基于化学元素的部分中药药性量化方法的比较研究

徐钦涌[1, 2]，黄志帮[3]，姚思梦[4]，陈远方[5]，宁小英[2]，侯政昆[6*]，陈新林[7]

1. 东莞市滨海湾中心医院 中医科，广东 东莞 523900
2. 广州中医药大学第一临床医学院，广东 广州 510405
3. 汕头大学医学院第一附属医院揭阳浩泽医院，广东 揭阳 522021
4. 佛山市中医院，广东 佛山 528099
5. 南方医科大学第三附属医院，广东 广州 510630
6. 广州中医药大学第一附属医院 脾胃病科，广东 广州 510405
7. 广州中医药大学基础医学院，广东 广州 510006

摘 要：目的 基于中药的化学元素含量，通过不同分析方法获得中药药性的分类准确率，并对不同方法的分类准确率进行比较。**方法** 从《中药理论量化与应用研究》中获得目标中药，采用 Excel 对目标中药的化学元素数据进行提取，基于 IBM SPSS Statistics 26 软件对单个中药药性和化学元素进行 2 个独立样本非参数检验，将具有统计学意义的关联元素作为自变量，通过二元 Logistic 回归分析、决策树算法、人工神经网络等方法对因变量（药性）进行分类预测。运用此类算法得到中药药性的相关变量、分类准确率及模型函数系数，并比较不同方法的分类效果。**结果** 建立了含有 105 味中药、42 种化学元素的初步元素数据库，对中药药性进行统计，获得四气、五味、归经的药性变量。通过非参数检验得到药性的相关因素，寒性的相关因素有 Be、Sr、Ca、La；苦味的相关因素是 Mn、Ni、K、Ca、V、Si、Co、Zn；脾经的相关因素有 Ni、Bi、Co、Be、Eu、Ce、Nd、V、Pr、Sm、La、Dy。几种算法对寒性、苦味、脾经的分类预测准确率：二元 Logistic 回归分析分别是 87.6%、91.4%、81.4%；决策树模型训练集分别为 77.8%、87.7%、78.1%，检验集分别为 69.7%、65.0%、62.5%；人工神经网络模型训练集分别为 74.1%、73.7%、74.0%，检验集分别为 54.5%、72.4%、67.9%。**结论** 基于单因素分析获得药性的相关因素，通过二元 Logistic 回归、决策树、人工神经网络分析，揭示了中药药性与化学元素间存在一定联系。从分类准确率来看，决策树与神经网络训练集的准确率均高于检验集。决策树训练集、检验集平均分类准确率均高于神经网络。二元 Logistic 回归分类的准确率虽高于神经网络和决策树，但二元 Logistic 回归没有区分训练集和检验集。

关键词：中药药性量化；化学元素；数据挖掘；二元 Logistic 回归分析；决策树算法；人工神经网络

中图分类号：R285；Q212；TP181 **文献标志码：**A **文章编号：**0253 - 2670(2024)17 - 5964 - 08

DOI: 10.7501/j.issn.0253-2670.2024.17.022

Comparative study on quantitative methods of medicinal properties of some traditional Chinese medicines based on chemical elements

XU Qinyong[1, 2], HUANG Zhibang[3], YAO Simeng[4], CHEN Yuanfang[5], NING Xiaoying[2], HOU Zhengkun[6], CHEN Xinlin[7]

1. Department of Traditional Chinese Medicine, Binhaiwan Central Hospital of Dongguan, Dongguan 523900, China
2. The First Clinical Medical School of Guangzhou University of Chinese Medicine, Guangzhou 510405, China
3. The First Affiliated Hospital of Shantou University Medical College, Jieyang Haoze Hospital, Jieyang 522021, China
4. Foshan Hospital of Traditional Chinese Medicine, Foshan 528099, China
5. The Third Affiliated Hospital of Southern Medical University, Guangzhou 510630, China
6. Department of Spleen and Stomach Diseases, the First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510405, China
7. College of Basic Medical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

收稿日期：2024-02-16

基金项目：国家重点研发计划项目“基于证素辨识和状态可测原理的动静态中医临床评价方法学构建与示范研究”（2023YFC3503002）；广州中医药大学第一附属医院中青年骨干人才培育项目（09005650008）；广东省中医药信息化重点实验室项目（2021603）；广东省教育厅高校科研项目（2020ZDZX3011）

