

近红外光谱技术应用于中药四类味觉分类辨识的可行性分析

王小鹏¹, 张璐^{2,3,4}, 陈鹏举¹, 王艳丽^{2,3,4}, 李涵¹, 桂新景^{1,2,3,4}, 刘瑞新^{2,3,4*}, 李学林^{2,3,4,5*}

1. 河南中医药大学, 河南 郑州 450046

2. 河南中医药大学第一附属医院 药学部, 河南 郑州 450000

3. 河南省中药临床应用、评价与转化工程研究中心, 河南 郑州 450000

4. 河南中医药大学 呼吸疾病中医药防治省部共建协同创新中心, 河南 郑州 450046

5. 郑州众生实业集团有限公司, 河南 郑州 450001

摘要: 目的 探讨近红外光谱 (near infrared spectroscopy, NIRS) 技术用于不同味觉中药分类辨识的可行性。方法 以分别具有苦、甜、酸、咸 4 种味道的 35 种饮片水煎液和 12 种常用食品类成分溶液为研究载体, 获取其 NIRS 信息作为自变量 (X), 以《中国药典》2020 年版一部饮片性状项下味觉描述结合口尝结果作为标杆信息 (Y), 比较 5 种光谱预处理方法, 然后利用主成分分析-判别分析 (principal component analysis-discriminant analysis, PCA-DA)、偏最小二乘-判别分析 (partial least squares-discriminant analysis, PLS-DA)、 K -近邻算法 (K -nearest neighbor algorithm, KNN) 分别对中药苦、甜、酸、咸 4 类味觉进行模型辨识探讨, 并基于留一法交互验证结果的混淆矩阵 (confusion matrix, CM) 和敏感性、特异性、精度等指标对模型的性能进行综合评价。结果 标准正态变量变换 (standard normal variable transformation, SNV) 是相对更有效的预处理方法, 以预处理后的光谱数据建立的 PCA-DA 模型为最优辨识模型, 其对苦与非苦、甜与非甜、酸与非酸、咸与非咸、四分类辨识的留一法交互验证正判率分别为 89.4%、93.6%、87.2%、97.9%、87.2%。四分类辨识混淆矩阵也以 PCA-DA 模型性能较好, 对苦、甜、酸、咸的分类正确率分别为 87%、94%、73%、100%。PCA-DA 模型的敏感性、特异性、精度分别平均为 0.89、0.91、0.88, 均极显著优于 PLS-DA 和 KNN 模型 ($P < 0.01$)。结论 基于 NIRS 技术初步建立了中药苦、甜、酸、咸 4 类味觉的分类辨识模型, 可为中药五味的定性辨识研究提供新的方法参考。

关键词: 中药; 近红外光谱; 五味; 味觉辨识; 化学计量学; 苦; 甜; 酸; 咸; 主成分分析-判别分析; 偏最小二乘-判别分析; K -近邻算法; 标准正态变量变换

中图分类号: R283.6

文献标志码: A

文章编号: 0253-2670(2023)04-1076-11

DOI: 10.7501/j.issn.0253-2670.2023.04.007

Feasibility analysis of near-infrared spectroscopy technology applied to classification and identification of four kinds of taste in traditional Chinese medicine

WANG Xiao-peng¹, ZHANG Lu^{2,3,4}, CHEN Peng-ju¹, WANG Yan-li^{2,3,4}, LI Han¹, GUI Xin-jing^{1,2,3,4}, LIU Rui-xin^{2,3,4}, LI Xue-lin^{2,3,4,5}

1. Henan University of Chinese Medicine, Zhengzhou 450046, China

2. Department of Pharmacy, the First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450000, China

3. Henan Engineering Research Center for Clinical Application, Evaluation and Transformation of Traditional Chinese Medicine, Zhengzhou 450000, China

4. Co-construction Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases by Henan & Ministry of Education, Henan University of Chinese Medicine, Zhengzhou 450046, China

5. Zhengzhou Zhongsheng Industrial Group Co., Ltd., Zhengzhou 450001, China

收稿日期: 2022-09-20

基金项目: 国家重点研发计划中医药现代化重点专项课题 (2017YFC1703400); 国家重点研发计划中医药现代化重点专项课题 (2017YFC1703402); 河南省中医药拔尖人才培养项目资助 (2019ZYBJ07); 河南省高层次人才特殊支持“中原千人计划”——“中原青年拔尖人才”项目 (ZYQR201912158); 河南省卫生健康中青年学科带头人专项 (HNSWJW-2020014); 河南省科技攻关项目 (222102310377); 河南省中医药科学研究专项 (2021JDZY104); 河南省中医药科学研究专项 (2021JDZY106); 2022 年协同创新中心研究生科研创新基金项目 (协同中心[2022]002 号)

作者简介: 王小鹏 (1994—), 硕士研究生, 从事中药饮片临床应用现代化研究。E-mail: wxp940320@163.com

***通信作者:** 李学林 (1960—), 教授, 主任药师, 博士生导师, 博士后合作导师, 主要从事中药应用形式研究。E-mail: xuelinli450000@163.com
刘瑞新 (1980—), 博士, 主任药师, 研究生导师, 博士后合作导师, 从事中药饮片临床应用现代化关键技术研究。

E-mail: liuruixin7@163.com

Abstract: Objective To explore the feasibility of near-infrared spectroscopy (NIRS) technology for classification and identification of traditional Chinese medicine with different tastes. **Methods** The 35 kinds of traditional Chinese medicine decoctions and 12 kinds of common food ingredient solutions with four tastes of bitterness, sweetness, sourness and saltiness were taken as research object. The spectral information of samples which was obtained by near-infrared spectroscopy technology was used as independent variable (X), the taste description of Chinese medicinal decoction pieces feature in the first part of *Chinese Pharmacopoeia* (2020 edition) combined with the results of traditional human taste panel method were used as benchmarking information (Y). After compared five kinds of spectral pretreatment methods, three chemometric methods including principal component analysis-discriminant analysis (PCA-DA), partial least square-discriminant analysis (PLS-DA) and KNN (K -nearest neighbor) were used to establish the identification model of four kinds of traditional Chinese medicine tastes, respectively. The performance of models was evaluated synthetically by confusion matrix (CM), sensitivity, specificity and precision of leave-one-out cross validation results. **Results** Standard normal variate transformation (SNV) was a relatively effective pretreatment method. The PCA-DA model established with the pretreatment spectral data was the optimal identification model, and the accuracy of leave-one-out cross validation of bitterness or non-bitterness, sweetness or non-sweetness, sourness or non-sourness, saltiness and non-saltiness and four-class identification were 89.4%, 93.6%, 87.2%, 97.9% and 87.2%, respectively. The PCA-DA model performed better in confusion matrix of four-class identification, and the classification accuracy of bitterness, sweetness, sourness and saltiness were 87%, 94%, 73% and 100%, respectively. The average sensitivity, specificity and precision of PCA-DA model were 0.89, 0.91 and 0.88, respectively, which were significantly higher than PLS-DA and KNN models ($P < 0.01$). **Conclusion** In this study, the identification models to bitterness, sweetness, sourness and saltiness of traditional Chinese medicine were established by near-infrared spectroscopy technology, which provided a new method for qualitative identification of five flavors of traditional Chinese medicine.

