Drug Components-Disease Network Related to Acute Lung Injury Inference Based on Forest Graph-embedded Deep Feedforward Network

Bin Yang

Zaozhuang University

Wenzheng Bao (✉ baowz55555@126.com)

Xuzhou University of Technology

Jinglong Wang

Zaozhuang University

Baitong Chen

Xuzhou University of Technology

Research Article

Keywords: network pharmacology, acute lung injury, deep learning algorithm, target

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1238402/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Acute lung injury (ALI) is a serious respiratory disease, which can lead to acute respiratory failure or death. It is closely related to the pathogenesis of New Coronavirus pneumonia (COVID-19). Many researches showed that traditional Chinese medicine (TCM) had a good effect on its intervention, and network pharmacology could play a very important role.

Results: In order to construct "disease-gene-target-drug" interaction network more accurately, deep learning algorithm is utilized in this paper. Two ALI-related target genes (REAL and SATA3) are considered, and the active and inactive compounds of the two corresponding target genes are collected as training data, respectively. Molecular descriptors and molecular fingerprints are utilized to characterize each compound. Forest graph embedded deep feed forward network (forgeNet) is proposed to train and identify 19 compounds in Erhuang decoction (EhD) and Dexamethasone (DXMS).

Conclusions: The experiment results show that forgeNet performs better than support vector machines (SVM), random forest (RF) and gcForest.

1. Background

Internal and external etiology can lead to self-stable regulation disorder, which could change a series of metabolisms, functions and structures. Abnormal life activity processes are manifested as abnormal symptoms, signs and behavior [1–2]. Under certain conditions, the abnormal life activity processes caused by the disturbance of homeostasis after the damage of the disease cause the disease [3–4]. Traditional Chinese medicine (TCM) has been utilized to treat diseases for thousands of years [5–7]. Traditional Chinese medicine is a kind of material with the function of rehabilitation and health care, which could be utilized to prevent, treat and diagnose diseases under the guidance of TCM theory [8–11].

Traditional Chinese medicine mainly comes from natural medicine and its processed products, including plant medicine, animal medicine, mineral medicine and some chemical and biological products [12–13]. The most important feature of traditional Chinese medicine in treating diseases is to pay attention to the adjustment of the functions of viscera and organs, and the balance and coordination between them. The focus of traditional Chinese medicine treatment is not that the human body is infected with the specific bacteria, virus and other pathogenic factors, but the specific reaction of the human body after these pathogenic factors act on the human body [14–15]. The purpose of treatment is to enhance the disease resistance and recovery ability of human body. To kill bacteria and relieve symptoms are mainly achieved by enhancing the body's own functions. In recent years, traditional Chinese medicine has certain advantages in the treatment of pneumonia [16], shock [17], convulsion [18], hemorrhage [19], acute respiratory failure [20], renal failure [21], heart failure [22], cerebrovascular accident [23], etc. it is not only effective, but also safe and simple, with few adverse reactions.

In the past decade, with the rapid development of sequencing technology, a large number of genomics data such as genomics, proteomics, metabonomics and so on, have been generated, which has led to the

changes in the research of traditional Chinese medicine for diseases. Network pharmacology has been proposed, which was developed on the basis of the rapid development of systems biology and computer technology, generating the "disease-gene-target-drug" interaction network. Through network analysis, we can systematically and comprehensively observe the intervention and influence of drugs on the disease network, reveal the mystery of the synergistic effect of multi branch drugs on the human body, and find out the multi-target new drugs with high efficiency and low toxicity. Network pharmacology of traditional Chinese medicine has become a new idea for drug mechanism research and new drug development [24–28]. Lu et al. utilized network pharmacology and molecular docking technology to study the mechanism of Shaoyao Decoction in the treatment of ulcerative colitis, and found that Shaoyao decoction can improve the pathological damage of colon [29]. Liu et al. collected the main active components of *Portulacae Herba*, constructed interaction network of target proteins of liver cancer, and found that ketones may be the main material basis of its anti-liver cancer, which is related to the regulation of MAPK signaling pathway [30]. Liu et al. utilized network pharmacology to screen 102 active components of Danzhi Xiaoyao Powder, 147 corresponding targets and 52 intersecting targets with insomnia, and obtained the key components, key targets and key pathways of Danzhi Xiaoyao Powder in the treatment of insomnia [31]. Yang et al. presented network pharmacology to analyze the potential anti-tumor mechanisms of the main active components of *Prunella vulgaris* systematically at the molecular level [32]. Shen et al. discussed the possible mechanism of Wuling Powder in the treatment of diabetic nephropathy by network pharmacology, and found that Wuling Powder may reduce renal cell damage by regulating apoptosis related proteins, such as Caspases family protein and BCL2 Protein family [33].

