Repositioning Molecules of Chinese Medicine to Targets of SARS-Cov-2 by Deep Learning Method

Tao Song[1,2], Yue Zhong[1], Mao Ding[3,*], Renteng Zhao[4],
Qingyu Tian[1], Zhenzhen Du[1], Dayan Liu[1], Jiali Liu[1], Yufeng Deng[1]
[1]College of Computer Science and Technology, China University of Petroleum(East China), Qingdao 266580, China
[2]Department of Artificial Intelligence, Faculty of Computer Science, Polytechnical University of Madrid,
Campus de Montegancedo, Boadilla del Monte 28660, Madrid, Spain
[3]Department of Intensive Care Unit, The Second Hospital of Shandong University, Jinan 250033, China
[4]Trinity earth Technology Co.Ltd
*Correspondence: Mao Ding, 18264181312@163.com
Received: date; Accepted: date; Published: date

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313151

***Abstract*—Traditional Chinese medicine has been used to treat and prevent infectious diseases for thousands of years, and has accumulated a large number of effective prescriptions. Deep learning methods provide powerful applications in calculating interactions between drugs and targets. In this study, we try to use the method of deep learning to reposition molecules of Chinese medicines (CMs) and the targets of syndrome coronavirus 2 (SARS-CoV-2). A deep convolution neural network with residual module (DCNN-Res) is constructed and trained on KIBA dataset. The accuracy of predicting the binding affinity of drug-target pairs is 85.33%. By ranking binding affinity scores of 433 molecules in 35 CMs to 6 targets of SARS-Cov-2, DCNN-Res recommends 30 possible repositioning molecules. The consistency between our result and the latest research is 0.827. The molecules in Gancao and Huangqin have a strong binding affinity to targets of SARS-CoV-2, which is also consistent with the latest research.**

***Index Terms*—SARS-CoV-2, Chinese medicine, deep convolutional neural network, drug reposition**

I. Introduction

Coronavirus disease 2019 (COVID-19) is an unprecedented global threat caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The sudden outbreak and accelerated spreading of SARS-CoV-2 infection have caused substantial public concerns [1]. At the time of writing this manuscript, about 5,867,771 new infections are reported daily. According to the World Health Organization (WHO), the COVID-19 outbreak has resulted in 55,928,327 confirmed cases including with 1,344,003 deaths (https://covid19.who.int).

There is still no vaccine or drug specifically for COVID-19. So, researchers place hope on drug repositioning to find effective drugs, such as Remdesivir. On the other side, researchers also hope to find Chinese medicines (CMs) to fight the virus. SARS-CoV-2 is closely related to the SARS-CoV virus. The CoV spike glycoprotein (S Protein) is a key target for vaccines, therapeutic antibodies, and diagnostics [2]. Envelope protein (E Protein) of coronaviruses is a structural protein existing in both monomeric and homo-pentameric form. It has been related to a multitude of roles including virus infection, replication, dissemination and immune response stimulation [3]. And the exonuclease activity of Nsp14 provides possible proofreading ability to RNA polymerase makes coronaviruses different from other RNA viruses allowing coronaviruses to maintain their relatively large genome size [4]. The viral 3-chymotrypsin-like cysteine protease (3CLPro) enzyme controls coronavirus replication [5] and researchers find that ACE2 could be the host receptor for the virus [6]. In addition, the membrane glycoprotein (M Protein) is also the main structural gene of the virus [7]. Therefore, these six proteins are selected as targets to reposition molecules of CMS.

Drug-target interaction is ideally used in the repositioning of existing drugs. It can be divided into two categories: using classification methods to predict probabilities of drug-target interaction (DTI) and regression methods to predict drug-target binding affinity values (DTA). Since it is difficult to determine the negative samples in the classification methods, as well as the drug-target binding information, such as dissociation constant (K_d), can be directly used in the prediction of DTA. Therefore, the prediction of DTA is more popular at present.

