

doi: 10.12452/j.fxcxb.25120902

# 基于气相色谱-离子迁移谱及机器学习的不同产地香附饮片的快速鉴别

卢腾飞<sup>1,2</sup>, 刘青<sup>1</sup>, 董红敬<sup>2\*</sup>, 李佳<sup>1</sup>, 李明坤<sup>1,2</sup>, 王晓<sup>2</sup>, 张艳艳<sup>3\*</sup>

(1. 山东中医药大学药学院, 山东 济南 250300; 2. 齐鲁工业大学(山东省科学院)山东省分析测试中心天然产物分离提取共性技术创新与应用山东省工程研究中心, 山东 济南 250014; 3. 山东中医药大学附属医院风湿免疫科, 山东 济南 250011)

**摘要:** 采用气相色谱-离子迁移谱(GC-IMS)分析不同产地香附饮片中的挥发性成分, 通过偏最小二乘判别分析(PLS-DA)筛选具有显著性差异的特征成分; 基于特征性成分, 采用线性核支持向量机(SVM-L)等9种机器学习算法构建不同产地香附的区分模型。结果从5个省份的香附饮片中共鉴定出59种挥发性成分, 主要包括醇类、酯类、醛类、酮类、不饱和烃类等; 通过PLS-DA筛选出1, 4-二甲基苯、3-甲基-1-戊醇、乙酰乙酸乙酯等10种特征成分; 9种机器学习模型均对香附产地表现出良好的预测性能(Accuracy=1)。该研究为不同产地香附饮片的快速鉴别提供了一种简单、快速的方法, 同时可为香附饮片质量评价体系的构建提供参考依据。

**关键词:** 气相色谱-离子迁移谱; 香附; 挥发性成分; 产地; 机器学习

**中图分类号:** O657.7; R284.1 **文献标识码:** A **文章编号:** 1004-4957(2026)05-0951-09

## Rapid Identification of *Cyperi Rhizoma* Decoction Pieces from Different Geographical Origins Based on GC-IMS Combined with Machine Learning

LU Teng-fei<sup>1,2</sup>, LIU Qing<sup>1</sup>, DONG Hong-jing<sup>2\*</sup>, LI Jia<sup>1</sup>, LI Ming-kun<sup>1,2</sup>, WANG Xiao<sup>2</sup>, ZHANG Yan-yan<sup>3\*</sup>

(1. College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250300, China; 2. Shandong Engineering Research Center for Innovation and Application of General Technology for Separation of Natural Products, Shandong Analysis and Test Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China; 3. Department of Rheumatology and Immunology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250011, China)

**Abstract:** Gas chromatography-ion mobility spectrometry (GC-IMS) was employed to analyze the volatile components of *Cyperi Rhizoma* from different origins. Partial least squares-discriminant analysis (PLS-DA) was applied to screen differential characteristic components. Based on these characteristic components, 9 machine learning algorithms such as SVM-L were constructed to develop rapid discrimination models for *Cyperi Rhizoma* from different origins. Fifty-nine volatile compounds were identified in *Cyperi Rhizoma* decoction pieces from five provinces, including alcohols, esters, aldehydes, ketones, and unsaturated hydrocarbons. Ten characteristic components were screened by PLS-DA analysis, such as 1, 4-dimethylbenzene, 3-methyl-1-pentanol, and ethyl acetoacetate. Furthermore, nine machine learning models all demonstrated excellent predictive performance (Accuracy=1), indicating good potential for practical application. This study provides a simple and rapid method for discrimination and identification of *Cyperi Rhizoma* from different geographical origins, while also offering a reference for establishing a quality evaluation system for *Cyperi Rhizoma*.

**Key words:** gas chromatography-ion mobility spectrometry (GC-IMS); *Cyperi Rhizoma*; volatile components; geographical origins; machine learning

收稿日期: 2025-12-09; 修回日期: 2026-01-14

基金项目: 齐鲁工业大学(山东省科学院)科教产融合试点工程重大创新类项目(2025ZDZX07); 国家现代农业产业技术体系项目(CARS-21)

\* 通讯作者: 董红敬, 博士, 副研究员, 研究方向: 中药药效物质及质量控制, E-mail: donghongjing\_2006@163.com  
张艳艳, 博士, 主任医师, 研究方向: 风湿免疫, E-mail: 405784910@qq.com

网络首发日期: 2026-03-19

香附(*Cyperi Rhizoma*, CR)为莎草科莎草属植物莎草(*Cyperus rotundus* L.)的干燥根茎,其性平、味辛、微苦,具有疏肝解郁、理气宽中、调经止痛的功效,常用于治疗肝气郁结、胸胁胀痛、脘腹胀满等症。现代研究表明,香附主要含有挥发油、黄酮、生物碱、酚类、萜类等成分,其中,挥发油是香附的主要质量评价指标<sup>[1]</sup>。目前,2025版《中国药典》以挥发油总含量作为香附的质量评价指标<sup>[2]</sup>,难以实现不同产地的有效鉴别。香附在我国山东、河南、浙江、广东等地均有产出,不同产地香附的化学成分种类和含量存在显著差异<sup>[3-4]</sup>。因此亟需构建一种高效、快速、准确的香附产地鉴别方法。

气相色谱-离子迁移谱(GC-IMS)是一种将气相色谱和离子迁移谱相结合的检测技术,具有分离度优良、检测灵敏度高、样品前处理简单等特点,已被广泛用于中药品种鉴定、产地溯源等方面<sup>[5-6]</sup>。机器学习算法是一种常用的大数据分析技术,主要根据训练得到的监督类型进行分类,相比于传统多元数据分析方法,其预测能力高、稳定性好<sup>[7-8]</sup>。GC-IMS结合机器学习已广泛用于西洋参<sup>[9]</sup>、生姜<sup>[10]</sup>、天麻<sup>[11]</sup>等饮片的精准鉴别。然而,目前尚无GC-IMS结合机器学习用于香附产地快速鉴别的相关报道。

本研究采用GC-IMS对不同产地香附的挥发性成分(VOCs)及相对含量进行分析,采用偏最小二乘判别分析(PLS-DA)筛选特征成分;随后,基于特征成分结合机器学习算法构建不同产地香附饮片的快速区分模型。本研究旨在为香附饮片产地鉴别提供一种简便、快速的方法,并为不同产地香附饮片质量评价体系的建立提供参考。

## 1 实验部分

### 1.1 样品与仪器

共收集来自广东省广州市(GD)、河南省滑县(HE)、湖南省会同县(HU)、江苏省太仓市(JS)和山东省肥城市(SD)5个产地的香附生品,晒干后切片,每个省份收集25批。GAS Flavour Spec气相离子迁移谱(GC-IMS, G. A. S. 公司, 德国);梅特勒ME204万分之一电子天平(梅特勒-托利多仪器(上海)有限公司, 中国)。

