

谷传凯, 褚璇, 刘洪利, 等. 基于高光谱技术的金线莲多糖与黄酮含量的无损检测 [J]. 食品工业科技, 2025, 46(7): 227–234. doi: 10.13386/j.issn1002-0306.2023100154

GU Chuankai, CHU Xuan, LIU Hongli, et al. Non-destructive Detection of Polysaccharide and Flavonoid Contents in *Anoectochilus roxburghii* Using Hyperspectral Technology[J]. Science and Technology of Food Industry, 2025, 46(7): 227–234. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2023100154

· 分析检测 ·

基于高光谱技术的金线莲多糖与黄酮含量的无损检测

谷传凯¹, 褚璇^{1,*}, 刘洪利¹, 韦鸿钰¹, 牟英辉², 马稚昱¹
(1. 仲恺农业工程学院机电工程学院, 广东广州 510225;
2. 华南农业大学工程学院, 广东广州 510642)

摘要: 为了无损、快速地实现金线莲叶片中多糖与黄酮含量水平的判别, 本研究以在不同光周期下 (10、12、14、16、18、20 h/d) 栽培的金线莲为样本, 通过高光谱技术获取叶片像素光谱数据, 运用传统机器学习 (PCA-LDA 与 PCA-SVM) 与深度学习方法 (1D CNN 及其优化方法) 构建了相应的判别模型。结果表明, 1D CNN 模型在训练集、验证集与独立测试集上的判别准确率优于 PCA-LDA 与 PCA-SVM 模型, 分别为 99.99% 与 99.89%、99.98% 与 99.78%、91.62% 与 87.92%。通过在 1D CNN 模型中引入 Dropout 层, 模型的泛化能力得到增强, 多糖与黄酮含量水平在独立验证集上的判别准确率分别提升至 98.92% 与 95.67%。基于此, 本研究进一步构建了多糖与黄酮含量水平的可视化图像, 通过颜色直观展示了不同含量水平的判别结果。研究证实了高光谱技术可用于评估金线莲叶片中多糖与黄酮的含量水平, 研究结果可为金线莲的品质控制提供技术支撑。

关键词: 无损检测, 深度学习, 机器学习, 金线莲, 多糖与黄酮含量

中图分类号: O433.4; TQ936.2 文献标识码: A 文章编号: 1002-0306(2025)07-0227-08

DOI: 10.13386/j.issn1002-0306.2023100154



本文网刊:

Non-destructive Detection of Polysaccharide and Flavonoid Contents in *Anoectochilus roxburghii* Using Hyperspectral Technology

GU Chuankai¹, CHU Xuan^{1,*}, LIU Hongli¹, WEI Hongyu¹, MU Yinghui², MA Zhiyu¹

(1. College of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China;

2. College of Engineering, South China Agricultural University, Guangzhou 510642, China)

Abstract: This study aimed to rapidly and non-destructively evaluate the levels of polysaccharides and flavonoids in *A. roxburghii* leaves under various photoperiods (10, 12, 14, 16, 18, and 20 h/d). Hyperspectral imaging was employed to acquire pixel-level spectral data from the leaves, and discriminant models for content levels were developed using traditional machine learning methods (PCA-LDA and PCA-SVM) and deep learning approaches (1D CNN and its optimized version). The findings revealed that the 1D CNN model outperformed the PCA-LDA and PCA-SVM models in terms of discrimination accuracy on the training, validation, and independent test sets, achieving 99.99% and 99.89%, 99.98% and 99.78%, and 91.62% and 87.92%, respectively. The introduction of a Dropout layer in the 1D CNN model enhanced its generalization capability, increasing the discrimination accuracy for polysaccharide and flavonoid content levels on the independent test set to 98.92% and 95.67%, respectively. Additionally, visualization images depicting the discrimination results for different compound levels were constructed, providing an intuitive representation. This study validates the feasibility of hyperspectral imaging in evaluating polysaccharide and flavonoid levels in *A. roxburghii* leaves

收稿日期: 2023-10-19

基金项目: 国家自然科学基金青年科学基金项目 (32102087); 广州市科技计划项目 (2023A04J1667)。

作者简介: 谷传凯 (1998-), 男, 硕士研究生, 研究方向: 作物表型检测, E-mail: 201987769@qq.com。

* 通信作者: 褚璇 (1990-), 女, 博士, 副教授, 研究方向: 基于高光谱技术的农产品无损检测技术, E-mail: cauchx0105@163.com。

cultivated under various photoperiods, and the research results can provide technical support for the quality control of *A. roxburghii*.

Key words: non-destructive testing; deep learning; machine learning; *Anoectochilus roxburghii*; polysaccharides and flavonoids contents

金线莲(*Anoectochilus roxburghii*)是一种多年生兰科植物,亦是一种珍稀中草药,在医药、保健品、化妆品、食品等领域具有广泛应用^[1-4]。多糖和黄酮类化合物是其主要的活性成分,也是评价其品质的关键指标^[5-8],因此实现这些活性成分含量的评估至关重要。传统的理化检验方法,如紫外分光光度法、高效液相色谱法及气相色谱-质谱联用法等,尽管在准确性、可靠性和灵敏度方面表现出色,但通常需要较为昂贵的设备,复杂的试剂处理以及耗时的检测过程,不适用于大规模快速检测。因此,开发一种无损、快速的检测方法,对于确保金线莲品质和指导其栽培具有重要意义。

高光谱成像技术将图像信息和光谱信息相融合,被广泛应用于植物中叶绿素、氮素、水分及农药残留等成分监测^[9-12]。在多糖与黄酮等有效成分检测方面,高光谱成像技术也发挥重要作用,如应用于灵芝粉^[13]、桑果^[14]、杏鲍菇^[15]与小麦^[16]中的多糖,枇杷叶片^[17]、刺五加叶片^[18]中的黄酮,以及菊花^[19]中总多糖与总黄酮的含量预测等。除了实现有效成分的检测,也有研究对有效成分的含量分布进行了可视化显示。例如,余克强等^[20]实现了尖椒叶片中氮素分布的可视化分布显示;石吉勇等^[21]与李芳等^[22]均通过颜色差异可视化展示了黄酮在银杏叶片中的分布规律。尽管目前关于金线莲多糖与黄酮高光谱检测的报道较少,但上述研究为金线莲中有效成分的检测提供了参考。

