

电子鼻技术应用于白及及其近似饮片快速辨识的可行性分析

李涵¹, 王艳丽², 范雪花¹, 李海洋¹, 侯富国¹, 桂新景^{1,2}, 施钧瀚²,
张璐², 刘瑞新^{2*}, 李学林^{2*}

(1. 河南中医药大学, 郑州 450046; 2. 河南中医药大学第一附属医院, 郑州 450000)

[摘要] 目的:探讨电子鼻技术应用于白及及其近似饮片快速辨识的可行性。方法:收集134批白及及其近似饮片(白及45批、天麻30批、玉竹30批、黄花白及29批)作为待测样品,使用PEN3型电子鼻采集样品嗅觉感官数据作为自变量 X ,基于2020年版《中华人民共和国药典》和地方标准的鉴别结果,以及各饮片高效液相色谱法(HPLC)指纹图谱和原始采购信息,获得辨识模型的标杆数据 Y ,分别采用主成分分析-判别分析(PCA-DA)、偏最小二乘法-判别分析(PLS-DA)、最小二乘法-支持向量机(LS-SVM)及K-最近邻(KNN)4种化学计量学方法建立45批白及与89批非白及的二分类辨识模型和上述4种饮片的四分类辨识模型 $Y=F(X)$ 。结果:经留一法交互验证,在二分类辨识中,上述4种模型分类正判率分别为97.01%、97.01%、98.51%和97.01%;在四分类辨识中,这4种模型分类正判率分别为97.76%、89.55%、98.51%和97.01%。二分类和四分类辨识模型的最高正判率均可达到98.51%,且均以LS-SVM算法为最优,最优核函数分别选择径向基核函数和线性核函数。最优模型判别结果良好,没有未分类样本。结论:电子鼻技术能够准确、快速地辨识白及及其近似饮片,可为其他中药饮片的快速质量评价提供新思路和新方法。

[关键词] 电子鼻技术; 白及; 辨识模型; 中药饮片; 天麻; 玉竹; 黄花白及

[中图分类号] R22;R28;R94;O657 **[文献标识码]** A **[文章编号]** 1005-9903(2023)13-0157-09

[doi] 10.13422/j.cnki.syfjx.20221447 **[增强出版附件]** 内容详见<http://www.syfjxzz.com> 或<http://cnki.net>

[网络出版地址] <https://kns.cnki.net/kcms/detail/11.3495.R.20221107.1733.003.html>

[网络出版日期] 2022-11-08 09:34:01

Analysis on Feasibility of Electronic Nose Technology for Rapid Identification of Bletillae Rhizoma and Its Approximate Decoction Pieces

LI Han¹, WANG Yanli², FAN Xuehua¹, LI Haiyang¹, HOU Fuguo¹, GUI Xinjing^{1,2}, SHI Junhan²,
ZHANG Lu², LIU Ruixin^{2*}, LI Xuelin^{2*}

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

2. The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450000, China)

[Abstract] **Objective:** To investigate the feasibility of applying electronic nose technology to rapidly identify Bletillae Rhizoma and its approximate decoction pieces. **Method:** A total of 134 batches of Bletillae Rhizoma and its approximate decoction pieces, including 45 batches of Bletillae Rhizoma, 30 batches of Gastrodiae Rhizoma, 30 batches of Polygonati Odorati Rhizoma and 29 batches of Bletillae Ochraceae Rhizoma, were collected as test samples. The olfactory sensory data of the samples were collected by PEN3

[收稿日期] 2022-08-11

[基金项目] 国家自然科学基金项目(81773892);国家重点研发计划中医药现代化重点专项(2017YFC1703400,2017YFC1703401,2017YFC1703402);河南省中医药拔尖人才培养项目(重点项目)(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);河南省卫生健康委员会国家中医临床研究基地科研专项(2021JDZY104)

[第一作者] 李涵,在读硕士,从事中药质量评价及临床应用研究,E-mail:1531409875@qq.com

[通信作者] *刘瑞新,博士,主任药师,从事中药饮片临床应用现代化关键技术研究,Tel:0371-66233562,E-mail:liuruixin7@163.com;

*李学林,硕士,主任药师,从事中药应用形式研究,Tel:0371-66245142,E-mail:xuelinli450000@163.com

electronic nose as the independent variable(X). Based on the identification results of the 2020 edition of *Chinese Pharmacopoeia* and local standards, as well as the high performance liquid chromatography(HPLC) fingerprint and original purchase information of 134 batches of the decoction pieces, the benchmark data Y of the identification model were obtained, and four chemometric methods of principal component analysis-discriminant analysis(PCA-DA), partial least squares-discriminant analysis(PLS-DA), least square-support vector machine (LS-SVM) and K-nearest neighbor(KNN) were used to establish the binary identification model for 45 batches of *Bletillae Rhizoma* and 89 batches of non-*Bletillae Rhizoma* and the quadratic identification model of the four kinds of decoction pieces, that is, $Y=F(X)$. **Result:** After leave-one-out cross validation, the positive discrimination rates of the above four models were 97.01%, 97.01%, 98.51% and 97.01% in the binary identification, and 97.76%, 89.55%, 98.51% and 97.01% in the quadratic identification, respectively. The highest positive discrimination rate could reach 98.51% for the binary and quadratic identification models, and LS-SVM algorithm is both the optimal one, the most suitable kernel functions were chosen as radial basis function and linear kernel function, respectively. The optimal models discriminated well with no unclassified samples. **Conclusion:** Electronic nose technology can accurately and rapidly identify *Bletillae Rhizoma* and its approximate decoction pieces, which can provide new ideas and methods for rapid quality evaluation of other decoction pieces.

