

Prediction of traditional Chinese medicine prescriptions based on multi-label resampling

Chinese
medicine
prescription
prediction

213

Xiaomei Jiang

*National Pilot School of Software, Yunnan University, Kunming, China and
Yunnan Key Laboratory of Software Engineering, Yunnan University,
Kunming, China*

Shuo Wang

School of Computer Science, University of Birmingham, Birmingham, UK

Wenjian Liu

School of Data Science, City University of Macau, Macau, China, and

Yun Yang

*National Pilot School of Software, Yunnan University, Kunming, China and
Yunnan Key Laboratory of Software Engineering, Yunnan University,
Kunming, China*

Received 18 April 2023
Revised 16 August 2023
Accepted 17 August 2023

Abstract

Purpose – Traditional Chinese medicine (TCM) prescriptions have always relied on the experience of TCM doctors, and machine learning (ML) provides a technical means for learning these experiences and intelligently assists in prescribing. However, in TCM prescription, there are the main (Jun) herb and the auxiliary (Chen, Zuo and Shi) herb collocations. In a prescription, the types of auxiliary herbs are often more than the main herb and the auxiliary herbs often appear in other prescriptions. This leads to different frequencies of different herbs in prescriptions, namely, imbalanced labels (herbs). As a result, the existing ML algorithms are biased, and it is difficult to predict the main herb with less frequency in the actual prediction and poor performance. In order to solve the impact of this problem, this paper proposes a framework for multi-label traditional Chinese medicine (ML-TCM) based on multi-label resampling.

Design/methodology/approach – In this work, a multi-label learning framework is proposed that adopts and compares the multi-label random resampling (MLROS), multi-label synthesized resampling (MLSMOTE) and multi-label synthesized resampling based on local label imbalance (MLSOL), three multi-label oversampling techniques to rebalance the TCM data.

Findings – The experimental results show that after resampling, the less frequent but important herbs can be predicted more accurately. The MLSOL method is shown to be the best with over 10% improvements on average because it balances the data by considering both features and labels when resampling.

Originality/value – The authors first systematically analyzed the label imbalance problem of different sampling methods in the field of TCM and provide a solution. And through the experimental results analysis, the authors proved the feasibility of this method, which can improve the performance by 10%–30% compared with the state-of-the-art methods.

Keywords TCM prescription, Multi-label classification, Label imbalance, Resampling

Paper type Research paper

© Xiaomei Jiang, Shuo Wang, Wenjian Liu and Yun Yang. Published in *Journal of Electronic Business & Digital Economics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This work is funded by the Yunnan Basic Research Program for Distinguished Young Youths Project (No. 202101AV070003); the Yunnan Basic Research Program for Key Project (No. 202201AS070131) and funded by the Youth Science Fund of NSFC No. 62206239.



1. Introduction

TCM is the most traditional medicine culture in China. It is a medical system with a unique theoretical style of diagnosis and treatment characteristics gradually formed in the long-term practice. TCM records the rich experience and theoretical knowledge of Chinese in fighting against diseases for thousands of years. It has two prominent advantages, in the overall concept and treatment based on syndrome differentiation. They enable the treatment of TCM to fundamentally cure disease, adjust the overall state and finally achieve balance in human bodies. However, the development of TCM also faces some challenges, such as the shortage of TCM doctors because of the long training period of TCM talents. In the past five years, the ML and deep learning (DL) models have been widely used in the TCM field.

A number of automated TCM approaches have been proposed and applied to assist doctors' diagnosis and medical treatment and alleviate the shortage issue of TCM doctors. TCM prescription is the first and foremost way to treat diseases. It does not only emphasize the role of certain herbs but also emphasizes the ingenious combination of multiple herbs so that symptoms can be alleviated and pathogens can be eradicated. Prescription regularity can play a great role in clinical practice, new herbs discovery and the inheritance of TCM. Understanding how doctors prescribe through AI learning models and exploring the regularity of these prescriptions can intelligently assist in prescribing. Some brilliant achievements have been made in the previous work, however, because they did not solve label imbalance, this approach did not perform well. Since most prescription prediction tasks are considered multi-label classification (MLC) tasks, we consider using a multi-label dataset (MLD) imbalance processing method to solve it. The resampling method is a data-level processing method that can be flexibly applied in different algorithms, and resampling includes oversampling and undersampling. In DL, more data is more conducive to the learning of the model. Therefore, in this paper, the multi-label oversampling method is used to process the data imbalance of the TCM dataset, and in the proposed MLC framework, called ML-TCM, exploring various resampling methods in the previous work. Thus, this paper provides for the first time the multi-label framework combining resampled data balance methods to explore the data label imbalance in this field and to experimentally verify its possibility. Through the analysis of experimental results and conclusions, we prove the feasibility of this idea and provide a new possibility to improve prediction accuracy in this field. Through this method, we can improve the performance by 10%–30% compared with the state-of-the-art methods.

The rest of this paper is organized as follows: [Section 2](#) describes related work. [Section 3](#) introduces the ML-TCM framework. [Section 4](#) introduces the TCM dataset. [Section 5](#) presents the results and analysis. [Section 6](#) draws conclusions and points out future work.

