基于红外光谱结合机器学习方法的牛膝不同炮制品及炮制程度的判别分析

田瀚举[1,2]，杨颜溶[1,2]，贾　豪[1,2]，李莹莹[1,2]，段浩瀚[1,2]，赵新梅[1,2]，张春亚[1,2]，雷敬卫[1,2*]，谢彩侠[1,2]，杨春静[1,2,3]，龚海燕[1,2*]

1. 河南中医药大学药学院，河南 郑州 450046
2. 河南省中药质量控制与评价工程技术研究中心，河南 郑州 450046
3. 河南中医药大学第三附属医院，河南 郑州 450046

摘　要：目的　采用红外光谱技术结合机器学习算法建立牛膝 *Achyranthes bidentata* 炮制品类别与炮制程度的定性判别模型。**方法**　采集不同炮制品与不同炮制程度牛膝的中红外光谱（mid infrared spectroscopy，MIRS），运用 BP 神经网络（back propagation neural network，BPNN）、遗传算法优化 BP 神经网络（GA-BP）、随机森林（random forest，RF）、径向基神经网络（radial basis function network，RBFN）、卷积神经网络（convolutional neural networks，CNN）等机器学习算法建立牛膝炮制品类别与炮制程度的定性判别模型；采集不同炮制品与不同炮制程度牛膝的近红外光谱（near infrared spectroscopy，NIRS），使用 TQ Analyst 软件中的判别分析法建立牛膝炮制品类别与炮制程度的定性分析模型。**结果**　机器学习算法模型结果显示 CNN 判别模型较优秀，BPNN、RF 及 RBFN 性能相近，GA-BP 模型性能相对较差。3 个 NIRS 定性模型结果显示验证集准确率均为 100%，可准确预测炮制品类别与炮制程度。**结论**　通过红外光谱技术建立的定性分析模型可作为牛膝炮制品类别与炮制程度的鉴别手段。同时提供了快速、无损的检测手段及可靠的数据分析方法，为中药材炮制品类别与炮制程度精准识别提供新的方法参考。

关键词：牛膝；炮制品；炮制程度；红外光谱；正交偏最小二乘法-判别分析；机器学习算法

中图分类号：R283.6　　**文献标志码：**A　　**文章编号：**0253 - 2670(2023)22 - 7387 - 15

DOI: 10.7501/j.issn.0253-2670.2023.22.015

Discrimination analysis of different processed products and processing degree of *Achyranthis Bidentatae Radix* based on infrared spectroscopy combined with machine learning methods

TIAN Han-ju[1,2], YANG Yan-rong[1,2], JIA Hao[1,2], LI Ying-ying[1,2], DUAN Hao-han[1,2], ZHAO Xin-mei[1,2], ZHANG Chun-ya[1,2], LEI Jing-wei[1,2], XIE Cai-xia[1,2], YANG Chun-jing[1,2,3], GONG Hai-yan[1,2]

1. School of Pharmacy, Henan University of Chinese Medicine, Zhengzhou 450046, China
2. Henan Engineering Technology Research Center for TCM Quality Control and Evaluation, Zhengzhou 450046, China
3. Third Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450046, China

Abstract: Objective　To establish a qualitative discrimination model for the type and degree of processing of Niuxi (*Achyranthes bidentata*, AB) using infrared spectroscopy and machine learning algorithms. **Methods**　The infrared spectra of AB with different processing types and degree was collected, and various machine learning algorithms, including back propagation neural network (BPNN), genetic algorithm-optimized BP neural network (GA-BP), random forest (RF), radial basis function network (RBFN), and convolutional neural networks (CNN) were used to establish a qualitative discrimination model for the type and degree of processed products of AB. The near-infrared spectra (NIRS) of AB with different processing types and degree was collected, and TQ Analyst software was used to establish a qualitative analysis model for the type and degree of processed products of AB. **Results**　The results of the machine learning algorithm models showed that the CNN discriminative model was superior, the BPNN, RF and RBFN

收稿日期：2023-05-29

基金项目：国家重点研发计划“中医药现代化研究”重点专项项目（2018YFC1707000）；河南省中医药科学研究专项课题（2022ZY1156）

作者简介：田瀚举，男，硕士研究生，研究方向为中药质量分析研究。E-mail: tianhanju@163.com

***通信作者：**雷敬卫，男，教授，研究方向为中药质量分析研究。Tel: (0371)65955281　E-mail: 925390812@qq.com

龚海燕，女，副教授，研究方向为中药质量分析研究。Tel: (0371)65575838　E-mail: ghy_mz@163.com

had similar performance, and the GA-BP model had relatively poor performance. The three NIRS qualitative models had validation accuracies of 100%, indicating that they could accurately predict the type and degree of processed products of AB. **Conclusion** The qualitative analysis model developed in this study by infrared spectroscopy can be used as a means to identify the type and degree of processed products of AB. It also provides a rapid and non-destructive means of testing and a reliable method for data analysis, with view to providing a new method of reference for the accurate identification of the type and degree of preparation of Chinese herbal processed products.

Key words: *Achyranthes bidentata* BL.; processed product; processing degree; infrared spectroscopy; orthogonal partial least squares-discriminant analysis; machine learning algorithm

牛膝 *Achyranthis Bidentatae Radix* 为苋科牛膝属植物牛膝 *Achyranthes bidentata* BL.的干燥根[1]，最早出自《神农本草经》，其根入药，具有补肝肾、强筋骨、活血化瘀的功效[2]，主要含有皂苷类、甾酮类、多糖类等化合物[3]。现国内有三大牛膝产区：内蒙赤峰、河北安国和河南焦作[4]。牛膝炮制历史悠久，古代炮制方法有酒制（酒渍、酒浸、酒煮、酒洗、酒炒、酒蒸等）、炒制、焙制、炙制、药汁制等[5]。现代临床所用的牛膝饮片主要为牛膝生品、酒牛膝、盐牛膝等[6]。牛膝生品经酒炙后能增强活血祛瘀、通经止痛的作用，盐炙后能增强补肝肾、强筋骨作用[7]。

红外光谱法作为一种快速无损分析技术，且具有样品制备简单、无污染、经济实惠等特点，在诸多领域均有应用[8-10]。随着化学计量学和机器学习算法与红外光谱技术的结合，复杂的样品光谱信息得以有效可视化，成为中药快速鉴别及质量评价的一种有效手段[11]，目前，该技术已广泛应用于中药材产地溯源研究[12-17]。

本课题组前期采用红外光谱技术开展了牛膝产地的快速识别研究[18]，在此基础上本研究通过采集 3 个产地的牛膝生品，不同炮制程度的酒牛膝和盐牛膝近红外光谱（near infrared spectroscopy，NIRS）和中红外光谱（mid infrared spectroscopy，MIRS）信息，结合 BP 神经网络（back propagation neural network，BPNN）、遗传算法优化 BP 神经网络（GA-BP）、随机森林（random forest，RF）、径向基神经网络（radial basis function network，RBFN）、卷积神经网络算法开展对牛膝炮制类别与炮制程度研究，建立适合的定性判别模型，为牛膝炮制类别与炮制程度的精准识别提供方法支撑。

1 仪器与材料

1.1 仪器与试剂

INVENIOS 型傅里叶变换红外光谱仪，德国 Bruker 公司；Nicolet 6700 型傅里叶红外光谱仪，美国 Thermo Fisher 公司；Spectrum for Window 软件（版本 3.02），美国 Pekin Elmer 公司；Matlab 软件（版本 R2022b），美国 MathWorks 公司；FW-4A 型粉末压片机，天津市拓扑仪器有限公司；FW-100 型高速万能粉碎机，北京科伟永兴仪器有限公司；101-3AB 型点热恒温鼓风干燥箱，北京中兴伟业仪器有限公司；ME204E/OL 型万分之一天平，上海梅特勒-托利多仪器有限公司。

溴化钾，光谱纯，天津市科密欧化学试剂有限公司；无水乙醇，分析纯，天津市致远化学试剂有限公司；黄酒，酒精度≥10.0% vol，批号 20220616D，浙江古越龙山绍兴酒股份有限公司；精纯盐，河南省盐业集团有限公司。

1.2 样品

牛膝样品于 2021 年 12 月采集自道地产区河南省焦作市西陶镇、非道地产区河北省安国市西佛落镇与内蒙古自治区赤峰市喀喇沁旗牛家营子镇，共计 15 个批次，均为 1 年生，所有样品经河南中医药大学陈随清教授鉴定为苋科牛膝属植物牛膝 *A. bidentata* Bl.的干燥根。

牛膝除去杂质，洗净，润透，除去残留芦头，切段，干燥得到牛膝生品，粉碎后过 3 号和 9 号筛，贮藏备用。取牛膝生品，照参照《中国药典》2020 年版四部 0213 炮制通则中酒炙法[19]，加黄酒 10% 拌匀，焖透，置炒锅内，文火炒制，炒至表面颜色略深，偶见焦斑，微有酒香气，制备炮制不及、炮制适中（酒牛膝）和炮制过 3 种不同程度，粉碎后过 3 号和 9 号筛，贮藏备用。

取牛膝生品，参照《中国药典》2020 年版四部 0213 炮制通则中盐炙法，加食盐 2%，用 10%蒸馏水溶解拌匀，焖透，置炒锅内，文火炒制，炒至表面色深，略有焦斑，制备炮制不及、炮制适中（盐牛膝）和炮制太过 3 种不同程度，粉碎后过 3 号和 9 号筛，贮藏备用。具体样品信息见表 1，部分样品示图见图 1。

表 1　样品信息

Table 1　Information of samples

产地	生品	酒牛膝			盐牛膝		
		炮制不及	炮制适中	炮制太过	炮制不及	炮制适中	炮制太过
河南焦作	S1～S5	JB1～JB5	JS1～JS5	JG1～JG5	YB1～5	YS1～YS5	YG1～YG5
内蒙古赤峰	S6～S10	JB6～JB10	JS6～JS10	JG6～JG10	YB6～YB10	YS6～YS10	YG6～YG10
河北安国	S11～S15	JB11～JB15	JS11～JS15	JG11～JG15	YB11～YB5	YS11～YS15	YG11～YG15