[Keywords] electronic nose technology; *Bletillae Rhizoma*; identification model; decoction pieces; *Gastrodiae Rhizoma*; *Polygonati Odorati Rhizoma*; *Bletillae Ochraceae Rhizoma*

白及是兰科植物白及 *Bletilla striata* 的干燥块茎,始载于《神农本草经》,味苦、甘、涩,性微寒,功效收敛止血、消肿生肌,具有止血、保护胃黏膜、抗溃疡、抗菌、抗炎与促进伤口愈合等药理活性^[1]。此外,白及还常用作生物医药材料^[2]、药物制剂辅料^[3]、工业制胶原料、化妆品原料^[4]等。白及药材在自然状态下繁殖率低,而且由于人工过度采挖,其野生资源逐年减少。近年来白及种苗繁育和种植技术的发展在一定程度上解决了其野生资源减少的问题,但随着多种组织培育技术的成熟,白及药材的来源变得多样化,因此市场上存在白及饮片品种混杂、品质参差不齐的现象^[5],如在白及中掺入同属植物黄花白及^[6]或与白及有相似性状特征的天麻次品和玉竹碎片^[7]等。伪品的流通会在一定程度上影响白及饮片的临床疗效,甚至可能会因掺伪用量不当引发一些潜在的用药安全事故。因此,为控制饮片质量,保证临床的安全用药,有必要对白及进行真伪鉴别。

目前,对白及的真伪鉴别方法有传统经验鉴别、显微鉴别^[8]、薄层色谱法(TLC)鉴别^[9]、近红外光谱鉴别^[10-11]、DNA条形码技术鉴别^[12-13]、指纹图谱鉴别^[14-15]等,不同检测方法各有优势。例如,传统经验鉴别快速,显微、TLC、液相色谱结果客观准确且可定性表征,近红外光谱灵敏度高、可无损检测,DNA条形码技术标准严格、可靠性强。然而,这些

技术均存在一定不足,经验鉴别主观性强且难以量化,显微、TLC、液相色谱鉴别方法前处理繁琐耗时,DNA条形码技术操作难度相对较大。因此有必要探索新的准确、快速的饮片辨识方法。

电子鼻的工作原理是模拟人体嗅觉器官,通过气味传感器将样品气味信号转变为电信号,进而快速获得样品整体气味信息。该技术具有准确、灵敏、无损、实时的优势^[16-17]。由于不同样品所含成分对传感器的响应不同,因此可通过多种模式识别不同样品中的微小差异,从而实现对样品的区分。目前该技术已应用于中药材的产地鉴别、品质评价、气味差异鉴别、炮制品鉴别和掺伪鉴别等研究中^[18-20]。因此,本研究选取白及饮片及其近似品黄花白及、天麻、玉竹饮片作为研究载体,利用电子鼻技术获取不同样品嗅觉感官数据,借助化学计量学方法分别建立二分类(白及与非白及)和四分类(白及、黄花白及、天麻、玉竹)辨识模型,以模型留一法交互验证正判率为指标,探讨电子鼻技术用于白及其近似饮片辨识的可行性,为其他饮片的快速鉴别提供参考。

1 材料

PEN3型电子鼻(德国Airsense公司),10种金属氧化物传感器及其敏感物质见表1;UltiMate 3000型高效液相色谱仪[赛默飞世尔科技(中国)有限公司],FW-100型高速万能粉碎机(北京科伟永兴仪器

有限公司),BSA2245-CW型万分之一电子天平和CP225D型十万分之一电子分析天平(德国Sartorius公司),DM1000型电子显微镜[徕卡显微系统(上海)贸易有限公司],TLC Visualizer 2型薄层色谱数码成像系统(瑞士卡玛公司),GZX-9146MBE型电热鼓风干燥箱(上海博迅实业有限公司医疗设备厂),SRJX-4-13型高温箱式电阻炉(沈阳市节能电炉厂)。

表1 PEN3型电子鼻10根传感器及其敏感物质

Table 1 Sensitive substances of 10 sensors in PEN3 electronic nose

编号	传感器	敏感物质
S1	W1C	芳烃化合物
S2	W5S	氮氧化合物
S3	W3C	氨、芳香成分
S4	W6S	氢化物
S5	W5C	烯烃、芳族、极性分子
S6	W1S	烷类
S7	W1W	无机硫化物
S8	W2S	醇类、部分芳香族化合物
S9	W2W	芳香成分、有机硫化物
S10	W3S	烷烃、脂肪族

1,4-二[4-(葡萄糖氧)苄基]-2-异丁基苹果酸酯(militarine)对照品(上海源叶生物科技有限公司,批号K1809B72711,纯度 $\geq 98\%$),天麻素对照品(纯度96.7%)和白及对照药材(中国食品药品检定研究院,批号分别为110807-201809、121261-201706),水为超纯水,甲醇、乙腈、磷酸为色谱纯,其他试剂均为分析纯。实验用45批白及、30批天麻、30批玉竹和29批黄花白及待测样本的相关信息见增强出版附加材料。所有样本均经河南中医药大学第一附属医院施钧瀚副主任药师鉴定。31~134号样本是从药材市场上购买后根据2020年版《中华人民共和国药典》(以下简称《中国药典》)(一部)“白及、天麻、玉竹”项下及2015年版《四川省中药饮片炮制规范》“黄花白及”项下自行净制和切制的饮片。白及、天麻、玉竹、黄花白及饮片编号分别为1~45、46~75、76~105、106~134。指纹图谱及相似度分析采用ChemPattern 2017 Pro软件。

2 方法

2.1 嗅觉信息的获取 取1~134号样本粉碎后过六号筛,精密称取样本粉末2.0 g于100 mL烧杯中,双层保鲜膜封口,放置30 min后测试。采用直接顶空吸气法将进样针头插入含样品的密封样品杯中,

