Disease-related compound identification based on deeping learning method

Bin Yang¹, Wenzheng Bao²,*, Jinglong Wang¹, Baitong Chen*, Naoki Iwamori⁵, Jiazi Chen⁵* & Yuehui Chen⁶

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. 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. The experimental results show that forgeNet performs better than support vector machines (SVM), random forest (RF), logical regression (LR), Naive Bayes (NB), XGBoost, LightGBM and gcForest. forgeNet could identify 19 compounds in Erhuang decoction (EhD) and Dexamethasone (DXMS) more accurately.

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. Under certain conditions, the abnormal life activity processes caused by the disturbance of homeostasis after the damage of the disease cause the disease. Traditional Chinese medicine (TCM) has been utilized to treat diseases for thousands of years. 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.

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. 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. 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, shock, convulsion, hemorrhage, acute respiratory failure, renal failure, heart failure, cerebrovascular accident, 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 Portulacaceae 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\(^{31}\). Yu 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 Caspase 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 ALI-related target genes (REAL and SATA3), which have been verified biology experiment. The active compounds are collected from BindingDB database for two key target genes as positive samples. The inactive compounds are generated from DUD-E as the negative samples. The different molecular descriptors and molecular fingerprints are utilized to characterize each compound, which form the full feature set and contain 374 features. With full feature set collected, forest graph embedded deep feed forward network is trained, which is utilized to identify the compounds in Erhuang decoction (EhD) and Dexamethasone (DXMS) for the treatment of acute lung injury.

### Methods

**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. 1. From Fig. 1, 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

The flowchart of feature graph construction is depicted in Fig. 2. Before the labeled training data are input into classifier, the features of the data need to be extracted. In forgeNet, the used forest $\xi$ contains $p$ decision tree (DT). With the labeled training data, the forest is fitted and $p$ DT are generated ($\xi(\theta) = \{T_1(\theta_1), T_2(\theta_2), \ldots, T_p(\theta_p)\}$, $\theta_i$ is a parameter). Meanwhile if binary tree is regarded as a special case of directed graph, we can gain the following graph set.

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

where $V_i$ and $E_i$ represents vertex set and edge set of $G_i$.

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

$$G(V, E) = \bigcup_{i=1}^{p} G_i. \quad (2)$$

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\textsuperscript{12}. Every layer of GEDFN is introduced as followed.

$$Z_1 = \sigma (X(W_{in}\Theta) + b_{in}),$$

$$\ldots$$

$$Z_{k+1} = \sigma (Z_k W_k + b_k),$$

$$\ldots$$

$$Z_{out} = \sigma (Z_l W_l + b_l),$$

$$y = \text{soft max}(Z_{out} W_{out} + b_{out}). \quad (3)$$

where $X$ is input data, $Z_k$ is the $k-th$ hidden layers, $\Theta$ denotes Hadamard product, $W_k$ and $b_k$ are the weights and bias of the $k-th$ hidden layer, respectively. $\sigma(\cdot)$ is an activation function, which could be sigmoid, hyperbolic tangent or rectifiers.
**Inference algorithm.** In order to construct "disease-gene-target-drug" interaction network more accurately, an ALI-related compound identification based on deep learning model and target genes is proposed. The flowchart is depicted in Fig. 3 and the detailed process is given as follows.

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 loosely related to ALI diseases in the literature. Then the BindingDB database (http://www.bindingdb.org/bind/index.jsp) is searched for the known active compounds of these two key target genes. The active ligands are screened with the condition that IC50 < 5000 nmol 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. Active and inactive compound sets form the dataset. The structure of each compound is Simplified Molecular Input Line Entry System (SMILE), so the molecular descriptors and molecular fingerprints of each compound must be obtained as the feature vectors. In this paper, RDKit package is utilized to create the molecular descriptors and molecular fingerprints of each ligand. Molecular descriptors contains 208 features, such as topological polar surface area (TPSA) descriptor, number of valence electros, number of radical electrons, charge information and number of Aliphatic Carbocycles. MACCS fingerprints contains 166 molecular characteristic sites, such as Atom Pairs, topological torsions.

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.

**Experiments**

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, RF, logical regression (LR), Naive Bayes (NB), XGBoost, LightGBM and gcForest) are utilized to identify the compounds associated with diseases. Five evaluation criteria of classifier performance are utilized, which are SN, SP, Kappa, MCC and F1, respectively.

