaluate the performance of DeepCoVDR and the competitive methods. The final experimental results come from the average of five 5-fold cross-validations.

For the model evaluation of regression experiment, we use three metrics: Pearson's correlation coefficient (PCC) (Benesty et al. 2009), Spearman's correlation coefficient (SCC) (Myers and Sirois 2004) and root mean squared error (RMSE). PCC is a statistic reflecting the linear correlation degree of observed and predicted logarithm IC50. SCC is a nonparametric measure of ranked values and uses a monotone function to describe the relationship between two variables observed and predicted logarithm IC50. And the RMSE is a traditional measure to calculate the level of accuracy.

For the model evaluation of the classification experiment, we use four common metrics: the area under the receiver operating characteristic curve (AUC), the area under the precision-recall curve (AUPR), accuracy (ACC), and F1 score. The F1 score is a harmonic average of accuracy and recall.

### 4.2 Performance comparison on regression experiments

The performance comparison of DeepCoVDR on regression experiment is shown in Table 1. DeepTTA is the state-of-art transformer-based method to predict cancer drug response accurately. To further prove the high performance of DeepCoVDR, we train DeepTTA on our SARS-CoV-2 dataset and add it to the comparison on the regression experiment. From the comparison results, we can have the following analysis: (i) Among the eight methods, DeepCoVDR outperforms other methods, including DeepTTA, by achieving Pearson's correlation of 0.942, Spearman's correlation of 0.881, and RMSE of 0.509. Fig. 2 shows the *P*-values of PCC and SCC between each method and other methods. The *P*-value between DeepCoVDR and other methods is less than .05, which proves the credibility of our experimental conclusion and that the improvements of DeepCoVDR have statistical significance; (ii) The methods based on deep learning, including DNN, GCN, GAT, DeepTTA, and DeepCoVDR, perform better than the methods based on traditional machine learning. That shows deep learning has the ability to discover complex information in high-dimensional data. (iii) The methods that only use fingerprints as a drug feature perform worse than the methods that use chemical structures. That shows chemical structures contain more detailed and higher-level drug information.

For runtime analysis, we benchmarked DeepCoVDR and DeepTTA on the same server with a single RTX3090 GPU.

**Table 1.** Performance comparison of DeepCoVDR and baseline methods on regression experiments.

Methods	PCC	SCC	RMSE
SVM	0.470 ± 0.039	0.438 ± 0.045	1.616 ± 0.008
RF	0.812 ± 0.036	0.730 ± 0.033	0.921 ± 0.061
XGB	0.847 ± 0.032	0.783 ± 0.028	0.826 ± 0.087
DNN	0.887 ± 0.018	0.798 ± 0.027	0.843 ± 0.040
GCN	0.878 ± 0.025	0.805 ± 0.031	0.626 ± 0.063
GAT	0.829 ± 0.024	0.733 ± 0.022	0.926 ± 0.219
DeepTTA	0.910 ± 0.018	0.858 ± 0.024	0.588 ± 0.075
DeepCoVDR	<b>0.942 ± 0.011</b>	<b>0.881 ± 0.017</b>	<b>0.509 ± 0.031</b>

The training costs of DeepCoVDR and DeepTTA are 3.40 h and 1.72 h, respectively. It is reasonable that DeepCoVDR requires more training time because it utilizes a more complex deep transfer learning approach to improve regression performance.

### 4.3 Performance comparison on classification experiments

In this study, we binarize the IC50 values through the threshold 20  $\mu$ M (Zaliani et al. 2022). In order to prove the classification performance of our model, we use t-SNE (Van der Maaten and Hinton 2008) to visualize the classification results of DeepCoVDR. As shown in Fig. 3a–c, the green and blue dots respectively represent the drugs with and without antiviral activity in our dataset. It can be clearly seen that with the increase in the number of training epochs, the dots representing different types of samples are gradually distinguished, and the dots of the same type are gradually gathered. This represents that our model can extract and distinguish hidden knowledge in features. Then, we use AUC, AUPR, ACC, and F1 scores to measure the model's performance. Figure 3d shows the ROC curves of DeepCoVDR and other baseline methods. More results are shown in Fig. 3e. DeepCoVDR outperforms other baseline methods by achieving significantly higher AUC, AUPR, ACC, and F1 scores of 0.954, 0.956, 0.962, and 0.946. From the above results, we can see that DeepCoVDR has reached the state-of-art of all comparison methods, demonstrating the superiority of DeepCoVDR when performing classification tasks.