2. Related work

In ML, the TCM prescription prediction is a prediction task that aims to automatically generate a TCM prescription (i.e. Chinese herbs) based on text symptom descriptions as inputs. This task faces several challenges. Foremost, unlike the Western medicine system, TCM regards the human body as an organic whole system. A series of symptoms of patients are interdependent and interactive. Since different symptoms are related, it is inappropriate to treat a patient's several symptoms separately. Besides, the treatment process includes a large number of complex knowledge in the field of TCM, such as herbal compatibility. Therefore, it is difficult to describe the treatment process comprehensively and accurately. Last but not the least, the lack of digital TCM data and open medical

records makes the research difficult. Nevertheless, some achievements have been made. This task has been formulated as a text generation task or a recommended task. In text generation, some researchers use a topic model, automatically extracting potential theme structures containing symptom information and corresponding TCM information (Jiang, Zhou, Zhang, & Chen, 2012; Yao, Zhang, Wei, Zhang, & Jin, 2018; Wang, Zhang, Wang, & Chen, 2019). Others use sequence-to-sequence (seq2seq) generation models to complete the task of prescription prediction (TCMseq2seq) (Li, Yang, & Sun, 2018; Wang, Poon, & Poon, 2019). For example, the attention-herb (Liu *et al.*, 2019) model uses a long short-term memory network (LSTM) to encode and decode symptoms and herbs. On this basis, a knowledge graph(KG) is added (Li, Liu, Yang, Huang, & Lv, 2020) and the attention mechanism is considered (Liu, Luo, *et al.*, 2022), then a dual-branch guidance strategy combined with an attention mechanism that integrates the TCM background knowledge into a seq2seq structure to help generate prescriptions (Hou *et al.*, 2023). Different from ordinary text, the order of the herbal medicines in prescription has no effect. When making a prescription prediction, however, these models will focus on the order of herbal medicines without fully considering the diversity and complexity of the compatibility of herbal medicines in prescription. For recommendation, treated symptoms as users and herbs as recommended items (Li, Wang, & He, 2021; Jin, Zhang, He, Wang, & Wang, 2020, 2021; Dong *et al.*, 2021; Zhao *et al.*, 2022; Rong, Li, Sun, & Sun, 2022). Their approaches modeled the interaction between herbs and symptoms, and used bipartite graphs (to capture co-occurrence patterns between symptoms), the embedding of symptoms and inducing a group of symptoms into a whole symptom representation. Based on the idea of integrated learning, a multi-layer information fusion graph convolution approach (KDHR) generates symptom and herbs' feature representation with rich information and low noise (Yang, Rao, Yu, & Kang, 2022). A meta-path-guided graph attention network tried to provide interpretable herb recommendations (Jin, Ji, Shi, Wang, & Yang, 2023). There is a phenomenon of label imbalance in the TCM dataset. The basic principle of the composition of TCM prescriptions is "Jun-Chen-Zuo-Shi", which means that different herbs play different roles in prescribing (Yao *et al.*, 2018). Among them, the "Jun" herb plays a major therapeutic role, which can be regarded as the main herb, while the rest can be regarded as auxiliary herbs to assist and strengthen the effect of the main herb. In a prescription, there are often more auxiliary than the main one and the auxiliary often appears in other prescriptions. This leads to different frequencies of different herbs in the prescriptions, namely, imbalanced labels (herbs). This phenomenon manifested in the actual prescription prediction is auxiliary herbs have a high probability of appearing in the predicted prescription, and main herbs that appear less often have not been predicted. For ML models, this imbalance leads to the model being biased toward the label that appears more frequently (majority), resulting in bias and poor prediction performance. There are few existing methods to solve this problem, only tried reweighting. Works (Jin *et al.*, 2020, 2023) using the frequency of label (herb) occurrence as a weight to add to the mean squared error (MSE) loss function was proposed to overcome the label imbalance, but no major progress has been made.

MLC can accomplish the task that an instance can be associated with a set of labels simultaneously and is mostly used in text, emotion and scene classification. With the deepening of a combination of the medical field and AI, MLC is also more widely used in medicine, such as TCM diagnosis of Parkinson's disease (Peng, Fang, Wang, & Xie, 2015, 2017), ECG (Ge *et al.*, 2021), hypertension (Weng *et al.*, 2018) and AIDS (Zhang *et al.*, 2022). MLC is also used for disease prediction (Pham *et al.*, 2022), so much as be applied to clinical decision support systems (Khan & Shamsi, 2021). These successful application examples represent the TCM prescription prediction as MLC is quite feasible, and because each instance of the TCM prescription data itself has a label set (multiple labels), the reason is

not mutually exclusive between labels, is the MLD, so this paper also adopts the MLC method for the prediction of TCM prescription label combination.

One of the challenges in MLC is the imbalanced distribution of MLD, in which labels are unevenly distributed in label space. The label distribution in MLD is normally described by label cardinality (LCard) and the imbalance degree of MLD is measured by imbalance ratio (IR), both are based on the labels' frequency. LCard is the ratio of label frequency to total instances. MaxIR, MinIR and MeanIR are the maximum, minimum and average values of IR per label (IRLbl) (Tarekegn, Giacobini, & Michalak, 2021) for all labels, which can reflect the distribution of different labels in the entire label set. The calculation of IRLbl in Eq (1). Let M be MLD, m be the numbers of M , in which there is a set of labels L and $\lambda, \lambda' \in L$, Y is the label set of the i th instance. For label λ , $h(\lambda, Y_i)$ is the frequency in the labels set of the i th instance in Eq. (2).

$$\text{IRLbl}(\lambda) = \frac{\max_{\lambda' \in L} (\sum_{i=1}^m h(\lambda', Y_i))}{\sum_{i=1}^m h(\lambda, Y_i)} \quad (1)$$

$$h(\lambda, Y_i) = \begin{cases} 1 & \lambda \in Y_i \\ 0 & \lambda \notin Y_i \end{cases} \quad (2)$$

After calculating m instances, we can calculate the frequency of all labels in L in M , and we can get a maximum value among these frequencies. IRLbl is the ratio between max frequency and labels λ 's frequency, where IRLbl is 1 for the most frequent label and a larger value for the rest of the labels. The higher value of IRLbl, the higher the imbalance level of the related label. Therefore, based on the imbalance of TCM data, the solution of MLD imbalance also provides us with corresponding ideas to solve this problem.

