

RESEARCH

Open Access



# Developing an artificial intelligence method for screening hepatotoxic compounds in traditional Chinese medicine and Western medicine combination

Zhao Chen<sup>1†</sup>, Mengzhu Zhao<sup>1†</sup>, Liangzhen You<sup>1†</sup>, Rui Zheng<sup>1</sup>, Yin Jiang<sup>1</sup>, Xiaoyu Zhang<sup>1</sup>, Ruijin Qiu<sup>1</sup>, Yang Sun<sup>1</sup>, Haie Pan<sup>1</sup>, Tianmai He<sup>1</sup>, Xuxu Wei<sup>1</sup>, Zhineng Chen<sup>2</sup>, Chen Zhao<sup>3\*</sup> and Hongcai Shang<sup>1\*</sup>

## Abstract

**Backgrounds:** Traditional Chinese medicine and Western medicine combination (TCM-WMC) increased the complexity of compounds ingested.

**Objective:** To develop a method for screening hepatotoxic compounds in TCM-WMC based on chemical structures using artificial intelligence (AI) methods.

**Methods:** Drug-induced liver injury (DILI) data was collected from the public databases and published literatures. The total dataset formed by DILI data was randomly divided into training set and test set at a ratio of 3:1 approximately. Machine learning models of SGD (Stochastic Gradient Descent), kNN (k-Nearest Neighbor), SVM (Support Vector Machine), NB (Naive Bayes), DT (Decision Tree), RF (Random Forest), ANN (Artificial Neural Network), AdaBoost, LR (Logistic Regression) and one deep learning model (deep belief network, DBN) were adopted to construct models for screening hepatotoxic compounds.

**Result:** Dataset of 2035 hepatotoxic compounds was collected in this research, in which 1505 compounds were as training set and 530 compounds were as test set. Results showed that RF obtained 0.838 of classification accuracy (CA), 0.827 of F1-score, 0.832 of Precision, 0.838 of Recall, 0.814 of area under the curve (AUC) on the training set and 0.767 of CA, 0.731 of F1, 0.739 of Precision, 0.767 of Recall, 0.739 of AUC on the test set, which was better than other eight machine learning methods. The DBN obtained 82.2% accuracy on the test set, which was higher than any other machine learning models on the test set.

**Conclusion:** The DILI AI models were expected to effectively screen hepatotoxic compounds in TCM-WMC.

**Keywords:** AI, DILI, Machine learning, Deep learning, TCM-WMC, Safety assessment

## Introduction

Drug-induced liver injury (DILI), one of the commonest and serious adverse drug reactions, is the dominant cause for terminating clinical trials or withdrawing new drugs [1]. Statistical analysis showed that 15 of the 47 withdrawn drugs were due to hepatotoxicity, accounting for 31.9% [2]. An observational cohort study showed that DILI was the main cause for acute liver injury (ALF),

<sup>†</sup>Zhao Chen, Mengzhu Zhao and Liangzhen You contributed equally to this work

\*Correspondence: zyyzc2007@163.com; shanghongcai@126.com

<sup>1</sup> Key Laboratory of Chinese Internal Medicine of Ministry of Education, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China  
<sup>3</sup> Present Address: Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China  
Full list of author information is available at the end of the article



accounting for 11% in the USA between 1998 and 2013 [3]. The incidence rate of DILI is 23.8 per 100 thousand people in China [4], attracted more research attentions.

The pathogenesis of DILI is not completely clear. Many risk factors lead to DILI, such as the role of host genetic, immunologic, and metabolic factors as well as drug and environmental effects [5, 6]. Although many methods have been developed to predict and assess the risk of DILI, mature and highly accurate methods are still insufficient. Classic methods, including experiments in vitro and in vivo, played a crucial role in predicting DILI. However, drug combinations have become common as multiple diseases coexisting. TCM-WMC have also been increasingly recognized [7–9], covering almost all clinical therapeutic areas in China [10], which increased the complexity of compounds ingested. Hence, it's essential to develop an effective method to screen hepatotoxicity of compounds in TCM-WMC.

With the development of new computing technologies, artificial intelligence (AI) models have been widely used in cheminformatics [11], medical imaging [12], diagnostics [13], bioinformatics [14], and other fields, provided new ideas for screening hepatotoxic compounds. Machine learning and deep learning methods have been increasingly applied to screen hepatotoxic compounds, which treat the high-dimensional chemical structure information as vectors and calculate for prediction or classification purpose in an efficient way [15–17]. In this study, DILI dataset is collected to establish AI models for screening hepatotoxic compounds in TCM-WMC.