本研究旨在快速、无损地评估金线莲中多糖与黄酮含量水平。以不同光周期下生长的金线莲为对象,基于可见近红外波段下的高光谱数据,结合传统机器学习与深度学习技术,建立多糖和黄酮含量水平的判别模型;绘制出多糖与黄酮不同含量水平的叶片中可视化图像,实现成分含量水平的直观显示。预期结果将为金线莲栽培的精细管理和品质调控提供参考。

1 材料与方法

1.1 材料与仪器

金线莲 地缘(厦门)生物科技有限公司;芦丁、无水葡萄糖 纯度均不低于 99%,中国索莱宝科技公司;氢氧化钠、无水乙醇 天津市百世化工有限公司;苯酚、浓硫酸 广州化学试剂厂;所有分离用有机溶剂均为国产分析纯。

高光谱成像系统 配有 GaiaSky mini 近红外高光谱相机与特制暗箱,中国四川双利合谱科技有限公司;UV-1800 紫外光-可见光分光光度计 日本岛津公司;KQ5200E 超声波清洗器 昆山市超声仪器有

限公司;DK-8AD 电热恒温水槽 上海一恒科学仪器有限公司;HR/T16M 台式高速冷冻离心机 湖南赫西仪器装备有限公司。

1.2 实验方法

1.2.1 样本准备 金线莲组培苗经过两周炼苗处理,挑选长势一致的健康苗株移栽至广东省仲恺农业工程学院人工栽培室进行培养。培养基质配制比例为 6:2:1(基质:蛭石:树皮),环境条件控制在日间温度 24 ℃、夜间 20 ℃,相对湿度 60%;光源采用 1:1 比例的蓝光和红光(B1R1),光强设置为 $40 \pm 2 \mu\text{mol}/(\text{m}^2 \cdot \text{s})$,光周期分别设置为 10、12、14、16、18 和 20 h/d,共 6 种处理。约 100 d 后,从每个光周期处理下选取 5 株金线莲进行光谱采集与分析。

1.2.2 高光谱数据采集 如图 1 所示,高光谱成像系统配置包括一台相机,安装于暗箱顶,与升降平台保持 50 cm 的距离。暗箱四角均匀分布四个 50 W 的钨卤灯光源,以 45°的角度投射光线。高光谱相机的光谱范围及分辨率分别为 386~1016 nm 与 3.5 nm,图像像素为 696×700。采集光谱时,系统设置为漫反射模式,曝光时间与推扫速度分别设定为 9.98 ms 和 0.8 mm/s。为降低光照不均和暗电流等因素的影响,采集前先获取标准黑、白背景数据,用于后续叶片数据黑白校正。其中,白背景通过拍摄标准白板获得,而黑背景则通过覆盖高光谱成像仪的镜头拍摄获得。最后,选取金线莲植株的第二片子叶,将叶片朝向一致摆放于样品台上,依次采集其高光谱图像。

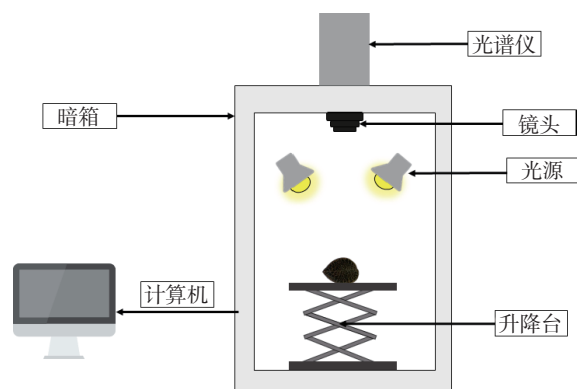


图 1 高光谱相机系统示意图

Fig.1 Schematic of the hyperspectral camera system

1.2.3 多糖与黄酮含量的理化测定 参考黄瑞平等^[23]的方法完成金线莲中多糖与黄酮的测定。首先以无水葡萄糖和芦丁为标准品,制备相应的多糖和黄酮标准溶液。使用紫外光-可见光分光光度计在 490 nm 与 510 nm 的波长下,分别测定所配制不同质量浓度的标准品溶液的吸光值,以试剂空白溶液为

参比,质量浓度为横坐标,吸光度为纵坐标,绘制标准曲线,据此得到多糖和黄酮含量测定的回归方程(式(1)和(2))。为确保测定的准确性,要求回归方程的决定系数 $R^2 > 0.999$ 。

$$Y_p = 12.78X_p - 0.06, R^2 = 0.9993 \quad \text{式 (1)}$$

$$Y_f = 56.88X_f - 0.01, R^2 = 0.9996 \quad \text{式 (2)}$$

式中, Y_p 与 Y_f 分别为检测多糖与黄酮时的吸光度值; X_p 与 X_f 分别为样本多糖与黄酮的质量浓度, mg/mL。

1.2.3.1 多糖含量测定 取金线莲干样,粉碎过 40 目筛,称取 0.5 g(精度保持在 0.01 g 范围内),置于 10 mL 玻璃试管中并添加 5 mL 蒸馏水,将试管置于 80 °C 水浴锅中加热处理 90 min,然后冷却至室温,继续添加蒸馏水,并定容至 10 mL,提取液摇匀后进行过滤。为排除初段滤液中可能存在的杂质对测定结果的影响,弃初段滤液 1 mL,收集余下滤液。吸取余下滤液 1 mL,加入 7.5 mL 无水乙醇,以 6000 r/min 的转速进行离心,仅保留试管中的沉淀,并用 90% 乙醇进行二次洗涤和二次离心(6000 r/min),留取沉淀物并加水溶解,精确定容至 10 mL,可得多糖检测样品液。从样品液中吸取 0.2 mL,按多糖标准品的测量方法进行吸光度测定,带入回归方程(1)计算。进一步基于公式(3)可得金线莲样品中多糖含量。

$$T_p = \frac{X_p \times V_{ip} \times \frac{V_{bp}}{V_{sp}} \times \frac{V_{ip}}{V_{ip}}}{m_p} \quad \text{式 (3)}$$

