Briefings in Bioinformatics, 2025, **26(2)**, bbaf117
https://doi.org/10.1093/bib/bbaf117
Problem Solving Protocol

GiGs: graph-based integrated Gaussian kernel similarity for virus–drug association prediction

Yixuan Jin[1,‡], Juanjuan Huang[1,2,‡], Xu Sun[1], Yabo Fang[1], Jiageng Wu[1], Jianshi Du[3,*], Jiwei Jia[1,4,*], Guoqing Wang[2,*]

[1]Department of Computational Mathematics, School of Mathematics, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China
[2]State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Key Laboratory for Zoonosis Research of the Ministry of Education, College of Basic Medicine, Jilin University, No. 126 Xinmin Street, Changchun 130021, China
[3]China-Japan Union Hospital of Jilin University, Key Laboratory and Engineering Laboratory of Lymphatic Surgery Jilin Province, No. 126 Sendai Street, Changchun 130033, China
[4]Jilin National Applied Mathematical Center, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China

[*]Corresponding authors. Guoqing Wang, State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Key Laboratory for Zoonosis Research of the Ministry of Education, College of Basic Medicine, Jilin University, No. 126 Xinmin Street, Changchun 130021, China. E-mail: qing@jlu.edu.cn; Jiwei Jia, Department of Computational Mathematics, School of Mathematics, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China; Jilin National Applied Mathematical Center, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China. E-mail: jiajiwei@jlu.edu.cn; Jianshi Du, Key Laboratory and Engineering Laboratory of Lymphatic Surgery Jilin Province, China-Japan Union Hospital of Jilin University, No. 126 Sendai Street, Changchun 130033, China. E-mail: dujs@jlu.edu.cn

[‡]Yixuan Jin and Juanjuan Huang contributed equally to this work.

Abstract

The prediction of virus–drug associations (VDAs) is crucial for drug repositioning, contributing to the identification of latent antiviral drugs. In this study, we developed a graph-based integrated Gaussian kernel similarity (GiGs) method for predicting potential VDAs in drug repositioning. The GiGs model comprises three components: (i) collection of experimentally validated VDA information and calculation virus sequence, drug chemical structure, and drug side effect similarity; (ii) integration of viruses and drugs similarity based on the above information and Gaussian interaction profile kernel (GIPK); and (iii) utilization of similarity-constrained weight graph normalization matrix factorization to predict antiviral drugs. The GiGs model enhances correlation matrix quality through the integration of multiple biological data, improves performance via similarity constraints, and prevents overfitting and predicts missing data more accurately through graph regularization. Extensive experimental results indicated that the GiGs model outperforms five other advanced association prediction methods. A case study identified broad-spectrum drugs for treating highly pathogenic human coronavirus infections, with molecular docking experiments confirming the model's accuracy.

Keywords: VDAs; graph regularization; matrix factorization; drug repositioning; antiviral drugs

Introduction

Drug repositioning is the treatment of new diseases using existing drugs, with the advantages of low cost and short development cycles [1–4]. However, the traditional wet-laboratory experiments remain a bottleneck due to their high expense and being time-consuming, and computational methods provide a new testable hypothesis for drug repositioning [5, 6]. The prediction of biological entity association, such as drug–target interaction, microbe–drug association, and drug–disease association prediction, has become a critical component of drug repositioning. In recent years, various computational methods have been developed to predict these problems, including Multi-TransDTI, DeepMPF, MDSVDNV, and AMDGT [7–10].

Moreover, the determination of virus–drug associations (VDAs) is crucial for drug repositioning and can help identify potential antiviral drugs. At present, some computational models have been developed to predict VDAs. Zhou *et al*. developed the VDA-KATZ method with the calculation of the number of connecting paths between given virus and drug, yet it may not fully encapsulate the intricate biological relationships [11]. Meng *et al*. developed the similarity constrained probabilistic matrix factorization (SCPMF) to identify new VDAs, with the introduction of similarity data as constraints for drugs and viruses, potentially limiting its applicability to novel or rare VDAs due to data scarcity [12]. Xu *et al*. put forward the weight regularization matrix factorization (WRMF) for the predictive identification of drugs that are effective against given viruses, while the VDAs data may be incomplete [13]. Qu *et al*. presented the MHBVDA, which integrates the Matrix Decomposition with Heterogeneous Graph Inference (MDHGI) and Bounded Nuclear Norm Regularization (BNNR), to identify potential VDAs, but due to the possibility of not being the optimal model parameters, it may lead to biased results [14]. Su *et al*. presented a VDA-DLCMNMF model for identifying drugs to treat new viruses but might have high computational complexity [15]. Given these limitations, it is particularly important to develop an effective and accurate model to predict potential VDAs.

Received: November 11, 2024. **Revised:** February 10, 2025. **Accepted:** February 27, 2025

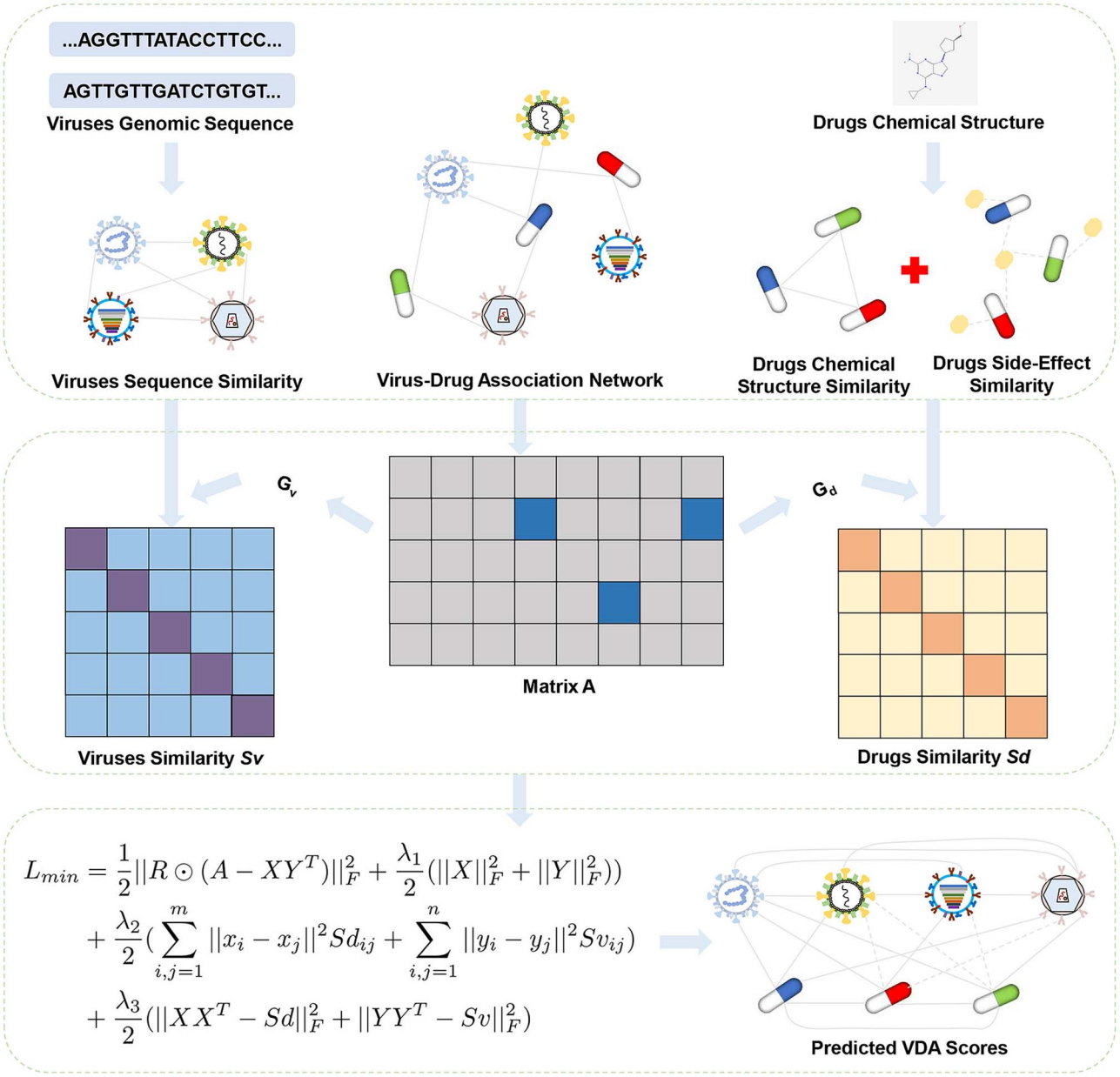

Figure 1. The network framework of GiGs.