In the recent years, data mining methods have been applied to extract useful information from lots of TCM data [33]. Ren et al. utilized data mining methods to screen out 47 prescriptions, and found out 14 core drugs and 7 new prescriptions in order to search the medication rules and mechanism of TCM in the treatment of carotid atherosclerosis (CAS) [34]. Ga et al. utilized data mining method to select the top five active components of each Tibetan medicine with high frequency and network pharmacology was utilized to analyze the mechanism of Tibetan medicine in the treatment of high altitude polycythemia [35]. In order to study the medication rule of TCM intervention in iron death, Ou et al. constructed target-compound, compound-TCM, target-compound-TCM network, and frequency statistics was utilized to show that bitter and pungent herbs were the main herbs that could interfere with iron death, while cold herbs were the main ones, which mainly belonged to liver and lung meridians [36]. Pan et al. reprocessed a large number of Chinese medicine prescriptions for the treatment of primary liver cancer, and by analysis of data mining and network pharmacology medication regularity of effective traditional Chinese medicine prescriptions in the treatment of primary liver cancer was obtained [37]. Zheng et al. presented four classifiers to infer compound-target interaction network in the process of network pharmacology analysis [38].

In order to better mine omics data and construct "disease-gene-target-drug" interaction network, deep learning model was utilized in this paper. Taking acute lung injury (ALI) disease as an example, we selected two disease-related target genes (**REAL** and **SATA3**). The active and inactive compounds of the two target genes combined are collected. Molecular descriptors and molecular fingerprints are utilized to

characterize each compound, which contain 374 features. Forest graph embedded deep feed forward network is utilized to train and identify new compounds target genes related.

2. Results

In this section, active and inactive ligands of two key target genes: REAL and SATA3 about ALI disease are collected. For REAL, 966 ligands are collected, which contain 146 positive samples and 820 negative samples (Data1). For SATA3, 193 active ligands and 1210 inactive ligands are collected (Data2). Molecular descriptors and molecular fingerprints of each ligand could be obtained, which contains 374 features. In order to better reflect the effectiveness of forgeNet, three classical classifiers (SVM [42], RF [43] and gcForest [44]) are utilized to identify the compounds associated with diseases. Five evaluation criteria of classifier performance are utilized, which are *SN*, *SP*, *ACC*, *MCC* and *F1*, respectively.

2.1 Model test

In order to test the generalization and stability of forgeNet, leave-one-out, 3-fold, 5-fold and 10-fold cross validation methods are utilized. With Data1, the inference performances of four methods with leave-one-out, 3-fold, 5-fold and 10-fold cross validation methods are listed in Table 1, Table 3, Table 5 and Table 7, respectively. With Data2, the inference performances of four methods with leave-one-out, 3-fold, 5-fold and 10-fold cross validation methods are listed in Table 2, Table 4, Table 6 and Table 8, respectively. From the results, it could be seen that gcForest has the highest *SN* performances among four methods, which reveal that gcForest could identify more true active ligands. RF could obtain the higher *SP* than SVM, gcForest and forgeNet, which show that RF could identify true more inactive ligands. ForgeNet has the best *ACC*, *MCC* and *F1* performances among four methods. The results reveal that forgeNet could identify more true active and inactive ligands than SVM, RF and gcForest. And when two categories have very different sizes, forgeNet performs best. *F1* performances show that on the whole forgeNet could infer components-disease network more accurately than other three classifiers.