式中, T_p 为多糖含量(mg/g); V_{ip} 为检测溶液的总体积(mL); V_{bp} 为最终样品溶液的体积(mL); V_{sp} 为用于测量的样品溶液体积(mL); V_{ip} 为金线莲干样制备的溶液体积(mL); V_{ip} 是离心的剩余滤液体积(mL); m_p 为金线莲干样粉末的质量(g)。

1.2.3.2 黄酮含量测定 取金线莲干样,粉碎过 40 目筛,称取 0.5 g,置于 25 mL 的玻璃试管中,加入无水乙醇 20 mL,进行超声处理 2 h(功率 350 W,频率 50 Hz,温度 80 °C)后用滤纸过滤。随后用无水乙醇定容至 10 mL,得黄酮检测样品液。取 2 mL 样品液,按照测量黄酮标准品的方法测量其吸光度,然后带入回归方程(2)进行计算。进一步基于公式(4)可得金线莲样品中黄酮含量。

$$T_f = \frac{X_f \times V_{if} \times \frac{V_{vf}}{V_{sf}}}{m_f} \quad \text{式 (4)}$$

式中, T_f 为黄酮含量(mg/g); V_{if} 为检测溶液的总体积(mL); V_{vf} 为最终样品溶液的体积(mL); V_{sf} 为用于测量的样品溶液体积(mL); m_f 为金线莲干样粉末的质量(g)。

1.3 数据处理

为准确判定金线莲中多糖和黄酮的含量水平,本研究从叶片的高光谱图像中提取了每个像素的光

谱数据,采用机器学习与深度学习方法,分别构建了用于区分不同含量水平的判别模型,并对模型性能进行了评估。

1.3.1 高光谱数据校正 基于叶片的原始高光谱数据以及标准白背景与黑背景数据,通过式(5)对叶片高光谱数据进行校正。

$$R = \frac{R_0 - B}{W - B} \quad \text{式 (5)}$$

式中, R 为校正后的叶片高光谱图像; R_0 为原始采集的叶片高光谱图像; W 与 B 分别为标准白背景和黑背景图像。

1.3.2 光谱数据提取 使用 ENVI 软件对校正后的叶片高光谱数据进行主成分分析(Principal Component Analysis, PCA)。基于前两个主成分(PC1 和 PC2)生成的得分图,配合软件中的感兴趣区域选取(Region of Interest, ROI)工具,将主成分得分图中散点与叶片像素进行对应标记,实现叶片区域的掩膜选取与背景信息的移除。随后通过 MATLAB R2016a 软件提取叶片区域每个像素点的光谱数据。

1.3.3 模型建立与评估

1.3.3.1 传统机器学习建模 基于提取的叶片像素光谱数据,建立多糖与黄酮含量水平的主成分判别分析(Principal Component Analysis-Discriminant Analysis, PCA-LDA)和支持向量机(Support Vector Machine, SVM)判别模型。PCA-LDA 建模方法是在 PCA 处理后,选取关键主成分(Principal Component, PCs)结合 LDA 线性分类器以构建分类模型。SVM 模型则通过特定的传递核函数将数据映射到高维特征空间,并构建最优超平面以实现不同类别的划分^[24]。此外,本研究在建立 SVM 模型过程中将 PCA 与 SVM 进行了结合^[25]。对原始数据进行降维处理后,将 PCA 处理后的数据输入 SVM 模型,实现判别。

1.3.3.2 深度学习建模 除了传统的机器学习方法,本研究还采用了卷积神经网络(Convolutional Neural Network, CNN)来判别金线莲中多糖与黄酮的含量水平。CNN 主要由输入层、卷积层、激活层、池化层、全连接层和输出层等构成。输入层接收金线莲的一维像素光谱数据,而输出层负责预测多糖、黄酮的含量水平。由于输入为一维光谱数据,因此本研究选用一维卷积来提取数据特征,并生成特征映射。池化层用于减小特征图的尺寸,降低数据维度^[26],提升模型的计算效率。全连接层则将这些特征映射到最终的类别输出。此外在 CNN 中引入激活函数来增强网络的非线性能力^[27]。对于多糖和黄酮含量水平的判别,构建的 1D CNN 模型结构相同,均包含一个卷积层(Conv1d)、一个最大池化层(Max Pooling)、一个过渡层(Transition Layer)和两个全连接层(Fully Connected Layer)。

其中,卷积层激活函数、内核的大小、内核数量

与步幅分别设置为 ReLU、4×1、32 与 1。池化层采用最大池化法,使用 2×1 的池化核,并设置输出单元数量为 16。过渡层的作用是将多维数据展平,实现从卷积层到全连接层的平滑过渡。最后一个全连接层采用 SoftMax 激活函数,输出模型的预测结果。在模型编译阶段,选用 Categorical_Crossentropy 作为损失函数,以评估模型性能。优化器选用 Adam,以提升收敛速度,确保模型能够有效执行分类任务。训练过程中,训练集被分为 16 批次,每个批次包含 16(Batch Size)样本。构建的模型网络结构和详细参数如表 1 所示。

表 1 1D CNN 模型主要结构参数
Table 1 Main structural parameters of 1D CNN model

层数	输入	参数	输出	步长	激活函数
1	256×1×1	Conv1d: 64×4×1	1×256×64	1	ReLU
2	256×64×1	Dense	256×16×1	—	ReLU
3	256×16×1	MaxPooling1D: 128×2×1	128×16×1	—	—
4	128×16×1	Flatten	2048×1×1	—	—
5	2048×1×1	Dense	4×1	—	SoftMax

1.3.3.3 模型评估 模型的判别性能通过计算训练集、验证集和独立验证集的总体准确率来评估,即正确分类样本占总样本的百分比。总体准确率的计算方法如式(6)所示:

$$\text{Accuracy}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100 \quad \text{式 (6)}$$

式中, TP(True Positive)表示实际为正类被正确预测为正类的数量; FN(False Negative)表示实际为正类却被错误预测为负类的数量; FP(False Positive)表示实际为负类却被错误预测为正类的数量; TN (True Negative)表示实际为负类被正确预测为负类的数量。

