



LETTER OPEN

TCMBank-the largest TCM database provides deep learning-based Chinese-Western medicine exclusion prediction

Signal Transduction and Targeted Therapy (2023)8:127

; <https://doi.org/10.1038/s41392-023-01339-1>**Dear Editor,**

In the modernization of traditional Chinese medicine (TCM), two key aspects are determining the active ingredients in herbs and elucidating the mechanism of action between the active ingredients and targets. The construction of a comprehensive and highly-reliability TCM database is highly desirable.

Since its establishment in 2011, our TCM Database@Taiwan¹ has been used extensively and heavily cited, and it also has been included in the ZINC database.² Using natural language processing, we set up a knowledge graph and molecular signaling pathways to establish a TCM database, TCMBank (<https://TCMBank.cn/>), which extends from TCM Database@Taiwan and includes 9192 herbs, 61,966 ingredients, 15,179 targets, and 32,529 diseases. The updated TCMBank expanded the number of herbal ingredients from 32,364 to 61,966 (unduplicated), and two new data fields, targets, and diseases, have been added. The number of herbs with connection information is 9010, and the average number of connection edges of herbs is 16.05. The number of ingredients with connection information is 54,676, and the average number of connection edges of herbs is 5.26. TCMBank provides 3D structures of herbal ingredients in mol2 format and provides cross-reference links to external public databases, such as CAS, DrugBank, PubChem, MeSH, OMIM, DO, ETCM,³ HERB,⁴ etc. At present, TCMBank is the most comprehensive, downloadable, and largest non-commercial TCM database, and comparisons of data size between TCMBank and other TCM-related databases can be viewed in Fig. 1a. TCMBank provides a convenient website for users to freely explore the relationship between herbs, ingredients, gene targets, and related pathways or diseases (Fig. 1b). Figure 1c shows the process of establishing the TCMBank, including text mining strategy, intelligent document identification module, etc. All TCM-related information must be manually verified by volunteers at least twice to ensure the reliability of TCMBank data.

Adverse reactions between Chinese-Western medicines can lead to increased medical costs and even death. It is estimated that more than 10% of patients need to take five drugs at the same time, and 20% of elderly patients need to take at least ten drugs at the same time, which greatly increases the medical risk caused by the mutual exclusion of Chinese-Western medicines. The identification of mutually exclusive reaction of Chinese-Western medicines mainly relies on biochemical assays in clinical. However, this process is very manpower and material consuming.

AI-based prediction of mutual exclusion of Chinese-Western medicines requires a large number of pairs of Chinese medicine and Western medicine with adverse reaction labels. There is a lack of mutual exclusion datasets for Chinese-Western medicines, while there are currently two real-world public drug-drug interactions (DDI) datasets: DrugBank and TWOSIDES. In previous works, we first proposed two models, 3DGT-DDI⁵ and SA-DDI,⁶ on the DDI datasets to predict the interaction between the two compounds.

Supplementary Tables S1-S6 shows that 3DGT-DDI and SA-DDI achieve state-of-the-art performance on two public DDI datasets. Then, we extended the prediction results of the above two models to the prediction of mutual exclusion of Chinese-Western medicines. TCMBank provides the world's largest herb-ingredient-target-disease mapping information. Benefiting from the big data drive of TCMBank, we used the DDI model to predict the mutual exclusion of Chinese-Western medicines for unsupervised learning. For a pair of traditional Chinese medicine and Western medicine, we query the active ingredients contained in TCM according to TCMBank. Assuming that all ingredients in the TCM do not have adverse reactions with Western medicine, it is determined that there is no mutually exclusive reaction between them. If one or more ingredients in the TCM have adverse reactions with Western medicine, they have a mutually exclusive reaction. In this way, we use an AI-assisted DDI prediction model to produce the prediction results of the mutual exclusion of Chinese-Western medicine.

The prediction results of the AI-assisted model have not been verified by actual clinical or biochemical tests. In the future, we will combine AI-assisted models for mutual exclusion prediction of Chinese-Western medicines, NLP and knowledge graph technology in text mining to develop a comprehensive database of combined Chinese-Western medicines. We will use IDIM module to search the mutual exclusion reaction of Chinese-Western medicine predicted by an AI-assisted model, and download, analyze the literature. Knowledge graph, keyword extraction and automatic summarization will be used to assist researchers to manually check the mutually exclusive information of Chinese-Western medicine contained in the literature. We will publish a comprehensive database of combined Chinese-Western medicines, which is a future work.