### 4.4 Drug response prediction on the cancer dataset

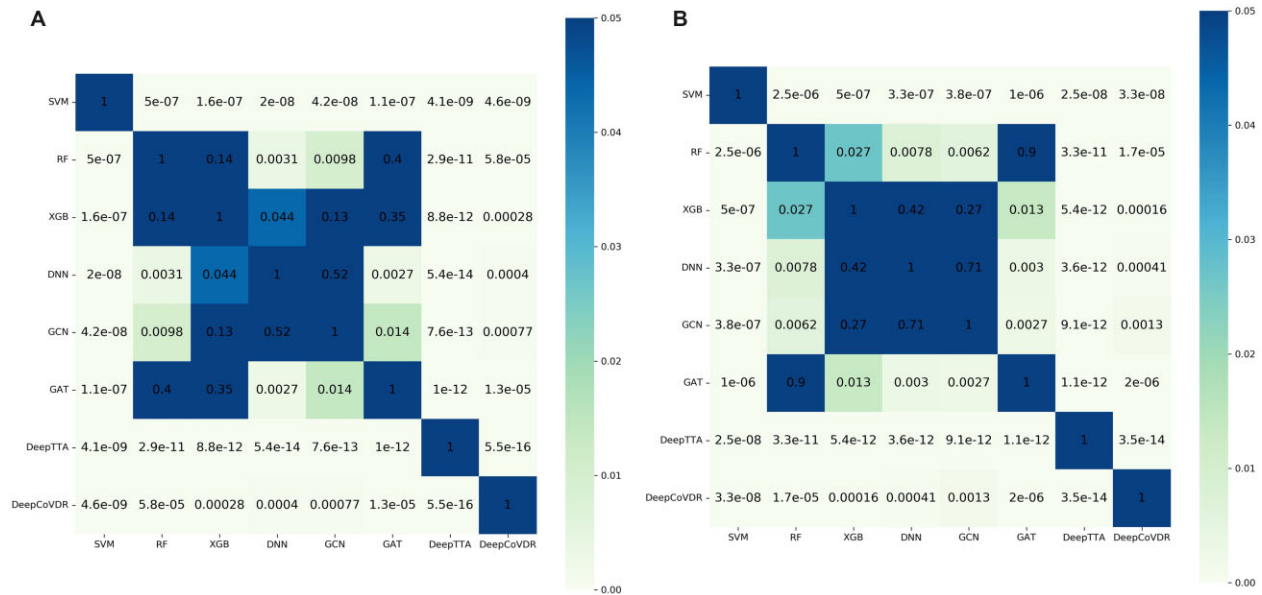
Before we use the SARS-CoV-2 dataset to predict the drug response, we first train DeepCoVDR on the cancer dataset. To demonstrate the high performance of DeepCoVDR, we compare the performance of our model with some advanced cancer drug response models on the cancer dataset. As shown in Table 2, DeepCoVDR achieves state-of-the-art performance on PCC and SCC among the existing models where DeepTTA is the state-of-art transformer-based method. Specifically, the highest PCC 0.946 and SCC 0.935 proves a strong agreement between observed and predicted IC50, and the high performance on the cancer dataset shows our model has strong generalization ability.