**Model test.** In order to test the generalization and stability of forgeNet, threefold, fivefold and tenfold cross validation methods are utilized. For each cross validation method, 10-repeat experiments are implemented. Identification averaged performances (Mean ± SD) of eight methods with Data1 and Data2 by threefold cross validation, fivefold and tenfold cross validation methods are listed in Tables 1, 2 and 3, respectively. For Table 1, with Data1 it could be seen that NB algorithm has the best SN performance, which is 0.9111 ± 0.021. In terms of SP, Kappa, MCC and F1, LightGBM performs better than other seven methods and forgeNet has the second better performances. With Data2, NB also obtain the highest SN performance, which shows that this method could identify more true ALI-related compounds than other methods, but NB also obtain the worst SP performance, which reveals that this method identifies most of compounds as related ones. From Tables 2 and 3, we also see that with Data1 and Data2, NB algorithm could obtain the best SN performances by fivefold cross validation and tenfold cross validation methods. forgeNet could obtain the highest SP, which shows that this method could
identify more unrelated-disease compounds. Although forgeNet can identify less true related compounds than NB, this method could obtain the higher accuracy according to MCC performances. Kappa performances show that forgeNet can make the prediction results more consistent with the actual classification ones with the unbalanced data.

F1 performances show that on the whole forgeNet could infer components-disease network more accurately than other seven classifiers. Standard Deviation performances of forgeNet also show that this method could obtain the more stable performances.

Receiver operating characteristic (ROC) and Precision-Recall (PR) curves are two important curves to evaluate the performance of machine learning algorithm. ROC curve is based on false positive rate (FPR) and true positives rate (TPR). PR curve is based on Recall and Precision. Area under curve (AUC) is defined as the area under the ROC curve or PR curve surrounded by the coordinate axis. PR and ROC curves of eight methods with Data1 and Data2 by threefold cross validation, fivefold and tenfold cross validation methods are depicted in Figs. 4, 5 and 6, respectively. From Fig. 4, it could be seen that LightGBM performs best with Data1 in terms of PR and ROC curves. forgeNet could obtain the second better performances. With Data2, forgeNet could obtain the best ROC and PR curves. From Fig. 5, with Data1, gcForest, LightGBM and forgeNet have the similar PR and ROC curves. From AUC values, it could be seen that forgeNet performs better than gcForest and LightGBM.

### Table 1. Identification performances of eight methods with Data1 and Data2 by threefold cross validation method. Significant values are in bold.

| Datasets | Methods | SN       | SP       | Kappa   | MCC        | F1          |
|----------|---------|----------|----------|---------|------------|-------------|
| Data1    | RF      | 0.8789 ± 0.038 | 0.8886 ± 0.0312 | 0.6385 ± 0.043 | 0.6597 ± 0.031 | 0.7042 ± 0.031 |
|          | SVM     | 0.8541 ± 0.018 | 0.898 ± 0.022 | 0.6413 ± 0.042 | 0.657 ± 0.035 | 0.7053 ± 0.032 |
|          | LR      | 0.8856 ± 0.018 | 0.9043 ± 0.012 | 0.6736 ± 0.025 | 0.6894 ± 0.022 | 0.7136 ± 0.012 |
|          | NB      | 0.9111 ± 0.021 | 0.6129 ± 0.014 | 0.2822 ± 0.008 | 0.3761 ± 0.01 | 0.446 ± 0.006 |
|          | gcForest | 0.9103 ± 0.024 | 0.9154 ± 0.021 | 0.7145 ± 0.04 | 0.7924 ± 0.033 | 0.7648 ± 0.031 |
|          | XGBoost | 0.8563 ± 0.034 | 0.8837 ± 0.03 | 0.6147 ± 0.06 | 0.635 ± 0.05 | 0.6848 ± 0.04 |
|          | LightGBM | 0.8809 ± 0.037 | 0.9388 ± 0.022 | 0.7523 ± 0.036 | 0.7794 ± 0.03 | 0.7936 ± 0.028 |
|          | ForgeNet | 0.902 ± 0.021 | 0.9277 ± 0.019 | 0.7369 ± 0.027 | 0.747 ± 0.023 | 0.7819 ± 0.024 |

### Table 2. Identification performances of eight methods with Data1 and Data2 by fivefold cross validation method. Significant values are in bold.