There are many methods for dealing with data imbalance in the field of text and images, but their processing methods are not suitable for this task (Yang & Jiang, 2015, 2018; Yang, Hu, Zhang, & Wang, 2021). Resampling and reweighting are two types of universal solutions to the imbalance of MLD. Resampling is a data-level solution, including oversampling and undersampling, while reweighting is to deal with an imbalance on the algorithm level. Reweighting methods rebalance labels by adjusting the loss values of different labels during training, such as CPNL (Wu, Tian, & Liu, 2018), UCML (Dou, Song, Wei, & Zhang, 2022) and SMGCN (Jin *et al.*, 2020). Among them, the SMGCN method is a reweighting method to balance the imbalance of TCM labels on the classifier, but very little effect has been achieved. The multi-label resampling method is a more flexible method, and the balance effect on the MLD is obvious. Currently, common multi-label resampling methods include multi-label undersampling deletes majority labels instances to reduce the imbalance of data sets, such as MLRUS (Charte, Rivera, del Jesus, & Herrera, 2015; Charte, Rivera, del Jesus, & Herrera, 2015) and oversampling usually copies or synthesizes new instances with minority labels to achieve the effect of balancing a data labels distribution, such as MLROS (Charte *et al.*, 2015a), MLSMOTE (Charte *et al.*, 2015b) and MLSOL (Liu, Blekas, & Tsoumakas, 2022). Given the reasons in Section 1, which are the three main methods adopted in this paper, the differences and application effects are described in detail in Section 3.2, A part and Section 4, respectively.

3. Proposed framework

In this section, we define the problem of herb prediction in Section 3.1, and then introduce the learning framework multi-label-traditional Chinese medicine (ML-TCM). Table 1 summarizes the notations used in this section.

Table 1.
Notations

Notation	Description
D	Prescription data set
(s_set, h_set)	A instance in D
S	Full symptom set
H	Full herb set
s_set	A symptom set as input
h_set	A true prescription set as output
ph_set	The predicted herbal set
Z_S	Characteristic representation of S
Z_H	Characteristic representation of H
$Z(s_set)$	Representation of symptoms set s_set
$R(h_set)$	The predicted probabilistic vector of h_set
θ	A set of trainable parameters of function $g(\cdot)$

3.1 Problem definition

In a prescription data set D that contains a symptom set S and an herb set H, an instance is expressed with s_set (a subset of S) and h_set (a subset of H). The length of the symptom sets and herb sets is not fixed. The goal is to learn a prediction function $g(x)$, by entering the symptom set (s_set) in the prescription, learning to train the herbal label set (h_set) in the existing prescription, to predict the set of herbal labels corresponding to the new symptom set. That is $R(h_set) = g(s_set, H|\theta)$. Where $R(h_set)$ is a probability vector, in which each number is the probability of prescribing the corresponding herb, the function $g(x)$ parameters θ changed through training.

3.2 ML-TCM framework

ML-TCM consists of three key modules as shown in Figure 1, including data imbalance processing, GNN learning of KDHR (the most comprehensive model so far) and prescription prediction.

3.2.1 Label imbalance processing. After the previous analysis, multi-label resampling methods were first used to balance the labels of the TCM data. The three most commonly used oversampling methods for tackling label imbalance in MLC problems are applied and compared, which are the random resampling MLROS, random synthesized resampling method MLSMOTE and synthesized resampling MLSOL based on local label imbalance. The same is that they used to balance the data set in the data preprocessing stage by increasing the number of instances of minority classes or decreasing majority classes. The difference between the three methods is that MLROS copies instances containing minority labels randomly, which is simple and basic. MLSMOTE uses k-nearest neighbor to generate new instances with minority labels. MLSOL, which was recently proposed based on MLSMOTE, puts more emphasis on locality, that is, the balance of labels in k instances similar to an instance.

3.2.2 Graph neural network learning. This part is a classification algorithm, referring to KDHR, which proves the superiority of the classification algorithm through their research. Of course, this part can also be replaced with other classification algorithms as in the sampling method in part A. GNN can learn dependencies of instances, labels and between labels and instances. As one specific type of GNN, GCN uses convolution operation and can be applied to graph embedding (GE). They not only utilize the structure information of the graphs but also consider the characteristic information of nodes in the graph. Therefore, KDHR proposed to learn the co-occurrence relationship between herb-herb, herb-symptom and symptom-symptom in the dataset by creating and capturing the information of bipartite graphs and using the herbal KG to make more detailed characterization of the information of herbal labels. The fusion of this information can better obtain the representation data for each symptom, and each herb, as the basis for the learning of the features and the classification of

