基于近红外光谱技术结合深度学习快速鉴别滇黄精干燥方法和产地

胡晓燕[1,2]，苏俊宇[1]，沈　涛[3]，杨绍兵[1*]，王元忠[1*]

1. 云南省农业科学院 药用植物研究所，云南 昆明　650200
2. 云南中医药大学中药学院，云南 昆明　650500
3. 玉溪师范学院化学生物与环境学院，云南 玉溪　653100

摘　要：目的　探讨傅里叶变换近红外（fourier transform near infrared，FT-NIR）光谱技术在滇黄精 *Polygonatum kingianum* 质量评价中的应用，基于不同干燥处理方法、不同产地滇黄精的一维光谱与二维光谱构建鉴别模型，为滇黄精质量评价提供新方法。**方法**　研究采集了 198 份不同干燥方式和产地的滇黄精样品，将其进行不同干燥处理，分析产地、干燥处理方法等因素对滇黄精质量的影响；采集并分析滇黄精 FT-NIR 光谱和二维相关光谱（two-dimensional correlation spectroscopy，2DCOS）特征，采用主成分分析（principal component analysis，PCA）和层次聚类分析（hierarchical clustering analysis，HCA）可视化样品的分类趋势，并基于 FT-NIR 光谱建立偏最小二乘判别分析（partial least squares discriminant analysis，PLS-DA）和支持向量机（support vector machine，SVM）2 种传统模型；基于同步、异步和综合 2DCOS 图像建立残差神经网络（residual neural network，ResNet）模型，比较 3 种模型的整体鉴别能力和模型性能。**结果**　同步 2DCOS 图像结合 ResNet 可以成功鉴别滇黄精不同产地和干燥处理方法，准确率达到 100%，整体鉴别能力和模型性能均显著优于 PLS-DA 和 SVM 传统模型。**结论**　基于 FT-NIR 光谱构建的 ResNet 鉴别模型可以快速准确鉴别滇黄精的产地和干燥处理方法，为滇黄精质量的快速检测提供科学依据。

关键词：滇黄精；二维相关光谱；残差神经网络；干燥方法；产地

中图分类号：R286.2　　**文献标志码：**A　　**文章编号：**0253 - 2670(2025)18 - 6761 - 12

DOI: 10.7501/j.issn.0253-2670.2025.18.025

Rapid identification *Polygonatum kingianum* of drying method and geographical origin based on near infrared spectroscopy combined with deep learning

HU Xiaoyan[1,2], SU Junyu[1], SHEN Tao[3], YANG Shaobing[1], WANG Yuanzhong[1]

1. Institute of Medicinal Plants, Yunnan Academy of Agricultural Sciences, Kunming 650200, China
2. College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China
3. College of Chemistry, Biology and Environment, Yuxi Normal University, Yuxi 653100, China

Abstract: Objective　To explore the application of Fourier transform near infrared (FT-NIR) spectroscopy technology in the quality evaluation of Dianhuangjing (*Polygonatum kingianum*). Based on different drying treatment methods and one-dimensional and two-dimensional spectra of *P. kingianum* from different origins, a discrimination model was further constructed to provide a new method for the quality evaluation of *P. kingianum*. **Method**　In this study, a total of 198 samples of *P. kingianum* from five regions of Yunnan Province were collected and subjected to different drying treatments in order to analyze the effects of factors such as different origins and drying treatments on the quality of *P. kingianum*. Collect and analyze FT-NIR spectra and two-dimensional correlation spectroscopy (2DCOS) features of *P. kingianum*, and then use principal component analysis (PCA) and hierarchical clustering analysis (HCA) to visualize the classification trend of the samples. Based on FT-NIR spectra, establish two traditional models: partial least squares discriminant analysis (PLS-DA) and support vector machine (SVM). Establish residual neural network (ResNet) models based on synchronous, asynchronous, and integrative 2DCOS images to compare the overall discrimination ability and model performance of the three models. **Results**　Synchronized 2DCOS images combined with ResNet can successfully identify the different origins and drying treatments of *P. kingianum*, with an accuracy of 100%, and the overall identification ability and model performance are

收稿日期：2025-03-02

基金项目：云南省重大科技专项计划（202502AS100009）；云南省农业基础研究联合专项重点项目（202401BD070001-018）

作者简介：胡晓燕，女，硕士研究生，研究方向为中药资源评价。E-mail: h2280493670@126.com

***通信作者：**杨绍兵，副研究员，从事中药栽培学研究。E-mail: ysb9-116@163.com

王元忠，研究员，硕士生导师，从事中药资源开发与利用研究。E-mail: boletus@126.com

significantly better than that of PLS-DA and SVM traditional models. **Conclusion** The ResNet identification model based on FT-NIR spectroscopy can rapidly and accurately identify the origin and drying method of *P. kingianum*, and provide a scientific basis for the rapid detection of the quality of *P. kingianum*.

Key words: *Polygonatum kingianum* Coll. et Hemsl.; two-dimensional correlation spectroscopy; residual neural network; drying methods; origin identification

滇黄精 *Polygonatum kingianum* Coll. et Hemsl.属百合科（Liliaceae A. L. Jussieu）黄精属植物，主要分布在云南、四川、贵州等地区。滇黄精具有丰富的药用和食用价值，是著名的药食同源物种[1]。其具有补气养阴、健脾、润肺、益肾的功效[2]。现代药理学研究表明滇黄精的主要化学成分含多糖、甾体皂苷、黄酮等，具有抗肿瘤、抗氧化和抗衰老作用。影响滇黄精质量的因素众多，如产地、品种、生长年限和加工方式等[3]。这些可变因素导致其化学成分发生变化，造成质量差异。随着现代食品加工技术的发展，滇黄精常被制成茶叶、干果零食或粉末，作为营养补充剂添加到食品中。常见的滇黄精加工方法包括阴干、烘烤、蒸制、酒蒸、九蒸九制等[4]。黄精生品含有较多的黏液质，对咽喉有一定刺激性，经反复蒸制之后，不仅可消除其刺激性及不良反应，还利于有效成分积累，增强补益作用，便于长期保存[5]。随着我国经济持续发展，市场对滇黄精的需求不断增加，且售卖品多为饮片或粉末。因此，亟待建立一种快速有效的分析方法协助鉴别不同干燥方法和产地的滇黄精样品，上述工作对保障中药质量和疗效、促进中药市场健康发展具有重要意义。

核磁共振、高效液相色谱、质谱等传统鉴别技术虽具有高灵敏度和准确性，但操作复杂耗时，且仪器设备购买费用昂贵。近红外光谱技术（near-infrared，NIR）是一种快速、简单和无损的分析工具，不涉及化学品使用或复杂的样品制备[6]，已广泛应用于纺织、农业、医药等领域[7]。化学计量学对于挖掘有价值的数据和构建高性能模型至关重要。传统化学计量学算法包括支持向量机（support vector machine，SVM）、偏最小二乘判别分析（partial least squares discriminant analysis，PLS-DA）等。随着大数据时代的发展，人工智能（artificial intelligence，AI）在数据分析方面具有不可替代的作用。深度学习（deep learning，DL）是现阶段人工智能技术中采用的主要研究方法，在图像分类和物体识别上具有独特优势[8]。该技术不需要复杂的光谱预处理，且在建模过程中无需人工提取特征，提高了建模效率。卷积神经网络（convolutional neural networks，CNN）是主要用于图像识别的深度学习算法，广泛应用于医学图像分析领域。残差神经网络（residual network，ResNet）作为一种优秀的 CNN 网络，具有良好的收敛性和准确性，能够解决传统卷积神经网络中梯度消失与梯度爆炸问题[9]。二维相关光谱（two-dimensional correlation spectroscopy，2DCOS）可以有效地解决光谱重叠，提高视光谱分辨率。因此，深度学习结合 2DCOS 建模比传统光谱结合化学计量学建模分类鉴别性能更优，两者相结合用于药用植物的鉴别可发挥各自优势，改善鉴别效果[10]。