Key words: traditional Chinese medicine; near-infrared spectroscopy; five flavors; tastes identification; chemometrics; bitterness; bitterness; sweetness; sourness; saltiness; principal component analysis-discriminant analysis; partial least squares-discriminant analysis; K -nearest neighbor algorithm; standard normal variable transformation

中药五味是指药物具有苦、甘、酸、辛、咸 5 种基本味道^[1], 是中药药性理论的核心内容之一, 也是中医遣方用药、辨证论治的重要理论基础。现代的五味药性研究以“功能味”为主^[2-4], 即根据该药的功效作用进行反推确定, 该研究对于指导临床遣药组方和保障中药的安全、合理应用具有重要意义。然而, 中药的“口尝味”也不容小觑, “口尝味”作为中药五味药性的原始定义, 反映了人们对于药物最直观、最真实的味觉感受, 故有“非口不能味也”, 这也是五味界定的最主要的依据之一^[2]。通过对中药五味(口尝味)进行分类辨识研究, 将有助于丰富中药五味定性定量辨识技术和手段, 促进中药五味药性理论的现代化发展。

目前中药口尝味的常用辨识方法有口尝法^[5-6]、仿生技术评价法^[7]等。其中口尝法具有检测成本低、能直观反应味觉感受等优势, 是中药五味评价应用最早、最广泛的方法, 如《荀子·正名篇》:“甘、苦、咸、淡、辛、酸、奇味以口异”, 但其同时存在易疲劳、操作繁琐费时、主观性强、重复性差、安全性差等不足, 无法满足当前中药味觉评价的需求。近年来备受青睐的电子舌、电子鼻等现代仿生技术^[8], 相比口尝法具有简便快捷、客观性强、灵敏度高、安全性好、可智能检测等特点, 已广泛应用于中药

滋味、气味等的客观评价^[9-10], 为中药五味定性、定量辨识研究提供了可参照的技术方法。然而, 该类仿生技术的传感器数量有限, 如目前大多电子舌一般需要 5~20 个传感器来组成阵列^[11], 相比于人类 9000 多个味蕾^[12], 传感器数量显得较少, 所获得的味觉信息相对有限。能否应用一种能够获取丰富的味觉相关成分信息的方法对中药五味进行辨识研究? 近红外光谱(near infrared spectroscopy, NIRS)技术为此提供了可能。NIRS 技术系指通过测定物质在 NIRS 区的特征光谱并利用化学计量学方法提取相关信息, 对物质进行定性、定量分析的一种光谱分析技术。该技术主要利用 O-H、N-H、C-H 等含氢基团化学键伸缩振动的倍频及合频吸收 NIRS 特点^[13], 可以获得绝大多数物质的物理、化学、生物学信息^[14], 并用于中药生产过程中在线监测^[15]。中药饮片的味觉信息与其内在的物质基础密切相关, 如苦味饮片多含有生物碱、苷类、黄酮等成分^[16], 甜味饮片多含有糖类、蛋白质、氨基酸类等成分^[17], 酸味饮片多含有有机酸、鞣质等成分, 咸味饮片多含有无机盐、矿物质类等成分^[18]。目前, 已有相关学者基于 NIRS 对红茶风味^[19]、咖啡口感^[20]进行评价, 并对中药及食品的味觉相关成分如龙胆苦苷^[21]、总多糖^[22]、总酸^[23]等进行定量辨识研究,

模型的整体预测性能较好 ($R^2 > 0.8$), 表明 NIRS 信息与中药饮片及食品类成分味觉表达的物质基础具有一定的相关性, 将该技术用于不同味觉中药分类辨识具有一定的可行性。因此, 本研究基于《中国药典》2020 年版一部筛选 35 种代表性中药饮片, 并补充 12 种味觉信息明确的常用食品类成分, 拟采用 NIRS 技术结合化学计量学对其苦、甜、酸、咸 4 类味觉进行分类辨识研究, 以期对中药五味定性辨识研究提供新的方法参考。

1 仪器与材料

1.1 仪器

Nicolet 6700 傅里叶变换近红外光谱仪, 配 InGaAs 检测器、Opus 5.5 分析软件和 TQ analyst 软

件, 美国 Thermo Nicolet 公司; LCD-A200 型电子天平, 福州华志科学仪器有限公司; CP225D 型十万分之一电子天平, 德国 Sartorius 公司; SDHC07-210 型电磁炉, 浙江绍兴苏泊尔生活电器有限公司; LXJ-IIB 型低速大容量多管离心机, 上海安亭科学仪器厂; HK250 型科导台式超声清洗器, 上海科导超声仪器有限公司; TD360 型红外测温仪, 深圳市泰克曼电子有限公司; AM-5250B 型磁力搅拌器, 天津奥特赛恩斯仪器有限公司; HHS 电热恒温水浴锅, 常州普天仪器制造有限公司。

1.2 材料

中药饮片及食品类样品信息如表 1 所示, 其中蔗糖、无水柠檬酸、氯化钠 (NaCl) 既是样品, 也

表 1 样品信息

Table 1 Information of samples

编号	类别	名称	《中国药典》性状 项下味觉描述	批号	用量/ g	编号	类别	名称	《中国药典》性状 项下味觉描述	批号	用量/ g
1	苦味	龙胆	气微, 味甚苦	140503	45.00	25	甜味	白茅根	气微, 味微甜	161225	195.00
2		黄柏	气微, 味极苦	160704	75.00	26		猫爪草	气微, 味微甘	1702132	225.00
3		苦参	气微, 味极苦	1703122	67.50	27		蜂蜜	气芳香, 味极甜	1609305	225.00
4		黄连	气微, 味极苦	1702041	35.00	28		麦芽糖*	—	20170519	4.00
5		穿心莲	气微, 味极苦	150820	75.00	29		饴糖*	—	20170427	4.00
6		五加皮	气微香, 味微辣而苦	161225	75.00	30		葡萄糖*	—	20170321	4.00
7		防己	气微, 味苦	161208	75.00	31		红糖*	—	20170405	4.00
8		连翘	气微香, 味苦	161224	105.00	32		蔗糖*	—	20170513	8.00
9		青风藤	气微, 味苦	1701211	90.00	33	酸味	木瓜	气微清香, 味酸	161224	75.00
10		野菊花	气芳香, 味苦	1702151	120.00	34		醋五味子	果肉气微, 味酸	170401	40.00
11		黄芩	气微, 味苦	1703171	65.00	35		乌梅	气微, 味极酸	170401	90.00
12		预知子	气微香, 味苦	1701181	60.00	36		南五味子	果肉气微, 味酸	160608	40.00
13		莲子心	气微, 味苦	160901	105.00	37		瓜蒌皮	味淡、微酸	160615	195.00
14		泽泻	气微, 味微苦	1704141	80.00	38		炒山楂	味酸、微甜	1703192	105.00
15		槐花	气微, 味微苦涩	1703042	75.00	39		黑醋*	—	7191187116	5.00
16	甜味	甘草	气微, 味甜而特殊	170401	60.00	40		白醋*	—	9535521243	5.00
17		炙甘草	具焦香气, 味甜	170401	60.00	41		乳酸*	—	20170609	0.07
18		龙眼肉	气微香, 味甜	161201	45.00	42		L-苹果酸*	—	004-00345	0.07
19		罗汉果	气微, 味甜	170303	120.00	43		无水柠檬酸*	—	2016121503A	0.07
20		枸杞子	气微, 味甜	161101	90.00	44	咸味	食盐*	—	20170415D16	4.00
21		芦根	气微, 味甘	161209	225.00	45		NaCl*	—	1160904	2.30
22		红芪	气微, 味微甜	161110	195.00	46		玄明粉	气微, 味咸	150715	60.00
23		使君子仁	气微香, 味微甜	170525	105.00	47		芒硝	气微, 味咸	161115	120.00
24		党参	有特殊香气, 味微甜	170401	195.00						