作者简介：徐钦涌，男，硕士，研究方向为机器学习、中医药研究。E-mail: 875045212@qq.com

***通信作者：**侯政昆，男，博士，主任医师，教授，研究方向为中医药防治脾胃病研究。E-mail: fenghou5128@126.com

Abstract: Objective Based on the content of chemical elements in traditional Chinese medicine (TCM), the classification accuracy of TCM medicinal properties was obtained through decision tree and neural network, and the classification accuracy of different methods was compared. **Methods** The target TCM was obtained from the *Quantitative and Applied Research of TCM Theory*. Excel was used to extract the chemical element data of the target TCM, and two independent samples of non-parametric tests were performed on the medicinal properties and chemical elements of a single TCM based on IBM SPSS Statistics 26 software, and statistically significant associated elements were taken as independent variables. Binary logistic regression analysis, decision tree algorithm, artificial neural network and other methods were used to classify and predict the dependent variable (medicinal properties). Such algorithms were used to obtain the relevant variables, classification accuracy and model function coefficients of TCM medicinal properties, and the classification effects of different methods were compared. **Results** A preliminary element database containing 105 TCMs and 42 chemical elements was established. The drug properties of TCM were statistically analyzed, and the drug properties variables of four *qi*, five flavours and channel tropism were obtained. The correlation factors of the medicinal properties were obtained by non-parametric tests. The cold nature related elements were Be, Sr, Ca, La. The bitter flavor related elements were Mn, Ni, K, Ca, V, Si, Co, Zn; the spleen meridian related elements were Ni, Bi, Co, Be, Eu, Ce, Nd, V, Pr, Sm, La, Dy. Binary logistic regression analysis was used to obtain regression models. The overall accuracies of classification were 87.6%, 91.4%, 81.4% for cold nature, bitter flavor, spleen meridian, respectively. In the training samples of the decision tree model, the classification accuracies were 77.8%, 87.7%, 78.1% for cold nature, bitter flavor and spleen meridian, respectively. The accuracies of the classification of the samples tested were 69.7%, 65.0%, 62.5% for cold nature, bitter flavor and spleen meridian, respectively. In the training samples of the artificial neural network, the classification accuracies were 74.1%, 73.7%, 74.0% for cold nature, bitter flavor and spleen meridian, respectively. In the tested samples, the classification accuracies were 54.5%, 72.4%, 67.9% for cold nature, bitter flavor and spleen meridian, respectively. **Conclusion** Based on the univariate analysis of the relevant factors of medicinal properties, binary logistic regression, decision tree and artificial neural network analysis revealed that there is a certain relationship between the medicinal properties of TCM and chemical elements. From the perspective of classification accuracy, the accuracy of the decision tree and neural network training set is higher than that of the test set. In the comparison of the two methods, the average classification accuracy of the decision tree training set and the test set is higher than that of the neural network. Although the accuracy of binary logistic regression classification is higher than that of neural network and decision tree, binary logistic regression does not distinguish between the training set and the test set.

Key words: quantification of medicinal properties of traditional Chinese medicine; chemical elements; data mining; statistical analysis; binary Logistic regression analysis; decision tree algorithm; artificial neural network

中药药性指中药的性能，是对中药作用性质和特征的高度概括，也是阐明中药疗效机制的理论依据。中药药性作为中医理论体系的重要组成部分，主要包括四性（四气）、五味、归经、升降浮沉及毒性等内容[1]。传统的中药药性理论由于受到古代医家认识水平的限制，因此更偏向于主观性，然而部分古代医家对药性的细化程度已经有了初步的认识和描述，如大热、微温、大寒、微寒等概念在一定程度体现了药性的定量化[2]。随着现代科学技术的发展，许多新技术方法应用中药量化领域，使得中药药性理论得到快速的发展[3-4]。

本研究主要结合统计学方法及机器学习，以中药的化学元素为基础，运用非参数检验、二元 Logistic 回归、决策树、神经网络等方法，分析不同分类方法的预测准确率，从而为后期中医临床处方的客观化和标准化提供具有可行性的思路与方法。

1 数据来源与处理

1.1 数据来源

基于文献计量学分析，本研究采用管竞环主编的《中药理论量化与应用研究》[5]作为数据来源（管竞环教授团队对其临床常用的 105 味中药的微量元素进行数据分析和提取），选取文献中公开的中药化学元素信息进行数据处理和分析。《中药理论量化与应用研究》中记录研究者从药材产地获取道地药材，委托专业机构鉴别药材的真伪，并对药材进行清洗、风干、切片、碾碎获得备用标本，使用电感耦合原子发射光谱法[6]（inductively coupled plasma-atomic emission spectrometry，ICP-AES）测量中药标本的化学元素含量。

1.2 数据处理

对文献中的中药化学元素数据进行提取和整理，并将数据录入到 Excel 表格中，形成初步中药化学元素数据库。随后对元素数据进行核对，进一步明确每味中药所对应的元素数据与来源数据一致。将药性的分类数据列入表中，并核对药性的分类是否正确。在药性的二分类变量中，数值“0”与标签“否”代表药物不具有该药性，数值“1”和标签“是”代表药物具有此药性。