Deep learning methods provide powerful applications in calculating interactions between drugs and targets. The DeepDTA model [8] uses convolutional neural network to represent features of drug molecules and proteins to predict binding affinity. It is considered to be the best prediction model at present [9]. However, the loss of the model fluctuates greatly when training for many times. The DeepDR [10] model constructs multiple heterogeneous networks of drugs and diseases, and learns their feature representations through autoencoder, so as to reposition drugs for diseases. However, it does not consider the side effects of drugs and diseases. The GraphDTA model [11] is improved from the DeepDTA model, and uses graph convolutional neural network to represent the feature representations of drug molecules, but its calculation is too large. If a subtle model is well trained by a large scale of DT pairs data and obtains stable output, it can be used to

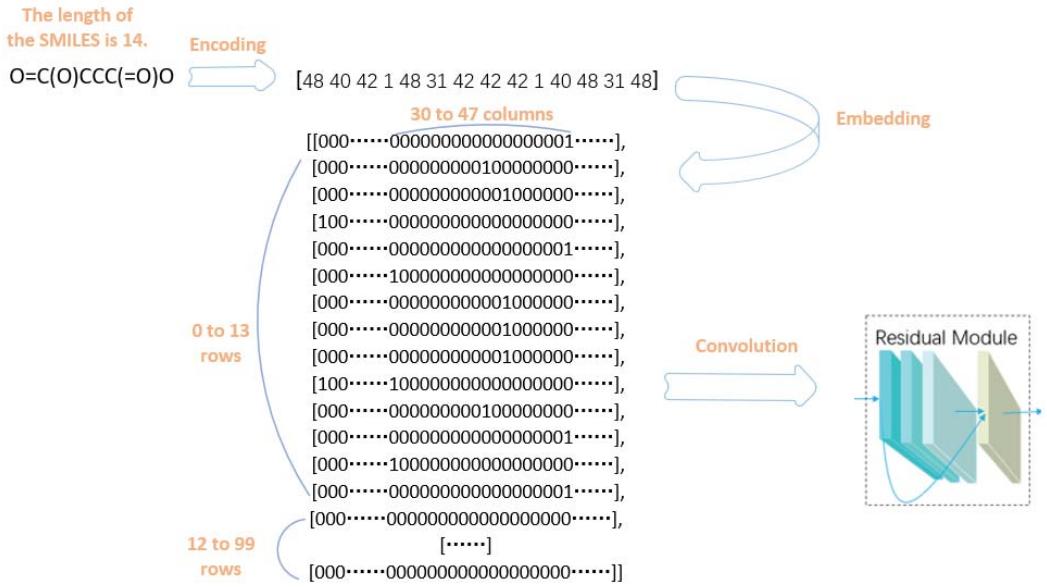

Fig. 1. Molecular encoding diagram. The SMILES sequence is encoded by labels, so that each character is represented by a unique integer. After that, it is expanded to a matrix of 100*64, which enters the convolutional layers learning the characteristic representations. The sequences of proteins are encoded in the same way.

predict the binding affinity values between drugs and specific targets. On the other hand, the prediction of the relationship between targets and drugs has been widely calculated by deep learning methods. However, the relationship between molecules of CMs and targets has not been studied widely. At present, most of the research on CMs still lies in the literature mining of traditional Chinese medicine prescriptions and CMs. For example, the latest SMGCN model [12] uses multi-graph convolution network to recommend CMs in syndromes.

In this study, we implement the deep learning approach to reposition CMs to six targets of COVID19, in order to find the effective molecules of CMs for the virus. Specifically, a deep convolution neural network with residual [13], [14] module (DCNN-Res) is constructed based on the affinity data of drug protein binding in KIBA dataset. The accuracy of DCNN-Res model is 85.33%, slightly lower than that of DeepDTA model (86.42%). However, the two models are compared with the high-frequency CMs in 576 prescriptions [15], the coverage of DeepDTA model is 79.7%, while that of DCNN-Res model is 86.7%. Gancao and Huangqin rank first and second in the recommended drugs respectively, which is also consistent with the frequency of CMs obtained from 576 prescriptions.

II. Materials and Methods

A. Data Source

KIBA dataset contains 229 proteins, 2111 drugs and binding affinities of 50,181 drug-protein pairs [8]. Binding affinities data of drug-protein pairs in KIBA are applied as training set. KIBA dataset comprehensively contains three indexes related to binding affinity: dissociation constant, inhibition constant, and the half maximal inhibitory concentration. KIBA method [16] is used to combine the three indexes to prevent the disturbance caused by data loss.