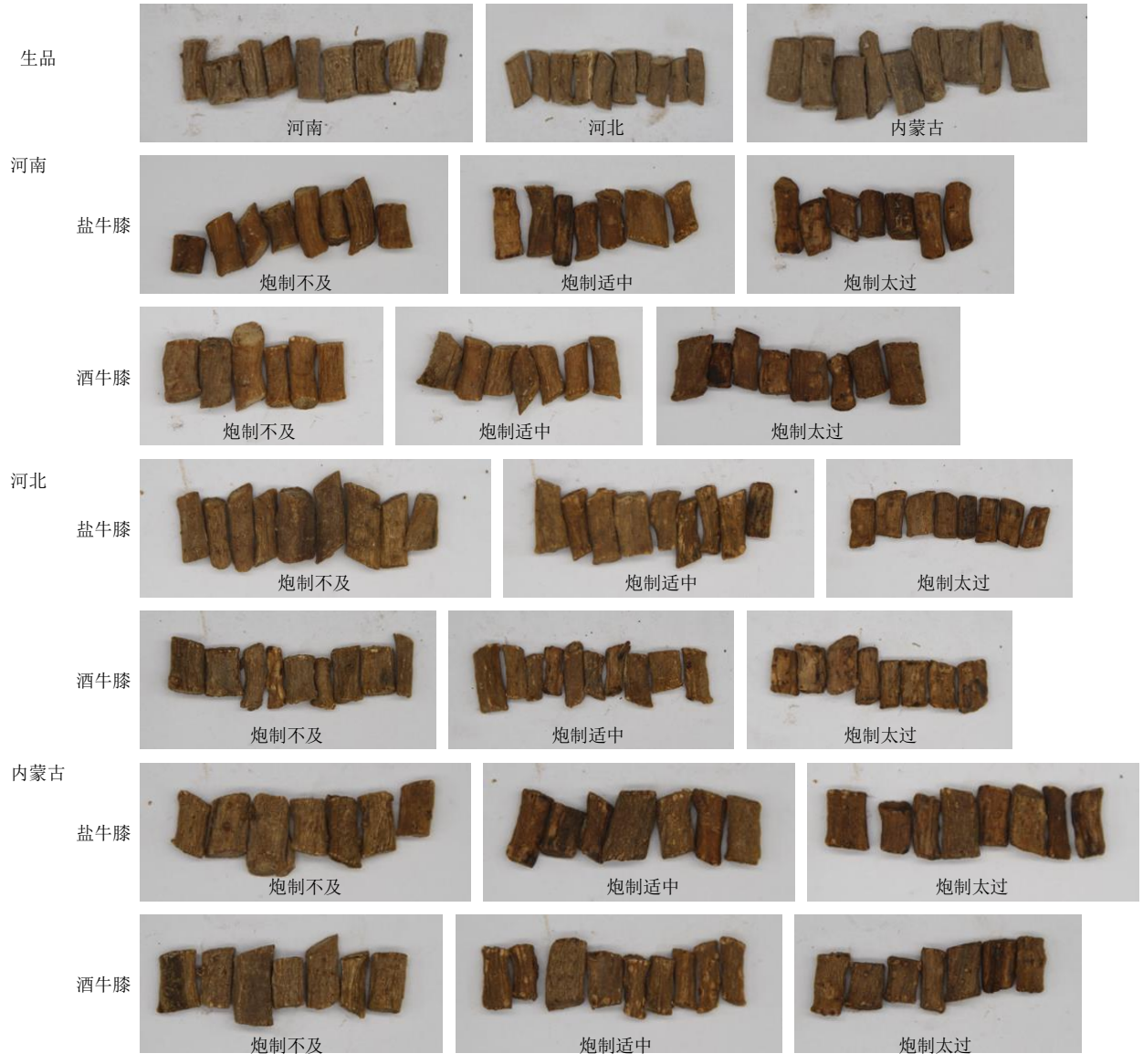


图 1　不同产地来源牛膝生品及炮制品

Fig. 1　Raw and processed products of *A. bidentata* from different producing areas

2　方法

2.1　MIRS 信息的采集

称取样品粉末（过 9 号筛）约 2 mg 与干燥溴化钾以 1∶100 研磨混匀，取适量混合均匀的样品置于专用压片模具中，用 8 MPa 的压力压制 30 s，压成均匀半透明的薄片，取出，置红外光谱仪中采集各样品 MIRS 图。光谱扫描范围 400～4000 cm^{-1}，每张光谱扫描次数 16 次每秒，光谱分辨率为 4 cm^{-1}，扫描速度 0.2 cm^{-1}，扫描时扣除 CO_2 和 H_2O，室温 20～25 ℃，相对湿度 25%～35%。每张图谱重复扫描 3 次，取其平均光谱，每份样品扫描 3 张图谱。

2.2　NIRS 信息的采集

称取样品粉末（过 3 号筛）约 6 g，置于石英样品杯中，混合均匀，轻轻压平，以空气为背景，扣除背景采集光谱图，采用积分球漫反射，分辨率为 8 cm^{-1}，扫描 64 次，扫描范围为 4000～12 000 cm^{-1}，温度范围为 25～30 ℃，空气湿度为 25%～30%。每张图谱重复扫描 3 次，取其平均光谱，每份样品扫描 3 张图谱。

2.3　光谱信息的预处理

MIRS 信息均采用 Spectrum for window 3.02 软件对各样品采集的原始 MIRS 进行处理，采用 TQ Analyst 软件对 NIRS 进行多元信号修正（multiple signal correction，MSC）、标准正则变换（standard normal variate transform，SNV）、一阶导数（first derivative）、二阶导数（second derivative）、SG 平滑（Savitzky-Golay，SG）、ND 平滑（Norris derivative，ND）。

2.4　数据处理

使用 GraphPad Prism 软件绘制牛膝生品、酒牛膝和盐牛膝平均相对峰高柱状图，使用 Matlab 软件构建不同炮制品和不同炮制程度分类模型，将数据样本随机拆分成训练集（70%）和测试集（30%），运用 BPNN、遗传算法优化 BP 神经网络（GA-BP）、随机森林（random forest，RF）、径向基神经网络（radial basis function network，RBFN）、卷积神经网络（convolutional neural networks，CNN）等算法构建分类模型。使用 TQ 软件建立不同炮制品和不同炮制程度牛膝近红外定性分析模型。

3　结果与分析

3.1　样品红外光谱

MIRS 进行透过率与吸光度转换、基线校正、归一化处理，计算得到 14 个共有峰（图 2），对 1 号峰进行归一化之后，牛膝生品的 2～14 号峰经酒炙与盐炙后相对峰高均升高，且酒牛膝增长幅度大于盐牛膝，结果见图 3。不同产地牛膝样品的原始 MIRS 及不同炮制品（以河南为例）原始 MIRS 如图 4、5 所示，原始 NIRS 及不同炮制品（以河南为例）原始 NIRS 如图 6、7 所示。

3.2　MIRS 判别模型的选择及建立

采用 BPNN、GA-BP、RF、RBFN、CNN 等算法建立牛膝不同炮制品及不同炮制品不同炮制程度 MIRS 分类判别模型。

BPNN 是一种按照误差逆向传播算法训练的多

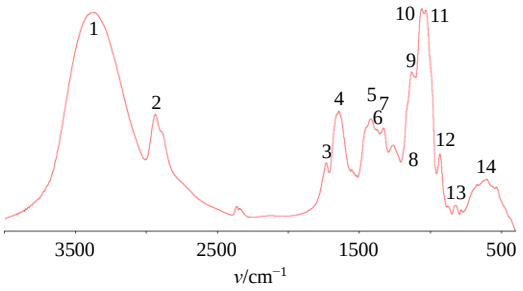


图 2　MIRS 共有峰示意图

Fig. 2　Common peak schematic of MIRS

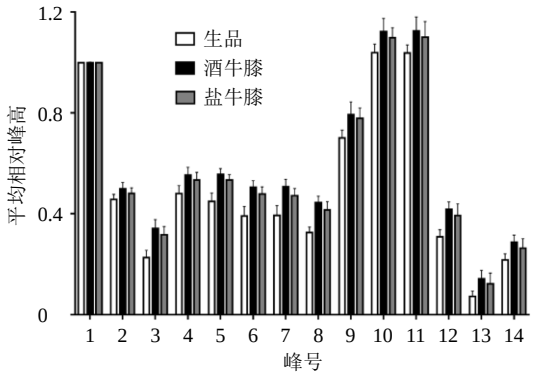


图 3　牛膝不同炮制品的平均相对峰高柱状图

Fig. 3　Bar chart of average relative peak heights of different processed products of *A. bidentata*

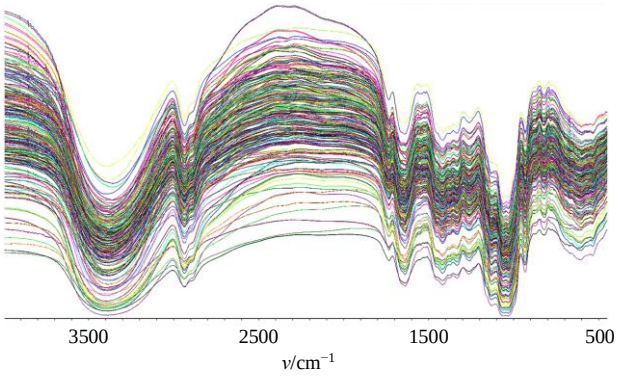


图 4　牛膝样品的原始 MIRS

Fig. 4　Original MIRS of *A. bidentata* samples

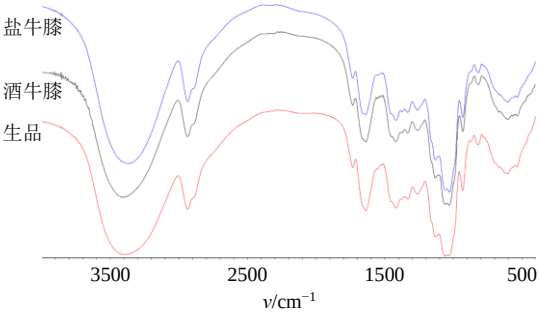


图 5　河南牛膝不同炮制品 MIRS 示意图

Fig. 5　MIRS schematic of different processed products of *A. bidentata* from Henan

