Drug Repositioning for SARS-CoV-2 Based on Graph Neural Network*

1st Haifeng Liu
Department of Computer Science
Dalian University of Technology
LiaoNing, China
liuhaifeng@mail.dlut.edu.cn

2nd Hongfei Lin*
Department of Computer Science
Dalian University of Technology
LiaoNing, China
hflin@dlut.edu.cn

3rd Chen Shen
Department of Computer Science
Dalian University of Technology
LiaoNing, China
shenchenioi@mail.dlut.edu.cn

4th Liang Yang
Department of Computer Science
Dalian University of Technology
LiaoNing, China
liang@dlut.edu.cn

5th Yuan Lin
Faculty of Humanities and Social Sciences
Dalian University of Technology
LiaoNing, China
zhlin@dlut.edu.cn

6th Bo Xu
Department of Computer Science
Dalian University of Technology
LiaoNing, China
xubo@dlut.edu.cn

7th Zhihao Yang
Department of Computer Science
Dalian University of Technology
LiaoNing, China
yangzh@dlut.edu.cn

8th Jian Wang
Department of Computer Science
Dalian University of Technology
LiaoNing, China
wangjian@dlut.edu.cn

9th Yuanyuan Sun
Department of Computer Science
Dalian University of Technology
LiaoNing, China
syuan@dlut.edu.cn

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313236

***Abstract*—Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the strain of coronavirus that causes coronavirus disease 2019 (COVID-19), which leads to over 800,000 deaths and is still no specific medicines. Drug repositioning aiming to infer potential drugs for diseases and achieve much attention during the SARS-CoV-2 epidemic. However, find a specific drug of SARS-CoV-2 is still a large challenge that cannot be addressed well with current methods. To overcome this problem, we present a novel drug repositioning framework of heterogeneous graph convolutional networks for SARS-CoV-2. The deep2CoV model can effectively search the potential drugs for SARS-CoV-2, which reduce the number of clinical trials and drug development cycles. The experimental results demonstrate the effectiveness and feasibility of our proposed deep2CoV framework.**

***Index Terms*—SARS-CoV-2, COVID-19, Drug Repositioning, Graph Neural Network.**

I. Introduction

SARS-CoV-2 is the strain of coronavirus that causes COVID-19, the respiratory illness responsible for the SARS-CoV-2 pandemic. As to August 2020, SARS-CoV-2 has spread to over 150 countries, and resulted in 19 million people infected with over 800,000 deaths (fatality rate of 3.6%), and still have more than 6.7 million confirmed cases, which cost the global economy with an estimated over 100 billion dollars. The timeline of the COVID-19 pandemic showcases the urgent need for fast development of effective treatments for new diseases. However, there are currently no effective medications against the SARS-CoV-2 virus. Several national and international research groups are working on the development of vaccines to prevent and treat the SARS-CoV-2, but effective vaccines are not available yet. There is an urgent need for the development of effective prevention and treatment strategies for the SARS-CoV-2 outbreak. Drug repositioning has become a hopeful method to treat SARS-CoV-2.

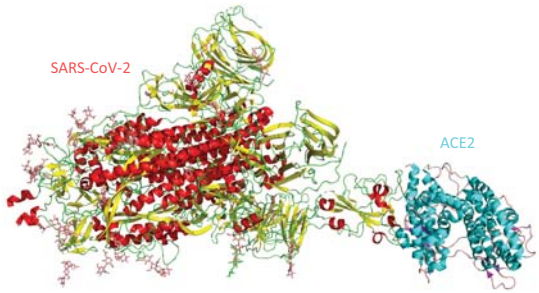

Fig. 1. SARS-CoV-2 & ACE2

In recent years, an increasing number of machine learning methods have been proposed for drug repositioning, For example, Menden et al. [1] proposed machine-learning models to predict the response of cancer cell lines to drug treatment through IC_{50} values. The network-based analysis is another research trends for drug repositioning. Zeng et al. [2] developed a networks-based deep learning approach to integrate heterogeneous network for Alzheimer's disease and Parkinson's disease. Luo et al. [3] proposed a random

* Corresponding author.