Ren et al. have obtained 40 kinds of CMs from 576 prescriptions with the key words of “Warm diseases (Wenbing)”, “Pestilence (Wenyi or Yibing)” or “Epidemic diseases (Shiyi)”, and select 35 CMs (including 433 molecules) with oral bioavailability (OB) $> 30\%$ and drug-likeness (DL) > 0.18 as candidate compounds for further analysis [15]. These traditional Chinese medicine molecules are stored in SDF format are transformed into CSV format, from which we get the SMILES sequence of each molecule for calculation. Meanwhile, target proteins in form of FASTA sequences are used for binding affinity calculations. The FASTA sequences of six targets (S Protein, E Protein, M Protein, Nsp14, 3CLPro and ACE2) are obtained from UniProt database.

B. Molecules Encoding

There are 64 unique characters in SMILES and 25 unique characters in FASTA sequences [8]. We take the SMILES sequence encoding of molecules as an example. Firstly, each character in molecules is mapper into a unique integer by using label encoding. In this way, a one-dimensional vector is obtained. In KIBA dataset, the length of SMILES sequences is concentrated in 0 to 100. So, the maximum length in the model of the SMILES sequence is set to 100. Then, the one-dimensional vector is stretched to a binary matrix of 100*64 by using “one-hot” encoding strategy. Each row represents a symbol in the SMILES sequence. It is encoded with only one significant bit set to ”1” and the rest of 63 positions are set to “0”. The encoding process of a molecule in Renshen is shown in Fig. 1.

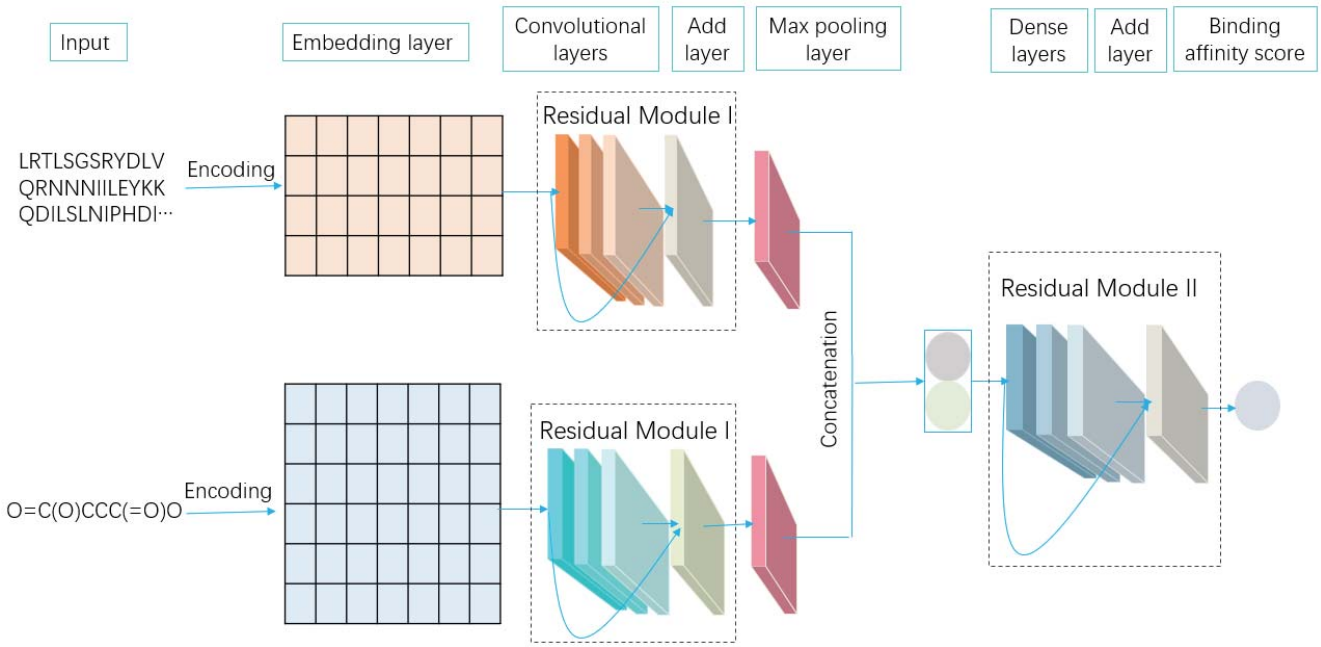

Fig. 2. The structure of DCNN-Res model.

As well, proteins in form of FASTA sequences can be encoded in a similar way. The difference is that the longest of protein sequences is 1000, and finally a matrix of 1000 * 25 is obtained. If the length of SMILES sequence is less than 100 (or FASTA sequence is less than 1000), the row of matrix is filled with "0". The encoding strategy is the one widely used in molecular information processing from [8].