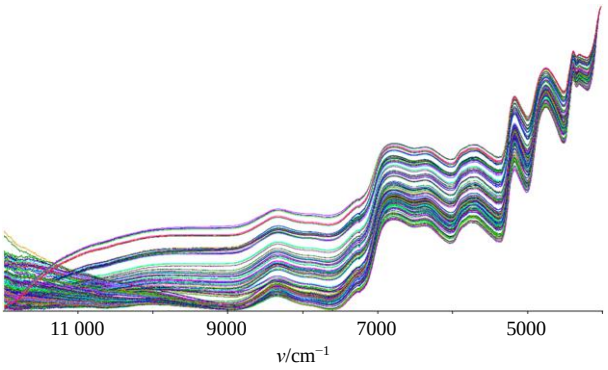


图 6 牛膝样品的原始 NIRS

Fig. 6 Original NIRS of *A. bidentata*

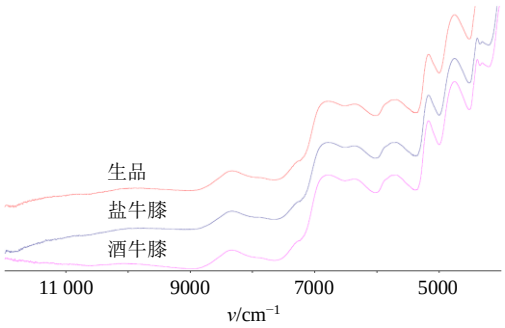


图 7 河南牛膝不同炮制品的 NIRS 示意图

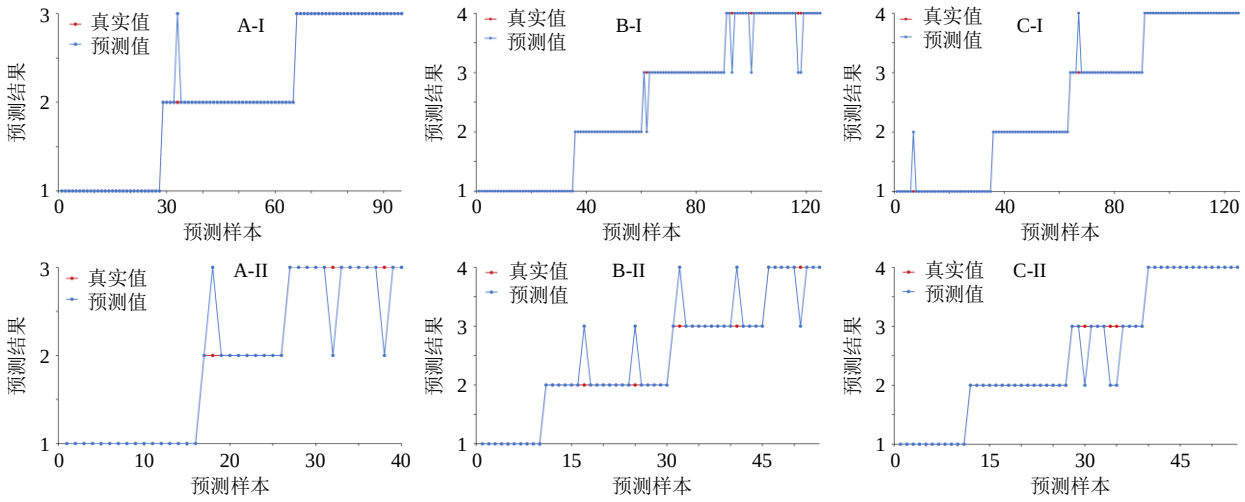
Fig. 7 NIRS schematic of different processed products of *A. bidentata* from Henan

层前馈神经网络，具有很强的非线性映射能力、适应能力和学习能力[20]，具备任意复杂的分类模式和良好的多维函数映射的能力[21]，是应用最为广泛的人工神经网络之一[22]；与 BPNN 因在模型连接权值和阈值选取时具有随机性，从而易于陷入局部最优解不同。遗传算法（genetic algorithm，GA）是模拟自然界中遗传机制及物种进化的过程中形成的一种并行随机搜索优化方法，二者相结合得到的 GA-BP 算法发可以做到优化可行域内 BP 神经网络模型连接权值和阈值选取的随机性，有效增强模型的泛化能力和收敛性[23]；以决策树为核心的多分类 RF 算法作为一种典型的多分类器算法，可以很好地对数据进行集成学习[24]，同时根据数据的多样性进行分类处理，故此，RF 算法拥有非常强大的适用性，可以在许多领域进行广泛应用，特别是针对一些非线性高维数据，随机森林算法也可以很快地进行数据处理[25]，此外，RF 算法对噪声和随机误差的防控非常到位，可以极大地减少因数据产生的误差，从而降低了数据处理难度，节约了大量的人力物力，帮助数据得到快速、准确的分析；RBFN 具有唯一最佳逼近、训练简洁、学习收敛速度快等良好性能，并且具有很强的非线性拟合能力，可逼近任意的非线性函数，具有较好的泛化能力，现已成功应用于语音识别、自动控制、信息图像处理和故障诊断等多个领域[26]；CNN 是一种常见的文本分类模型，是由卷积层、池化层、全连接层组成的人工神经网络结构[27]。相对于传统的多层感知神经网络，其卷积层具有局部链接、权值共享以及池化操作既能够有效地提取特征，大幅度地简化了网络的复杂度[28]。

以 BPNN 为例，将已划分好的不同炮制品的数据集导入 Matlab R2022b 软件，模型判别结果如图 8 所示，不同炮制品判别模型中训练集的准确率为 98.9%，而测试集的准确率为 92.5%，训练集与测试集准确率均大于 90%。对于不同炮制品的不同炮制程度 BPNN 模型判别结果显示：酒牛膝与盐牛膝不同炮制程度模型训练集的准确率分别为 96.0%和 98.4%，测试集的准确率分别为 92.6%和 94.4%，两个模型训练集与测试集准确率均大于 90%，说明基于该样本集建立的 BPNN 模型适用于酒牛膝、盐牛膝不同炮制程度的预测判别。对比 GA-BP 的模型判别结果，随着迭代次数的增加，判别模型错误率呈现降低趋势（图 9）。不同炮制品的 GA-BP 判别模型中训练集的准确率为 93.6%，测试集准确率为 90%。对于不同炮制品的不同炮制程度 GA-BP 模型判别结果显示：酒牛膝与盐牛膝不同炮制程度模型训练集的准确率分别为 90.5%和 96.8%，测试集准确率分别为 90.7%和 94.4%，如图 10 所示。CNN 模型迭代曲线如图 11 所示，CNN、RBFN、RF 模型判别准确率结果见图 12～14。

3.3 MIRS 判别模型性能评估

在机器学习中，混淆矩阵作为一个误差矩阵，常用来可视化地评估监督学习算法的性能，是机器学习中总结分类模型预测结果的情形分析表，以矩阵形式将数据集中的记录按照真实的类别与分类模型预测的类别判断 2 个标准进行汇总。分类的正确性可以通过计算正确预测样本属于此样本数量（true positives，TP），正确预测的样本数量不属于此样本集数量（true negatives，TN），和样本被错误地预测为此样本数量（false positives，FP）以及样本被错误地预测为不属于样本数量（false negatives，FN）来进行衡量，评价指标主要包括准确度（accuracy）、精确度（precision）、召回率（recall）等[29]。其中精确度可以用来衡量模型的整体有效性，即预测正确的结果占总样本的百分比；样本类别于模型判别结



A-不同炮制品　B-酒牛膝不同炮制程度　C-盐牛膝不同炮制程度　I-训练集　II-测试集，下图同

A-different processed products　B-different processing degree of wine-processed *A. bidentata*　C-different processing degree of salt-processed *A. bidentata*　I-training set　II-testing set, same as below tables

图 8　BPNN 模型判别准确率

Fig. 8　Discrimination accuracy of BPNN model

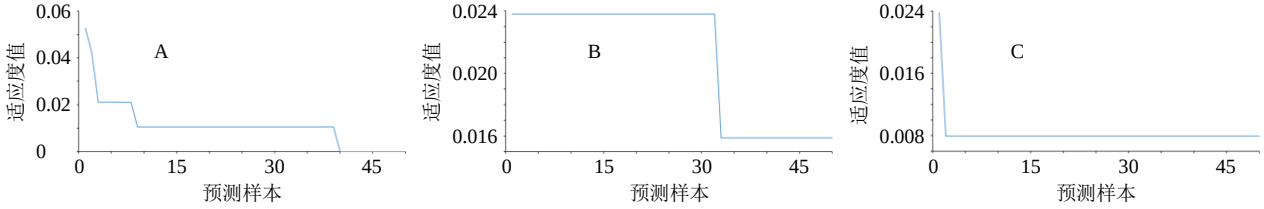


图 9　GA-BP 模型适应度曲线

Fig. 9　Fitness curve of the GA-BP model

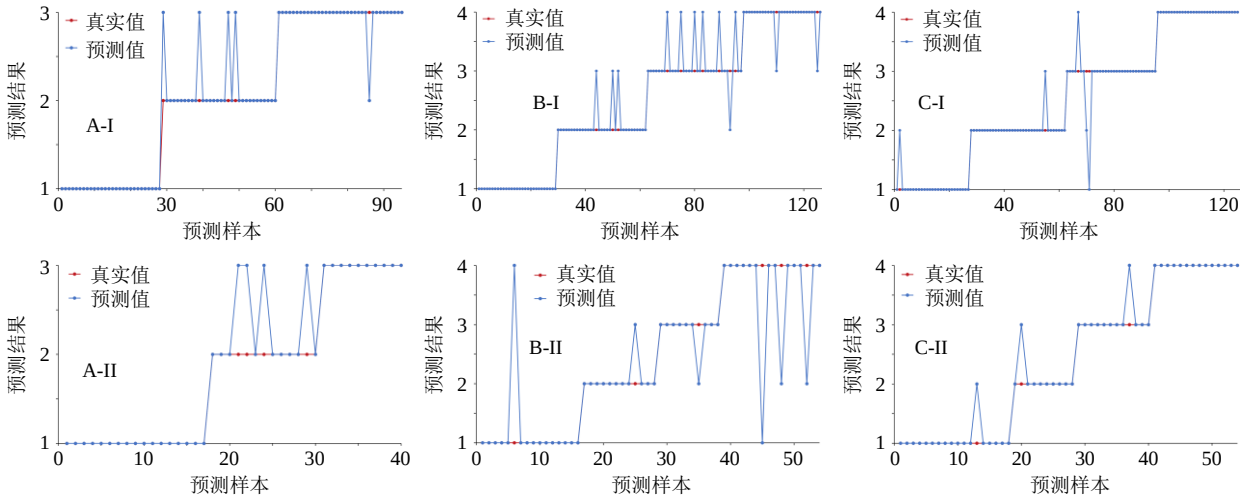


图 10　GA-BP 模型判别准确率

Fig. 10　Discrimination accuracy of GA-BP model

果的一致性则可以通过精密度来衡量；召回率即在实际为样本中被预测为该样本的概率。本实验以混淆矩阵结合准确度、精确度以及召回率评估模型性能，其数据越接近 1，模型的性能越好。具体计算公式如下。