*表示食品类样品, 其中除黑醋和白醋为体积分数外, 其余均为质量分数

*indicates food sample, the dosage unit of black vinegar and white vinegar is volume fraction, and the dosage unit of other food samples is mass fraction

分别为甜、酸、咸味的标准参比物质。盐酸小檗碱 (berberine hydrochloride, BH) 为苦味的标准参比物质, 批号为 140406, 质量分数 98.62%, 购自于四川省玉鑫药业有限公司。

2 方法与结果

2.1 口尝评价

2.1.1 样品溶液的制备

(1) 口尝参比样品溶液的制备: 分别取 BH、

蔗糖、无水柠檬酸、NaCl 适量, 加纯化水配制成一系列不同质量浓度的溶液, 分别作为苦、甜、酸、咸味的参比溶液。经口尝实验志愿者对多个质量浓度进行预试后, 确定对应各味觉等级的参比溶液质量浓度。结果见表 2。

(2) 中药饮片水煎液的制备: 以龙胆为例, 精密称取 45 g, 置于适宜容器内, 加水 2000 mL, 浸泡 30 min, 电磁炉加热煎煮 (功率 2100 W), 沸腾

表 2 4 类味觉参比物质味觉程度 (*I*) 的定性描述、等级及定量范围

Table 2 Qualitative description, grade and quantitative range of taste degree (*I*) of four kinds of taste reference substances

味觉类别	参比物质	参比溶液质量浓度/(mg·mL ⁻¹)	口感描述	赋予 <i>I</i> 等级	<i>I</i> 取值范围
苦味	BH	0.00	几乎没有苦味	I	0.5~1.5
		0.01	略有苦味	II	1.5~2.5
		0.05	可接受的苦味	III	2.5~3.5
		0.10	很苦, 但还可以忍受	IV	3.5~4.5
		0.50	不能忍受的苦味	V	4.5~5.5
甜味	蔗糖	0.00	没有甜味	I	0.5~1.5
		40.00	略有甜味	II	1.5~2.5
		80.00	正常甜味, 乐于接受	III	2.5~3.5
		150.00	很甜, 但还可接受	IV	3.5~4.5
		290.00	不乐意接受的甜味	V	4.5~5.5
酸味	无水柠檬酸	0.00	没有酸味	I	0.5~1.5
		0.30	略有酸味	II	1.5~2.5
		0.70	中等酸味, 乐于接受	III	2.5~3.5
		1.70	很酸, 但还可接受	IV	3.5~4.5
		3.30	极酸, 无法忍受, 眼睛几乎眯成一条缝	V	4.5~5.5
咸味	NaCl	0.00	没有咸味	I	0.5~1.5
		10.00	略有咸味	II	1.5~2.5
		23.00	中等咸味	III	2.5~3.5
		48.00	很咸, 但还可接受	IV	3.5~4.5
		71.00	极咸, 无法忍受	V	4.5~5.5

后转小火 (功率 600 W), 煎煮 20 min, 滤过, 滤液另器收集备用; 二煎加水 2 L, 沸腾后煎煮 10 min, 余同一煎。合并滤液, 混匀, 冷却至室温, 4000 r/min (离心半径 19.6 cm) 离心 15 min, 取上清液, 定容至 4 L, 灌装, 压盖, 灭菌 (105 °C 流通蒸汽灭菌 45 min), 备用。中药饮片的用量以《中国药典》2020 年版规定用量平均值的 10 倍量为依据, 各饮片具体用量见表 1。

(3) 食品类样品溶液的制备: 以麦芽糖为例, 精密称取麦芽糖适量, 加纯化水溶解, 定容, 配置成 4% 的麦芽糖溶液。各食品类样品的用量通过口尝预试验确定, 见表 1。

2.1.2 志愿者筛选与口感标准化培训 参考文献[6], 选择 27 名健康志愿者作为受试者, 在试验前签订知情同意书。使用口尝参比溶液 (表 2) 对志愿者进行口感标准化培训, 目的是使志愿者具有统一的味觉等级感受, 以便于口尝样品时进行标准的味觉评价。

2.1.3 口尝试验 27 名志愿者根据自己的口尝感受, 结合口尝参比溶液, 对“2.1”项下制备的待测样品溶液进行口感评价, 评价结果记录于预先设计好的“口感评价表”中, 口尝测试具体步骤参照文献方法[24]。

2.1.4 离群值的处理 本实验对象为生物样本, 参

与口尝试验的志愿者间存在个体化差异, 为保证整体实验数据的准确性, 采用 Grubbs 检验法对数据进行异常值的循环检验和剔除。考虑口尝试验中过大和过小的异常值均有存在的可能, 采用双侧检验进行剔除, 选择检出水平为 0.1、剔除水平为 0.05。

2.1.5 口尝评价结果 经 Grubbs 检验循环剔除异常值后, 苦、甜、酸、咸 4 类样品的口尝结果如表 3 所示。

表 3 样品的口尝结果 ($\bar{x} \pm s$, $20 \leq n \leq 26$)

Table 3 Taste results of samples ($\bar{x} \pm s$, $20 \leq n \leq 26$)

