

基于多源数据融合的丹参饮片产地溯源研究

傅饶¹, 史亚博¹, 李铭轩¹, 李昱¹, 屈凌芸², 毛春芹¹, 郭志俊³, 陆兔林¹, 赵晓莉¹

(1. 南京中医药大学药学院, 江苏 南京 210023; 2. 华润三九医药股份有限公司, 广东 深圳 518000; 3. 华润三九现代中药制药有限公司, 广东 惠州 516007)

摘要: **目的** 探究不同产地丹参饮片色泽与气味变化, 结合现代机器学习技术实现产地快速区分。 **方法** 采用智能传感技术结合主成分分析(PCA)、判别分析、判别因子分析(DFA)、成分热图、相关分析、机器学习等多种数据分析方法, 对不同产地丹参饮片的颜色和气味进行了定量表征。 **结果** 基于色泽信息建立了丹参饮片产地判别函数。通过对气味信息的分类和筛选, 鉴定出 10 种不同的标志物: 乙醇、二硫化碳、环戊烷、3-甲基呋喃、丙二醇、正壬烷、苯酚、1,5-辛二烯酮、1,8-桉叶醇和葫芦巴内酯。同时发现饮片的颜色和气味之间存在显著的相关性。此外, 基于数据融合的理念, 研究建立了子空间聚类等分类模型, 与单一色泽分析相比, 分类准确率提高到 94.4%。 **结论** 研究证实了智能感官技术在中药产地溯源中的可行性和优越性, 为丹参饮片的产地快速区分与质量控制提供了新的方法和见解。

关键词: 丹参; 产地溯源; 快速识别; 智能感官; 机器学习; 数据融合

中图分类号: R284

文献标志码: A

文章编号: 1672-0482(2024)12-1414-10

DOI: 10.14148/j.issn.1672-0482.2024.1414

引文格式: 傅饶, 史亚博, 李铭轩, 等. 基于多源数据融合的丹参饮片产地溯源研究[J]. 南京中医药大学学报, 2024, 40(12): 1414-1423.

Research on Traceability of *Salvia Miltiorrhiza* Bge. Origin Based on Multi Source Data Fusion

FU Rao¹, SHI Yabo¹, LI Mingxuan¹, LI Yu¹, QU Lingyun², MAO Chunqin¹, GUO Zhijun³, LU Tulin¹, ZHAO Xiaoli¹

(1. School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; 2. China Resources Sanjiu Pharmaceutical Co., Ltd., Shenzhen 518000, China; 3. China Resources Sanjiu Modern Traditional Chinese Medicine Pharmaceutical Co., Ltd., Huizhou 516007, China)

ABSTRACT: OBJECTIVE To explore the color and odor changes of *Salvia miltiorrhiza* Bge. slices from different origins, and combine modern machine learning technology to achieve rapid differentiation of origins. **METHODS** Intelligent sensory technology was used to quantify the color and represent the odor of *Salvia miltiorrhiza* Bge. slices from different geographical origins. Various data analysis methods including principal component analysis (PCA), discriminant analysis, discriminant factor analysis (DFA), component heat maps, correlation analysis, machine learning and so on, were employed to establish a discrimination function for distinguishing the origin of *Salvia miltiorrhiza* Bge. slices based on color data. **RESULTS** Classification and screening of odor information led to the identification of 10 differential markers: ethanol, carbon disulfide, cyclopentane, 3-methylfuran, propylene glycol, nonane, phenol, 1,5-octadienone, 1,8-cineole, and sotolon. It was also found that there was a significant correlation between the color and odor of the slices. Furthermore, based on the concept of data fusion, the study established classification models such as subspace clustering, and compared to single-color discriminant analysis, the classification accuracy was improved to 94.4%. **CONCLUSION** The feasibility and superiority of intelligent sensory technology in classifying the geographical origin of TCM is confirmed, providing new methods and insights for quality control of *Salvia miltiorrhiza* Bge. slices.

KEYWORDS: *Salvia Miltiorrhiza* Bge.; origin traceability; rapid recognition; intelligent sensory; machine learning; data fusion

丹参为唇形科植物丹参 *Salvia miltiorrhiza* Bge. 的干燥根及根茎, 性微寒, 味苦, 归心、肝经, 为活血调经的常用药^[1], 具有多种药理作用, 包括抗凝血、抗炎、抗氧化、抗纤维化和抗肿瘤等作用^[2-5]。丹参有着悠久的药用历史, 在临床上得到了广泛的

应用^[6]。然而, 目前市场上可买到的丹参饮片的质量存在着显著差异^[7]。丹参饮片的质量受到如地理、气候和品种等因素的影响, 使产地成为了一个重要的决定因素^[8]。丹参饮片在我国多个地区生产, 包括山东、山西、河南、四川和云南。山东以种植药

收稿日期: 2024-03-26

基金项目: 国家重点研发计划(2023YFC3504200); 江苏省研究生培养创新工程研究生科研与实践创新计划(KYCX24_2259)

第一作者: 傅饶, 男, 硕士研究生, E-mail: 2361678033@qq.com

通信作者: 陆兔林, 男, 教授, 主要从事中药炮制研究, E-mail: ltl2021@njucm.edu.cn;

赵晓莉, 女, 教授, 主要从事中药炮制研究, E-mail: leah_zhao@njucm.edu.cn

材而闻名,尤其是山东临沂,分布着大量优质丹参药材种植基地。《本草经疏》记载:“北方产者佳,俗名逐马”^[9],而四川中江丹参以其优质而闻名。现代研究证实,不同产地的丹参饮片中化学成分的类型和数量存在显著差异^[10]。因此,质量评估在确保中药资源的安全性和有效性方面发挥着至关重要的作用。老药师传统“辨状论质”主要从颜色、气味等角度入手来评判中药饮片最佳品质,但其评价多依靠主观经验,缺乏一定的客观性与一致性。根据《中国药典》2020 年版记载^[1],丹参的表皮呈棕红色,气微,其颜色和气味是重要的特征指标。研究表明,丹参粉末的颜色与其活性成分(隐丹参酮、丹参酮 I、丹参酮 II A、丹酚酸 B)的含量之间存在相关性^[11],颜色可以部分反映样品的质量。然而,目前对丹参气味的研究主要集中在其挥发油和花朵上^[12-13],对丹参饮片的研究有限,这表明迫切需要一种快速准确的方法来模拟人类视觉、嗅觉,并将其以数据形式量化,以进行丹参饮片的质量识别。