准确度＝(TP＋TN)/(TP＋TN＋FP＋FN)

精确度＝TP/(TP＋FP)

精确度＝TP/(TP＋FN)

不同炮制品预测输出有 3 个类别，其中 1 代表生品，2 代表酒牛膝，3 代表盐牛膝。不同炮制品不同炮制程度预测输出有 4 个类别，其中 1 代表生品，2 代表炮制不及，3 代表炮制适中，4 代表炮制过。

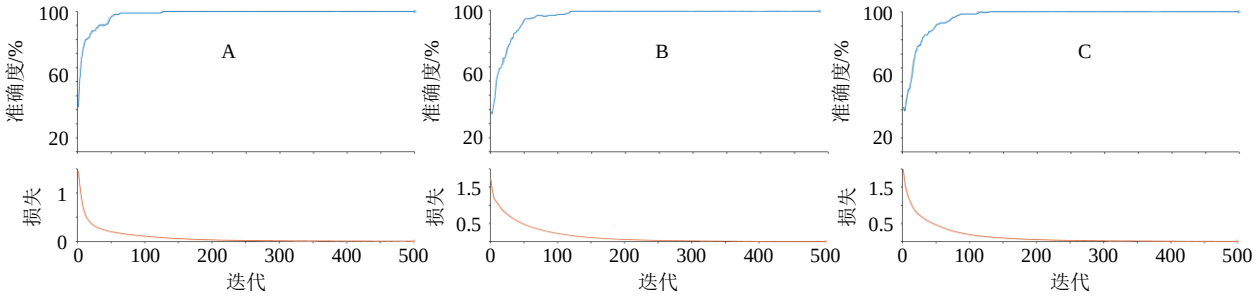


图 11　CNN 模型迭代曲线

Fig. 11　Iteration curve of CNN model

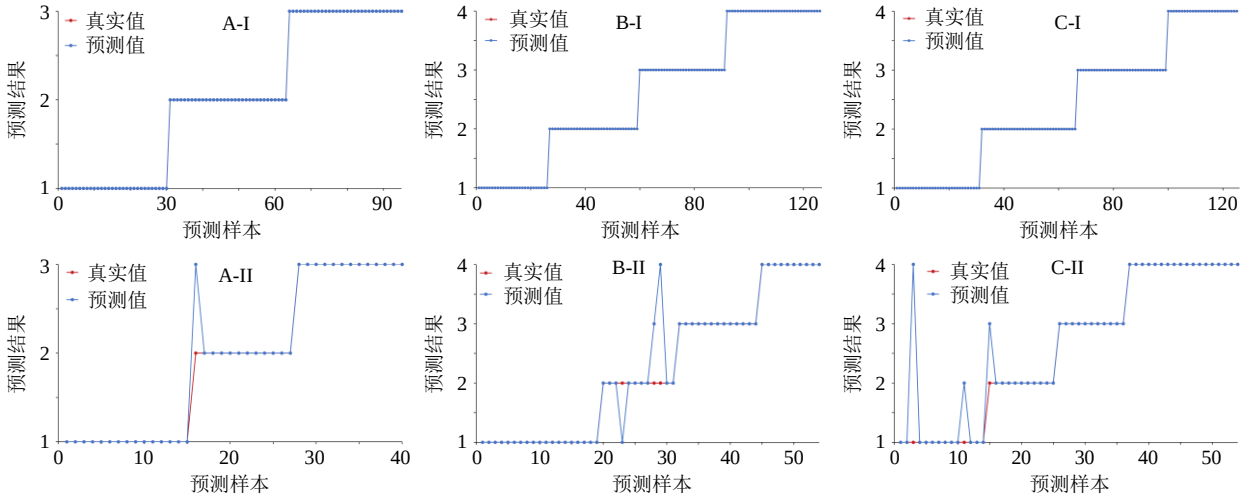


图 12　CNN 神经网络模型判别准确率

Fig. 12　Discrimination accuracy of CNN model

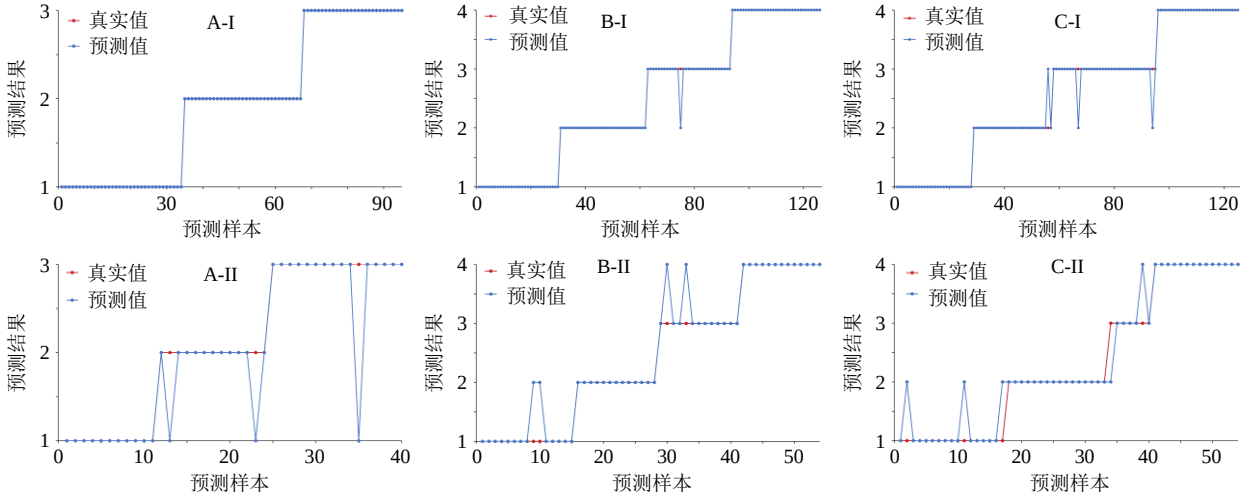


图 13　RBFN 模型判别准确率

Fig. 13　Discrimination accuracy of RBFN model

模型混淆矩阵可视化见图 15～19，评价指标数值见表 2～4。

上述结果表明，5 种算法对于训练集以及预测集的判别准确率除 GA-BP 外均在 0.90 以上，展现了良好的分类性能，但是不同模型之间判别性能有较大差异。例如，在对于不同炮制品的判别模型中，CNN 模型性能极佳，对于训练集以及预测集的判别成功率分别达到了 1.00 和 0.98，且二者差距较小，说明该模型在当前样本量下面对不同数据集时鲁棒性较佳。反观 GA-BP 算法，虽然弥补了 BPNN 算法易陷入局部极小、收敛速度慢的缺点，但在本样本集建立的模型判别结果中不难看出，对于不同炮