2 研究工具与方法

2.1 研究工具

使用 IBM SPSS Statistics 26 软件进行统计分析，对中药药性和化学元素进行单因素分析（两独立样本非参数检验）、多因素分析（二元 Logistic 回归分析）、机器学习分析（决策树与人工神经网络分析），对分析结果进行检验和对比，分析不同模型的预判准确率及变量对模型的重要性。采用 Microsoft Office Excel（v.2016）对化学元素的数据源进行录入及整理，同时作为中介软件对 SPSS 的分析结果进行导入及处理，制作表格及部分图片。

2.2 研究方法

本研究采用二分类方法对目标中药的化学元素数据库进行提取，获取中药主要化学元素的量化数据源。对单个中药药性和化学元素进行非参数检验，将非参数检验所获得的具有统计学意义的关联化学元素作为下一步药性分析的自变量。通过二元 Logistic 回归分析、决策树算法、人工神经网络等统计学分析及机器学习方法获得与药性具有关联的化学元素，对因变量（药性）及自变量的关联性进行预判。运用此类算法得到中药的四气、五味、归经等药性的相关变量的判别率及函数变量系数，并比较不同方法的判别效果。

3 结果

3.1 化学元素及药性变量

本研究对中药药性进行统计，获得四气、五味、归经的药性变量。从《中药理论量化与应用研究》中获得含有 105 味中药、42 种化学元素的初步元素数据库。将每一个药性作为一个数据表，每个数据表包含 105 味中药及每味药物所包含的 42 种化学元素，共获得 22 个数据表。105 味中药分别是肉桂子、桑葚子、巴戟天、白花蛇舌草、厚朴、虎杖、槐米、黄柏、黄连、黄藤、黄芩、火麻仁、桔梗、橘红、金樱子、九节菖蒲、菊花、连翘、白木耳、白芍、白术、白芷、覆盆子、高良姜、葛根、狗脊、瓜蒌皮、红豆蔻、红花、红蚤休、柏子仁、北沙参、草果、草乌、柴胡、车前子、川芎、郁李仁、云木香、泽泻、浙贝母、天南星、土茯苓、党参、地肤子、独活、鹅不食草、鄂贝母、防己、佛手、佛手花、凌霄花、豆蔻壳、肉豆蔻、枳壳、羌活、龙胆草、麻黄、麦冬、密蒙花、明党参、木通、牛蒡子、牵牛子、秦皮、秦艽、蛇床子、生半夏、生地黄、生附子、升麻、使君子、紫苏子、太子参、桃仁、天麻、乌药、吴茱萸、五味子、细辛、仙茅、香橼皮、小茴香、辛夷、苦杏仁、玄参、元胡、鸦胆子、砂仁壳、砂仁、山茱萸、山柰、川楝子、刺蒺藜、生大黄、丹参、牡丹皮、肉苁蓉、当归、紫草、茯苓、莞蔚子、菟丝子、葶苈子、槟榔。42 种元素分别是 Be、Si、V、Cu、Sr、Hg、Pr、Tb、Yb、F、P、Mn、Zn、Cd、Bi、Nd、Dy、Lu、Na、Cl、Fe、As、Sb、Y、Sm、Ho、Mg、K、Co、Se、I、La、Eu、Er、Al、Ca、Ni、Br、Ba、Ce、Gd 及 Tm。管竞环教授团队[7-9]通过 SPSS 分别对 105 味中药的 42 种元素进行分布检验，发现 42 种元素在每味药物中的分布均为偏态分布，不能使用正态分布的分析方法对数据进行统计分析。

3.2 统计分析与数据挖掘

3.2.1 两独立样本非参数检验　单因素分析可以初步探索预测变量与响应变量的关系，并且当样本量不是很大的时候可以通过单因素分析删除部分无关的预测变量。本研究中化学元素的总体分布为非正态，故使用非参数检验中的曼-惠特尼 U 检验。通过该检验得到化学元素与因变量的相关性，将在各个药性二分类变量（“是”与“否”）中差异具有统计学意义（$P<0.05$）的变量列于表中。

本研究以四气的寒性，五味的苦味，归经的脾经为例具体分析。如表 1～3 所示，与寒性具有统计学意义的独立相关因素有 Be、Sr、Ca、La；与苦味有统计学意义的独立相关因素有 Mn、Ni、K、Ca、V、Si、Co、Zn；与脾经有统计学意义的独立相关因素有 Ni、Bi、Co、Be、Eu、Ce、Nd、V、Pr、Sm、La、Dy。每个药性其他不显著相关元素不列于表中，但不能说明这些元素与药性变量无相关性。通过单因素分析得到与因变量具有统计学意义的关联性自变量，将筛选出来的自变量作为预测变量进入到后面的预测模型中。