### 1.2 样品制备与挥发性成分分析

精密称取0.5 g香附饮片粉末于20 mL顶空进样瓶。采用GC-IMS对香附挥发性成分进行分析,分析条件为:分析时间为30 min;色谱柱为MXT-WAX(30 m×0.53 mm, 1.0 μm);柱温为60 °C;载气/漂移气为N<sub>2</sub>;离子迁移谱(IMS)温度为60 °C。自动顶空进样单元条件:进样量为100 μL;孵化时间为15 min;孵化温度为80 °C;进样针温度为85 °C;孵化转速为500 r·min<sup>-1</sup>。气相色谱条件:载气流速采用梯度模式(0~2 min, 2 mL·min<sup>-1</sup>; 2~5 min, 从2 mL·min<sup>-1</sup>升至10 mL·min<sup>-1</sup>; 5~25 min, 从10 mL·min<sup>-1</sup>升至100 mL·min<sup>-1</sup>; 25~30 min, 100 mL·min<sup>-1</sup>)。

### 1.3 GC-IMS数据分析

采用GC-IMS自带的Laboratory Analytical Viewer(LAV)分析软件,提取特征峰信号强度,将三维数据降维为二维数据矩阵。分析前,需对该矩阵进行归一化处理。选择指示性化合物去除冗余信息,提高模型的泛化能力。将每个样本中所有特征峰的信号强度之和标准化为1,消除样本间的量纲差异,突出特征峰的相对比例关系。结合仪器内置的NIST数据库和IMS数据库,对香附中的挥发性化合物进行定性分析,并构建不同产地香附饮片的挥发性成分表达矩阵。采用R-studio软件对数据进行主成分分析(PCA)和偏最小二乘判别分析。

### 1.4 机器学习分析

线性核支持向量机(SVM-L)、多项式核支持向量机(SVM-P)、径向核支持向量机(SVM-R)、Sigmoid核支持向量机(SVM-S)、朴素贝叶斯(NB)、K近邻(KNN)、随机森林(RF)、极端梯度提升(XGB)以及线性判别分析(LDA)均通过R语言实现。机器学习算法的详细信息如表1所示。机器学习算法可预测4种结果,分别为真阳性(TP)、真阴性(TN)、假阳性(FP)和假阴性(FN)。其中,TP指模型将阳性样本正确预测为阳性的情况;TN指模型将阴性样本准确预测为阴性的情况;FP指模型将阴性样本错误预测为阳性的情况;FN指模型将阳性样本错误预测为阴性的情况。采用精确度(Precision)、召回率(Recall)、F1分数(F1 score)和准确率(Accuracy)对模型的整体性能进行综合评估<sup>[8]</sup>。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

精确度：阳性结果占所有预测结果的比例；召回率：真阳性样本占实际阳性样本总数的比例；F1分数综合了精确度与召回率；准确率：用于评估预测结果的整体正确性。

## 2 结果与讨论

### 2.1 不同产地香附饮片的挥发性成分分析

2.1.1 GC-IMS分析 图1为香附挥发性成分的GC-IMS三维图，X轴代表离子漂移时间，Y轴代表气相色谱保留时间。每个点均代表1种挥发性成分，其颜色越红，表明离子流信号越强，成分的相对含量越高；颜色越蓝，表明离子流信号强度越弱，成分的相对含量越低<sup>[12]</sup>。在相对漂移时间1.0~2.0 ms、保留时间200~1 800 s范围内，信号点密集，说明香附中挥发性成分主要分布在该区域。

采用GC×IMS Library Search软件内置NIST和IMS数据库，结合挥发性成分的漂移时间和保留时间，对挥发性成分进行定性分析。从香附中共鉴定出59种挥发性成分(表2)，主要包括15种酯类、14种醇类、7种醛类、7种酮类、7种不饱和烃类及9种其他类型挥发性成分。

酯类是香附中含量最高的一类成分，不同产地香附饮片中酯类成分的含量存在差异。例如， $\gamma$ -丁内酯(VOC8)(图2A)在GD组中含量最高，该化合物具有抗癌、抗炎、抗氧化、免疫抑制和神经保护作用<sup>[13]</sup>。乙酸香茅酯(VOC12)(图2C)在HE组中含量最高，该成分是一种萜烯短链脂肪酸酯类化合物，具有良好的镇痛作用<sup>[14]</sup>。醇类是一种具有花香和果香特征的挥发性成分<sup>[15]</sup>，是香附中含量较多的第二类挥发性成分，在GD组和HU组中含量最高。其中， $\alpha$ -松油醇(VOC4)具有抗氧化、抗炎、抗菌作用<sup>[16]</sup>，其在GD组香附饮片中含量较高(图2B)；HU组则主要富含(E)-2-己烯-1-醇(VOC19)(图2D)、1-己醇(VOC28)(图2E)、3-戊醇(VOC50)(图2F)等醇类成分。酮类是香附饮片中含量较高的第三类挥发性成分，不同产地香附饮片中酮类成分存在差异。例如，具有抗菌、抗病毒及抗寄生虫等活性的香芹酮(VOC6)(图2G)<sup>[17]</sup>在GD组香附饮片中含量最高。中药材中醛类物质的产生途径主要有不饱和脂肪酸氧化、微生物发酵<sup>[18]</sup>。醛类是香附饮片中含量较高的第四类挥发性成分，主要在GD组和HU组中含量最高。有些醛类具有抗菌活性，例如，糠醛(VOC25)具有良好的葡萄球菌和枯草芽孢杆菌抑制活性(图2H)<sup>[19]</sup>，其在GD组中含量最高；苯甲醛(VOC15)具有明显的抗真菌作用(图2I)<sup>[20]</sup>，其在HU组中含量最高。

表1 机器学习模型的参数

Table 1 Parameters of the machine learning model

Model	R package	R function	Parameter
SVM-L	e1071	svm()	kernel="linear", cost=1
SVM-P	e1071	svm()	kernel="polynomial", degree=4, coef0=1
SVM-R	e1071	svm()	kernel="radial", gamma=0.5
SVM-S	e1071	svm()	kernel="sigmoid", gamma=0.1, coef0=0.5
KNN	knn	knn()	k=2
NB	e1071	naivebayes()	laplace=0
XGB	xgboost	xgboost()	booster="gblinear", eta=0.5, num_class=9, objective="multi: softmax", nround=50
RF	randomforest	randomforest()	ntree=30
LDA	mass	lda()	tol = 1.0×10 <sup>-4</sup>