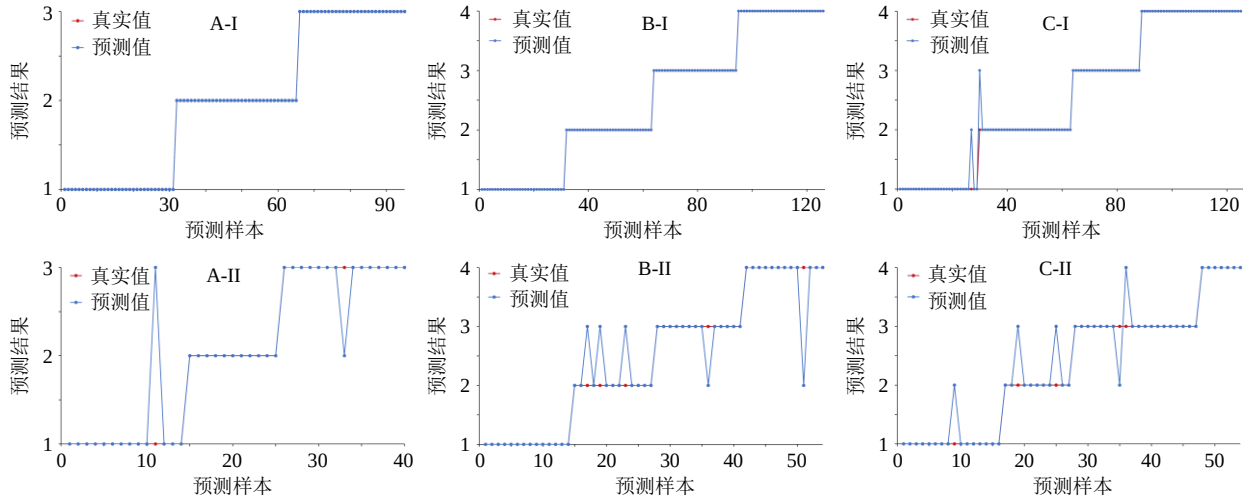


图 14　RF 模型判别准确率

Fig. 14　Discrimination accuracy of RF model

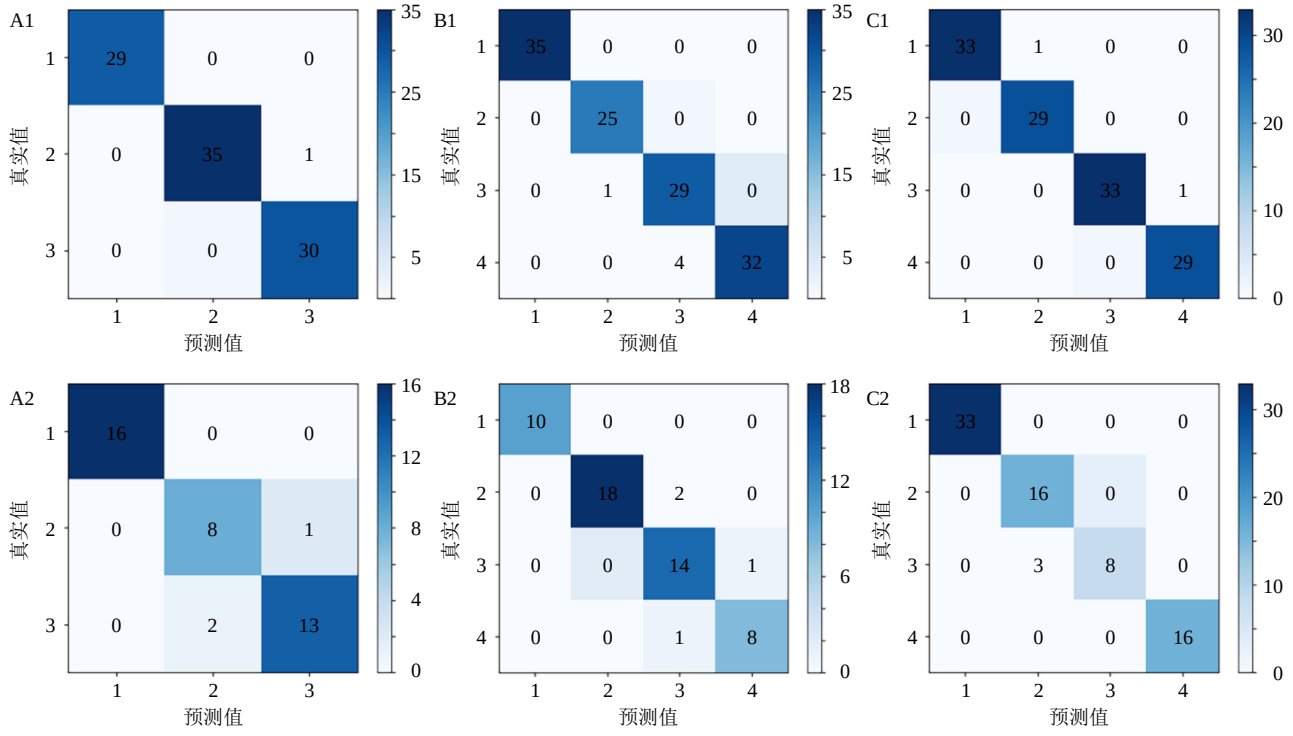


图 15　不同炮制品 (A)、酒牛膝不同炮制程度 (B)、盐牛膝不同炮制程度 (C) 的 BPNN 模型混淆矩阵 (1 训练集、2 测试集)

Fig. 15　BPNN model confusion matrix of different processed products (A), different processing degree of wine-processed *A. bidentata* (B), different processing degree of salt-processed *A. bidentata* (C) (1 training set, 2 testing set)

制品判别模型以及炮制品不同炮制程度判别模型中，BPNN 模型整体优于 GA-BP 模型。

由此可见，不同建模方法对于数据集特征提取逻辑不同，应根据数据集特性选择合适的建模方法进行判别以及分析。

3.4　牛膝不同炮制品 NIRS 定性模型建立

本实验采用判别分析法建立不同炮制品 NIRS 定性判别模型，以及不同炮制品不同炮制品程度的 NIRS 定性判别模型，以性能系数（PI）和误判例数为评价指标，PI 值越大，误判例数越小，说明 NIRS 定性模型的判别分析结果越准确。本实验考察了光谱预处理方法对定性模型的影响，得不同炮制品定性模型的最佳预处理条件为 SNV＋SG；不同炮制品光谱预处理结果见表 5。以 PI 和误判例数为评价指标，考察不同波段对 NIRS 定性模型的影响，得不同炮制品定性模型的最佳波段为 4250～5150 cm^{-1}，

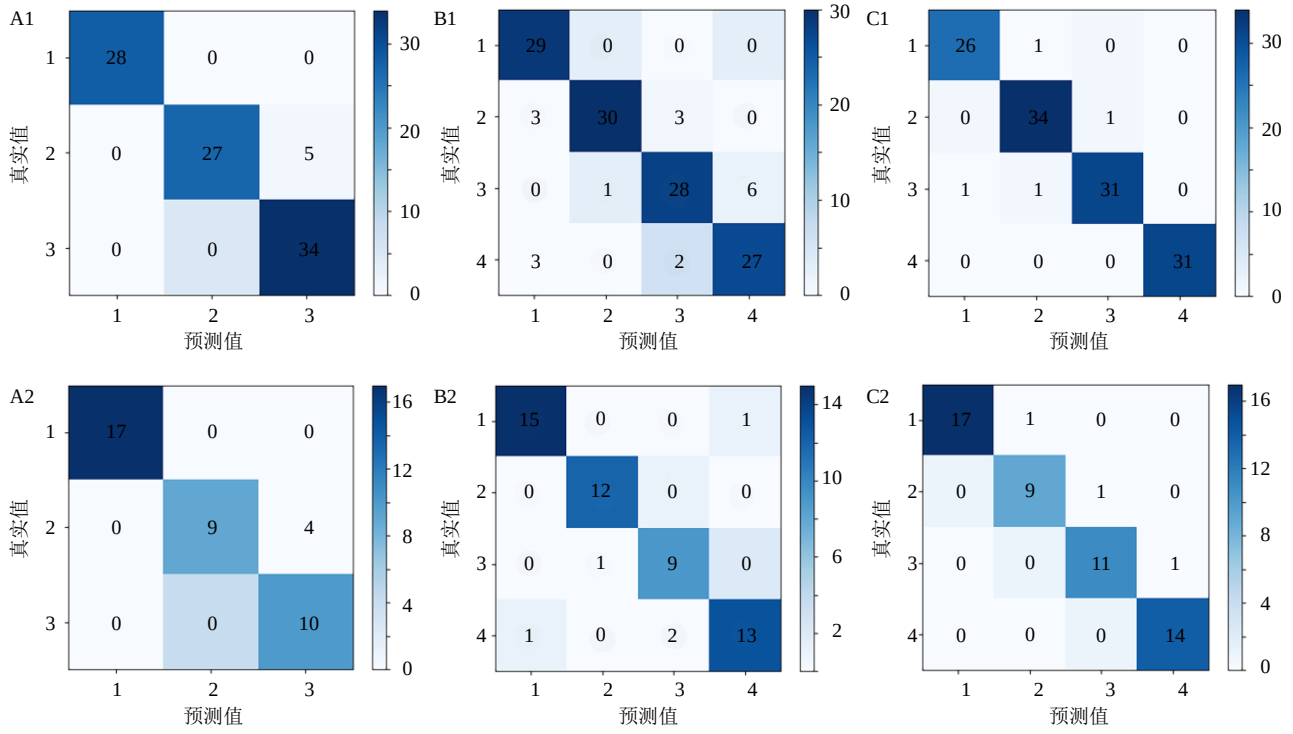


图 16　不同炮制品 (A)、酒牛膝不同炮制程度 (B)、盐牛膝不同炮制程度 (C) 的 GA-BP 模型混淆矩阵 (1 训练集、2 测试集)

Fig. 16　GA-BP model confusion matrix of different processed products (A), different processing degreesof wine-processed *A. bidentata* (B), different processing degree of salt-processed *A. bidentata* (C) (1 training set, 2 testing set)

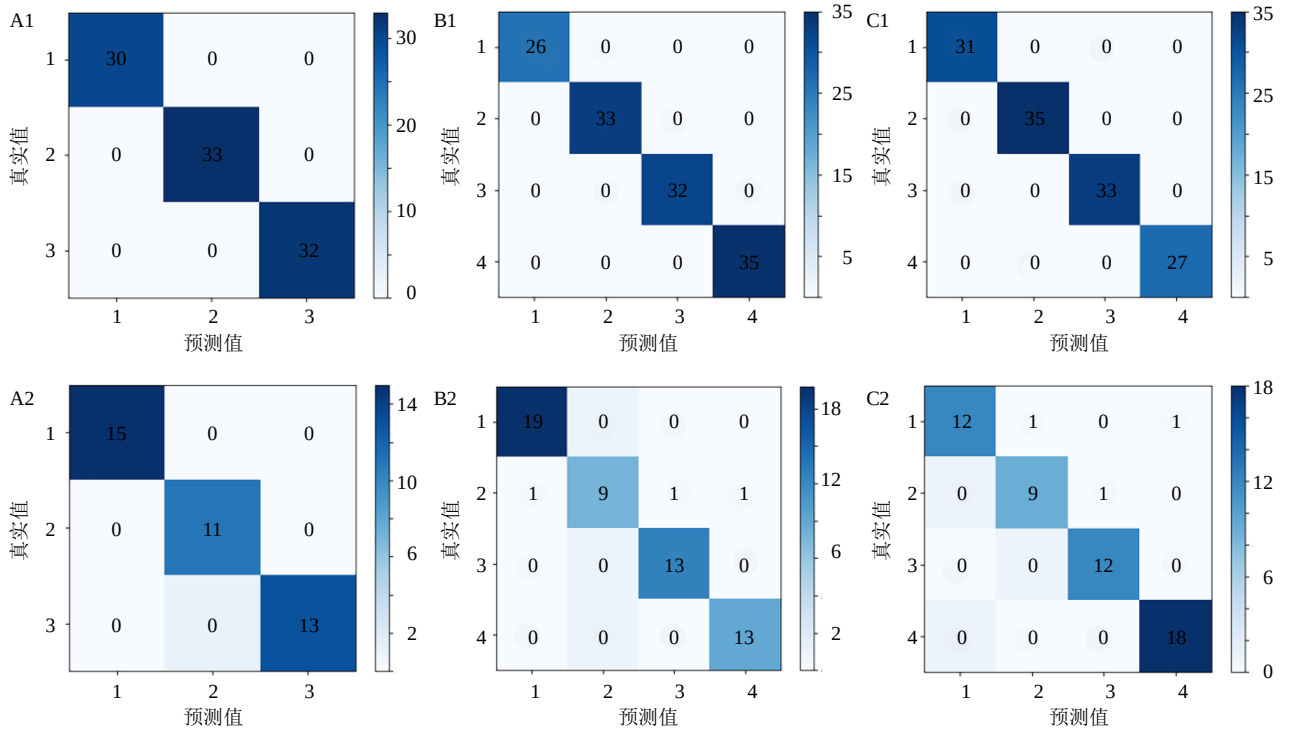


图 17　不同炮制品 (A)、酒牛膝不同炮制程度 (B)、盐牛膝不同炮制程度 (C) 的 CNN 模型混淆矩阵 (1 训练集、2 测试集)

Fig. 17　CNN model confusion matrix of different processed products (A), different processing degree of wine-processed *A. bidentata* (B), different processing degree of salt-processed *A. bidentata* (C) (1 training set, 2 testing set)

不同炮制品不同波段分析结果见表 6。采用 TQ Analyst 软件，根据 NIRS 最佳预处理方法及最佳的光谱波段进行判别分析，建立不同炮制品的定性分析模型（图 20）。