3.2.2 二元 Logistic 回归分析　将 105 味中药的 42 种元素数据变量用 IBM SPSS Statistics 26 软件进行二元 Logistic 回归分析，以四气、五味、归经的 22 个变量作为因变量，以 42 种化学元素建立二元 Logistic 回归方程模型。以寒性、苦味、脾经为例，列出具有统计学意义的化学元素及各药性的分析结果。

在四气药性中，寒性方程中的变量见表 4。在寒性预测模型中，具有统计学意义（$P<0.05$）的影响元素有 Si、Co。由表 5 可知，寒性“否”的预测

表 1 两独立样本非参数检验 (分组变量：寒，*n*＝105)

Table 1 Two independent samples nonparametric test (grouping variable: cold, *n* = 105)

自变量	M (P25, P75)	曼-惠特尼 U	*Z*	*P* 值
Be	0.030 (0.020, 0.060)	955.000	−2.411	0.016
Sr	23.500 (11.400, 43.850)	950.000	−2.378	0.017
Ca	4 338.000 (1619.000, 7469.000)	981.500	−2.171	0.030
La	0.344 (0.132, 0.825)	1 013.500	−1.961	0.049

表 2 两独立样本非参数检验 (分组变量：苦，*n*＝105)

Table 2 Two independent samples nonparametric test (grouping variable: bitter, *n* = 105)

自变量	M (P25, P75)	曼-惠特尼 U	*Z*	*P* 值
Mn	35.630 (19.750, 87.520)	857.500	−3.305	0.001
Ni	1.160 (0.530, 2.050)	906.000	−2.994	0.003
K	9 913.000 (6 405.000, 14 869.000)	958.500	−2.656	0.008
Ca	4 338.000 (1 619.000, 7 469.000)	984.000	−2.492	0.013
V	0.630 (0.220, 1.270)	989.500	−2.467	0.014
Si	3 038.000 (1 752.500, 5 833.500)	989.500	−2.457	0.014
Co	0.300 (0.150, 0.585)	1 024.500	−2.264	0.024
Zn	21.800 (12.750, 39.050)	1 054.000	−2.043	0.041

表 3 两独立样本非参数检验 (分组变量：脾经，*n*＝105)

Table 3 Two independent samples nonparametric test (grouping variable: spleen meridian, *n* = 105)

自变量	M (P25, P75)	曼-惠特尼 U	*Z*	*P* 值
Ni	1.160 (0.530, 2.050)	827.000	−3.186	0.001
Bi	0.016 (0.006, 0.034)	865.000	−2.942	0.003
Co	0.300 (0.150, 0.585)	920.500	−2.609	0.009
Be	0.030 (0.020, 0.060)	939.500	−2.515	0.012
Eu	0.012 (0.005, 0.024)	960.500	−2.312	0.021
Ce	0.591 (0.241, 1.335)	973.500	−2.223	0.026
Nd	0.264 (0.091, 0.582)	975.000	−2.214	0.027
V	0.630 (0.220, 1.270)	978.500	−2.200	0.028
Pr	0.095 (0.044, 0.178)	983.000	−2.161	0.031
Sm	0.056 (0.033, 0.144)	999.500	−2.053	0.040
La	0.344 (0.132, 0.825)	1012.000	−1.971	0.049
Dy	0.035 (0.017, 0.091)	1014.000	−1.958	0.050

表 4 二元 Logistic 回归方程中的变量 (寒)

Table 4 Variables in binary Logistic regression equation (cold)

自变量	β	*SE*	Wald χ^2 值	*df*	*P* 值
Si	−0.001	0.001	4.213	1	0.040
Co	−21.328	9.420	5.126	1	0.024

表 5 二元 Logistic 回归分类预测 (寒)[a]

Table 5 Prediction of binary Logistic regression classification (cold)[a]

因变量		寒		准确率/%
		否	是	
寒	否	59	5	92.2
	是	8	33	80.5
总体百分比				87.6

[a] 分界值为 0.500。

[a]boundary value is 0.500.