### 4.5 Performance comparison on drug repurposing

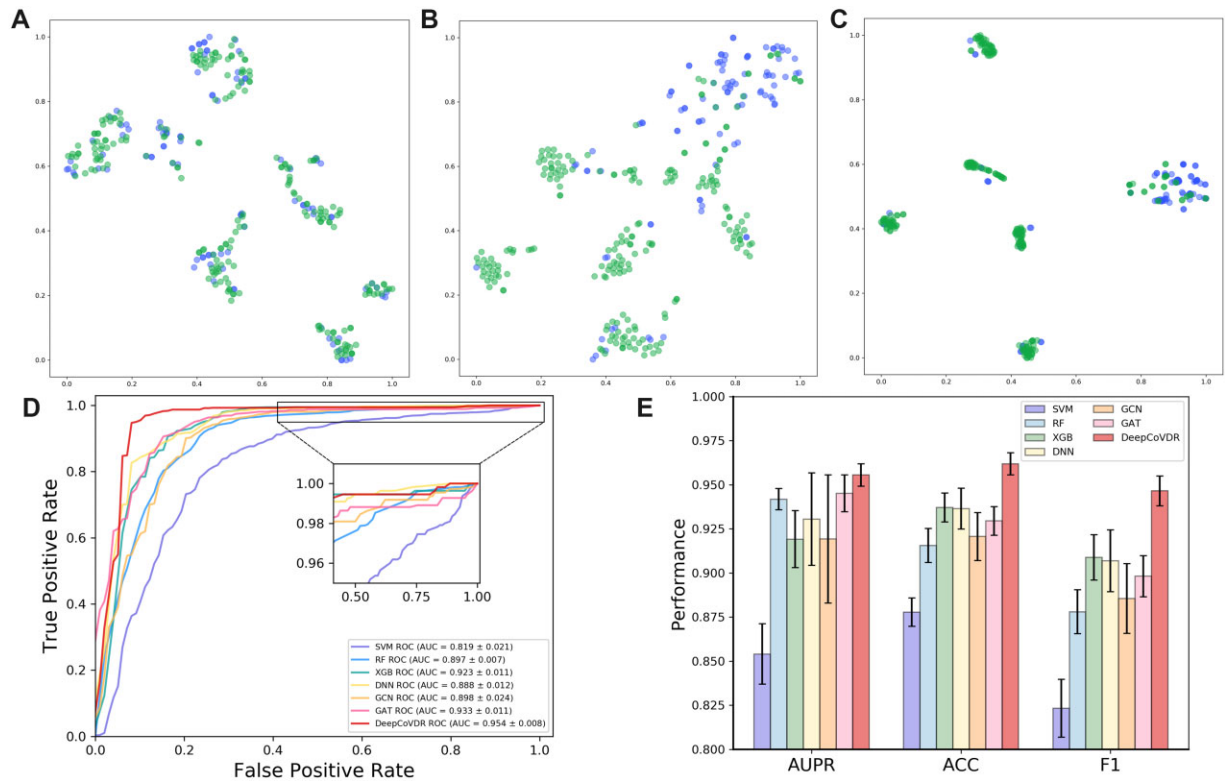
To test DeepCoVDR's ability of repurposing, we compare DeepCoVDR with COVIDVS-1 and COVIDVS-2 on the same ReFRAME actives dataset (Wang et al. 2021a) which contains 17 identified ReFRAME (Riva et al. 2020) actives. There is no overlap between the ReFRAME actives dataset and our SARS-CoV-2 dataset. The result is shown in Fig. 4. Among 17 compounds, 15 got predicted scores more than 0.8 by DeepCoVDR, approximately 82% of all ReFRAME actives. While 6 got predicted scores greater than 0.8 by COVIDVS-2, and no compounds got predicted scores greater than 0.8 by COVIDVS-1. The results demonstrate that our method has a strong ability to screen anti-SARS-CoV-2 drugs.

### 4.6 Ablation study

To further investigate the importance of components, transfer learning, graph transformer framework, gene expression, and the cross-attention module, we design the following variants of DeepCoVDR:



**Figure 2.** Heat maps of  $P$ -value obtained by performing  $t$ -test on the results of regression experiment. (a) is for Pearson's correlation and (B) is for Spearman's correlation.



**Figure 3.** Performance of DeepCoVDR in the classification setting. (a–c) are visualization of the final embeddings of DeepCoVDR, and they are from epochs 20, 100, and 1000 respectively. The green and blue dots respectively represent the drugs with and without antiviral activity in our dataset. (d) shows the receiver operating characteristic (ROC) curve of the seven comparing methods. (e) shows the AUPR, ACC, and F1 scores of DeepCoVDR and baseline methods.