目前，滇黄精研究主要集中于化学成分和药理作用方面，药用资源质量评价研究较少。此外，不同数据集的预处理和特征提取方法无法统一，需要大量时间去验证，降低鉴别效率。因此，本研究不考虑预处理和特征提取，直接利用原始光谱数据建模。通过 PLS-DA、SVM 和 ResNet 算法比较，探讨并验证深度学习结合 2DCOS 在滇黄精干燥方法和产地鉴别中是否具备优越性。研究为滇黄精不同干燥处理及地理来源样品的鉴别提供准确高效的现代化方法，并为药材质量快速评价提供科学依据。

1 仪器与材料

1.1 仪器

配有傅里叶变换附件的近红外光谱仪（Thermo Fisher Scientific Inc.，美国）；电热恒温烘箱（上海亿恒科学仪器有限公司）；FW-100 型高速万能粉碎机（天津市华鑫仪器厂）；100 目不锈钢筛（北京中西泰安技术服务有限公司）。

1.2 材料

本研究采集云南省、广西省和四川省共 198 份栽培滇黄精样品，所有样品均为 5 年生，产地环境信息见表 1；其中，广西省样品和四川省样品用于最优模型验证。所采样品由云南省农业科学院药用植物研究所张金渝研究员鉴定为滇黄精 *P. kingianum* Coll. et Hemsl.；同时将 168 份云南省样品进行 2 种干燥处理，广西省样品进行直接干燥处理，四川省样品进行蒸后干燥处理，作为外部验证集实验材料（表 2）。

表 1 滇黄精产地环境信息

Table 1 Environmental information of origin of *P. kingianum*

产地	经度/(°)	纬度/(°)	海拔/m	活动积温/℃	年均温/℃	日照时数/h	年均降水量/mm
云南保山	99.25	25.19	1 692	4 334.9	16.7	2 241	831
云南楚雄	101.48	25.01	1 822	4 564.0	16.8	2 697	757
云南昆明	102.54	24.76	1 985	4 502.8	15.5	2 703	897
云南临沧	99.78	23.46	1 092	7 189.5	20.3	2 057	974
云南曲靖	103.77	25.58	1 882	4 380.7	15.9	2 092	657
广西	109.60	25.79	158	5 881.5	19.0	1 109	1 926
四川	108.96	28.45	365	4 973.8	16.7	874	1 862

表 2 滇黄精样品信息

Table 2 Sample information of *P. kingianum*

产地	样本数量		总数
	直接干燥	蒸后干燥	
云南保山	14	14	28
云南楚雄	27	27	54
云南昆明	8	8	16
云南临沧	15	15	30
云南曲靖	20	20	40
广西	15	0	15
四川	0	15	15

所有样品分成 2 部分，一部分去除地上部分和须根后，将样品洗净、切片，然后在 55 ℃烘箱中烘至恒定质量。另一部分洗净后切片，然后均匀的铺在蒸锅的铺帘上，蒸至透心滋润时取出，最后将其放至 55 ℃烘箱中烘至恒定质量。随后，将其全部研磨成粉末，过 100 目筛，并在室温下储存在贴有标签的 PE 密封袋中，避光保存用于后续分析。

2 方法

2.1 FT-NIR 光谱采集

样品粉末使用 FT-NIR 光谱仪进行扫描。在积分球漫反射模式下，每次取约 5 g 样品置于样品杯中。扫描范围为 10 000～4 000 cm^{-1}，分辨率设置为 8 cm^{-1}，每个光谱累计扫描 32 次。每个样品平行测量 3 次，使用 OMNIC 软件获得平均光谱。为了消除空气信息的干扰，在每次扫描前，将实验室空气（H_2O 和 CO_2）光谱记录为背景并自动相减。光谱测定过程中，保持实验室温度 25 ℃，空气相对湿度 30%。最后，利用 SIMCA14.1 软件对采集的光谱数据进行分析，用于后续建模。

2.2 数据可视化分析

主成分分析（principal component analysis，PCA）是常用的无监督化学计量学方法，它将原始变量和相关变量转换为不相关变量，称为主成分（PCs），可视化分析样本的分类趋势[11]。层次聚类分析（hierarchical clustering analysis，HCA）是根据所有数据点的距离来确定每个类别的数据之间的差异，从而确定它们之间的相似度[12]。本研究基于云南省内不同产地和干燥方法样品的平均原始光谱信息，进行 PCA 与 HCA。PCA 采用 SIMCA 14.1 软件，HCA 采用 Origin 2021 软件操作。

2.3 2DCOS 光谱图像生成

一维光谱存在谱峰重叠和特征峰不明显的缺点，而广义 2DCOS 通过增加维数可以提高光谱分辨率，提取谱峰的特征信息。本研究广义 2DCOS 由离散广义二维相关光谱算法获得[13]。根据 Noda 理论，分析光谱强度随扰动的动态变化，利用相关性分析区分同步与异步的谱带响应，并最终绘制谱图[13]。

本研究中，所有样本通过 SIMCA 14.1 转换成光谱数据。利用 Matlab R2017a 软件自动生成同步、异步和综合 2DCOS 图像。在构建深度学习模型之前，根据 Kennard-stone（KS）算法划分数据集，每种干燥方法或产地的 60%样品作为训练集进行建模，30%的样本作为测试集，剩余 10%的样本作为外部验证集以验证模型性能。

2.4 模型建立

PLS-DA 是有监督的线性分类方法，通过提取样本之间的差异信息进行分类。采用敏感性（sensitivity，SEN）、特异性（specificity，SPE）、准确率（accuracy，ACC）、约登指数（Youden index，*J*）、精密度（precision，*P*）多个参数评估分类模型性能。SEN 表示正确预测目标类样本的能力，SPE 反映了拒绝与目标类别不匹配的样本的能力。*J* 表示模型检测真阳性或真阴性样本的总能力，是评价准确性的总结性指标[14]。这些参数值越接近 1，模型性能越好，错误率越低。通过假

阳性（false positive，FP）、假阴性（false positive，FN）、真阴性（true negative，TN）、真阳性（true positive，TP）计算参数[14]。

本研究采用 SIMCA 14.1 软件进行 200 次置换检验，基于 R^2 截距和 Q^2 截距结果，验证 PLS-DA 模型的鲁棒性和拟合程度[15]。通常，R^2 截距小于 0.4，Q^2 截距小于 0，表明模型没有过拟合的风险。本研究采用 SIMCA 14.1 软件建立 PLS-DA 模型，并使用 K-S 算法对数据集进行划分，不同干燥方法的数据集包括 70%训练集（118）和 30%测试集（50），不同产地的数据集分为 70%训练集（117）和 30%测试集（51）。