## Methods

### Method design

DILI dataset was collected from public databases and published literatures. Nine machine learning models and a deep learning model were constructed with combined DILI dataset. A better performance model would be chosen to screen hepatotoxic compounds in TCM-WMC.

### DILI dataset collection

The compounds in DILI combined dataset were retrieved from the DILIrank [18], LiverTox [19], LTKB [20], Hepatox [21]. The annotations in DILIrank were assigned four different severity classes by considering DILI-related market withdrawals and warnings [18]. LiverTox contains comprehensive and evidence-based information on drug, dietary supplement, and herbal-induced liver injury [19]. Liver Toxicity Knowledge Base (LTKB) contains drugs whose potential to cause DILI in humans using the FDA-approved prescription drug labels [20]. Hepatox is a data base on the hepatotoxic drugs file published every year in Gastroentérologie Clinique et Biologique [21]. The keywords of "liver damage", "Drug-induced

liver injury (DILI)", "hepatotoxicity", "liver toxicity", "liver failure", "liver injury", "hepatitis", "jaundice", "cholestasis", "liver protection", "hepatoprotective", "hepatoprotection", "Herb-induced liver injury (HILI)" were searched in PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>), Nature(<https://www.nature.com/>), Science Online (<http://www.sciencemag.org/>), Elsevier Science Direct (<https://www.Sciencedirect.com>), Springer (<https://link.springer.com/>), Wiley (<https://onlinelibrary.wiley.com/>), Oxford Academic (<https://academic.oup.com/journals/>) and other publishers' databases to search the relevant literatures with DILI dataset. The search time was limited to 1999–2021. Duplicates from different sources and compounds without structures were excluded.

