

Research Article

An Ontology-Based Artificial Intelligence Model for Medicine Side-Effect Prediction: Taking Traditional Chinese Medicine as an Example

Yuanzhe Yao,¹ Zeheng Wang ^{1,2}, Liang Li,¹ Kun Lu,³ Runyu Liu,¹ Zhiyuan Liu,¹ and Jing Yan⁴

¹School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

²School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia

³Faculty of Medicine, Ludwig Maximilian University of Munich, Munich 81377, Germany

⁴The First Clinical Medical College, Zhejiang Chinese Medicine University, Hangzhou 310053, China

Correspondence should be addressed to Zeheng Wang; zenwang@outlook.com

Received 26 March 2019; Revised 30 June 2019; Accepted 30 July 2019; Published 1 October 2019

Academic Editor: Michele Migliore

Copyright © 2019 Yuanzhe Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this work, an ontology-based model for AI-assisted medicine side-effect (SE) prediction is developed, where three main components, including the drug model, the treatment model, and the AI-assisted prediction model, of the proposed model are presented. To validate the proposed model, an ANN structure is established and trained by two hundred forty-two TCM prescriptions. These data are gathered and classified from the most famous ancient TCM book, and more than one thousand SE reports, in which two ontology-based attributions, hot and cold, are introduced to evaluate whether the prescription will cause SE or not. The results preliminarily reveal that it is a relationship between the ontology-based attributions and the corresponding predicted indicator that can be learnt by AI for predicting the SE, which suggests the proposed model has a potential in AI-assisted SE prediction. However, it should be noted that the proposed model highly depends on the sufficient clinic data, and hereby, much deeper exploration is important for enhancing the accuracy of the prediction.

1. Introduction

Artificial intelligence is a modern technology that is utilized in various fields of medicine [1–3]. At the meantime, traditional Chinese medicine (TCM) is now widely considered as a promising alternative medicine for complementary treatment in cancers or chronic diseases due to the effective methodology practically developed by generations of doctors for almost 4000 years [4]. Based on previous verification, it is undeniable that there are many correlations between the TCM syndromes and western diseases, turning out novel approaches for enhancing the treatment efficiency and developing medicines regarding TCM methodologies [5]. Unfortunately, hindered by the remarkable gap between the modern informatics and the fundament of TCM—ancient

Chinese philosophy, such correlations are still too elusive to be formulated precisely.

Recently, in order to figure out the deep connection between modern science and TCM, the research combining TCM with AI for valid knowledge acquisition and mining attracts great attention, thereby leading to many profound works, such as ontology information system design [6], latent tree models design [7], TCM warehouse for AI application [8], and digital knowledge graph development [2]. Especially, in the view of algorithms, these AI-assisted techniques can be recognized by two different approaches: pattern classification and knowledge mining. The former technology attempts to recognize the correct pathological information such as pulse condition [9–14] and tongue diagnosis [15] of an individual patient. However, the later

one, knowledge mining, mainly focuses on finding out various kinds of hidden relationships in the knowledge, for example, the relationships between symptom and symptom, symptom and syndrome, and syndrome and disease [16–20]. In addition, it should be noted that there are many other studies that deserve attention as well, such as classifying herbs by the convolutional neural network model [21] using the deep learning mode to explore the relationship between herbal property and action [22].

On the contrary, researchers face, however, many difficulties in setting up AI for TCM in terms of directly interpreting the TCM semantic system (almost recorded by ancient Chinese doctrines) into structured database. However, in this way, considerable workload must be undertaken by limited numbers of experts who are proficient in both AI and TCM to translate the TCM terminologies and then formulate the modern model thereof. In contrast, as shown in Figure 1, using TCM methodology, but not the modern one, in overcoming the barriers of modern science, designing new medicine, for example, is relatively lacking and thus of significant worth to explore.

In this paper, an ontology-based model is developed to train AI for drug side-effect (SE) prediction, in which the methodology of TCM including syndromes differentiation is applied to determine the ontology-based attributions and optimize the AI components, and consequently, form a novel scheme of effectively predicting the medicine’s attribution. Here, limited by the shortage of accurate clinic experiment data of modern medicine, TCM data in famous ancient books are used to verify the model which shows a tremendous potential in medicine discovery. The paper is organized as follows: in Section 2, three main components, including the drug model, the treatment model, and the AI-assisted prediction model, are established to introduce how to use TCM theory to explore the modern drugs; then in Section 3, an artificial neural network- (ANN-) based AI model is established and trained by the collected data; and in Section 4, the prediction performance in the proposed model framework is shown and discussed.