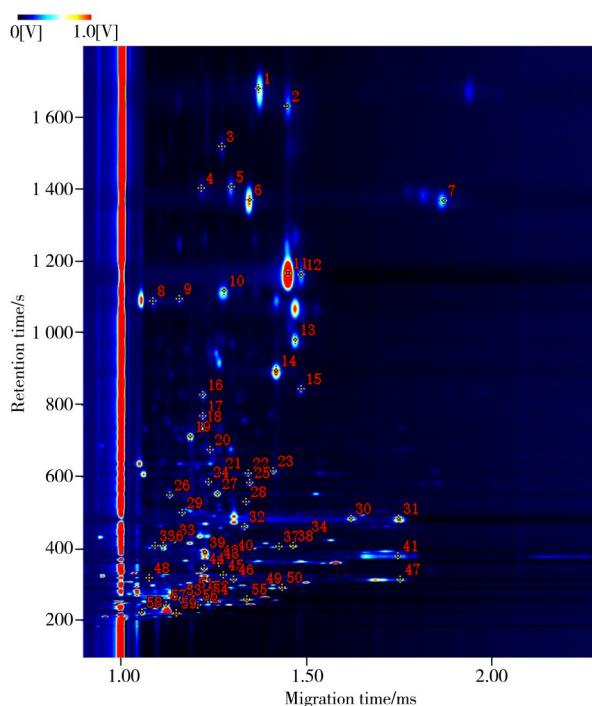


图1 香附挥发性成分的GC-IMS三维谱图

Fig. 1 3D GC-IMS spectrum of VOCs in *Cyperi Rhizoma*

表2 不同产地香附的挥发性有机化合物信息

Table 2 Information of VOCs in *Cyperi Rhizoma* from different origins

ID	Compound	CAS	Molecular formula	Relative molecular mass	Retention index	Retention time/s	Migration time/ms	Type
VOC1	(E, E)-2, 4-壬二烯醛	C5910872	C <sub>9</sub> H <sub>14</sub> O	138.2	1 777.3	1 679.141	1.373 5	醛类
VOC2	(3E, 6E)- $\alpha$ -法尼烯	C502614	C <sub>15</sub> H <sub>24</sub>	204.4	1 766.1	1 629.565	1.450 6	不饱和烃类
VOC3	水杨酸乙酯	C118616	C <sub>9</sub> H <sub>10</sub> O <sub>3</sub>	166.2	1 739.5	1 517.478	1.274 7	酯类
VOC4	$\alpha$ -松油醇	C98555	C <sub>10</sub> H <sub>18</sub> O	154.3	1 709.7	1 401.081	1.219 4	醇类
VOC5	$\alpha$ -松油醇-D	C98555	C <sub>10</sub> H <sub>18</sub> O	154.3	1 710.9	1 405.392	1.300 4	醇类
VOC6	香芹酮	C99490	C <sub>10</sub> H <sub>14</sub> O	150.2	1 701.0	1 368.748	1.349 2	酮类
VOC7	香芹酮-D	C99490	C <sub>10</sub> H <sub>14</sub> O	150.2	1 700.5	1 366.593	1.871 0	酮类
VOC8	$\gamma$ -丁内酯	C96480	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	86.1	1 615.6	1 088.532	1.089 3	酯类
VOC9	(E)-2-辛烯-1-醇	C18409171	C <sub>8</sub> H <sub>16</sub> O	128.2	1 617.8	1 094.999	1.160 7	醇类
VOC10	2-苯乙醛	C122781	C <sub>8</sub> H <sub>8</sub> O	120.2	1 625.1	1 116.554	1.279 8	醛类
VOC11	(Z)-3-壬烯-1-醇	C10340235	C <sub>9</sub> H <sub>18</sub> O	142.2	1 641.3	1 166.130	1.452 4	醇类
VOC12	乙酸香茅酯	C150845	C <sub>12</sub> H <sub>22</sub> O <sub>2</sub>	198.3	1 639.9	1 161.819	1.488 1	酯类
VOC13	1-辛醇	C111875	C <sub>8</sub> H <sub>18</sub> O	130.2	1 576.7	980.757	1.472 2	醇类
VOC14	(E)-2-壬烯醛	C18829566	C <sub>9</sub> H <sub>16</sub> O	140.2	1 542.4	894.536	1.420 6	醛类
VOC15	苯甲醛	C100527	C <sub>7</sub> H <sub>6</sub> O	106.1	1 521.1	844.960	1.488 1	醛类
VOC16	芳樟基乙酸酯	C115957	C <sub>12</sub> H <sub>20</sub> O <sub>2</sub>	196.3	1 513.1	827.014	1.223 3	酯类
VOC17	1, 2-丙二醇二乙酸酯	C623847	C <sub>7</sub> H <sub>12</sub> O <sub>4</sub>	160.2	1 485.9	768.897	1.223 3	酯类
VOC18	$\beta$ -香茅醛	C106230	C <sub>10</sub> H <sub>18</sub> O	154.3	1 471.8	740.348	1.220 1	醛类
VOC19	(E)-2-己烯-1-醇	C928950	C <sub>6</sub> H <sub>12</sub> O	100.2	1 456.6	710.779	1.190 2	醇类
VOC20	2, 6-二甲基-3-乙基吡嗪	C13925070	C <sub>8</sub> H <sub>12</sub> N <sub>2</sub>	136.2	1 436.8	674.074	1.242 5	其他类
VOC21	$\delta$ -壬内酯	C3301948	C <sub>9</sub> H <sub>16</sub> O <sub>2</sub>	156.2	1 398.9	608.819	1.272 2	酯类
VOC22	(E)-2-辛烯醛	C2548870	C <sub>8</sub> H <sub>14</sub> O	126.2	1 398.9	608.819	1.345 9	醛类
VOC23	2-庚基呋喃	C3777717	C <sub>11</sub> H <sub>18</sub> O	166.3	1 403.8	616.976	1.413 1	其他类
VOC24	(Z)-3-己烯-1-醇	C928961	C <sub>6</sub> H <sub>12</sub> O	100.2	1 384.2	585.368	1.238 3	醇类
VOC25	糠醛	C98011	C <sub>5</sub> H <sub>4</sub> O <sub>2</sub>	96.1	1 382.9	583.329	1.349 1	醛类
VOC26	4-羟基-4-甲基-2-戊酮	C123422	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	116.2	1 360.1	548.662	1.134 7	酮类
VOC27	二丙基二硫醚	C629196	C <sub>6</sub> H <sub>14</sub> S <sub>2</sub>	150.3	1 363.5	553.760	1.263 8	其他类
VOC28	1-己醇	C111273	C <sub>6</sub> H <sub>14</sub> O	102.2	1 347.9	531.063	1.339 6	醇类