2 结果与分析

2.1 多糖与黄酮含量的分析

不同光周期处理下多糖与黄酮含量的均值如图 2 所示。由图可知,光周期对多糖和黄酮含量累积具有极显著影响($P < 0.01$)。在不同光周期处理下,多糖含量的变化范围为 43.36~79.75 mg/g,黄酮含量介于 15.76~19.56 mg/g。随着光照时间的延长,两种成分的含量均呈现先增后减的趋势,且均在 14 h/d 光照处理下达到最高值。基于多重比较的结果,金线莲中多糖含量被分为三个水平分组: I 组(14 h/d 处理组), II 组(10、12 及 20 h/d 处理组)和 III 组(16、18 h/d 处理组);黄酮含量则被分为四个水平分组: I 组(14 h/d 处理组), II 组(12 h/d 处理组), III 组(16 h/d 处理组)和 IV 组(10、18 及 20 h/d 处理组)。

2.2 光谱特征分析

为比较不同光周期下叶片的光谱特征,参照 1.2.2 小节所述方法,提取了叶片区域每个像素点的光谱数据,并计算了不同光周期下所有叶片像素光谱

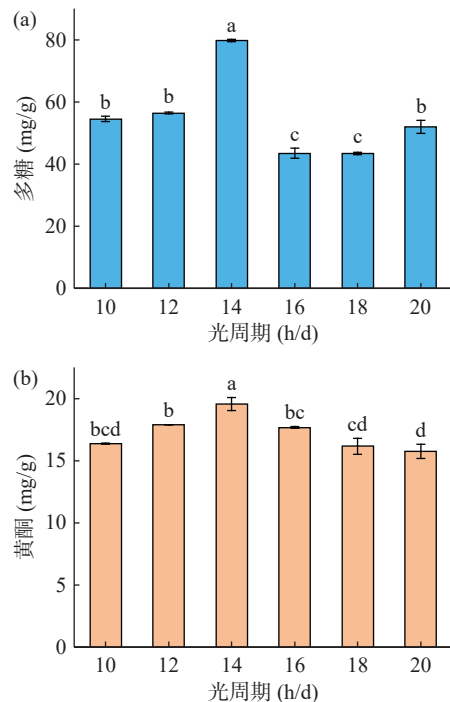


图 2 不同光周期下金线莲多糖(a)与黄酮(b)含量

Fig.2 Contents of polysaccharides (a) and flavonoid (b) in *A. roxburghii* under various photoperiods

注: 不同的字母表示各组均值在 0.01 置信水平上差异显著。

的平均值,绘制的不同光周期样本的平均光谱如图 3 所示。光谱曲线呈现的趋势与一般绿色植物叶片的光谱特征相吻合。在约 560 nm 波长处的波峰与花青素成分相关^[28];在 680~780 nm 范围内,光谱反射率急剧上升并保持在较高水平,这一现象体现了植物光谱的“红边”效应,即从红光波段过渡到近红外波段时,反射率显著提高,光谱曲线呈现出一条直线的现象。这一变化主要是由于叶绿素在红光波段的强烈吸收,以及在近红外波段叶片内部结构对光的散射和反射强烈所致^[29-30]。此外,970 nm 处的波谷与叶片中水分的 O-H 键伸缩的二阶泛音相关^[31]。在整个波段范围内,不同光周期下叶片的光谱曲线显示出较高的重合度,无法从原始光谱曲线上直接体现出光周期影响下的多糖或黄酮含量水平的差异。

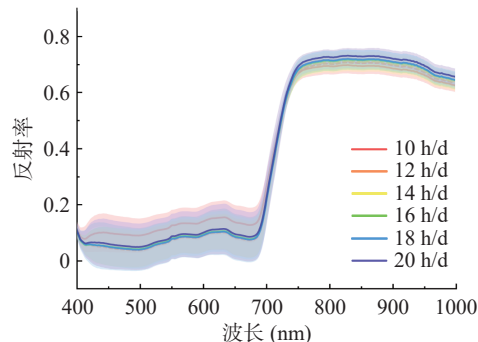


图 3 不同光周期处理下金线莲叶片的光谱

Fig.3 Spectra of *A. roxburghii* leaves under different photoperiods

为进一步探讨基于近红外光谱数据区分多糖与

黄酮含量水平的可行性,从每个含量水平的叶片中分别随机抽取一片,并对叶片区域所有像素的光谱数据进行主成分分析,以主成分得分来表征不同多糖与黄酮含量水平的信息。通过 MATLAB 软件绘制每个像素的前三个主成分得分的空间分布图,如图 4 所示。从图中可以看出,不同多糖含量水平叶片的像素得分沿着 PC2 轴大致呈从负到正方向分布趋势,但各水平间的边界不够清晰,仍存在较多的交叉;对于不同黄酮含量水平的叶片,样本间的区分并不明显,水平 I 类与 IV 类叶片的像素呈现分离趋势,水平 II 与 III 类叶片像素间则交叉重叠,难以区分。鉴于此,为进一步实现多糖与黄酮含量水平的判别,本研究进一步运用机器学习与深度学习方法,通过建立相应的判别模型进行深入分析。

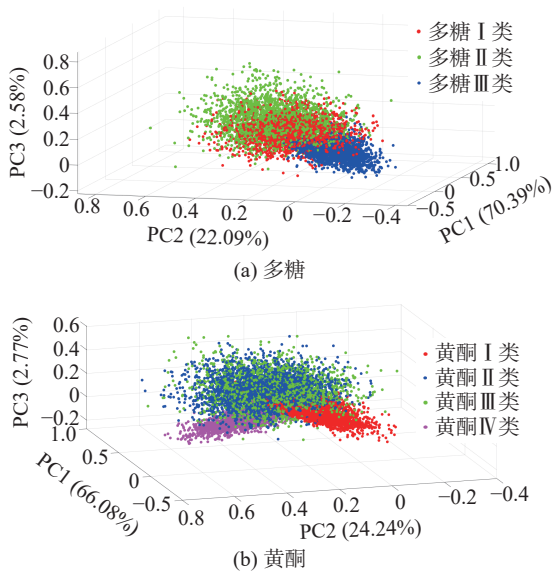


图 4 不同有效成分含量水平下每个叶片像素的主成分散点图

Fig.4 Principal components scatter plots of each pixel in leaves at different levels of bioactive component