序号	名称	味觉	味觉值	序号	名称	味觉	味觉值
1	龙胆	苦味	4.5±0.5	25	白茅根	甜味	1.2±0.3
2	黄柏		4.4±0.7	26	猫爪草		1.2±0.3
3	苦参		4.2±1.1	27	蜂蜜		2.2±0.5
4	黄连		4.7±0.6	28	麦芽糖		1.3±0.3
5	穿心莲		3.0±0.8	29	饴糖		1.6±0.3
6	五加皮		3.4±0.9	30	葡萄糖		1.9±0.4
7	防己		1.9±0.8	31	红糖		2.0±0.6
8	连翘		3.2±0.5	32	蔗糖		3.0±0.0
9	青风藤		3.8±0.5	33	木瓜	酸味	3.4±0.3
10	野菊花		4.1±0.7	34	醋五味子		4.7±0.3
11	黄芩		2.3±0.6	35	乌梅		4.9±0.3
12	预知子		2.0±0.5	36	南五味子		3.8±0.4
13	莲子心		4.3±0.6	37	瓜蒌皮		2.6±0.6
14	泽泻		2.2±0.6	38	炒山楂		4.2±0.3
15	槐花		1.4±0.4	39	黑醋		4.2±0.6
16	甘草	甜味	1.7±0.2	40	白醋		4.4±0.4
17	炙甘草		1.9±0.4	41	乳酸		3.2±0.4
18	龙眼肉		1.4±0.3	42	L-苹果酸		2.4±0.4
19	罗汉果		2.5±0.6	43	无水柠檬酸		3.0±0.0
20	枸杞子		1.4±0.2	44	食盐	咸味	3.7±0.6
21	芦根		1.2±0.3	45	NaCl		3.0±0.0
22	红芪		1.5±0.4	46	玄明粉		1.7±0.5
23	使君子仁		1.0±0.3	47	芒硝		2.2±0.5
24	党参		1.7±0.4				

2.2 NIRS 信息的采集

取中药饮片水煎液及食品类样品溶液适量, 采用透射方式采集其光谱信息。检测条件为光谱范围 10 000~4000 cm^{-1} , 分辨率 8.0 cm^{-1} , 扫描次数 32 次。环境条件为温度 $(25 \pm 2)^\circ\text{C}$, 相对湿度 50%~60%。每个样品平行扫描 3 次, 取其平均光谱作为样品的 NIRS。47 种样品的 NIRS 图见图 1-a。

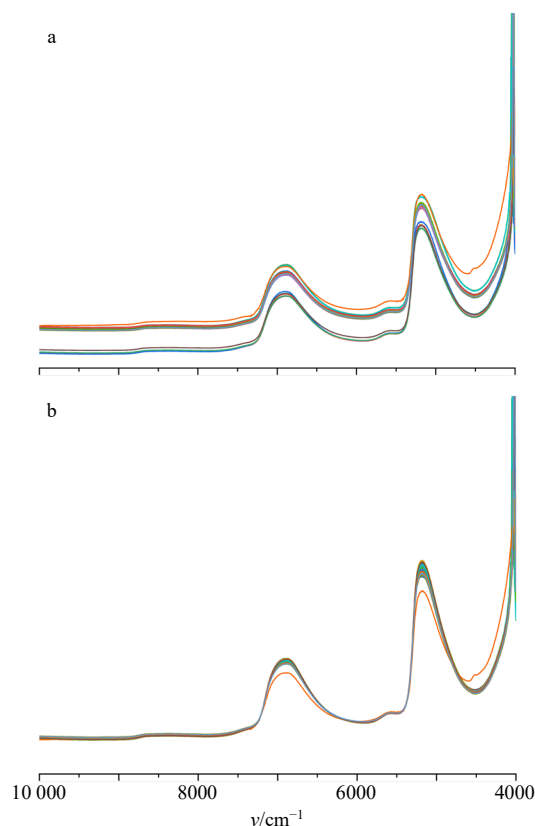


图 1 47 种样品的原始光谱曲线 (a) 和经 SNV 处理后的光谱曲线 (b)

Fig. 1 Original spectral curves (a) and spectral curves after standard normal variate pre-processing (b) of 47 kinds of samples

2.3 NIRS 预处理

由于样本形态、表面特征以及仪器噪声等因素的影响, 采集的光谱信息通常会包含一些冗余信息。为了较少或消除各种非目标因素的干扰, 提高模型的性能, 需要对光谱进行预处理。常用的光谱预处理方法主要包括一阶导数 (first derivative, 1st)、二阶导数 (second derivative, 2nd)、标准正态变量变换 (standard normal variate transformation, SNV)、多元散射校正 (multiplicative signal correction, MSC)、SG (Savitzky-Golay) 平滑等, 其中导数可以消除仪器背景或漂移对信号的影响, SNV 和 MSC 用于消除样品分布不均导致的散射对光谱的干扰, SG 平滑能够降低光谱中随机噪声的影响^[25]。本研究使用主成分分析-判别分析 (principal components analysis-discriminant analysis, PCA-DA) 方法结合 5 种常用的光谱预处理方法建立中药苦与非苦、甜与非甜、酸与非酸、咸与非咸及苦、甜、酸、咸四分类 5 种分类辨识模型, 对比分析不同预处理方法对

模型判别能力的影响,并以留一法交互验证的分类正确率结果选择光谱预处理方法。另外,由于极端波长区域的噪声通常比较显著^[16-18],如图 1-a 所示,在 10 000~8000 cm⁻¹ 波段范围内无明显的特征信息,而在 4500~4000 cm⁻¹ 波段范围内噪声信号比较强烈。因此,为了降低该波段的噪声干扰,将 NIRS 范围缩小至 8000~4500 cm⁻¹ 的波数区域。

PCA-DA 结合不同预处理方法的模型判别结果如表 4 所示,除苦与非苦分类模型中 SNV 与其他几种预处理方法的结果相当,咸与非咸分类模型以不预处理和 SG 平滑的结果较好外,其余均以 SNV 预处理的结果较好,且相比不预处理(87.2%、78.7%、80.9%),SNV 预处理(93.6%、87.2%、87.2%)提高了模型 5%~9% 的准确率,可见经 SNV 预处理能够提升模型的性能。经 SNV 预处理后的 NIRS 图见图 1-b,从图中可知,经过 SNV 预处理后,能够消除或减少散射对光谱的干扰,有助于光谱中特征信息的获取。因此,选取 SNV 作为 5 种分类模型的预处理方法。

2.4 中药四类味觉分类辨识模型的建立

本研究分别采用 PCA-DA、偏最小二乘-判别分析(partial least squares-discriminant analysis, PLS-DA)、*K*-近邻算法(*K*-nearest neighbor, KNN) 3 种常用算法建立中药四类味觉的分类辨识模型,并对 3 种方法建立的模型进行对比分析。PCA-DA 算法是在 PCA 分析的基础上应用 DA,主要利用主成分分析的原理,通过最大化类间方差的比率和最小化类内方差的比率来进一步压缩高维数据,从而探索能够解释数据集主要趋势的变量的组合^[26-27]; PLS-DA 是一种基于 PLS 回归模型的判别分析算法^[28],该算法根据已知样品集的特性建立判别模型,对未知样品进行预测分析,通过建立光谱数据与类别特征之间的回归模型进行判别分析^[29]; KNN 算法是在近邻算法(nearest neighbor, NN)的基础上改进的分类方法,通过查询前 *K* 个与测试数据最相似的训练数据,并根据 *K* 个最近邻数据的特征判断测试数据的类别^[30]。建模过程以 MATLAB 2016b 数据处理软件进行,PCA-DA 的主成分数、PLS-DA 的潜变量数以及 KNN 的 *K* 值均通过留一法交互验证程序选择。