- **NoTransfer** remove transfer learning.
- **NoGraphTransformer** use transformer instead of graph transformer.
- **NoGene** remove the gene expression feature.
- **NoCrossAttention** remove cross-attention module.

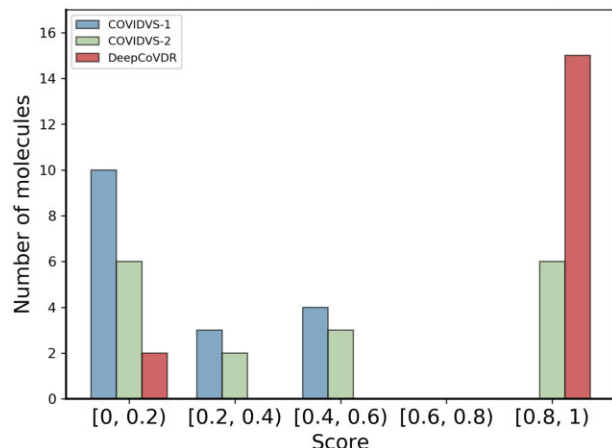
The results are shown in Table 3. The experiment on NoTransfer shows that transfer learning can successfully

transfer the knowledge of drugs and gene expression values from cancer to predict COVID-19 drug response. The performance of NoGraphTransformer shows the importance of graph transformers in drug information mining. When we remove gene expression values, the degradation of model performance indicates the information on the cell line is useful for prediction. And the experiment on NoCrossAttention demonstrates the cross-attention module is effective. The training time of

**Table 2.** Performance comparison of DeepCoVDR and advanced methods on cancer dataset.

Methods	PCC	SCC
MOLI (Sharifi-Noghabi et al. 2019)	0.813 ± 0.007	0.782 ± 0.005
CDRscan (Chang et al. 2018)	0.871 ± 0.004	0.852 ± 0.003
tCNNs (Liu et al. 2019)	0.910 ± 0.009	0.889 ± 0.008
DeepCDR (Liu et al. 2020)	0.923 ± 0.005	0.898 ± 0.008
DeepTTA	0.941 ± 0.003	0.914 ± 0.004
DeepCoVDR	<b>0.946 ± 0.001</b>	<b>0.935 ± 0.005</b>

Note: The results of the comparison method are from DeepTTA.

**Figure 4.** Distribution of scores for 17 ReFRAME actives predicted with COVIDVS-1, COVIDVS-2, and DeepCoVDR. The results of COVIDVS-1, COVIDVS-2 are from Wang et al. (2021a).**Table 3.** Ablation study.

Methods	PCC	SCC
NoTransfer	0.910 (↓ 3.40%)	0.873 (↓ 0.91%)
NoGraphTransformer	0.912 (↓ 3.18%)	0.861 (↓ 2.27%)
NoGene	0.922 (↓ 2.12%)	0.853 (↓ 3.18%)
NoCrossAttention	0.925 (↓ 1.80%)	0.853 (↓ 3.18%)
DeepCoVDR	<b>0.942</b>	<b>0.881</b>

Note: The content in parentheses shows the performance delta percentage of each variant of DeepCoVDR.

NoTransfer, NoGraphTransformer, NoGene, and NoCrossAttention is 3.24 h, 1.31 h, 2.79 h, and 2.29 h, respectively. Although the training time of DeepCoVDR (3.40 h) is slightly longer, the performance is significantly improved.