2.3 基于传统机器学习的含量水平的判别模型

建立多糖与黄酮含量水平判别模型前,首先构建模型训练所需的训练、验证与独立验证样本集。将每个光周期处理下的叶片均按 3:1:1 比例划分,得到 6 个光周期处理下共用于训练、验证和独立验证集的叶片数分别为 18、6 和 6。进一步提取每个叶片上所有的像素光谱数据,作为最终的训练、验证、独立验证集样本数据。在进行 PCA 处理后,分别选择 LDA 与 SVM 分类器建立基于机器学习方法的判别模型,并以 PCA 所得的主成分作为分类器输入。表 2 展示了 PCA-LDA 与 PCA-SVM 模型的输入变量和对多糖与黄酮含量水平的判别结果。

表 2 中可以看出,在针对多糖和黄酮两种成分水平的判别中,PCA-LDA 模型在训练集与验证集上的准确率虽均超过了 90%,但在独立验证集中,该模型对多糖含量水平的判别准确率降至 87.40%,对黄酮含量的判别准确率也仅为 76.25%。PCA-SVM 模

表 2 多糖与黄酮含量水平不同判别模型下的测试结果

Table 2 Results of different discrimination models for polysaccharide and flavonoid content levels

成分	建模方法	输入变量数	训练集(%)	验证集(%)	独立验证集(%)
多糖	PCA-LDA	10 PCs	92.43	92.00	87.40
	PCA-SVM	12 PCs	99.94	83.00	77.55
	1D CNN	256 波长	99.99	99.98	91.62
黄酮	PCA-LDA	10 PCs	99.44	91.52	76.25
	PCA-SVM	12 PCs	97.40	80.00	79.07
	1D CNN	256 波长	99.89	99.78	87.92

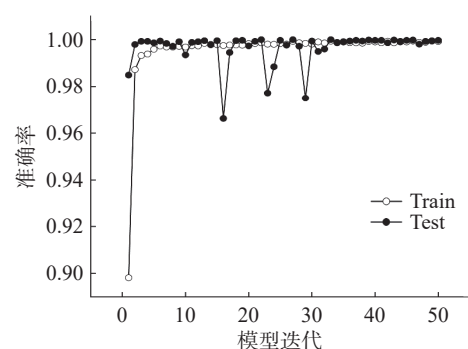
型在训练集展现出高于 97% 的判别准确率,但在独立测试集上的准确率仅在 80% 左右,表明存在过拟合现象。这可能是由于训练过程中输入的光谱数据量大而模型结构相对简单,导致的模型识别能力不足^[32]。此外,PCA-SVM 模型对于独立验证集多糖与黄酮含量的准确率分别为 77.55% 与 79.07%。这些结果均表明了利用光谱数据实现两种成分含量水平的判别是可行的,但 PCA-LDA 与 PCA-SVM 模型的泛化性能较差。因此,后续采用了深度学习方法进一步挖掘数据特征并建立模型。

2.4 基于深度学习的含量水平的判别模型

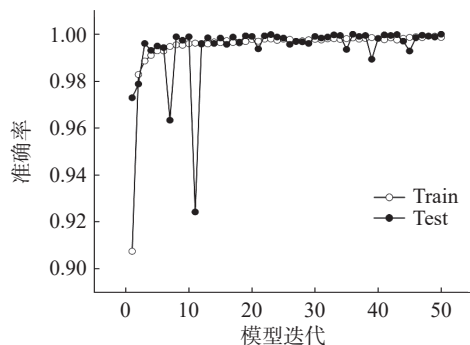
2.4.1 一维卷积神经网络(1D CNN)判别模型 1D CNN 模型在多糖与黄酮含量水平的判别结果如表 2 所示。所建模型在训练集与验证集上对两种成分含量水平的判别准确率均超过了 99%;而对于独立验证集,判别准确率分别降至 91.62% 与 87.92%。尽管与传统建模方法相比,1D CNN 表现出更好的判别效果,但其泛化能力仍需进一步提升。绘制其准确率与损失函数曲线,如图 5 所示。对多糖与黄酮含量预测两个模型,在训练集中准确率和损失函数曲线均较为平滑,且随着迭代次数的增加分别呈现增加/降低并逐渐稳定的趋势。但在验证集的准确率和损失函数在迭代过程中却出现显著波动,尤其在多糖模型中更为明显。这可能是由于模型训练过程中过度拟合了数据中的误差信息,如随机噪声等,导致验证集上的性能不稳定,出现了过拟合现象^[33]。

2.4.2 优化后的 1D CNN 判别模型 为了优化模型,将 Dropout 引入模型构建,其可以随机丢弃模型中的部分神经元以及其连接,以防止单元过度共同适应。在优化的 1D CNN 模型中,卷积层和池化层的参数保持不变,仅在最后的展平层与输出层之间添加了 Dropout 层。多糖判别模型的训练分批为 64 批次,每批 64 个样本。黄酮判别模型则分为 16 批次,每批 16 个样本。

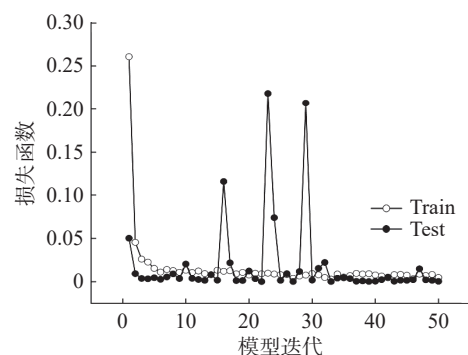
优化后的模型在训练集、验证集和独立验证集上对多糖含量水平的判别准确率分别达到 99.95%、99.97% 和 98.92%,相应的,黄酮含量水平的判别准确率分别为 99.93%、99.94% 和 95.67%。与未优化的模型相比,独立验证集上的准确率分别提高了 7.30% 和 7.75%,说明 Dropout 对模型有较好的修正