VOC29	1, 2, 3-三甲基苯	C526738	C <sub>9</sub> H <sub>12</sub>	120.2	1 324.8	499.155	1.168 4	其他类
VOC30	3-甲基-1-戊醇	C589355	C <sub>6</sub> H <sub>14</sub> O	102.2	1 313.2	483.938	1.623 6	醇类
VOC31	2-庚醇	C543497	C <sub>7</sub> H <sub>16</sub> O	116.2	1 312.9	483.447	1.751 1	醇类
VOC32	2-辛酮	C111137	C <sub>8</sub> H <sub>16</sub> O	128.2	1 295.8	461.848	1.336 9	酮类
VOC33	1-甲氧基-2-丙醇乙酸酯	C108656	C <sub>6</sub> H <sub>12</sub> O <sub>3</sub>	132.2	1 259.5	424.540	1.147 4	酯类
VOC34	乙酸(4-甲基苯基)甲酯	C2216457	C <sub>10</sub> H <sub>12</sub> O <sub>2</sub>	164.2	1 269.9	434.849	1.503 8	酯类
VOC35	2-乙氧基乙醇	C110805	C <sub>4</sub> H <sub>10</sub> O <sub>2</sub>	90.1	1 242.2	407.850	1.094 1	醇类
VOC36	环戊酮	C120923	C <sub>5</sub> H <sub>8</sub> O	84.1	1 242.2	407.850	1.116 8	酮类
VOC37	(3-甲基丁氧基)-乙酸-2-丙醇酯	C67634008	C <sub>10</sub> H <sub>18</sub> O <sub>3</sub>	186.3	1 241.2	406.869	1.428 7	酯类
VOC38	2-丁烯酸己酯	C19089920	C <sub>10</sub> H <sub>18</sub> O <sub>2</sub>	170.3	1 242.2	407.850	1.466 3	酯类
VOC39	柠檬烯	C138863	C <sub>10</sub> H <sub>16</sub>	136.2	1 222.0	389.197	1.228 6	不饱和烃类
VOC40	柠檬烯-D	C138863	C <sub>10</sub> H <sub>16</sub>	136.2	1 214.3	382.324	1.304 6	不饱和烃类
VOC41	(Z)-罗勒烯	C470826	C <sub>10</sub> H <sub>18</sub> O	154.3	1 212.0	380.361	1.750 2	不饱和烃类
VOC42	2-戊基呋喃	C3777693	C <sub>9</sub> H <sub>14</sub> O	138.2	1 198.2	368.335	1.265 3	其他类
VOC43	丙基苯	C103651	C <sub>9</sub> H <sub>12</sub>	120.2	1 187.6	359.532	1.266 1	其他类
VOC44	$\alpha$ -水芹烯	C99832	C <sub>10</sub> H <sub>16</sub>	136.2	1 169.2	344.705	1.227 2	不饱和烃类
VOC45	2-异丁基-3-甲基吡嗪	C13925069	C <sub>9</sub> H <sub>14</sub> N <sub>2</sub>	150.2	1 144.7	325.941	1.278 2	其他类
VOC46	$\beta$ -蒎烯	C127913	C <sub>10</sub> H <sub>16</sub>	136.2	1 130.2	315.284	1.306 6	不饱和烃类
VOC47	D- $\beta$ -蒎烯	C127913	C <sub>10</sub> H <sub>16</sub>	136.2	1 129.2	314.589	1.755 2	不饱和烃类
VOC48	1, 4-二甲基苯	C106423	C <sub>8</sub> H <sub>10</sub>	106.2	1 136.2	319.686	1.079 8	其他类
VOC49	2-甲基-1-丙醇	C78831	C <sub>4</sub> H <sub>10</sub> O	74.1	1 094.3	291.191	1.381 1	醇类
VOC50	3-戊醇	C584021	C <sub>5</sub> H <sub>12</sub> O	88.1	1 097.1	292.349	1.436 9	醇类
VOC51	3, 5-二甲基-1, 2-环戊二酮	C13494070	C <sub>7</sub> H <sub>10</sub> O <sub>2</sub>	126.2	1 052.5	274.974	1.197 3	酮类
VOC52	2-甲基丁酸乙酯	C7452791	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	130.2	1 032.7	267.561	1.239 4	酯类
VOC53	乙酸丙酯	C109604	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	102.1	1 003.1	256.904	1.165 7	酯类
VOC54	乙酸2-甲基-1-丙酯	C110190	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	116.2	1 012.2	260.148	1.237 7	酯类
VOC55	2-丁醇	C78922	C <sub>4</sub> H <sub>10</sub> O	74.1	1 008.3	258.758	1.341 2	醇类
VOC56	2-甲基丙酸乙酯	C97621	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	116.2	965.1	243.700	1.209 2	酯类
VOC57	2-戊酮	C107879	C <sub>5</sub> H <sub>10</sub> O	86.1	964.4	243.468	1.123 7	酮类
VOC58	2-乙基呋喃	C3208160	C <sub>6</sub> H <sub>8</sub> O	96.1	898.6	222.155	1.058 5	其他类
VOC59	乙酰乙酸乙酯	C141979	C <sub>6</sub> H <sub>10</sub> O <sub>3</sub>	130.1	896.4	221.460	1.152 2	酯类