C. DCNN-Res Model

The structure of DCNN-Res model is shown in Fig. 2. It has two tunnels for processing SMILES and FASTA sequences, each tunnel consists of residual module I and a max pooling layer. And residual module I is composed of three convolutional layers and an add layer. The activation function between convolutional layers is ReLU. While the results from the add layer are input to the max pooling layer after being activated by the ELU function. The first convolution kernel is 32, the second convolution kernel is 32*2, and the third convolution kernel is 32*3 [8]. And the loss function is measured by Mean Squared Error (MSE).

The max pooling layer compresses the learned features. The feature representations of drugs and proteins are spliced together, and enter into residual module II composed of three dense layers, two dropout layers and an add layer. The dropout layer is between two dense layers, and the third dense layer is followed by the add layer. Residual module II is responsible for nonlinear transformation of the extracted feature representations and extracting the correlation between features. The first two dense layers are followed by dropout layer (dropout rate is 0.1), which randomly "delete" hidden neurons to prevent over fitting, and finally map to the output space. The output of the model is a 433 * 6 matrix. Each element of the matrix represents the binding affinity score of the molecule and viral target.

III. Results

A. DCNN-Res Model Training

The training set is 50181 pairs of drug and protein relationship pairs in the KIBA dataset. The cross-validation method is used in order to find the most appropriate parameters for the model, rather than just a set of parameters. The training set is divided into five equal subsets, four of which are used as training set and one as validation set. Each subset is used in turn as the training and validation set. The 256 small batch data is used to update the weights of neural networks, as well Adam optimization algorithm (with learning rate is 0.001) is applied to optimize the model. And after 100 epochs, the result of binding affinity scores between molecules and the targets is generated.

In the training process, consistency index (CI) denoted in (1) and (2) is used to evaluate the training performance, and MSE [17] explained in (3) is used as the loss function to measure the error of each epoch.

$$CI = \frac{1}{Z} \sum_{(\delta_i > \delta_j)} h(b_i - b_j) \tag{1}$$

$$h(x) = \begin{cases} 1 & x > 0 \\ 0.5 & x = 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

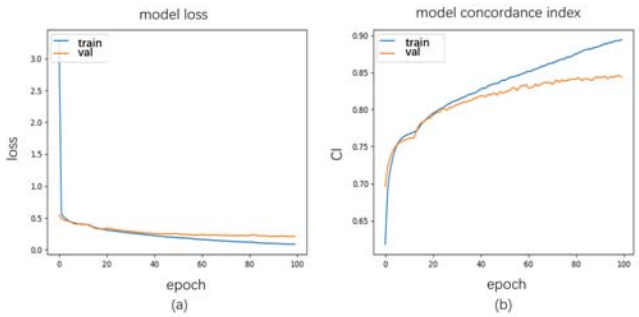

Fig. 3. Training optimal result graph: (a) Diagram of the loss of training set and validation set; (b) Diagram of the accuracy of training set and validation set.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{3}$$

where δ_i andδ_j represents the true binding affinity values of the i-th and j-th drug-protein relationship pairs respectively; while b_i and b_j represents the binding affinity prediction values of the i-th and j-th drug-protein relationship pairs respectively. It is denoted by Z the normalization constant, indicating the number of data pairs with different labels. Function h (x) is a segmented function, which measures the consistency between the predicted value and the real value according to the relationship between the predicted values of two groups of drug-protein relationships. If b_i is greater than b_j, then the value of function h (x) is 1. This indicates that the relationship between the predicted value is consistent with the actual one. If b_i is less than b_j, then h (x) is with value 0. It indicates that the predicted value is completely opposite to the actual value. If b_i is equal to b_j, then h (x) is 0.5, which means the prediction result is not completely wrong (but not consistent with the actual one). Meanwhile, n represents the number of training samples (i.e. 50,181 drug-protein interaction pairs), p_i represents the prediction vector of the i-th drug-protein relationship pair, y_i denotes the real vector of the i-th drug-protein relationship pair.

The model is trained five times in total. We select the group of values the parameters with the best accuracy (concordance index) in the validation set as the optimal parameters of the model, and the results are shown in Fig. 3.