Another interesting future study will be to predict the mutually exclusive reaction of a group of multiple (more than two) Chinese-Western medicines. In the real world, the patient obviously intakes many more than two TCM or western medicine. This will require the development of new algorithms to consider the mutual exclusion of multiple drug combinations. Based on knowledge of medicinal chemistry, a drug is an entity composed of different functional groups/chemical substructures that determine their pharmacokinetic, pharmacodynamic properties, and the mutual exclusion of Chinese-Western medicine. We think that the interaction of substructure is regarded as the causal relationship of the interaction of Chinese-Western medicine, so as to establish a network of drug interactions or a network of interactions between multiple drugs (Fig. 1d), in which compounds as nodes and their causal relationships as edges. The nodes corresponding to all the ingredients in a TCM form a sub-network. We will predict whether TCM or Western medicine has mutual exclusion reaction according to whether there are edges between their corresponding sub-networks (Fig. 1e). Details of possible causal learning models are described in supplementary materials.

Received: 28 August 2022 Revised: 8 January 2023 Accepted: 28 January 2023

Published online: 31 March 2023

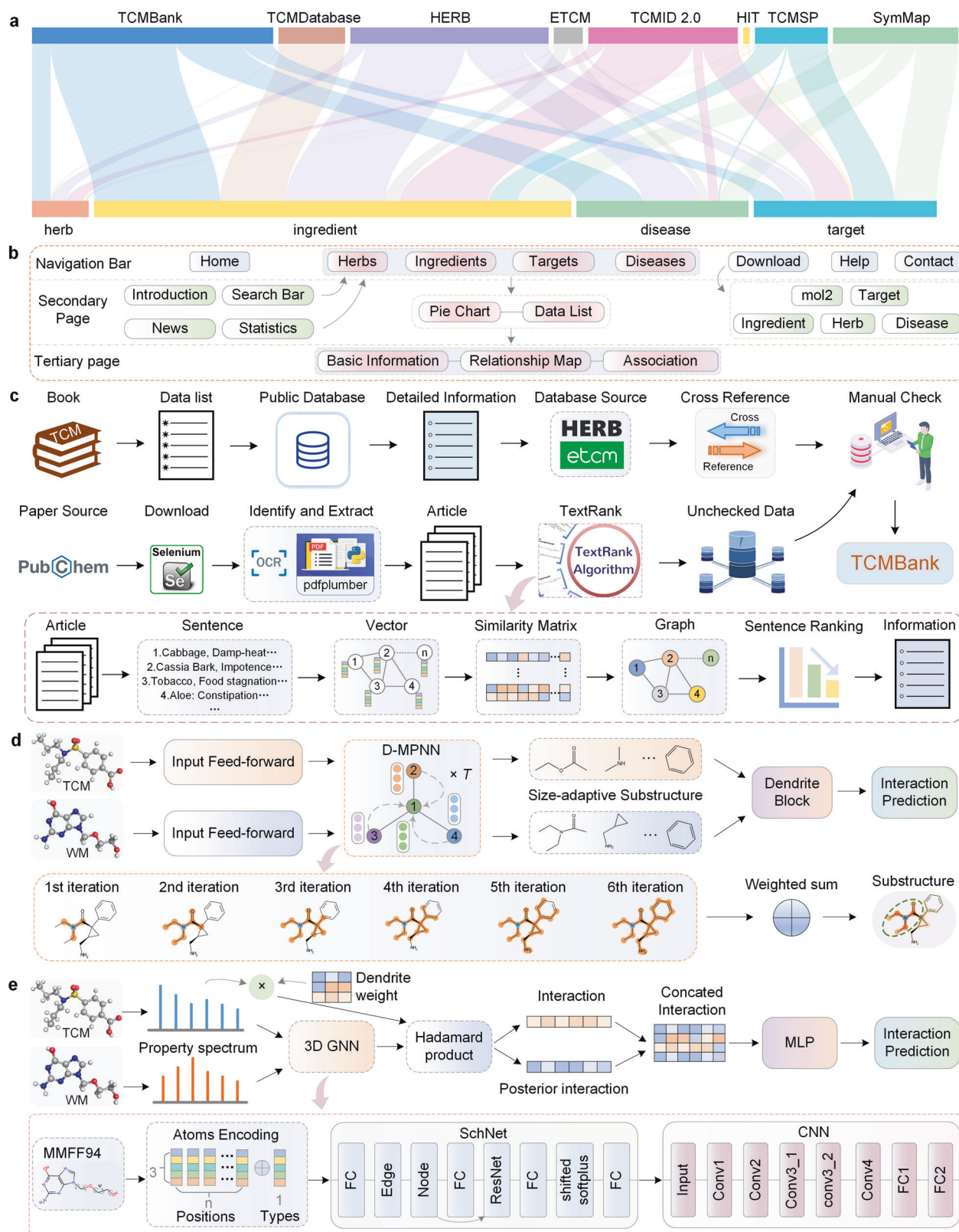


Fig. 1 Comprehensive analysis of TCMBank, the largest database of traditional Chinese medicine. **a** The comparison of data sizes between TCMBank and other TCM-related databases, where TCMBank is the richest at herb, ingredient and disease. **b** The composition of TCMBank website, including the navigation bar, the home page, secondary page, and tertiary page. **c** A schematic diagram of the data processing framework and objectives in TCMBank. **d** Schematic diagram of adaptive substructure-aware module based on graph neural network for drug functional group extraction. **e** Mutual exclusion prediction of Chinese-Western medicines based on causal learning. TCM traditional Chinese medicine, WM western medicine, D-MPNN direct message passing neural network, 3D GNN three-dimensional graph neural network, MLP multilayer perceptron, MMFF94 Merck molecular force field, 1994 version, CNN convolutional neural network

We developed TCMBank (<https://TCMBank.cn/>) to aggregate earlier studies dispersed in various forms of sources and create a comprehensive and reliable information system for Chinese medicine. TCMBank enables research on the molecular mechanism of herbal medicine and promotes the discovery of new drug molecules and corresponding potential molecular targets. The advantages of TCMBank include: (1) TCMBank is currently the largest downloadable and non-commercial database. (2) TCMBank provides up-to-date TCM-related information through continuous updates of the intelligent document recognition module. (3) TCMBank provides a large amount of herb/ingredient information with properties, physical and chemical properties, and 3D structure, as well as its target/disease information. We hope that TCMBank can meet the increasing needs for data resources related to TCM modernization and provide strong support for future advancement in the modernization of TCM.