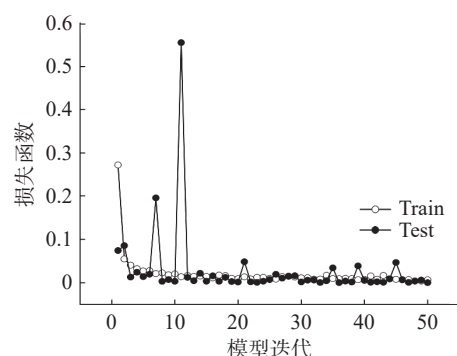
(a) 多糖1D CNN模型的准确率



(b) 黄酮1D CNN模型的准确率



(c) 多糖1D CNN模型的损失函数

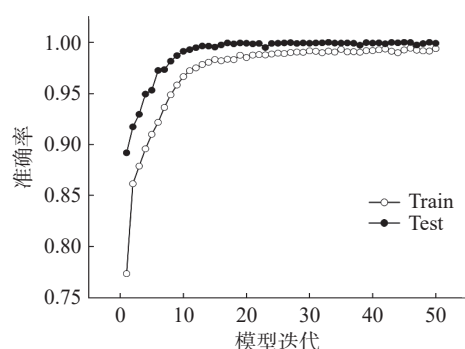


(d) 黄酮1D CNN模型的损失函数

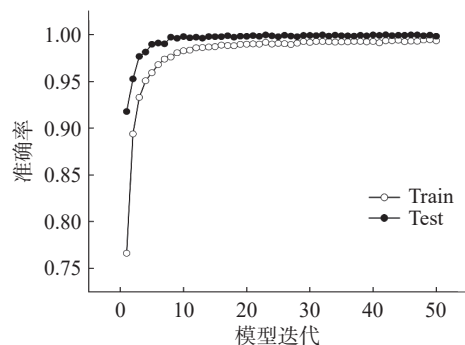
图5 1D CNN模型对多糖与黄酮含量水平的判别性能与模型迭代次数的关系

Fig.5 Relationship between discrimination performance of the 1D CNN model on polysaccharide and flavonoid content levels and the number of model iterations

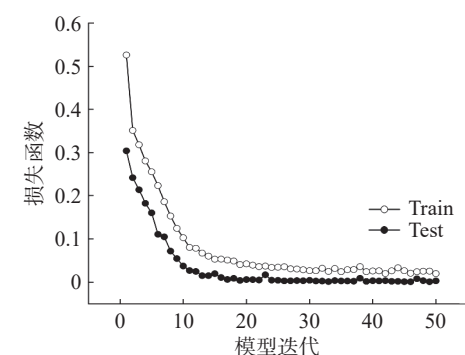
作用。此外,优化后模型的准确率和损失函数曲线如图6所示。可以看出,随着迭代次数的增加,准确率稳步提升,损失函数逐渐下降至稳定状态。与优化前相比,验证集的性能指标得到改善,训练集与验证集的曲线更为平滑。



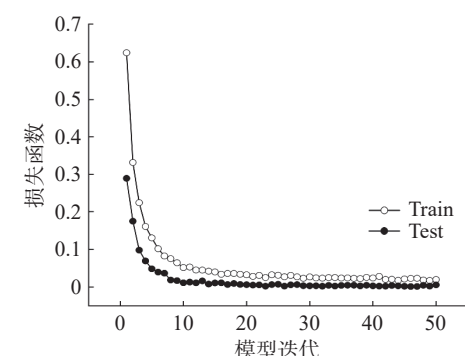
(a) 优化后多糖判别模型的准确率



(b) 优化后黄酮判别模型的准确率



(c) 优化后多糖判别模型的损失函数



(d) 优化后黄酮判别模型的损失函数

图6 优化后 1D CNN 模型对多糖与黄酮含量水平的判别性能与模型迭代次数的关系

Fig.6 Relationship between discrimination performance of the optimized 1D CNN model on polysaccharide and flavonoid content levels and the number of model iterations

2.5 成分含量水平的可视化判别

多糖与黄酮含量水平的判别模型均基于像素光谱数据构建,将像素判别结果映射回相应像素位置,通过不同颜色标记预测值类别,实现叶片含量水平的可视化判别,其中蓝色表示低含量水平,红色表示高

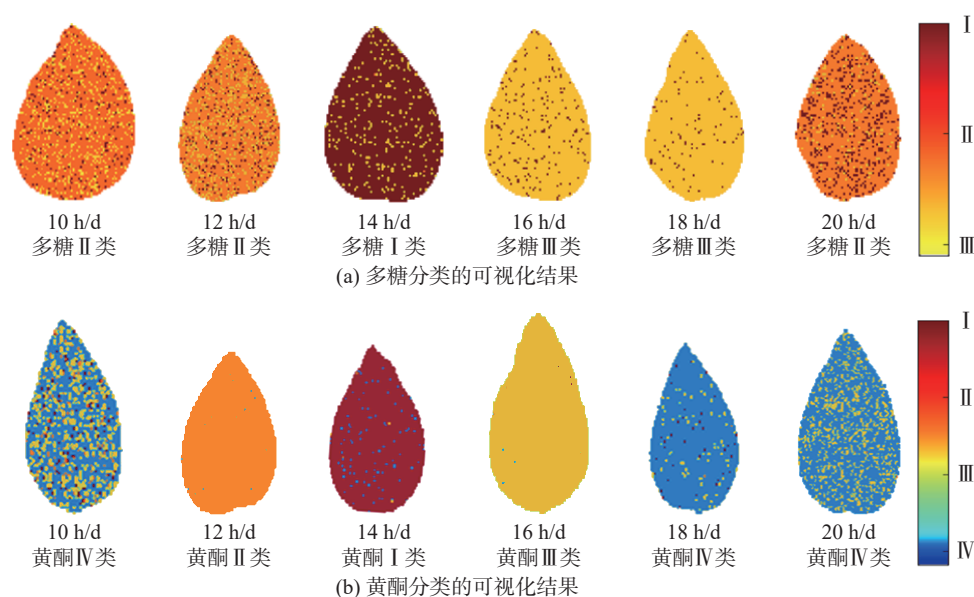


图 7 金线莲叶片多糖和黄酮含量水平的可视化判别

Fig.7 Visualization of polysaccharide and flavonoid content in *A. roxburghii*

含量水平。在不同光周期处理下,独立验证样本叶片的可视化结果如图 7 所示。尽管叶片上存在部分像素的误判,但并不影响对叶片整体含量水平的准确判别。在多糖含量判别方面,14 h/d 光照处理下的叶片被判定为 I 类,10、12、20 h/d 处理下的叶片被判定为 II 类,16、18 h/d 下的叶片被判定为 III 类;黄酮含量判别中,14 h/d 下的叶片被判定为 I 类,12 h/d 下的叶片被判定为 II 类,16 h/d 下的叶片被判定为 III 类,10、18 与 20 h/d 下的叶片被判定为 IV 类。这些判别结果与实际分组一致,证明了模型对多糖与黄酮含量水平具有良好的预测能力。此外,可视化图像不仅直观显示了叶片中多糖与黄酮的含量水平,也可形象地展示不同含量水平叶片的空间分布。