#### 4.7 Screening FDA-approved drugs to identify novel anti-SARS-CoV-2 drugs

Finding potential targets from existing drugs is still an important way of finding a drug that can treat SARS-CoV-2. We applied DeepCoVDR to screen the 1871 FDA-approved drugs from (Skuta et al. 2017). We removed the drugs that are present in our SARS-CoV-2 dataset, and finally, 1853 drugs are left. To improve the performance, we integrated five models through ensemble technique (Dietterich 2000) technology and used the ensemble model for screening. To verify the

**Table 4.** Drugs predicted by DeepCoVDR in FDA-approved drugs with the support of literature.

Compound name	Status	Ranking position	PMID
Orlistat	Preclinical	6	34580494
Ritonavir	Phase 3	8	33243253
Artesunate	Preclinical	17	34272426
Folic acid	Phase 3	23	36060767
Artemether	Preclinical	28	34527599
Valproic acid	Phase 4	31	34635175
Adefovir dipivoxil	Preclinical	34	34440200
Remdesivir	Approved (FDA)	41	32366720
Lopinavir	Phase 4	61	33290767
Diltiazem	Preclinical	76	35176124
Bortezomib	Preclinical	96	36335206

reliability of the prediction results of DeepCoVDR, we conducted a nonexhaustive quality search on the top 100 compounds and found that at least 11 compounds have been proven effective against SARS-CoV-2 in previous studies (Table 4).

We select the following examples for detailed descriptions. Orlistat, a US FDA-approved drug, is used for the treatment of obesity by preventing the absorption of dietary fat and fatty acid synthase to reduce lipid synthesis (Kridel et al. 2004), which also significantly inhibits the infectivity of the SARS-CoV-2 virus with an EC<sub>50</sub> of 0.39 μM in vitro (Chu et al. 2021). Remdesivir is a broad-spectrum antiviral compound and was approved for the treatment of COVID-19 in 2020. With strong and stable anti-SARS-CoV-2 activity, remdesivir was used as a reference drug in several Vitro studies with IC<sub>50</sub> of ~10 μM (Jeon et al. 2020; Jang et al. 2021). Lopinavir, ranked at position 61, is an antiretroviral protease inhibitor against HIV infections, and its antiviral effects on SARS-CoV-2 are found to have IC<sub>50</sub> of 13.16 μM in vitro (Raj et al. 2021). Ritonavir, ranked at position 8, is an HIV-1 protease inhibitor and CYP3A inhibitor used in combination with other antivirals to treat HIV infection and shows some ability in the inhibition of SARS-CoV-2 M<sup>pro</sup> with IC<sub>50</sub> of 13.7 nM (Mahdi et al. 2020). Notably, Ritonavir may be regarded as a fixed-dose combination product with other drugs to treat COVID-19, such as lopinavir or nirmatrelvir. Nirmatrelvir/ritonavir, commercially named Paxlovid, was granted emergency use authorization by the US FDA for the treatment of COVID-19 in December 2021. Recent research shows that symptomatic COVID-19 patients treated with nirmatrelvir plus ritonavir, have a lower risk of progression to severe COVID-19 compared with placebo (Hammond et al. 2022). However, despite the antiviral effects on SARS-CoV-2 of lopinavir and ritonavir in vitro, a randomized trial found lopinavir/ritonavir was not effective in treating COVID-19 patients in 2020 (Cao et al. 2020). Therefore, more effort is still needed to discover other effective antiviral medications. And above all, our model can identify certain potential candidates with significant antiviral activity in vitro.

## 5 Conclusion and discussion

In this article, we propose a novel deep learning model to predict COVID-19 drug response based on graph transformer, cross-attention, and transfer learning. To the best of our knowledge, DeepCoVDR is the first computing method that can accurately predict COVID-19 drug response value, and it is also the first work to apply graph transformer and cross-attention in a drug response prediction tasks. Our model



achieves state-of-art performance in all metrics, according to the comparison experiment results of regression and classification tasks, compared with baseline methods. DeepCoVDR also achieves high performance in cancer drug response prediction, showing its high versatility. In addition, we demonstrate DeepCoVDR can repurpose existing drugs for the treatment of SARS-CoV-2 through screening FDA-approved drugs to identify novel anti-SARS-CoV-2 drugs.

In the future, we will explore more multisource heterogeneous information for COVID-19 drug response prediction, e.g. the priority of gene expression value in cell lines that is significant in revealing the potential therapeutic targets of anti-SARS-CoV-2 drugs.

## Acknowledgements

We are grateful for resources from the High Performance Computing Center of Central South University.