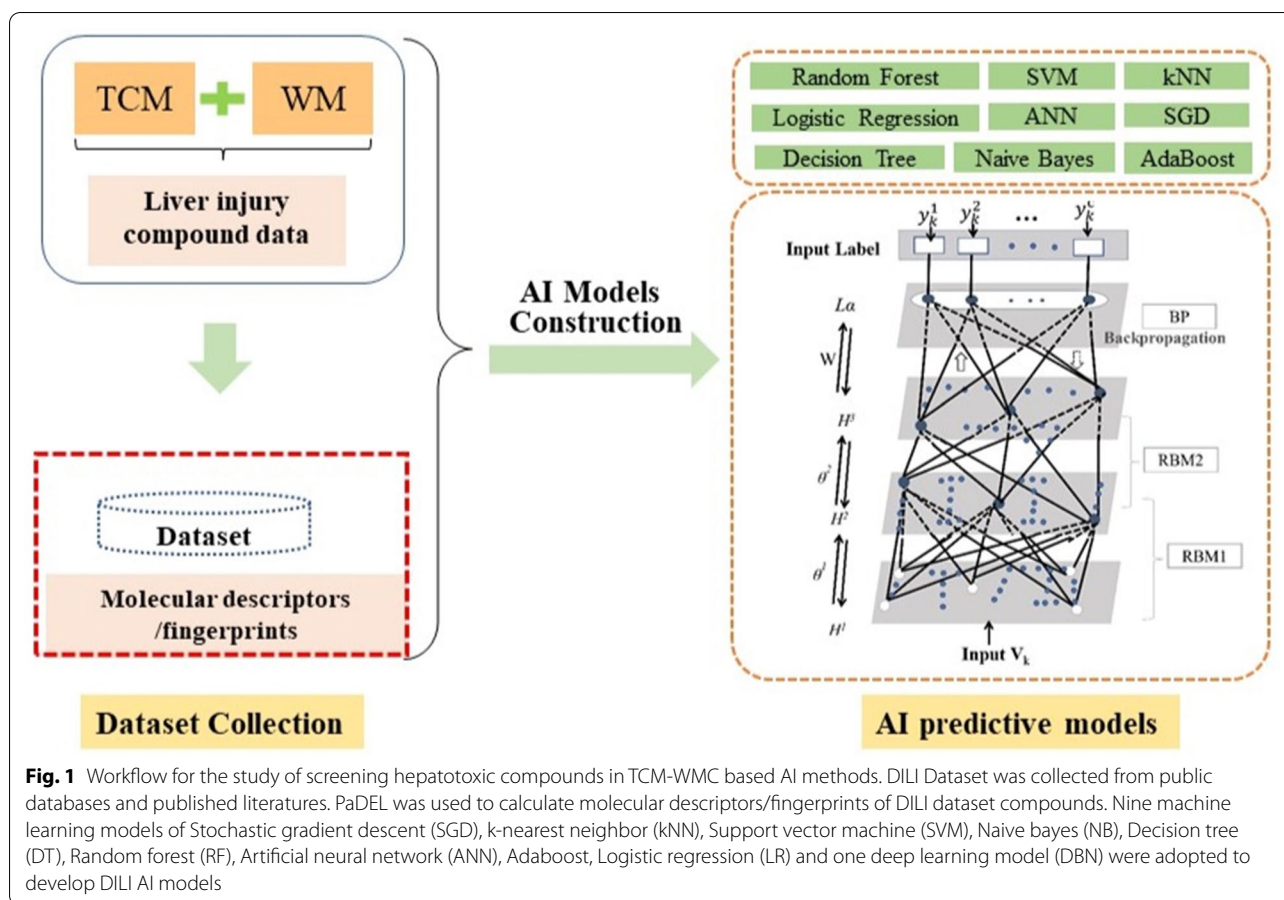
### AI model construction

Chemical structures of compounds were coded with SMILES (simplified molecular input line entry system). PaDEL-Descriptor software [22] was used to calculate the molecular descriptor and fingerprint of each compound based on SMILES string. PaDEL-1D and 2D descriptors of all compounds were calculated using PaDEL-Descriptor software (Yap, 2011). PaDEL-1D and 2D contained 1444 descriptors, including atom type electrotopological state (Estate) descriptors, Crippen's logP, and molecular linear free energy.

The machine learning (ML) methods of SGD (Stochastic Gradient Descent), kNN (k-Nearest Neighbor), SVM (Support Vector Machine), NB (Naive Bayes), DT (Decision Tree), RF (Random Forest), ANN (Artificial Neural Network), Adaboost, LR (Logistic Regression) were adopted to build liver injury AI models. Two restricted Boltzmann machines (RBM) of deep belief network (DBN) were also constructed in this research. All these AI methods were trained on the same dataset, which was randomly divided into training set and test set at a ratio of 3:1 approximately. The workflow for the study of screening hepatotoxic compounds in TCM-WMC based AI methods was showed in Fig. 1.

### Statistics for model evaluation measures

Five important model evaluation measures for ML methods, including classification accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), F1 score (Eq. 4), and area under the curve (AUC) of receiver-operating characteristic (ROC) were applied to assess the performance of each model. Therein, AUC represented the area under the ROC (Receiver operating characteristic) curve and the coordinate axis, CA represented the classification accuracy, Precision was how close the measured values that were to each other, Recall represented the recall rate. And the calculation formula of F1 score was as the Eqs. (4).