使用电子鼻对样品进行检测。检测条件设定为采样时间80 s,传感器自清洗时间80 s,传感器归零时间5 s,样品准备时间5 s,进样流速400 mL \cdot min⁻¹。每个样品采集3次,取平均值,得到10根传感器嗅觉信息矩阵 $X(10 \times 134)$ 。

2.2 标杆信息的获取

2.2.1 基于2020年版《中国药典》及地方标准的辨识 参照2020年版《中国药典》(一部)“白及”项下的性状、鉴别(显微和TLC)、水分和灰分检查方法,对134个样本进行鉴别和测定;并结合2015年版《四川省中药饮片炮制规范》^[21]所载黄花白及及*B. ochracea*、2009年版《甘肃省中药炮制规范》^[22]和2009年版《甘肃省中药材标准》^[23]中小白及(实为黄花白及*B. ochracea*,而非植物学白及属小白及*B. formosana*)项下的性状鉴定方法,对4种饮片进行辨识。

2.2.2 高效液相色谱法(HPLC)指纹图谱 检测条件设定为Shim-pack GIST C₁₈-AQ色谱柱(4.6 mm \times 250 mm,5 μ m),流动相0.1%磷酸水溶液(A)-乙腈(B)梯度洗脱(0~5 min,5%~20%A;5~10 min,20%~24%A;10~20 min,24%~31.5%A;20~25 min,31.5%~35%A;25~30 min,35%~42%A;30~45 min,42%~60%A),流速1.0 mL \cdot min⁻¹,检测波长280 nm,柱温30 $^{\circ}$ C,进样量10 μ L。对HPLC指纹图谱方法进行方法学考察,涉及精密度试验[取14号样品供试品溶液按上述色谱条件重复进样6次,以militarine为参照峰计算各共有峰相对保留时间和相对峰面积的相对标准偏差(RSD)]、重复性试验(制备14号样品供试品溶液6份,按上述色谱条件测定,以militarine为参照峰计算各共有峰相对保留时间和相对峰面积的RSD)、稳定性试验(取2号样品供试品溶液按上述色谱条件分别于制备后0、2、4、8、12、24 h进样,以militarine为参照峰,计算各共有峰相对保留时间和相对峰面积的RSD)。对所有样本进行HPLC分析并建立指纹图谱,建立各类样本的共有模式图谱并标明共有峰。以分离度较好、峰面积较大的militarine为参照峰,以45批白及和对照药材为代表性样本生成共有模式,以该共有模式为对照指纹图谱,采用共有峰加权法对134批样品进行相似度分析(其中militarine权重为2.5,天麻素权重为1,其他组分权重均为1)。

2.3 分类辨识模型 $Y=F(X)$ 的构建及优化 以电子鼻采集的134个样本的嗅觉信息矩阵为 X 值,以2020年版《中国药典》和地方标准鉴别、HPLC指

图谱和原始采购信息的综合结果为标杆 Y 值,建立 X 和 Y 之间的关系模型 $Y=F(X)$ 。利用 Matlab® 2016b 平台建模时,二分类辨识模型的构建中 Y 设定 1 为白及饮片,2 为非白及饮片;四分类辨识模型的构建中 Y 设定 1 为白及饮片,2 为天麻饮片,3 为玉竹饮片,4 为黄花白及饮片。建立 134 个样本的主成分分析-判别分析(PCA-DA)、偏最小二乘法-判别分析(PLS-DA)、最小二乘法-支持向量机(LS-SVM)和 K-最近邻(KNN)二分类和四分类辨识模型,以留一法交互验证后的分类结果作为电子鼻辨识结果,根据留一法交互验证后的模型正判率选择最优模型。建模及验证方法参考文献[24]。

2.4 嗅觉传感器对分类贡献度的追踪

2.4.1 基于 Wilk's Lambda 值变异度分析 以 PEN3 型电子鼻的 10 根嗅觉传感器为原始变量,通过 Wilk's Lambda 柱状图可分析各传感器携带的变异信息,从而追踪各传感器对模型分类的贡献度大小。

2.4.2 模型原始变量载荷图追踪 根据 2.3 项下二分类及四分类的 PLS-DA 或 PCA-DA 模型,分别建立其模型原始变量载荷图,同时结合各传感器 Wilk's Lambda 值,找到对模型分类结果贡献度较大的传感器。

3 结果与分析

3.1 基于 2020 年版《中国药典》及地方标准检测结果的辨识 根据性状、显微、TLC 鉴别及水分、灰分检查的综合结果,45 批白及饮片符合 2020 年版《中国药典》(一部)“白及”项下标准,不符合 2015 年版《四川省中药饮片炮制规范》、2009 年版《甘肃省中药炮制规范》和 2009 年版《甘肃省中药材标准》“黄花白及”项下标准。30 批天麻饮片和 30 批玉竹饮片均不符合 2020 年版《中国药典》(一部)“白及”项下标准,29 批黄花白及饮片既符合 2020 年版《中国药典》(一部)“白及”项下标准,又符合 2015 年版《四川省中药饮片炮制规范》、2009 年版《甘肃省中药炮制规范》和 2009 年版《甘肃省中药材标准》“黄花白及”项下标准。106~120 号黄花白及样本性状特征与白及相同;121~134 号黄花白及样本与白及比较,木化严重,质感有明显差异,性状特征与白及不同,符合地方标准中规定的黄花白及项下标准;29 批黄花白及样本与白及对照药材薄层色谱极为相似,难以区别;29 批黄花白及样本显微特征与白及相同,但其显微背景较白及浑浊。可见,仅靠 2020 年版《中国药典》及地方标准,部分黄花白及饮片不易辨识。