2. Methodology

2.1. Ontology-Based Drug Model. The artificial intelligence model proposed in this paper is based on ontology that considers the essence of a certain entity as a combination of several fundamental attributions with corresponding values and relationships [2, 23, 24]. Such attributions are not only the definite properties which are already completely recognized by researchers but also the latent properties including unknown information and relationships.

For example, as shown in Figure 2, each drug has certain attributions including the definite ones and latent others, which are all involved in a certain prescription with sufficient records of clinical effects. In addition, assuming our prepared ontology system is complete and exclusive, a new drug or prescription which contains attributions we have already recorded can be depicted easily in the ontology-based semantic system, where we could focus on the superficial relationship between such attributions, or in another word labels, and

effects caused thereby. In other words, we avoid literally figuring out the ingredient and other deeper properties of each attribution in the new drug, and therefore, the attribution-effect pair is crucial and could be easily converted into an AI scheme such as ANN to handle the prediction of the treatment. Moreover, the proposed ontology-based attribution model could be revised by more accurate clinic records automatically with AI assistance due to the intentionally fuzzy and dynamic defined latent attributions. In this paper, as discussed later, two items, hot and cold, are presented as the fundamental attributions of any medicines.

2.2. Ontology-Based Treatment Model. Based on the proposed drug model, it can be depicted, as in Figure 3, that the model of the treatment procedure via a certain prescription X. This prescription contains several drugs including the attributions of known ingredients and the latent attributions. As shown in Figure 3, the latent attributions own the capability of influencing the group of indicators with different unknown paths and efficiencies. In another word, in this model, the results of the treatment of a certain patient X that is defined as the positive or negative change of the corresponding indicator are the comprehensive synthesis of the effects induced by various latent attributions. Therefore, this procedure could be interpreted into TCM-based semantic entities: attribution-indicator pairs performing the effects. It should be noted that the different attributions maybe dominate in influencing the same indicator. Furthermore, the model is compatible with the known ingredients or explored attributions and the effects thereof.

2.3. AI-Assisted Prediction Model. Based on the aforementioned drug/prescription and treatment models, as illustrated in Figure 4, the SE prediction of new drug X is realized by comprehensive consideration of the involved ontology-based latent attributions with their influential factors (IFs) revised by sufficient medicines’ clinic records that contain, for instance, the attribution no. 3 and X, where the revision procedure could be undertaken by an AI scheme such as ANN. Also, the same AI scheme could predict the SE with the trained pattern.

It should be noted that the IFs must be linked with the corresponding attributions and indicators which means the trained model is consisted of IFs’ indicator vectors but not the isolated IFs as the input. In this way, the ontology-based model that the latent attributions with corresponding IFs influence a certain indicator is established. Next, we will generate an AI scheme to validate our proposed model by determining two latent attributions, which are hot and cold of the prescription, and a simple indicator: whether the prescription causes SE or not when this prescription is used in a right way.

3. Experiment Detail

According to the analysis in Section 2, it is the key for establishing the proposed model that determines the attributions and obtains the IFs’ indicator vectors. However, owing to the lack of related theory, generating the

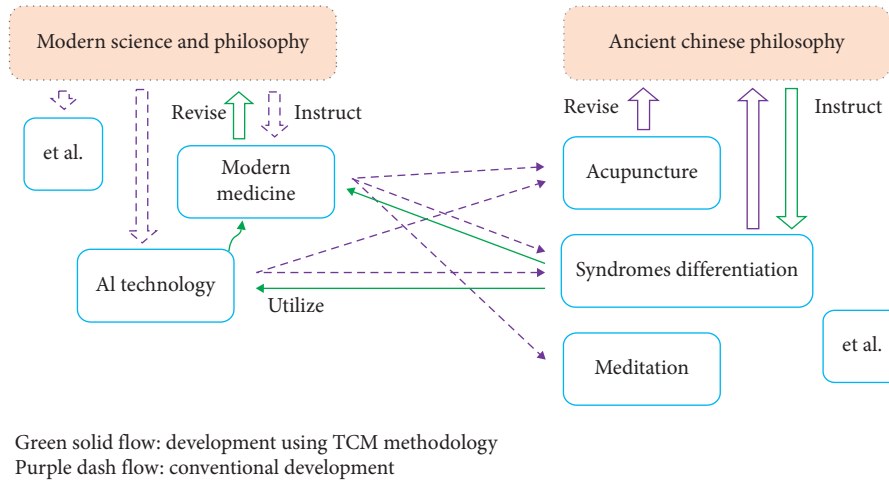


FIGURE 1: The development procedure based on modern science and the TCM-based ontology.