准确率为 92.2%，“是”的预测准确率为 80.5%，总体准确率为 87.6%。

在五味药性中，苦味方程中的变量见表 6。苦

表 6　二元 Logistic 回归方程中的变量（苦）

Table 6　Variables in binary Logistic regression equation (bitter)

自变量	β	*SE*	Wald χ^2 值	*df*	*P* 值
P	−0.001	0.001	4.380	1	0.036
V	−14.082	6.130	5.276	1	0.022
Fe	0.036	0.016	5.152	1	0.023
Co	−10.812	5.473	3.902	1	0.048
Br	1.363	0.556	6.004	1	0.014
Y	95.746	46.451	4.249	1	0.039
Dy	−862.897	342.753	6.338	1	0.012
Ho	5 511.590	2 085.435	6.985	1	0.008

味的回归模型显著影响因变量的元素有 P、V、Fe、Co、Br、Y、Dy、Ho。由表 7 可知，苦味“否”的预测准确率为 91.8%，“是”的预测准确率为 91.1%，总体准确率为 91.4%。

在归经药性中，脾经方程中的变量见表 8。脾经的回归模型中，显著影响因变量的元素有 Ni、I、La、Ce、Pr、Dy、Ho。由表 9 可知，脾经“否”的预测准确率为 95.3%，“是”的预测准确率为 85.4%，总体准确率 91.4%。

3.2.3　决策树分析　在四气药性中，寒性有 6 个解释变量：Be、Sr、Ca、La、Si、Co。寒性的决策树

表 7　二元 Logistic 回归分类预测（苦）[a]

Table 7　Prediction of binary Logistic regression classification (bitter)[a]

因变量		苦		准确率/%
		否	是	
苦	否	45	4	91.8
	是	5	51	91.1
总体准确率/%				91.4

[a] 分界值为 0.500。

[a]boundary value is 0.500.

表 8　二元 Logistic 回归方程中的变量（脾经）

Table 8　Variables in binary Logistic regression equation (spleen meridian)

自变量	β	*SE*	Wald χ^2 值	*df*	*P* 值
Ni	−6.907	3.429	4.056	1	0.044
I	−15.263	6.486	5.538	1	0.019
La	−29.127	14.064	4.289	1	0.038
Ce	−16.315	7.485	4.751	1	0.029
Pr	556.654	235.018	5.610	1	0.018
Dy	852.616	404.867	4.435	1	0.035
Ho	−5 507.149	2 811.855	3.836	1	0.050

表 9　二元 Logistic 回归分类预测（脾经）[a]

Table 9　Prediction of binary Logistic regression classification (spleen meridian)[a]

因变量		脾经		准确率/%
		否	是	
脾经	否	61	3	95.3
	是	6	35	85.4
总体准确率/%				91.4

[a] 分界值为 0.500。

[a]boundary value is 0.500.

预测模型第 1 层按 La 拆分，即分类树的 2 个初始分支的一级分裂，La 变量标准化的重要性为 100%。Sr 是二级分裂的决定因素，变量标准化的重要性为 35.9%。Si 和 La 是三级分裂的决定因素，Si 变量标准化的重要性是 38.0%。其余变量的重要性如图 1 所示。树模型的分类正确率见表 10，训练集的准确率为 77.8%，检验集的准确率为 69.7%。

五味药性中，苦味的解释变量是 Mn、Ni、K、Ca、V、Si、Co、Zn、P、Fe、Dy、Ho、Br、Y。决

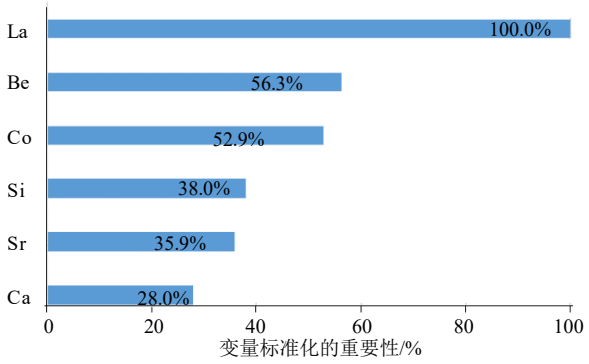

图 1　寒性决策树模型中自变量的标准化重要性

Fig. 1　Importance of standardization of independent variables in cold nature decision tree model

表 10　决策树模型分类预测（寒）

Table 10　Prediction of decision tree model classification (cold)

样本	指标	预测		
		否	是	准确率/%
训练	否	37	10	78.7
	是	6	19	76.0
	总体准确率/%	59.7	40.3	77.8
检验	否	12	5	70.6
	是	5	11	68.8
	总体准确率/%	51.5	48.5	69.7

策树预测模型的第 1 层按 Si 拆分，Si 变量标准化的重要性为 95.9%。Ca、P 是二级分裂的决定因素，变量标准化的重要性分别为 39.8%、30.4%。Fe 是三级分裂的决定因素，Fe 变量标准化的重要性是 100%。苦味决策树模型训练集的预测准确率为 87.7%，检验集的预测准确率为 65.0%。