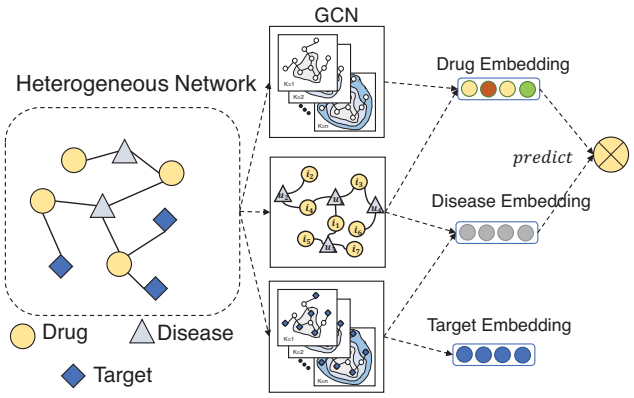

Fig. 2. Our framework for SARS-CoV-2 drug repositioning

walk model with a heterogeneous network to predict potential drug for diseases. Text mining of semantic inference approach mined and retrieved possible drug information via the amount of literature. Chen et al. [4] developed a statistical model to assess drug-target associations from a semantic network. However, the above methods only consider part of drug-disease association, the drug-drug networks often were ignored.

As we know, there are some challenges for SARS-CoV-2 drug repositioning. The most current methods often focus on only one aspect of drug repositioning, such as drug-target interaction prediction [2] or drug-disease association prediction [5], which lack of comprehensive consideration of multiple interactive networks among drugs, targets, and diseases. Therefore, it needs to mine hidden associations among the drug-target-disease heterogeneous network.

In this paper, we propose a novel and effective graph-based deep learning model (deep2CoV) to systematically infer new drug-disease relationships for SARS-CoV-2 drug repositioning. The underlying concept of deep2CoV is to fuse diverse information from SARS-CoV-2 drug-drug network, drug-disease network, and drug-target network, and infer potential drug for SARS-CoV-2 by a collective graph convolutional network. The advantages of deep2CoV can be summarized as follows: (1) deep2CoV integrate diverse information from the drug-drug network, drug-disease network, and drug-target network, which can reduce the data sparsity of SARS-CoV-2 drug repositioning; (2) deep2CoV used graph convolution function to aggregate local nodes information, which boosts the drug, disease and target non-linear feature interaction.

II. MATERIALS AND METHODS

In this study, we propose a novel drug repositioning approach, named deep2CoV, to prioritize candidate drugs for SARS-CoV-2. The framework is showed in Fig. 2.

A. Dataset

Since there is currently no public drug repositioning dataset on SARS-CoV-2, we build our dataset from multiple open sources, including DrugBank[1], repoDB[2], News reports[3] and some literature[4]. In total, we constructed three networks: (1) drug-disease interactions. The interactions between drugs and diseases come from the repoDB[2] dataset, which provides some "Approved" drug-disease information for drug repositioning by Drug Central and ClinicalTrials.gov. (2) drug-target interactions. The drug-target interactions consist of two parts, the first part drug-target interactions obtained from DrugBank[1] database, which is a biomedical database with detailed drug data and includes known drug indications and direct or indirect gene targets. The second part of the dataset comes from manual labeling. Since there is currently no drug-target interaction data on SARS-CoV-2, we look through SARS-CoV-2 related literature[4] and labeled 70 pairs drug-target information about SARS-CoV-2 for this study. 3) drug-drug interactions. The interactions between drugs and drugs were obtained from literature [2], which provided 290,836 clinically reported drug-drug interactions. The statistic of drugs, diseases, targets, and all the interactions constructed in this study shown in Table I.

TABLE I
STATISTICS OF THE DATASET USED IN THIS STUDY

Dataset	Drug	Diseases	Targets
	7,365	1,230	4,596
	Drug-Drug Interactions	Drug-Disease Interactions	Drug-Target Interactions
	290,836	6,680	25,012

B. Model training on the heterogeneous network

Based on the constructed heterogeneous network, deep2CoV simulates the process of the graph convolutional network on the heterogeneous network to predict candidate drugs for the SARS-CoV-2 disease under consideratio.

1) Embedding Layer: First, we introduce the disease, drug and target initialize representation embedding with a vector initializer, denoted as $\boldsymbol{p} \in \mathbb{R}^{L\times D}$ for diseases, $\boldsymbol{q} \in \mathbb{R}^{M\times D}$ for drugs, $\boldsymbol{r} \in \mathbb{R}^{N\times D}$ for targets, where D means embedding dimensions, the L, M, N mean numbers of disease, drugs and targets. We define $\boldsymbol{p_a}$ means the embedding of disease a, $\boldsymbol{q_b}$ means the embedding of drug b, $\boldsymbol{r_c}$ means the embedding of target c.