分光光度计是一种光学测量仪器,它用电子眼取代人类视觉,将人眼对于红绿蓝的感知转变为电子眼对于亮度(L^*)、红绿轴(a^*)和黄蓝轴(b^*)的感知,大大减少了人眼的主观性判断误差,消除不同标准引起的不一致。分光光度计在选择中药材的最佳采收期、识别中药材的不同产地和评估中药材质量等方面发挥着至关重要的作用^[14-15]。超快速气相电子鼻是一种利用传感器技术模拟人类嗅觉系统的设备,采用顶空进样法,可对中药气味进行快速、无损的分析,被广泛用于各种药材的鉴定和质量评价^[16-17]。

本文基于分光光度计和快速气相电子鼻 2 种现代智能传感技术,研究了不同产地的丹参饮片,以量化颜色信息和定性描述气味,为丹参饮片的质量控制研究提供支持。

1 材料

1.1 仪器

CM-5 色度计(Konica Minolta,日本);Heracles Neo 超快速气相电子鼻,配备 PALRSI 自动顶空进样器、MXT-5 非极性柱、MXT-1701 中极性柱(Alpha MOS,法国);GH-380N 高纯氢发生器、GA-380N 低噪声气泵(北京中兴汇力科技有限公司有限公司);HL-250 高速多功能粉碎机(上海赛耐机械有限公司);FA1104N 电子天平(上海静海仪器有限公司)。

1.2 样品和试剂

共采集河南、四川、山东、山西和云南省 54 批丹参饮片(表 1)。所有样品均经南京中医药大学陈建伟教授鉴定为丹参干燥根和根茎;系列烷烃的混合物($nC_6 \sim nC_{16}$,批号:A10142930)购自美国 RESTEK 有限公司。

表 1 丹参饮片来源

Table 1 Source of <i>Salvia miltiorrhiza</i> Bge. slices	
产地	编号
河南省	HN1~6
四川省	SC1~6
山东省	SD1~15
山东省临沂市	SDLY1~15
山西省	SX1~6
云南省	YN1~6

2 方法和结果

2.1 基于电子眼的颜色分析

2.1.1 样品制备 将来自不同产地的丹参饮片样品分别破碎,过 3 号筛,然后置于 $-20\text{ }^{\circ}\text{C}$ 冰箱中保存。

2.1.2 测量条件 使用 CIE D65/ 10° 在透射模式下测定丹参粉末的颜色参数:光源/孔径 $\phi 3\text{ mm}$ 、波长范围 $360 \sim 740\text{ nm}$ 、波长间隔 10 nm 、脉冲氙灯观测者条件,并选择镜面成分排除(SCE)模式(30 nm)。仪器零点校正(放置透明片、黑管)和用户零点校正(放透明膜、白板)。该仪器使用透明片和黑管进行零校准,以及使用透明膜和白板进行用户校准^[18]。

2.1.3 样品检测 取适量的粉末样品并均匀地散布在试验皿上,测量并记录 L^* 、 a^* 和 b^* 。 L^* 的值越大表示颜色越亮、越浅, a^* 的值越大表示颜色越红, b^* 的值越大表示黄色越多。

2.1.4 方法学考察 为了保证颜色分析方法的稳定性和可靠性,对所建立的方法进行了方法学研究。样品粉末(SD1)按照“2.1.3”中的方法平行测量 6 次,并分别检查该方法的精密性和重复性。此外,按照“2.1.3”项中概述的检测方法,在室温下以 0、2、4、6、8、10、12、24 h 的时间间隔对粉末样品(SD1)进行测量,以检查该方法的稳定性。将测得的 L^* 、 a^* 和 b^* 值的相对标准偏差(RSD)值用作评价指标。分析结果表明,在精密度、重复性和稳定性实验中, L^* 、 a^* 和 b^* 的 RSD 值均小于 3%,表明该方法具有

良好的精密度和重复性,以及样品在室温下 24 h 内具有稳定性。

2.1.5 数据采集与处理 所使用的颜色测量软件是 Spectra Magic NX,利用 CIE $L^* a^* b^*$ 颜色空间系统。数据分析采用 SPSS 26.0 软件和 SIMCA 软件。

表 2 不同产地丹参饮片颜色测定结果($\bar{x}\pm s, n=3$)

Table 2 Results of color determination of *Salvia miltiorrhiza* Bge. slices from different geographical origins ($\bar{x}\pm s, n=3$)

色度值	HN	SC	SD	SDLY	SX	YN
L^*	59.54±0.16	49.63±0.25	54.79±3.39	56.82±4.38	58.03±0.20	52.49±0.42
a^*	8.51±0.98	9.21±0.09	9.76±0.54	9.85±1.23	10.70±0.02	12.93±0.26
b^*	17.02±0.13	16.77±0.37	18.62±0.26	17.83±0.65	18.78±0.15	20.61±0.31

2.1.7 秩和检验 Kruskal-Wallis H 秩和检验通常用于比较连续变量的多个独立样本。这是一种非参数方法,用于测试 2 个以上样本是否来自相同的概率分布^[19]。根据不同的产地,将收集到的丹参饮片分为 6 类。Kruskal-Wallis H 秩和检验的结果表明, L^* 、 a^* 和 b^* 的 P 值都小于 0.01,也就是说,它们都拒绝了原始假设。这表明样品粉末的色度值可以根据其来源显著地分为 6 类(表 3)。

表 3 Kruskal-Wallis H 秩和检验
Table 3 Kruskal-Wallis H rank sum test

组别	χ^2	原假设	显著性
L^*	24.446	在样本类别中,相同分布的 L^*	$P<0.01$
a^*	32.017	在样本类别中,相同分布的 a^*	$P<0.01$
b^*	41.843	在样本类别中,相同分布的 b^*	$P<0.01$

2.1.8 主成分分析(PCA) PCA 是一种常用的数据降维方法,通过线性变换将高维数据转换为低维数据,从而最大限度地减少数据的信息损失^[20]。根据 PCA 得分图,可以直观地观察不同产地的有效分类性能,如图 1 所示。主成分 1(PC1)和主成分 2(PC2)的累积贡献率达到 92%,表明它们能够区分与不同产地相关的气味信息。