In this work, we developed the graph-based integrated Gaussian kernel similarity (GiGs) method for predicting potential VDAs (Fig. 1). With this method, VDAs information is collected and virus sequence similarities, drug chemical structures, and side effect similarities are calculated. Then, comprehensive integration is conducted based on this information and a Gaussian interaction profile kernel (GIPK). We used the GiGs model to predict drugs for the potential treatment of viral infections. The GiGs model has three advantages: (i) it improves correlation matrix quality by integrating diverse biological information data; (ii) it employs similarity constraints that reflect the essence of VDAs, thereby enhancing predictive performance; and (iii) it introduces graph regularization to prevent overfitting and to accurately predict missing data, thus providing a robust framework for VDAs prediction. When applied to the DrugVirus dataset, the GiGs model showed superior performance compared to five other advanced association prediction models. Its applicability was further verified through VDA2 and HDVD datasets. In addition, we conducted a case study for broad-spectrum anticoronavirus therapy to demonstrate the reliability of GiGs, which led to the predictive identification of mefloquine, anisomycin, camostat, hydroxychloroquine, and chlorpromazine. This study presents a novel method for the prediction of potential VDAs to accelerate drug repositioning.

Materials and methods

Materials

The performance of GiGs was assessed by using three various datasets. The DrugVirus dataset is provided by Long *et al.*, including 95 viruses, 175 drugs, and 933 VDAs [16]. The VDA2 dataset includes 69 viruses, 128 drugs, and 770 VDAs [17]. The HDVD dataset is a human VDA database supported by experiments, including 34 viruses, 219 drugs, and 455 VDAs [12].

The VDA network matrix is defined as A ($m \times n$), where variables m and n represent the number of drugs and viruses,

respectively. If drug d_i is associated with virus v_j, then A (i, j) is 1, or it is 0.

Viruses' genomic sequence similarity

MAFFT provides different strategies including progressive alignment, iterative refinement, and structural alignment methods [18]. Thus, it is used to compute the similarity between virus sequences. A matrix K_v is utilized to represent the similarity between virus sequences, where K_v (v_i, v_j) indicates the degree of sequence similarity between virus v_i and v_j.

Drugs' chemical structure similarity

We used SIMCOMP to compute the similarity of drugs' chemical structure [19]. Firstly, we downloaded the MOL files of the drugs from the Kyoto Encyclopedia of Genes and Genomes (KEGG) DRUG database and then imported these MOL files into SIMCOMP, which can calculate the global similarity of two drugs. Construct a matrix K_{d1} to store the similarity between drugs structure, where K_{d1} (i, j) represents the degree of structure similarity between drug d_i and d_j.

Drugs' side effect similarity

We obtained drugs' side effect data for the model from SIDER database. M (i) represents side effects between drug d_i and other drugs. A matrix K_{d2} stores side effects similarity, where K_{d2} (i, j) is the degree of side effects similarity between drug d_i and d_j. Jaccard similarity is an indicator used to compare the degree of similarity between two sets, typically used to measure the similarity of intersection and union between two sets. Here, the Jaccard score computes drugs' side effect similarity using following equation:

$$K_{d2}(i,j) = \frac{|M(i) \cap M(j)|}{|M(i) \cup M(j)|}$$

GIPK similarity of viruses and drugs

We constructed matrix G_v and G_d in our model to further calculate the degree of viruses and drugs' similarity based on GIPK. For the VDA matrix A, AV (v_j) is the j-th column vector of A and AV (d_i) is the i-th row vector of A. G_v (v_i, v_j) for viruses and G_d (d_i, d_j) for drugs can be calculated using the following equations, respectively:

$$G_v(v_i, v_j) = \exp\left(-\gamma_v \left\|AV(v_i) - AV(v_j)\right\|^2\right)$$

$$G_d(d_i, d_j) = \exp\left(-\gamma_d \left\|AV(d_i) - AV(d_j)\right\|^2\right)$$

Parameters γ_v and γ_d are calculated using the following formulas:

$$\gamma_v = \gamma_v' \Big/ \left(\frac{1}{n}\sum_{i=1}^{n} \left\|AV(v_i)\right\|^2\right)$$
$$\gamma_d = \gamma_d' \Big/ \left(\frac{1}{m}\sum_{i=1}^{m} \left\|AV(d_i)\right\|^2\right)$$

$\|\cdot\|^2$ represents L2-norm, γ'_v and γ'_d are 1.

Comprehensive viruses' and drugs' similarity

We compute the final viruses' similarity matrix Sv by integrating G_v and K_v by the following equation:

$$Sv = \omega_1 G_v + (1 - \omega_1) K_v$$

where the parameter ω_1 balances the importance between G_v and K_v.

Similarly, we compute the final drugs' similarity matrix Sd by integrating K_{d1}, $K_{d2,}$ and G_d:

$$Sd = \omega_2 G_d + (1 - \omega_2)\frac{K_{d1} + K_{d2}}{2}$$

where the parameter ω_2 balances the importance among K_{d1}, $K_{d2,}$ and G_d.

GiGs

The VDAs prediction problem was considered as a matrix completion problem, predicting unknown interaction terms. Firstly, we calculate the similarity of virus sequences, collect experimentally validated VDAs information, and calculate the similarity of drug chemical structures and side effects. Then, based on the above information and GIPK, a comprehensive integration is performed to obtain the similarity matrix of the virus and the drug. Finally, combining similarity information, the graph regularization matrix factorization method is used to predict missing values in the correlation matrix.

We decompose the VDA matrix $A \in R^{m\times n}$ into $X \in R^{m\times k}$ and $Y \in R^{n\times k}$, $k \in N$ is the dimensionality of the latent feature vector, and the matrix factorization minimization problem is denoted:

$$L_{min} = \frac{1}{2}\left\|A - XY^T\right\|_F^2 + \frac{\lambda}{2}\left(\|X\|_F^2 + \|Y\|_F^2\right)$$

where $\|\cdot\|_F$ is the F-norm, and λ is the regularization parameter.

In the VDA database, we can obtain the correlation between viruses and drugs. Due to the fact that there are only positive samples, we propose a similarity-constrained weight graph regularization matrix factorization model for problems with only positive feedback to prevent overfitting and improve generalization ability. The algorithm form is as follows:

$$R = \begin{cases} 1+\beta & A_{ij} = 1 \\ 1 & A_{ij} = 0 \end{cases}$$

$$L_{min} = \tfrac{1}{2}\left\|R \odot \left(A - XY^T\right)\right\|_F^2 + \tfrac{\lambda_1}{2}\left(\|X\|_F^2 + \|Y\|_F^2\right) + \tfrac{\lambda_2}{2}\left(\sum_{i,j=1}^{m}\|x_i - x_j\|^2 Sd_{ij} + \sum_{i,j=1}^{n}\|y_i - y_j\|^2 Sv_{ij}\right) + \tfrac{\lambda_3}{2}\left(\left\|XX^T - Sd\right\|_F^2 + \left\|YY^T - Sv\right\|_F^2\right).$$

Assign weight R to the training samples; the parameter β regulates the impact of positive samples in training, where λ_1, λ_2, and λ_3 are regularization parameters, and x_i, y_j represent the i-th and j-th rows vector of matrices X and Y.