Table 1
Performances of four methods with Data1 by leave-one-out method.

	SN	SP	ACC	MCC	F1
SVM	0.664384	0.968293	0.92236	0.679827	0.72119
RF	0.534247	0.99878	0.928571	0.696718	0.693333
gcForest	0.828767	0.960976	0.940994	0.774779	0.809365
ForgeNet	0.80137	0.981707	0.954451	0.816616	0.841727

Table 2
Performances of four methods with Data2 by leave-one-out method.

	SN	SP	ACC	MCC	F1
SVM	0.549223	0.975207	0.916607	0.610521	0.644377
RF	0.533679	0.997521	0.933713	0.692319	0.688963
gcForest	0.735751	0.952893	0.923022	0.679871	0.72449
ForgeNet	0.751295	0.978512	0.947256	0.768381	0.796703

Table 3
Performances of four methods with Data1 by 3-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.589041	0.964634	0.907867	0.612385	0.659004
RF	0.356164	0.995122	0.898551	0.538431	0.514851
gcForest	0.842466	0.876829	0.871636	0.61047	0.664865
ForgeNet	0.69863	0.97561	0.933747	0.727021	0.761194

Table 4
Performances of four methods with Data2 by 3-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.264249	0.980165	0.881682	0.374255	0.380597
RF	0.316062	0.997521	0.903778	0.517658	0.474708
gcForest	0.580311	0.704132	0.687099	0.207576	0.337858
ForgeNet	0.466321	0.975207	0.905203	0.543783	0.57508

Table 5
Performances of four methods with Data1 by 5-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.60274	0.962195	0.907867	0.615696	0.664151
RF	0.39726	0.997561	0.906832	0.585928	0.563107
gcForest	0.787671	0.937805	0.915114	0.688823	0.737179
ForgeNet	0.691781	0.97439	0.931677	0.71832	0.753731

Table 6
Performances of four methods with Data2 by 5-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.388601	0.957025	0.878831	0.414907	0.46875
RF	0.139896	1	0.881682	0.350741	0.245455
gcForest	0.585492	0.835537	0.80114	0.348734	0.447525
ForgeNet	0.502591	0.966942	0.903065	0.544829	0.587879

Table 7
Performances of four methods with Data1 by 10-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.630137	0.968293	0.917184	0.654572	0.69697
RF	0.486301	0.997561	0.92029	0.655719	0.648402
gcForest	0.80137	0.947561	0.925466	0.721602	0.764706
ForgeNet	0.760274	0.980488	0.947205	0.785198	0.813187

Table 8
Performances of four methods with Data2 by 10-fold cross validation method.

	SN	SP	ACC	MCC	F1
SVM	0.455959	0.961157	0.891661	0.487193	0.536585
RF	0.378238	0.999174	0.913756	0.581579	0.546816
gcForest	0.601036	0.928099	0.883108	0.518083	0.585859
ForgeNet	0.595855	0.973554	0.921597	0.640386	0.676471

Receiver operating characteristic (ROC) curve is based on false positive rate (FPR) and true positives rate (TPR). Area under curve (AUC) is defined as the area under the ROC curve surrounded by the coordinate axis. In general ROC and AUC are utilized to evaluate the classifiers. With Data1, ROC curve and AUC performances of four methods with 3-fold, 5-fold, 10-fold cross validation methods and leave-one-out are depicted in Fig. 1, Fig. 2, Fig. 3 and Fig. 4, respectively. With Data2, ROC curve and AUC performances of four methods with 3-fold, 5-fold, 10-fold cross validation methods and leave-one-out are depicted in Fig. 5, Fig. 6, Fig. 7 and Fig. 8, respectively. From the results, in most of cases, forgeNet could obtain the best AUC performances among four classifiers. With Data1 and Data2, RF performs best in terms of ROC and AUC by leave-one-out method.