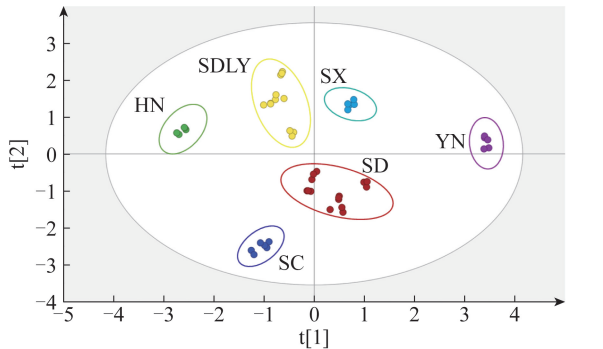


图 1 不同产地丹参饮片色度值的 PCA 得分

Fig. 1 PCA scores of chromaticity values of

Salvia miltiorrhiza Bge. slices from different origins

2.1.6 色度值测定结果 按照“2.1.3”项中的检测方法,使用 CM-5 分光光度计测量来自不同产地的丹参粉末的色度值。每个样品被测量 3 次,并获得平均值。结果如表 2 所示。

2.1.9 判别分析 在建立模型之前,对数据进行了多重共线性分析。由于存在更精确或高度相关的关系,线性回归模型中的解释变量有时会扭曲模型估计,或使其难以准确估计,因此有必要评估自变量之间的相互关系,通常使用容差或方差膨胀因子(VIF,容差的倒数)作为评估指标。VIF 值越高表示共线问题越明显,模型越不准确。通常,VIF 值超过 10 表明某个自变量与其他自变量之间存在严重的共线关系^[21]。使用 SPSS 26 软件对粉末色度值 L^* 、 a^* 和 b^* 进行分析。结果表明,VIF 值分别为 1.091、2.308 和 2.202,这表明粉末色度值 L^* 、 a^* 和 b^* 之间不存在共线。因此,所构建的线性模型不会导致失真。

Fisher 判别法是一种基于方差分析原理构建判别函数,有效区分多个群体的线性判别法。通过人工建立分类标准,可以将新样本分配到各自的类别^[22]。在这项研究中,54 个样本根据其来源分为 6 组。选择每个样本的粉末色度值(L^* 、 a^* 和 b^*)作为自变量,并使用 54 个样本的数据作为训练样本。利用 Fisher 判别方法,建立了数学判别模型,快速识别丹参饮片的来源,并进行了交叉验证。将验证数据中的 L^* 、 a^* 和 b^* 三个变量分别带入判别式(1)~(6)中计算,各产地判别式中结果最大值即为产地判别结果。该模型实现了 90.74% 的高分类准确率,成功地识别了丹参饮片的产地。结果汇总在表 4 中。

$$Y_{HN} = 17.904L^* + 37.871a^* + 150.823b^* - 1979.765 \quad (1)$$

$$Y_{SC} = 16.442L^* + 37.354a^* + 145.071b^* - 1798.210 \quad (2)$$

$$Y_{SD} = 18.123L^* + 40.433a^* + 160.744b^* - 2192.562 \quad (3)$$

$$Y_{SDLY} = 18.100L^* + 40.597a^* + 156.038b^* - 2107.068 \quad (4)$$

$$Y_{SX} = 18.853L^* + 43.097a^* + 163.739b^* - 2316.996 \quad (5)$$

$$Y_{YN} = 19.235L^* + 47.678a^* + 175.412b^* - 2622.105 \quad (6)$$

表 4 不同产地的丹参饮片颜色判别交叉验证结果

Table 4 Cross-validation results of distinguishing the colors of *Salvia miltiorrhiza* Bge. slices from different geographical origins

项目	类别	预测值						
		HN	SC	SD	SDLY	SX	YN	总计
原始	计数	HN	6	0	0	0	0	6
		SC	0	6	0	0	0	6
		SD	0	0	14	1	0	15
		SDLY	0	0	0	15	0	15
		SX	0	0	0	0	6	6
		YN	0	0	0	0	6	6
	判别率/%	HN	100	0	0	0	0	100
		SC	0	100	0	0	0	100
		SD	0	0	93.3	6.7	0	100
		SDLY	0	0	0	100	0	100
		SX	0	0	0	0	100	100
		YN	0	0	0	0	100	100
交叉验证	计数	HN	5	0	0	1	0	6
		SC	0	6	0	0	0	6
		SD	0	0	13	2	0	15
		SDLY	0	0	0	15	0	15
		SX	0	0	0	1	5	6
		YN	0	1	0	0	5	6
	判别率/%	HN	83.3	0	0	16.7	0	100
		SC	0	100	0	0	0	100
		SD	0	0	86.7	13.3	0	100
		SDLY	0	0	0	100	0	100
		SX	0	0	0	13.3	83.3	100
		YN	0	13.3	0	0	83.3	100

2.1.10 不同产地的丹参饮片色度值参考范围 参考范围采用百分位法确定,不同产地丹参饮片的双边 90%参考值范围如表 5 所示。根据参考值范围与颜色判别分析相结合,可以对不同产地的丹参饮片进行数字化区分和表征^[23]。

表 5 不同产地丹参饮片双边 90%色度值的参考范围
Table 5 Reference range for 90% chromaticity values of *Salvia miltiorrhiza* Bge. slices from different origins on both sides

样本	<i>L</i> [*]	<i>a</i> [*]	<i>b</i> [*]
HN	59.32~59.75	8.42~8.63	16.90~17.19
SC	49.29~49.98	9.13~9.35	16.29~17.19
SD	50.50~60.13	9.29~10.76	18.27~19.04
SDLY	51.25~61.89	8.07~11.62	17.07~18.95
SX	57.71~58.31	10.67~10.74	18.60~18.96
YN	52.02~52.96	12.59~13.13	20.27~20.99

2.2 基于电子鼻的气味分析

2.2.1 样品制备 将来自不同产地的丹参样品粉碎,过 3 号筛,然后置于-20 ℃冰箱中保存。

2.2.2 超快速气相电子鼻测量条件的建立 在对快速气相电子鼻进行文献考证后,选择称样量、孵育温度、孵化时间和注射体积 4 个参数进行单因素研究^[24-26]。

2.2.2.1 称样量考察 分别精密称定 SD1 样品 0.5、1、1.5、2、2.5 g,在孵化温度 80 ℃,孵化时间 20 min,进样体积 5 000 μL 条件下进样,每个水平平行 3 次,在称样量达到 2 g 时各色谱峰强度达到饱和。

2.2.2.2 孵化温度考察 确定称样量后考察孵化温度,设定孵化温度为 60、65、70、75、80 min,在称样量 2 g,孵化时间 20 min℃,进样体积 5 000 μL 条件下进样,每个水平平行 3 次,在孵化时间为 20 min 时各色谱峰强度达到饱和。

2.2.2.3 孵化时间考察 确定孵化温度后考察孵化时间,设定孵化时间为 15、20、25、30、35 min,在称样量 2 g,孵化温度 75 ℃,进样体积 5 000 μL 条件下进样,每个水平平行 3 次,在孵化时间为 25 min