The accuracy of the model is 0.8533, MSE is 0.197. DeepDTA is considered to be the state-of-the-art methods [14, 8]. The accuracy of DeepDTA model is 0.8624, MSE is 0.1965. The numerical results of the two models are equivalent. We compare the CI and MSE of DCNN-Res and DeepDTA model in five training processes with the best, worst and average performance. The results are shown in Table I. The CI of DCNN-Res model is slightly lower than DeepDTA method. Although the best MSE still appears in DeepDTA, the worst and average performance of MSE in DCNN-Res is better than DeepDTA, which also shows the advantage of residual module.

B. The Recommended Repositioning Regulatory Molecules to Targets

The inputs of DCNN-Res are the labels encoding of molecules of CMs in form of SMILES sequences and the targets of SARS-CoV-2 in form of FASTA sequences. These labels enter a residual module to learn feature representations. For repositioning regulatory molecules to the targets of SARS-Cov-2, the trained DCNN-Res model is used to calculate the binding affinity score of each potential drug-target pair. The binding affinity scores of 433*6 DT pairs are calculated.

By ranking binding affinity scores of all the potential DT pairs, we recommend six Chinese medicine molecules for each target protein, in addition to seven molecules recommended for the 3CLPro. After removing the repeated molecules, a total of 31 molecules are recommended, which are distributed in 18 CMs. The results are shown in Table II. Here we only give Compound CID of the molecules and the pinyin name and Latin name of Chinese medicine to which it belongs. The SMILES sequences of molecules can be obtained in PubChem database (https://pubchem.ncbi.nlm.nih.gov/) according to these given Compound CID.

Among the 18 CMs, there are 10 CMs which also appear in the top 20 list of CMs ranked by Ren [12]. And then, we calculate the frequency of each Chinese medicine in Table II. And CMs with the same frequency is classified into the same category (see Table III). In the ranking recommended by DCNN-Res model, Gancao and Huangqin are in the first and second place respectively, which is consistent with the results of Ren.

DeepDTA method is also used to recommend 35 Chinese medicine molecules and 18 CMs. The molecules are shown in Table IV, while the categories of CMs are shown in Table V.

In order to compare our results with those of Ren precisely., we use normalized discounted cumulative gain (nDCG) [18] method to calculate the similarity of recommendation list. We find the top 20 CMs in Ren's result, among which Shigao ranked 11 and Shexiang ranked 16 are excluded from our repositioning CMs set because they cannot be taken in large doses for a long time. Therefore, the two CMs are also removed in the calculation of similarity. According to our regulations, the relevance grade (rel) [18] of CMs corresponding to ranking which are recommended by DCNN-Res model and the top 20 Chinese medicine recommended by Ren is set to "4". The rel of CMs which exist in the top 20 CMs recommended by Ren, but the wrong ranking traditional Chinese medicine are set to "2". The rel of CMs which are invisible in the top 20 of Ren's result are set to "0". Here, the rels of 18 CMs recommended by DCNN-Res model are given, which are 4, 4, 0, 0, 2, 2, 0, 0, 2, 0, 0, 2, 0, 2, 0, 2, 2, 4.

Equation (4) is used to calculate the discounted cumulative gain (DCG). It aims to make the top results more influential. After that, 18 rels are arranged in descending order, and (5) is used to calculate the ideal discounted cumulative gain (IDCG). It is the maximum DCG value under ideal condition. Finally,

TABLE I
COMPARISON BETWEEN DCNN-RES AND DEEPDTA

Model	CI			MSE			nDCG
	Worst	*Average*	*Best*	*Worst*	*Average*	*Best*	
DeepDTA	**0.8568**	**0.8589**	**0.8624**	0.2327	0.2109	**0.1965**	0.797
DCNN-Res	0.8435	0.8476	0.8533	**0.2113**	**0.2072**	0.1970	**0.827**

Bold is the better one.

DCG is normalized to nDCG by (6).

$$DCG = \sum_{i=1}^{m} \frac{rel_i}{_l og_2(i+1)} \tag{4}$$

$$IDCG = \sum_{i=1}^{|REL|} \frac{REL_i}{_l og_2(i+1)} \tag{5}$$

$$nDCG = \frac{DCG}{IDCG} \tag{6}$$

where rel_i is the i-th rel value and m is the total number of rels. REL represents the set obtained by descending rels, while $|REL|$ indicates the number of non-zero elements in the REL set. REL_i represents the i-th rel value after descending.