以四分类 PCA-DA 模型为例,当选择主成分数为 14、15、16 时,判别错误率最低;前 6 个主成分可以解释原变量 99% 以上的信息,见图 2。辨识结

表 4 不同预处理方法的 5 种中药味觉分类辨识 PCA-DA 模型的判别结果

Table 4 Discriminant results of five kinds of PCA-DA methods for taste identification with different preprocessing algorithms

辨识模型	预处理方法	主成分数	全样本模型	交互验证
			正判率/%	正判率/%
苦与非苦	不预处理	10	93.6	89.4
	SG 平滑	10	93.6	89.4
	一阶导数	10	89.4	83.0
	二阶导数	14	93.6	89.4
	SNV	15	95.7	89.4
	MSC	15	95.7	89.4
甜与非甜	不预处理	19	95.7	87.2
	SG 平滑	14	95.7	83.0
	一阶导数	19	91.5	72.3
	二阶导数	19	93.6	78.7
	SNV	17	95.7	93.6
	MSC	17	95.7	89.4
酸与非酸	不预处理	17	97.9	78.7
	SG 平滑	15	97.9	80.9
	一阶导数	11	87.2	78.7
	二阶导数	20	91.5	76.6
	SNV	19	97.9	87.2
	MSC	19	97.9	87.2
咸与非咸	不预处理	11	100.0	100.0
	SG 平滑	11	100.0	100.0
	一阶导数	10	95.7	93.6
	二阶导数	7	95.7	89.4
	SNV	14	100.0	97.9
	MSC	7	100.0	97.9
四分类	不预处理	18	95.7	80.9
	SG 平滑	20	100.0	85.1
	一阶导数	15	80.9	59.6
	二阶导数	10	74.5	63.8
	SNV	14	93.6	87.2
	MSC	12	93.6	85.1

果如表 5 所示,在全样本建模时,除咸与非咸分类辨识模型中 PLS-DA 模型(100.0%)与 PCA-DA 模型(100.0%)性能相当外,余下 4 种分类辨识模型中的 PLS-DA 模型(100.0%、100.0%、100.0%、100.0%)性能均优于 PCA-DA 模型(95.7%、95.7%、97.9%、93.6%),但经留一法交互验证后,除苦与

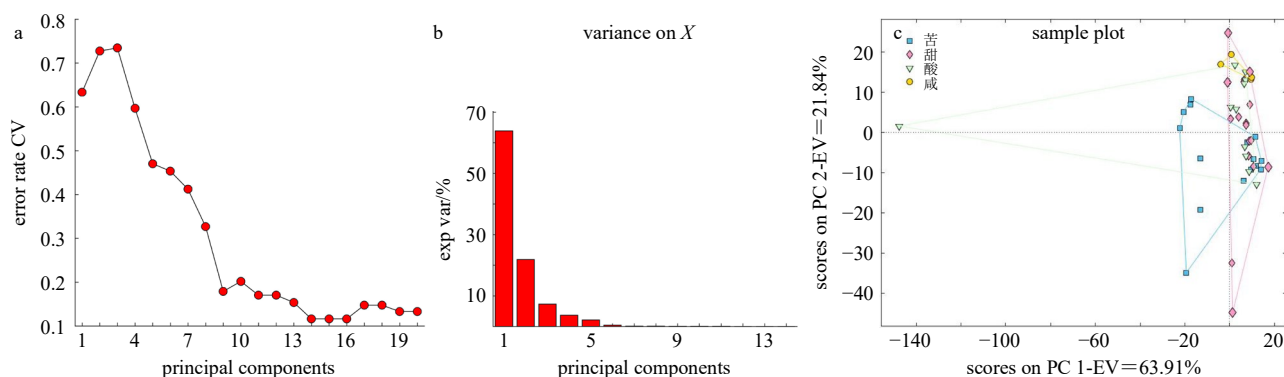


图 2 四分类 PCA-DA 模型主成分优选结果 (a)、主成分贡献百分比 (b) 及得分图 (c)

Fig. 2 Principal component optimization results (a), principal component contribution percentage (b) and principal component scores (c) of four-class identification PCA-DA model

表 5 不同预处理方法与化学计量学方法成对组合的 5 种辨识模型的判别结果

Table 5 Discriminant results of five kinds of identification models paired by different pretreatment methods and chemometrics methods

辨识模型	预处理方法+ 化学计量学方法	全样本模型 正判率/%	交互验证 正判率/%
苦与非苦	SNV+PCA-DA	95.7	89.4
	SNV+PLS-DA	100.0	91.5
	SNV+KNN	85.1	85.1
甜与非甜	SNV+PCA-DA	95.7	93.6
	SNV+PLS-DA	100.0	87.2
	SNV+KNN	72.3	72.3
酸与非酸	SNV+PCA-DA	97.9	87.2
	SNV+PLS-DA	100.0	89.4
	SNV+KNN	74.5	74.5
咸与非咸	SNV+PCA-DA	100.0	97.9
	SNV+PLS-DA	100.0	97.9
	SNV+KNN	100.0	100.0
四分类	SNV+PCA-DA	93.6	87.2
	SNV+PLS-DA	100.0	74.5
	SNV+KNN	66.0	66.0

非苦、酸与非酸分类辨识模型中 PLS-DA 模型 (91.5%、89.4%) 略优于 PCA-DA 模型 (89.4%、87.2%) 外, 其余 3 种分类辨识模型均以 PCA-DA 较优或二者相当, 尤其在四分类辨识时, PCA-DA 模型 (87.2%) 明显优于 PLS-DA 模型 (74.5%), 表明 PCA-DA 模型具有相对更好的分类性能。将 PCA-DA 模型与 KNN 模型相比, 除咸与非咸全样本建模时以 KNN 模型性能略好外, 其余均以 PCA-DA 模型较好。因此, 经综合对比分析, 考虑

选择以 SNV 预处理的 NIRS 建立的 PCA-DA 模型为中药四类味觉分类辨识的最优模型。

2.5 中药四类味觉分类辨识模型的性能评估

模型的性能通常采用准确率进行评价, 但当分类数据不平衡时, 准确率就难以全面反映模型的性能^[31]。因此, 本研究采用混淆矩阵以及敏感性、特异性和精度等指标来更加客观、全面地评价模型的性能。混淆矩阵又叫误差矩阵, 是数据分析中对分类模型预测结果的一种评价方式, 可以直观反映模型对不同种类样品的分类性能好坏, 其每一行表示样本的真实分类结果, 每一列表示样本的实际预测结果^[32]。敏感性、特异性、精度等是常用的模型性能评价指标, 其数据越大 (越接近 1), 模型的性能越好。

$$\text{敏感性} = TP / (TP + FN)$$

$$\text{特异性} = TN / (TN + FP)$$

$$\text{精度} = TP / (TP + FP)$$

TP 是真阳性, FP 代表假阳性, TN 代表真阴性, FN 表示假阴性

如图 3 所示, 以中药味觉四分类辨识模型为例, 图中显示了不同模型判别结果的混淆矩阵, 其中 PCA-DA 对中药的苦、甜、酸、咸 4 种味觉的样品均有较强的识别能力, 除对酸味样品的判别结果较低外, 判别准确率均在 0.85 以上, 且准确率均优于 PLS-DA 和 KNN, 表明 PCA-DA 对不同味觉中药辨识能力相对更佳, 显示了良好的分类性能。