SVM 是有监督的非线性模式识别方法，能够识别未知样本。利用映射方法将样本数据映射到高维空间，求解最优超平面实现分类[16]。最优惩罚参数（c）和核参数（g）影响模型的分类准确率。c 值越大，模型的容错率越低，模型越容易过拟合，c 值越小，模型越容易欠拟合。g 值越大，支持向量越少，模型训练速度越快，g 值越小，模型训练速度越慢。本研究采用交叉验证进行网格搜索选择最优参数 c 和 g，利用 Matlab R2017a 软件建立 SVM 模型来鉴别不同干燥方法和产地的滇黄精。与 PLS-DA 模型类似，采用 K-S 算法划分训练集和测试集。

ResNet 通过将输入信息直接传递给输出，目标函数变为 $F(x)=H(x)-x$。当输入与输出的维度一致时，恒等模块被用来建立模型。如果输出 $F(x)$ 的维度与输入 x 不一致，则添加卷积核大小为 1×1 的卷积模块匹配输入与输出的维度。本研究建立了 12 层的 ResNet 模型，输入数据为同步、异步和综合 2DCOS 光谱图像。采用交叉熵损失函数曲线和精度曲线是评价模型的参数。其中，交叉熵损失函数来说明模型的收敛效果，其值越接近于零，模型的收敛性越好。通过训练集和测试集的准确率曲线来评价模型的识别能力，该值越接近 1，模型的识别能力越强。

3 结果与分析

3.1 FT-NIR 光谱分析

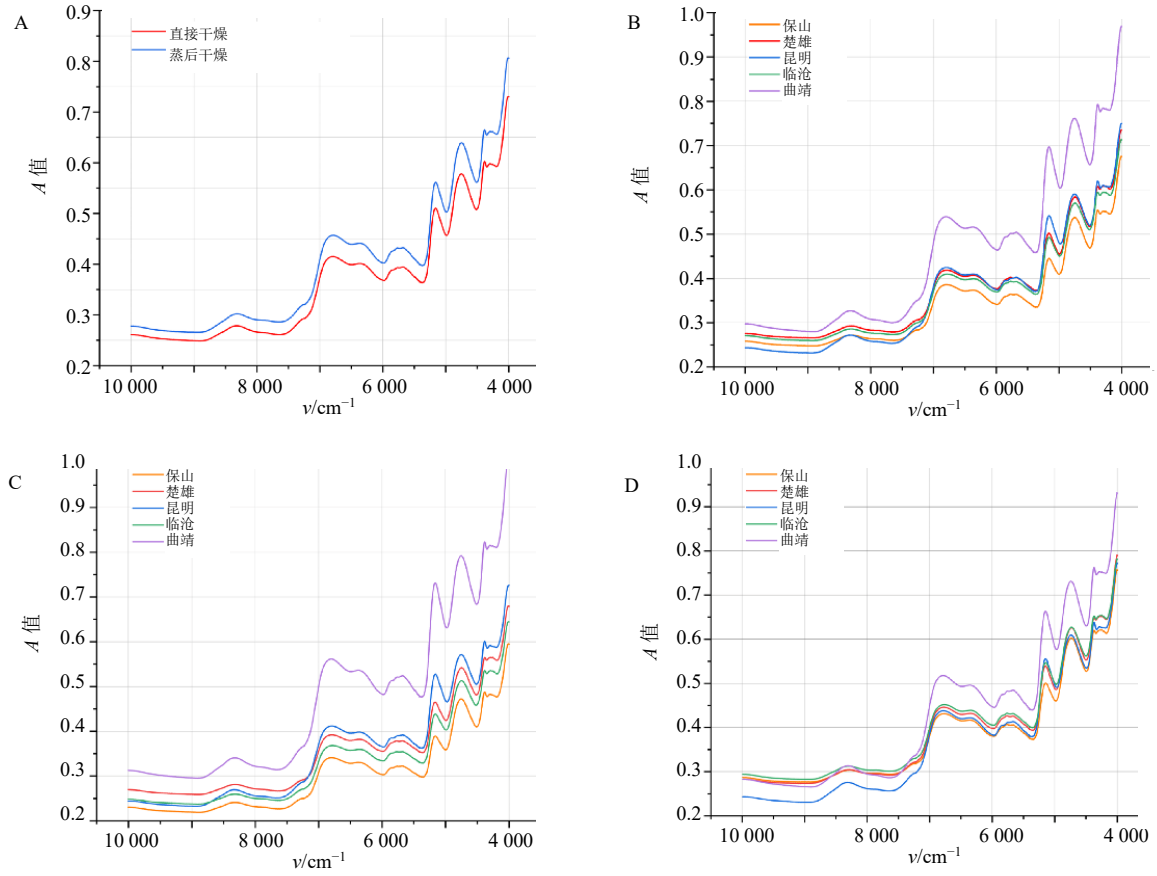
所有 FT-NIR 平均光谱如图 1 所示，其中图 1-A、B 分别代表了不同干燥方法和 5 个产地的平均光谱。图 1-C、D 显示了每种干燥方法的 5 个地区的平均光谱，包括直接干燥和蒸后干燥。综合来看，滇黄精平均光谱的特征吸收带主要表现在 8 500～8 000 cm^{-1} 和 7 000～4 000 cm^{-1} 2 个波段。CH_2 和 CH_3 基团 C-H 拉伸振动的第 2 个泛音在 8 380～8 230 cm^{-1} 附近产生了较大的宽带。位于 6 800、6 300、4 295 cm^{-1} 处的频谱峰分别代表了 O-H 和 C-H 拉伸振动的第 1 个泛音。在 5 170 cm^{-1} 附近有 1 个明显的峰值是 O-H 拉伸振动和 H-O-H 转化的多糖组合带。接近 4 750 cm^{-1} 的峰值是由 C-H 转变模式的第二泛音引起的，这归因于多糖、纤维素和生物碱。吸收带 4 400 cm^{-1} 属于 O-H 和 C-O 拉伸的结合带。

2 种干燥方法的平均光谱如图 1-A 所示。两者的吸收峰趋势相似，但各官能团引起的振动吸收强度相差很大。这表明 2 种干燥方法的滇黄精总体化学成分相似，但每个化学成分含量积累不同。样品蒸后干燥的吸收峰强度强于直接干燥。5 个地区的平均光谱如图 1-B 所示。在 10 000～4 000 cm^{-1} 内，曲靖的吸收强度最高，说明曲靖的滇黄精样品在该波数范围内具有较强的吸收信号。在 10 000～7 500 cm^{-1} 内，不同地区的吸收峰大小依次为曲靖＞楚雄＞临沧＞保山＞昆明，而在 7 500～4 000 cm^{-1} 内，昆明的吸收峰强度大于保山地区，同时在该波数下，曲靖、楚雄和临沧的吸收峰变化趋势和吸收强度相同，表明其化学成分含量相似。从图 1-C 和 D 中可以看出，不同产地和干燥方法的样品存在差异。在 7 000～4 000 cm^{-1} 下，曲靖直接干燥和蒸后干燥样品的吸收峰强度均明显高于其他地区。