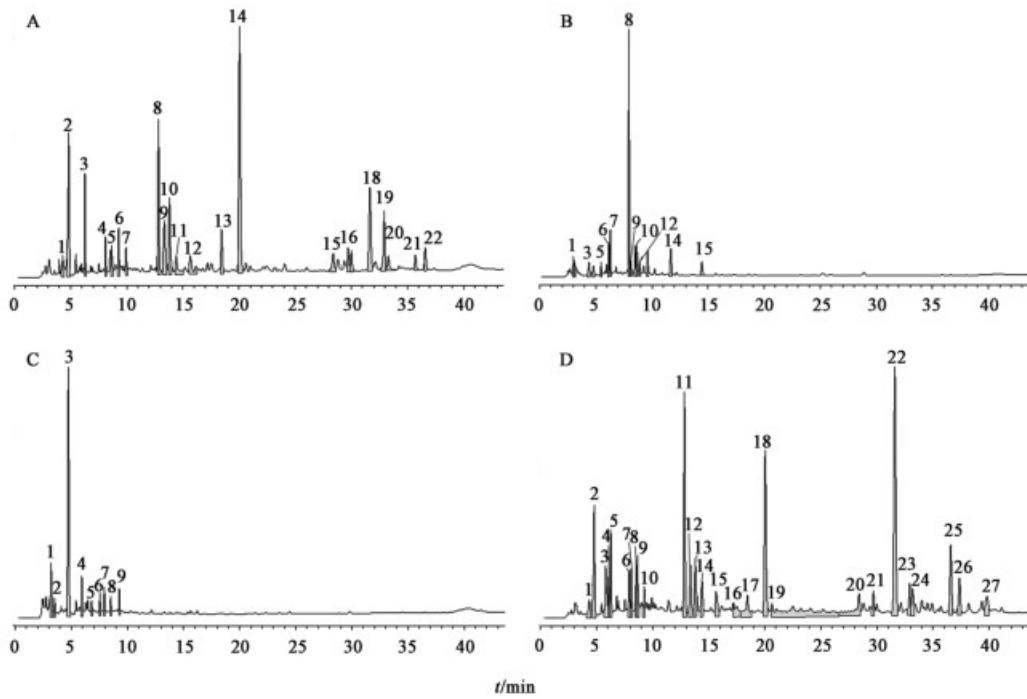
3.2 HPLC 指纹图谱相似度 在 HPLC 指纹图谱方法学考察时,精密度试验中 8 个共有峰相对保留时间和相对峰面积的 RSD 均 $\leq 0.1\%$,说明仪器精密度良好;重复性试验中 8 个共有峰相对保留时间和相对峰面积的 RSD 分别为 $\leq 0.1\%$ 、2.3%,表明该方法具有较好重复性;稳定性试验中 8 个共有峰相对保留时间和相对峰面积的 RSD 分别为 0.1%、0.3%,表明供试品溶液 24 h 内稳定性良好。每类样本的 HPLC 指纹图谱见增强出版附加材料,共有模式见图 1,根据对照品指认出白及共有模式中 3 号峰为天麻素、14 号峰为 militarine,天麻共有模式中 7 号峰为天麻素,黄花白及共有模式中 5 号峰为天麻素、18 号峰为 militarine。由图 1 可知,白及与天麻、玉竹的指纹图谱差异明显。45 批白及的指纹图谱相似度 0.88~0.98(平均值 0.931),30 批天麻的指纹图谱相似度 0.2~0.34(平均值 0.263,与白及共有模式图谱比较, $P < 0.01$),30 批玉竹的指纹图谱相似度 0.1~0.2(平均值 0.134,与白及共有模式图谱比较, $P < 0.01$),29 批黄花白及的指纹图谱相似度 0.7~0.88(平均值 0.822,与白及共有模式图谱比较, $P < 0.05$),说明根据 HPLC 指纹图谱相似度结果,白及饮片和黄花白及饮片能够明显区分。

综合 2020 年版《中国药典》及地方标准辨识、HPLC 指纹图谱相似度辨识结果及原始采购信息,二分类中 1~45 号样本为白及,46~134 号样本为非白及;四分类中 1~45 号样本为白及,46~75 号样本为天麻,76~105 号样本为玉竹,106~134 号样本为黄花白及。

3.3 电子鼻对白及及其近似饮片的定性判别

3.3.1 二分类 PCA-DA 辨识 当选择 5 个主成分时,模型分类错误率最低,前 5 个主成分解释变异之和达 99% 以上,能解释样本大部分变异信息。45 个白及饮片均被正确分类,在 89 个非白及饮片中,56、57 号天麻和 76、97 号玉竹被错误分类,模型主成分得分图见增强出版附加材料,两类饮片在二维空间有重叠情况,说明白及饮片与非白及饮片在前 2 个主成分所表征的变异信息上有相似之处,模型交互验证正判率 97.01%,不存在未分类样本,说明模型区分度好,可用于白及饮片与非白及饮片的辨识。

3.3.2 二分类 PLS-DA 辨识 当潜变量数为 6 个时,模型性能达到最优,前 6 个潜变量可以解释样本 99% 以上的变异信息。PLS-DA 模型潜变量得分图详见增强出版附加材料,白及样本与非白及样本有少量重叠,说明两类样本在前 2 个潜变量所表征的



注:A.白及;B.天麻;C.玉竹;D.黄花白及

图1 各类饮片HPLC指纹谱的共有模式

Fig. 1 Common patterns of HPLC fingerprint of four kinds of samples

变异信息上有部分相似之处。45个白及样本中有2个被判定为非白及,非白及饮片中有2个被错判为白及,模型交互验证正判率97.01%,无未分类样本,说明该PLS-DA模型也可实现白及饮片与非白及饮片的辨识。