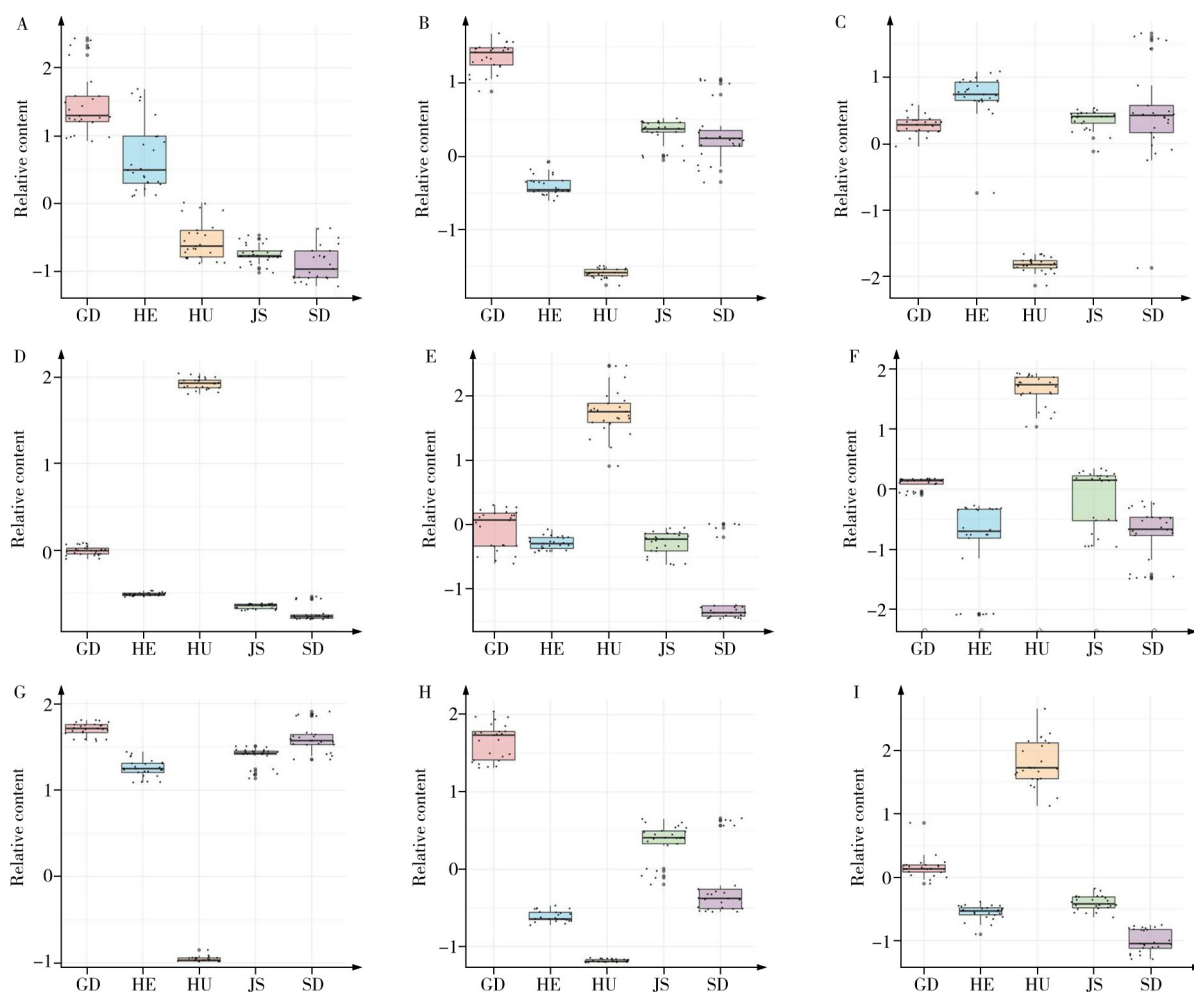


图2 不同产地香附中代表性挥发性成分的相对含量：VOC8(A)、VOC4(B)、VOC12(C)、VOC19(D)、VOC28(E)、VOC50(F)、VOC6(G)、VOC25(H)、VOC15(I)

Fig. 2 Relative contents of representative volatile components in *Cyperi Rhizoma* from different origins; VOC8(A), VOC4(B), VOC12(C), VOC19(D), VOC28(E), VOC50(F), VOC6(G), VOC25(H), VOC15(I)

图3为不同产地香附的挥发性成分热图，颜色与其相对含量有关，越红表明挥发性成分的相对含量越高；越蓝表明挥发性成分的相对含量越低。图3显示，不同产地香附的挥发性成分含量存在差异，其中，GD组香附中挥发性成分的相对含量更高，其次为HE组、JS组和HU组。GD组香附含有较高含量的VOC46~VOC54，而HU组中上述VOCs的含量显著低于GD组；SD组香附含有更高含量的VOC16~VOC58；与SD组相比，JS组含有较高含量的VOC49~VOC31；HE组含有较高含量的VOC46~VOC13。

2.1.2 不同产地香附的差异挥发性成分分析 PCA是一种常用的数据降维技术，其目的是通过正交线性变换将高维数据转化为低维数据，同时保留原始数据的信息<sup>[21]</sup>。如图4A所示，主成分1(Dim1)可解释45.5%的方差，主成分2(Dim2)可解释26%的方差，该模型共可解释71.5%的方差，表明该模型在降维的同时可有效保留主要的原始数据信息。采用PCA方法分析，GD组、HU组和SD组能够被明显区分，然而HE、JS和SD组距离较近，区分度较差，这可能归因于产地的地理位置和气候条件相近。河南地处中原，属温带季风气候，夏季高温多雨，冬季寒冷干燥，光照充足<sup>[22]</sup>；江苏跨南北气候带，淮河以北属温带季风气候<sup>[23]</sup>；山东地处黄河下游，属温带季风气候，降水集中在夏季，春秋干旱，光照充足，昼夜温差大<sup>[24]</sup>；三省同属温带季风气候，地理与气候特征存在显著共性，这可能导致HE、JS和SD组距离较近，区分度较差。

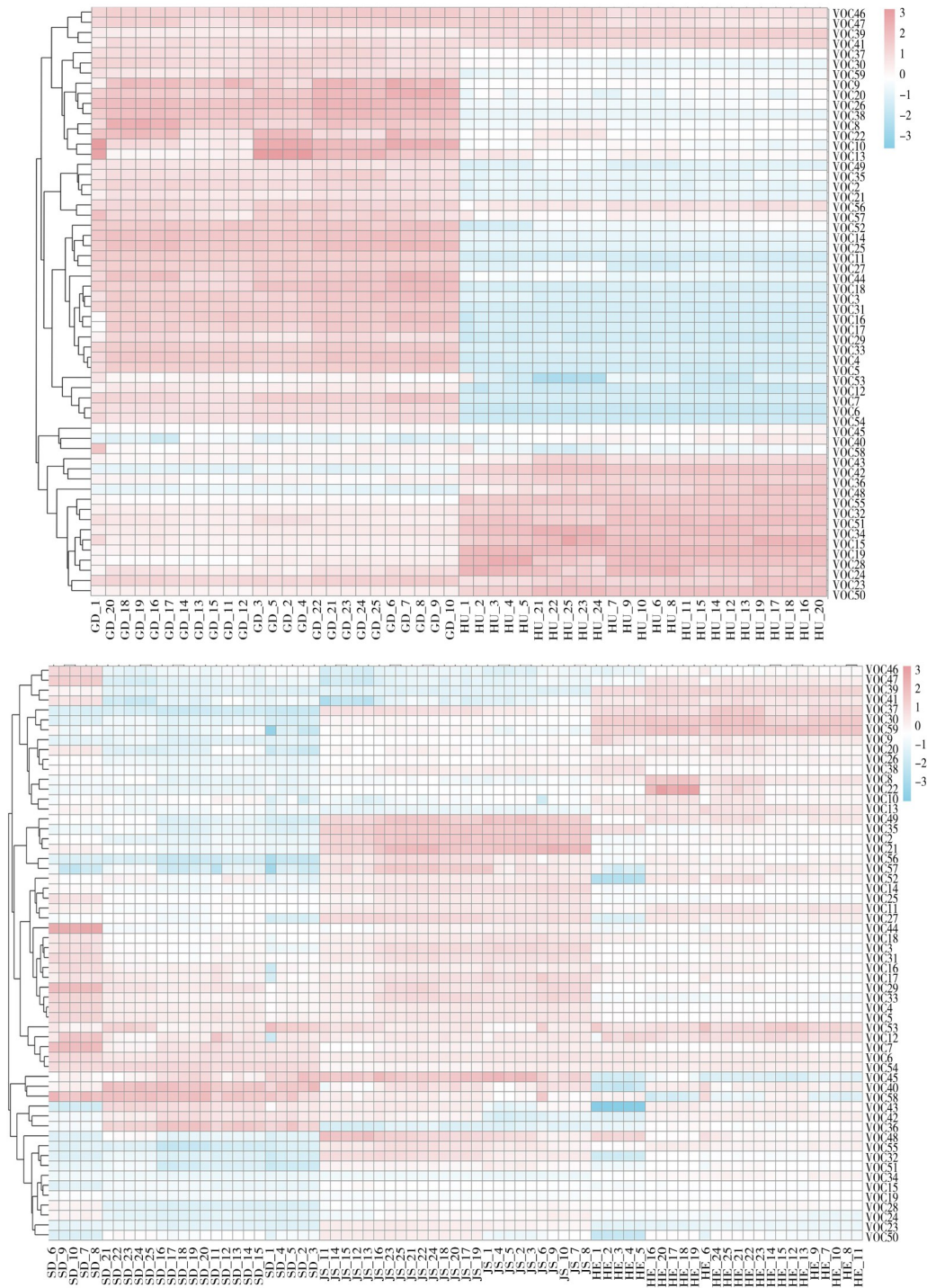


图 3 不同产地香附挥发性成分相对含量的聚类热图

Fig. 3 Cluster heatmaps of the relative contents of volatile components in *Cyperi Rhizoma* from different origins