The nDCG of DCNN-Res model is 0.827, while that of DeepDTA method is 0.797. Therefore, the result of DCNN-Res model is more similar to those of Ren. The result also shows that the deep learning network with residual module has better effect in practical problems. It is shown in Table I.

The biggest difference between the deep learning model and Ren's method is the ranking of Renshen. Ren ranked Renshen at 21. But in our results, its performance is similar to that of Huangqin. We consider that every molecule of CMs are chosen as the experimental object, rather than the complex compound.

TABLE II
RECOMMENDED REPOSITIONING REGULATORY MOLECULES TO TARGETS OF SARS-COV-2 USING DCNN-RES MODEL.

Target	Compound CID	Chinese name	Latin name[a]
S Protein	64982	Banxia	Pinelliae Rhizoma
	33934	Renshen	Ginseng Radix Et Rhizoma
	10207	Dahuang	Rhei Radix Et Rhizome
	10168	Dahuang	Rhei Radix Et Rhizome
	78407230	Dihuang	Rehmanniae Radix
	70698143	Baishao	Paeoniae Radix Alba
E Protein	21599928	Renshen	Ginseng Radix Et Rhizoma
	64982	Huangqin	Scutellariae Radi
	182232	Huangqin	Scutellariae Radi
	441913	Shengma	Cimicifugae Rhizoma
	3286789	Mahuang	Ephedra Herba
	443023	Huoxiang	Pogostemonis Herba
M Protein	285342	Renshen	Ginseng Radix Et Rhizoma
	44257530	Gancao	Glycyrrhizae Radix Et Rhizoma
	5280343	Huanglian	Coptidis Rhizoma
	5281781	Huoxiang	Pogostemonis Herba
	44575944	Zhimu	Anemarrhenae Rhizoma
	392442	Gancao	Glycyrrhizae Radix Et Rhizoma
Nsp14	101866715	Fangfeng	Saposhnikoviae Radix
	442534	Shengma	Cimicifugae Rhizoma
	5281617	Mahuang	Ephedra Herba2
	5280343	Zhizi	Gardeniae Fructus
	5281703	Danggui	Angelicae Sinensis Radix
	14135325	Huangqin	Scutellariae Radi
3CLPro	73402	Fuling	Poria
	15380912	Gancao	Glycyrrhizae Radix Et Rhizoma
	10336244	Gancao	Glycyrrhizae Radix Et Rhizoma
	5282184	Fangfeng	Saposhnikoviae Radix
	73299	Fuling	Poria
	5280863	Baishao	Paeoniae Radix Alba
	5280863	Zhizi	Gardeniae Fructus
ACE2	5280343	Huanglian	Coptidis Rhizoma
	5317652	Gancao	Glycyrrhizae Radix Et Rhizoma
	5280343	Huoxiang	Pogostemonis Herba
	443758	Jiegeng	Platycodonis Radix
	5280343	Gancao	Glycyrrhizae Radix Et Rhizoma
	10212	Baizhi	Angelicae Dahuricae Radix

[a] The Latin names by Chinese Pharmacopoeia (2015 Edition).

TABLE III
CLASSIFICATION TABLE OF CMS RECOMMENDED BY DCNN-RES MODEL.

Category	Chinese name of CMs	Frequency
1	**Gancao**	5
2	**Huangqin**	3
2	Renshen	3
2	Huoxiang	3
3	*Dahuang*	2
3	*Baishao*	2
3	Shengma	2
3	Mahuang	2
3	*Huanglian*	2
3	Zhizi	2
3	Fuling	2
3	*Fangfeng*	2
4	Banxia	1
4	*Dihuang*	1
4	Zhimu	1
4	*Danggui*	1
4	*Jiegeng*	1
4	**Baizhi**	1

Bold indicates that the ranking is consistent with Ren's results.
Italics indicates presence in the top 20 of Ren.

IV. DISCUSSION

In the public health emergencies like COVID-19, the public place high hopes for traditional Chinese medicine. The combination of CMs and deep learning is a new attempt of our team.