3.3.3 二分类LS-SVM辨识 以留一法交互验证后的正判率为指标,优化LS-SVM的核函数类型,在二分类定性判别模型中,分别对线性核函数、多项式核函数及径向基核函数进行优选,3种核函数运行后的交互验证正判率结果显示,线性核函数正判率96.27%,多项式核函数正判率94.78%,径向基核函数正判率98.51%。其中径向基核函数的模型交互验证正判率最大,此时模型对134个样本的判别结果显示,45个白及样本中有2个被误判为非白及,89个非白及样本均被正确分类,模型未分类样本数量为0个。

3.3.4 二分类KNN辨识 当最邻近值数量(k)为1、2、4个时,模型错误率均最低,为0.033。以交互验证正判率及模型参数为指标,进一步优选 k ,详见表2。结果发现当 k 为1或2时,模型正判率较高,达97.01%。

3.3.5 四分类PCA-DA辨识 当主成分数为6个时,PCA-DA模型性能最优,前6个主成分能解释样本99%以上的变异信息。45个白及样本中有1个

表2 不同 k 值时的交互验证正判率及模型参数

Table 2 Positive discrimination rates and model parameters for cross validation at different k values

k	正判率/%	精确度	灵敏度	特异性值
1	97.01	0.955 6	0.955 6	0.977 5
2	97.01	0.955 6	0.955 6	0.977 5
4	96.27	0.916 7	0.977 8	0.955 1

被误判为黄花白及,30个天麻样本和29个黄花白及样本均被正确分类,30个玉竹样本中有1个被误判为白及,1个被误判为天麻。由模型主成分得分图(增强出版附加材料)可知,白及与天麻、玉竹、黄花白及均存在重叠现象,与玉竹重叠部分最多,说明白及与天麻、玉竹、黄花白及在该模型前两个主成分所代表的变异信息上有相似之处,与玉竹饮片的相似最多。玉竹样本和黄花白及样本没有重叠,表明这2种饮片在前2个主成分所表征的大部分变异信息上无相同之处。该模型不存在未分类样本,交互验证正判率97.76%,因此,可以用做四类饮片的区分。

3.3.6 四分类PLS-DA辨识 当潜变量数为5个时,模型性能最优,前5个潜变量能解释99%以上的样本变异信息。45个白及样本中有2个被误分类为黄花白及,有2个样本未分类,天麻、玉竹、黄花白及

样本辨识结果中均无误分类的现象,但分别有3、5、2个样本未分类,将未分类样本看作模型判错样本,所建PLS-DA模型交互验证正判率89.55%。由模型潜变量得分图(增强出版附加材料)可知,白及样本与其他三类样本均有重叠现象,与玉竹重叠部分最多。但由于存在未分类样本,因此该模型对白及饮片与其他3种饮片的四分类辨识性能还有待改进。

3.3.7 四分类LS-SVM辨识 在构建四分类辨识模型时,优选核函数过程同3.3.3项下,3种核函数运行后的交互验证正判率结果显示,线性核函数正判率98.51%,多项式核函数正判率无结果,径向基核函数正判率97.76%。当核函数类型为线性核函数时,模型交互验证正判率最大,此时判别结果混淆矩阵见表3。

表3 LS-SVM模型的留一法交互验证判别

Table 3 Classification of LS-SVM model with leave-one-out cross validation

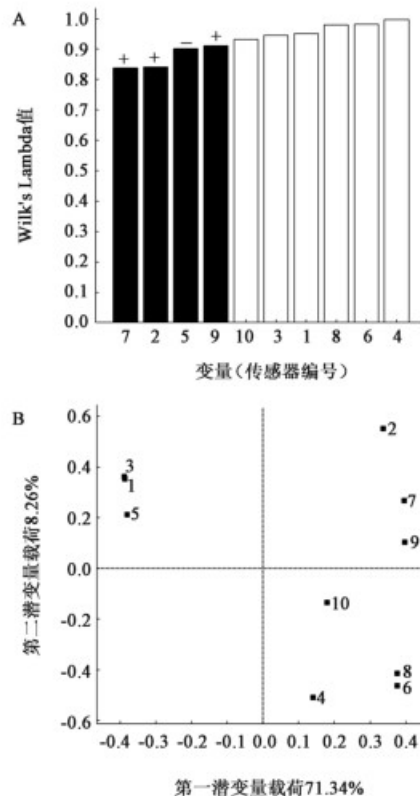
饮片	白及/个	天麻/个	玉竹/个	黄花白及/个
白及	45	0	0	0
天麻	0	29	1	0
玉竹	0	1	29	0
黄花白及	0	0	0	29

3.3.8 四分类KNN辨识 当k为1或2时,模型的错误率均最小,为0.028,此时交互验证正判率和模型参数均相同(模型正判率97.01%,精确度0.9556,灵敏度0.9556,特异性值0.9775)。因此,k优选为1或2,四分类辨识模型正判率97.01%。

综上所述,二分类辨识中4个模型均没有出现未分类样本,其中LS-SVM模型的正判率最高,为98.51%,因此,选择该模型作为PEN3型电子鼻辨识白及其近似饮片的最优二分类模型。在四分类辨识中,仅PLS-DA模型出现未分类样本,LS-SVM模型仍是最优模型,其正判率98.51%,故选择该模型作为PEN3型电子鼻辨识白及其近似饮片的最优四分类模型。

3.4 PEN3型电子鼻各嗅觉传感器(原始变量)对模型贡献度分析 综合PEN3型电子鼻的辨识结果,二分类辨识模型选择PLS-DA模型,四分类辨识模型选择PCA-DA模型。PEN3型电子鼻10根传感器所携带变异信息的Wilk's Lambda值见图2A和图3A,Wilk's Lambda值越小,表明变量携带变异信息越多。由图2A可知,二分类辨识模型中2号、