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

## Results

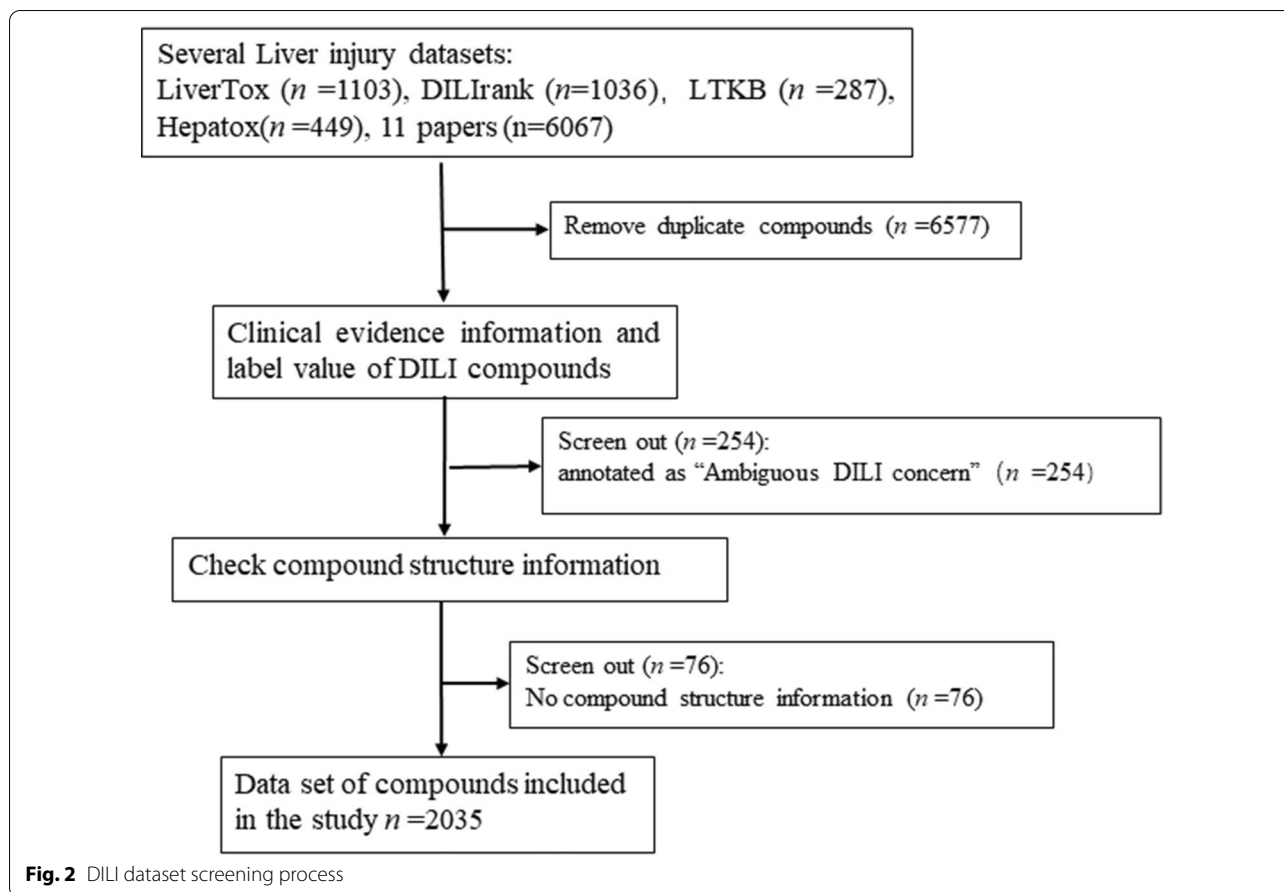
Two thousand eight hundred and seventy-five compounds were obtained from 4 DILI-related databases, and 6067 compounds were from 11 datasets in published DILI-related literatures [23–33]. After excluding the duplicate compounds, 2365 compounds was obtained. Subsequently, 254 drugs were also excluded annotated as “Ambiguous DILI concern” in DILIRank. Drugs without structure information were excluded after checking their

structure information. At last, a total of 2035 liver injury compounds were collected in this research, as shown in Fig. 2.

In the total dataset, 1505 compounds were as the training set and 530 compounds were as the test set. The ratio of the number of liver injury and non-liver injury compounds in the training set was 1125:400, and the number ratio of liver injury and non-liver injury compounds in the test set was 380:130, respectively.