## Conflict of interest

The authors declare that they do not have any conflict of interest.

## Funding

This work was supported by the National Natural Science Foundation of China under grants Nos. 62272490 and 61972422.

## References

- Avorn J. The \$2.6 billion pill-methodologic and policy considerations. *N Engl J Med* 2015;372:1877–9.
- Benesty J, Chen J, Huang Y *et al.* Pearson correlation coefficient. In: Benesty J. and Kellermann W. (eds.), *Noise Reduction in Speech Processing*. Heidelberg: Springer, 2009, 1–4.
- Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- Cao B, Wang Y, Wen D *et al.* A trial of lopinavir–ritonavir in adults hospitalized with severe COVID-19. *N Engl J Med* 2020;382:1787–99.
- Cha Y, Erez T, Reynolds IJ *et al.* Drug repurposing from the perspective of pharmaceutical companies. *Br J Pharmacol* 2018;175:168–80.
- Chang Y, Park H, Yang H-J *et al.* Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep*, 2018;8:1–11.
- Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. ACM, New York, NY, USA, pp.785–794, 2016.
- Chu J, Xing C, Du Y *et al.* Pharmacological inhibition of fatty acid synthesis blocks SARS-CoV-2 replication. *Nat Metab* 2021;3:1466–75.
- Daly AK. Pharmacogenetics: a general review on progress to date. *Br Med Bull* 2017;124:65–79.
- Deepthi K, Jereesh A S, Liu Y *et al.* A deep learning ensemble approach to prioritize antiviral drugs against novel coronavirus SARS-CoV-2 for covid-19 drug repurposing. *Appl Soft Comput* 2021;113:107945.
- Dietterich TG. Ensemble methods in machine learning. In: Kittler J. and Roli F. (eds.), *International Workshop on Multiple Classifier Systems*. Heidelberg: Springer, 2000, 1–15.
- Gilmer J, Schoenholz S S, Riley P F *et al.* Neural message passing for quantum chemistry. In: *International Conference on Machine Learning*, pp.1263–72. PMLR, 2017.
- Gordon DE, Jang GM, Bouhaddou M *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;583:459–68.
- Guan W-J, Ni Z-Y, Hu Y *et al.*; China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;382:1708–20.
- Hammond J, Leister-Tebbe H, Gardner A *et al.*; EPIC-HR Investigators. Oral nirmatrelvir for high-risk, nonhospitalized adults with covid-19. *N Engl J Med* 2022;386:1397–408.
- Hearst MA, Dumais ST, Osuna E *et al.* Support vector machines. *IEEE Intell Syst Their Appl* 1998;13:18–28.
- Jang WD, Jeon S, Kim S *et al.* Drugs repurposed for covid-19 by virtual screening of 6,218 drugs and cell-based assay. *Proc Natl Acad Sci USA* 2021;118:e2024302118.
- Jayk Bernal A, Gomes da Silva MM, Musungaie DB *et al.*; MOVE-OUT Study Group. Molnupiravir for oral treatment of covid-19 in nonhospitalized patients. *N Engl J Med* 2022;386:509–20.
- Jeon S, Ko M, Lee J *et al.* Identification of antiviral drug candidates against SARS-CoV-2 from fda-approved drugs. *Antimicrob Agents Chemother* 2020;64:e00819–20.
- Jiang L, Jiang C, Yu X *et al.* Deeptta: a transformer-based model for predicting cancer drug response. *Brief Bioinf* 2022;23:bbac100.
- Kim S, Chen J, Cheng T *et al.* Pubchem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;47:D1102–D1109.
- Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *5th International Conference on Learning Representations (ICLR)*, 2017.
- Kridel SJ, Axelrod F, Rozenkrantz N *et al.* Orlistat is a novel inhibitor of fatty acid synthase with antitumor activity. *Cancer Res* 2004;64:2070–5.
- Kubinyi H. Similarity and dissimilarity: a medicinal chemist's view. *Perspect Drug Disc Des* 1998;9:225–52.
- Liu P, Li H, Li S *et al.* Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinf*, 2019;20:1–14.
- Liu Q, Hu Z, Jiang R *et al.* Deepcdr: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 2020;36:i911–8.
- Mahdi M, Mótýán JA, Szojka ZI *et al.* Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease. *Virol J* 2020;17:1–8.
- Meng Y, Jin M, Tang X *et al.* Drug repositioning based on similarity constrained probabilistic matrix factorization: covid-19 as a case study. *Appl Soft Comput* 2021;103:107135.
- Morgan HL. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *J Chem Doc* 1965;5:107–13.
- Myers L, Sirois MJ. Spearman correlation coefficients, differences between. *Encycl Statist Sci* 2004;12:1–2.
- Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: *Proc. 27th International Conference on Machine Learning*, pp. 807–814, 2010.
- Owen DR, Allerton CMN, Anderson AS *et al.* An oral SARS-CoV-2 mpro inhibitor clinical candidate for the treatment of covid-19. *Science* 2021;374:1586–93.
- Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22:1345–59.
- Pedregosa F *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- Peng L, Shen L, Xu J *et al.* Prioritizing antiviral drugs against SARS-CoV-2 by integrating viral complete genome sequences and drug chemical structures. *Sci Rep* 2021;11:1–11.
- Raj V, Park JG, Cho K-H *et al.* Assessment of antiviral potencies of cannabinoids against SARS-CoV-2 using computational and in vitro approaches. *Int J Biol Macromol* 2021;168:474–85.
- Riva L, Yuan S, Yin X *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* 2020;586:113–9.
- Sharif-Askari NS, Sharif-Askari FS, Mdkhana B *et al.* Enhanced expression of immune checkpoint receptors during SARS-CoV-2 viral infection. *Mol Ther Methods Clin Dev* 2021;20:109–21.