We used the gradient descent algorithm to train a model's parameter, and the iterative steps are as follows:

$$x_i^{new} = x_i \frac{(R\cdot(AY)) + 2(\lambda_2 + \lambda_3)\left((Sd)X\right)_i}{\left(R\cdot\left(XY^TY\right)\right)_i + \lambda_1 x_i + 2\lambda_2\sum_j Sd_{ij}\cdot x_i + 2\lambda_3\left(XX^TX\right)_i}$$

$$y_j^{new} = y_j \frac{\left(R\cdot\left(A^TX\right)\right) + 2(\lambda_2 + \lambda_3)\left((Sv)Y\right)_j}{\left(R\cdot\left(YX^TX\right)\right)_j + \lambda_1 y_j + 2\lambda_2\sum_i Sv_{ij}\cdot y_j + 2\lambda_3\left(YY^TY\right)_j}$$

According to the above formula, iteratively update matrices X and Y until the objective function reaches a local minimum. The final predicted VDA matrix is $A^* = XY^T$, where the i-th column of A^* represents the degree of correlation between v_i and other drugs.

Evaluation metrics

The area under the curve (AUC), area under the precision recall curve (AUPR), accuracy, and specificity are used as evaluation metrics for estimating the predictive performance of algorithms. ROC curves are plotted using true- and false-positive rates. PR

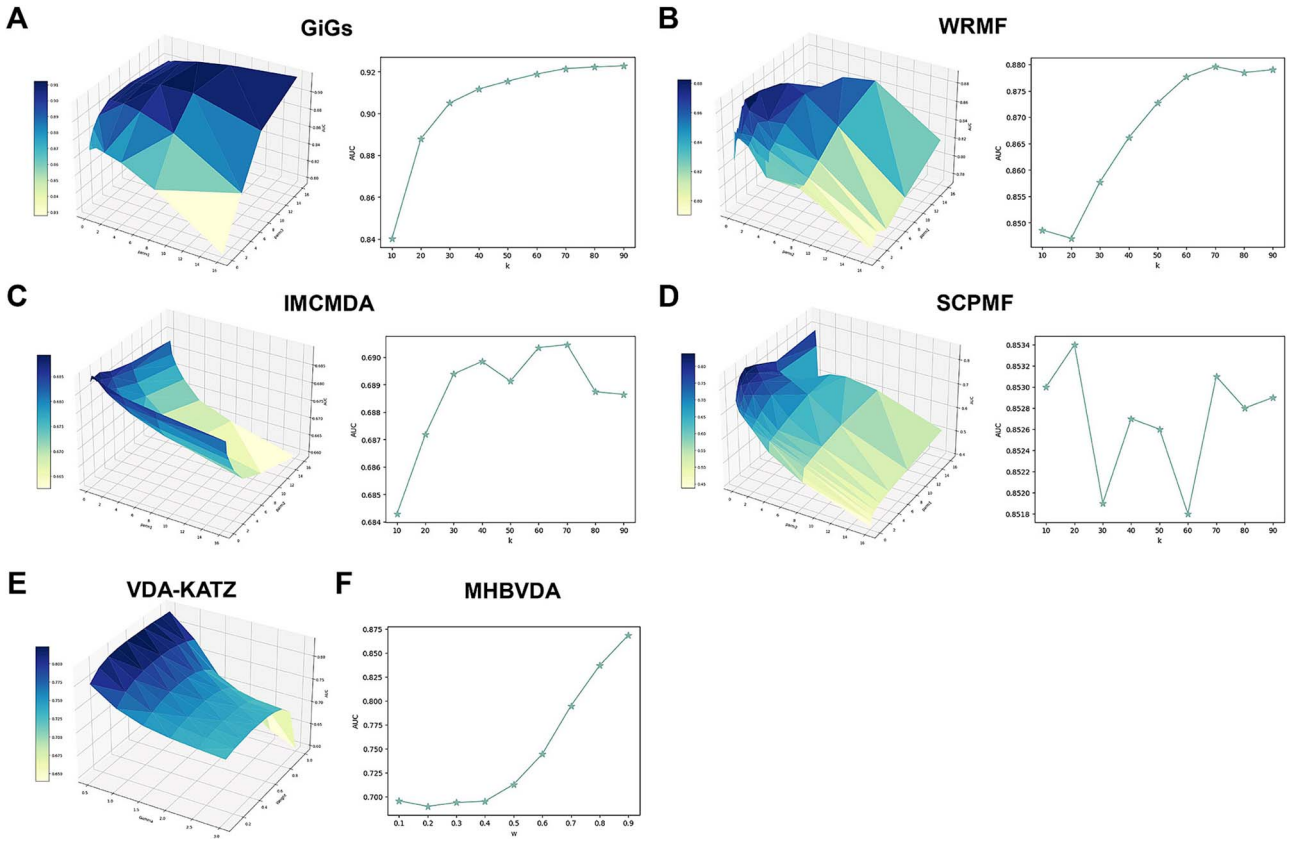

Figure 2. The impact of parameters on GiGs and other comparative algorithms performance on the DrugVirus dataset. (A–D) The regularization parameters and latent feature vector of GiGs, WRMF, IMCMDA, and SCPMF. (E, F) The impact of parameters on VDA-KATZ and MHBVDA performance.

curves are plotted using precision and recall. The calculation formula for accuracy and specificity is as follows:

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

$$specificity = \frac{TN}{TN + FP}$$

TP and TN represent the number of positive and negative samples predicted correctly, FP and FN represent the number of positive and negative samples predicted incorrectly.

Molecular docking

Molecular docking can simulate the interactions between small molecules and proteins [20], so we used it to explore the binding mode of unconfirmed drugs predicted by the GiGs model with PLpro. The 2D structures of benztropine (CID: 1201549) and trametinib (CID: 11707110) were obtained from the PubChem database. The PLpro structure of Middle East respiratory syndrome coronavirus (MERS-CoV) was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database, with Protein Data Bank (PDB) ID of 4RNA, and used AutoDockTools to remove ligands, dehydrate, and hydrogenate. Subsequently, AutoDock was used for docking and calculating the binding energy of protein–ligand complexes. Following this, we choose the best docking model to visualize and display interaction diagram using PyMOL and Discovery Studio.

Results

Parameter setting on the DrugVirus dataset

We performed GiGs and five other association prediction methods (WRMF, Inductive Matrix Completion for MiRNA–Disease Association prediction (IMCMDA) [21], SCPMF, VDA-KATZ, and MHBVDA) on the DrugVirus dataset for parameter setting. We conducted five-fold cross-validation (CV) on the training dataset and selected parameters with the best AUCs (Table 1). Five parameters needed to be determined for the GiGs algorithm: a weight parameter (α), three regularization parameters (λ_1, λ_2, and λ_3), and a latent feature vector (k). The performance of GiGs was optimal when $\alpha = 1$, $\lambda_1 = 16$, $\lambda_2 = 0.125$, $\lambda_3 = 16$, and $k = 70$ (Fig. 2A). The selection of these parameters is crucial for the model's performance. Specifically, α is a weight parameter that regulates the proportion of positive samples during training, with the step size increasing from 1 to 10. The regularization parameters λ_1, λ_2, and λ_3 control the regularization term, graph regularization constraints, and similarity constraints, respectively. In training, consider combining $\{2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4\}$. The latent feature vector k is used to explore low-dimensional spaces for drug and virus projections, and this parameter must be less than the minimum value between the number of rows and columns of the correlation matrix. During the training process,

Table 1. Parameter setting for different methods on the DrugVirus dataset.