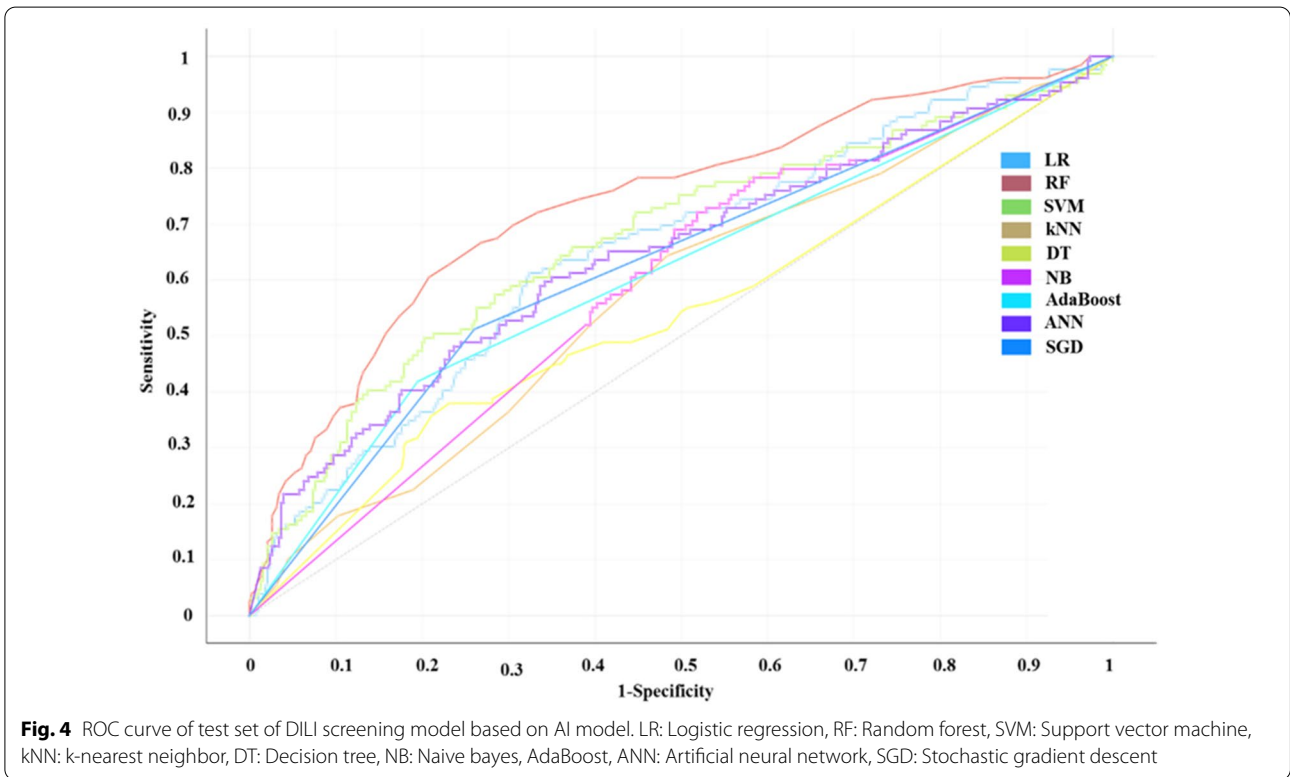
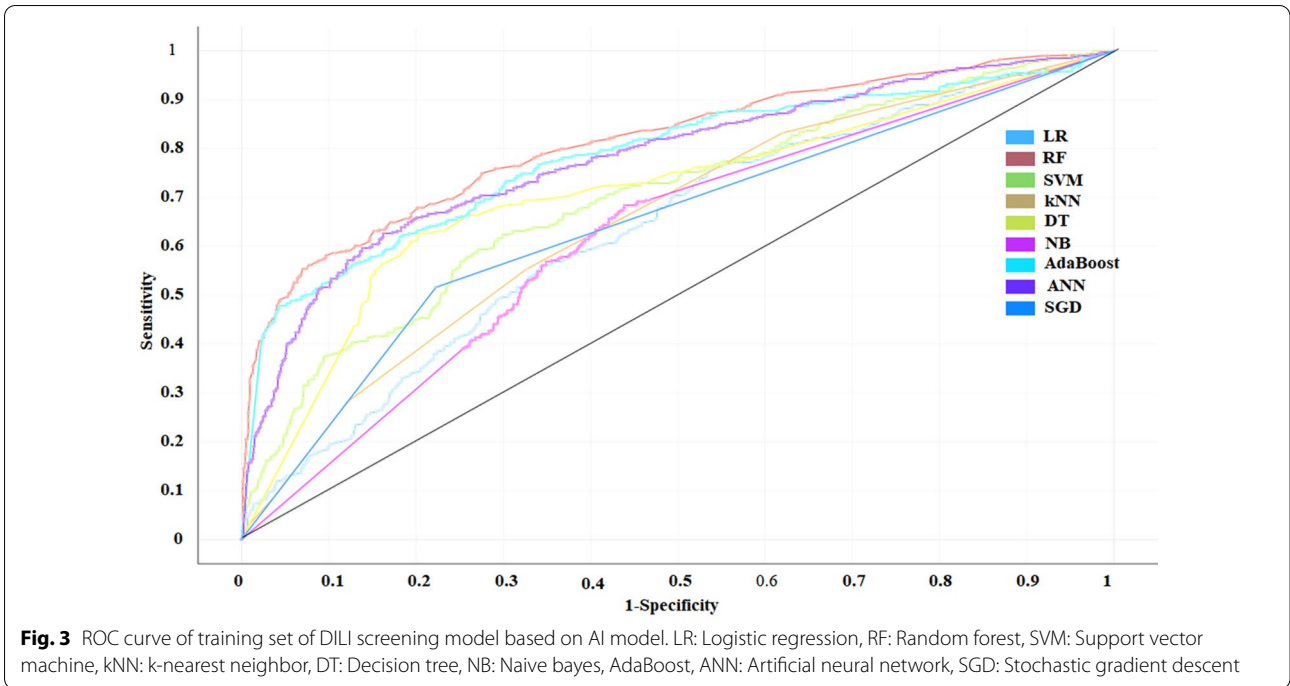
As listed in Table 1, CA ranged from 0.686 to 0.838, F1 ranged from 0.671 to 0.827, Precision ranged from 0.685 to 0.832, Recall ranged from 0.686 and 0.838, AUC ranged from 0.621 to 0.814 (Receiver-operating characteristic curve, ROC) for each model on the training set. CA ranged from 0.675 to 0.767, F1 ranged from 0.636 to 0.731, Precision ranged from 0.555 to 0.739, Recall ranged from 0.675 and 0.767, AUC ranged from 0.544 to 0.739 (ROC) for each model on the test set.