5号、7号、9号传感器携带变异信息较多,对模型分类贡献度较大,4号传感器携带变异信息最少,对白及和非白及饮片的分类贡献度最小。结合图2B可知,2号、7号、9号为正相关变量,5号为负相关变量,4号传感器在潜变量1上的载荷值最小,虽然在潜变量2上的载荷值较大,但由于潜变量1解释了样本大部分的变异信息(71.34%),因此,4号变量所携带的变异信息仍较少,对两类样本的区分贡献度不大。由图3A可知,四分类辨识模型中3~6号传感器所携带的变异信息较多,2号传感器所携带变异信息最少。结合图3B可知,4号、6号传感器为正相关变量,5号、3号传感器为负相关变量,2号传感器距离潜变量载荷图原点较近,说明该传感器特征属性的波动对样本的分类贡献不大,其余9个传感器变量整体分布距离原点较远,说明均携带有显著的变异信息。



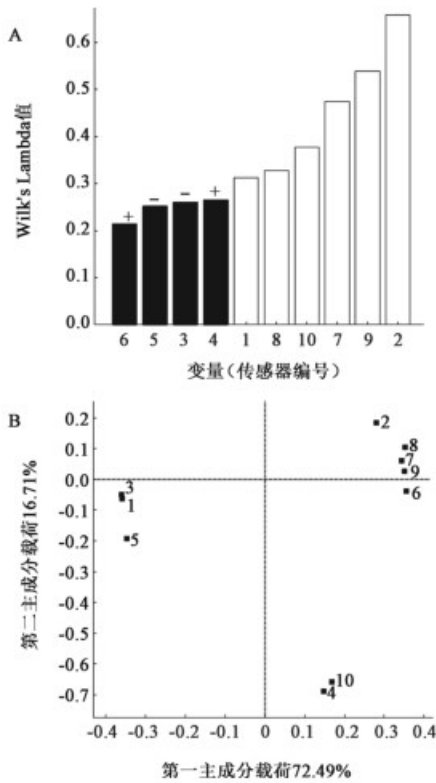
注:+,传感器信息值与结果呈正相关;-传感器信息值与结果呈负相关(图3同);A.Wilk's Lambda值柱状图;B.PLS-DA二分类辨识模型载荷图

图2 白及二分类辨识模型的变量贡献度分析

Fig. 2 Variable contribution analysis of binary identification model of Bletillae Rhizoma

4 讨论

4.1 模型判别结果及其机制分析 在样本二分类



注:A.Wilk's Lambda 值柱状图;B.PCA-DA 四分类辨识模型载荷图

图3 白及四分类辨识模型的变量贡献度分析

Fig. 3 Variable contribution analysis of quadratic identification model of Bletillae Rhizoma

辨识结果中,PCA-DA 模型和 PLS-DA 模型显示出了相同的判别能力,两者正判率均为 97.01%,无未分类样本;然而,在四分类辨识中,PLS-DA 模型有 12 个未分类样本,正判率较低,在本课题组前期的电子鼻辨识川贝母研究中,PLS-DA 模型在进行多分类辨识时也出现了未分类情况^[18],笔者推测这与模型的判别机制有关,相较于 2 种类别,将标杆信息划分为 4 种类别会使 PLS-DA 模型的辨识标准更加严苛,从而导致模型无法区分某些特征性不强的样本。PCA-DA 和 PLS-DA 分别是基于主成分回归与偏最小二乘回归的判别分析方法^[25],其中 PCA-DA 能够简化多维数据中相互重叠的样本信息,因此其对于多分类情况中某些成分影响较小的分类更适用^[26];PLS-DA 算法允许确定指定样本是否属于特定的预定义类别^[27],在一定程度上具有主观性,因此推测未分类样本与这种模型机制有关。

本实验中对 LS-SVM 的核函数进行了优选,白及、天麻、玉竹、黄花白及样本所含挥发性成分不同,因此电子鼻响应数据难以具有规律性,导致模型在对样本数据进行转换时并不全是线性的,也有

非线性存在的可能。LS-SVM 借助其核函数能较好地处理非线性数据的分类问题。常见核函数有线性核函数、多项式核函数和径向基核函数,其中多项式核函数具有较强全局特征表现能力,但学习能力较弱;径向基核函数具有局部特征表现能力,泛化能力弱,但学习能力较强^[28]。本实验的 LS-SVM 模型在二分类辨识中,使用径向基核函数时的正判率最高,使用多项式核函数时正判率最低;在四分类辨识中,多项式核函数甚至未能计算出判别结果,这是由于其参数多,当样本种类较多时,多项式的阶数较高导致计算复杂度过大而无法计算。当选择线性核函数时,四分类辨识的 LS-SVM 模型结果最优,线性核函数的特点是简洁,对于线性可分的情况其分类效果较好,笔者推测是由于将样本分为四类较将样本分为两类更具有线性特征,因为在将天麻、玉竹、黄花白及 3 种饮片的三类不同特征规划为 1 种特征的过程,很难实现线性,而四类样本的特征分别映射,更适合线性核函数简洁、线性可分的特点。