归经药性中，脾经的解释变量是 Ni、Bi、Co、Be、Eu、Ce、Nd、V、Pr、Sm、La、Dy、I、Ho。决策树预测模型的第 1 层按 Ni 拆分，变量标准化的重要性为 100%。I 是二级分裂的决定因素，变量标准化的重要性是 56.9%。Bi 是三级分裂的决定因素，变量标准化的重要性是 89.9%。脾经决策树模型训练集的预测准确率为 78.1%，检验集的预测准确率为 62.5%。

3.2.4 人工神经网络分析 将药性作为因变量，化学元素作为自变量，选用系统自动的多层感知器神经网络模型进行数据分析。协变量的重标度方法为正态化，隐藏层激活函数为双曲正切，输出层激活函数为 Softmax。在四气药性中，寒性神经网络有 6 个输入节点，1 个隐含层神经元，2 个输出节点。自变量与决策树模型一致。重要性从大到小排列依次是 Co（0.228）、Be（0.204）、Ca（0.185）、La（0.180）、Si（0.121）、Sr（0.083），标准化重要性分别是 100.0%、89.3%、81.1%、79.0%、52.9%、36.3%。寒性模型总体预测准确率见表 11，训练集的预测分类准确率为 72.1%，检验集的预测分类准确率为 54.5%。

五味药性中，苦味人工神经网络有 14 个输入节点，5 个隐含层神经元，2 个输出节点。自变量重要性从大到小排列依次是 K、Mn、Ca、V、Si、Ni、Dy、P、Co、Fe、Zn、Br、Y、Ho，各变量标准化重要性分别是 100.0%、85.1%、83.5%、64.4%、60.1%、42.3%、38.4%、37.1%、34.7%、32.3%、22.9%、22.6%、19.6%、14.1%。苦味模型训练集的总体预测分类准确率为 73.7%，检验集总体预测分类准确率为 72.4%。

表 11 人工神经网络模型分类预测 (寒)

Table 11 Prediction of artificial neural network model classification (cold)

样本	指标	预测		
		否	是	准确率/%
训练	否	22	12	64.7
	是	5	22	81.5
	总体准确率/%	44.3	55.7	72.1
检验	否	14	12	53.8
	是	8	10	55.6
	总体准确率/%	50.0	50.0	54.5

归经药性中，脾经神经网络有 14 个输入节点，2 个隐含层神经元，2 个输出节点。自变量重要性从大到小排列依次是 Bi、Ni、I、Dy、Co、V、Be、Eu、Ce、Nd、La、Sm、Pr、Ho，各变量标准化重要性分别是 100.0%、94.4%、84.4%、55.3%、49.1%、35.5%、33.4%、26.7%、21.3%、19.8%、18.4%、13.9%、13.2%、4.7%。脾经模型训练集的总体预测分类准确率为 74.0%，检验集总体预测分类准确率为 67.9%。

3.2.5 判别分析 判别分析是一种分类方法，指在已知判别的情况下，对未知类别的观测量归类到已知类别的多元分析法[10]。本研究采用 Fisher 判别分析法，对寒性药物进行判别。因 Fisher 判别分析属于分类判别，故需对数据进行标准化处理，等级范围为 1～10 个等级，等级差相等，并将元素数值取整数（四舍五入）。对北沙参、浙贝母、丹参、黄连、白芍、柴胡、黄芩、白花蛇舌草、菊花、连翘、枳壳 11 味寒性药进行训练，并对麦冬、大黄 2 味寒性药进行预判。训练过程中，因没有寒性分级为 2、3、8 级的药物，因此分级为寒性等级 1、4、5、6、7、9 共 6 个等级组别。判别中给予 4 个函数进行预判，各函数特征值如图 2 所示，函数 1 能较好地判别变量的数值。分类变量之间，同一类别的变量间距离越近、不同类别间的变量距离越远，说明分类特征越明显。函数分类如图 3 所示，在函数 1 所在的横轴上，各类别变量间的组质心距离较远，函数 1 分类更具有显著性。Fisher 判别分析结果如表 12 所示，该判别方法将未分组的麦冬、大黄 2 味药分别判为 4 组和 6 组，即 2 味中药的寒性量化整数值分别为 4 和 6，这与临床用药经验比较相符。且该判别分析对初始分组案例中的 72.7%的变量进行了正确分类。

4 讨论与展望

中药药性在临床运用中常遇到不统一甚至相反的问题。如枸杞在《药性论》中被记载其味甘，平。《本草蒙筌》则记载其味甘、苦，气微寒，无毒。《景岳全书》言其味甘、微辛，气温。同一种药就有 3 种说法，对于中医药的运用及传承造成阻碍。另一方面，自然界植物药、动物药数以万计，中药成分复杂，配伍灵活多变，在与不同药物联合使用时效果可能会起到相反的作用，如中药“十八反”“十九畏”