2) Heterogeneous network modeling: In addition to the disease-drug graph, the drug-drug graph and drug-target graph provides a great opportunity to learn disease and drug representations from different perspectives. The GNNs-based network analysis employ graph neural networks to aggregate graph feature of neighboring nodes, which makes the aggravated representation more powerful. In this paper, we extend the graph neural network into a heterogeneous with graph convolutional network for drug repositioning.

Disease-Drug Aggregation. To learn representations of diseases in the disease-drug graph, we introduce to utilize the

[1]https://www.drugbank.ca/
[2]http://apps.chiragjpgroup.org/repoDB/
[3]https://www.who.int/emergencies/diseases/novel-coronavirus-2019
[4]https://bigd.big.ac.cn/ncov/publication

graph convolutional networks to aggregate local neighbors of diseases as follows,

$$\boldsymbol{p_a^g} = Agg_{drug2disease}\{\frac{1}{|R_a|}\sum_{b\in R_a}\boldsymbol{q_b}\} \quad (1)$$

where R_a means disease a associations drug b set, and $Agg\{*\}$ is a mean-based aggregation operation of GNNs [6].

Drug-Drug Aggregation. To consider these heterogeneous strengths of drug-drug interactions, we introduce to differentiate drugs' local neighbors during aggregation operation in graph neural networks as follows,

$$\boldsymbol{q_b^g} = \sigma(\boldsymbol{W}\cdot Agg_{drug2drug}\{\sum_{d\in DDI_b} w_{d,b}(\boldsymbol{q_d}\odot\boldsymbol{q_b})\}) \quad (2)$$

where DDI_b means drug b interactions set of drug d in drug-drug network, the σ means relu activation function, and the $\odot$ denote the element-wise product, $\boldsymbol{W}$ is a trainable parametric matrix. Similarly, $w_{d,b}$ is the important weight between drug d and drug b, here we used GCN laplacian matrix [6] as weight matrix.

In addition to the directly connected drugs, distant drugs can also be beneficial. Therefore, we introduce to aggregate DDI information through k-layer aggregation as follows,

$$\boldsymbol{q_b^{g^{k+1}}} = \sigma(\boldsymbol{W}\cdot Agg_{drug2drug}\{\sum_{d\in DDI_b} w_{d,b}(\boldsymbol{q_d^{g^k}}\odot\boldsymbol{q_b^{g^k}})\}) \quad (3)$$

where $\boldsymbol{q_b^{g^k}}$ denotes the drug b representation after k-layer aggregation operation. The drug initial representation $\boldsymbol{q_b^{g^k}}$ is equal to $\boldsymbol{q_b^g}$ when k=0.

The Drug-Target aggregation is similar to Drug-Drug aggregation. Here we ignore it.

C. Prediction Component

As the drug-drug and drug-target provide important signals to understand drugs interactions information, we propose to obtain the final drug representation $\boldsymbol{q_b^f}$ as follows:

$$\boldsymbol{q_b^f} = \boldsymbol{q_b^{g^{k+1}}} \oplus \boldsymbol{q_b^{t^{k+1}}} \quad (4)$$

where $\oplus$ indicates summation operation.

With the disease and drug representation (e.g., $\boldsymbol{p_a^g}$ and $\boldsymbol{q_b^f}$), we perform score prediction via the inner product as follows,

$$r_{a,b} = \boldsymbol{p_a^{g^T}}\boldsymbol{q_b^f} \quad (5)$$

D. Model Training

To learn the model parameters of our proposed model, we adopt the pair-wise loss as our objective for the SARS-CoV-2 drug repositioning task as follows,

$$\min_{\theta} Loss = \sum_{a=1}^{L}\sum_{(b,d)\in R_a} -\sigma(r_{a,b}-r_{a,d}) + \lambda\|\theta\|^2 \quad (6)$$

where $\sigma(\cdot)$ is a sigmoid function. L denote the number of disease for training. θ denotes all trainable model parameters in our deep2CoV framework. λ is a regularization parameter that controls the complexity of disease and drug graph representation matrices. R_a denotes disease a interactive drugs set. By optimizing the loss function, all parameters can be tuned via backward propagation.