Methods	Parameters
GiGs	$\alpha = 1, \lambda_1 = 16, \lambda_2 = 0.125, \lambda_3 = 16, k = 70$
WRMF	$\alpha = 1, \lambda_1 = 2, \lambda_2 = 0.125, k = 70$
IMCMDA	$\lambda_1 = 0.25, \lambda_2 = 0.5, k = 70$
SCPMF	$\lambda_1 = 2, \lambda_2 = 0.125, k = 20$
VDA-KATZ	$\beta = 0.02, \gamma = 0.5, \omega = 0.5$
MHBVDA	$\omega_1 = 0.9, \omega_2 = 0.1$

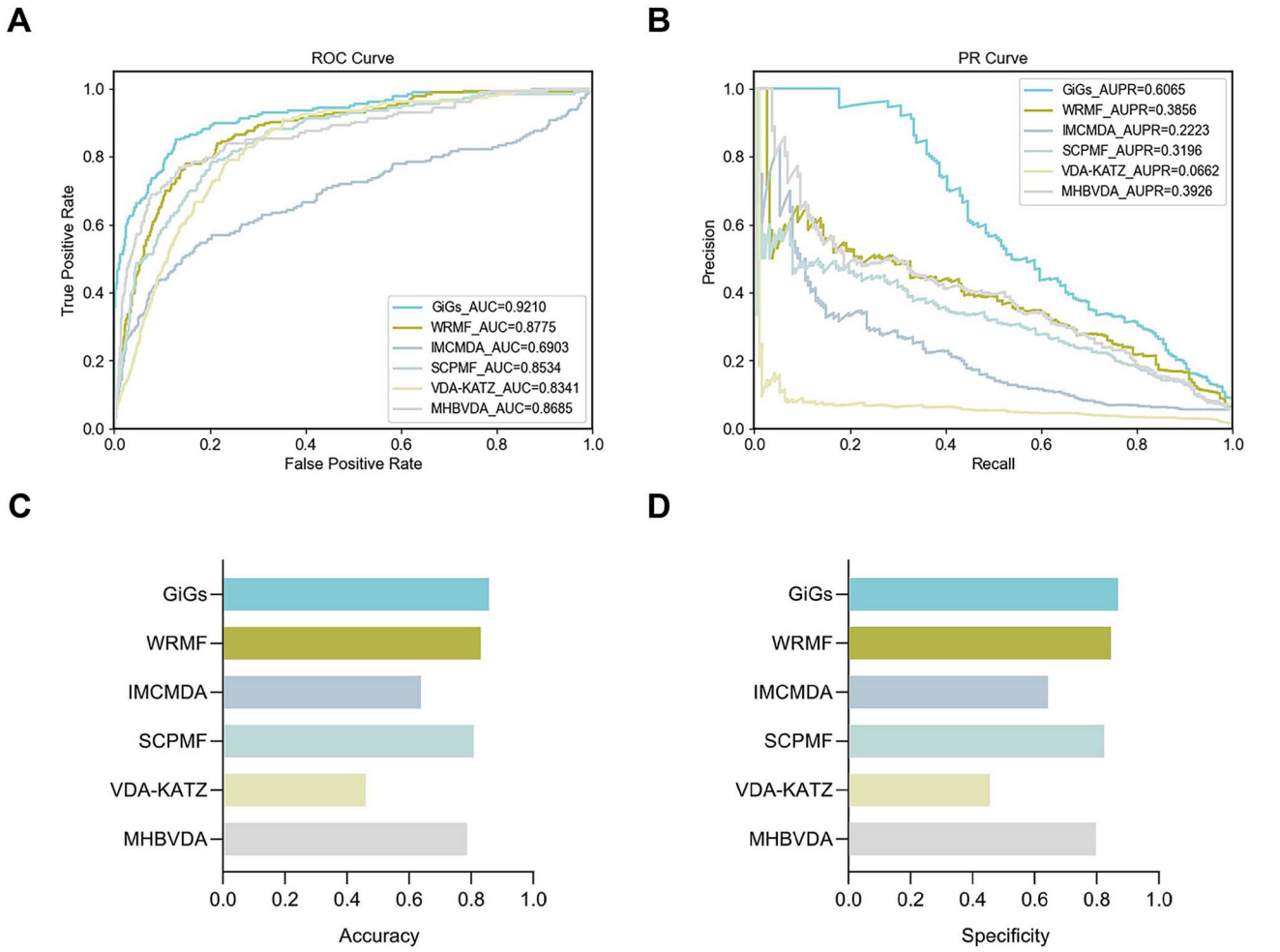

Figure 3. The performance of GiGs and other prediction methods under five-fold CV on the DrugVirus dataset. (A) ROC curve and AUC value. (B) PR curve and AUPR value. (C) Accuracy. (D) Specificity.

the parameters increase from 10 to 90 in steps of 10. To ensure fair comparison, the parameters of the other models were also optimized [WRMF algorithm, $\alpha = 1$, $\lambda_1 = 2$, $\lambda_2 = 0.125$, and $k = 70$ (Fig. 2B); IMCMDA algorithm, $\lambda_1 = 0.25$, $\lambda_2 = 0.5$, and $k = 70$ (Fig. 2C); SCPMF algorithm, $\lambda_1 = 2$, $\lambda_2 = 0.125$, and $k = 20$ (Fig. 2D); VDA-KATZ algorithm, $\beta = 0.02$, $\gamma = 0.5$, and $\omega = 0.5$ (Fig. 2E); MHBVDA algorithm, $\omega_1 = 0.9$ and $\omega_2 = 0.1$ (Fig. 2F)].

Performance under five-fold CV on the DrugVirus dataset

The AUC, AUPR, accuracy, and specificity were used to assess the performance of algorithms. The GiGs algorithm showed excellent performance under five-fold CV on the DrugVirus dataset (Fig. 3, Table 2). The AUC metric is a robust measure of a model's ability to distinguish between positive and negative samples in classification tasks, GiGs achieved an AUC of 0.9210, outperforming other algorithms (WRMF, 0.8775; IMCMDA, 0.6903; SCPMF, 0.8534; VDA-KATZ, 0.8341; MHBVDA, 0.8685; Fig. 3A). The high AUC score suggests that the GiGs has a strong capability to discriminate between samples effectively, which is crucial for accurately predicting potential antiviral drugs. GiGs also excelled in AUPR (0.6065) compared to other models (WRMF, 0.3856; IMCMDA, 0.2223; SCPMF, 0.3196; VDA-KATZ, 0.0662; MHBVDA, 0.3926; Fig. 3B), suggesting its enhanced ability to accurately identify positive samples, a critical aspect in VDA prediction. Moreover, GiGs showed the greatest accuracy (0.8575) and specificity (0.8688; Fig. 3C and D), suggesting its robustness in predictive performance across both positive and negative predictions. These results indicate that the optimized GiGs algorithm more accurately detected drugs of potential use for the treatment of novel viruses.

Performance under five-fold CV on the VDA2 and HDVD datasets

We tested the applicability of GiGs on the VDA2 and HDVD datasets, finding optimal parameters of $\alpha = 1$, $\lambda_1 = 8$, $\lambda_2 = 0.125$, $\lambda_3 = 16$, $k = 40$ for VDA2 (Fig. 4A and B, Table 3) and $\alpha = 1$, $\lambda_1 = 0.25$, $\lambda_2 = 0.125$, $\lambda_3 = 4$, and $k = 24$ for HDVD (Fig. 4C and D, Table 3). The optimal parameter values for the other five algorithms were determined by grid search (Figs S1 and S2, Table 3). AUC and AUPR values can comprehensively evaluate performance, which is more important than other evaluation metrics. The GiGs algorithm demonstrated superior performance compared to five other algorithms across two datasets, as indicated by higher AUC and AUPR values. On the VDA2 dataset, GiGs achieved an AUC of 0.9253 and AUPR of 0.6807, outperforming WRMF (AUC: 0.8359, AUPR: 0.3871), IMCMDA (AUC: 0.6103, AUPR: 0.1811), SCPMF (AUC: 0.8478, AUPR: 0.2559), VDA-KATZ (AUC: 0.8325, AUPR: 0.0586), and MHBVDA (AUC: 0.8198, AUPR: 0.3797; Fig. 5A and B). On the HDVD dataset, GiGs further excelled with an AUC of 0.9553 and AUPR of 0.7331, surpassing WRMF (AUC: 0.8712, AUPR: 0.4882), IMCMDA (AUC: 0.5236, AUPR: 0.1262), SCPMF (AUC: 0.8573, AUPR: 0.4899),

Table 2. Experimental results under five-fold CV on the DrugVirus dataset.

Methods	AUC	AUPR	Accuracy	Specificity
GiGs	**0.9210**	**0.6065**	**0.8575**	**0.8688**
WRMF	0.8775	0.3856	0.8310	0.8466
IMCMDA	0.6903	0.2223	0.6390	0.6439
SCPMF	0.8534	0.3196	0.8085	0.8234
VDA-KATZ	0.8341	0.0662	0.4600	0.4561
MHBVDA	0.8685	0.3926	0.7871	0.7980

Bold values represent the best performance values.