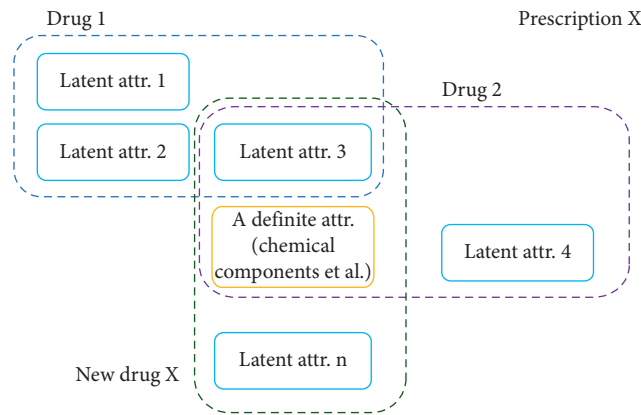


FIGURE 2: Ontology-based drug model and latent attributions thereof.

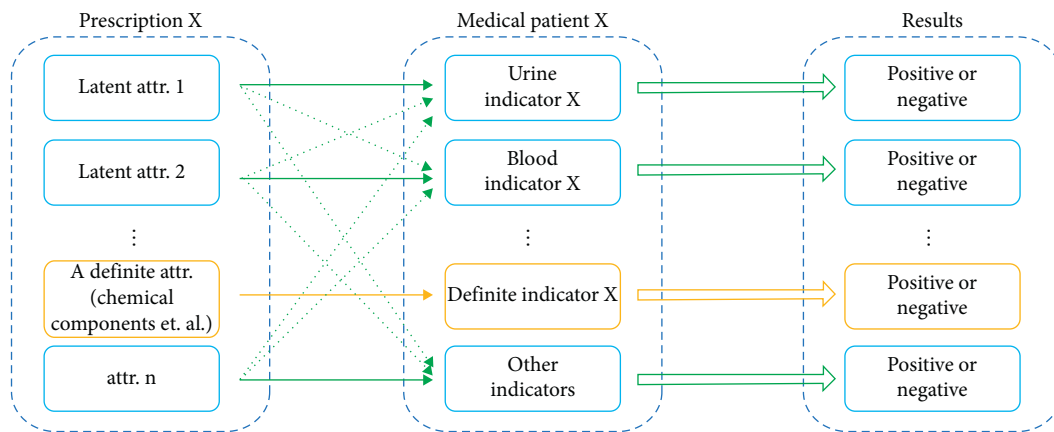


FIGURE 3: Ontology-based treatment model concerning the attribution-indicator relationships.

attributions directly, comprehensively, and exclusively is very hard. Therefore, we follow the theory of TCM which has the advantage in the matured ontology-based semantic system that can determine the attributions spontaneously. For example, hot and cold are two main attributions cataloged by TCM theory, where all the drugs must contain one out of these two attributions, leading to a charming

approach for determining the latent attributions of western drugs in the same way.

As shown in Figure 5, after the identification of the attributions and indicators, we should establish and train the AI model. Here, we gathered the detailed data, including 150 effective prescriptions, the dosages thereof, and the corresponding indicators from the famous ancient TCM book