3.2 主成分分析（principal components analysis，PCA）和 HCA

基于不同干燥方法 FT-NIR 光谱数据的 PCA 得分图结果如图 2-A 所示。从总体分析来看，不同干燥方法的样品很难用 FT-NIR 光谱特征进行区分。部分直接干燥与蒸后干燥的样品分离，但部分直接干燥的样品被蒸后干燥的样品点覆盖。结果表明，不同干燥方法的滇黄精样品的化学信息相似。主成分数（PC1 和 PC2）的前 2 个分量占总方差的 99.4%（87.3%和 12.1%），说明原始数据的大部分信息被保留了下来。基于不同产地 FT-NIR 光谱数据的 PCA 得分图结果如图 2-D 所示，不同地区的大多数样本聚在一起，表明不同产地之间的化学信息存在显著差异。

不同干燥方法和产地的 2 个主成分数的载荷图如图 2-B 和 E 所示。PC1 和 PC2 几乎强调了所有的光谱信号，呈现出与原始 FT-NIR 光谱接近的特征。在图 2-B 中，PC1 和 PC2 分别在 10 000～4 000 cm^{-1}



A-不同干燥方法；B-不同产地；C-直接干燥的不同地区；D-蒸后干燥的不同地区。

A-different drying methods; B-different regions; C-different regions with direct drying; D-different regions with drying after steaming.

图 1　滇黄精原始平均光谱图

Fig. 1　Original mean spectra of *P. kingianum*

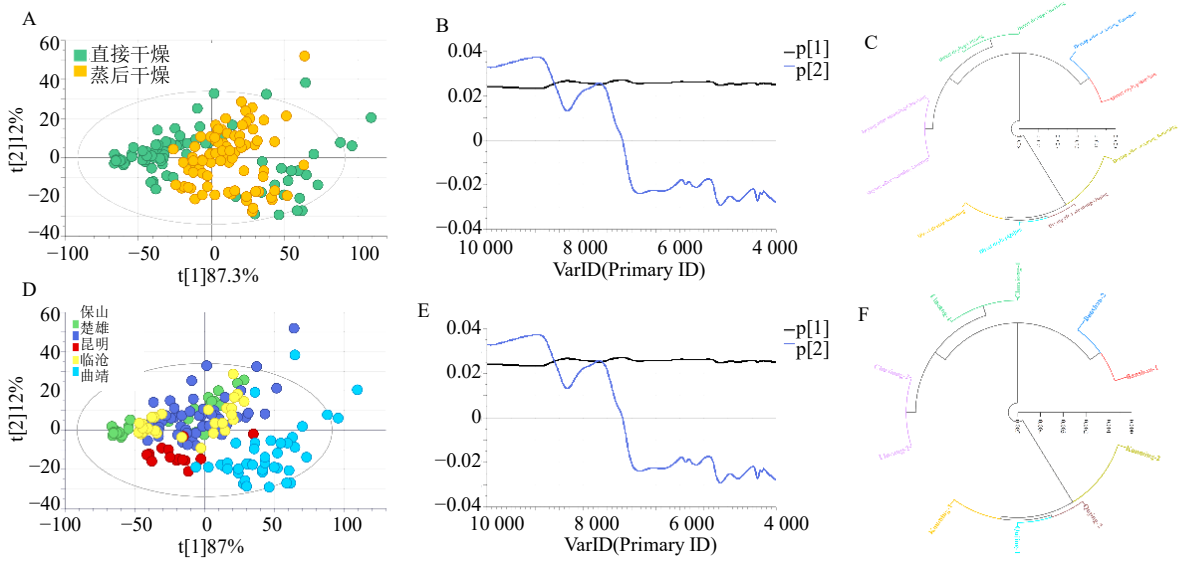


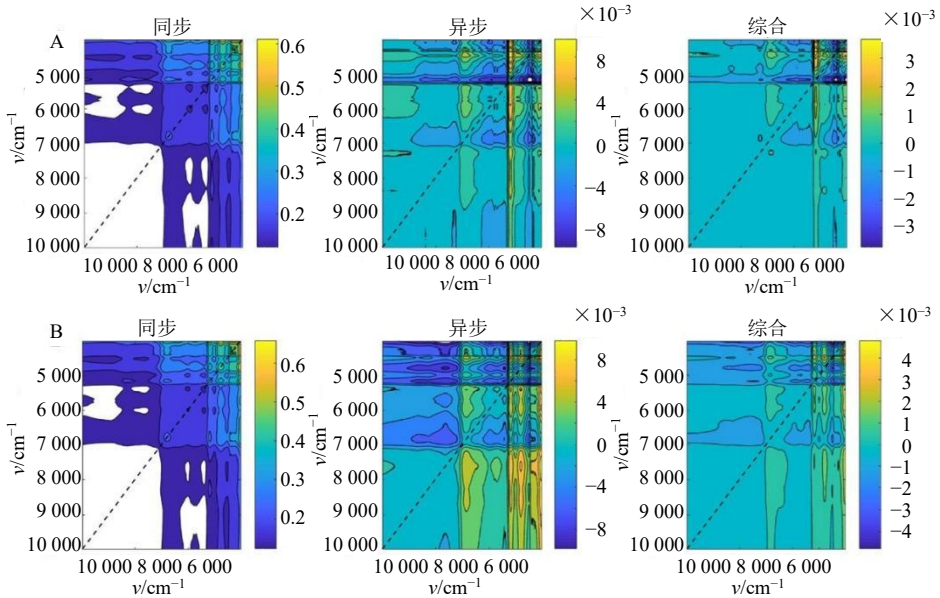
图 2　不同干燥方法和产地的 PCA 得分图 (A、D)、载荷图 (B、E) 和 HCA (C、F)

Fig. 2　Score plots (A and D); loading plots (B and E) of PCA and HCA (C and F) for different drying methods and regions

和 10 000～7 200 cm^{-1} 内对样本分布产生积极影响的变量。PC2 在 7 200～4 000 cm^{-1} 内对样本分布产生负面影响的变量。由于不同产地的 2 个主成分数占比总方差与不同干燥方法相同，因此 PC1 和 PC2 对样本分布产生积极或负面影响的变量的波数范围一致（图 2-E）。为了评估不同分类样本之间的相似性，建立了 HCA。不同干燥方法和产地的 HCA 结果如图 2-C 和 F 所示，不同干燥方法的聚类效果是混乱的，但 5 个地区的聚类效果较好。结果表明，不同地区的滇黄精样品存在明显差异。

3.3 2DCOS 光谱图像分析

本研究绘制了云南省 168 个样品，共 1008 张 2DCOS 图像，包括不同干燥方法（图 3）和不同产地（图 4）滇黄精的同步、异步和综合 2DCOS 图像。在同步 2DCOS 中，位于对角线上的峰被称为自动峰，反映了样品中每个化学基团在不同区域的相关光谱变化对扰动变量的敏感性。位于对角线两侧的峰称为交叉峰，反映了与其频率相对应的强度振动的相关性。异步 2DCOS 用于表征 2 个频谱信号之间的差异程度，在对角线两侧都是反对称的，完全由交叉峰组成，没有自动峰。综合 2DCOS 的特征比异步 2DCOS 清晰，可以提供分辨率相同的相关谱。从图 3、4 可以看出，同步 2DCOS 的峰值特征信息比较明显，而异步和综合 2DCOS 图像信息碎片化，光谱信息较少。此外，不同产地的 2DCOS 比不同干燥方法的 2DCOS 差异更大，这与一维光谱 FT-NIR 所呈现的结果一致。



A-直接干燥；B-蒸后干燥。

A-direct drying; B-drying after steaming.