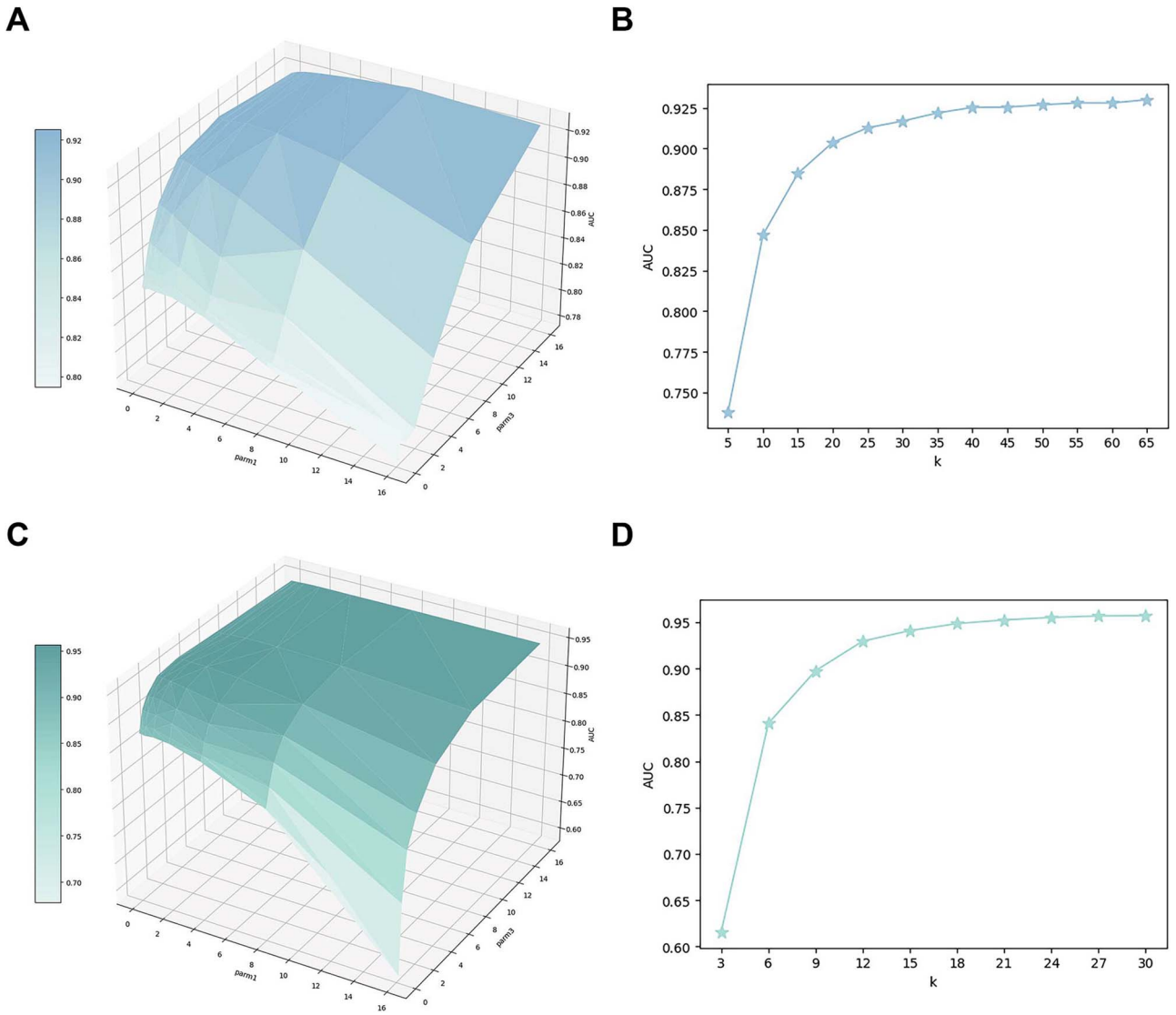

Figure 4. The impact of parameters on GiGs performance on the VDA2 and HDVD datasets. (A, B) The regularization parameters and latent feature vector of GiGs on the VDA2 datasets. (C, D) The regularization parameters and latent feature vector of GiGs on the HDVD datasets.

Table 3. Parameter setting for different algorithms on the VDA2 and HDVD datasets.

Methods	VDA2 dataset	HDVD dataset
GiGs	$\alpha = 1$, $\lambda_1 = 8$, $\lambda_2 = 0.125$, $\lambda_3 = 16$, $k = 40$	$\alpha = 1$, $\lambda_1 = 0.25$, $\lambda_2 = 0.125$, $\lambda_3 = 4$, $k = 24$
WRMF	$\alpha = 1$, $\lambda_1 = 8$, $\lambda_2 = 0.25$, $k = 30$	$\alpha = 5$, $\lambda_1 = 8$, $\lambda_2 = 1$, $k = 24$
IMCMDA	$\lambda_1 = 4$, $\lambda_2 = 8$, $k = 15$	$\lambda_1 = 0.5$, $\lambda_2 = 0.125$, $k = 9$
SCPMF	$\lambda_1 = 2$, $\lambda_2 = 0.125$, $k = 25$	$\lambda_1 = 2$, $\lambda_2 = 0.125$, $k = 12$
VDA-KATZ	$\beta = 0.04$, $\omega = 0.2$, $\gamma = 1.0$	$\beta = 0.04$, $\omega = 0.6$, $\gamma = 2.5$
MHBVDA	$\omega_1 = 0.9$, $\omega_2 = 0.1$	$\omega_1 = 0.9$, $\omega_2 = 0.1$

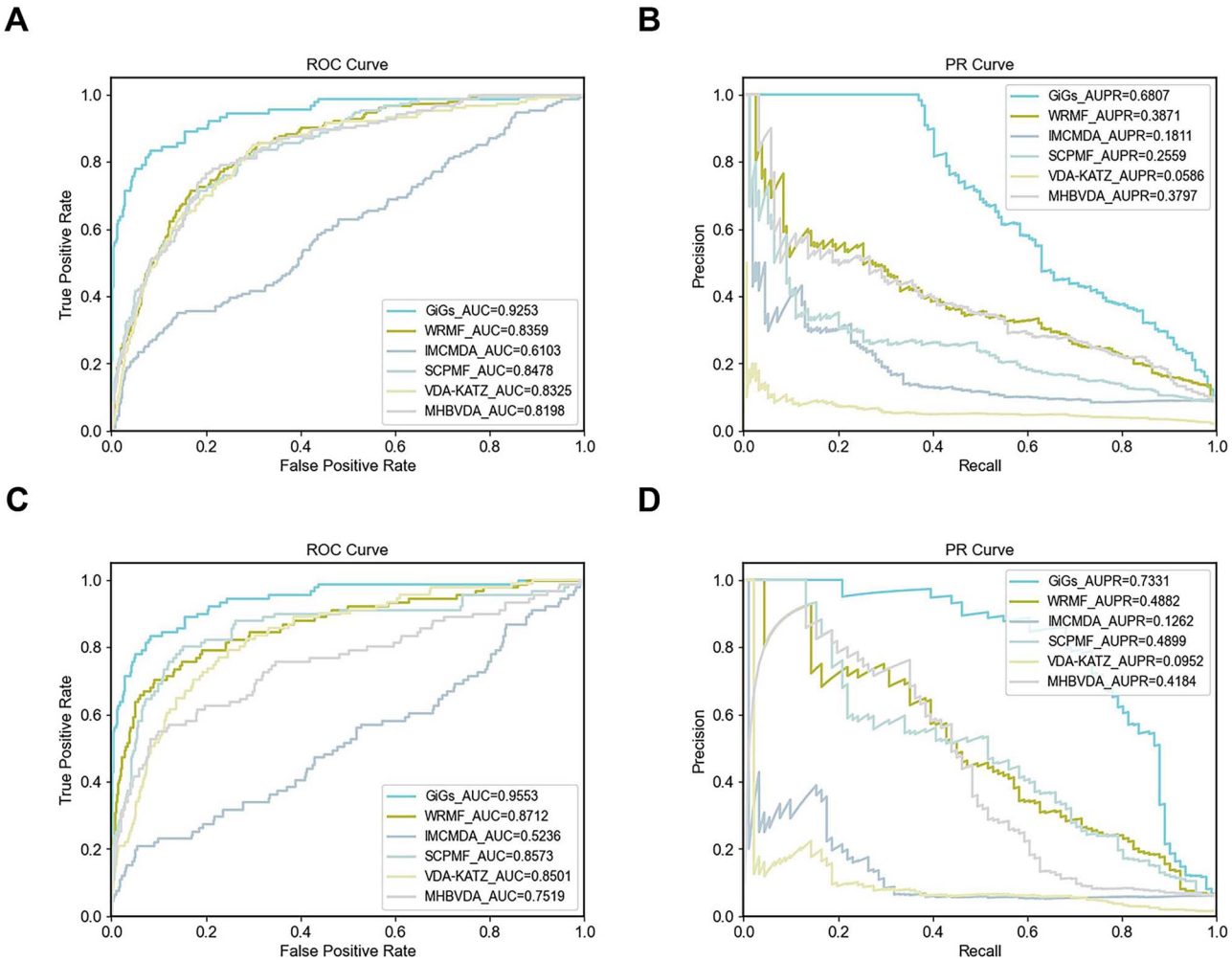

Figure 5. The performance of GiGs and other methods under five-fold CV on the VDA2 and HDVD datasets. (A, C) ROC curve and AUC value. (B, D) PR curve and AUPR value.