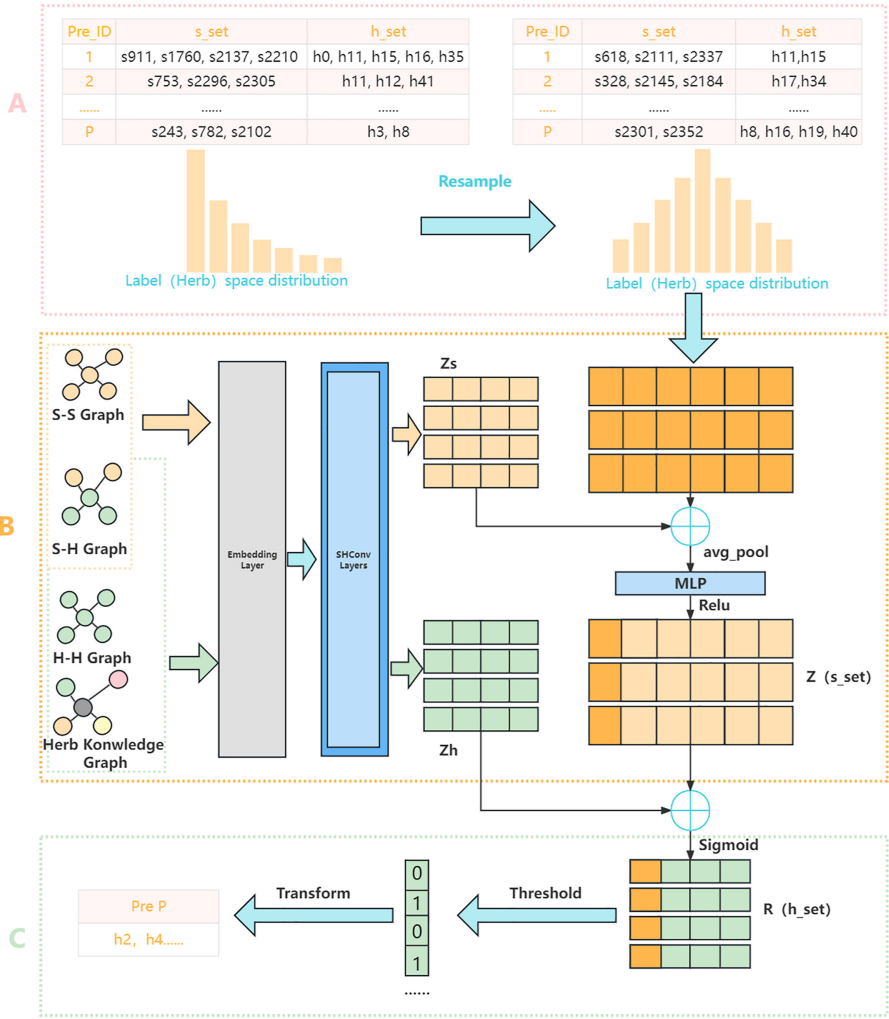


Figure 1. ML-TCM framework: A describes the imbalance processing of data, B shows the model built by previous work and C is the final multi-label prescription prediction

the labels. This part describes the previous work based on the GNN algorithms, KDHR proposed a new convolutional layer, which is called “SHConv” in our framework, as shown in Eq (3) and Eq (4). Z_S and Z_H are the symptom and herb characteristics, respectively.

$$Z_S = SHConv(S - S Graph, S - H Graph) \quad (3)$$

$$Z_H = SHConv(H - H Graph, S - H Graph, e^{kg}) \quad (4)$$

Construction graphs using H-H graph as the example is shown in Eq (5), where T is set as a threshold such as 5, used to represent the strength of the relationship between two herbs. With the common frequency in a prescription to measure if the frequency is greater than the threshold T, means the relationship is strong, there are edges between two herbs, in the storage matrix value of 1, otherwise the value is 0.

$$H - H \text{ Graph} = \begin{cases} 1 & \text{if herb pair frequency greater than } T \\ 0 & \text{if otherwise} \end{cases} \quad (5)$$

The remaining S-S Graph and S-H Graph are created as in H-H Graph, and the specific information of the graphs is shown in Table 3. The herbal KG is based on the TCM theory containing five attributes: category, five elements, meridian, smell and nature, forming 107 entities, 5 relationships and 322 triples. And obtain e^{hg} by an embedding method such as one-hot encoding.

3.2.3 Prescription prediction. After obtaining the characteristic representation Z_S of all symptoms, we can represent any instance's symptom set as Z_{s_set} through Eq 6.

$$Z_{s_set} = \text{ReLU}(\text{MLP}(\text{GAP}(Dv * Z_S))) \quad (6)$$

Dv represents the one-hot vector of symptom set in prescription mutual with Z_S in Eq 3. The global average pool (GAP) layer mapping multidimensional symptom representation into a low-dimensional space to improve the generalization ability of the model, and use the linear activation function Relu to correct the previous result.

As a prescription prediction task, the final output of the framework is the herb set with probabilities $R(h_set)$, resulting by Z_{s_set} (in Eq 6) and Z_H (in Eq 4) interactions. We use the sigmoid function to normalize the probability output of each label, as shown in Eq 7.

$$R(h_set) = \text{Sigmoid}(Z_{s_set} \cdot Z_H) \quad (7)$$

4. Prescription data and label imbalance

In this section, we introduce the multi-label prescription dataset and graphs. By the way, due to the minimal performance improvement of the KG, our method does not consider it. Then, explain the label imbalance of this dataset.

4.1 Prescription datasets

TCM prescriptions are the main means of guiding clinical disease prevention and treatment. So far, a large number of TCM prescriptions have been collected, which not only provides a reference for clinicians but also brings opportunities for using computational models to discover prescription patterns. To achieve this, the dataset we use in this work contains herbs and symptoms. The source of our raw data is consistent with KDHR. After data extraction and filtering, an example is shown in Table 2. The right-hand side of the table is the corresponding herb set (model output) for treating the left-hand side symptoms (model inputs).

If the original data is randomly divided, labels that appear less frequently in the herb set will not be guaranteed to be divided into the training set to be sampled. Therefore, we removed labels that appear less than ten times, resulting in sub-dataset 1 containing 389 symptoms and 330 herbs, and also divided the dataset according to the labels. Due to a large number of labels and the heavy workload of detailed analysis, we filtered the sub-dataset of 43 commonly used herbs (including 380 symptoms). The detailed information on the two sub-

Symptoms	Prescription
Alternating hot and cold, spit, unconscious, phlegm, cannot sleep, sweat, cough, asthma	Arranthus, Ejiao, Pinellia ternata, Alum, croton, apricot

Table 2.
Sample of
prescription data

datasets, as well as the graph and dataset partitioning created based on the dataset, are shown in [Table 3](#).

4.2 Label imbalance

This section explores the explanation of label imbalance in the TCM prescription data. As shown in [Figure 2](#), the frequency percentage of herbs in dataset 1. There are significant differences in the proportion of different labels that can be seen, with a few accounting for 6% and a few approaching 0%, indicating a significant degree of imbalance in this dataset.

[Figure 3](#) depicts the label imbalance in dataset 2, and [Figure 3 \(a\)](#) shows the frequency percentage of occurrence of each herb, based on the frequency and proportion of each label appearing in it, can be considered herb ID 8 (h8), 11 (h11), 12 (h12), 20 (h20), 34 (h34) as the majority (labels with high frequency) in the labels, while 5 (h5), 18 (h18), 28 (h28), 36 (h36) and 42 (h42) can be considered as a minority (labels with a low frequency). [Figure 3 \(b\)](#) shows the changes in MaxIR/MeanIR/MinIR (values = 1) of dataset 2 after imbalanced processing. The difference between the MaxIR and MinIR before sampling is significant, after sampling, the gap between them is reduced through different sampling methods, most significant on MLROS. Those indicate the imbalance phenomenon of the data, and resampling can ameliorate the degree of imbalance in the dataset.