时各色谱峰强度达到饱和。

2.2.2.4 进样量考察 确定孵化时间后考察进样体积,设定进样体积为 1 000、2 000、3 000、4 000、5 000 μL ,在称样量 2 g、孵化温度 75 $^{\circ}\text{C}$ 、孵化时间 25 min 条件下进样,每个水平平行 3 次,在进样体积 5 000 μL 时各色谱峰强度达到饱和。

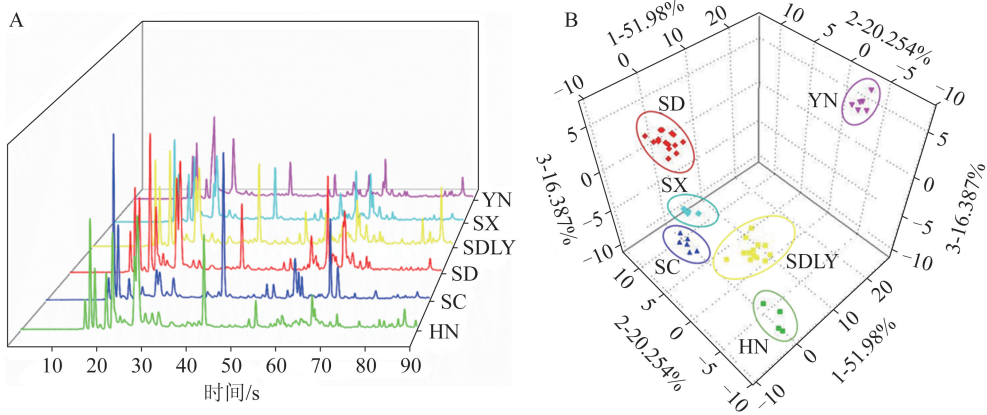
综上,单因素考察后最终确定测量条件为:称样量 2 g、孵化温度 75 $^{\circ}\text{C}$ 、孵化时间 25 min、进样体积 5 000 μL 。

2.2.3 方法学考查 为了保证气味分析方法稳定性和可靠性,对所建立的方法进行了方法学研究。样品粉末(SD1)按照“2.2.2”项中的方法平行测量 6 次,并分别检查了该方法的精密度和重复性。此外,使用“2.2.2”项中概述的检测方法,在室温下以 0、2、4、6、8、10、12、24 h 的时间间隔对 SD1 进行测量,以检查该方法的稳定性。分别使用 MXT-1701 与 MXT-5 这 2 个色谱柱下 6 个共有主成分峰的相对保留时间和相对峰面积 *RSD* 值作为评价指标。分析结果表明,在精密度、重复性和稳定性实验中,各峰的相对保留时间和相对峰面积的 *RSD* 值分别小于 0.2% 和 5.0%,表明该方法具有良好的精密度

和重复性,样品在室温下 24 h 内具有良好的稳定性。

2.2.4 气味信息的提取与分析 根据“2.2.2”项中的条件,使用 Heracles NEO 超快速气相电子鼻检测不同产地丹参的气味色谱峰。每个样品平行测试 3 次,在扣除空白组参考后获得气味色谱峰。结果如图 2A 所示,可以观察到来自不同产地的丹参饮片的气味分布存在显著差异。

2.2.4.1 判别因子分析(DFA) 为了进一步阐明来自不同产地的丹参饮片之间的嗅觉差异,使用 Alpha Soft 17.0 对原始气味信息进行 DFA。DFA 是一种多元统计分析方法,主要用于探讨不同群体之间的差异和分类问题^[27]。它是在 PCA 的基础上发展起来的,PCA 将原始变量转换为一组新的不相关变量(称为判别因子)。然后利用这些判别因子对样本进行分类。结果如图 2B 所示,判别因子 1、2 和 3 占总方差的 88.621%,这意味着良好的分类性能。DFA 判别模型可以进一步证明不同产地的丹参饮片之间存在显著的气味差异,并且可以有效地进行区分。



注:A. 气味色谱瀑布图;B. 气味 DFA 得分
图 2 不同产地丹参饮片气味信息

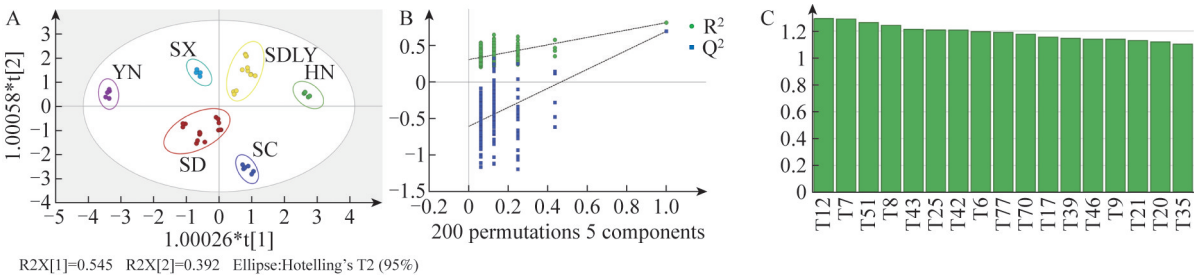
Fig. 2 Odor information of *Salvia miltiorrhiza* Bge. slices from different origins

2.2.4.2 气味差异标记辨识 校准每批样品的保留时间,并在 SIMCA-P 14.1 软件中使用其峰面积作为变量,建立不同产地丹参样品气味信息的偏最小二乘判别分析(PLS-DA)模型。PLS-DA 是一种基于偏最小二乘回归(PLSR)和判别分析(DA)的多元统计分析方法^[28]。它主要用于探索多组变量之间的差异和分类问题,具有抗多重共线性和增强信噪比等优点。结果如图 3 所示。图 3A 的得分散点

图表明,模型中的所有数据都在 95% 的置信区间内,聚类效果良好,差异明显,表明不同来源的丹参饮片之间存在一定的气味差异。将分类 *Y* 矩阵变量设置为随机排列 200 次进行排列检验,结果如图 3B 所示。可以看出, R^2 拟合线在 *Y* 轴上的截距小于 0.3,表明了模型的可靠性。此外, Q^2 拟合线在 *Y* 轴上的截距小于 0.05,表明模型中不存在过拟合现象^[29]。

可变投影重要性(VIP)值是指机器学习模型中每个特征对模型预测结果的重要性水平。在 PLS-DA 模型中,VIP 值大于 1.0 表示与平均水平相比,相应的变量对整个模型的贡献更大。它在影响样本分类方面具有统计学意义,可以作为一种差异标记^[30]。使用 SIMCA-P 14.1 软件的 VIP 分析程序,