2.2 Compound screening

19 active chemical compounds (Neoglycyrol, Uralenol, Syringic acid 4- β -D-Glucopyranoside, Gancaonin N, Chrysin -6-C-glucoside-8-C-arabinoside, Chrysin-6-C-arabinoside 8-C-glucoside Liquiritin, Baicalin, Isomer of Baicalin, Oroxylin A-7-O- β -D-glucuronide, Chrysin-7-O-glucuronide, Isoliquiritin, Wogonoside, Liquiritigenin, Baicalein, Isoliquiritigenin, Wogonin, Oroxylin A, and Glycyrrhetic acid) in Erhuang decoction (EhD) can dock with ALI related target genes and have high potential biological activity, which have been proved in the reference [39]. Dexamethasone (DXMS) is used as control drug. Molecular descriptors and molecular fingerprints are also utilized to obtain the features of 20 chemical compounds. Data1 and Data2 are utilized as the training sets in order to predict 20 chemical compounds, respectively. The prediction ranks are listed in Table 9. By ranking results, we can see that DXMS ranks last by forgeNet on average, which is consistent with the results of molecular docking in the past research [39]. Thus the results reveal that forgeNet could screen the chemical compounds more accurately than SVM, RF and gcForest.

Table 9
Prediction ranks of 20 chemical compounds by SVM, RF, gcForest and forgeNet.

		D1	D2	D1	D2	D1	D2	D1	D2
ID	Chemical compounds	SVM	SVM	RF	RF	gcForest	gcForest	forgeNet	forgeNet
1	Neoglycyrol	11	13	18	8	12	12	11	8
2	Uralenol	12	15	12	6	16	6	4	7
3	Syringic acid 4-β-D-Glucopyranoside	3	4	7	9	17	11	1	17
4	Gancaonin N	13	16	19	10	15	10	12	6
5	Chrysin -6-C-glucoside-8-C-arabinoside	14	19	15	18	11	19	10	18
6	Chrysin-6-C-arabinoside 8-C-glucoside	15	18	14	19	9	20	9	20
7	Liquiritin	9	6	8	12	4	13	17	13
8	Baicalin	17	9	5	14	8	14	3	11
9	Isomer of Baicalin	18	10	10	15	19	15	5	10
10	Oroxylin A-7-O-β-D-glucuronide	16	17	13	16	10	16	8	14
11	Chrysin-7-O-glucuronide	7	7	9	13	13	9	2	9
12	Isoliquiritin	10	5	16	20	7	17	13	12
13	Wogonoside	20	20	11	17	20	5	6	16
14	Liquiritigenin	19	14	20	11	1	18	19	15
15	Baicalein	2	11	17	7	6	4	14	3
16	Isoliquiritigenin	6	2	1	2	18	1	7	1
17	Wogonin	4	12	3	5	5	7	16	5
18	Oroxylin A	5	8	4	4	2	8	15	4
19	Glycyrrhetic acid	8	1	6	1	3	2	18	2
20	DXMS	1	3	2	3	14	3	20	19

3. Discussions

In order to test the influence of different feature sets on the identification results, we utilized molecular descriptors as control feature set. Molecular descriptors and molecular fingerprints make up full feature set. With these two feature sets, SVM, RF, gcForest and forgeNet are utilized by 3-fold, 5-fold, 10-fold and leave-one-out methods. The AUC and F1 results are depicted in Fig. 9 and Fig. 10, respectively. From the results, it could be seen that full feature set could improve the compound identification accuracy of methods.

4. Conclusions

Network pharmacology has become a frontier and hot spot in the field of traditional Chinese medicine research. This research method can effectively predict the effective components, target and side effects of drugs, and is conducive to the process of modernization of traditional Chinese medicine. In order to construct "disease-gene-target-drug" interaction network more accurately, forest graph embedded deep feed forward network is utilized to infer "disease-compound" network in this paper. According to acute lung injury, two ALI-related target genes (REAL and SATA3) are selected, and the active and inactive compounds of the two corresponding target genes are collected, respectively. Molecular descriptors and molecular fingerprints are utilized to characterize each compound. By leave-one-out, 3-fold, 5-fold and 10-fold cross validation methods, forgeNet has the better performance than SVM, RF and gcForest in terms of *SN*, *SP*, *ACC*, *MCC*, *F1*, AUC and ROC curves. ForgeNet is also utilized to identify 19 compounds in Erhuang decoction (EhD) and Dexamethasone (DXMS) and the results reveal that forgeNet could infer the compounds of disease related more accurately. We also test the influence of different feature sets on the identification results and find the feature set based on molecular descriptors and molecular fingerprints could improve the compound identification accuracy of methods.