3 结论

本研究针对不同光环境下栽培的金线莲,利用叶片像素光谱数据,分别建立了传统机器学习模型(PCA-LDA 与 PCA-SVM)以及深度学习模型(1D CNN 与优化的 1D CNN),并验证了高光谱技术在判别光周期影响下金线莲中多糖与黄酮含量水平方面的可行性。结果表明,传统机器学习模型(PCA-LDA 与 PCA-SVM)的判别效果相似,而一维卷积神经网络(1D CNN)在提高判别准确率方面表现出显著优势,在训练集、验证集与独立验证集上对多糖与黄酮含量水平的判别准确率分别为 99.99% 与 99.89%、99.98% 与 99.78%、91.62% 与 87.92%。通过在 1D CNN 模型中引入 Dropout 层进行优化,进一步缓解了过拟合问题,多糖与黄酮含量水平的判别准确率在训练集、验证集与独立验证集中分别为 99.95% 与 99.93%、99.97% 与 99.94%、98.92% 与 95.67%。此外,研究还构建了多糖与黄酮含量水平判别的可视化图像,显示不同光周期下叶片的预测结果与实际含量水平分组相符。

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

参考文献

- [1] ZENG B, SU M, CHEN Q, et al. Antioxidant and hepatoprotective activities of polysaccharides from *Anoectochilus roxburghii* [J]. *Carbohydrate Polymers*, 2016, 153: 391–398.
- [2] CUI S C, YU J, ZHANG X H, et al. Antihyperglycemic and antioxidant activity of water extract from *Anoectochilus roxburghii* in experimental diabetes [J]. *Experimental and Toxicologic Pathology*, 2013, 65(5): 485–488.
- [3] WANG X X, HE J M, WANG C L, et al. Simultaneous structural identification of natural products in fractions of crude extract of the rare endangered plant *Anoectochilus roxburghii* using 1H NMR/RRLC-MS parallel dynamic spectroscopy [J]. *International Journal of Molecular Sciences*, 2011, 12: 2556–2571.
- [4] LV T W, TENG R D, SHAO Q S, et al. DNA barcodes for the identification of *Anoectochilus roxburghii* and its adulterants [J]. *Planta*, 2015, 242: 1167–1174.
- [5] 应震, 杨燕萍, 林恩义, 等. 林下仿野生栽培和设施栽培金线莲生物量及药用成分含量变化研究 [J]. *浙江林业科技*, 2024, 44(2): 95–99. [YING Z, YANG Y P, LIN E Y, et al. Variation of biomass and medicinal components in *Anoectochilus roxburghii* with different cultivation methods [J]. *Journal of Zhejiang Forestry Science and Technology*, 2024, 44(2): 95–99.]
- [6] 蔡文燕, 肖华山, 范秀珍. 金线莲研究进展 [J]. *亚热带植物科学*, 2003, 32(3): 68–72. [CAI W Y, XIAO H S, FAN X Z. A review of research on *Anoectochilus roxburghii* [J]. *Subtropical Plant Science*, 2003, 32(3): 68–72.]
- [7] WU Y, LIU C, JIANG Y, et al. Structural characterization and hepatoprotective effects of polysaccharides from *Anoectochilus zhejiangensis* [J]. *International Journal of Biological Macromolecules*, 2022, 198: 111–118.
- [8] SHEN Y Y, QING S S, MENG J X, et al. Effects of light quality on morphology, enzyme activities, and bioactive compound contents in *Anoectochilus roxburghii* [J]. *Frontiers in Plant Science*, 2017, 8: 857.
- [9] 郑涛, 刘宁, 孙红, 等. 基于高光谱成像的马铃薯叶片叶绿素