图 3 不同干燥方法滇黄精样品的二维相关光谱图像

Fig. 3 Two dimensional correlation spectral images of *P. kingianum* samples from different drying methods

3.4 PLS-DA 建模分析

本研究基于 FT-NIR 光谱数据建立了不同干燥方法和产地的 PLS-DA 模型。模型的 R^2 截距均小于 0.4，不存在过拟合风险。表 3 显示了 PLS-DA 模型结果的参数，以此评价模型的性能。在不同干燥方法的 PLS-DA 模型结果中，所有参数值均大于 0.9，且基于直接干燥训练集和测试集的 SEN（0.95 和 0.96）高于蒸后干燥的样本（0.93 和 0.93），表明直接干燥模型正确预测目标样本的能力较强。同时在混淆矩阵（图 5-A）中，对于直接干燥方法的训练集和测试集，53 个样本中只有 3 个样本分类错误，27 个样本中只有 1 个样本分类错误，分类正确率分别达到 94.64%和 96.43%。此外在不同产地的 PLS-DA 模型结果中，5 个地区训练集的模型参数值均为 1，混淆矩阵的分类准确率达到 100%，表明模型具有较好的识别能力。但在不同产地的测试集（图 5-B）中，昆明地区的各参数值均低于 0.9，混淆矩阵的分类准确率较低。因此，需要建立其他模型提升整体识别能力。通过 2 种模型对比得出，无论是模型参数结果还是识别准确率，基于不同产地数据的 PLS-DA 模型优于不同干燥方法数据的 PLS-DA 模型。

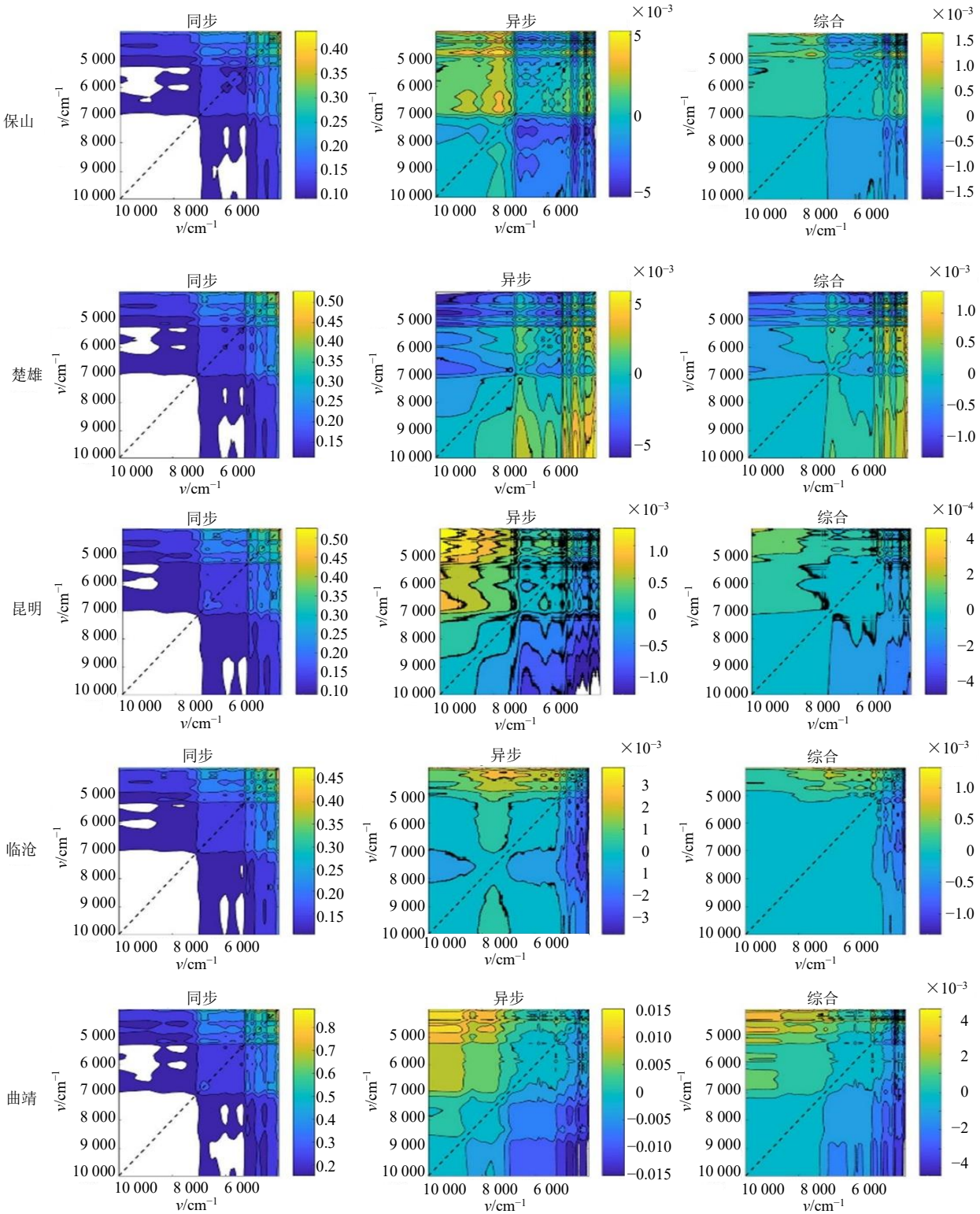


图 4　不同产地滇黄精样品的二维相关光谱图像

Fig. 4　Two dimensional correlation spectral images of *P. kingianum* samples from different origins

3.5　SVM 模型

表 4 显示了所有 SVM 模型的最优 *c* 和 *g*，以及训练集和测试集的准确率。虽然基于不同干燥方法和产地的分类准确率较高，这些模型中的所有样本被完全正确分类，除了不同干燥方法模型中 2 个样本被错误分类。但其 Best *g* 值（185 363.8 和 92 681.9）

表 3 不同干燥方法和不同产地建立的 PLS-DA 模型结果参数

Table 3 Parameters of PLS-DA established for different drying methods and regions

类别		训练集					测试集				
		SEN	SPE	*J*	*P*	ACC	SEN	SPE	*J*	*P*	ACC
不同干燥方法	直接干燥	0.95	0.93	0.88	0.93	0.94	0.96	0.93	0.89	0.93	0.95
	蒸后干燥	0.93	0.95	0.88	0.95	0.94	0.93	0.96	0.89	0.96	0.95
不同产地	保山	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	楚雄	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	昆明	1.00	1.00	1.00	1.00	1.00	0.83	0.83	0.67	0.88	0.83
	临沧	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	曲靖	1.00	1.00	1.00	1.00	1.00	0.93	0.93	0.86	0.94	0.93

A

训练集			
	直接干燥	蒸后干燥	准确率
直接干燥	53	3	94.64%
蒸后干燥	4	52	92.86%
总数	57	55	93.75%
测试集			
	直接干燥	蒸后干燥	准确率
直接干燥	27	1	96.43%
蒸后干燥	2	26	92.86%
总数	29	27	94.64%