5. Methods

5.1 forgeNet

Forest graph-embedded deep feedforward network (forgeNet) is a novel machine learning algorithm, which has been successfully applied to solve classification problem with TCGA RNA-seq data. The flowchart of forgeNet is depicted in Fig. 11. From Fig. 12, it could be seen that this method contains two parts: feature graph construction and deep neural network. Compared with deep learning models, forgeNet solves the dimension problem of biological data and is more robust. The algorithm is described as follows [39].

- Step 1: feature graph construction

Before the labeled training data are input into classifier, the features of the data need to be extracted. In forgeNet, the used forest ξ contains p decision tree (DT). With the labeled training data, the forest is fitted

and p DT are generated ($\xi(\theta) = \{T_1(\theta_1), T_2(\theta_2), \dots, T_p(\theta_p)\}$, θ_i is a parameter). Meanwhile if binary tree is regarded as a special case of directed graph, we can gain the following graph set.

$$\Phi = \{G_1(V_1, E_1), K, G_i(V_i, E_i), K, G_N(V_p, E_p)\}. \quad 1$$

Where V_i and E_i represents vertex set and edge set of G_i .

To integrate the directed graph set Φ , the final aggregated graph can be gained by the following formula.

$$G = \bigcup_{i=1}^N G_i.$$

2

- Step 2: deep neural network

The feature graph obtained the previous step are embedded into this part. With the processed features graph-embedded deep feedforward networks (GEDFN) is used to train and make the classification for the unknown data [12]. Every layer of GEDFN is introduced as followed.

$$\begin{aligned} Z_1 &= \sigma(X(W_{in} \Theta G) + b_{in}), \\ K \\ Z_{k+1} &= \sigma(Z_k W_k + b_k), \\ K \\ Z_{out} &= \sigma(Z_i W_i + b_i), \\ y &= \text{softmax}(Z_{out} W_{out} + b_{out}). \end{aligned}$$

3

Where X is input data, Z_k is the k - th hidden layers, Θ denotes Hadamard product, W_k and b_k are the weights and bias of the k - th hidden layer, respectively.

5.2 Inference algorithm

(1) Data preparation. Two key target genes: signal transducer and activator of transcription 3 (STAT3), and nuclear transcription factor- κ B/p65 (nuclear factor kappa, B/p65, REAL) were proved to be mainly involved in the key pathways related to acute lung injury (ALI), and losely related to ALI diseases in the literature [40]. Then, the BindingDB database is searched for the known active compounds of two key target genes [38]. The active ligands are screened with the condition that $IC_{50} < 5000 \text{ nmol} \cdot \text{L}^{-1}$. The collected active compounds are labeled as positive samples. In order to collect the negative samples, 20% of the active ligands are randomly selected and uploaded to DUD-E database (<http://dude.docking.org/>) to generate the inactive ligands [41]. In order to obtain the molecular descriptors and molecular fingerprints of each ligand, the active and inactive ligands collected are uploaded as the feature vectors.

(2) Model training. According to the collected data, the feature vector of each ligand is used as input for forgeNet. After training phase, the unknown compounds are screened for the target disease.

Declarations

Acknowledgements

This work was supported by the talent project of "Qingtian scholar" of Zaozhuang University, Jiangsu Provincial Natural Science Foundation, China (No. SBK2019040953), Youth Innovation Team of Scientific Research Foundation of the Higher Education Institutions of Shandong Province, China (No. 2019KJM006), the Key Research Program of the Science Foundation of Shandong Province (ZR2020KE001), the PhD research startup foundation of Zaozhuang University (No.2014BS13), and Zaozhuang University Foundation (No. 2015YY02).