VDA-KATZ (AUC: 0.8501, AUPR: 0.0952), and MHBVDA (AUC: 0.7519, AUPR: 0.4184; Fig. 5C and D). These results underscore the effectiveness of GiGs in capturing complex drug–disease interactions and its reliability for drug repositioning applications.

Case study: Prediction of potentially effective broad-spectrum anticoronavirus drugs

No drug currently used in clinical practice effectively prevents or treats highly pathogenic human coronavirus infections. We applied GiGs to discover drugs for the potential treatment of broad-spectrum coronaviruses from the DrugVirus dataset and selected the top 20 drugs based on VDA probability (Fig. S3A–C and Tables S1–S3). 20, 18, and 20 of the top 20 drugs correlated with SARS-CoV, MERS-CoV, and SARS-CoV-2, and have been validated, according to the existing literature (Fig. 6A–C). In addition, five of the identified drugs (mefloquine, anisomycin, camostat, hydroxychloroquine, and chlorpromazine) had broad-spectrum anticoronavirus effects (Fig. 6D, Fig. S3D).

Molecular docking

As shown in Table S2, benztropine [22] and trametinib [23] lack evidence for treating MERS-CoV. Therefore, we performed molecular docking with these drugs and PLpro, an effective target to develop broad-spectrum anticoronavirus drugs [24], and their binding sites are shown in Fig. 7. The blue spiral sections represent the structures of MERS-CoV PLpro proteins, and the green rod-shaped structures represent unconfirmed drugs. The binding energy of PLpro–benztropine was −7.189 kcal/mol, which was attributed to van der Waals forces (PHE83, LEU71, LYS139, GLY12, ASN36, PHE35, GLY37, PHE34, and ALA88), hydrogen bonds (TYR72 and ASP11), electrostatic interactions (ASP147), and hydrophobic interactions (LYS87, TYR55, and TYR84) (Fig. 7A and B). The binding energy of PLpro–trametinib was −7.257 kcal/mol, which was attributed to van der Waals forces (HIS171, SER247, GLY248, THR308, THR249, TYR279, VAL276, PHE269, and ILE272), hydrogen bonds (SER167), hydrophobic interactions (VAL210, ARG168, and PRO250), and Pi-Anion (ASP165) (Fig. 7C and D). These binding patterns revealed interaction between these drugs and multiple binding sites on PLpro, indicating that the drugs predicted by the GiGs model might have inhibitory effects on MERS-CoV. These findings demonstrate the reliability of the GiGs model in the identification of antiviral drugs that can be used to target highly pathogenic human coronaviruses and other emerging infectious diseases.

Discussion

Computational methods can be used to quickly identify repurposable drugs to treat diseases [25]. Here, we developed the

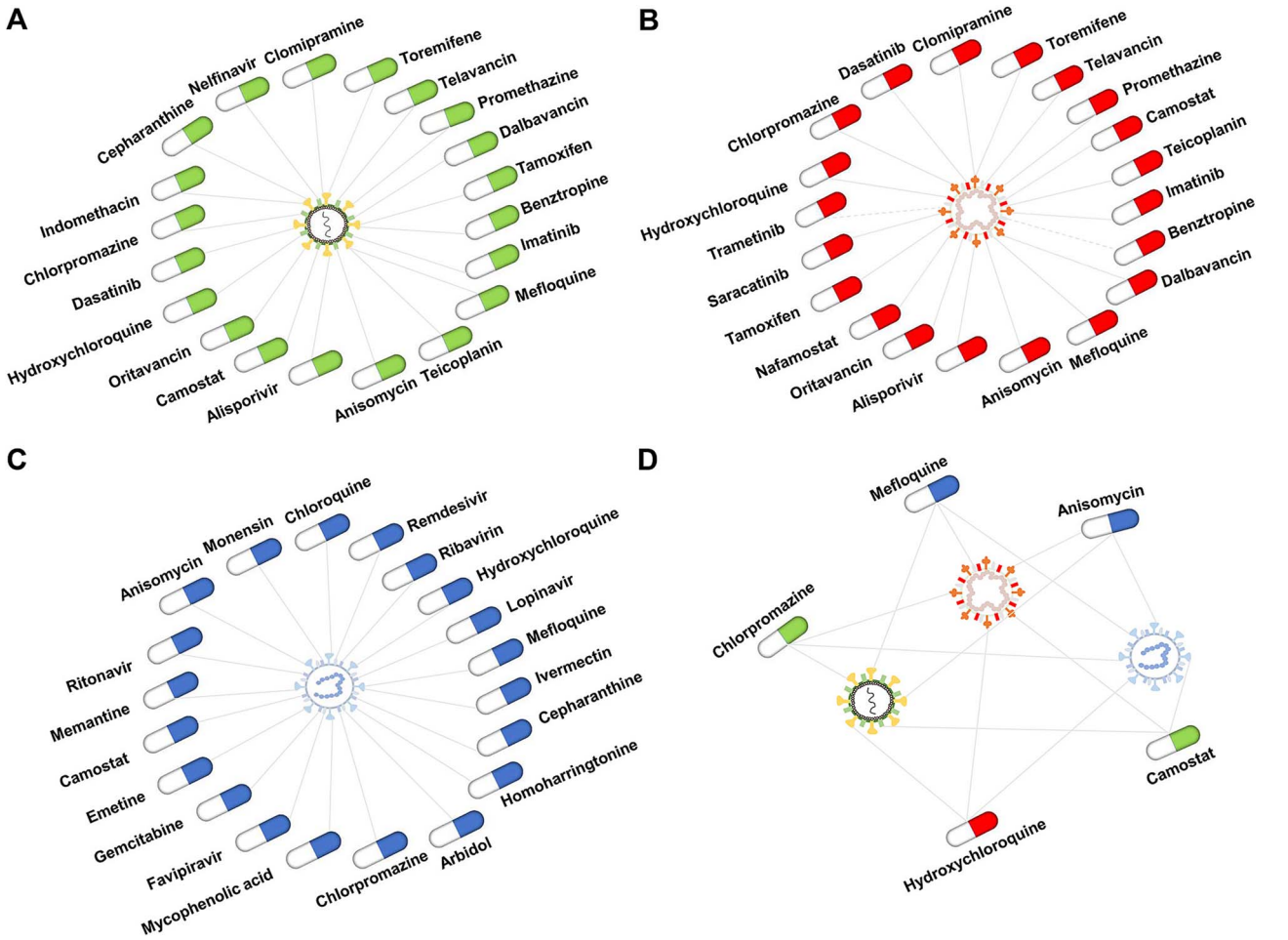

Figure 6. Predicted drugs against highly pathogenic human coronaviruses by GiGs on the DrugVirus dataset. (A) Predicted top 20 drugs against SARS-CoV. (B) Predicted top 20 drugs against MERS-CoV. (C) Predicted top 20 drugs against SARS-CoV-2. (D) Predicted broad-spectrum anti-coronavirus drugs. Solid lines represent verified drugs and dashed lines represent unverified drugs.