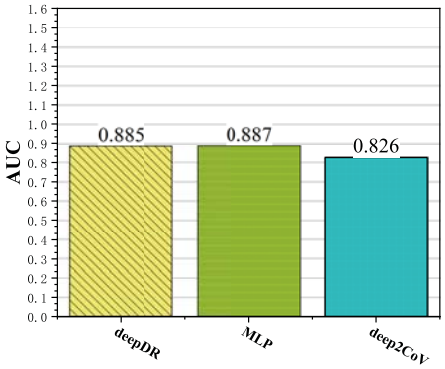

Fig. 3. Performance of experiment dataset on AUC

III. Results

A. Baseline of deep2CoV

We compare our method to some pairwise association prediction methods, including deepDR [2] and MLP [7] methods, which used deep learning model for drug reponsitioning. The detailed introduction is as follows.

- **deepDR [2]:** deepDR model constructed nine network for drug repositioning. Since there is currently no SARS-CoV-2 drug side effect association, we only used drug-disease interactions of deepDR for drug repositioning.
- **MLP [7]:** MLP model is one of the deep learning model, which used transcriptomic data for predicting drug repositioning. The MLP consists of three or more layers with one or more hidden layers of nonlinearly-activating nodes, here we used two hidden layer for training.
- **deep2CoV:** Our deep2CoV framework, which contained three interaction network, including drug-drug, drug-target and drug-disease interactions.

B. Performance of deep2CoV on the dataset

We use the area under the receiver operating characteristic curve (AUC) and area under the precision-recall curve (AUPR) to evaluate the deep2CoV performance. During the 5-fold cross-validation, we randomly selected a subset of 20% of the clinically reported drug-disease pairs with three times number of randomly sample unknown pairs as the test set. The remaining 80% clinically reported drug-disease pairs with three times the number of randomly sampled unknown pairs were used to train the model. The detailed performance is shown in Fig. 3 and Fig. 4.

As shown in Fig. 3 and Fig. 4, we can find that: (1) The MLP achieves worse performance than other deep model baseline models, means that MLP has some limitations in a heterogeneous network, the possible reason is that drug-disease interactions have graph structure information, but the

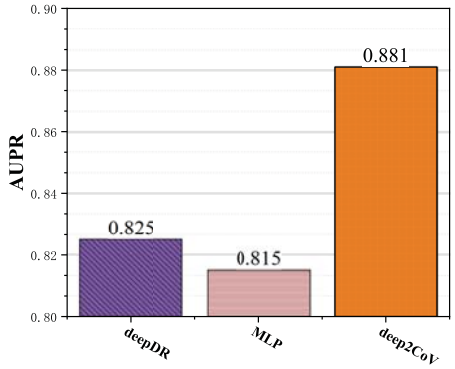

Fig. 4. Performance of experiment dataset on AUPR

TABLE II
TOP 10 PREDICTED DRUG FOR SARS-COV-2

DrugBank ID	Drug Name	Target UniProt ID	References
DB01212	Ceftriaxone	P0A3M6, Q9NSA0	[8]
DB00537	Ciprofloxacin	P43702, P43700	[9]
DB00319	Piperacillin	Q75Y35, P0A3M6	[10]
DB00681	Amphotericin B	Q9UKR5	[11]
DB00254	Doxycycline	Q63X76	[12]

MLP hard to learn local graph information. (2) The deepDR achieves better performance than MLP on AUPR, which verifies that graph representation learning can automatically produce effective representations for drug repositioning. (3) Our deep2CoV achieves the best performance on AUPR and achieves comparable performance on AUC. Our deep2Cov achieves worst performance than deepDR in AUC, the mainly reason is data imbalance. The effective drugs for diseases are far less than the number of ineffective drugs, and the AUC indicator is often invalid on the imbalanced data, but the AUPR indicator is more convincing.

C. SARS-CoV-2 Potential Drug Prediction

We focus on the top predicted drug-disease association which remains unrecorded in the dataset. We list the top 10 potential therapeutic associations for SARS-CoV-2 in Tab. II, and check up on these predictions according to literature and publications.