Funding

This work was supported by the talent project of "Qingtian scholar" of Zaozhuang University, Jiangsu Provincial Natural Science Foundation, China (No. SBK2019040953), Youth Innovation Team of Scientific Research Foundation of the Higher Education Institutions of Shandong Province, China (No. 2019KJM006), the Key Research Program of the Science Foundation of Shandong Province (ZR2020KE001), the PhD research startup foundation of Zaozhuang University (No.2014BS13), and Zaozhuang University Foundation (No. 2015YY02).

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author.

Declarations Ethics approval and consent to participate

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors have no conflict of interest to declare.

References

1. Hall C.B., Simões E.A.F., Anderson L.J. (2013) Clinical and Epidemiologic Features of Respiratory Syncytial Virus. In: Anderson L., Graham B. (eds) Challenges and Opportunities for Respiratory Syncytial Virus Vaccines. Current Topics in Microbiology and Immunology, vol 372. pp. 39-57. Springer, Berlin, Heidelberg.
2. Michael, A, Beckles, et al. Initial Evaluation of the Patient With Lung Cancer*: Symptoms, Signs, Laboratory Tests, and Paraneoplastic Syndromes. *Chest*, 2003, 123(1):97S-104S.
3. Sobhani M, Bechara A. A somatic marker perspective of immoral and corrupt behavior. *Soc Neurosci*, 2011, 6(5-6):640-652.
4. Ekpenyong C. Abnormal Serum Uric Acid Levels in Health and Disease: A Double-Edged Sword. *American Journal of Internal Medicine*, 2014, 2(6):113-130.
5. Bensoussan A, Myers S P, Carlton A L. Risks associated with the practice of traditional Chinese medicine: an Australian study. *Archives of Family Medicine*, 2000, 9(10):1071-8.
6. Li S, Zhang Z Q, Wu L J , et al. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. *let Systems Biology*, 2007, 1(1):51-60
7. Wang X, Hui S, Zhang A , et al. Potential role of metabolomics approaches in the area of traditional Chinese medicine: as pillars of the bridge between Chinese and Western medicine. *Journal of Pharmaceutical & Biomedical Analysis*, 2011, 55(5):859-868
8. Xue R, Fang Z, Zhang M, et al. TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Research*, 2013, 41(D1):D1089-D1095.
9. Hou Y, Wang Y, J Zhao, et al. Correction: Smart Soup, a Traditional Chinese Medicine Formula, Ameliorates Amyloid Pathology and Related Cognitive Deficits. *PLoS ONE*, 2020, 15(8):e0237035
10. Ren, Beida, Cheng, et al. Possible mechanisms underlying treatment of Alzheimer's disease with Traditional Chinese Medicine: active components, potential targets and synthetic pathways of Bulao Elixir. *Journal of Traditional Chinese Medicine*, 2020, 40(3):145-157
11. Xu W, Towers A D, Li P, et al. Traditional Chinese medicine in cancer care: perspectives and experiences of patients and professionals in China. *European Journal of Cancer Care*, 2010, 15(4):397-403.
12. Che C T, Man S W, Lam C. Natural Products from Chinese Medicines with Potential Benefits to Bone Health. *Molecules*, 2016, 21(3):239
13. Zhang J, Han J, Oyeleye A, et al. Extraction methods of natural products from Traditional Chinese medicines. *Methods Mol Biol*, 2015, 1263:177-185
14. Alromaima A, Liao Y, Feng J, et al. Advances in the treatment of novel coronavirus disease (COVID-19) with Western medicine and traditional Chinese medicine: A narrative review. *Journal of Thoracic Disease*, 2020, 12(10):6054-6069
15. Matsuo R, Ball M A, Kobayashi M, et al. Effects of a traditional Chinese herbal medicine, Kanzo-bushi-to, on the resistance of thermally injured mice infected with herpes simplex virus type 1. *International Journal of Immunopharmacology*, 1994, 16(10):855-863