结果如图 3C 所示,有 17 个色谱峰的 VIP 值超过 1。进一步分析来自不同产地的这些峰值的方差,结果表明,10 个峰值的 *P* 值小于 0.05,存在显著差异。因此可以确定,10 个色谱峰对应成分可作为不同产地的丹参饮片的气味差异标记。



注: A. PLS-DA 得分散点图; B. 置换检验图; C. VIP 值图

图 3 不同产地丹参饮片 PLS-DA 模型

Fig. 3 PLS-DA model of *Salvia miltiorrhiza* Bge. slices from different origins

2.2.4.3 气味成分的识别和差异成分的确证 通过计算不同产地的丹参饮片差异色谱峰的保留指数,并将其与 Archembase 数据库进行比较,确定了每个来源的丹参饮片的差异标记。共鉴定出 10 种不同成分,即乙醇、二硫化碳、环戊烷、3-甲基呋喃、丙二醇、正壬烷、苯酚、1,5-辛二烯酮、1,8-桉叶醇和葫芦巴内酯,结果如表 6 所示。绘制了这些差异成分的相对含量,如图 4A 所示。不同产地的气味

成分存在显著差异。例如,部分产地显示几乎不存在葫芦巴内酯和苯酚成分,而主要常见组分如环戊烷、二硫化碳、3-甲基呋喃和 1,5-辛二烯酮的含量随样品产地不同而变化显著。道地产区(四川、山东、山西)的 1,8-桉叶醇含量相对较高。以上结果表明,基于超快速气相电子鼻的丹参饮片分类结果较好,该方法可以有效地区分不同的产地。

表 6 不同产地丹参饮片中的差异成分及感官描述信息

Table 6 Differential marker components and sensory description information in *Salvia miltiorrhiza* Bge. slices from different origins

编号	分子式	保留指数		化合物	相关指数	气味信息
		MXT-5	MXT-1701			
1	C ₂ H ₆ O	442	563	乙醇	87.75	酒精性、刺鼻、甜味
2	CS ₂	553	629	二硫化碳	89.60	芳香、燃烧、硫磺、甜味
3	C ₅ H ₁₀	587	586	环戊烷	94.84	温和、甜美
4	C ₅ H ₆ O	653	858	3-甲基呋喃	91.75	轻质和芳香族
5	C ₃ H ₈ O ₂	737	923	丙二醇	90.42	酒精性、焦糖化、无气味
6	C ₉ H ₂₀	904	908	正壬烷	98.19	烷烃、杂醇、汽油
7	C ₆ H ₆ O	963	1 200	苯酚	88.58	刺激性、塑料、橡胶、甜味
8	C ₈ H ₁₂ O	995	1 084	1,5-辛二烯酮	93.18	泥土味、发霉
9	C ₁₀ H ₁₈ O	1 040	1 084	1,8-桉叶醇	99.31	樟脑、草药、甘草、薄荷糖、甜
10	C ₆ H ₈ O ₃	1 112	1 351	葫芦巴内酯	94.03	燃烧、焦糖化、咖啡、辛辣、甜

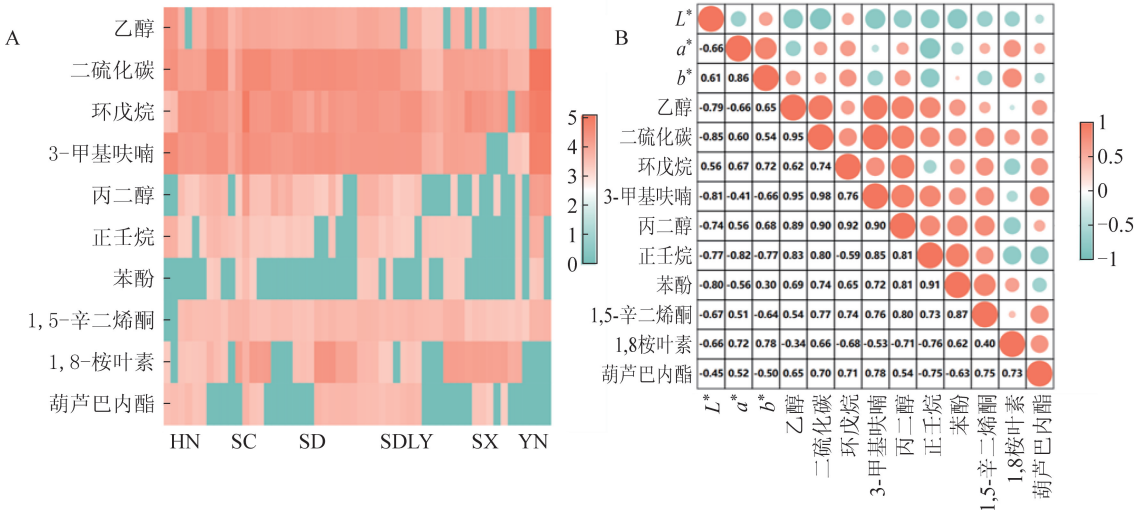
2.3 颜色与气味的相关性分析

为了探索丹参饮片的颜色和气味之间的潜在关系,将不同来源的丹参饮片的比色值(*L*^{*}、*a*^{*}、*b*^{*})和气味成分数据导入 SPSS 26.0 软件中。由于数据

已分组且为有序形式,因此采用了 Spearman 相关性分析^[31]。结果如图 4B 所示,红色表示正相关,绿色表示负相关,圆圈大小表示相关强度。通常认为,当相关系数的绝对值大于 0.5 时,这两个因素之间存

在显著的相关性。在丹参饮片的色度值中,光度的代表性指标 L^* ,与气味成分含量呈负相关的是 1,8-桉叶醇、1,5-辛二烯酮、苯酚、正壬烷、丙二醇、3-甲基呋喃、二硫化碳、乙醇,与之呈正相关的是环戊烷。红绿代表指数 a^* ,与气味成分呈负相关的是苯酚、正壬烷、乙醇,而与气味成分呈正相关的是葫芦

巴内酯、1,8-桉叶醇、1,5-辛二烯酮、丙二醇、环戊烷、二硫化碳。黄蓝代表指数 b^* ,与气味成分呈负相关的有 1,5-辛二烯酮、正壬烷、3-甲基呋喃、葫芦巴内酯,与气味组分呈正相关的有 1,8-桉叶醇、丙二醇、环戊烷、二硫化碳、乙醇。上述成分的 P 值均小于 0.05,具有统计学意义。



注:A. 不同产地差异成分含量热图;B. 色泽-气味相关性热图

图 4 不同产地丹参饮片气味差异成分热图

Fig. 4 Hot map of odor differential components in *Salvia miltiorrhiza* Bge. slices from different origins