基于 PCA-DA、PLS-DA、KNN 3 种算法的 5 种分类辨识模型性能评价结果如表 6 所示, 在苦与非苦、甜与非甜、酸与非酸、甜与非甜 4 种二分类辨识模型中, PCA-DA 和 PLS-DA 的敏感性、特异性、精度差异较小, 二者基本相当, 然而在四分类

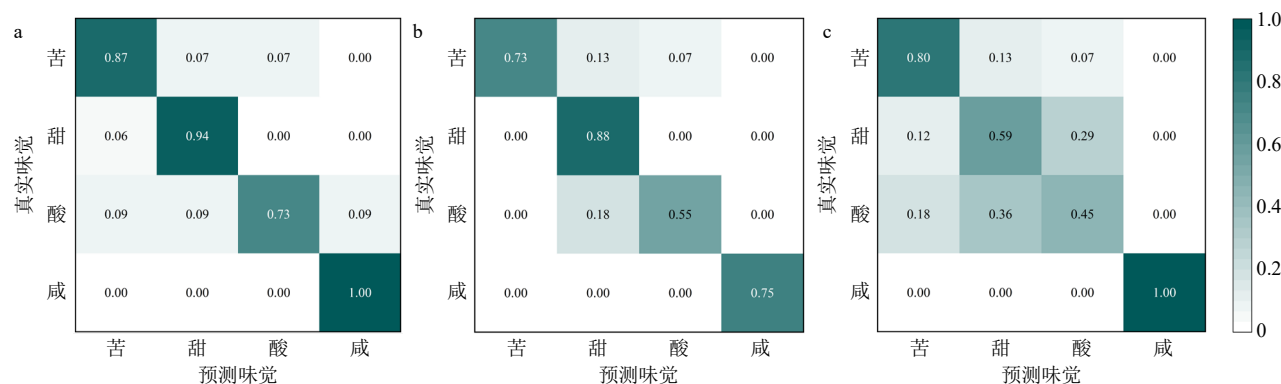


图 3 中药味觉四分类辨识 PCA-DA (a)、PLS-DA (b)、KNN (c) 模型判别结果的混淆矩阵

Fig. 3 Confusion matrix of PCA-DA (a), PLS-DA (b), and KNN (c) models of four-class identification of taste in traditional Chinese medicine

表 6 基于不同化学计量学方法的 5 种辨识模型的敏感性、特异性和精度

Table 6 Sensitivity, specificity and precision of five kinds of identification models by different chemometrics methods

辨识模型	类别	敏感性			特异性			精度		
		PCA-DA	PLS-DA	KNN	PCA-DA	PLS-DA	KNN	PCA-DA	PLS-DA	KNN
苦与非苦	苦	0.80	0.80	0.80	0.94	0.97	0.88	0.86	0.92	0.75
	非苦	0.94	0.97	0.88	0.80	0.80	0.80	0.91	0.91	0.90
甜与非甜	甜	0.88	0.82	0.59	0.97	0.90	0.80	0.94	0.82	0.63
	非甜	0.97	0.90	0.80	0.88	0.82	0.59	0.94	0.90	0.77
酸与非酸	酸	0.64	0.73	0.45	0.94	0.94	0.83	0.78	0.80	0.45
	非酸	0.94	0.94	0.83	0.64	0.73	0.45	0.89	0.92	0.83
咸与非咸	咸	1.00	1.00	1.00	0.98	0.98	1.00	0.80	0.80	1.00
	非咸	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
四分类	苦	0.87	0.73	0.80	0.94	0.81	0.88	0.87	0.65	0.75
	甜	0.94	0.88	0.59	0.93	0.70	0.80	0.89	0.63	0.63
	酸	0.73	0.55	0.45	0.97	0.86	0.83	0.89	0.55	0.45
	咸	1.00	0.75	1.00	0.98	0.86	1.00	0.80	0.33	1.00
平均值		0.89	0.84	0.77	0.91	0.86	0.82	0.88	0.77	0.76
SD		0.11	0.13	0.20	0.10	0.10	0.16	0.06	0.19	0.20

辨识时, PCA-DA 的性能明显优于 PLS-DA, 原因可能为 PLS-DA 用于四分类辨识时, 泽泻、芦根等 7 种样品未分类。

在 5 种分类辨识模型中, PCA-DA 的敏感性、特异性、精度均优于 KNN。结合图 3 和表 5, 相比 PLS-DA 和 KNN 算法, NIRS 技术结合 PCA-DA 算法建立的中药不同味觉分类辨识模型较为精准, 且性能良好。

2.6 NIRS 信息对模型贡献度分析

为了进一步明确 NIRS 信息对模型的贡献度差异, 本研究通过载荷图对模型贡献度较高的光谱范围进行追踪解析。载荷图表示原始变量信息对主成分的综合作用的贡献度, 图中每 1 个点表示 1 个波

数变量, 变量距离原点越远, 其解释的变异信息越大^[33]。由于主成分载荷表示原始变量与主成分间的相关性大小, 与自变量数据矩阵有关, 而不同模型的 X 数据矩阵输入是相同的, 仅标杆信息不同, 因此, 苦与非苦、甜与非甜、酸与非酸、咸与非咸以及四分类辨识 5 种模型的载荷图基本一致。

如图 4 所示, 由于第 1 主成分 (PC1) 能够解释原变量 63.91% 的信息, 而第 2 主成分仅解释原变量 21.84% 的信息, 因此, 以 PC1 为主对贡献较大的光谱信息进行追踪。对模型分类贡献率最大的波数范围是 7390~7274、6700~6534 cm^{-1} , 分别对应于 CH_3 和 CH_2 伸缩振动的组合频、游离 NH 和氢键键合 NH 伸缩振动的一级倍频。