Above model results showed that RF could achieved the best results than other machine learning methods on both the training set and the test set (Figs. 3 and 4). DBN obtained 82.2% accuracy on test set, which the number of hidden layers was 100, the batchsize was 25, and the

**Table 1** AI models for drug-included liver injury

Dataset	ML method	AUC	CA	F1	Precision	Recall
Training set	SGD	0.647	0.709	0.715	0.723	0.709
	kNN	0.654	0.722	0.698	0.690	0.722
	SVM	0.785	0.795	0.791	0.788	0.795
	DT	0.710	0.756	0.760	0.764	0.756
	RF	0.814	0.838	0.827	0.832	0.838
	Adaboost	0.785	0.792	0.788	0.785	0.792
	ANN	0.621	0.737	0.671	0.685	0.737
	LR	0.746	0.776	0.757	0.757	0.761
	NB	0.632	0.686	0.694	0.711	0.686
Test set	SGD	0.627	0.682	0.694	0.712	0.682
	kNN	0.574	0.745	0.636	0.555	0.745
	SVM	0.669	0.747	0.712	0.710	0.747
	DT	0.544	0.680	0.679	0.678	0.680
	RF	0.739	0.767	0.731	0.739	0.767
	Adaboost	0.614	0.708	0.707	0.707	0.708
	ANN	0.647	0.694	0.696	0.697	0.694
	LR	0.656	0.733	0.694	0.688	0.733
	NB	0.598	0.675	0.648	0.705	0.675

LR Logistic regression, RF Random forest, SVM Support vector machine, kNN k-nearest neighbor, DT Decision tree, NB Naive bayes, ANN Artificial neural network, SGD Stochastic gradient descent



learning rate was 1 with 400 iterations. This accuracy was higher than that of nine machine learning models on the test set.

**Discussion**

In clinical practice, DILI lacks specific and sensitive diagnostic criteria, and the critical part of diagnosis depends

on eliminating a series of diseases through blood testing. Once DILI events occur, the drug will be discontinued for the patient or even withdrawn from the market. Thus, it is urgent to develop a new approach to screen DILI.

Multiple compounds entered into the human body when TCM combined with WM. The complexity of the chemical compounds of TCM makes DILI extremely complicated in TCM-WMC, also increasing the risk of metabolic DILI. Early warning of compounds with hepatotoxicity has a vital clinical medication significance in TCM-WMC. Traditional DILI identification relies on animal experiments and clinical trials, risk factor assessment and case reports, etc. Above methods were not only inefficient but costly in terms of manpower, material, and financial resources. Besides, there may be a gap between experimental animals and mankind, reducing applicability of hepatotoxicity test results in humans. By comparison, recent DILI AI models were constructed based on mathematical models. For instance, quantitative structure–activity relationship (QSAR) model, one representative, was able to speculate on the specific physical, chemical and biological properties of compounds on the basis of structural information of known compounds, thereby achieving qualitative or quantitative screening of unknown compound.

It was challenging and crucial to systematically optimize the description form and their combinations of compounds. SMILES describes a three-dimensional chemical structure with a string of characters. Chemical structures can be characterized by a set of numerical values called molecular fingerprints or descriptors. These compound molecular descriptors or molecular fingerprints with high-dimensional information need to be processed with the application of information technologies, such as AI. Therefore, in this study, AI models was built for screening hepatotoxic compounds in TCM-WMC, with the 9 machine learning models (SGD, kNN, SVM, NB, DT, RF, ANN, Adaboost, LR) and one deep learning model (DBN). We found that DBN model had a better model performance than other 9 machine learning methods.