5. Experiment

5.1 Evaluation metrics and benchmark

In order to keep consistent with the previous work, we considered the metrics shown in [Eq \(8-10\)](#) to evaluate the performance of the model, and K = 5,10,20, also.

Table 3.
Detailed information of datasets

Dataset	Symptoms (Nums)	Herbs (Nums)	LCard (Average)	S-S Graph	H-H (edge nums)	S-H	Train	Dev	Test
1	389	330	0.378	10,786	1,624	51,450	2,994	999	999
2	380	43	0.023	2,232	1,178	13,636	2,706	902	902

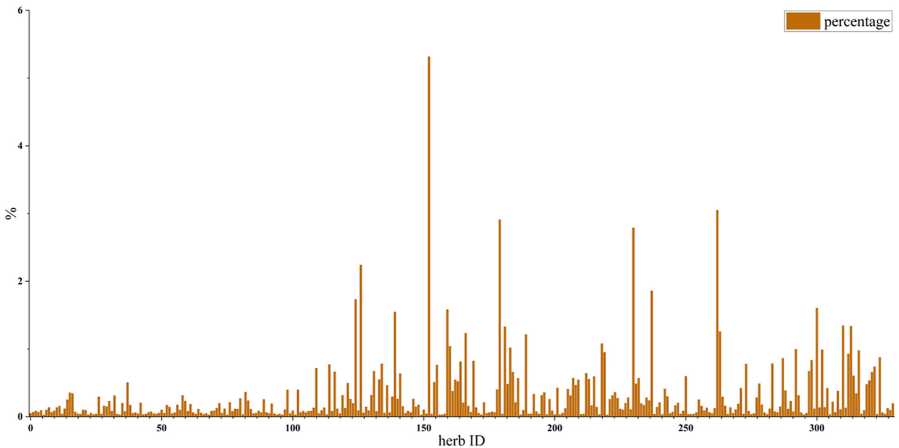


Figure 2.
Frequency percentage of occurrence of each herbs (labels) in TCM dataset 1

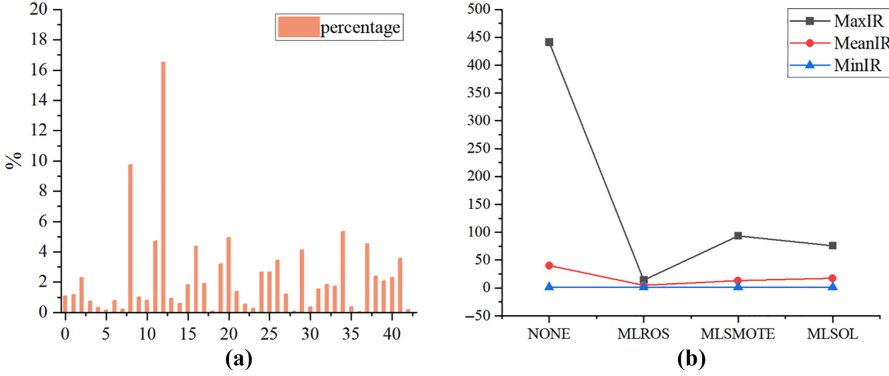


Figure 3.
Depicts the label
imbalance of dataset 2

$$Precision@K = \frac{1}{n} \sum_{i=1}^n \frac{|h_set \cap K_set|}{|K_set|} \quad (8)$$

$$Recall@K = \frac{1}{n} \sum_{i=1}^n \frac{|h_set \cap K_set|}{|h_set|} \quad (9)$$

$$F1 - Score@K = \frac{2 \cdot Precision@K \cdot Recall@K}{Precision@K + Recall@K} \quad (10)$$

$$Accuracy = \frac{n_{correct}}{n_{total}} \quad (11)$$

Here, h_set is the real herb set in the prescription, ph_set is the predicted herb set by the model, n is the number of prescriptions (i.e. the number of instances in the dataset) and is the number of elements in a set. The metrics Precision@K (Eq 8), Recall@K (Eq 9) and F1-Score@K (Eq 10), and only evaluate the first K of the predicted prescription. For instance, an example in prescriptions, $h_set = (h_1, h_2, h_3, h_4, h_5, h_6)$, $ph_set = (h_1, h_2, h_3, h_6, h_7, h_8, h_4, h_{10})$. When $K = 5$, the five predicted herbs with the highest probability of model output are $K_set = (h_1, h_2, h_3, h_7, h_6)$, $Precision@5 = 4/5$, $Recall@K = 4/6$. When considering the predictive performance of a single label category, the accuracy (Eq 11) metric used, is the ratio of correctly predicted instances to the total number of instances.

We compare our method with the previous algorithm, detailed information in section 1.

- (1) MLKNN: MLKNN (Tarekegn *et al.*, 2021) is the basic solution for MLC, as a baseline. In order to verify whether is effective the processing of data imbalance, also.
- (2) KDHR: KDHR (Yang *et al.*, 2022) is the most comprehensive herbal medicines recommendation based on KG and GCN, and has achieved well.
- (3) MGAT: MGAT (Jin *et al.*, 2023) is the latest method and tries to interpret the TCM prescription prediction through the meta-path KG.
- (4) ML-TCM: On the basis of KDHR, the method of processing data imbalance is added.

5.2 Experimental setup

We use PyTorch to implement a DL model and experiments on the Intel (R) Core (TM) i7-10750H CPU @ 2.60GHz, 32GB of memory. In the stage of data imbalance processing, three

sampling methods MLROS, MLSMOTE and MLSOL mentioned in 3.2 part A. Their parameters settings are MLROS (sample_size = 100), MLSMOTE ($k = 10$, $n = 2500$) and MLSOL (max replication ratio 25%, sample_size = 200). Training and learning stage, valid prescription data was randomly divided 6:2:2, detailed data is shown in Table 2. Based on the preliminary, we set parameters lr = 0.0002 and batch size = 512 (dataset1), batch size = 32 (dataset2) and 200 epochs, 30 times.