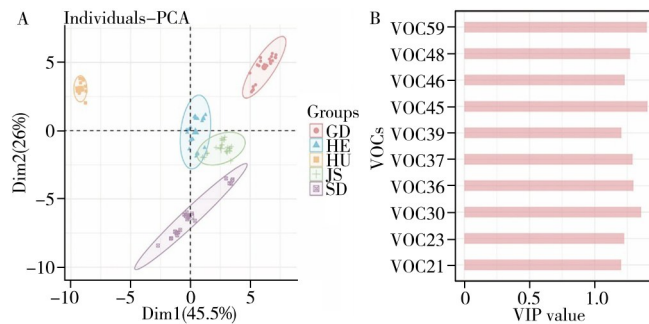


图 4 不同产地香附的主成分分析图(A)与差异成分VIP值(B)

Fig. 4 Principal component analysis (PCA) of *Cyperi Rhizoma* from different origins (A), and VIP values of differential components (B)

PLS-DA 是一种判别分析方法，常用于处理分类和判别问题，能够有效区分组间观测值，并识别导致组间差异的影响变量<sup>[25]</sup>。采用  $Q^2$  (表3)、 $R^2$  (表4)、交叉验证 (图5) 等验证 PLS-DA 模型的可靠性。采用 PLS-DA 对不同产地香附的差异挥发性成分进行分析 (图4B)，共筛选出变量重要性投影 (VIP) > 1.2 的 10 种特征成分，包括乙酰乙酸乙酯 (VOC59)、环戊酮 (VOC36)、 $\beta$ -蒎烯 (VOC46)、柠檬烯 (VOC39)、(3-甲基丁氧基)-乙酸-2-丙烯醇酯 (VOC37)、1, 4-二甲基苯 (VOC48)、3-甲基-1-戊醇 (VOC30)、2-庚基咪喃 (VOC23)、 $\delta$ -壬内酯 (VOC21) 和 2-异丁基-3-甲基吡嗪 (VOC45)。其中， $\beta$ -蒎烯具有抗炎、抗疟疾、抗菌及抗氧化等作用<sup>[26]</sup>；柠檬烯属于不饱和烃类，具有抗氧化、抗炎、抗肿瘤、神经保护活性<sup>[27]</sup>。此外，VOC46 在 GD 组中含量最高 (图 6C)；VOC59、VOC37 和 VOC30 在 HE 组中含量最高 (图 6A、E、G)；VOC39、VOC48 和 VOC23 在 HU 组中含量最高 (图 6D、F、H)；VOC21 和 VOC45 在 JS 组中含量最高 (图 6I、J)；VOC36 在 SD 组中含量最高 (图 6B)。

表3 PLS-DA 模型的  $Q^2$   
Table 3  $Q^2$  of the PLS-DA model

	GD	HE	HU	JS	SD
Principal component 1	0.286	-0.003 18	0.816	0.059 0	-0.017 7
Principal component 2	0.544	-0.007 46	0.693	0.003 33	0.647
Principal component 3	0.137	0.152	0.042 0	0.930	0.260
Principal component 4	0.632	0.962	0.773	0.515	0.847

表4 PLS-DA 模型的  $R^2$   
Table 4  $R^2$  of the PLS-DA model

	$R^2X$	$R^2Y$
Principal component 1	0.454	0.243
Principal component 2	0.714	0.480
Principal component 3	0.823	0.721
Principal component 4	0.899	0.960

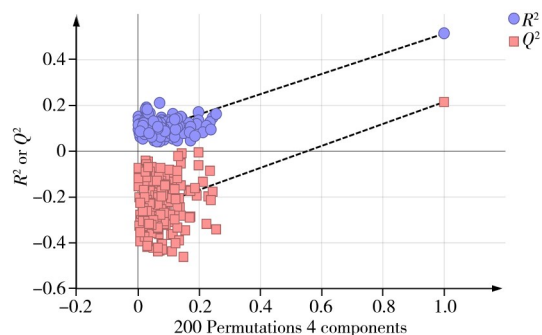


图5 PLS-DA 交叉验证图  
Fig. 5 PLS-DA cross-validation plot

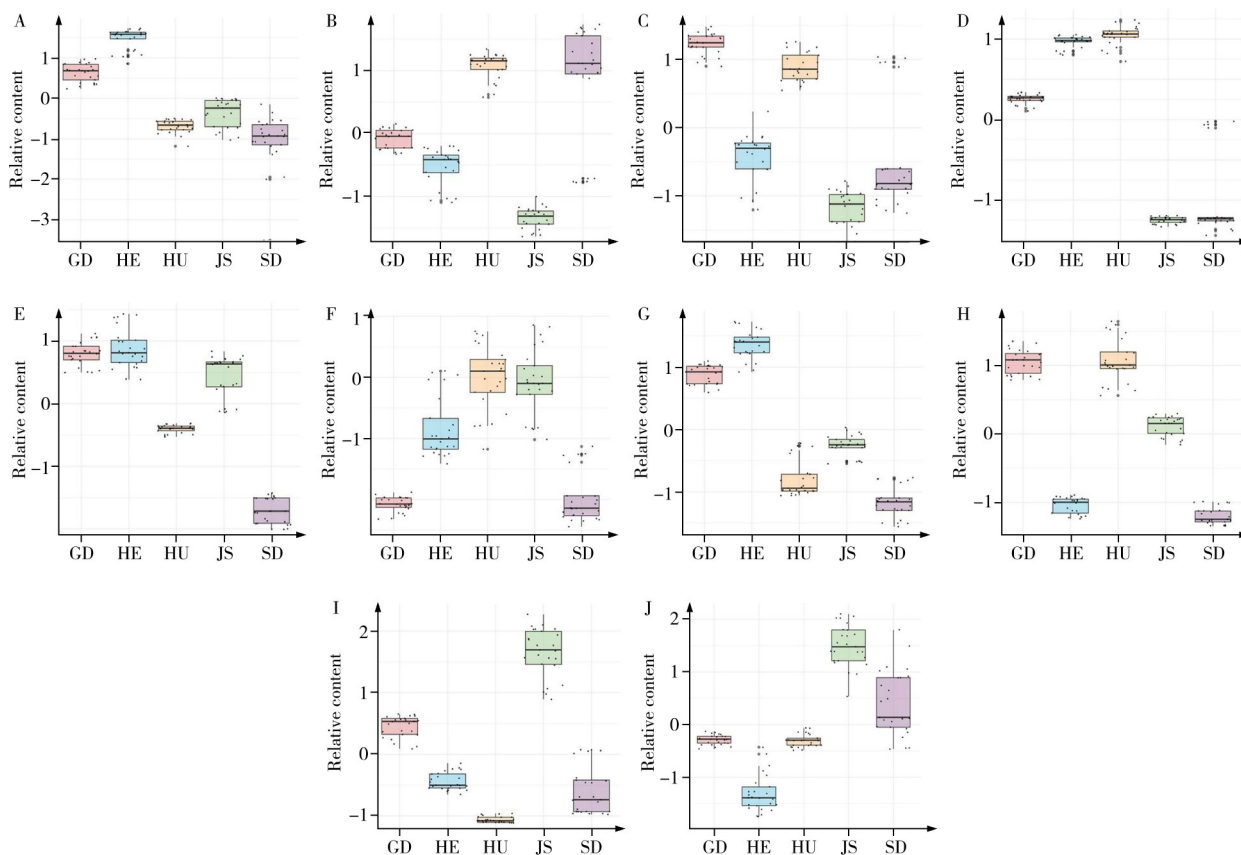


图6 不同产地香附中差异成分的相对含量：VOC59(A)、VOC36(B)、VOC46(C)、VOC39(D)、VOC37(E)、VOC48(F)、VOC30(G)、VOC23(H)、VOC21(I)、VOC45(J)