16. Lin S K, Tsai Y T, Lo P C, et al. Traditional Chinese medicine therapy decreases the pneumonia risk in patients with dementia. *Medicine*, 2016, 95(37):e4917
17. Chen H C, Chen W C, Lin K H, et al. Simultaneous use of traditional Chinese medicine (Si-Ni-Tang) to treat septic shock patients: study protocol for a randomized controlled trial. *Trials*, 2011, 12:199
18. Meng F Y. Therapeutic Effect of Combination of Traditional Chinese Medicine and West Medicine for Convulsion after Burn Injury in Children. *Chinese General Practice*, 2014(24):2890-2891.
19. Hobbs V. Traditional Chinese medicine for hemorrhage. *Midwifery Today with International Midwife*, 1997(43):30
20. Wu T, Yang X, Zeng X, et al. Traditional Chinese medicine in the treatment of acute respiratory tract infections. *Respiratory Medicine*, 2008, 102(8):1093-1098
21. Wang H Q, Wei L, Wu W, et al. Reevaluation on systematic assessment of traditional Chinese medicine for chronic renal failure. *Chinese Traditional and Herbal Drugs*, 2014, 45(5):738-744.
22. Fu T C, Lin Y C, Chang C M, et al. Validation of a new simple scale to measure symptoms in heart failure from traditional Chinese medicine view: a cross-sectional questionnaire study. *BMC complementary and alternative medicine*, 2016, 16:342
23. Sun L H, Li Y J, Fu G L, et al. Nursing implementation with Traditional Chinese Medicine based on syndrome differentiation for cerebrovascular accident patients with sleep disorders. *Nursing and Rehabilitation Journal*, 2013, 12(5):467-468.
24. Gu J, Gui Y, Chen L, et al. Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *Plos One*, 2013, 8(4):e62839.
25. Jian L, Lu C, Jiang M, et al. Traditional Chinese Medicine-Based Network Pharmacology Could Lead to New Multicomponent Drug Discovery. *Evid Based Complement Alternat Med*, 2012, 2012(12):149762.
26. Zhang G B, Li Q Y, Chen Q L, et al. Network Pharmacology: A New Approach for Chinese Herbal Medicine Research. *Evid Based Complement Alternat Med*. 2013, 2013:621423
27. Zhang A, Hui S, Yang B, et al. Predicting new molecular targets for rhein using network pharmacology. *BMC Systems Biology*, 2012, 6(1):20
28. Da C H, Pei G X. Network Pharmacology: A Rosetta Stone for Traditional Chinese Medicine. *Drug Development Research*, 2015, 75(5):299-312.
29. Lu Ai-ni, Wang De-long, Zhao Fang, Chen Rui-jie, Chen Wei-qiao, Zheng Hong-bin, Ji Xu-ming, Mechanism of Shaoyao Decoction in treatment of ulcerative colitis based on network pharmacology and molecular docking technology, *Chinese Traditional and Herbal Drugs*, 2020, 51(23): 6035-6044.
30. Liu Jiahui, Sun Baoguo, Kuang Weihong, Zhou houming, Study on Mechanism of Portulacae Herba in the Treatment of Hepatocellular Carcinoma based on Network Pharmacology and Molecular Docking, *Journal of Chinese Medicinal Materials*. 2020, 43(12): 3017-3023.
31. Liu Yi, Li Fei, Jia Yuejin, Hao Shifei, Exploration of Mechanisms of Danzhixiaoyao Powder in Treatment of Insomnia Disorder Based on Network Pharmacology and Molecular Docking, *Clinical*