| Datasets | Methods | SN       | SP       | Kappa   | MCC        | F1          |
|----------|---------|----------|----------|---------|------------|-------------|
| Data1    | RF      | 0.8733 ± 0.024 | 0.918 ± 0.016 | 0.6965 ± 0.025 | 0.7076 ± 0.02 | 0.749 ± 0.019 |
|          | SVM     | 0.8596 ± 0.023 | 0.905 ± 0.02 | 0.6599 ± 0.03 | 0.6737 ± 0.028 | 0.7198 ± 0.026 |
|          | LR      | 0.872 ± 0.019 | 0.9212 ± 0.021 | 0.7041 ± 0.04 | 0.7144 ± 0.034 | 0.7549 ± 0.032 |
|          | NB      | 0.9222 ± 0.009 | 0.5912 ± 0.008 | 0.2687 ± 0.004 | 0.3678 ± 0.004 | 0.4373 ± 0.003 |
|          | gcForest | 0.9158 ± 0.019 | 0.9233 ± 0.012 | 0.7353 ± 0.03 | 0.7475 ± 0.03 | 0.7813 ± 0.02 |
|          | XGBoost | 0.8669 ± 0.04 | 0.901 ± 0.032 | 0.6575 ± 0.048 | 0.6744 ± 0.037 | 0.7184 ± 0.036 |
|          | LightGBM | 0.898 ± 0.03 | 0.9379 ± 0.02 | 0.7603 ± 0.04 | 0.7681 ± 0.03 | 0.8008 ± 0.03 |
|          | ForgeNet | 0.8999 ± 0.018 | 0.9426 ± 0.014 | 0.7703 ± 0.02 | 0.7769 ± 0.021 | 0.8086 ± 0.02 |

| Datasets | Methods | SN       | SP       | Kappa   | MCC        | F1          |
|----------|---------|----------|----------|---------|------------|-------------|
| Data2    | RF      | 0.8848 ± 0.034 | 0.9069 ± 0.0241 | 0.6504 ± 0.039 | 0.6711 ± 0.029 | 0.7084 ± 0.03 |
|          | SVM     | 0.7935 ± 0.02 | 0.8974 ± 0.011 | 0.5842 ± 0.02 | 0.5982 ± 0.018 | 0.6615 ± 0.016 |
|          | LR      | 0.8409 ± 0.021 | 0.8758 ± 0.028 | 0.5717 ± 0.048 | 0.5967 ± 0.039 | 0.6464 ± 0.037 |
|          | NB      | 0.9383 ± 0.008 | 0.3601 ± 0.009 | 0.1123 ± 0.005 | 0.2207 ± 0.007 | 0.3154 ± 0.004 |
|          | gcForest | 0.8379 ± 0.03 | 0.8965 ± 0.03 | 0.6137 ± 0.05 | 0.6321 ± 0.04 | 0.6774 ± 0.04 |
|          | XGBoost | 0.8647 ± 0.036 | 0.8803 ± 0.022 | 0.5934 ± 0.031 | 0.6194 ± 0.024 | 0.6626 ± 0.023 |
|          | LightGBM | 0.8968 ± 0.027 | 0.9002 ± 0.028 | 0.6565 ± 0.046 | 0.6783 ± 0.034 | 0.7136 ± 0.035 |
|          | ForgeNet | 0.8956 ± 0.026 | 0.9044 ± 0.017 | 0.6633 ± 0.035 | 0.6829 ± 0.03 | 0.7192 ± 0.028 |
With Data2, LightGBM and forgeNet have the similar PR and ROC curves. Figure 6 also shows that forgeNet could perform better than other classifiers for compound identification.

**Compound screening for traditional Chinese medicine prescription.** Erhuang decoction (EhD) is a traditional heat clearing and detoxifying prescription, which is composed of Radix Scutellariae, Rhizoma Cop-
tidis and licorice. 19 active chemical compounds (Neoglycyrol, Uralenol, Syringic acid 4-β-d-Glucopyranoside, Gannaconin 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 Glycyrrhetinic acid) in EhD can dock with ALI related target genes and have high potential biological activity, which have been proved in the reference. 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. SVM, RF and gcForest are selected as comparison methods. The prediction ranks are listed in Table 4. 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. Thus the results reveal that forgeNet could screen the chemical compounds more accurately than SVM, RF and gcForest. We also analyze the mechanism of action of the highly ranked compounds. Glycyrrhizin has a protective effect on acute lung injury through the activation and increase of Nrf2 nuclear translocation. Baicalin plays a role in regulating the inflammatory response of ALI by stimulating regulatory T cells and inhibiting the release of IL6 and interleukin-23, which could lead to the decrease of Th17 (T helper cell 17) cells in order to affect the immune balance between Th17 and Treg response. Baicalein can down regulate the mRNA expression of STAT3 and STAT4 in T cell JAK STAT signal pathway in order to promote T cell proliferation, and play an immune and anti-inflammatory role.