In this study, a deep learning model DCNN-Res is constructed and trained by approved binding affinities from KIBA dataset. The well-trained AI model are then used to predict the binding affinity of 433 Chinese medicine molecules with 6 targets of SARS-CoV-2. The recommended result of DCNN-Res models is compared with those of Ren using complex network screening. We find that Gancao and Huangqin both performed well in these models. Meanwhile, our results also show that CMs play a therapeutic role in multiple targets. In the following work, we will implement molecular docking of the recommended molecules and conduct biological experiments on the selection of effective molecules. Besides, merged LSTM neural networks [19] is expected to encode drugs and proteins. Membrane computing [20]–[22] and DNA encoding [23]–[25] can be considered in future research to explore the mystery of drug molecules acting on the human body as well.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (Grant Nos. 61873280, 61873281, 61672033, 61672248, 61972416), Taishan Scholarship (tsqn201812029), Natural Science Foundation of Shandong Province (No. 2019GGX101067, ZR2019MF012), Fundamental Research Funds for the Central Universities (18CX02152A, 19CX05003A-6), Foundation of Science and Technology Development of Jinan(201907116).

REFERENCES

[1] B. Shen, X. Yi, Y. Sun, X. Bi, J. Du, C. Zhang et al., "Proteomic and metabolomic characterization of COVID-19 patient sera," Cell, vol. 182, pp. 59-72, July 2020.

[2] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona et al., "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," Sci, vol. 367, pp. 1260-1263, March 2020.

[3] B. Tiloccaa, A. Soggiubc, M. Sanguinettide, G. Babinif, F. D. Maiode, D. Brittia et al., "Immunoinformatic analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19," Microbes Infect, vol. 22, pp. 182-187, May 2020.

[4] J. Jain, S. Gaur, Y. Chaudhary, and R. Kaul, "The molecular biology of intracellular events during Coronavirus infection cycle," Disease, May 2020.

[5] M. T. Qamar, S. M. Alqahtani, M. A. Alamri, and L. Chen, "Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants," J Pharm Anal, 2020.

[6] Y. Cao, L. Li, Z. Feng, S. Wan, P. Huang, X. Sun et al., "Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations," Cell Discov, vol. 6, February 2020.

[7] T. P. Velavan, and C. G. Meyer, "The COVID-19 epidemic," TROP MED INT HEALTH, vol. 25, pp. 278-280, February 2020.

[8] Ö. Hakime, Ö. Arzucan, and Ö. Elif, "DeepDTA: Deep Drug-Target Binding Affinity Prediction," Bioinformatics, vol. 34, January 2018.

[9] K. Huang, T. Fu, C. Xiao, L. M. Glass, and J. Sun, " DeepPurpose: a Deep Learning Based Drug Repurposing Toolkit," unpublished.

[10] 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, May 2019.

[11] T. Nguyen, H. Le, and S. Venkatesh, "GraphDTA: prediction of drug–target binding affinity using graph convolutional networks," unpublished.

[12] Y. Jin, W. Zhang, X. He, X. Wang, and X. Wan, "Syndrome-aware Herb Recommendation with Multi-Graph Convolution Network," unpublished.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.

[14] S. Pang, S. Qiao, T. Song, J. Zhao, and Z. Pan, "An Improved Convolutional Network Architecture Based on Residual Modeling for Person Re-Identification in Edge Computing," IEEE Access, 2019.

[15] X. Ren, X.Shao, X. Li, X.Jia, T. Song, W, Zhou et al., "COVID-19 from Traditional Chinese Medicine (TCM) by using a data-driven approach," J Ethnopharmacol, vol. 258, April 2020.

[16] J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg et al., "Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis," J Chem Inf Modelm. vol. 54, pp. 735-743, February 2014.

[17] S. Kansal, P. P. Bansod, and A. Kumar, " Prediction of instantaneous heart rate using adaptive algorithms," INT J ADAPT CONTROL, vol. 2, pp. 267-281, 2019.

[18] E. Yilmaz, E. Kanoulas, and J. A. Aslam, " A Simple and Efficient Sampling Method for Estimating AP and NDCG," Proc of Sigir, pp. 603-610, July 2008.

[19] T. Song, J. Jiang, W. Li, and D. Xu, "A Deep Learning Method with Merged LSTM Neural Networks for SSHA Prediction," IEEE J-STARS, 2020.

[20] T. Song, L. F. Macías-Ramosb, L. Pan, and M. J. Pérez-Jiménez, "Time-free Solution to SAT Problem Using P Systems with Active Membranes," THEOR COMPUT SCI, vol. 529, pp. 61-68, Aprial 2014.

[21] J. He, J. Xiao, X. Shi, and T. Song, "A Membrane-Inspired Algorithm with Memory Mechanism for Knapsack Problem," Journal of Zhejiang Univ-Sci C, pp. 612-622, 2013.