- Journal of Traditional Chinese Medicine, 33(6): 1104-1112, 2021.
32. Yang Yi, Guo Chun, Lin Xiaoyuan, Yi Jian, Shao Le, Long Hongping, The Anti-tumor Mechanism of Main Chemical Components of *Prunellae Spica* Based on Network Pharmacology, *Anti-tumor Pharmacy*, December 2020, 10(6): 675-680.
 33. Shen Xinhui, Yang Yufeng, Shi Yan, Network Pharmacological Study on the Potential Targets and Mechanisms of Wuling Powder for Diabetic Nephropathy, *Journal of Traditional Chinese Medicine*, 2020, 61(24): 2190-2196.
 34. D Zhang, Xia C, Xu C, et al. Improving Distantly-Supervised Named Entity Recognition for Traditional Chinese Medicine Text via A Novel Back-labeling Approach. *IEEE Access*, 2020, 8:145413-145421.
 35. Ren Xing, Yang Jing, Zhang Wantong, Li Rui, Gao Rui, Medication Rules and Mechanisms of Carotid Atherosclerosis Based on Data Mining and Network Pharmacology, *Traditional Chinese Drug Research & Clinical Pharmacology*, 2021, 32(4): 518-525
 36. Ga Z C, San Z J, Guo W C, et al. Network pharmacology research on high frequency use of Tibetan medicine in treatment of HAPC based on data mining. *China journal of Chinese materia medica*, 2019, 44(21):4756-4767.
 37. Haiya Ou, Xiaopeng Ye, Shu Li, et al. Study on Medication Rules of Herbs in the Regulation of Ferroptosis Based on Network Pharmacology and Data Mining. *Chinese Journal of Modern Applied Pharmacy*, 2019, 36(18): 2317-2324.
 38. Pan Shu-mao, Zhang Xin-you, Zhang Yu-jiao, Niu Xiao-lu, Wu Di-yao, Li Xiu-yun, Ding Liang, The Medication Rules of Traditional Chinese Medicine Prescription against Primary Hepatic Carcinoma Based on Data Mining and Network Pharmacology, *Science Technology and Engineering*, 2021, 21(13) : 05275-12.
 39. Zheng Yi-fu, Kong Ling-lei, Jia Hao, Zhang Bao-yue, Wang Zhe, Xu Lü-jie, Liu Ai-liu, Du Guan-hua, Network pharmacology study on anti-stroke of Xiaoshuan Tongluo formula based on systematic compound-target interaction prediction models, *Acta Pharmaceutica Sinica* 2020, 55(2): 256-264.
 40. Kong Y, Yu T. forgeNet: a graph deep neural network model using tree-based ensemble classifiers for feature graph construction. *Bioinformatics*, 2020, 36(11):3507-3515.
 41. Wang Jing-long, Yang Bin, Zheng Dan-dan, Zhang Li-hua, Fu Xian-jun, Zhang Yong-qing, Sun Xiu-mei, Zhao Zhong-xi, The possible mechanisms of Erhuang decoction in the treatment of acute lung injury based on network pharmacology, *Acta Pharmaceutica Sinica* 2021, 56(1): 244 -256
 42. Mysinger M M, Carchia M, Irwin J J, et al. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 2012, 55(14):6582
 43. Furey T S, Cristianini N, Duffy N, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 2000, 16(10):906-14
 44. Breiman L. Random forest. *Machine Learning*, 2001, 45:5-32
 45. Liu H, Zhang N, Jin S, et al. Small sample color fundus image quality assessment based on gcforest. *Multimedia Tools and Applications*, 2020, 80:17441–17459

Figures

Figure 1

ROC Performances of four methods with Data1 by 3-fold cross validation method.

Figure 2

ROC Performances of four methods with Data1 by 5-fold cross validation method.

Figure 3

ROC Performances of four methods with Data1 by 10-fold cross validation method.

Figure 4

ROC Performances of four methods with Data1 by leave-one-out method.

Figure 5

ROC Performances of four methods with Data2 by 3-fold cross validation method.

Figure 6

ROC Performances of four methods with Data2 by 5-fold cross validation method.

Figure 7

ROC Performances of four methods with Data2 by 10-fold cross validation method.

Figure 8

ROC Performances of four methods with Data2 by leave-one-out method.

(a)

(b)

(c)

(d)

Figure 9

AUC performances of four methods by 3-fold (a), 5-fold (b), 10-fold (c) and leave-one-out (d) methods with full feature set (blue) and control feature set (red).

(a)

(b)

(c)

(d)

Figure 10

F1 performances of four methods by leave-one-out (a), 3-fold (b), 5-fold (c), and 10-fold (d) methods with full feature set (blue) and control feature set (red).

Figure 11

The flowchart of forgeNet algorithm.

Figure 12

The flowchart of drug components-disease network related to acute lung injury inference.