DATA AVAILABILITY

TCMBank database is available at <https://TCMBank.cn/>. Any other information required to reanalyze the data reported in this paper are available upon request.

ACKNOWLEDGEMENTS

We acknowledge the help of Hsin-Yi Chen, Shiyue Cheng, Zhicheng Yang, Liwei Yang, Rui Huang, Hao Liu, Hui Peng, Binghao Cheng, and Mingqing Lin for their hard work in data processing and website construction. This work was supported by the National Natural Science Foundation of China (Grant No. 62176272).

AUTHOR CONTRIBUTIONS

L.Q. and C.Y.-C.C. conceived and supervised the project. L.Q., G.C., and H.H. performed data analysis. Z.Y. and L.Z. interpreted the results. L.Q., G.C., H.H., K.Z., and C.Y.-C.C. wrote the paper with input from all the other authors. All authors have read and approved the article.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41392-023-01339-1>.

Competing interests: The authors declare no competing interests. K.Z. is one of the Editors-in-Chief of Signal Transduction and Targeted Therapy, but he has not been involved in the process of the manuscript handling.

Qiuji Lv¹, Guanxing Chen¹, Haohuai He¹, Ziduo Yang¹, Lu Zhao², Kang Zhang³ and Calvin Yu-Chian Chen¹✉

¹Artificial Intelligence Medical Research Center, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China; ²The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China and ³Center for Innovations and Biomedicine, Faculty of Medicine, Macao University of Science and Technology, Macao, China
These authors contributed equally: Qiuji Lv, Guanxing Chen, Haohuai He

Correspondence: Calvin Yu-Chian Chen (chenyuchian@mail.sysu.edu.cn)

REFERENCES

- Chen, C. Y.-C. TCM Database@ Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE* **6**, 15939 (2011).
- Irwin, J. J. et al. Zinc20-a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).
- Xu, H.-Y. et al. ETCM: an encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res.* **47**, D976–D982 (2019).
- Fang, S. et al. HERB: a high-throughput experiment-and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res.* **49**, D1197–D1206 (2021).
- He, H., Chen, G. & Chen, C. Y.-C. 3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction. *Brief. Bioinf.* **23**, 134 (2022).
- Yang, Z., Zhong, W., Lv, Q. & Chen, C. Y.-C. Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network. *Chem. Sci.* **13**, 8693–8703 (2022).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023