5.3 Results and analysis

In this part, firstly, we present our experimental results. In Table 4, the underlined number indicates the highest result and the “improvement” shows the percentage of improvement from MLSOL compared to the one without resampling. The results show that all the sampling methods improve all the performance metrics compared to the model without sampling. In particular, the MLSOL sampling method has the most significant improvement, with an average improvement of 10.3% across the metrics. The possible reason is that MLROS randomly reproduces instances, including the majority labels if the output set contains both minority and majority labels so that the minority-class instances are still likely to be under-trained. MLSMOTE considers the global label imbalance, the synthesis process is to assign the label vector to the instance, but during which noise may be introduced. MLSOL is based on local imbalance, where synthetic new examples with both features and labels, so as to achieve the best sampling effect, and the best model performance. Comparing the two datasets, the improvement of the metrics is very similar, while MLSMOTE and MLROS are different. The improvement effect of MLROS on dataset 2 is higher than MLSMOTE. On dataset 1, MLSMOTE performs better. It implies that the choice of resampling could depend on the number of labels.

Here, we compare the performance of the baseline and ML-TCM on dataset 2. We use statistical analysis software SPSS for experimental results and use the Waller-Duncan method to carry out single-factor ANOVA to obtain the significance test results and a significance level of 0.05, shown in Table 5. The mean value and significance level are ranked from a-d, where a, is the best performance and d is the lowest. A different letter means there are significant differences between models, while the same has no difference. We can see that has significant differences and there is a 10%–30% improvement in all metrics. We also studied a variety of classifiers, including KDHR and MLKNN, TCMseq2seq. Although the resampling methods of the optimal combination of different classifiers are different, all the classifiers showed a significant improvement after adding the resampling method to balance the data.

Next, we further look into predictive accuracy over certain labels, including the top five majority herb labels and the top five minority herb labels. Table 6 illustrated their accuracy and compares, the underline is the best. The accuracy of each label in predicted prescriptions is calculated based on the predicted top ten herbs of dataset 2. From the table, the label (herb)’s accuracy ascends after sampling, in the majority, the prediction accuracy increases maximum with MLROS while the minority is MLSOL. For the majority, the prediction accuracy without sampling is not much different from that before sampling. However, we can see that the accuracy of the minority labels is improved significantly by resampling. The accuracy of the minority is significantly improved on h36, especially for the model using the MLSOL sampling method. After sampling, the accuracy of the instances in the majority labels h8 and h34 decreased in the MLSOL method, due to the performance trade-off between minority and majority labels.

In order to further explore whether resampling has any actual effect on prescription prediction, we take a group of symptoms as instances to analyze the compatibility rules of

Data set	Model	Precision			Recall			F1-score		
		P@5	P@10	P@20	R@5	R@10	R@20	F@5	F@10	F@20
1	NORESAM	0.2180	0.1650	0.1150	0.3900	0.5768	0.7844	0.2796	0.2566	0.2005
	MLROS	0.2496	0.1865	0.1294	0.3736	0.5553	0.7745	00.2993	0.2792	0.2218
	MLSMOTE	0.2801	0.2036	0.1374	0.4286	0.6117	0.8218	0.3388	0.3055	0.2355
	MLSOL	0.3556	0.2479	0.1519	0.5343	0.7354	0.9052	0.4270	0.3708	0.2602
	Improvement	13.8%	8.3%	3.7%	14.4%	15.9%	12.1%	14.8%	11.4%	6.0%
2	NORESAM	0.2171	0.1649	0.1150	0.3875	0.5744	0.7854	0.2783	0.2562	0.2006
	MLROS	0.2713	0.1981	0.1332	0.4050	0.5875	0.7948	0.3250	0.2963	0.2282
	MLSMOTE	0.2766	0.2019	0.1365	0.4261	0.6096	0.8174	0.3354	0.3033	0.2340
	MLSOL	0.3590	0.2501	0.1525	0.5367	0.7391	0.9090	0.4302	0.3737	0.2612
	Improvement	14.1%	8.5%	3.7%	14.9%	16.5%	12.4%	15.1%	11.8%	6.0%

Table 4.
Changes in metrics for
different sampling
methods

herbs predicted by different models in Table 7. The underline means that the herbs predicted by the model are herbs in the real prescription.

From the instances, we can easily see that the hit rate of prediction can be improved by sampling. In addition, we pay more attention to this instance, because the real prescription contains a minority herb “Astragalus membranaceus” (h14). From the prescriptions predicted by different methods, the model did not predict the minority without sampling, but after sampling, the minority was predicted. In addition, MLSMOTE’s incorrectly predicted herbs “Chinese herbaceous peony”(h37), MLSOL’s “Chaihu”(h16) and Banxia(h20) belong to

Table 5.
Metrics’ average and
significance test results
of the four models

	N	F@5	F@10	F@20
MLKNN	30	0.16 ± 0.00 c	0.06 ± 0.00 c	0.09 ± 0.00 d
KDHR		0.28 ± 0.01 b	0.26 ± 0.00 b	0.20 ± 0.00 c
MGAT		0.28 ± 0.05 b	0.29 ± 0.13 b	0.19 ± 0.01 b
ML-TCM (MLSOL)		0.43 ± 0.01 a	0.37 ± 0.01 a	0.26 ± 0.00 a

Table 6.
Label-based accuracy
over the top five
majority and minority
labels

Methods	Accuracy									
	Majority					Minority				
	h8	h11	h12	h20	Labels	h34	h5	h18	h28	h36
NORESAM	0.462	0.642	0.472	0.477	0.615	0.935	0.835	0.817	0.767	0.911
MLROS	<u>0.486</u>	<u>0.764</u>	<u>0.558</u>	<u>0.602</u>	<u>0.682</u>	0.953	0.945	0.960	0.979	0.884
MLSMOTE	<u>0.455</u>	<u>0.714</u>	<u>0.532</u>	<u>0.556</u>	<u>0.527</u>	0.961	0.937	0.975	0.916	0.949
MLSOL	0.433	0.658	0.545	0.566	0.531	<u>0.983</u>	<u>0.977</u>	<u>0.985</u>	<u>0.988</u>	<u>0.977</u>