Basic machine learning models become progressively better at making predictions or decisions, which still need some guidance. More specifically, deep learning is considered to be the evolution of machine learning. It uses a programmable neural network to enable machines to make accurate decisions without humans help. DBN is composed of multi-layer unsupervised restricted Boltzmann machine (RBM) network and one layer supervised back propagation (BP) network.. The training process of DBN can be implemented from low to high level to train multilayer RBMs. Each RBM layer

was trained by using the hidden unit (H) of the previous layer as the input/visible unit (V). The descriptors and fingerprints of drug or compounds were used as input  $V_k$ , and the binary classification with one layer supervised back propagation (BP) network of liver injury is used as input label. The combination of unsupervised RBM and supervised BP implements the supervised prediction of hepatotoxic compounds. After the training of the multi-layer RBMs, the liver injury of the prototype compounds/metabolic compounds of TCM-WMC is gradually categorized, to make a rapid warning for hepatotoxicity of compounds in TCM-WMC.

This study only focuses on the risk of liver injury caused by a single compound of TCM-WMC, and provides a path for the prediction of hepatotoxicity of more new compounds formed through the interaction between compounds from TCM and WM. In future research, mechanisms of liver injury in TCM-WMC needs to be further explained. We will further focus on the liver injury possible risk caused by TCM-WMC metabolites or interaction among complex TCM-WMC compounds by calculating more compounds data. Overall, this study will provide guidance for the safe utilization of TCM-WMC and improve the diagnosis and treatment ability of clinicians.

## Conclusion

We compared the DILI models' performance of different machine learning and deep learning models, and found that DBN model had better model performance than other 9 machine learning methods. Therefore, this method of DBN for screening hepatotoxic compounds in TCM-WMC may be helpful to guide the clinical standards and safe medication, and avoid the risk of liver injury in the clinical combination.

## Abbreviations

TCM-WMC: Traditional Chinese medicine and western medicine combination; DILI: Drug-induced liver injury; kNN: K-nearest neighbor; SVM: Support vector machine; NB: Naive bayes; DT: Decision tree; RF: Random forest; ANN: Artificial neural network; LR: Logistic regression; DBN: Deep belief network; CA: Classification accuracy; AUC: Area under the curve; AI: Artificial intelligence; ALF: Acute liver injury; ML: Machine learning; SGD: Stochastic gradient descent; ANNs: Artificial neural networks; RBM: Restricted Boltzmann machine; ROC: Receiver operating characteristic; QSAR: Quantitative structure–activity relationship.

## Acknowledgements

Not applicable.

## Author contributions

ZC, CZ and HS conceived and designed the experiments; ZC, MZ, XZ, YS, HP and RQ collected the data; ZC, RZ, LY, YJ, TH, XW and ZNC contributed reagents/materials/analysis tools/codes; ZC constructed the dataset; ZC and MZ wrote this manuscript; ZC, MZ, LY, RZ, RQ, CZ and HS revised the manuscript. All authors read and approved the final manuscript.

**Funding**

This study was supported by National Key R&D Program of China (2019YFC1710400; 2019YFC1710405); National Natural Science Foundation of China (No.82104694; No.81803963), China Postdoctoral Science Foundation (No. 2020M670236; No. 2021T140075).

**Availability of data and materials**

All data used in the presented study can get from the corresponding author upon request.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All the authors have agreed that the manuscript published in Chinese Medicine.

**Competing interests**

The authors declare that they have no conflicts of interest.

**Author details**

<sup>1</sup>Key Laboratory of Chinese Internal Medicine of Ministry of Education, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China. <sup>2</sup>School of Computer Science, Fudan University, Shanghai, China. <sup>3</sup>Present Address: Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China.