Fig. 6 Relative contents of differential components in *Cyperi Rhizoma* decoction pieces from different origins: VOC59(A), VOC36(B), VOC46(C), VOC39(D), VOC37(E), VOC48(F), VOC30(G), VOC23(H), VOC21(I), VOC45(J)

## 2.2 机器学习算法区分模型的构建

基于筛选得到的 10 种特征成分, 分别采用 9 种机器学习模型(KNN、LDA、NB、RF、XGB、SVM-L、SVM-P、SVM-R 和 SVM-S)对不同产地香附饮片进行区分。通过混淆矩阵(图 7)计算精确率、召回率、F1 分数和准确率。所有机器学习模型在训练集和测试集中的上述评估指标均为 1, 表明上述模型具有良好的分类能力, 这主要归因于其数学特性。例如, SVM-P 通过最大化决策边界与数据点之间的间隔(即“间隔最大化”技术)启动分类过程<sup>[28]</sup>, 使得 SVM-P 构建的分类器不仅鲁棒性强, 还具有较高的泛化能力, 能有效降低过拟合风险; LDA 则通过最大化类间离散度与类内离散度的比值, 在降维空间中寻找能最大程度分离不同类别数据的最优投影方向, 进而构建高效分类器, 提升预测准确率<sup>[29]</sup>。

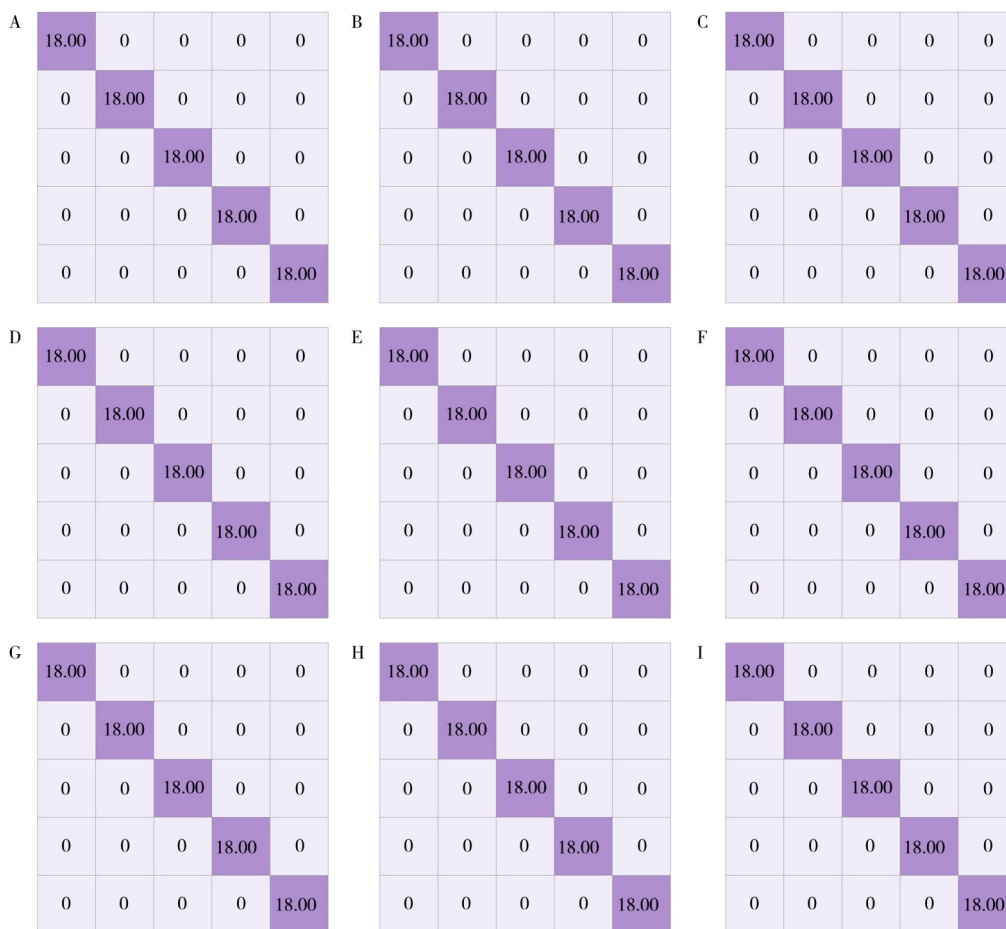


图 7 机器学习模型的混淆矩阵: KNN(A)、LDA(B)、NB(C)、RF(D)、SVM-L(E)、SVM-P(F)、SVM-R(G)、SVM-S(H)、XGB(I)

Fig. 7 Confusion matrix of a machine learning model: KNN(A), LDA(B), NB(C), RF(D), SVM-L(E), SVM-P(F), SVM-R(G), SVM-S(H), XGB(I)

## 3 结 论

本研究采用 GC-IMS 从香附中鉴定出 59 种挥发性成分, 主要包括酯类、醇类和醛类等成分。其中, GD 组香附中挥发性成分的含量最高, 其次为 HE 组、JS 组、HU 组和 SD 组。随后, 采用 PLS-DA 从不同产地香附中筛选出 10 个特征成分。基于这些特征成分, 构建了 9 种机器学习区分模型, 在训练集和测试集中均表现出良好的预测能力(Accuracy=1), 表明其在香附产地快速区分中具有较高的应用价值。机器学习是传统多元统计分析的升级, 其便捷性和准确性优于传统多元统计分析。PCA 与 PLS-DA 均基于线性降维与判别假设, 无法有效捕捉数据中的非线性特征, 且 PCA 仅做无监督降维, 无法主动筛选与产地高度相关的特征<sup>[21]</sup>; PLS-DA 虽能通过 VIP 筛选特征, 但筛选过程受主成分个数的影响较大, 存在主观性<sup>[25]</sup>。机器学习算法可突破线性假设的限制, 通过特征空间映射或多层非线性拟合, 挖掘挥发性成分数据与产地间的隐藏关联, 具有强鲁棒性, 可有效抵抗噪声干扰, 实现更高精度、更具

解释性的产地鉴别<sup>[30]</sup>。本研究为香附的产地溯源提供了一种精准、简便、快速的鉴别方法,同时可为香附质量评价方法的建立及临床用药的稳定性提供参考依据。