B

训练集						
	保山	楚雄	昆明	临沧	曲靖	准确率
保山	19	0	0	0	0	100%
楚雄	0	36	0	0	0	100%
昆明	0	0	11	0	0	100%
临沧	0	0	0	20	0	100%
曲靖	0	0	0	0	27	100%
总数	19	36	11	20	27	100%
测试集						
	保山	楚雄	昆明	临沧	曲靖	准确率
保山	9	0	0	0	0	100%
楚雄	0	18	0	0	0	100%
昆明	0	0	4	1	0	66.67%
临沧	0	0	0	10	0	100%
曲靖	0	0	0	1	12	85.71%
总数	9	18	4	12	12	90.48%

图 5 不同干燥方法 (A) 和不同产地 (B) 的混淆矩阵

Fig. 5 Confusion matrix for different drying methods (A) and different origins (B)

表 4 不同干燥方法和不同产地建立的 SVM 模型结果参数

Table 4 Parameters of SVM established for different drying methods and regions

类别	Best *c*	Best *g*	准确率/%	
			训练集	测试集
不同干燥方法	185 363.8	0.000 172 63	98.21	96.43
不同产地	92 681.9	0.000 690 53	92.04	100.00

均异常高，说明这 2 类模型性能较差可能存在过拟合的情况，不能很好的鉴别不同干燥方法和产地的滇黄精。图 6 为不同干燥方法和产地的 SVM 模型最优超平面和分类结果图。此外，不同产地的 SVM 模型测试集的准确率达到 100%，表明鉴别不同地区比不同干燥方法更容易实现，结果与 PLS-DA 模型一致。综上，虽然 SVM 模型具有非线性、高维数据等优势，但本研究中未经预处理的数据不适用于 SVM 模型。

3.6 ResNet 模型

不同干燥方法和产地的 ResNet 模型的精度曲线和交叉熵损失函数结果如图 7 所示。此外，使用

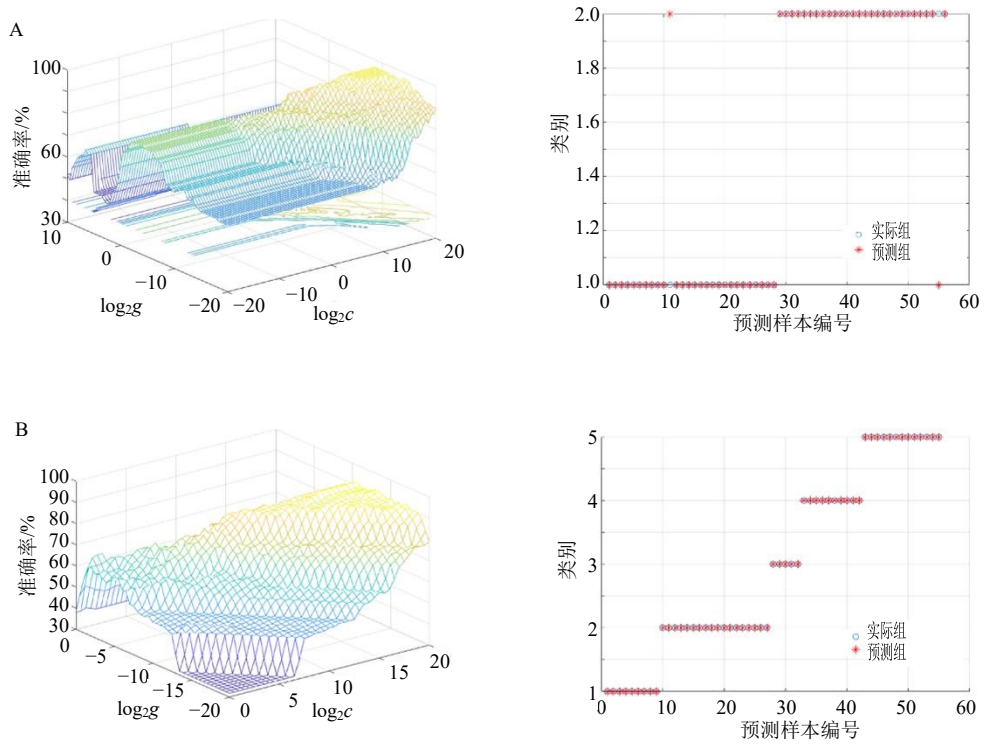


图 6　基于不同干燥方法 (A) 和不同地区 (B) 的 SVM 模型最优分离超平面图和分类结果

Fig. 6　Optimal separation hyperplanes and classification results of SVM models based on different drying methods (A) and different regions (B)

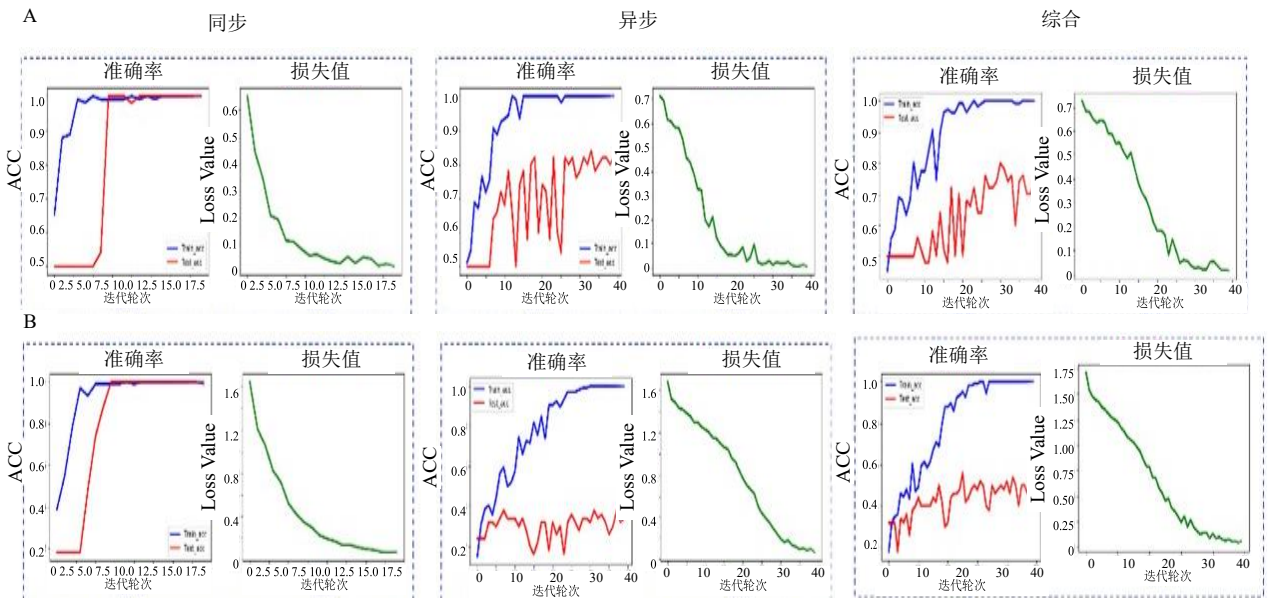


图 7　不同干燥方法 (A) 和产地 (B) 的 ResNet 模型的精度曲线和交叉熵损失函数

Fig. 7　Accuracy curves and cross-entropy cost function of ResNet model for drying methods (A) and regions (B)