Table 7.
Comparison of
predicted and real
prescriptions using
different methods

Symptom	True prescription	Methods	Predict prescription
Anorexia	Rehmannia glutinosa	NORESAM	Chinese herbaceous peony
Whitening of lips	Schisandra chinensis		Schisandra chinensis
Numbness of limbs	Poria cocos		Poria cocos, Chaihu, Adzuki Beans
	Jujube		Licorice, Rehmannia glutinosa
	Licorice		Cinnamon, Scape seed, Banxia
	Astragalus membranaceus	MLROS	Poria cocos, Schisandra chinensis
	Cinnamon		Licorice, Rehmannia glutinosa
	Ginger		Astragalus membranaceus, Cinnamon
			Anemarrhena asphodeloides, Ginger
			Jujube, Alisma orientalis
			Poria cocos, Rehmannia glutinosa
			Licorice, Jujube
		Chinese herbaceous peony	
		Schisandra chinensis	
		Astragalus membranaceus	
		Alisma orientalis, Ginger, Chaihu	
		MLSOL	Licorice, Schisandra chinensis
			Poriacocos, Rehmannia glutinosa
			Astragalus membranaceus, Cinnamon
			Jujube, Ginger
			Chaihu, Baikal Skullcap

the majority, while the minority predicted takes precedence over the majority. This indicates that through sampling, the visibility of a few categories in the model is improved and can be appropriately predicted. On the other hand, according to the TCM theory, symptoms of the patient including loss of appetite, pale lips and numbness in limbs, are caused by insufficient qi and blood and weak spleen and stomach function. The effect of “Astragalus membranaceus” is to strengthen the spleen and replenish qi, mainly for the treatment of qi and blood loss and spleen weakness. It is a key therapeutic herb for this prescription, and its prediction plays a crucial role in the overall efficacy of the prescription. This indicates that the model with resampling is able to find an herb that belongs to a minority but is the main herb.

6. Conclusion and future work

This paper provides a multi-label prediction framework ML-TCM, based on resampling to learn the rules of TCM prescriptions and to mine the knowledge between symptoms and diseases. We also dealt with the imbalance distribution of labels in the data of herbs by using multi-label resampling techniques, which is the very first work in this area. By resampling on the existing GCN model, we effectively improve the performance of herbs prediction, and we have conducted a detailed analysis from the perspectives of quantity and quality, both theoretically and practically. It shows by balancing the label distribution of the TCM dataset, it is beneficial to learn the prescription rules more accurately and achieve good prescription prediction results.

In the future, according to our experimental results, when the current resampling method samples a large number of labels and a small number of labels, the performance of sampling a large number of labels is affected. This problem also needs to be solved urgently. In addition, we will continue to explore new ways of balancing the distribution of herb datasets with a large number of labels in order to make further contributions to research in this field.

References

- Charte, F., Rivera, A. J., del Jesus, M. J., & Herrera, F. (2015a). Addressing imbalance in multilabel classification: Measures and random resampling algorithms. *Neurocomputing*, 163, 3–16.
- Charte, F., Rivera, A. J., del Jesus, M. J., & Herrera, F. (2015b). Mlsmote: Approaching imbalanced multilabel learning through synthetic instance generation. *Knowledge-Based Systems*, 89, 385–397.
- Dong, X., Zheng, Y., Shu, Z., Chang, K., Yan, D., Xia, J., . . . Zhu, X. (2021). *TCMPR: TCM prescription recommendation based on subnetwork term mapping and deep learning*. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 3776–3783). IEEE.
- Dou, J., Song, Y., Wei, G., & Zhang, Y. (2022). Fuzzy information decomposition incorporated and weighted relief-f feature selection: When imbalanced data meet incompleteness. *Information Sciences*, 584, 417–432.
- Ge, Z., Jiang, X., Tong, Z., Feng, P., Zhou, B., Xu, M., . . . Pang, Y. (2021). Multi-label correlation guided feature fusion network for abnormal ecg diagnosis. *Knowledge-Based Systems*, 233, 107508.
- Hou, J., Song, P., Zhao, Z., Qiang, Y., Zhao, J., & Yang, Q. (2023). “Tcm prescription generation via knowledge source guidance network combined with herbal candidate mechanism”, *Computational and Mathematical Methods in Medicine*, Vol. 2023, 3301605.
- Jiang, Z., Zhou, X., Zhang, X., & Chen, S. (2012). Using link topic model to analyze traditional Chinese medicine clinical symptom-herb regularities. In 2012 IEEE 14th international conference on e-health networking, applications and services (Healthcom) (pp. 15–18). IEEE.
- Jin, Y., Zhang, W., He, X., Wang, X., & Wang, X. (2020). Syndrome-aware herb recommendation with multi-graph convolution network. In 2020 IEEE 36th International Conference on Data Engineering (ICDE) (pp. 145–156). IEEE.