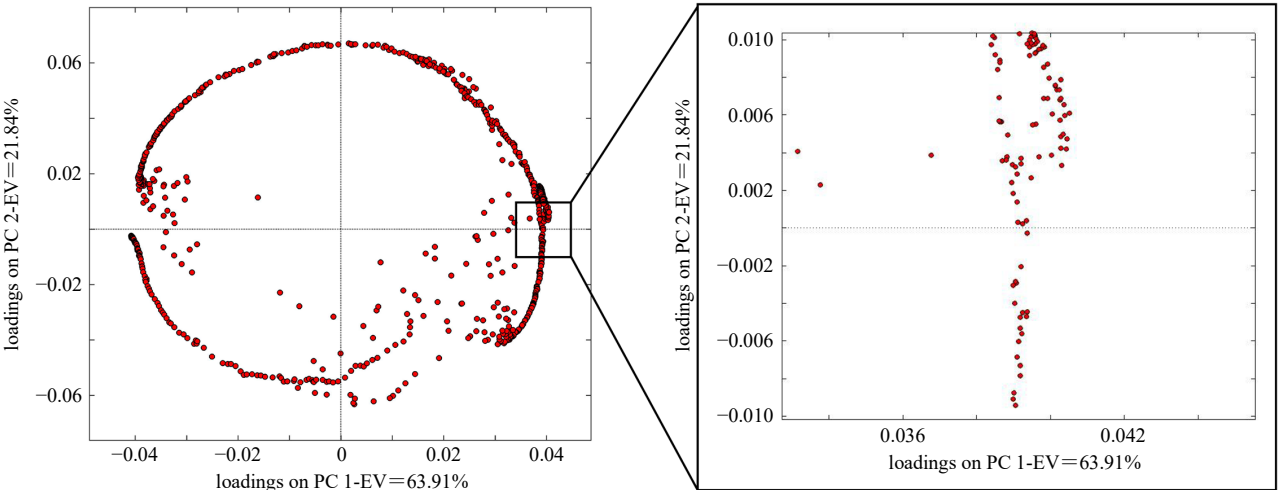


图 4 中药四类味觉分类辨识模型的变量贡献度分析

Fig. 4 Variables contribution analysis of four tastes identification model in traditional Chinese medicine

3 讨论

3.1 研究载体的选择依据

由于中药五味中苦、甜、酸、咸 4 味与辛味的感受机制不同,前者由味觉器官及受体感知和传递,而后者由嗅觉器官及受体感知和传递^[3];现代生理学也认为前者与机体的其它感觉信息一样,有其专门的神经传入通路,而后者则是热感与痛感的综合感觉^[34]。因此,本研究初步选择了味觉感受机制相同的苦、甜、酸、咸味类样品进行分类辨识研究。

本研究以《中国药典》一部饮片性状项下的味觉描述为依据,以饮片“气较微、味强且无兼味”为标准进行研究载体的选择,共筛选了 80 种饮片。然而,部分饮片水煎液的口尝味觉值较低,不符合建模的要求,剔除后样本数量较少,直接建模可能效果不佳。为了扩大样本的数量,将部分味微及含兼味的饮片纳入筛选范围,包括泽泻、炒山楂等 8 种,由于甜、酸、咸味类样本数量相对较少,进一步增补了 12 种味觉明确的常用食品类成分,其中包含葡萄糖、蔗糖、乳酸、L-苹果酸、无水柠檬酸、NaCl 等单体成分。虽然单体类成分与成分复杂的中药水煎液相比具有一定的差异性,但其味觉信息明确(表 1),且在建立 5 种模型辨识中均未错分类(表 7),表明补充的单体类样品不影响模型的辨识能力。因此,最终以 35 种饮片和 12 种食品类成分为研究载体。

3.2 样本错分类原因分析

以样品在 PCA-DA、PLS-DA 和 KNN 3 种算法所建立模型中被错分次数 ≥ 2 为依据,对样品的错分类情况进行统计分析。如表 7 所示,其中青风藤、

表 7 5 种辨识模型中错分类样品统计

Table 7 Statistics of misclassified samples in five kinds of identification models

辨识模型	样品数目	错分类别	样品名称
苦与非苦	4	苦错分为非苦	青风藤、莲子心、槐花
		非苦错分为苦	醋五味子
甜与非甜	3	甜错分为非甜	芦根、饴糖
		非甜错分为甜	莲子心
酸与非酸	5	酸错分为非酸	木瓜、醋五味子、南五味子
		非酸错分为酸	青风藤、食盐
咸与非咸	1	非咸错分为咸	醋五味子
四分类	5	苦错分为酸	青风藤
		苦错分为甜	莲子心
		酸错分为甜	木瓜、南五味子
		酸错分为咸	醋五味子

莲子心、醋五味子 3 种样品多次出现错分类的情况。分析错分类的原因可能为:①虽然总样本数量能够满足建模分析,但各类样本数量相对较少,且不同类别样品数量差异较大,如甜味样品最多,为 17 种,而咸味类样品仅 4 种,可能会影响样品不同类别的特征获取,从而影响模型的辨识能力;②中药饮片成分复杂,其味觉信息丰富,尤其含兼味的饮片,可能会对模型造成干扰。例如,甘草和炒山楂在《中国药典》性状项目味觉描述分别为“气微,味甜而特殊”“味酸、微甜”,虽然其分类正确,但可能会影响模型对特征信息的获取;又如五味子,

其名源自其具有酸、苦、甘、辛、咸 5 种味道,《唐本草》中记述五味子曰:“其果实五味,皮肉甘、酸,核中辛、苦,都有咸味。此则五味俱也”,因此其被错判为苦、甜、咸味类。

3.3 样品未分类原因分析

样品未分类的情况主要存在于四分类的 PLS-DA 判别模型中,这可能与 PLS-DA 的判别机制有关。由于 PLS-DA 是一种线性分析方法,要求变量之间存在高度的相关性,若建模时分类变量值与真实类别值的偏差大于 0.5,PLS-DA 将难以识别该样本^[10]。另外,PLS-DA 多用于处理二分类问题,当用于多分类辨识时,部分样品的特征信息可能不明显,不符合几种类别的任何一种,因此,最终显示为未分类。

3.4 NIRS 的不足分析

本研究基于 NIRS 技术结合不同化学计量学方法初步建立了中药四类味觉的分类辨识模型,表明 NIRS 技术用于中药味觉辨识研究具有良好的可行性。然而,该技术仍有不足之处:①NIRS 技术在获取待测样品的物质结构组成信息的同时,也难免会获取部分无关信息,而光谱预处理及波段选择等方式优化能力相对有限,残留的冗余信息可能会影响模型的性能,降低模型的辨识能力。②由于水在 NIRS 区具有较强的吸收,可能会干扰水溶液中目标成分的响应能力^[35],而本研究的待测样品为水溶液(或水煎液)形式,相比于固体形式样品,NIRS 可能对液体形式样品的响应较低,从而弱化不同样品之间的差异,影响模型的辨识能力。

综上,本研究基于 NIRS 技术建立了中药苦、甜、酸、咸四类味觉的分类辨识模型,建立的模型正判率较高,且模型整体性能较好,能够满足不同味觉中药分类辨识的目的,可为中药五味定性辨识研究提供新的方法参考。在后续的研究中,将尝试进一步提取特征光谱信息,并与电子鼻、电子舌等智能感官信息融合建模分析,同时将获取样品粉末的光谱信息,对比溶液型样品与粉末型样品对模型的贡献度差异,以期进一步提高模型的性能,丰富中药五味药性的辨识方法。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 马文凤,许浚,韩彦琪,等. 仿生技术在中药五味辨识研究中的进展与实践 [J]. 中草药, 2018, 49(5): 993-1001.