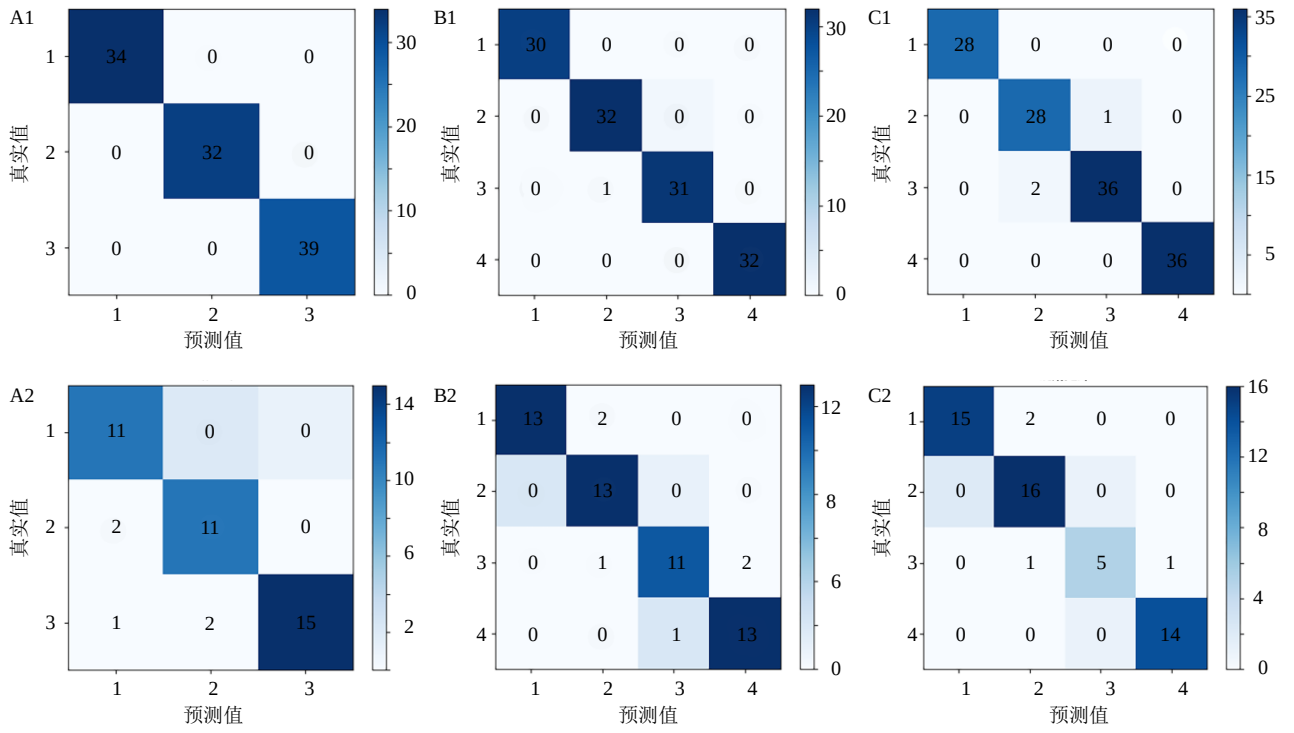


图 18　不同炮制品 (A)、酒牛膝不同炮制程度 (B)、盐牛膝不同炮制程度 (C) 的 RBFN 模型混淆矩阵 (1 训练集、2 测试集)

Fig. 18　RBFN model confusion matrix of different processed products (A), different processing degree of wine-processed *A. bidentata* (B), different processing degree of salt-processed *A. bidentata* (C) (1 training set, 2 testing set)

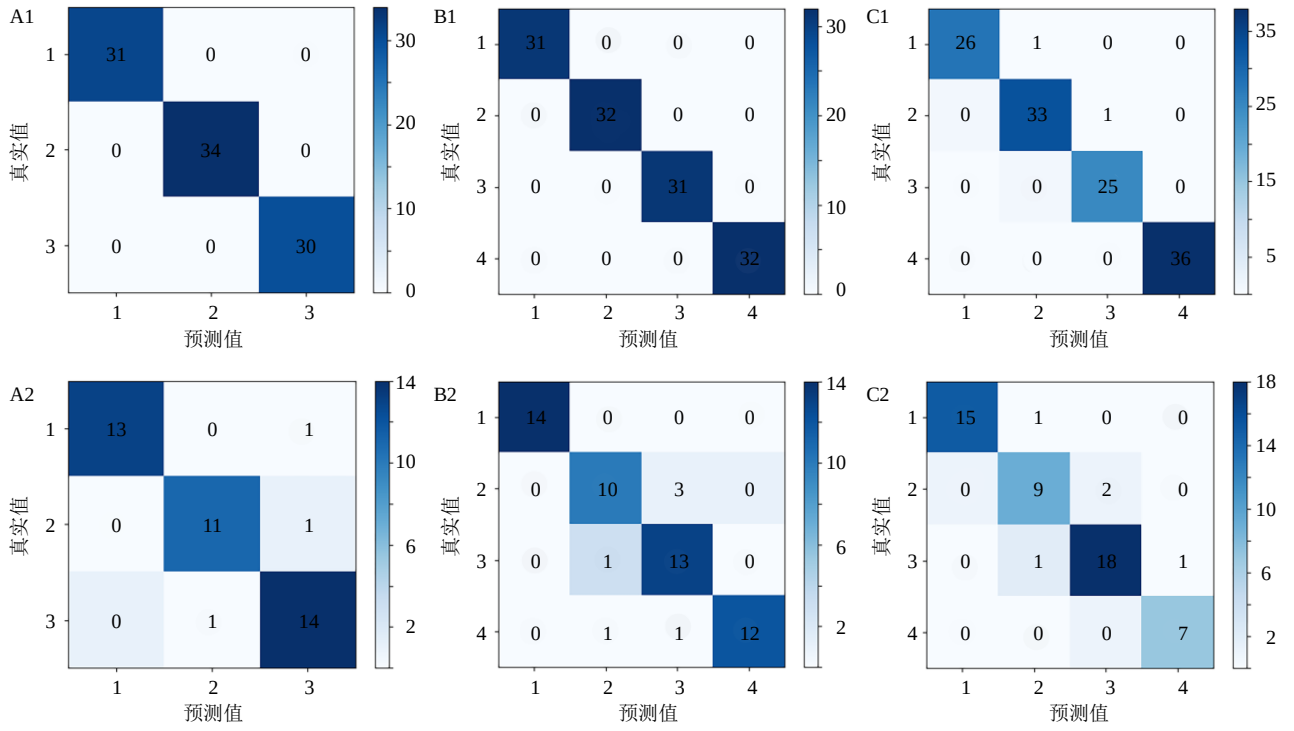


图 19　不同炮制品 (A)、酒牛膝不同炮制程度 (B)、盐牛膝不同炮制程度 (C) 的 RF 模型混淆矩阵 (1 训练集、2 测试集)

Fig. 19　RF model confusion matrix of different processed products (A), different processing degree of wine-processed *A. bidentata* (B), different processing degree of salt-processed *A. bidentata* (C) (1 training set, 2 testing set)

将验证集样品的 NIRS 图谱输入所建模型，结果显示，不同炮制品可被准确分为 3 类，正确率为 100%。本实验考察了光谱预处理方法对定性模型的影响，得酒牛膝不同炮制程度定性模型的最佳预处

表 2 不同炮制品判别模型评价指标

Table 2 Evaluation indicators of discrimination models of different processed products

算法模型	类别	训练集			预测集		
		精确度/%	召回率/%	准确度/%	精确度/%	召回率/%	准确度/%
BPNN	1	1.00	1.00	0.99	1.00	1.00	0.93
	2	1.00	0.97		0.80	0.89	
	3	0.97	1.00		0.93	0.87	
RBFN	1	1.00	1.00	1.00	0.79	1.00	0.93
	2	1.00	1.00		1.00	0.85	
	3	1.00	1.00		1.00	0.94	
GA-BP	1	1.00	1.00	0.94	1.00	1.00	0.90
	2	0.96	0.84		1.00	0.69	
	3	0.97	0.97		0.70	1.00	
RF	1	1.00	1.00	1.00	1.00	0.93	0.95
	2	1.00	1.00		0.92	1.00	
	3	1.00	1.00		0.93	0.93	
CNN	1	1.00	1.00	1.00	1.00	1.00	0.98
	2	1.00	1.00		1.00	0.92	
	3	1.00	1.00		0.93	1.00	

表 3 酒牛膝不同炮制程度判别模型评价指标

Table 3 Evaluation indicators of wine-processed *A. bidentata* discrimination models with different processing degree

算法模型	类别	训练集			预测集		
		精确率/%	召回率/%	准确率/%	精确率/%	召回率/%	准确率/%
BPNN	1	1.00	1.00	0.96	1.00	1.00	0.93
	2	0.96	1.00		1.00	0.90	
	3	0.88	0.97		0.82	0.93	
	4	1.00	0.89		0.89	0.89	
GA-BP	1	0.83	1.00	0.87	0.94	0.93	0.91
	2	0.97	0.83		0.92	1.00	
	3	0.85	0.80		0.82	0.90	
	4	0.82	0.84		0.93	0.81	
RF	1	1.00	1.00	1.00	1.00	1.00	0.91
	2	1.00	1.00		0.83	0.77	
	3	1.00	1.00		0.81	0.93	
	4	1.00	1.00		1.00	0.92	
RBFN	1	1.00	1.00	0.99	1.00	0.87	0.91
	2	0.97	1.00		0.81	1.00	
	3	1.00	0.97		1.00	0.79	
	4	1.00	1.00		0.87	1.00	
CNN	1	1.00	1.00	1.00	0.95	1.00	0.94
	2	1.00	1.00		1.00	0.75	
	3	1.00	1.00		0.93	1.00	
	4	1.00	1.00		0.91	1.00	

表 4 盐牛膝不同炮制程度判别模型评价指标

Table 4 Evaluation indicators of salt-processed *A. bidentata* discrimination models with different processing degree

算法模型	类别	训练集			预测集		
		精确率/%	召回率/%	准确率/%	精确率/%	召回率/%	准确率/%
BPNN	1	1.00	0.97	0.98	1.00	1.00	0.96
	2	0.97	1.00		0.84	1.00	
	3	1.00	0.97		1.00	0.73	
	4	0.97	1.00		1.00	1.00	
GA-BP	1	0.96	0.96	0.97	1.00	0.94	0.94
	2	0.94	0.97		0.90	0.90	
	3	0.97	0.94		0.92	0.92	
	4	1.00	1.00		0.93	1.00	
RF	1	1.00	0.97	0.98	1.00	0.94	0.91
	2	0.97	0.97		0.82	0.92	
	3	0.96	1.00		0.90	0.90	
	4	0.97	1.00		0.88	1.00	
RBFN	1	1.00	1.00	0.98	1.00	0.88	0.93
	2	0.93	0.97		0.84	1.00	
	3	0.97	0.95		1.00	0.71	
	4	1.00	1.00		0.93	1.00	
CNN	1	1.00	1.00	1.00	1.00	0.85	0.94
	2	1.00	1.00		0.90	0.90	
	3	1.00	1.00		0.92	1.00	
	4	1.00	1.00		0.95	1.00	