[22] B. Song, T. Song, and L. Pan, "Time-free solution to SAT problem by P systems with active membranes and standard cell division rules," NAT COMPUT, vol. 14, pp. 673-681, 2015.

[23] X. Li, H. Li, T. Song, A. Rodríguez-Patón, C. Chen, H. Zhao, and X. Shi, "Highly biocompatible drug-delivery systems based on DNA nanotechnology," J BIOMED NANOTECHNOL, vol. 13, pp. 745-757, 2017.

[24] X. Li, T. Song, Z. Chen, X. Shi, C. Chen, and Z. Zhang, "A Universal Fast Colorimetric Method for DNA Signal Detection with DNA Strand Displacement and Gold Nanoparticles," J NANOMATER, 2016.

[25] Z. Zhang, X. Zeng, T. Song, Z. Chen, X. Wang, and Y.Ye, "WormStep: An Improved Compact Graphical Representation of DNA Sequences Based on Worm Curve," J COMPUT THEOR NANOS, vol. 10, pp. 189-193, 2013.

TABLE IV
RECOMMENDED REPOSITIONING REGULATORY MOLECULES TO TARGETS OF SARS-COV-2 USING DEEPDTA METHOD.

Target	Compound CID	Chinese name	Latin name[a]
S Protein	15226717	Fuling	Poria
	5280863	Gancao	Glycyrrhizae Radix Et Rhizoma
	5271991	Huangqin	Scutellariae Radi
	5742590	Renshen	Ginseng Radix Et Rhizoma
	72703	Huanglian	Coptidis Rhizoma
	4166098	Lianqiao	Forsythiae Fructus
E Protein	91510	Gancao	Glycyrrhizae Radix Et Rhizoma
	69502	Baizhi	Angelicae Dahuricae Radix
	5281331	Tianhuafen	Trichosanthis Radix
	638072	Zhizi	Gardeniae Fructus
	5280794	Renshen	Ginseng Radix Et Rhizoma
	5280343	Mahuang	Ephedra Herba
M Protein	5319252	Chaihu	Bupleuri Radix
	5312521	Gancao	Glycyrrhizae Radix Et Rhizoma
	19009	Huanglian	Coptidis Rhizoma
	91462	Xuanshen	Scrophulariae Radix
	98608	Baizhi	Angelicae Dahuricae Radix
	439533	Mahuang	Ephedra Herba
Nsp14	5321865	Huangqin	Scutellariae Radi
	10715163	Chantui	Cicadae Periostracum
	6437642	Chantui	Cicadae Periostracum
	5318962	Fangfeng	Saposhnikoviae Radix
	156992	Huangqin	Scutellariae Radi
	5318980	Zhimu	Anemarrhenae Rhizoma
3CLPro	107876	Dahuang	Rhei Radix Et Rhizome
	389888	Banxia	Pinelliae Rhizoma
	17897	Baizhi	Angelicae Dahuricae Radix
	21599928	Renshen	Ginseng Radix Et Rhizoma
	268208	Gancao	Glycyrrhizae Radix Et Rhizoma
	5486699	Chaihu	Bupleuri Radix
ACE2	5318679	Gancao	Glycyrrhizae Radix Et Rhizoma
	10715163	Banxia	Pinelliae Rhizoma
	71629	Mahuang	Ephedra Herba
	10212	Zhizi	Gardeniae Fructus
	101577840	Shengma	Cimicifugae Rhizoma
	33934	Renshen	Ginseng Radix Et Rhizoma

[a] The Latin names by Chinese Pharmacopoeia (2015 Edition).

TABLE V
CLASSIFICATION TABLE OF CMS RECOMMENDED BY DEEPDTA METHOD.

Category	Chinese name of CMs	Frequency
1	**Gancao**	5
2	Renshen	4
3	*Huangqin*	3
3	*Baizhi*	3
3	Mahuang	3
4	**Chaihu**	2
4	Zhizi	2
4	Banxia	2
4	*Huanglian*	2
4	Chantui	2
5	*Dahuang*	1
5	Fuling	1
5	Zhimu	1
5	Lianqiao	1
5	Tianhuafen	1
5	**Xuanshen**	1
5	**Fangfeng**	1
5	Shengma	1

Bold indicates that the ranking is consistent with Ren's results.
Italics indicates presence in the top 20 of Ren.