- 分布可视化研究[J]. *农业机械学报*, 2017, 32(S1): 153-159, 340.
- [ZHENG T, LIU N, SUN H, et al. Visualization of chlorophyll distribution of potato leaves based on hyperspectral imaging technology[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 32(S1): 153-159, 340.]
- [10] 赵艳茹, 余克强, 李晓丽, 等. 基于高光谱成像的南瓜叶片叶绿素分布可视化研究[J]. *光谱学与光谱分析*, 2014, 34(5): 1378-1382. [ZHAO Y R, YU K Q, LI X L, et al. Study on SPAD visualization of pumpkin leaves based on hyperspectral imaging technology[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(5): 1378-1382.]
- [11] WU L, GONG Y J, BAI X P, et al. Nondestructive determination of leaf nitrogen content in corn by hyperspectral imaging using spectral and texture fusion[J]. *Applied Sciences*, 2023, 13(3): 1910-1927.
- [12] YE W, YAN T, ZHANG C, et al. Detection of pesticide residue level in grape using hyperspectral imaging with machine learning[J]. *Foods*, 2022, 11(11): 1609-1615.
- [13] CHEN Y, XIE M, ZHANG H, et al. Quantification of total polysaccharides and triterpenoids in *Ganoderma lucidum* and *Ganoderma atrum* by near infrared spectroscopy and chemometrics[J]. *Food Chemistry*, 2012, 135(1): 268-275.
- [14] YANG L, GAO H, MENG L, et al. Nondestructive measurement of pectin polysaccharides using hyperspectral imaging in mulberry fruit[J]. *Food Chemistry*, 2021, 334: 127614.
- [15] 宋镇, 姬长英, 张波. 基于光谱与图像信息的杏鲍菇多糖含量检测[J]. *华南农业大学学报*, 2019, 40(3): 104-110. [SONG Z, JI C Y, ZHANG B. Detection of polysaccharide content in *Pleurotus eryngii* based on spectral and image information[J]. *Journal of South China Agricultural University*, 2019, 40(3): 104-110.]
- [16] 张郡赫, 于海业, 党敬民. 基于双光谱技术的高温紫外胁迫下小麦多糖反演模型研究[J]. *光谱学与光谱分析*, 2023, 43(9): 2705-2709. [ZHANG J H, YU H Y, DANG J M. Research on inversion model of wheat polysaccharide under high temperature and ultraviolet stress based on dual-spectral technique[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(9): 2705-2709.]
- [17] 邹小波, 陈武, 徐艺伟, 等. 不同生长阶段枇杷叶总黄酮含量的二维分布检测[J]. *天然产物研究与开发*, 2016, 28: 354-358. [ZOU X B, CHEN W, XU Y W, et al. Content distribution of total flavonoid in loquat (*Eriobotrya japonica*) leaves at different growth stages[J]. *Natural Product Research and Development*, 2016, 28: 354-358.]
- [18] 王树文, 修成, 董元, 等. 不同时段刺五加叶片黄酮含量的高光谱无损估测[J]. *农机化研究*, 2022, 44(4): 119-126, 268. [WANG S W, XIU C, DONG Y, et al. Non-destructive estimation of flavonoids in leaves of *Acanthopanax senticosus* in different periods[J]. *Journal of Agricultural Mechanization Research*, 2022, 44(4): 119-126, 268.]
- [19] HE J, CHEN L, CHU B, et al. Determination of total polysaccharides and total flavonoids in *Chrysanthemum morifolium* using near-infrared hyperspectral imaging and multivariate analysis[J]. *Molecules*, 2018, 23(9): 2395.
- [20] 余克强, 赵艳茹, 李晓丽, 等. 高光谱成像技术的不同叶位尖椒叶片氮素分布可视化研究[J]. *光谱学与光谱分析*, 2015, 35(3): 746-750. [YU K Q, ZHAO Y R, LI X L, et al. Application of hyperspectral imaging for visualization of nitrogen content in pepper leaf with different positions[J]. *Spectroscopy and Spectral Analysis*, 2015, 35(3): 746-750.]
- [21] 石吉勇, 邹小波, 张德涛, 等. 不同颜色银杏叶总黄酮含量分布高光谱图像检测[J]. *农业机械学报*, 2014, 45(11): 242-245, 33. [SHI J Y, ZOU X B, ZHANG D T, et al. Determination of total flavonoid content distribution on different color *Ginkgo biloba* leaves[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(11): 242-245, 33.]
- [22] 李芳, 石吉勇, 张德涛, 等. 银杏叶中黄酮含量的叶面分布检测研究[J]. *食品工业科技*, 2015, 36(9): 270-272. [LI F, SHI J Y, ZHANG D T, et al. Determination of total flavonoid content distribution on *Ginkgo biloba* leaf[J]. *Science and Technology of Food Industry*, 2015, 36(9): 270-272.]
- [23] 黄瑞平, 黄颖桢, 陈菁瑛, 等. 不同月龄金线莲多糖和总黄酮含量的比较[J]. *热带生物学报*, 2012, 3(2): 174-176. [HUANG R P, HUANG Y Z, CHEN J Y, et al. Comparison of polysaccharide and flavonoid contents in *Anoictochilus roxburghii* (Wall.) Lindl under different cultivation forms and harvest time[J]. *Journal of Tropical Biology*, 2012, 3(2): 174-176.]
- [24] de SANTANA F B, OTANI S K, de SOUZA A M, et al. Comparison of PLS and SVM models for soil organic matter and particle size using vis-NIR spectral libraries[J]. *Geoderma Regional*, 2021, 27: e00436.
- [25] AMSARAJ R, AMBADE N D, MUTTURI S. Variable selection coupled to PLS2, ANN and SVM for simultaneous detection of multiple adulterants in milk using spectral data[J]. *International Dairy Journal*, 2021, 123: 105172.
- [26] 李灵巧, 潘细朋, 冯艳春, 等. 深度卷积网络的多品种多厂商药品近红外光谱分类[J]. *光谱学与光谱分析*, 2019, 39(11): 3606-3613. [LI L Q, PAN X P, FENG Y C, et al. Deep convolution network application in identification of multi-variety and multi-manufacturer pharmaceutical[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(11): 3606-3613.]
- [27] GUO Y H, SUN L, ZHANG Z H, et al. Algorithm research on improving activation function of convolutional neural networks [C]//In Proceedings of the 31st Chinese Control and Decision Conference. Nan Chang, China: 2019: 3582-3586.
- [28] BARANSKI R, BARANSKA M, SCHULZ H. Changes in carotenoid content and distribution in living plant tissue can be observed and mapped in situ using NIR-FT-Raman spectroscopy[J]. *Planta*, 2005, 222: 448-457.
- [29] 邹红玉, 郑红平. 浅述植被“红边”效应及其定量分析方法[J]. *遥感信息*, 2010(4): 112-116. [ZOU H Y, ZHENG H P. The Effect and method of quantitative analysis of "Red Edge" of vegetation[J]. *Remote Sensing Information*, 2010(4): 112-116.]
- [30] 王震, 李映雪, 吴芳, 等. 冠层光谱红边参数结合随机森林机器学习估算冬小麦叶绿素相对含量[J]. *农业工程学报*, 2024, 40(4): 171-182. [WANG Z, LI Y X, WU F, et al. Estimation of winter wheat chlorophyll relative content combining with canopy spectrum red edge parameters and random forest machine learning[J]. *Transactions of the CSAE*, 2024, 40(4): 171-182.]
- [31] YU K Q, ZHAO Y R, ZHU F L, et al. Mapping of chlorophyll and SPAD distribution in pepper leaves during leaf senescence using visible and near-infrared hyperspectral imaging[J]. *Transactions of the ASABE*, 2016, 59(1): 13-24.
- [32] 蒋国瑞, 司学峰. 基于代价敏感 SVM 的电信客户流失预测研究[J]. *计算机应用研究*, 2009, 26(2): 521-523. [JIANG G R, SI X F. Study of telecom customer churn prediction based on cost sensitive SVM[J]. *Application Research of Computers*, 2009, 26(2): 521-523.]
- [33] XU Q, ZHANG M, GU Z, et al. Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs[J]. *Neurocomputing*, 2019, 382: 69-74.