4.2 电子鼻传感器响应分析 PEN3 型电子鼻的原理是利用气体与传感器相互作用,使传感器活性材料导电性发生变化,从而产生响应值,传感器是由 10 种金属氧化物半导体型化学传感元件组成,不同传感器对应的敏感物质不同,响应值也不同。由图 2 和图 3 的传感器追踪可知,对样本二分类结果贡献度较大的原始变量是 2 号、5 号、7 号、9 号传感器,对样本四分类结果贡献度较大的原始变量是 3~6 号传感器,结合原始数据,在二分类辨识中,2 号、5 号、7 号、9 号传感器对应白及和非白及饮片的响应值差异均具有统计学意义($P < 0.01$);在四分类辨识中,虽然 3 号、6 号传感器对应白及和天麻饮片响应值差异无统计学意义,5 号传感器对应白及、天麻、玉竹饮片响应值差异无统计学意义,4 号传感器对应白及和黄花白及饮片响应值差异无统计学意义,但模型辨识结果并不是靠单一传感器,而是多个传感器的响应值经过运算后的综合结果。由 PEN3 型电子鼻各传感器对应的敏感物质可知,这些传感器响应的物质主要为无机硫类、氮氧化合物、烯烃芳香类、有机硫类、甲烷等短链烷烃、氨及芳香类、氢氧化物。结合文献研究,白及挥发油中含有酸、酯、醇、醛、多环烃及少量含氮化合物等^[29],天麻挥发油中含有 3-糠醛、甲氧基乙酸酐、N-甲基-2-吡咯甲酰胺等^[30],玉竹挥发性成分中含有酸、酯、饱和及不饱和烷烃、芳香族化合物等^[31],说明电子鼻对饮片辨识

是基于一定物质基础的。

4.3 误分类及未分类样本分析 PLS-DA四分类辨识模型中的未分类样本为35、44、52、53、57、76、80、83、84、103、129、133号样本,在这些未分类样本中,有部分样本在其他辨识模型中也存在误分类现象。例如,44号白及在PLS-DA和LS-SVM二分类模型中均被误分类,57号天麻和76号玉竹在PCA-DA和PLS-DA二分类模型中均被误分类,83号玉竹在PCA-DA四分类模型中被误分为天麻。因此,推测这些样本在包装、运输或储藏过程中由于温度、湿度变化导致其挥发性成分发生改变,从而使这些样本本身包含的分类信息与其同类饮片不同或不够丰富,易被模型错误识别。PLS-DA虽有降维、简化样本信息作用,但其机制是将数据转换成1组线性潜在变量用于分类变量预测,通过已知类别作为训练集建立分类模型,对未知样本进行预测^[32-33],训练集足够大时,未分类样本可能对2种辨识模型的性能影响不大,但当训练集较少时,未分类样本会干扰模型的判定边界,解决方法是增大训练集、标准化样本处理过程、减小组间差异等。

4.4 多源数据融合分析 多源数据融合将来自多个数据源的信息作为研究对象,通过化学计量学手段或机器学习方法实现对研究对象特定属性的精准识别,能够获得多维度、更全面的样本特征,提高模型对样本信息的判别和预测能力^[34-35]。中药饮片种类众多,成分复杂,对于含挥发性成分少或不含挥发性成分的饮片,单一电子鼻辨识无法获取样本更全面的信息,而联合使用其他智能感官如电子眼、电子舌等,通过采集样本视觉、味觉信息,构建二源或多源智能感官融合模型,能获得更准确的辨识结果^[36]。但在数据融合时应注意原始数据的量级大小,避免数量级的差异减慢算法迭代收敛速度及量级较大的属性占主导地位的现象,因此数据预处理方法的选择也很重要,应结合数据集自身特点(偏度是否较大、是否引入哑变量)、模型种类、问题类型(回归或是分类)等^[37],综合分析并选取1种适宜的预处理方法甚至多种方法联合使用。因此,后期可尝试通过融合其他类型电子鼻或超快速气相电子鼻、电子眼、电子舌、光谱数据、代谢组学数据等来对中药饮片进行辨识,为饮片质量评价提供参考。

在本实验中,二分类辨识和四分类辨识均以LS-SVM模型的正判率最高,且没有未分类样本,是PEN3型电子鼻辨识白及及其近似饮片的最优模

型,其余模型的正判率也均>89%。总体来看,基于饮片中挥发性物质,PEN3型电子鼻能够实现对白及及其近似饮片的相对准确、快速的辨识,可为中药饮片的定性辨识提供新思路和新方法,并为其他饮片的质量评价提供参考。