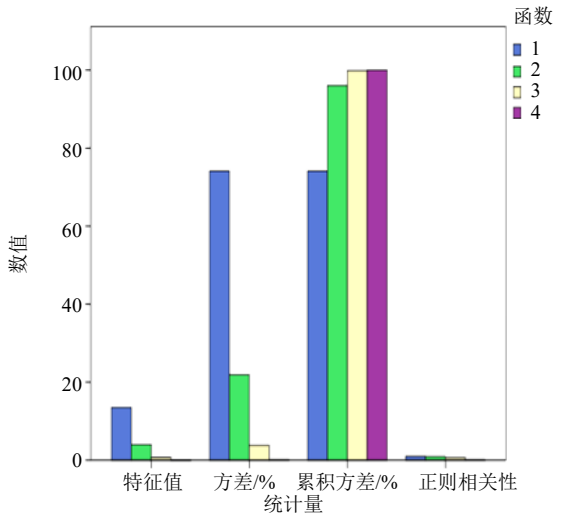

图 2　判别函数特征图

Fig. 2　Feature graph of discriminant function

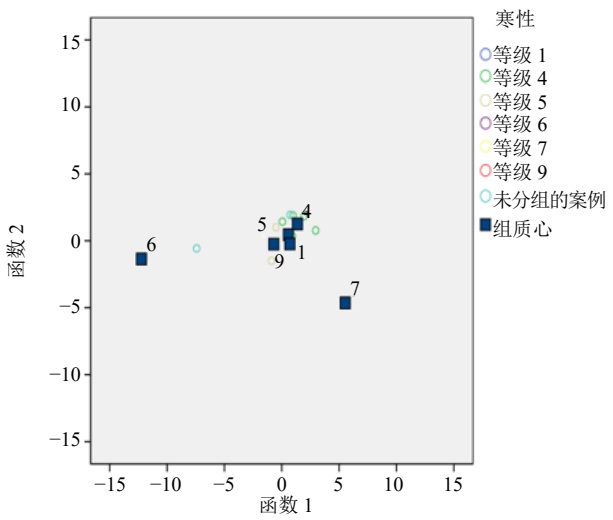

图 3　函数分类图

Fig. 3　Function classification graph

表 12　寒性药物 Fisher 判别分类结果 [a]

Table 12　Results of Fisher discriminate classification of cold drugs[a]

项目	分组	预测组成员						合计
		1	4	5	6	7	9	
计数	1	1	0	0	0	0	0	1
	4	2	3	0	0	0	0	5
	5	1	0	1	0	0	0	2
	6	0	0	0	1	0	0	1
	7	0	0	0	0	1	0	1
	9	0	0	0	0	0	1	1
	未分组的案例	0	1	0	1	0	0	2
药物数量占比/%	1	100.0	0	0	0	0	0	100.0
	4	40.0	60.0	0	0	0	0	100.0
	5	50.0	0	50.0	0	0	0	100.0
	6	0	0	0	100.0	0	0	100.0
	7	0	0	0	0	100.0	0	100.0
	9	0	0	0	0	0	100.0	100.0
	未分组的案例	0	50.0	0	50.0	0	0	100.0

a-已对初始分组案例中的 72.7%进行了正确分类。

a-72.7% of the initial grouping cases have been correctly classified.

等情况。对于未知药物四气、五味、归经的判断不能简单的一言概之，需要经过数据分析、实验探索才能取得人们认可。

自机器学习算法面世以来，基于机器学习探究微量元素与中药联系的研究已较为成熟。如刘进等[11]应用支持向量机预测中药药性，发现 Ca、Fe 元素对温热药识别较敏感。但该研究数据有限，仅纳入 7 种元素进行预测研究，样本量较少，存在一定的限制。杨波[12]从有机成分、无机成分着手，研究中药药性与化学成分的相关性。2011 年，龙伟[13]提出“计算中药学”的理念，旨在通过计算科学、数理统计学以及药物化学等现代科学技术方法来解决中药问题。其通过原创的重心处理技术，结合化学描述符计算和支持向量机算法构建了预测率超过 80%的中药寒热预测系统。多项研究表明，机器学习对中药药性研究可提供较大帮助[14]。

本研究前期基于文献计量学，研究人员纳入了管竞环教授团队的文献数据进一步分析。经过单因

素分析获得药性的相关因素，并将相关因素运用到分类预测模型中。通过二元 Logistic 回归[15]、决策树[16]、人工神经网络分析[17]，揭示了中药药性与化学元素间存在一定联系，并获得不同模型的自变量重要性及分类正确率。研究中将训练集和测试集的 100 余味中药微量元素数据输入 SPSS，系统将 70%数据作为训练集，30%数据作为测试集。本研究将因变量（药性）与自变量（化学元素）输入软件，运用不同分类方法对同一个药性进行分类。由于方法不同，SPSS 系统形成的模型方程不尽相同，自变量也有所不同。通过观察，发现同一个药性（如寒性）的不同分类方法所得出的关键自变量（化学元素）有一部分相同，可以认为这些相同的自变量与因变量存在较紧密的联系。