建立的 ResNet 模型对外部验证集进行分类，不同干燥方法和产地的外部验证集的分类结果展示在图 8 的混淆矩阵中。在本研究中，如图 7-A 所示，当迭代轮次为 12 时，训练集和测试集的准确率均达到 100%，loss 值为 0.026。结果表明，该模型具有较强的稳定性，可用于识别不同的干燥方法。异步和综合

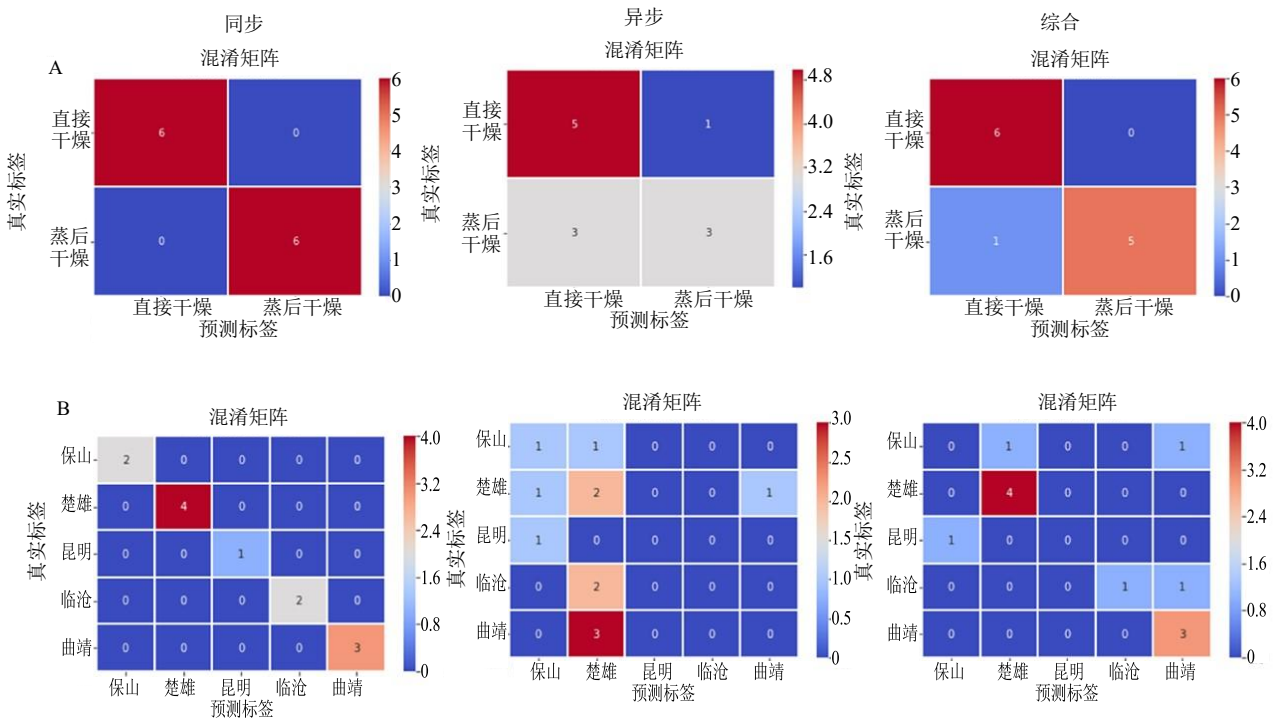


图 8　不同干燥方法 (A) 和产地 (B) 的 ResNet 模型的混淆矩阵

Fig. 8　Confusion matrix of ResNet model for drying methods (A) and regions (B)

2DCOS 图像在迭代轮次为 15 和 26 时，训练集的准确率达到 100%，但测试集的准确率未达到 100%，并且波动较大，说明模型的准确率和泛化能力较差。图 8-A 显示不同干燥方法的同步 2DCOS 所有样本分类正确，识别准确率为 100%。异步 2DCOS 中 4 个样本被错误分类，综合 2DCOS 中只有 1 个样本被错误分类。与同步 2DCOS 模型相比，异步和综合 2DCOS 模型的结果较差，这可能是因为同步 2DCOS 图像的颜色信息比异步和综合 2DCOS 光谱丰富，具有更清晰的特征峰，可以更好地表征化学信息。从图 7-B 同步 2DCOS 可以看出，当迭代轮次为 15 时，训练集和测试集的准确率为 100%，loss 值接近于零。异步和综合 2DCOS 图像模型测试集的准确率随迭代轮次的增加而波动，但未达到 100%。结果表明，当前还没有较好的异步和综合 2DCOS 模型用于滇黄精的产地鉴别。图 8-B 混淆矩阵结果发现同步 2DCOS 图像模型的所有样本分类正确，准确率达 100%。在异步 2DCOS 图像中，只有 3 个样本被正确分类，占样本总数的 25%，即使模型的泛化能力不是很好，但保山和楚雄地区的分类是正确的，均占 50%。而综合 2DCOS 图像的模型中楚雄和曲靖分类准确率达 100%。

综上，研究发现同步 2DCOS 光谱模型具有最好的鉴别效果，训练集、测试集和外部验证集准确率均为 100%，建模结果与图像直观分析一致。

3.7　最优模型验证

本研究将广西省直接干燥样品（15）与四川省蒸后干燥样品（15）作为外部验证集输入最优干燥方法鉴别模型，结果见图 9-A。同时，将广西省与四川省 2 个产地的样品划分为训练集（11）、测试集（2）、外部验证集（2），重新建立 ResNet 产地鉴别模型，混淆矩阵结果见图 9-B。结果表明，基于云南省不同干燥方法与产地样品 Synchronous-2DCOS 光谱建立的 ResNet 模型具有较好泛化能力，能有效识别广西省、四川省滇黄精样品的干燥方法。

4　讨论

本研究基于 FT-NIR 光谱构建了 3 种判别模型对不同干燥方法和产地的滇黄精进行鉴别研究。原始光谱与 2DCOS 光谱解析结果表明不同干燥方法和产地的滇黄精所含化学成分相似，但吸光度差异明显，表明不同干燥方法和产地的滇黄精化学成分含量不同。此外，干燥方法对样品光谱吸收峰强度和峰形的影响小于产地的影响作用，导致不同产地样品较不同干燥方法的样品更容易识别。侧面也暗示产地对滇黄精品质的影响可能大于干燥处理方