表 5 不同炮制品 NIRS 预处理结果

Table 5 Preprocessing results of NIRS of different processed *A. bidentata* products

光谱预处理方法	误判例数	PI	光谱预处理方法	误判例数	PI
SNV	1	92.554	MSC	1	93.004
SNV＋SG	1	95.560	MSC＋SG	1	93.014
SNV＋1stDer	1	92.554	MSC＋1ndDer	1	89.200
SNV＋SG＋1stDer	1	92.560	MSC＋SG＋1ndDer	1	89.394
SNV＋ND＋1stDer	1	92.277	MSC＋ND＋1ndDer	1	92.267
SNV＋2ndDer	50	80.322	MSC＋2ndDer	50	80.298
SNV＋SG＋2ndDer	1	91.794	SNV＋SG＋2ndDer	59	80.183
SNV＋ND＋2ndDer	1	91.794	SNV＋ND＋2ndDer	1	91.814

表 6 不同炮制品不同建模波段结果

Table 6 Results of different modeling bands for different processed *A. bidentata* products

建模波段/cm^{-1}	误判例数	PI	建模波段/cm^{-1}	误判例数	PI
4250～5000	1	93.014	4250～6000	1	90.768
4250～7000	2	91.615	4250～8000	3	91.005
4250～9000	1	90.909	4250～10 000	1	90.191
4250～11 000	4	89.505	4250～11 800	8	86.808
4250～5500	1	91.815	4250～5150	1	93.340

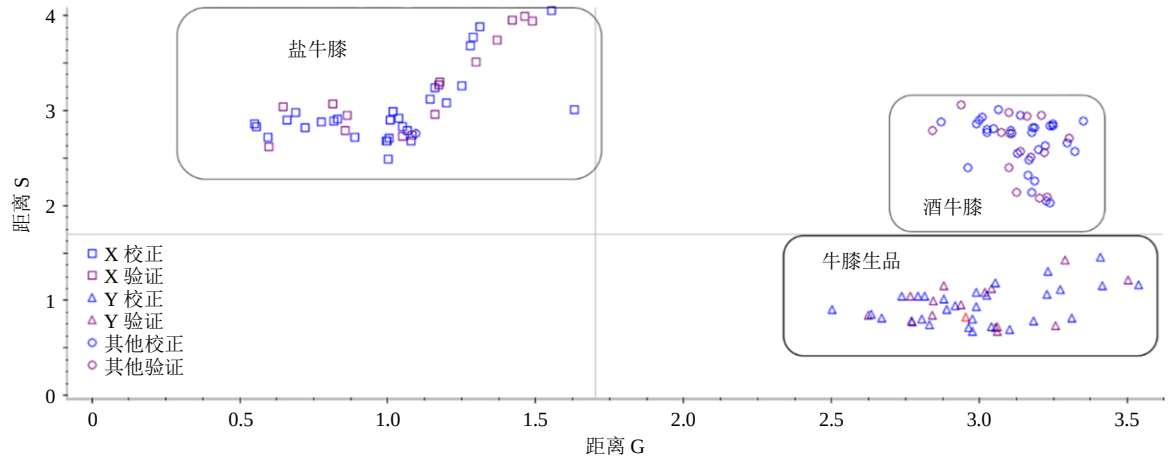


图 20　不同炮制品定性分析模型

Fig. 20　Qualitative analysis model for different processed *A. bidentata* products

理条件为 SNV＋ND＋1stDer，盐牛膝不同炮制程度定性模型的最佳预处理条件为 MSC＋SG。酒、盐牛膝不同炮制程度光谱预处理结果（表 7、8）。

表 7　酒牛膝不同炮制程度 NIRS 预处理结果

Table 7　Preprocessing results of NIRS of wine-processed *A. bidentata* with different processing degree

光谱预处理方法	误判例数	PI	光谱预处理方法	误判例数	PI
SNV	0	91.493	MSC	0	91.161
SNV＋SG	0	91.504	MSC＋SG	0	91.257
SNV＋1stDer	6	88.392	MSC＋1ndDer	5	88.229
SNV＋SG＋1stDer	5	88.936	MSC＋SG＋1ndDer	5	88.829
SNV＋ND＋1stDer	0	91.911	MSC＋ND＋1ndDer	0	91.714
SNV＋2ndDer	53	81.024	MSC＋2ndDer	54	81.256
SNV＋SG＋2ndDer	31	84.714	SNV＋SG＋2ndDer	31	84.674
SNV＋ND＋2ndDer	0	91.549	SNV＋ND＋2ndDer	0	91.391

表 8　盐牛膝不同炮制程度 NIRS 预处理结果

Table 8 Preprocessing results of NIRS of salt-processed *A. bidentata* with different processing degree

光谱预处理方法	误判例数	PI	光谱预处理方法	误判例数	PI
SNV	2	89.587	MSC	3	89.994
SNV＋SG	2	89.598	MSC＋SG	3	90.004
SNV＋1stDer	13	86.605	MSC＋1ndDer	12	86.521
SNV＋SG＋1stDer	10	86.873	MSC＋SG＋1ndDer	12	86.847
SNV＋ND＋1stDer	5	89.286	MSC＋ND＋1ndDer	5	89.207
SNV＋2ndDer	53	81.481	MSC＋2ndDer	55	81.401
SNV＋SG＋2ndDer	57	82.052	SNV＋SG＋2ndDer	59	81.936
SNV＋ND＋2ndDer	9	88.087	SNV＋ND＋2ndDer	7	88.019

以 PI 和误判例数为评价指标，考察不同波段对 NIRS 定性模型的影响，得酒牛膝不同炮制程度定性模型的最佳波段为 4150～5150 cm^{-1}，盐牛膝不同炮制程度定性模型的最佳波段为 4050～5000 cm^{-1}，酒、盐牛膝不同炮制程度、不同波段分析结果见表 9、10。

采用 TQ Analyst 软件，根据 NIRS 最佳预处理方法及最佳的光谱波段进行判别分析，建立酒、盐牛膝不同炮制程度的定性分析模型（图 21、22）。

表 9　酒牛膝不同炮制程度不同建模波段结果

Table 9　Results of different modeling bands for wine-processed *A. bidentata* with different processing degree

建模波段/cm^{-1}	误判例数	PI	建模波段/cm^{-1}	误判例数	PI
4150～5000	0	91.549	4150～6000	3	89.650
4150～7000	4	88.170	4150～8000	4	88.483
4150～9000	3	88.161	4150～10 000	4	87.065
4150～11 000	14	86.005	4150～11 800	22	84.733
4150～5200	0	91.664	4150～5150	0	91.762

表 10　盐牛膝不同炮制程度不同建模波段结果

Table 10　Results of different modeling bands for salt-processed *A. bidentata* with different processing degree

建模波段/cm^{-1}	误判例数	PI	建模波段/cm^{-1}	误判例数	PI
4050～5000	1	90.051	4050～6000	6	88.884
4050～7000	5	88.752	4050～8000	5	89.047
4050～9000	7	88.684	4050～10 000	6	88.134
4050～11 000	5	87.903	4050～11 800	11	86.076
4050～4500	23	86.152	4050～4900	2	89.083

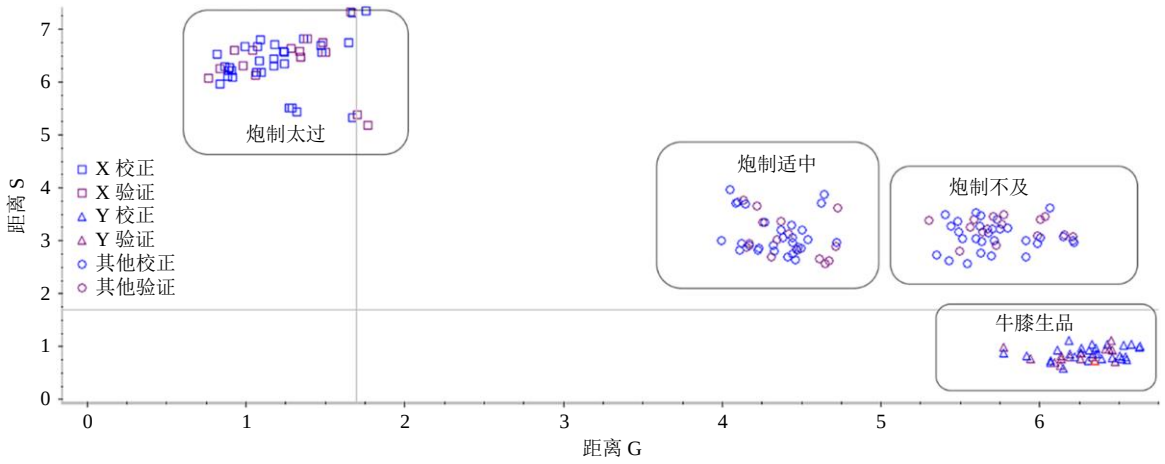


图 21　酒牛膝不同炮制程度定性分析模型

Fig. 21　Qualitative analysis model for wine-processed *A. bidentata* with different processing degree

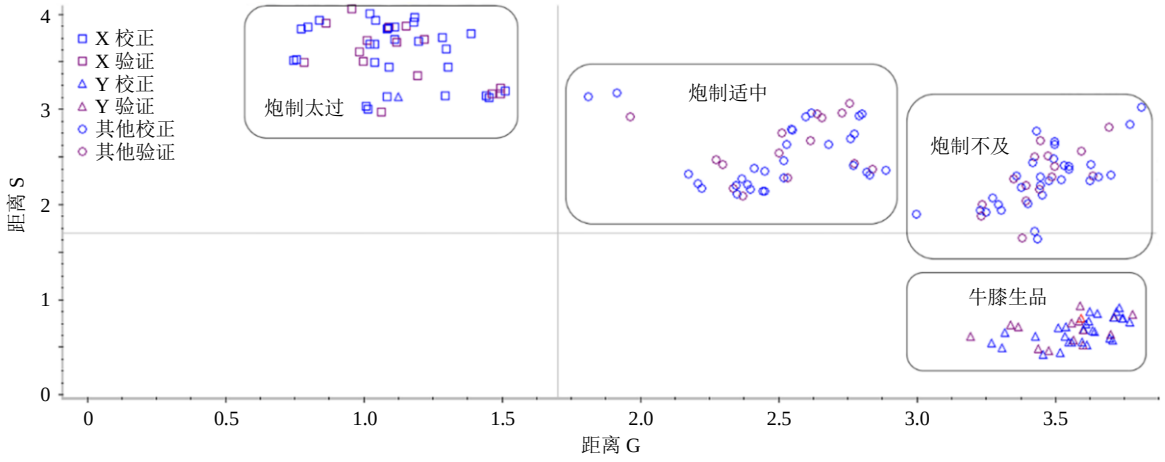


图 22　盐牛膝不同炮制程度定性分析模型

Fig. 22　Qualitative analysis model for salt-processed *A. bidentata* with different processing degree