从分类准确率来看，决策树与人工神经网络训练集的准确率均高于检验集。在这 2 种方法的比较中，决策树训练集、检验集平均分类准确率均高于人工神经网络。二元 Logistic 回归分类的准确率虽高于人工神经网络和决策树，但二元 Logistic 回归没有区分训练集和检验集。本研究将中药的药性及化学元素的数据库导入 SPSS 软件，选用系统判别分析方法，药性选入分组变量，定义范围是 1～10，42 种元素数据放入自变量，统计量函数系数选择 Fisher 和未标准化，运行软件可获得四气、五味、归经等药性变量的典型判别函数和 Fisher 线性判别函数。并基于函数特征值、判别结果调整参数。在后续研究中可采用德尔菲法邀请具有 20 年以上中药临床运用经验的专家，对判别出的中药药性、归经进行合理性评判，从而调整预测方程与判别系数。

本研究表明，通过 ICP-AES 提取中药中的微量元素，基于机器学习算法预测、判别药物药性，从而解决文献记载矛盾、未知药物药性判断的难题。是一种行之有效的科学方法。其不仅能让临床医师迅速识别中药的药性，指导用药。也给研究者提供更多的理论依据及实验数据。

但该研究也存在一定的局限性，所纳入的数据来源是管竞环教授基于实验室研究所获得的元素数据，中药样本量偏少，数据量不足，但数据较为完整、规范、统一，可以在后期的研究中对更多中药的微量元素进行分析提取，扩大中药的微量元素数据，使研究的样本量更加丰富。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 张铁军, 刘昌孝. 中药五味药性理论辨识及其化学生物学实质表征路径 [J]. 中草药, 2015, 46(1): 1-6.

[2] 胡文. 基于微量元素、有机成分和专家评阅结合的中药药性量化研究 [D]. 广州: 广州中医药大学, 2019.

[3] 侯政昆, 胡文, 刘凤斌, 等. 中医“症状-证型-中药-组方-评价”系统量化研究模式的分析探讨 [J]. 中华中医药杂志, 2018, 33(1): 14-18.

[4] 文艺, 李海文, 刘凤斌, 等. 中药药性量化研究的方法学进展 [J]. 中华中医药杂志, 2017, 32(3): 1181-1183.

[5] 管竞环, 朱宏斌, 马威. 中药理论量化与应用研究 [M]. 北京: 人民军医出版社, 2014: 282.

[6] 刘磊, 杨艳. 电感耦合等离子体-原子发射光谱法测定中药中微量元素 [J]. 光谱实验室, 2010, 27(5): 1964-1967.

[7] 管竞环, 李恩宽, 代行信, 等. 无机元素在植物类中药中分布规律的研究 [J]. 微量元素与健康研究, 1994, 11(3): 30-33.

[8] 汤学军, 管竞环. 中药辛、甘、苦味与稀土元素的关系 [J]. 微量元素与健康研究, 1994, 11(4): 24-26.

[9] 汤学军, 管竞环, 薛莎. 中药微量元素含量区间尺最佳分级问题的探讨 [J]. 广东微量元素科学, 1995, 2(5): 16-19.

[10] 陈希镇, 曹慧珍. 判别分析和SPSS的使用 [J]. 科学技术与工程, 2008, 8(13): 3567-3571.

[11] 刘进, 邓家刚, 覃洁萍. 应用支持向量机探讨中药无机元素与药性的相关性 [J]. 中药材, 2008, 31(12): 1933-1936.

[12] 杨波. 植物类中药寒热药性与化学成分相关性的文献研究 [D]. 济南: 山东中医药大学, 2010.

[13] 龙伟. “计算中药学” 在中药药性及复方研究中的应用 [D]. 北京: 北京协和医学院, 2011.

[14] 杨淇, 郝二伟, 侯小涛, 等. 基于药性理论的中药抗辐射预测模型的构建 [J]. 中草药, 2024, 55(8): 2684-2693.

[15] 于晓牧. logistic 回归多重共线性诊断方法的研究 [D]. 大连: 大连医科大学, 2010.

[16] 杨静, 张楠男, 李建, 等. 决策树算法的研究与应用 [J]. 计算机技术与发展, 2010, 20(2): 114-116.

[17] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述 [J]. 计算机工程与应用, 2021, 57(11): 57-69.

[责任编辑 潘明佳]