2.4 颜色和气味的数据融合分析

为了观察丹参饮片在其来源背景下的整体颜色和气味特征,将电子眼和电子鼻测定的所有原始数据导入 MATLAB 软件,在保持数据完整性的同时建立分类模型。决策树、判别分析、支持向量机、K 最近邻、子空间判别和神经网络等算法常用于数据的分类挖掘中,在计算机强大的计算能力的支持下,以数学函数或模型的形式处理给定的数据,并输出分类结果^[31-33]。

本研究建立了包括子空间判别在内的 6 个分类算法模型,通过内部交叉验证对算法准确性进行了 10 次评估,并通过受试者工作特征(ROC)曲线对算法性能进行了评估。分类模型的分类结果、混淆矩阵和 ROC 曲线如图 5 所示。混淆矩阵用于比较分类结果和实际测得值,可以把分类结果的精度显示在一个混淆矩阵里面。AUC 介于 0.5~1,较大的 AUC 代表了较好的性能,是用来度量分类模型好坏的标准。几种模型的混淆矩阵表明,子空间判别算法在 origin 分类中的准确率最高,达到 94.4%,支持向量机、K-最近邻等算法的准确率也达到了 90%左右。这表明不同产地的丹参饮片在颜色和气味上存

在显著差异,可以通过算法模型准确识别和区分。图 5 的 ROC 曲线显示了 6 个模型的性能,其中 6 个原点为正值。无论哪一个原点为正值,子空间判别的 AUC 都接近 1.00。结果表明,子空间判别模型对不同产地的丹参饮片进行分类是非常可靠的。

在各种算法中,都存在一个常见的问题,即河南和四川的样本被算法错误地识别为山东临沂。这主要是因为山东临沂的丹参栽培规模较大,涵盖了不同的栽培方法。与四川类似,山东临沂主要利用种子进行繁殖,而湖南等地主要以扦插形式进行繁殖。该算法对样品包装中只标注山东省的样本和准确标注山东省临沂市的样本进行了区分,这也表明准确标注山东临沂的丹参饮片与山东其他地域的丹参饮片间存在显著差异。

该研究基于数据融合的概念,在数据层面整合颜色和气味信息,追踪丹参饮片的产地,取得了良好的效果。与仅基于颜色信息的分类相比,融合过程后的准确率显著提高(90.74%→94.4%),从混淆矩阵与 ROC 曲线也可看出,数据融合后的模型在保持良好模型性能的同时,减少了对山东来源丹参饮片的误判。这也为丹参饮片的产地追溯提供了一种快

速准确的新方法。

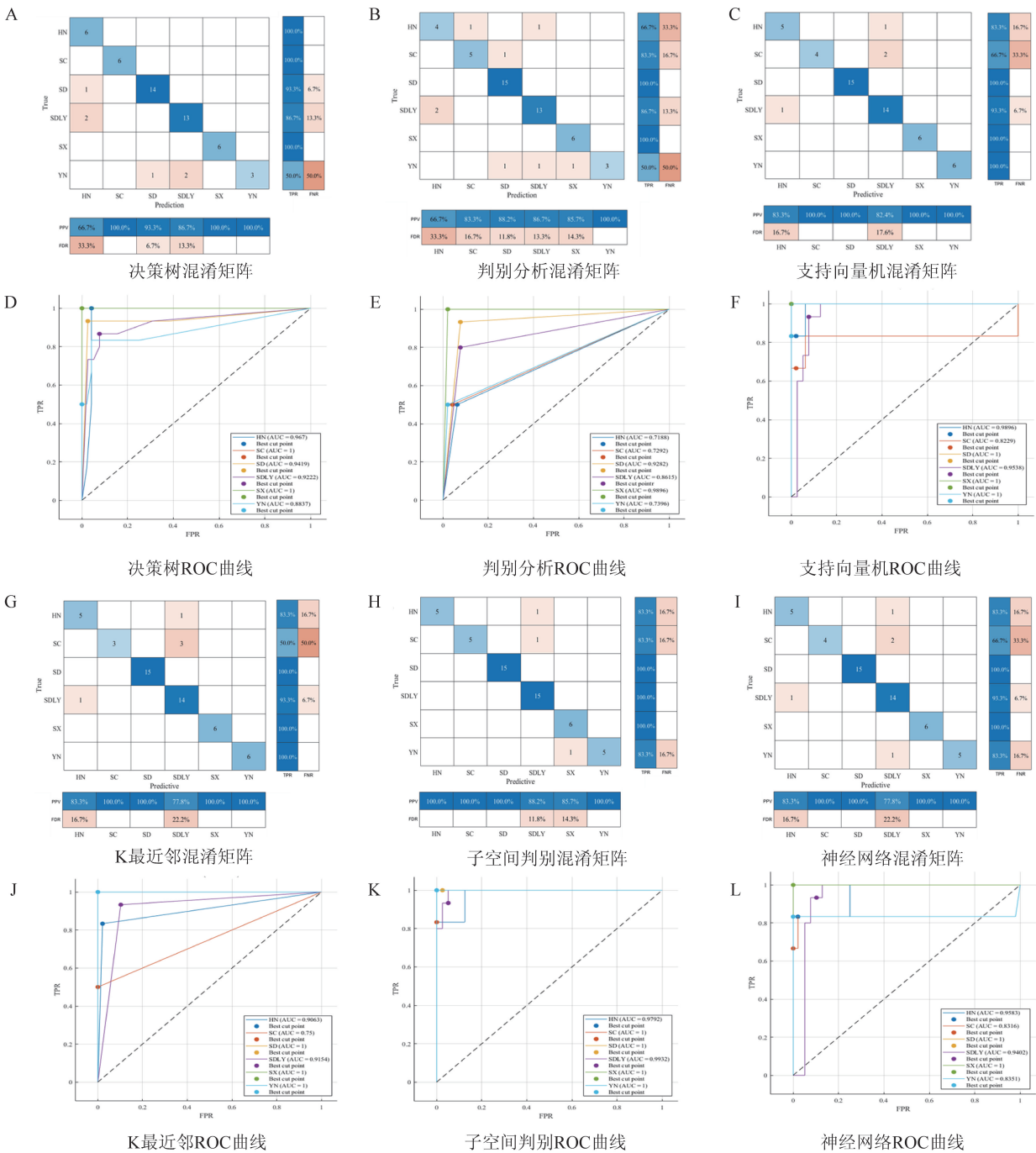


图 5 不同分类模型的混淆矩阵和 ROC 曲线图

Fig. 5 Confusion matrix and ROC curve of different classification models

3 讨论

颜色分析中的典型判别分析可以有效地区分不同产地的丹参饮片,除了丹参种植大省山东。临沂、淄博、济南、潍坊等地的大规模种植造成山东来源的丹参质量的显著差异^[34],这可能解释了为什么单一的电子眼颜色分析在区分山东来源丹参饮片时,将样品错误地划分为相邻其他产地。由于种植、土壤