#### 参考文献:

- [1] Wang F, Zhang S, Zhang J, Yuan F. *Front. Pharmacol.*, **2022**, 13: 965902.
- [2] Pharmacopoeia Commission of the People's Republic of China. *Pharmacopoeia of the People's Republic of China*. Volume I. Beijing: China Medical Science and Technology Press(国家药典委员会. 中华人民共和国药典. 一部. 北京: 中国医药科技出版社), **2020**.
- [3] Xu N, Mou Y Z, Li W B, Fu C M, Chen H L, Wang S Y, Lu J R. *China Pharm.* (许娜, 牟玉贞, 李文兵, 傅超美, 陈胡兰, 王世宇, 卢君蓉. 中国药房), **2020**, 31(23): 2833-2840.
- [4] Li Y Y, Ji L, Lei J W, Yang C J, Jia H, Zhang W F, Gong H Y, Liu W J, Xie C X. *Chin. Arch. Tradit. Chin. Med.* (李莹莹, 纪亮, 雷敬卫, 杨春静, 贾豪, 张维方, 龚海燕, 刘文俊, 谢彩侠. 中华中医药学刊), **2024**, 42(1): 194-199, 278.
- [5] Hu S M, Guo H Y, Lu H, Liu W, Wang X, Zhao H Q. *J. Instrum. Anal.* (胡思萌, 郭焕滢, 卢恒, 刘伟, 王晓, 赵恒强. 分析测试学报), **2024**, 43(11): 1843-1850.
- [6] Wang Y, Fan J L, Zhang M M, Ji W H, Wang X, Li L L. *J. Instrum. Anal.* (王燕, 范佳丽, 张敏敏, 纪文华, 王晓, 李丽丽. 分析测试学报), **2023**, 42(8): 976-983.
- [7] Zhang Y F, Zhang J B, Chen P, Shi Y B, Zhang K W, Mao C Q, Lu T L. *Chin. Tradit. Herb. Drugs*(张云飞, 张玖捌, 陈鹏, 史亚博, 张科卫, 毛春芹, 陆兔林. 中草药), **2024**, 55(12): 4185-4193.
- [8] Liu W W, Dong H J, Zhang M M, Liu S, Ma X H, Wang Z Q, Wang X. *Chin. Hosp. Pharm. J.* (刘纹纹, 董红敬, 张敏敏, 刘双, 马鑫慧, 王珍强, 王晓. 中国医院药学杂志), **2024**, 44(18): 2082-2089.
- [9] Mi Y Z, Dong H J, Wang X, Liu S, Jiang M, Liang Q, Chen J. *Rapid Commun. Mass Spectrom.*, **2025**, 39(8): 9991.
- [10] Liu S, Dong H J, Zhang M M, Geng W, Wang X. *J. Pharm. Anal.*, **2024**, 14(1): 149-151.
- [11] Sun H Y, Hao D Q, Li X S, Jin W G. *Food Mach.*, **2022**, 38(4): 58-64.
- [12] Zhang B Y, Sui Y M, Zhang X, Li Y J, Kong B H, Liu H T, Chen Q. *Food Sci.* (张碧莹, 隋雨萌, 张鑫, 李永杰, 孔保华, 刘昊天, 陈倩. 食品科学), **2025**, 46(1): 227-236.
- [13] Hur J, Jang J, Sim J. *Int. J. Mol. Sci.*, **2021**, 22(5): 2769.
- [14] Rios E R V, Rocha N F M, Carvalho A M R, Vasconcelos L F, Dias M L, Sousa D P, Sousa F C F, Fonteles M M F. *Chem. -Biol. Interact.*, **2013**, 203(3): 573-579.
- [15] Zhang Y R, Cui C X, He Y N, Li C J, Wang D X, Tian Q, Sheng W Q. *J. Henan Univ. Technol.: Nat. Sci. Ed.* (张玉荣, 崔春晓, 何雅楠, 李晨杰, 王殿轩, 田奇, 生吴琼. 河南工业大学学报: 自然科学版), **2023**, 44(2): 96-104.
- [16] Sales A, Felipe L D O, Bicas J L. *Food Bioprod. Process.*, **2020**, 13(8): 1261-1279.
- [17] Bouyahya A, Mechchate H, Benali T, Ghchime R, Charfi S, Balahbib A, Burkov P, Shariati M A, Lorenzo J M, Omari N E. *Biomolecules*, **2021**, 11(12): 1803.
- [18] Feng J M, Li S Y, You R Q, Bao X M, Chen X M, Huang H Z, Zheng C, Zhang D K. *Chin. Tradit. Herb. Drugs*(冯佳敏, 李师瑶, 游润秋, 包晓明, 陈心铭, 黄浩洲, 郑川, 张定堃. 中草药), **2025**, 56(15): 5596-5604.
- [19] Chai W M, Liu X, Hu Y H. *Int. J. Biol. Macromol.*, **2013**, 57: 151-155.
- [20] Kim J H, Chan K L, Mahoney N, Campbell B C. *Ann. Clin. Microbiol. Antimicrob.*, **2011**, 10: 23.
- [21] Kurita T. *Computer Vision*. Cham: Springer International Publishing, **2021**: 1013-1016.
- [22] Wu L, Chang J. The 35th Annual Meeting of the Chinese Meteorological Society, Hefei(吴璐, 常军. 第35届中国气象学会年会, 合肥), **2018**.
- [23] Ren Y F, Wang C Y, Zhang X H, Li S R, Qian B D. *Chin. J. Agrometeorol.* (任义方, 王春乙, 张旭晖, 李时睿, 钱半吨. 中国农业气象), **2024**, 45(10): 1160-1173.
- [24] Dong L J, Sun Y Y, Zhang L J, Wang X Y, Zhang C X, Yang B. *China Fruit News*(董丽娟, 孙阳阳, 张立娇, 王晓英, 张春香, 杨波. 中国果业信息), **2025**, 42(2): 64-65, 71.
- [25] Kalivodová A, Hron K, Filzmoose P, Najdekr L, Janečková H, Adam T. *J. Chemom.*, **2014**, 29(1): 21-28.
- [26] Park B B, An J Y, Park S U. *EXCLI J.*, **2021**, 22(20): 812-818.
- [27] Eddin L B, Jha N K, Meeran M F N, Kesari K K, Beiram R, Ojha S. *Molecules*, **2021**, 26(15): 4535.
- [28] Asaka M S. *Adv. Appl. Math.*, **2021**, 10(12): 4535-4541.
- [29] Liu S, Dong H J, Lu H, Li M, Wang X. *Appl. Food Res.*, **2025**, 5(1): 100760.
- [30] Zhang H W, Li Y E, Yu J W, Guo Q, Li M X, Li Y, Mei Q, Li L, Su L L, Mao C Q, Ji D, Lu T L. *China J. Tradit. Chin. Med.* (张翰文, 李月娥, 喻家伟, 郭强, 李铭轩, 李昱, 梅茜, 李林, 苏联麟, 毛春芹, 季德, 陆兔林. 中国中药杂志), **2025**, 50(13): 3605-3614.