Received: 14 February 2022 Accepted: 29 April 2022

Published online: 17 May 2022

**References**

- Chen M, et al. FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov Today*. 2011;16(15–16):697–703.
- Stevens JL, Baker TK. The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov Today*. 2009;14(3–4):162–7.
- Reuben A, et al. Outcomes in adults with acute liver failure between 1998 and 2013. *Ann Intern Med*. 2016;164(11):724–32.
- Shen T, et al. Incidence and etiology of drug-induced liver injury in mainland China. *Gastroenterology*. 2019;156(8):2230–2241.e11.
- Fontana RJ. Pathogenesis of idiosyncratic drug-induced liver injury and clinical perspectives. *Gastroenterology*. 2014;146(4):914–28.
- Jay H, Hoofnagle MD, Einar S, Björnsson MD. Drug-induced liver injury—types and phenotypes. *N Engl J Med*. 2019;381(3):264–73.
- Zheng R, et al. Clinical safety and efficacy of Wenxin Keli-amiodarone combination on heart failure complicated by ventricular arrhythmia: a systematic review and meta-analysis. *Front Physiol*. 2018;9:487.
- Han ZZ, et al. Is acupuncture effective in the treatment of COVID-19 related symptoms? Based on bioinformatics/network topology strategy. *Brief Bioinform*. 2021;22(5):bbab110.
- Qiu RJ, et al. Outcome reporting from clinical trials of non-valvular atrial fibrillation treated with traditional Chinese medicine or Western medicine: a systematic review. *BMJ Open*. 2019;9(8): e028803.
- Zheng R, Chen SQ, Shang HC. The safety evaluation method of herb-drug combination: manifestation-target alliance. *Mod Trad Chin Med Mater Med World Sci Technol*. 2018;20(10):1734–8.
- Diaz O, Dalton JAR, Giraldo J. Artificial intelligence: a novel approach for drug discovery. *Trends Pharmacol Sci*. 2019;40(8):550–1.
- Bi WL, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *Ca Cancer J Clin*. 2019;69(2):127–57.
- Tomasev N, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572(7767):116–9.
- Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*. 2017;18(5):851–69.
- Zhang L, et al. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today*. 2017;22(11):1680–5.
- Wang H, et al. Deep neural network models for predicting chemically induced liver toxicity endpoints from transcriptomic responses. *Front Pharmacol*. 2019;10:42.
- Xu Y, et al. Deep learning for drug-induced liver injury. *J Chem Inf Model*. 2015;55(10):2085–93.
- Chen M, et al. DILLrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov Today*. 2016;21(4):648–53.
- Hoofnagle JH, et al. LiverTox: a website on drug-induced liver injury. *Hepatology*. 2013;57(3):873–4.
- Thakkar S, et al. The Liver Toxicity Knowledge Base (LKTb) and drug-induced liver injury (DILI) classification for assessment of human liver injury. *Expert Rev Gastroent*. 2018;12(1):31–8.
- Quinton A, Lattry P, Biour M. Hepatox: database on hepatotoxic drugs. *Gastroenterol Clin Biol*. 1993;17(5 Pt 2):H116–20.
- Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem*. 2011;32(7):1466–74.
- O'Brien PJ, et al. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch Toxicol*. 2006;80(9):580–604.
- Xu JHJ, et al. Cellular imaging predictions of clinical drug-induced liver injury. *Toxicol Sci*. 2008;105(1):97–105.
- Greene N, et al. Developing structure-activity relationships for the prediction of hepatotoxicity. *Chem Res Toxicol*. 2010;23(7):1215–22.
- Rodgers AD, et al. Modeling liver-related adverse effects of drugs using kNearest neighbor quantitative structure activity relationship method. *Chem Res Toxicol*. 2010;23(4):724–32.
- Liew CY, Lim YC, Yap CW. Mixed learning algorithms and features ensemble in hepatotoxicity prediction. *J Comput Aided Mol Des*. 2011;25(9):855–71.
- Liu ZC, et al. Translating clinical findings into knowledge in drug safety evaluation—drug induced liver injury prediction system (DILiPs). *Plos Comput Biol*. 2011;7(12): e1002310.
- Low Y, et al. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem Res Toxicol*. 2011;24(8):1251–62.
- Chen M, et al. The liver toxicity knowledge base: a systems approach to a complex end point. *Clin Pharmacol Ther*. 2013;93(5):409–12.
- Zhu X, Kruhlik NL. Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. *Toxicology*. 2014;321:62–72.
- Huang SH, et al. Developing a QSAR model for hepatotoxicity screening of the active compounds in traditional Chinese medicines. *Food Chem Toxicol*. 2015;78:71–7.
- Chen MJ, et al. Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. *Toxicol Sci*. 2013;136(1):242–9.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