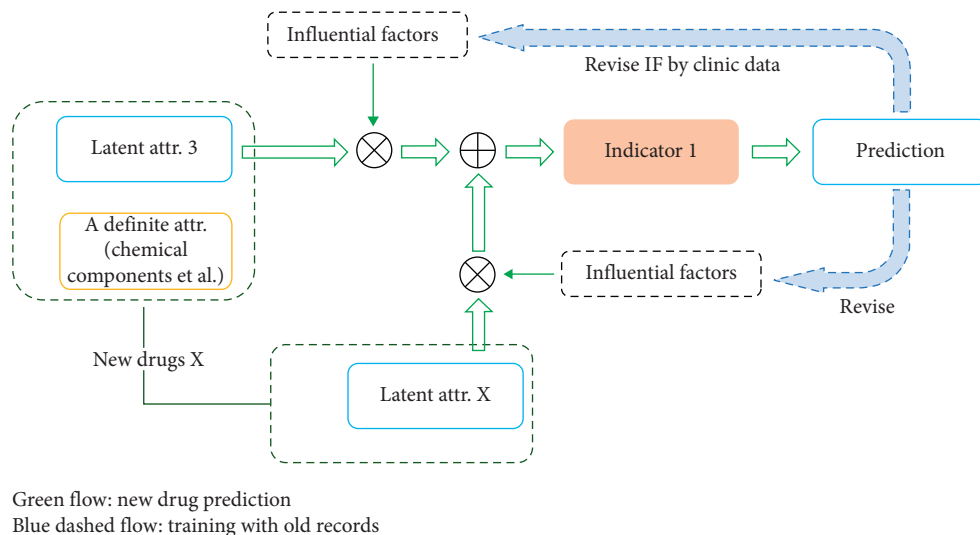


FIGURE 4: The network for training AI using proposed models.

Shanghanzabinglun (*Treatise on Cold Pathogenic and Miscellaneous Diseases*) which is considered as the origin of practical TCM prescription in clinic. In addition, as concluded before, according to the practice identification by ancient TCM doctors and the TCM standards published by Chinese government [25, 26], we labeled two ontology-based attributions that are hot and cold for describing the drugs' fundamental property, which is the first step of conducting the prediction as depicted in Figure 5. Thereafter, we assigned the IFs of each attribution equaling the total dosage of the drugs which own the corresponding attribution in the prescription. Since in the ontology-based labeling procedure, a drug must belong to one certain catalogue out of the two in total, using the summarized dosage to represent the IFs is reasonable; however, it needs more verification in future research. In addition, it should be noted that some prescriptions do not contain any drugs (for example, some uncatalogued pure chemical ingredients) associated with hot or cold, where for convenience these drugs could be considered as neutral ones and not affiliated with the two attributions mentioned before.

As shown in Figure 6(a), 242 effective prescriptions are dotted regarding the normalized total dosage, where the x -axis and y -axis in the figure represent the total hot dosage and the total cold dosage, respectively. These dosages are considered as the IF factors of the prescription. According to our best knowledge, because there are no reports indicating the 150 prescriptions gathered from the book *Shanhanzabinglun*, we consider these prescriptions are the safe prescriptions. In contrast, we gathered 92 unsafe prescriptions reported to frequently cause SE when they are used in a right way.

The distribution of the percentages of the safe prescriptions and the unsafe prescriptions features a huge difference in terms of whether the dosage is stronger than 500, which suggests the reported unsafe prescriptions own the characteristics that can be distinguished from the safe prescriptions. Therefore, we try to use the pattern recognition

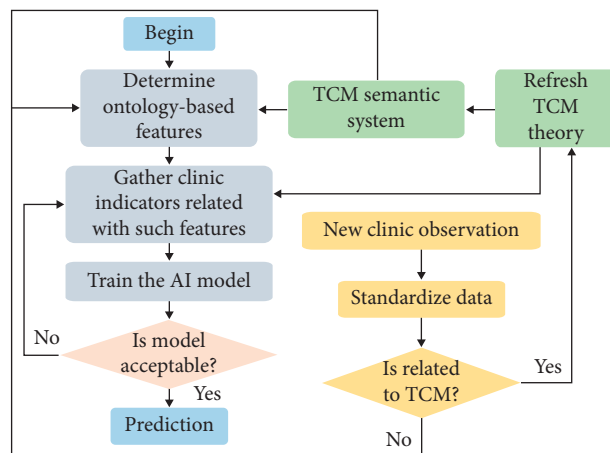


FIGURE 5: The SE prediction procedure of the proposed model.

method to build a simple classifier to predict which prescription is unsafe.

ANN is a classic model in pattern recognition tasks. Due to its good performance and simple form, it is widely used in solving nonlinear classification problems. Here, a multilayer ANN model is developed to learn how to recognize the special pattern from our collected prescription data.