环境和栽培技术等因素^[35],丹参饮片的其他来源表现出显著的颜色变化。典型判别分析与 PCA 模型分析的结果相似,可以获得更好的分类结果。在气味分析中,基于整体气味信息的 DFA 模型揭示各个产地间气味成分组内差异小、组间差异大,对于不同产地丹参饮片分类效果较为显著。经过 PLS-DA 建模和 VIP 值>1 的筛选,共鉴定出 10 种气味成分,

其中环戊烷、二硫化碳和 3-甲基呋喃的感官描述与《中国药典》2020 版中丹参的气味描述一致,表明电子鼻识别气味组分具有一定的准确性。

通过中药形态学鉴定进行质量评价是一种长期使用总结出来的质量评价方法。《仁寿堂药镜》记载^[36],优质丹参的特点是周皮呈红色,木质部呈紫色,突出了其外观颜色作为质量指标的重要性。本研究证实了丹参饮片的气味在质量控制中同样有着独特的作用,不同产地的丹参饮片存在明确的气味差异标记。近年来,智能传感技术的快速发展,以其快速、准确、方便的特点,在中药质量控制中得到了广泛的应用。本研究探索了使用电子眼和电子鼻在不同产地丹参饮片上进行快速识别,研究了丹参饮片的外部颜色和内部气味成分之间的相关性,从不同产地丹参样品中鉴定出 10 种气味成分作为丹参饮片的潜在差异标记,发现颜色和气味差异标记之间存在显著相关性。在此基础上,本研究建立并比较了 6 种分类模型在数据融合信息中的适用性。结果表明,与传统的分类算法相比,子空间判别等模型在保持较高模型性能的同时,显著提高了分类精度,这为进一步的质量控制提供了依据。

参考文献:

- [1] 国家药典委员会. 中华人民共和国药典:一部[S]. 北京:中国医药科技出版社,2020.
National Pharmacopoeia Commission. Pharmacopoeia of People's Republic of China; I [S]. Beijing:China Medical Science Press, 2020.
- [2] MAIONE F, FEO V D, CAIAZZO E, et al. Tanshinone II A, a major component of *Salvia miltiorrhiza* Bunge, inhibits platelet activation via Erk-2 signaling pathway[J]. J Ethnopharmacol, 2014, 155(2): 1236-1242.
- [3] GAO H W, HUANG L T, DING F, et al. Simultaneous purification of dihydrotanshinone, tanshinone I, cryptotanshinone, and tanshinone IIA from *Salvia miltiorrhiza* and their anti-inflammatory activities investigation[J]. Sci Rep, 2018, 8(1): 8460.
- [4] ZHANG J Q, JIN Q H, DENG Y P, et al. New depsides from the roots of *Salvia miltiorrhiza* and their radical-scavenging capacity and protective effects against H₂O₂-induced H9c2 cells[J]. Fitoterapia, 2017, 121: 46-52.
- [5] CAO Y, HUANG B, GAO C Q. *Salvia miltiorrhiza* extract dihydrotanshinone induces apoptosis and inhibits proliferation of glioma cells[J]. Bosn J Basic Med Sci, 2017, 17(3): 235-240.
- [6] 屠燕, 邱实, 董志颖, 等. 丹参炮制方法的历史沿革研究[J]. 中国中药杂志, 2021, 46(18): 4683-4688.
TU Y, QIU S, DONG Z Y, et al. Historical evolution of *Salviae Miltiorrhizae* Radix et Rhizoma processing methods[J]. China J Chin Mater Med, 2021, 46(18): 4683-4688.
- [7] 张萍, 郭晓晗, 金红宇, 等. 2022 年全国中药材及饮片质量分析概况[J]. 中国现代中药, 2023, 25(10): 2045-2054.
ZHANG P, GUO X H, JIN H Y, et al. Quality overview of Chinese medicinal materials and decoction pieces in 2022[J]. Mod Chin Med, 2023, 25(10): 2045-2054.
- [8] 陈谷. 不同产地丹参的形态学及 ISSR、SSR 分子标记分析