将验证集样品的 NIRS 图谱输入所建模型，结果显示，酒、盐牛膝不同炮制程度可被准确分为 4 类，正确率为 100%。

4　讨论

本实验通过采集不同炮制品以及炮制品不同炮制程度牛膝 MIRS 图谱，使用 4 种不同的机器学习算法建立判别模型，其结果显示，当前样本量下，不同炮制品判别模型中 CNN 模型性能较好，仅在预测集中 1 个样品被错误预测，并且训练集与预测集准确度差异较小，显示出较佳的鲁棒性，BPNN、RBFN 以及 RF 算法模型准确度相差较小性能相当，GA-BP 算法模型性能相对较差；而炮制品不同炮制程度判别模型中，CNN 模型效果最好，其次为 BPNN 模型，RF 与 RBFN 模型性能相近，GA-BP 模型较差。

使用 NIRS 技术采集不同炮制品以及不同炮制程度牛膝红外图谱，建立定性模型结果显示 3 个 NIRS 定性模型验证集准确率均为 100%，可准确预测炮制品类别与炮制程度。

由表 2～4 可知，GA-BP 算法虽然使用了遗传算法对 BPNN 进行了优化，弥补了一些方面的不足，但是在本样本数据集中并没有展现优于 BPNN 的效果，这可以归结于如下原因：GA-BP 对 BPNN 最核心的改进在于通过随机搜索的方法避免了模型的局部最优解，而这一改进在数据样本相对较少，数据特征并不复杂的情况下是很难起到作用的。因为对机器学习模型而言，在一个简单低维的特征空间中求解，往往其局部最优解正是全局最优解。因此，在本实验中，由于数据样本构造的特征空间较为简便，因此模型在能够很容易找到其全局最优解，进而在实验结果上呈现出 GA-BP 没有展现优于 BPNN 的效果。

另一方面，当数据量进一步扩大时，GA-BP 或许能够有效提升 BPNN 的效果。以上论断提示提示应当根据数据集特征选择合适的算法进行建模。同时，为了提高判别模型建立的效率、准确度以及鲁棒性，可以在建模前选择合适的数据预处理方法在建模前期对数据集进行预处理，进行去噪声、基线校正、散射校正等操作，同时结合如竞争自适应重加权采样算法（competitive adaptive reweighted sampling，CARS）[30]、投影算法（successive projections algorithm，SPA）、非信息变量剔除（uninformative variables elimination，UVE）、区间偏最小二乘法（interval partial least squares，iPLS）[31]等方法选择合适的建模波段，提高建模效率。

此外，从上文数据可以看出，虽然各算法在不同炮制品以及炮制品不同炮制程度模型判别过程中表现出相当的适应性，但是随着样本类别、数量的增加，其判别准确率均有不同程度的下降，可以通过如下手段提升判别的准确性：首先，对炮制工艺进行优化，保证炮制品工艺稳定，产品合格，在确保炮制品质量均一的前提下扩大样本量。其次，与光谱照相机等多光谱成像技术相结合[32]，提升产品信息维度，得到更为饱满的产品信息，最后，可以使用如图神经网络[33]以及具有时序属性的改进神经网络等深度学习算法进行建模，最终建立准确度高、适用性广、鲁棒性佳的判别模型。

随着中医药行业的高速发展，对中药材的需求量不断提升，中药饮片质量参差不齐已成为制约中医药产业健康发展的主要因素，红外光谱技术结合化学计量学和机器学习算法可实现快速对中药材产地进行溯源、炮制品以及不同炮制程度的判别，同时结合不同来源数据进行整合分析[34]，从而明确药材来源，保证药材质量。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 王小燕, 郭常润, 常军民, 等. 怀牛膝多糖的柱前衍生化-HPLC 指纹图谱建立及单糖成分含量测定 [J]. 中国药房, 2021, 32(3): 294-300.

[2] 唐维维, 梁献葵, 马驰虹, 等. 不同采收季节怀牛膝指纹图谱研究 [J]. 中药材, 2019, 42(9): 2079-2085.

[3] 纪亮, 刘倩茹, 梁献葵, 等. 不同规格怀牛膝不同极性部位 HPLC 指纹图谱 [J]. 中国药学杂志, 2020, 55(8): 580-587.

[4] 施之琪, 朱月琴, 曹琰, 等. 基于标准汤剂的牛膝配方颗粒质量评价研究 [J]. 中药新药与临床药理, 2019, 30(7): 863-869.

[5] 翁倩倩, 赵佳琛, 金艳, 等. 经典名方中牛膝类药材的本草考证 [J]. 中国现代中药, 2020, 22(8): 1261-1268.

[6] 李思懿, 张凤玲, 王晓倩. 牛膝炮制方法的历史沿革与现代研究 [J]. 中医药管理杂志, 2022, 30(3): 19-22.

[7] 陶益, 杜映姗, 黄苏润, 等. 牛膝不同炮制品中化学成分的 UPLC-Q-TOF/MS 分析 [J]. 中国实验方剂学杂志, 2017, 23(12): 1-5.

[8] 陈露萍, 徐芳芳, 张欣, 等. 基于偏最小二乘法建立大株红景天片素片硬度近红外光谱预测模型 [J]. 中草药, 2023, 54(8): 2446-2452.

[9] Xue J T, Liu Y F, Ye L M, *et al*. Rapid and simultaneous analysis of five alkaloids in four parts of *Coptidis Rhizoma* by near-infrared spectroscopy [J]. *Spectrochim Acta A Mol Biomol Spectrosc*, 2018, 188: 611-618.

[10] 黄志伟, 郭拓, 黄文静, 等. 近红外光谱技术在名贵中药材质量评价中的研究进展 [J]. 中草药, 2022, 53(20): 6328-6336.

[11] 姜泽明, 周甜甜, 卜洪洋, 等. 落叶松树皮原花青素生产过程的红外光谱分析 [J]. 光谱学与光谱分析, 2018, 38(1): 62-67.

[12] 田胜尼, 李亚楠, 胡艺璇, 等. 安徽齐云山石斛傅里叶红外光谱分析 [J]. 生物学杂志, 2021, 38(6): 65-69.

[13] 郑司浩, 赵莎, 曾燕, 等. 中药材品种与产地鉴别研究现状与思考 [J]. 中国现代中药, 2021, 23(12): 2037-2045.

[14] 李超, 李孟芝, 李丹霞, 等. 基于傅里叶变换红外光谱指纹技术的艾叶产地溯源研究 [J]. 光谱学与光谱分析, 2022, 42(8): 2532-2537.

[15] 王小鹏, 张璐, 陈鹏举, 等. 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 [J]. 中草药, 2023, 54(4): 1076-1086.

[16] 赖长江生, 周融融, 余意, 等. 基于近红外分析和化学计量学方法对不同产地灵芝快速鉴别及多糖含量测定的研究 [J]. 中国中药杂志, 2018, 43(16): 3243-3248.

[17] 张振宇, 常相伟, 严辉, 等. 基于近红外光谱分析技术的干姜质量快速评价研究 [J]. 中草药, 2022, 53(23): 7516-7523.

[18] 贾豪, 雷益铭, 张维方, 等. 牛膝药材的红外指纹图谱建立及多元统计分析 [J]. 中国药房, 2022, 33(2): 153-159.

[19] 中国药典 [S]. 四部. 2020: 31.

[20] 方翔, 侯淑萍, 刘琐, 等. 基于 BP 神经网络算法和公式法纠正黄疸对仪器测定血红蛋白的影响及探讨 [J]. 中国卫生检验杂志, 2022, 32(18): 2233-2236.

[21] Xie F Y, Fan H D, Li Y, *et al*. Melanoma classification on dermoscopy images using a neural network ensemble model [J]. *IEEE Trans Med Imaging*, 2017, 36(3):

849-858.

[22] 孙炬仁. 基于遗传算法优化 BP 神经网络下马铃薯产量预测模型 [J]. 农机化研究, 2023, 45(6): 53-57.

[23] 于旭峰, 李红梅, 卓伟, 等. 基于近红外光谱技术的马铃薯叶片含水率高效预测 [J]. 光学仪器, 2020, 42(4): 7-13.

[24] Lam C, Calvert J, Siefkas A, *et al*. Personalized stratification of hospitalization risk amidst COVID-19: A machine learning approach [J]. *Health Policy Technol*, 2021, 10(3): 100554.

[25] 汤卫东, 肖大军, 谈林涛, 等. 机器学习下随机森林算法在电网故障分析指挥系统中的应用 [J]. 计算技术与自动化, 2022, 41(3): 59-63.

[26] 冯麟涵, 杨俊杰, 焦立启. 基于 RBF 神经网络的船舶冲击谱速度数据挖掘与预报 [J]. 振动与冲击, 2022, 41(13): 189-194.

[27] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229-1251.

[28] 何力, 郑灶贤, 项凤涛, 等. 基于深度学习的文本分类技术研究进展 [J]. 计算机工程, 2021, 47(2): 1-11.

[29] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks [J]. *Inf Process Manag*, 2009, 45(4): 427-437.

[30] Li H D, Liang Y Z, Xu Q S, *et al*. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. *Anal Chim Acta*, 2009, 648(1): 77-84.

[31] Zou X B, Zhao J W, Povey M J W, *et al*. Variables selection methods in near-infrared spectroscopy [J]. *Anal Chim Acta*, 2010, 667(1/2): 14-32.

[32] 吴刚, 彭要奇, 周广奇, 等. 基于多光谱成像和卷积神经网络的玉米作物营养状况识别方法研究 [J]. 智慧农业: 中英文, 2020, 2(1): 111-120.

[33] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述 [J]. 计算机学报, 2020, 43(5): 755-780.

[34] 赵倩, 缪培琪, 李小莉, 等. 数据融合技术在中药分析领域中的应用进展 [J]. 中草药, 2023, 54(11): 3706-3714.

[责任编辑 郑礼胜]