[2] 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.

[3] 韩彦琪,许浚,龚苏晓,等. 基于味觉、嗅觉受体分子对接技术的中药性味物质基础研究的路径和方法 [J]. 中草药, 2018, 49(1): 14-19.

[4] 贺鹏,李海英,樊启猛,等. 超分子“印迹模板”理论解析中药五味 [J]. 中草药, 2019, 50(12): 2763-2770.

[5] 王盼盼,凌霄,马静,等. 口尝法在中药药性研究中的作用与改进思考 [J]. 中华中医药杂志, 2021, 36(1): 126-129.

[6] 桂新景,张璐,张璞,等. 基于经典人群口尝评价法的中药饮片水煎液味觉评价 [J]. 医药导报, 2021, 40(11): 1544-1551.

[7] Lin Z Z, Zhang Q, Liu R X, *et al.* Evaluation of the bitterness of traditional Chinese medicines using an E-tongue coupled with a robust partial least squares regression method [J]. *Sensors*, 2016, 16(2): 151.

[8] 刘涛涛,代悦,于淼,等. 基于智能感官分析技术的九蒸九晒大黄饮片气味表征 [J]. 中国实验方剂学杂志, 2022, 28(20): 116-121.

[9] Li X L, Gao X J, Liu R X, *et al.* Optimization and validation of the protocol used to analyze the taste of traditional Chinese medicines using an electronic tongue [J]. *Exp Ther Med*, 2016, 12(5): 2949-2957.

[10] 王艳丽,陈鹏举,桂新景,等. 基于电子舌的天然药物的四类味觉分类辨识研究 [J]. 中华中医药杂志, 2021, 36(1): 423-433.

[11] 刘瑞新,陈鹏举,李学林,等. 人工智能感官: 药学领域的新技术 [J]. 药物分析杂志, 2017, 37(4): 559-567.

[12] Kikut-Ligaj D, Trzecińska-Lorych J. How taste works: Cells, receptors and gustatory perception [J]. *Cell Mol Biol Lett*, 2015, 20(5): 699-716.

[13] 解育静,张家楠,朱冬宁,等. 肉桂中 4 种成分近红外定量分析模型的建立 [J]. 中国实验方剂学杂志, 2020, 26(2): 119-123.

[14] 田燕龙,王毅,王箫,等. 近红外光谱技术在微生物检测中的应用进展 [J]. 光谱学与光谱分析, 2022, 42(1): 9-14.

[15] 尚献召,孟昭平,李德坤,等. 近红外光谱技术在注射用益气复脉(冻干)生产过程分析系统中的应用进展 [J]. 药物评价研究, 2021, 44(11): 2314-2321.

[16] 葛改变,吴心悦,陈姣,等. 计算机虚拟筛选技术在中药苦味物质基础中的研究应用 [J]. 中国新药杂志, 2022, 31(13): 1273-1281.

[17] 张静雅,曹煌,龚苏晓,等. 中药甘味的药性表达及在临证配伍中的应用 [J]. 中草药, 2016, 47(4): 533-539.

[18] 张静雅,曹煌,龚苏晓,等. 中药咸味药性表达及在临

- 证配伍中的应用 [J]. 中草药, 2016, 47(16): 2797-2802.
- [19] Chen Q S, Chen M, Liu Y, *et al.* Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea [J]. *J Food Sci Technol*, 2018, 55(10): 4363-4368.
- [20] Baqueta M R, Coqueiro A, Valderrama P. Brazilian coffee blends: A simple and fast method by near-infrared spectroscopy for the determination of the sensory attributes elicited in professional coffee cupping [J]. *J Food Sci*, 2019, 84(6): 1247-1255.
- [21] 吴晓燕, 侯晓琳, 宿莹, 等. 便携式近红外测定龙胆中水溶性浸出物及龙胆苦苷含量 [J]. 天然产物研究与开发, 2020, 32(8): 1363-1369.
- [22] 杨泉女, 周权驹, 蒙高山, 等. 近红外反射光谱技术测定甜玉米总多糖和淀粉含量的研究 [J]. 吉林农业大学学报, 2019, 41(1): 108-114.
- [23] 梁楷, 闫裕峰, 郎繁繁, 等. 食醋总酸近红外快速检测方法的研究 [J]. 中国酿造, 2020, 39(2): 94-97.
- [24] 张璞, 张耀, 桂新景, 等. 基于经典人群口尝法和电子舌法的中药饮片水煎液苦度叠加规律研究 [J]. 中草药, 2021, 52(3): 653-668.
- [25] 范林宏, 何林, 谭超群, 等. 基于便携式近红外光谱技术快速检测川贝母掺伪问题 [J]. 中国实验方剂学杂志, 2022, 28(3): 131-138.
- [26] Wong K H, Razmovski-Naumovski V, Li K M, *et al.* Differentiating *Puerariae Lobatae Radix* and *Puerariae Thomsonii Radix* using HPTLC coupled with multivariate classification analyses [J]. *J Pharm Biomed Anal*, 2014, 95: 11-19.
- [27] Chua L S, Latiff N A, Lee S Y, *et al.* Flavonoids and phenolic acids from *Labisia pumila* (Kacip Fatimah) [J]. *Food Chem*, 2011, 127(3): 1186-1192.
- [28] Lee L C, Liong C Y, Jemain A A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps [J]. *Analyst*, 2018, 143(15): 3526-3539.
- [29] 崔腾飞, 杨晓玉, 丁佳兴, 等. 近红外高光谱成像技术对鸡蛋种类的鉴别 [J]. 食品工业科技, 2018, 39(5): 13-17.
- [30] 王婧怡, 陈胤佳, 袁野, 等. 面向 K-近邻学习模型的高效数据清洗框架 [J/OL]. 计算机科学与探索. [2023-01-29]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20221020.1348.004.html>.
- [31] 沈国芳, 黄俊航, 许麦成, 等. 基于近红外高光谱成像鉴别不同产地的红参 [J]. 世界中医药, 2021, 16(23): 3419-3423.
- [32] 郑洁, 茹晨雷, 张璐, 等. 基于近红外高光谱成像技术对不同产地苦杏仁和桃仁药材的鉴别 [J]. 中国中药杂志, 2021, 46(10): 2571-2577.
- [33] 陈琳, 唐志书, 宋忠兴, 等. 不同产地白芷药材 9 个呋喃香豆素成分的含量测定及其质量评价 [J]. 中国中药杂志, 2019, 44(14): 3002-3009.
- [34] 吴秀玲, 李大岩, 王丹丹, 等. 现代味觉电生理技术在中药五味定性和定量研究中的作用 [J]. 中国医药指南, 2009, 7(12): 195-196.
- [35] 王明, 刘新. 近红外技术在液态食品成分检测中的应用研究进展 [J]. 激光杂志, 2018, 39(10): 9-13.

[责任编辑 郑礼胜]