In order to use the ANN model to train this classifier, we represent each prescription into a vector. We analyze each prescription and identify dosages about every single herbal drug which form the prescription. According to "Chinese Pharmacopoeia," we can clarify hot/cold properties of each drug appearing in our collected prescriptions. We use the weighted BOW model to represent prescription, v_p [27, 28]. Furthermore, we generate a weighted matrix according to the BOW model, W . The matrix W has two columns, and each column in W represent a type of property and each row in W represent the hot or cold property on a single drug. We use this model-generated matrix as a linear operator to generate the input vector in

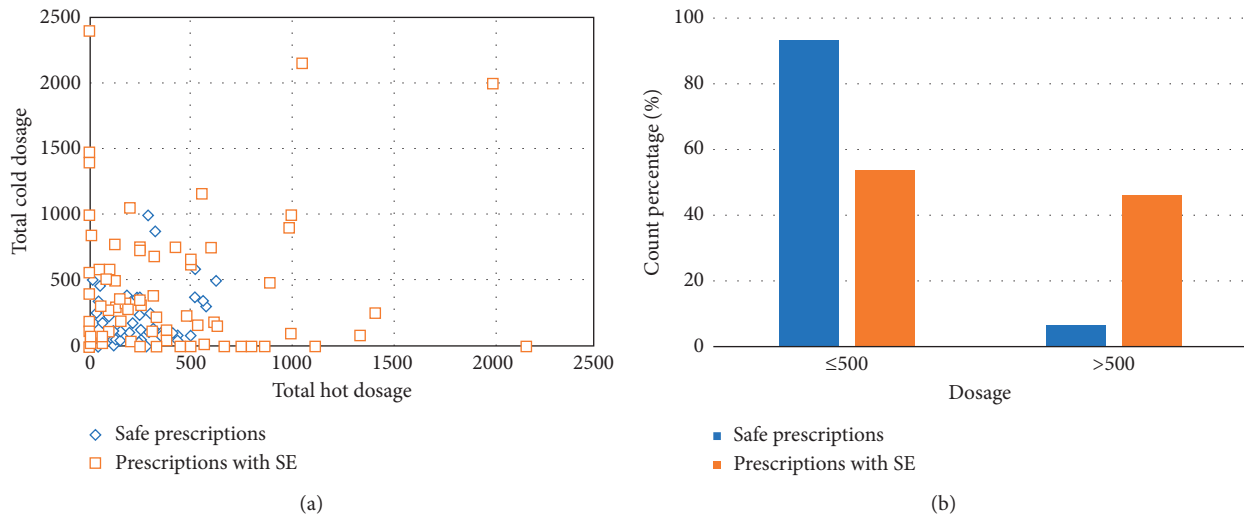


FIGURE 6: The counts of the hot/cold IF (counts) in the book.

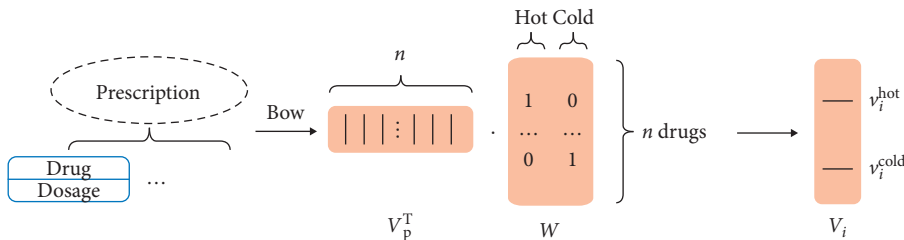


FIGURE 7: The schematic procedure of converting prescription into a vector.

the ANN model. Thus, the input vector of the ANN model can be expressed as follows:

$$v_i = W^T \cdot v_p, \quad (1)$$

where v_i is the input vector in the ANN model. The procedure is shown in Figure 7.

As shown in Figure 8, the model consists of 5 layers where the input and output layers both contain two units for receiving the dosage vectors and accordingly yielding the SE prediction vectors. The three hidden layers that totally have more than 60 units with enough parameters are used to fit the complex relationships among ontology-based attributes, which are cold and hot here, and the affections thereof.