GiGs model for the prediction of potential VDAs to accelerate drug repositioning. The predictive ability of the GiGs model was validated using five-fold CV on the DrugVirus, VDA2, and HDVD datasets. The performance of the model was superior to that of five other association prediction methods. In addition, we applied the GiGs model to identify potential drugs treating highly pathogenic human coronavirus infections, which would contribute to controlling the existing and newly emerging coronavirus diseases.

GiGs predicted that mefloquine, anisomycin, camostat, hydroxychloroquine, and chlorpromazine would have broad-spectrum anticoronavirus effects. Mefloquine, a quinoline compound used to treat malaria, has been shown to inhibit SARS-CoV, MERS-CoV, and SARS-CoV-2 with EC_{50} values of 2.9 ± 0.8, 7.416, and 2.2 ± 0.0 μM, respectively [26–28]. The antibiotic anisomycin is demonstrated to suppress SARS-CoV and MERS-CoV inflections ($EC_{50} = 0.191$ and 0.003 μM, respectively) and SARS-CoV-2 replication ($IC_{50} = 31.4$ nM) [28–30]. Camostat can obviously suppress SARS-CoV replication, inhibit MERS-CoV cell entry and growth, and suppress SARS-CoV-2 cell entry [31–34]. Hydroxychloroquine, an antimalarial drug, was found to suppress these three highly pathogenic human coronavirus infections [28, 35, 36]. Chlorpromazine, an antipsychotic, was shown to have antiviral activity against SARS-CoV and MERS-CoV and to suppress SARS-CoV-2 replication [28, 37, 38]. The molecular docking performed in the present study further demonstrated the potential therapeutic ability of predicted unconfirmed drugs.

The excellent performance of the GiGs model is attributable to three factors. First, the model introduction of similarity constraints improves model performance. Second, the addition of graph regularization to matrix factorization prevents overfitting, preserves the similarity between each drug and virus and its nearest neighbors, and increases the accuracy of the prediction of missing data in the matrix. Third, the integration of multiple biological data improves correlation matrix quality.

The GiGs model is a promising tool for drug repositioning, but there remains room for improvement. In subsequent work, we will seek more effective means of matrix factorization while ensuring accuracy. Furthermore, the addition of more biological information will improve prediction accuracy. We will also establish a larger VDA database to increase model applicability.

Conclusion

A novel VDAs prediction model called GiGs was developed for drug repositioning. The performance of GiGs was estimated and validated via application to the DrugVirus, VDA2, and HDVD datasets and was found to be superior to that of five other association prediction models. Mefloquine, anisomycin,

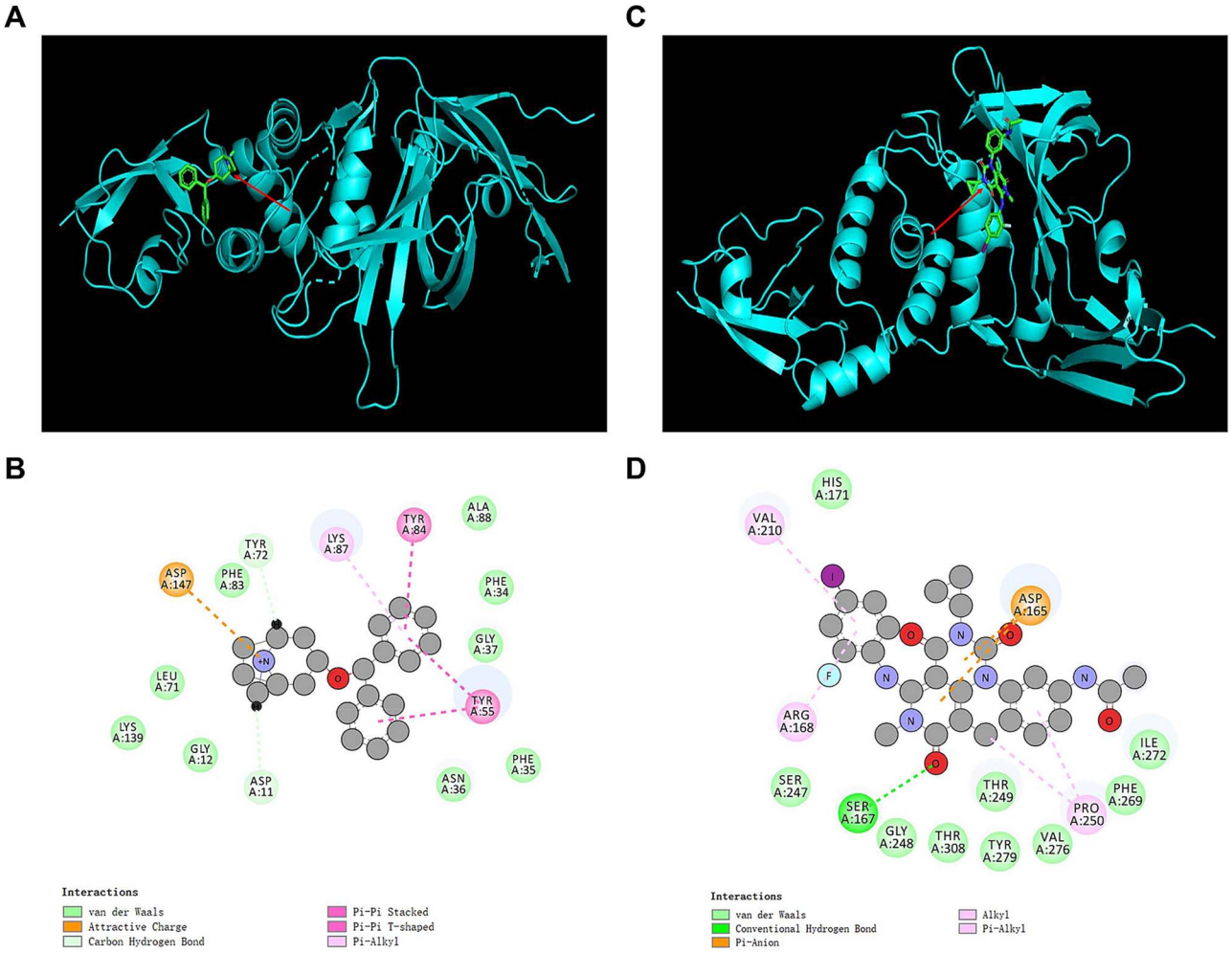

Figure 7. The molecular docking patterns between unconfirmed drugs and MERS-CoV PLpro. (A, B) PLpro-benztropine. (C, D) PLpro-trametinib.

camostat, hydroxychloroquine, and chlorpromazine were predicted to be effective broad-spectrum anticoronavirus drugs, providing new insight for the treatment of highly pathogenic human coronavirus infections. In summary, GiGs is a promising VDAs prediction method that can provide meaningful guidelines for treating diseases.

Key Points

- GiGs improves correlation matrix quality due to the integration of multiple biological data.
- The GiGs model shows optimal performance compared with five other association prediction models.
- GiGs is a promising model for predicting VDA and provides meaningful guidelines for treating diseases.

Author contributions

Yixuan Jin: formal analysis, investigation, validation, visualization, and writing-original draft. Juanjuan Huang: formal analysis, investigation, validation, visualization, and writing-original draft. Xu Sun: investigation, and validation. Yabo Fang: investigation, and validation. Jiageng Wu: investigation. Jianshi Du: conceptualization, writing-review and editing. Jiwei Jia: conceptualization, methodology, funding acquisition, and writing-review and editing. Guoqing Wang: conceptualization, methodology, funding acquisition, and writing-review and editing.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Funding

This work was supported by National Key R&D Program of China (2022YFF1203204); Project supported by the Joint Funds of the National Natural Science Foundation of China (U23A20269); Natural Science Foundation of Jilin Province (20210101481JC); and China Natural National Science Foundation (22341302).

Data availability

The codes and datasets of GiGs are available in its online supplementary material.

References

1. Zhou Y, Wang F, Tang J. *et al.* Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health* 2020;**2**:e667–76. https://doi.org/10.1016/S2589-7500(20)30192-8.