Performance test of different feature sets. 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 threefold, fivefold, tenfold and leave-one-out methods. The AUC and F1 results are depicted in Figs. 7 and 8, respectively. From the results, it could be seen that full feature set could improve the compound identification accuracy of methods.

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...
Figure 6. PR curves and ROC curves of eight methods with Data1 and Data2 by threefold cross validation method.

Table 4. Prediction ranks of 20 chemical compounds by SVM, RF, gcForest and forgeNet. Significant values are in bold.
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 threefold, fivefold and tenfold cross validation methods, the experimental results show that forgeNet has the better performance than SVM, RF, LR, NB, XGBoost, LightGBM and gcForest in terms of SN, SP, Kappa, MCC, F1, AUC, ROC curve and PR curve. 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.

In the further we will apply the method to prioritize the compounds in other ALI-related and other diseases related TCM prescriptions.
Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Received: 7 January 2022; Accepted: 15 November 2022
Published online: 29 November 2022

References

1. Hall, C. B., Simões, E. A. F. & Anderson, L. J. Clinical and epidemiologic features of respiratory syncytial virus. In Challenges and Opportunities for Respiratory Syncytial Virus Vaccines Current Topics in Microbiology and Immunology Vol. 372 (eds Anderson, L. & Graham, B.) 39–57 (Springer, 2013).
2. Michael, A. et al. Initial evaluation of the patient with lung cancer: Symptoms, signs, laboratory tests, and paraneoplastic syndromes. Chest 123(1), 97S–104S (2003).
3. Sobhani, M. & Bechara, A. A somatic marker perspective of immoral and corrupt behavior. Soc. Neurosci. 6(5–6), 640–652 (2011).
4. Ekpenyong, C. Abnormal serum uric acid levels in health and disease: A double-edged Sword. Am. J. Internal Med. 2(6), 113–130 (2014).
5. Bensoussan, A., Myers, S. P. & Carlton, A. L. Risks associated with the practice of traditional Chinese medicine: An Australian study. Arch. Fam. Med. 9(10), 1071–1078 (2000).
6. Li, S. et al. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. IET Syst. Biol. 1(1), 51–60 (2007).
7. Wang, X. et al. Potential role of metabolomics approaches in the area of traditional Chinese medicine: As pillars of the bridge between Chinese and Western medicine. J. Pharm. Biomed. Anal. 55(5), 859–868 (2011).
8. Xue, R. et al. TCMD: Traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res. 41(D1), D1089–D1095 (2013).
9. Hou, Y. et al. Correction: Smart Soup, a Traditional Chinese Medicine Formula, Ameliorates Amyloid Pathology and Related Cognitive Deficits. PLoS ONE 15(8), e0237035 (2020).
10. Ren, B. et al. Possible mechanisms underlying treatment of Alzheimer’s disease with Traditional Chinese Medicine: active components, potential targets and synthetic pathways of Bulao Elixir. J. Trad. Chin. Med. 40(3), 145–157 (2020).
11. Xu, W. et al. Traditional Chinese medicine in cancer care: perspectives and experiences of patients and professionals in China. Eur. J. Cancer Care 15(4), 397–403 (2010).
12. Che, C., Man, S. W. & Lam, C. Natural products from Chinese medicines with potential benefits to bone health. Molecules 21(3), 239 (2016).
13. Zhang, J. et al. Extraction methods of natural products from traditional Chinese medicines. Methods Mol. Biol. 1263, 177–185 (2015).
14. Alromaima, A. et al. Advances in the treatment of novel coronavirus disease (COVID-19) with Western medicine and traditional Chinese medicine: A narrative review. J. Thorac. Dis. 12(10), 6054–6069 (2020).
35. Ga, Z. C.
36. Haiya, Ou.
37. Pan, S.
38. Zheng, Y.
39. Kong, Y. & Yu, T.
40. Wang, J.
41. Mysinger, M. M.
42. Furey, T. S.
43. Breiman, L.
44. Liu, H.
45. Wang, G.
46. Liu, T.

Author contributions
W.B. conceived the method. B.Y. designed the method. J.W., N.I., J.C. and B.C. conducted the experiments. W.B. and Y.C. wrote the main manuscript text. All authors reviewed the manuscript.

Funding
This work was supported by the talent project of “Qingtan 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. 2019KJMO06), 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).

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to W.B. or J.C.