To train this ANN model, we prepared and washed 150 safe prescriptions from the book *Shanhanzabinglun* and 92 reported prescriptions that frequently cause SE as mentioned before [29–34]. For convenience, we adopted 10-fold cross-validation to train our model, and then we got a convincing result as shown in Section 4.

4. Results and Discussion

As seen in Table 1, where bold values highlight the average results obtained in this research, the average accuracy is 87% with a sensitivity rate and a specificity rate of 98% and 17%, respectively. It can be seen in this result that the proposed

classifier has a high performance on predicting positive items. Meanwhile, 87% accuracy also proves the high performance of the classifier. However, the low specificity rate means the classifier features poor capability of distinguishing negative items from all data.

This is because the negative data are less than the positive data. The lack of negative data leads to the failure of our ANN model, learning enough knowledge from the provided samples; therefore, the prediction on negative items is more inaccurate. Another reason that should be noted is the features we extracted from medicine data could not represent the typical ones in the classify decision process. Although the other result may be not good as anticipated, the sensitivity rate is out of expectation. The high accuracy on positive items strongly supports our hypothesis.

Hence, the proposed ontology-based SE prediction model is preliminarily verified by the ANN. However, in this procedure, we did not revise the IFs due to the lack of the dataset, resulting in a weak prediction accuracy. Furthermore, the determination of the attributions is relatively broad. In another word, the attributions may be classified into more detailed catalogue such as hot, warm, neutral, cool, and cold. In this way, the ANN could learn more features of the dataset and give more precise predictions. Besides, according to other factors such as different lengths of treatment, it is of great significance that

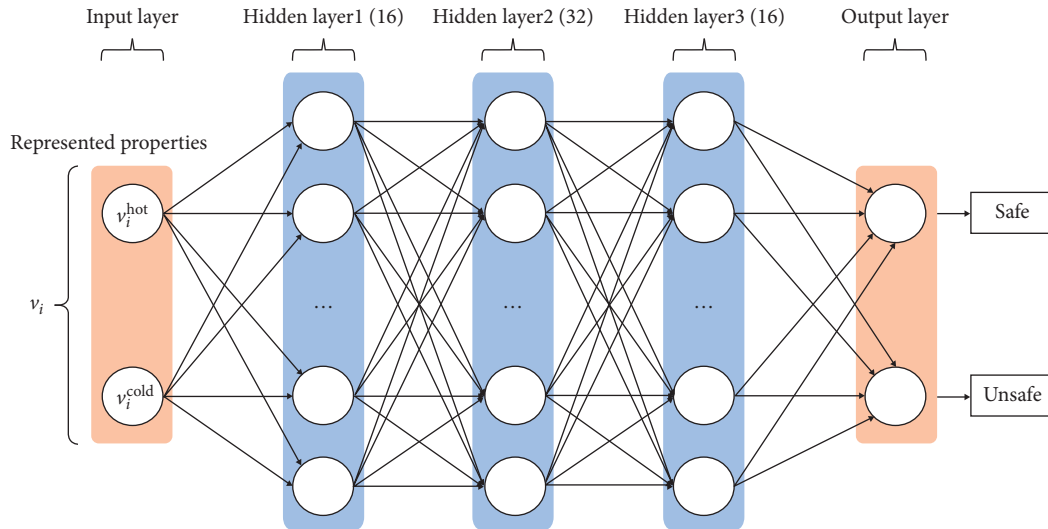


FIGURE 8: The schematic structure of the ANN and the dataflow.

TABLE 1: The results of 10-fold cross-validation.

Fold N	SE	SP	ACC
1	1.00	0.00	0.92
2	1.00	0.33	0.92
3	0.92	1.00	0.92
4	1.00	0.33	0.92
5	0.91	0.00	0.88
6	1.00	0.00	0.79
7	1.00	0.00	0.92
8	1.00	0.00	0.88
9	1.00	0.00	0.79
10	0.95	0.00	0.79
Average	0.98	0.17	0.87

IFs should be evaluated with a weight vector or even tensor, which will influence the results of prescriptions and should be studied in the next stage.