- [D]. 雅安: 四川农业大学, 2022.
CHEN G. Morphological and ISSR, SSR molecular markers analysis of *Salvia miltiorrhiza* from different habitats[D]. Ya'an: Sichuan Agricultural University, 2022.
- [9] 缪希雍. 神农本草经疏[M]. 太原: 山西科学技术出版社, 2013.
MIAO X Y. Annotations on *Shennong's materia medica* [D]. Taiyuan: Shanxi Science and Technology Press, 2013.
- [10] 王巧, 于永杰, 付海燕, 等. 基于多指标含量测定结合化学计量学的不同产地丹参品质差异分析[J]. 分析测试学报, 2023, 42(4): 389-401.
WANG Q, YU Y J, FU H Y, et al. Quality difference analysis of *salviae miltiorrhizae* Radix et rhizoma from different origins based on multi-index content determination combined with chemometrics [J]. J Instrum Anal, 2023, 42(4): 389-401.
- [11] 刘伟, 周冰谦, 王晓, 等. 丹参药材粉末色泽与有效成分含量的相关性[J]. 中华中医药杂志, 2019, 34(4): 1466-1470.
LIU W, ZHOU B Q, WANG X, et al. Correlational between the powder of color traits and active ingredient content in *Salvia miltiorrhiza* Bge[J]. China J Tradit Chin Med Pharm, 2019, 34(4): 1466-1470.
- [12] LIN R, MU F, LI Y, et al. *Salvia miltiorrhiza* and the volatile of *Dalbergia odorifera* attenuate chronic myocardial ischemia injury in a pig model: A metabonomic approach for the mechanism study [J]. Oxid Med Cell Longev, 2021, 2021: 8840896.
- [13] LIN C C, XING P Y, JIN H, et al. Loss of anthocyanidin synthase gene is associated with white flowers of *Salvia miltiorrhiza* Bge. f. *alba*, a natural variant of *S. miltiorrhiza* [J]. Planta, 2022, 256(1): 15.
- [14] 付小梅, 杨超, 吴志瑰, 等. 不同采收期梔子中 8 个有效成分含量动态变化与颜色的相关性研究[J]. 中国中药杂志, 2020, 45(13): 3191-3202.
FU X M, YANG C, WU Z G, et al. Correlation between color and content of eight components of *Gardeniae Fructus* at different harvest time [J]. China J Chin Mater Med, 2020, 45(13): 3191-3202.
- [15] CAI J Y, LIANG M, WEN Y D, et al. Analysis of tobacco color and location features using visible-near infrared hyperspectral data [J]. Spectrosc Spectr Anal, 2014, 34(10): 2758-2763.
- [16] 李铭轩, 秦宇雯, 李昱, 等. 基于 Heracles Neo 超快速气相电子鼻对不同基原郁金饮片的快速识别及差异标志物研究[J]. 中国中药杂志, 2023, 48(6): 1518-1525.
LI M X, QIN Y W, LI Y, et al. Rapid identification and differential markers of *Curcumae Radix* decoction pieces of different sources based on Heracles Neo ultra-fast gas phase electronic nose [J]. China J Chin Mater Med, 2023, 48(6): 1518-1525.
- [17] 陈鹏, 肖晓燕, 梅茜, 等. 基于仿生技术对苡苡仁麸炒过程中色泽气味变化研究[J]. 中草药, 2022, 53(14): 4285-4297.
CHEN P, XIAO X Y, MEI Q, et al. Study on color and odor change of *Coicis Semen* bran stir-fry based on bionic technology [J]. Chin Tradit Herb Drugs, 2022, 53(14): 4285-4297.
- [18] 贾奥蒙, 秦宇雯, 李昱, 等. 猪心血丹参性状特征与内在品质相关性研究[J]. 中草药, 2023, 54(9): 2765-2775.
JIA A M, QIN Y W, LI Y, et al. Research on correlation between morphological characteristics and internal quality of *Salvia miltiorrhiza* processed by porcine cardiac blood [J]. Chin Tradit Herb Drugs, 2023, 54(9): 2765-2775.
- [19] SHERWANI R A K, SHAKEEL H, AWAN W B, et al. Analysis of COVID-19 data using neutrosophic Kruskal Wallis H test [J]. BMC Med Res Methodol, 2021, 21(1): 215.
- [20] BEATTIE J R, ESMONDE-WHITE F W L. Exploration of principal component analysis: Deriving principal component analysis visually using spectra [J]. Appl Spectrosc, 2021, 75(4): 361-375.

- [21] KIM J H. Multicollinearity and misleading statistical results [J]. *Korean J Anesthesiol*, 2019, 72(6): 558–569.
- [22] CASASENT D, CHEN X W. Radial basis function neural networks for nonlinear Fisher discrimination and Neyman–Pearson classification[J]. *Neural Netw*, 2003, 16(5/6): 529–535.
- [23] 史亚博, 傅饶, 李铭轩, 等. 基于电子感官技术的醋香附饮片不同炮制程度判别及炮制终点量化研究[J]. *中国中药杂志*, 2023, 48(18): 5003–5013.
SHI Y B, FU R, LI M X, et al. Discrimination of different processing degrees and quantitative study of processing end point of vinegar-processing *Cyperi Rhizoma* pieces based on electronic sensory technology[J]. *China J Chin Mater Med*, 2023, 48(18): 5003–5013.
- [24] ZHANG K W, WANG J, FAN X C, et al. Discrimination between raw and ginger juice processed *Magnoliae officinalis* cortex based on HPLC and Heracles NEO ultra-fast gas phase electronic nose [J]. *Phytochem Anal*, 2022, 33(5): 722–734.
- [25] OTSUKA Y, NOJIRI M, KUSUMOTO N, et al. Production of flavorful alcohols from woods and possible applications for wood brews and liquors[J]. *RSC Adv*, 2020, 10(65): 39753–39762.
- [26] KOSTYRA E, KR6L K, KNYSAK D, et al. Characteristics of volatile compounds and sensory properties of mixed organic juices based on kiwiberry fruits[J]. *Appl Sci*, 2021, 11(2): 529.
- [27] NEZU T, HOSOMI N, YOSHIMURA K, et al. Predictors of stroke outcome extracted from multivariate linear discriminant analysis or neural network analysis[J]. *J Atheroscler Thromb*, 2022, 29(1): 99–110.
- [28] DEV K, HO C J H, BI R Z, et al. Machine learning assisted handheld confocal Raman micro-spectroscopy for identification of clinically relevant atopic eczema biomarkers[J]. *Sensors*, 2022, 22(13): 4674.
- [29] SUN B K, WANG R Y, LI B, et al. Rapid identification of polypeptide from carbapenem-resistant and susceptible *Escherichia coli* via orbitrap-MS and pattern recognition analyses [J]. *Chem Biodivers*, 2022, 19(9): e202200118.
- [30] WENG M F, YOU S M, LUO J W, et al. Antibacterial mechanism of polysaccharides from the leaves of *Lindera aggregata* (Sims) Kosterm. by metabolomics based on HPLC/MS[J]. *Int J Biol Macromol*, 2022, 221: 303–313.
- [31] SCHOBER P, BOER C, SCHWARTE L A. Correlation coefficients: Appropriate use and interpretation [J]. *Anesth Analg*, 2018, 126(5): 1763–1768.
- [32] OGUZ C, YAGANOGLU M. Detection of COVID-19 using deep learning techniques and classification methods [J]. *Inf Process Manag*, 2022, 59(5): 103025.
- [33] NIE F P, WANG Z, TIAN L, et al. Subspace sparse discriminative feature selection [J]. *IEEE Trans Cybern*, 2022, 52(6): 4221–4233.
- [34] 张玉. 山东地区丹参药材品质区划研究[D]. 济南: 山东中医药大学, 2019.
ZHANG Y. Study on the quality regionalization of Danshen medicinal materials in Shandong Province[D]. Jinan: Shandong University of Traditional Chinese Medicine, 2019.
- [35] 赵恒强, 张敏敏, 赵志国, 等. 多元素分析结合多元统计分析用于丹参产地溯源研究[J]. *天然产物研究与开发*, 2017, 29(10): 1752–1758.
ZHAO H Q, ZHANG M M, ZHAO Z G, et al. Determination of geographical origin of *Salvia miltiorrhiza* by multi-element analysis combined with multivariate statistical analysis[J]. *Nat Prod Res Dev*, 2017, 29(10): 1752–1758.
- [36] 郑二阳. 仁寿堂药镜[M]. 北京: 中国中医药出版社, 2024.
ZHEN E Y. *Renshou Tang medicine mirror* [M]. Beijing: China Traditional Chinese Medicine Press, 2024.

(编辑: 杨巍敏)