- Sharifi-Noghabi H, Zolotareva O, Collins CC *et al.* Moli: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 2019;**35**:i501–9.
- Skuta C, Popr M, Muller T *et al.* Probes & drugs portal: an interactive, open data resource for chemical biology. *Nat Methods* 2017;**14**:759–60.
- Spinner CD, Gottlieb RL, Criner GJ *et al.*; GS-US-540-5774 Investigators. Effect of remdesivir vs standard care on clinical status at 11 days in patients with moderate covid-19: a randomized clinical trial. *JAMA* 2020;**324**:1048–57.
- Van der Maaten L, Hinton G. Visualizing data using t-sne. *J Mach Learn Res* 9(Nov) 2008:2579–2605.
- Vaswani A, Shazeer N, Parmar N *et al.* Attention is all you need. *Adv Neural Inf Process Syst* 2017;**30**:5998–6008.
- Velicković P, Curull G, Casanova A *et al.* Graph attention networks. In: *6th International Conference on Learning Representations*, 2018.
- Wang S, Sun Q, Xu Y *et al.* A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2. *Brief Bioinf* 2021a;**22**:bbab211.
- Wang Y, Li P, Solanki K *et al.* Viral polymerase binding and broad-spectrum antiviral activity of molnupiravir against human seasonal coronaviruses. *Virology* 2021b;**564**:33–8.
- Wei X, Zhang P, Solanki K *et al.* Multi-modality cross attention network for image and sentence matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.10941–50, 2020.
- Willett P. The calculation of molecular structural similarity: principles and practice. *Mol Inform* 2014;**33**:403–13.
- Yang K, Swanson K, Jin W *et al.* Analyzing learned molecular representations for property prediction. *J Chem Inf Model* 2019;**59**:3370–88.
- Yang W, Soares J, Greninger P *et al.* Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;**41**:D955–D961.
- Yin W, Mao C, Luan X *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 2020;**368**:1499–504.
- Zaliani A, Vangeel L, Reinshagen J *et al.* Cytopathic SARS-CoV-2 screening on vero-e6 cells in a large-scale repurposing effort. *Sci Data* 2022;**9**:1–11.
- Zhou Y, Hou Y, Shen J *et al.* Network-based drug repurposing for novel coronavirus 2019-NCOV SARS-CoV-2. *Cell Discov* 2020;**6**:1–18.