5. Conclusion

An ontology-based model for AI-assisted medicine side-effect prediction is proposed in this paper. The drug, treatment, and prediction models are established to describe the methodology. In addition, the SE prediction is carried out and verified by the ANN, in which a simplified scheme containing latent attributions (cold and hot) and corresponding indicators (with or without SE) is investigated preliminarily. Clinic data coming from both safe and unsafe prescriptions are adopted to train the ANN and thereafter predict SE. The success of predicting whether a prescription will cause SE demonstrates the simplicity and effectiveness of this work, which should, however, be further improved as a powerful tool to predict more side-effect syndrome.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61370202.

References

- [1] Y. Feng, Z. Wu, X. Zhou, Z. Zhou, and W. Fan, "Knowledge discovery in traditional Chinese medicine: state of the art and perspectives," *Artificial Intelligence in Medicine*, vol. 38, no. 3, pp. 219–236, 2006.
- [2] T. Yu, J. Li, Q. Yu et al., "Knowledge graph for TCM health preservation: design, construction, and applications," *Artificial Intelligence in Medicine*, vol. 77, pp. 48–52, 2017.
- [3] X. Zhou, B. Liu, Z. Wu, and Y. Feng, "Integrative mining of traditional Chinese medicine literature and MEDLINE for functional gene networks," *Artificial Intelligence in Medicine*, vol. 41, no. 2, pp. 87–104, 2007.
- [4] M. Jiang, C. Lu, C. Zhang et al., "Syndrome differentiation in modern research of traditional Chinese medicine," *Journal of Ethnopharmacology*, vol. 140, no. 3, pp. 634–642, 2012.
- [5] X. Zhou, J. Menche, A. L. Barabási, and A. Sharma, "Human symptoms-disease network," *Nature Communications*, vol. 5, no. 1, 2014.
- [6] X. Zhou, Z. Wu, A. Yin, L. Wu, W. Fan, and R. Zhang, "Ontology development for unified traditional Chinese medical language system," *Artificial Intelligence in Medicine*, vol. 32, no. 1, pp. 15–27, 2004.
- [7] N. L. Zhang, S. Yuan, T. Chen, and Y. Wang, "Latent tree models and diagnosis in traditional Chinese medicine," *Artificial Intelligence in Medicine*, vol. 42, no. 3, pp. 229–245, 2008.
- [8] X. Zhou, S. Chen, B. Liu et al., "Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support," *Artificial Intelligence in Medicine*, vol. 48, no. 2-3, pp. 139–152, 2010.
- [9] H. Guo and B. Wang, "Research on TCM pulse condition identification using probabilistic neural networks," in

- Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics*, pp. 2352–2355, IEEE, Yantai, Chinadoi, October 2010.
- [10] A. C. Y. Tang, J. W. Y. Chung, and T. K. S. Wong, “Digitalizing traditional Chinese medicine pulse diagnosis with artificial neural network,” *Telemedicine and e-Health*, vol. 18, no. 6, pp. 446–453, 2012.
 - [11] D. Wang, D. Zhang, and J. C. Chan, “Feature extraction of radial arterial pulse,” in *Proceedings of the 2014 International Conference on Medical Biometrics*, pp. 41–46, IEEE, Shenzhen, China, June 2014.
 - [12] X. Hu, H. Zhu, J. Xu, D. Xu, and J. Dong, “Wrist pulse signals analysis based on deep convolutional neural networks,” in *Proceedings of the 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1–7, IEEE, Honolulu, HI, USA, May 2014.
 - [13] C. Y. Chung, Y. W. Cheng, and C. H. Luo, “Neural network study for standardizing pulse-taking depth by the width of artery,” *Computers in Biology and Medicine*, vol. 57, pp. 26–31, 2015.
 - [14] R. Velik, “An objective review of the technological developments for radial pulse diagnosis in traditional Chinese medicine,” *European Journal of Integrative Medicine*, vol. 7, no. 4, pp. 321–331, 2015.
 - [15] J. Hou, H. Y. Su, B. Yan, H. Zheng, Z. L. Sun, and X. C. Cai, “Classification of tongue color based on CNN,” in *Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 725–729, IEEE, Beijing, China, March 2017.
 - [16] S. Li, B. Zhang, D. Jiang, Y. Wei, and N. Zhang, “Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae,” *BMC Bioinformatics*, vol. 11, no. S11, 2010.
 - [17] Y. Wu, M. Jiang, J. Lei, and H. Xu, “Named entity recognition in Chinese clinical text using deep neural network,” *Studies in Health Technology*, vol. 216, pp. 624–628, 2015.
 - [18] L. Yao, Y. Zhang, B. Wei et al., “Discovering treatment pattern in traditional Chinese medicine clinical cases by exploiting supervised topic model and domain knowledge,” *Journal of Biomedical Informatics*, vol. 58, pp. 260–267, 2015.
 - [19] L. Yao, Y. Zhang, B. Wei, Z. Li, and X. Huang, “Traditional Chinese medicine clinical records classification using knowledge-powered document embedding,” in *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1926–1928, IEEE, Shenzhen, China, December 2016.
 - [20] C. X. Hong, Z. Y. Feng, C. X. Rong, L. Tian, W. Y. Wei, and M. Li, “The ontology-based knowledge representation modeling of the traditional-Chinese-medicine symptom,” in *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1345–1349, IEEE, Kansas City, MO, USA, November 2017.
 - [21] X. Sun and H. Qian, “Chinese herbal medicine image recognition and retrieval by convolutional neural network,” *PLoS One*, vol. 11, no. 6, Article ID e0156327, 2016.
 - [22] Z. Chen, Y. Cao, S. He, and Y. Qiao, “Development of models for classification of action between heat-clearing herbs and blood-activating stasis-resolving herbs based on theory of traditional Chinese medicine,” *Chinese Medicine*, vol. 13, no. 1, pp. 1–11, 2018.
 - [23] S. Lukman, Y. He, and S.-C. Hui, “Computational methods for traditional Chinese medicine: a survey, comput,” *Computer Methods and Programs in Biomedicine*, vol. 88, no. 3, pp. 283–294, 2007.
 - [24] C. Cao, H. Wang, and Y. Sui, “Knowledge modeling and acquisition of traditional Chinese herbal drugs and formulae from text,” *Artificial Intelligence in Medicine*, vol. 32, no. 1, pp. 3–13, 2004.
 - [25] GB/T 20348-2006, *Basic Theory Nomenclature of Traditional Chinese*, Standardization Administration of the People’s Republic of China, Beijing, China, 2006, in Chinese.
 - [26] X. Zhu-Fan, “On standard nomenclature of basic Chinese medical terms (VA),” *Chinese Journal of Integrative Medicine*, vol. 9, no. 4, pp. 306–307, 2003.
 - [27] J. Sivic and A. Zisserman, “Efficient visual search of videos cast as text retrieval,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 591–606, 2009.
 - [28] K. Weinberger, A. Dasgupta, J. Attenberg, J. Langford, and A. Smola, “Feature hashing for large scale multitask learning,” 2009, <http://arxiv.org/abs/0902.2206>.
 - [29] X. Z. Rong Fang and X. Zhao, “Analysis of 245 cases of adverse reactions of Chinese patent medicine,” *Journal of Yangtze University (Natural Science Edition)*, vol. 36, pp. 105–107, 2014, in Chinese.
 - [30] X. G. Shan Cao, Y. Xia, H. Qu, P. Fan, and X. Yang, “Analysis of 266 cases of adverse reactions of Chinese patent medicine,” *Guangxi Medical*, vol. 38, pp. 1315–1317, 2016, in Chinese.
 - [31] H. X. Xiaomei Jiang and Q. Zhang, “Adverse drug reactions analysis of Chinese patent medicine in years of 2010~2012,” *Journal of Traditional Chinese Medicine Management*, vol. 11, pp. 1806–1807, 2014, in Chinese.
 - [32] Y. Y. Yingyan Yan and Y. Mao, “Analysis of adverse reactions of Chinese patent medicines in our hospital from 2011 to 2014,” *Chinese Journal of Rural Medicine and Pharmacy*, vol. 3, pp. 51–52, 2016, in Chinese.
 - [33] A. L. Lihua Pan, Y. Huang, W. Hong, and X. Gu, “Common adverse reactions of traditional Chinese medicine and its proprietary Chinese Medicine,” *Lishizhen Medicine and Materia Medica Research*, vol. 13, pp. 685–686, 2002, in Chinese.
 - [34] F. K. Ping Wang, J. Wei, and X. Cao, “Analysis of 213 cases of adverse reactions of proprietary Chinese medicines,” *Chinese Journal of Clinical Rational Drug Use*, vol. 9, pp. 84–85, 2016, in Chinese.