2. Sonaye HV, Sheikh RY, Doifode CA. Drug repurposing: Iron in the fire for older drugs. *Biomed Pharmacother* 2021;**141**:111638. https://doi.org/10.1016/j.biopha.2021.111638.
3. Hassanali, Aragh A, Givehchian P, Moslemi Amirani R. *et al.* MiRAGE: Mining relationships for advanced generative evaluation in drug repositioning. *Brief Bioinform* 2024;**25**:bbae337. https://doi.org/10.1093/bib/bbae337.
4. Wei J, Zhuo L, Fu X. *et al.* DrugReAlign: A multisource prompt framework for drug repurposing based on large language models. *BMC Biol* 2024;**22**:226.
5. Avorn J. The $2.6 billion pill–methodologic and policy considerations. *N Engl J Med* 2015;**372**:1877–9. https://doi.org/10.1056/NEJMp1500848.
6. Zeng X, Zhu S, Lu W. *et al.* Target identification among known drugs by deep learning from heterogeneous networks. *Chem Sci* 2020;**11**:1775–97. https://doi.org/10.1039/C9SC04336E.
7. Wang G, Zhang X, Pan Z. *et al.* Multi-TransDTI: Transformer for drug-target interaction prediction based on simple universal dictionaries with multi-view strategy. *Biomolecules* 2022;**12**:644.
8. Ren ZH, You ZH, Zou Q. *et al.* DeepMPF: Deep learning framework for predicting drug-target interactions based on multi-modal representation with meta-path semantic analysis. *J Transl Med* 2023;**21**:48. https://doi.org/10.1186/s12967-023-03876-3.
9. Tan H, Zhang Z, Liu X. *et al.* MDSVDNV: Predicting microbe-drug associations by singular value decomposition and Node2vec. *Front Microbiol* 2023;**14**:1303585. https://doi.org/10.3389/fmicb.2023.1303585.
10. Liu J, Guan S, Zou Q. *et al.* AMDGT: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction. *Knowl-Based Syst* 2024;**284**111329. https://doi.org/10.1016/j.knosys.2023.111329.
11. Zhou L, Wang J, Liu G. *et al.* Probing antiviral drugs against SARS-CoV-2 through virus-drug association prediction based on the KATZ method. *Genomics* 2020;**112**:4427–34. https://doi.org/10.1016/j.ygeno.2020.07.044.
12. 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. https://doi.org/10.1016/j.asoc.2021.107135.
13. Xu J, Meng Y, Peng L. *et al.* Computational drug repositioning using similarity constrained weight regularization matrix factorization: A case of COVID-19. *J Cell Mol Med* 2022;**26**:3772–82. https://doi.org/10.1111/jcmm.17412.
14. Qu J, Song Z, Cheng X. *et al.* A new integrated framework for the identification of potential virus-drug associations. *Front Microbiol* 2023;**14**:1179414. https://doi.org/10.3389/fmicb.2023.1179414.
15. Su X, Hu L, You Z. *et al.* A deep learning method for repurposing antiviral drugs against new viruses via multi-view non-negative matrix factorization and its application to SARS-CoV-2. *Brief Bioinform* 2022;**23**:bbab526. https://doi.org/10.1093/bib/bbab526.
16. Long Y, Wu M, Kwoh CK. *et al.* Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* 2020;**36**:4918–27. https://doi.org/10.1093/bioinformatics/btaa598.
17. Andersen PI, Ianevski A, Lysvand H. *et al.* Discovery and development of safe-in-man broad-spectrum antiviral agents. *Int J Infect Dis* 2020;**93**:268–76. https://doi.org/10.1016/j.ijid.2020.02.018.
18. Katoh K, Misawa K, Kuma K. *et al.* MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66. https://doi.org/10.1093/nar/gkf436.
19. Hattori M, Tanaka N, Kanehisa M. *et al.* SIMCOMP/SUBCOMP: Chemical structure search servers for network analyses. *Nucleic Acids Res* 2010;**38**:W652–6. https://doi.org/10.1093/nar/gkq367.
20. Meng XY, Zhang HX, Mezei M. *et al.* Molecular docking: A powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* 2011;**7**:146–57. https://doi.org/10.2174/157340911795677602.
21. Chen X, Wang L, Qu J. *et al.* Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 2018;**34**:4256–65. https://doi.org/10.1093/bioinformatics/bty503.
22. Gelenberg AJ, Van Putten T, Lavori PW. *et al.* Anticholinergic effects on memory: Benztropine versus amantadine. *J Clin Psychopharmacol* 1989;**9**:180–5.
23. Wright CJ, McCormack PL. Trametinib: First global approval. *Drugs* 2013;**73**:1245–54. https://doi.org/10.1007/s40265-013-0096-1.
24. Yuan S, Gao X, Tang K. *et al.* Targeting papain-like protease for broad-spectrum coronavirus inhibition. *Protein Cell* 2022;**13**:940–53. https://doi.org/10.1007/s13238-022-00909-3.
25. Deepthi K, Jereesh AS, Yuansheng L. 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. https://doi.org/10.1016/j.asoc.2021.107945.
26. Engwerda AHJ, Maassen R, Tinnemans P. *et al.* Attrition-enhanced deracemization of the antimalaria drug mefloquine. *Angew Chem Int Ed Engl* 2019;**58**:1670–3. https://doi.org/10.1002/anie.201811289.
27. Persoons L, Vanderlinden E, Vangeel L. *et al.* Broad spectrum anti-coronavirus activity of a series of anti-malaria quinoline analogues. *Antivir Res* 2021;**193**:105127. https://doi.org/10.1016/j.antiviral.2021.105127.
28. Dyall J, Coleman CM, Hart BJ. *et al.* Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob Agents Chemother* 2014;**58**:4885–93. https://doi.org/10.1128/AAC.03036-14.
29. Shao E, Zhao S, Dong Y. *et al.* Anisomycin inhibits coxsackievirus B replication by promoting the lysosomal degradation of eEF1A1. *Antivir Res* 2023;**215**:105621. https://doi.org/10.1016/j.antiviral.2023.105621.
30. Huang CT, Chao TL, Kao HC. *et al.* Enhancement of the IFN-β-induced host signature informs repurposed drugs for COVID-19. *Heliyon* 2020;**6**:e05646. https://doi.org/10.1016/j.heliyon.2020.e05646.
31. Hart PA, Osypchuk Y, Hovbakh I. *et al.* A randomized controlled phase 2 dose-finding trial to evaluate the efficacy and safety of Camostat in the treatment of painful chronic pancreatitis: The TACTIC study. *Gastroenterology* 2024;**166**:658–666.e656. https://doi.org/10.1053/j.gastro.2023.12.008.
32. Zhao H, To KKW, Lam H. *et al.* Cross-linking peptide and repurposed drugs inhibit both entry pathways of SARS-CoV-2. *Nat Commun* 2021;**12**:1517. https://doi.org/10.1038/s41467-021-21825-w.
33. Shirato K, Kawase M, Matsuyama S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* 2013;**87**:12552–61. https://doi.org/10.1128/JVI.01890-13.
34. Hempel T, Raich L, Olsson S. *et al.* Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat. *Chem Sci* 2021;**12**:983–92. https://doi.org/10.1039/D0SC05064D.
35. Koranda FC. Antimalarials. *J Am Acad Dermatol* 1981;**4**:650–5. https://doi.org/10.1016/S0190-9622(81)70065-3.

36. Ou T, Mou H, Zhang L. *et al.* Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. *PLoS Pathog* 2021;**17**:e1009212. https://doi.org/10.1371/journal.ppat.1009212.
37. Zare M, Bazrafshan A. Chlorpromazine versus metiapine for schizophrenia. *Cochrane Database Syst Rev* 2017;**3**:Cd011655. https://doi.org/10.1002/14651858.CD011655.pub2.
38. Plaze M, Attali D, Prot M. *et al.* Inhibition of the replication of SARS-CoV-2 in human cells by the FDA-approved drug chlorpromazine. *Int J Antimicrob Agents* 2021;**57**:106274. https://doi.org/10.1016/j.ijantimicag.2020.106274.

Briefings in Bioinformatics, 2025, **26(2)**, bbaf117
https://doi.org/10.1093/bib/bbaf117
Problem Solving Protocol