- Jin, Y., Ji, W., Zhang, W., He, X., Wang, X., & Wang, X. (2021). "A kg-enhanced multi-graph neural network for attentive herb recommendation", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 19 No. 5, pp. 2560-2571.
- Jin, Y., Ji, W., Shi, Y., Wang, X., & Yang, X. (2023). Meta-path guided graph attention network for explainable herb recommendation. *Health Information Science and Systems*, 11(1), 5.
- Khan, S., & Shamsi, J. A. (2021). Health quest: A generalized clinical decision support system with multi-label classification. *Journal of King Saud University-Computer and Information Sciences*, 33(1), 45–53.
- Li, W., Yang, Z., Sun, X., (2018). Exploration on generating traditional Chinese medicine prescription from symptoms with an end-to-end method. *arXiv preprint arXiv:1801.09030*.
- Li, C., Liu, D., Yang, K., Huang, X., & Lv, J. (2020). Herb-know: Knowledge enhanced prescription generation for traditional Chinese medicine. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1560–1567). IEEE.
- Li, S., Wang, W., & He, J. (2021). Kgap: Knowledge-aware neural group representation learning for attentive prescription generation of traditional Chinese medicine. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 450–455). IEEE.
- Liu, B., Blekas, K., & Tsoumakas, G. (2022). Multi-label sampling based on local label imbalance. *Pattern Recognition*, 122, 108294.
- Liu, Z., Luo, C., Fu, D., Gui, J., Zheng, Z., Qi, L., & Guo, H. (2022). A novel transfer learning model for traditional herbal medicine prescription generation from unstructured resources and knowledge. *Artificial Intelligence in Medicine*, 124, 102232.
- Liu, Z., Zheng, Z., Guo, X., Qi, L., Gui, J., Fu, D., . . . Jin, L. (2019). Attentiveherb: A novel method for traditional medicine prescription generation. *IEEE Access*, 7, 139069–139085.
- Peng, Y., Fang, M., Wang, C., & Xie, J. (2015). Entropy chain multi-label classifiers for traditional medicine diagnosing Parkinson's disease. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 856–862). IEEE.
- Peng, Y., Tang, C., Chen, G., Xie, J., & Wang, C. (2017). Multi-label learning by exploiting label correlations for tcm diagnosing Parkinson's disease. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 590–594). IEEE.
- Pham, T., Tao, X., Zhang, J., Yong, J., Li, Y., & Xie, H. (2022). Graph-based multi-label disease prediction model learning from medical data and domain knowledge. *Knowledge-Based Systems*, 235, 107662.
- Rong, C., Li, X., Sun, X., & Sun, H. (2022). Chinese medicine prescription recommendation using generative adversarial network. *IEEE Access*, 10, 12219–12228.
- Tarekegn, A. N., Giacobini, M., & Michalak, K. (2021). A review of methods for imbalanced multi-label classification. *Pattern Recognition*, 118, 107965.
- Wang, X., Zhang, Y., Wang, X., & Chen, J. (2019). A knowledge graph enhanced topic modeling approach for herb recommendation. In *International Conference on Database Systems for Advanced Applications* (pp. 709–724). Springer.
- Wang, Z., Poon, J., & Poon, S. (2019). Tcm translator: A sequence generation approach for prescribing herbal medicines. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2474–2480). IEEE.
- Weng, H., Liu, Z., Maxwell, A., Li, X., Zhang, C., Peng, E., . . . Ou, A. (2018). Multi-label symptom analysis and modeling of tcm diagnosis of hypertension. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 1922–1929). IEEE.
- Wu, G., Tian, Y., & Liu, D. (2018). Cost-sensitive multi-label learning with positive and negative label pairwise correlations. *Neural Networks*, 108, 411–423.
- Yang, Y., & Jiang, J. (2015). Hybrid sampling-based clustering ensemble with global and local constitutions. *IEEE Transactions on Neural Networks and Learning Systems*, 27(5), 952–965.

-
- Yang, Y., & Jiang, J. (2018). Adaptive bi-weighting toward automatic initialization and model selection for hmm-based hybrid meta-clustering ensembles. *IEEE Transactions on Cybernetics*, 49(5), 1657–1668.
- Yang, Y., Hu, Y., Zhang, X., & Wang, S. (2021). Two-stage selective ensemble of cnn via deep tree training for medical image classification. *IEEE Transactions on Cybernetics*, 52(9), 9194–9207.
- Yang, Y., Rao, Y., Yu, M., & Kang, Y. (2022). Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neural Networks*, 146, 1–10.
- Yao, L., Zhang, Y., Wei, B., Zhang, W., & Jin, Z. (2018). A topic modeling approach for traditional Chinese medicine prescriptions. *IEEE Transactions on Knowledge and Data Engineering*, 30(6), 1007–1021.
- Zhang, J., Zhao, Q., Adeli, E., Pfefferbaum, A., Sullivan, E. V., Paul, R., . . . Pohl, K. M. (2022). Multi-label, multi-domain learning identifies compounding effects of hiv and cognitive impairment. *Medical Image Analysis*, 75, 102246.
- Zhao, W., Lu, W., Li, Z., Zhou, C., Fan, H., Yang, Z., Lin, X., . . . Li, C. (2022). Tcm herbal prescription recommendation model based on multi-graph convolutional network. *Journal of Ethnopharmacology*, 297, 115109.

About the authors

Xiaomei Jiang, is studying master's degree at the School of Software, Yunnan University, and the research direction is software engineering and smart medical care.

Shuo Wang (corresponding author), is an associate professor at the School of Computer Science, the University of Birmingham, UK. Her research interests include data stream classification, class imbalance learning and ensemble learning approaches in machine learning, and their applications in social media analysis, software engineering and fault detection, detailed in <https://phd-shuowang.weebly.com>. Her work has been published in internationally renowned journals and conferences, such as IEEE Transactions on Knowledge and Data Engineering and International Joint Conference on Artificial Intelligence (IJCAI). Shuo Wang is the corresponding author and can be contacted at: s.wang.2@bham.ac.uk

Wenjian Liu, Ph.D. of Communication and Information System, South China University of Technology, is now teaching at the School of Data Science, City University of Macau. His main research interests are cloud computing and big data analysis, smart city, smart medical care, detailed in <https://fds.cityu.edu.mo/members/143>.

Yun Yang, is the National Pilot School of Software, Yunnan University, Kunming, China, as Full Professor of Machine learning, Director of Yunnan Education Department Key Laboratory of Data Science and Intelligent Computing and Director of Kunming Key Laboratory of Data Science and Intelligent Computing, He serves as Associate Editor in *Journal of Yunnan University* (Natural Sciences Edition). His research interests include publishing over 100 papers on artificial intelligence, machine learning, data mining, pattern recognition, big data processing and analysis and so on, detailed in <http://www.sei.ynu.edu.cn/info/1023/1099.htm>.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com