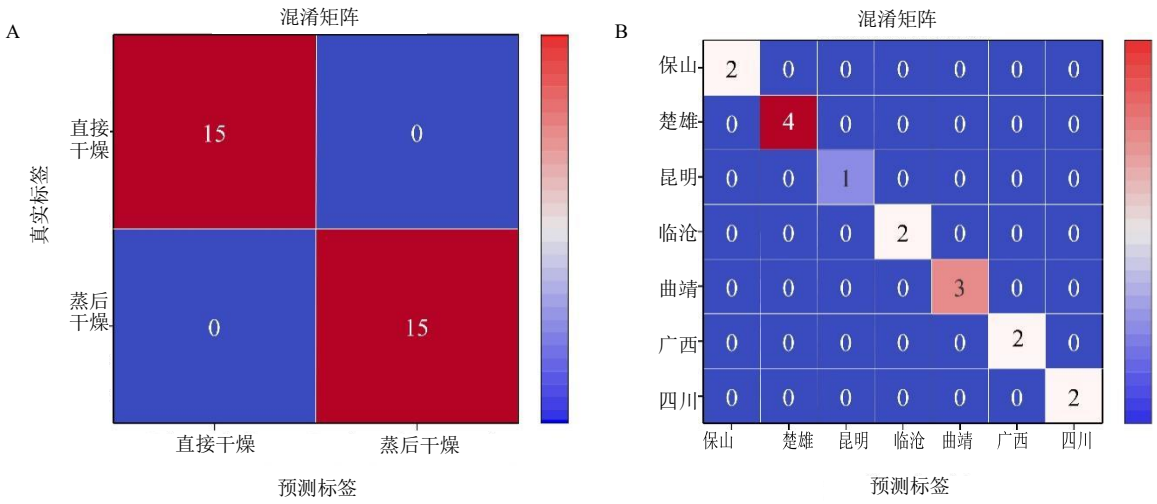


图 9　不同干燥方法 (A) 和产地 (B) 最优模型验证混淆矩阵

Fig. 9　Confusion matrix of optimal model verification for drying methods (A) and regions (B)

式。已有报道显示，海拔、产地年均温、年均降水等环境因子对滇黄精有效成分含量和药材品质有较显著的影响[17]。本研究样品地理来源涉及云南、四川和广西 3 省，云南境内，采样地又跨越了滇东、滇中及滇西地区，产地温度、降水、海拔等条件均有较大变化，这可能是导致不同产地样品 FT-NIR 光谱信息差异较大的原因，有待后续深入研究。

PLS-DA 和 SVM 是传统模式识别模型，在过去的研究中得到广泛的应用，ResNet 模型是深度学习中具有代表性的模型。目前已成功应用于中药和真菌，包括物种、储存期、地理可追溯性等[10]。本研究通过比较 3 种模型的鉴别能力和模型性能，发现 PLS-DA 模型鲁棒性与泛化能力较差，这可能与光谱数据未经预处理有关，光谱与预处理能有效提高 PLS-DA 模型性能，同时降低光谱采集过程中噪声和光散射影响，提高光谱分辨率[18-19]。SVM 模型的关键参数 Best *c* 值过高导致模型性能较差，可能存在过拟合的情况，Best *g* 则过低导致模型训练速度慢，增加时间成本[16]。而基于 FT-NIR 的同步 2DCOS 图像构建的 ResNet 模型具有省时、精度高、操作简单等优点[20]，不需要复杂的数据预处理便能快速准确鉴别滇黄精不同干燥方法和产地。

本研究基于 FT-NIR 光谱建立的滇黄精不同干燥方法和产地鉴别模型可为滇黄精质量评价提供新的技术支撑，为滇黄精实时快速检测和质量监控的实现提供了一种新的视角，为后期建立滇黄精质量数字化和智能化评价模式提供参考思路。

利益冲突　所有作者均声明不存在利益冲突

参考文献

[1] Xia M Q, Liu Y, Liu J J, *et al*. Out of the Himalaya-Hengduan Mountains: Phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere [J]. *Mol Phylogenet Evol*, 2022, 169: 107431.

[2] Zhao P, Zhao C C, Li X, *et al*. The genus *Polygonatum*: A review of ethnopharmacology, phytochemistry and pharmacology [J]. *J Ethnopharmacol*, 2018, 214: 274-291.

[3] Xu Y L, Zhang J Y, Wang Y Z. Recent trends of multi-source and non-destructive information for quality authentication of herbs and spices [J]. *Food Chem*, 2023, 398: 133939.

[4] 曾林燕, 宋志前, 魏征, 等. 黄精炮制过程中新产生成分分离及含量变化 [J]. 中草药, 2013, 44(12): 1584-1588.

[5] 秦宇雯, 张丽萍, 赵祺, 等. 九蒸九晒黄精炮制工艺的研究进展 [J]. 中草药, 2020, 51(21): 5631-5637.

[6] Stockl A, Lichti F. Near-infrared spectroscopy (NIRS) for a real time monitoring of the biogas process [J]. *Bioresour Technol*, 2018, 247: 1249-1252.

[7] Rossi G B, Lozano V A. Simultaneous determination of quality parameters in yerba mate (*Ilex paraguariensis*) samples by application of near-infrared (NIR) spectroscopy and partial least squares (PLS) [J]. *LWT*, 2020, 126: 109290.

[8] Richa, Patro B D K. Improved early detection accuracy for breast cancer using a deep learning framework in medical imaging [J]. *Comput Biol Med*, 2025, 187: 109751.

[9] Alaeddine H, Jihene M. Wide deep residual networks in

networks [J]. *Multimed Tools Appl*, 2023, 82(5): 7889-7899.

[10] Liu Z M, Yang S B, Wang Y Z, *et al*. Discrimination of the fruits of *Amomum* tsao-ko according to geographical origin by 2DCOS image with RGB and Resnet image analysis techniques [J]. *Microchem J*, 2021, 169: 106545.

[11] Zhou G S, Yuan Y C, Yin Y, *et al*. Hydrophilic interaction chromatography combined with ultrasound-assisted ionic liquid dispersive liquid-liquid microextraction for determination of underivatized neurotransmitters in dementia patients' urine samples [J]. *Anal Chim Acta*, 2020, 1107: 74-84.

[12] Brake F, Ross T, Holds G, *et al*. A survey of Australian oysters for the presence of human noroviruses [J]. *Food Microbiol*, 2014, 44: 264-270.

[13] Noda I. Two-trace two-dimensional (2T2D) correlation spectroscopy-A method for extracting useful information from a pair of spectra [J]. *J Mol Struct*, 2018, 1160: 471-478.

[14] Wang L Y, Wang Q, Cai H X, *et al*. Evaluation of fecal immunoassays for canine *Echinococcus* infection in China [J]. *PLoS Negl Trop Dis*, 2021, 15(3): e0008690.

[15] Yue J Q, Li W Y, Wang Y Z. Superiority verification of deep learning in the identification of medicinal plants: Taking *Paris polyphylla* var. *yunnanensis* as an example [J]. *Front Plant Sci*, 2021, 12: 752863.

[16] Dai Q, Sun D W, Xiong Z J, *et al*. Recent advances in data mining techniques and their applications in hyperspectral image processing for the food industry [J]. *Compr Rev Food Sci Food Saf*, 2014, 13(5): 891-905.

[17] 刘嘉灏. 不同产地滇黄精品质与生态因子和根际微生物相关性研究 [D]. 大理: 大理大学, 2023.

[18] Mishra P, Lohumi S. Improved prediction of protein content in wheat kernels with a fusion of scatter correction methods in NIR data modelling [J]. *Biosyst Eng*, 2021, 203: 93-97.

[19] Mishra P, Biancolillo A, Roger J M, *et al*. New data preprocessing trends based on ensemble of multiple preprocessing techniques [J]. *Trac Trends Anal Chem*, 2020, 132: 116045.

[20] He G, Yang S B, Wang Y Z. Suitable habitat prediction and identification of origin of *Lanxangia tsao*-ko [J]. *Comput Electron Agric*, 2024, 223: 109127.

[责任编辑　时圣明]