【利益冲突】 本文不存在任何利益冲突。

【参考文献】

- [1] 戴鸥,仰莲,周勤梅,等. 白及化学成分分离鉴定[J]. 中国实验方剂学杂志, 2018, 24(14): 43-47.
- [2] 朱峻霄,林亚蒙,杨野,等. 白及多糖在生物医药材料领域中的应用研究进展[J]. 中药材, 2018, 41(4): 1011-1014.
- [3] 朱珊梅,陈海鹰,范正达. 白及胶的制剂学研究进展[J]. 中国现代应用药学, 2019, 36(24): 3130-3135.
- [4] 蒋俊,陈红霞,汤兴利,等. 基于中医药美白理论的白及研发思考[J]. 中草药, 2017, 48(11): 2313-2320.
- [5] 张曼,韩亭亭,胡春芳,等. 白及产业现状及可持续发展策略[J]. 中草药, 2019, 50(20): 5103-5108.
- [6] 翟萌. 白及与其混淆品、伪品的生药学对比研究[D]. 成都:成都中医药大学, 2012.
- [7] 翟萌,马逾英,郑光雅,等. 市售白及与三种混淆品的对比鉴别[J]. 成都中医药大学学报, 2012, 35(3): 53-56.
- [8] 朱新焰,张彬若,杨竹雅,等. 白及与2种混淆品的对比鉴别研究[J]. 西南农业学报, 2019, 32(9): 2167-2173.
- [9] 李明华,程显隆,李宁新,等. 白及的质量问题及真伪鉴别方法[J]. 中国药事, 2018, 32(11): 1490-1499.
- [10] 鄢玉芬,李峰庆,刘珈羽,等. 近红外光谱技术鉴别白及粉及其混淆品[J]. 天然产物研究与开发, 2018, 30(7): 1219-1225.
- [11] 刘珈羽,李峰庆,郭换,等. 白及粉品种近红外快速定性鉴别模型的建立[J]. 成都中医药大学学报, 2018, 41(1): 34-37.
- [12] 赵丹,周涛,江维克,等. 基于ITS2序列SNP位点鉴定白及药材及其混淆品[J]. 中国中药杂志, 2015, 40(18): 3573-3578.
- [13] 陈美君,李峰庆,吕蒙,等. 白及与其混淆品ITS2序列二级结构比较与鉴别[J]. 中国实验方剂学杂志, 2017, 23(15): 46-52.
- [14] 迟明艳,黄勇,李勇军,等. UPLC结合化学计量学方法的白及指纹图谱分析[J]. 中国实验方剂学杂志, 2016, 22(14): 51-56.
- [15] 秦亚东,汪荣斌,方凤满,等. 基于化学计量学黄花白及和小白及HPLC指纹图谱分析[J]. 井冈山大学学报:自然科学版, 2020, 41(4): 21-25.
- [16] 刘涛涛,代悦,于森,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.
- [17] WANG T S, CHAO Y P, YIN F Z, et al. An E-nose and convolution neural network based recognition method for processed products of *Crataegi Fructus* [J]. *Comb Chem High Throughput Screen*, 2021, 24(7): 921-932.
- [18] 冯文豪, 田亮玉, 施钧瀚, 等. 电子鼻技术应用于川贝母真伪及规格辨识的可行性分析[J]. 中国实验方剂学杂志, 2021, 27(13): 108-118.
- [19] 黄得栋, 何微微, 晋玲, 等. 基于电子鼻技术区分不同产地的南五味子[J]. 中国实验方剂学杂志, 2017, 23(23): 22-26.
- [20] 拱健婷, 李莉, 丛悦, 等. 基于电子鼻和HS-GC-MS分析温郁金源3种中药气味差异[J]. 现代中药研究与实践, 2021, 35(4): 6-10.
- [21] 四川省食品药品监督管理局. 四川省中药饮片炮制规范[M]. 成都: 四川科学技术出版社, 2015: 5.
- [22] 甘肃省食品药品监督管理局. 甘肃省中药炮制规范[M]. 兰州: 甘肃文化出版社, 2009: 12-13.
- [23] 甘肃省食品药品监督管理局. 甘肃省中药材标准[M]. 兰州: 甘肃文化出版社, 2009: 13-14.
- [24] 刘瑞新, 郝小佳, 张慧杰, 等. 基于电子眼技术的中药川贝母真伪及规格的快速辨识研究[J]. 中国中药杂志, 2020, 45(14): 3441-3451.
- [25] 朱向荣, 李高阳, 黄绿红, 等. 近红外光谱与化学计量学方法用于镉污染稻米的定性鉴别[J]. 分析化学, 2015, 43(4): 599-603.
- [26] SUHANDY D, YULIA M, ABDULLAH A G, et al. The classification of arabica gayer coffee using UV-visible spectroscopy and PCA-DA method [J]. *MATEC Web Conf*, 2018, doi: 10.1051/mateconf/201819709002.
- [27] BORRAZ-MARTÍNEZ S, BOQUÉ R, SIMÓ J, et al. Development of a methodology to analyze leaves from *Prunus dulcis* varieties using near infrared spectroscopy [J]. *Talanta*, 2019, 204: 320-328.
- [28] 汤占军, 刘萍兰, 蒋鹏程, 等. 基于动力学与混合核函数LS-SVM的厌氧发酵产气量预测模型研究[J]. 安全与环境学报, 2020, 20(1): 277-282.
- [29] 金智伟, 杨维雄, 尹建华, 等. 响应曲面法优化白及挥发性成分提取工艺[J]. 桉树科技, 2021, 38(3): 47-53.
- [30] 邱红燕, 周雪, 吴林菁, 等. 天麻挥发性成分的顶空气质联用分析[J]. 时珍国医国药, 2019, 30(10): 2368-2369.
- [31] 权美平. 玉竹挥发油研究进展[J]. 粮食与油脂, 2019, 32(3): 3-5.
- [32] WANG Q Q, HUANG H Y, WANG Y Z. Geographical authentication of *Macrohyporia cocos* by a data fusion method combining ultra-fast liquid chromatography and Fourier transform infrared spectroscopy [J]. *Molecules*, 2019, 24(7): 1320.
- [33] YAO S, LI T, LIU H G, et al. Traceability of Boletaceae mushrooms using data fusion of UV-visible and FTIR combined with chemometrics methods [J]. *J Sci Food Agric*, 2017, 98(6): 2215-2222.
- [34] XU M, WANG J, ZHU L Y. The qualitative and quantitative assessment of tea quality based on E-nose, E-tongue and E-eye combined with chemometrics [J]. *Food Chem*, 2019, 289: 482-489.
- [35] BURATTI S, MALEGORI C, BENEDETTI S, et al. E-nose, E-tongue and E-eye for edible olive oil characterization and shelf life assessment: A powerful data fusion approach [J]. *Talanta*, 2018, 182: 131-141.
- [36] LAN Z W, ZHANG Y, SUN Y, et al. A mid-level data fusion approach for evaluating the internal and external changes determined by FT-NIR, electronic nose and colorimeter in *Curcuma Rhizoma* processing [J]. *J Pharm Biomed Anal*, 2020, 188: 113387.
- [37] 李颜平, 吴刚. 基于典型数据集的数据预处理方法对比分析[J]. 沈阳工业大学学报, 2022, 44(2): 185-192.

[责任编辑 刘德文]