As shown in Tab. II, we find some evidence to confirm ten therapeutic associations, For example, One SARS-CoV-2 patient was treated with Ceftriaxone, and his cough, fever, and fatigue were improved gradually, he completed the 5-day course of hydroxychloroquine and azithromycin, 7 days of ceftriaxone [8]. As reported in [9], ciprofloxacin and glycyrrhizic acid were selected based on their reported antiviral activity, safety, availability, affordability and subjected for Molecular Dynamics (MD) simulation. The MD simulation results indicate that ciprofloxacin may be repurposed against SARS-CoV-2. In patients in the ICU, piperacillin was the most commonly prescribed antibiotic, and the piperacillin was used for the treatment of SARS-CoV-2 in the study [10]. The studies [11] demonstrated that the amphotericin B and caspofungin were added, assuming a potential synergistic effect with liposomal amphotericin B for treat SARS-CoV-2.

IV. CONCLUSION

In this paper, we have proposed a drug-disease-target triplet association prediction method (deep2CoV) based on the graph convolutional network for SARS-CoV-2. We first build a SARS-CoV-2 drug repositioning dataset based on the existing data (e.g., drugbank, repoDB). Furthermore, we propose deep2CoV which uses graph convolutional neural networks to learn heterogeneous networks to obtain low-dimensional feature representation. Then we predict SARS-CoV-2 potential drugs based on the feature representation. Comparing our model to baselines, we can find our model outperforming the others. Moreover, we have validated most of the prediction drugs by literature reference.

ACKNOWLEDGMENTS

This work is partially supported by grant from the Natural Science Foundation of China (No. 61976036, No. 61772103, No. 61632011)

REFERENCES

[1] M. P. Menden, F. Iorio, M. J. Garnett, U. Mcdermott, C. H. Benes, P. J. Ballester, and J. Saezrodriguez, "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties," *PLOS ONE*, vol. 8, no. 4, 2013.

[2] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, and F. Cheng, "deepdr: a network-based deep learning approach to in silico drug repositioning," *Bioinformatics*, vol. 35, no. 24, pp. 5191–5198, 2019.

[3] H. Luo, J. Wang, M. Li, J. Luo, P. Ni, K. Zhao, F.-X. Wu, and Y. Pan, "Computational drug repositioning with random walk on a heterogeneous network," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 6, pp. 1890–1900, 2018.

[4] B. Chen, Y. Ding, and D. J. Wild, "Assessing drug target association using semantic linked data," *PLOS Computational Biology*, vol. 8, no. 7, 2012.

[5] D. E. Gordon, G. M. Jang, M. Bouhaddou, and J. X. et al, "A sars-cov-2 protein interaction map reveals targets for drug repurposing." *Nature*, vol. 583, no. 7816, pp. 459–468, 2020.

[6] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *International conference on learning representations*, 2017.

[7] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data," *Molecular Pharmaceutics*, vol. 13, no. 7, pp. 2524–2530, 2016.

[8] M. Khalid, T. Dernaika, L. Jacob, P. Annamaraju, and A. K. Guddati, "Weather the cytokine storm: A report of successful management of a colon cancer survivor and a critically ill patient with covid-19," *Case Reports in Oncology*, pp. 754–759, 2020.

[9] W. Wang, J. Yu, Y. He, Z. Wang, and F. Li, "Ambroxol inhibits mucoid conversion of pseudomonas aeruginosa and contributes to the bactericidal activity of ciprofloxacin against mucoid p. aeruginosa biofilms." *Apmis*, vol. 124, no. 7, pp. 611–618, 2016.

[10] B. Beović, M. Doušak, J. Ferreira-Coimbra, K. Nadrah, F. Rubulotta, M. Belliato, J. Berger-Estilita, F. Ayoade, J. Rello, and H. Erdem, "Antibiotic use in patients with covid-19: a 'snapshot'infectious diseases international research initiative (id-iri) survey," *Journal of Antimicrobial Chemotherapy*, 2020.

[11] C. Poignon, M. Blaize, C. Vezinet, A. Lampros, A. Monsel, and A. Fekkar, "Invasive pulmonary fusariosis in an immunocompetent critically ill patient with severe covid-19," *Clinical Microbiology and Infection*, 2020.

[12] E. Francini, S. T. Miano, A. I. Fiaschi, and G. Francini, "Doxycycline or minocycline may be a viable treatment option against sars-cov-2," *Medical Hypotheses*, vol. 144, p. 110054, 2020.