

基于紫外-可见-近红外光谱的苹果叶片褐斑病早期识别与诊断方法

张宏建^{1,2} 刘若飞¹ 王永贤¹ 马博¹ 田园³ 王金星^{1,4}

(1. 山东农业大学机械与电子工程学院, 泰安 271018; 2. 农业装备智能化山东省工程研究中心, 泰安 271018; 3. 山东省鱼台县清河镇农业综合服务中心, 济宁 272348; 4. 山东省设施园艺智慧生产技术装备重点实验室(筹), 泰安 271018)

摘要: 褐斑病是危害苹果叶片的真菌病害之一,严重影响苹果树的健康生长与果实品质。本文提出了一种结合紫外-可见-近红外光谱技术和机器学习算法的苹果叶片褐斑病早期识别与诊断方法,展现出较好的识别诊断效果。首先,利用日立 UH4150 型分光光度计(波长 200 ~ 3 050 nm)采集了健康、1 级病害和 3 级病害的苹果叶片高光谱数据。然后,采用标准正态变换(Standard normal variate, SNV)、多项式平滑滤波(Savitzky - Golay, SG)、均值标准化(Z-score standardization, ZS)、多元散射校正(Multiplicative scatter correction, MSC)和一阶导数(First derivative, FD)方法对原始光谱数据进行了预处理。接着,采用主成分分析(Principal component analysis, PCA)和递归特征消除(Recursive feature elimination, RFE)算法提取了与褐斑病叶片相关的特征波长。最后,建立了基于随机森林(Random forest, RF)算法与卷积神经网络(Convolutional neural network, CNN)的判别模型,评估了使用特征波长进行疾病早期诊断的能力。研究结果表明,SG - RFE - RF 模型表现出优异的诊断性能,其在建模集和预测集上的分类准确率分别为 96.03% 和 94.34%。该研究为开发苹果叶片褐斑病早期检测传感器提供了技术支持。

关键词: 苹果叶片; 褐斑病; 紫外-可见-近红外光谱; 机器学习; 早期诊断

中图分类号: TP391 文献标识码: A 文章编号: 1000-1298(2026)15-0287-09

OSID:



Early Identification and Diagnosis of Apple Leaf Brown Spot Disease Based on Ultraviolet - Visible - Near-infrared Spectroscopy

Zhang Hongjian^{1,2} Liu Ruofei¹ Wang Yongxian¹ Ma Bo¹ Tian Yuan³ Wang Jinxing^{1,4}

(1. College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian 271018, China

2. Shandong Engineering Research Center of Agricultural Equipment Intelligentization, Taian 271018, China

3. Shandong Province Yutai County Qinghe Town Agricultural Comprehensive Service Center, Jining 272348, China

4. Shandong Key Laboratory of Intelligent Production Technology and Equipment for Facility Horticulture, Taian 271018, China)

Abstract: Brown spot disease is one of the fungal diseases affecting apple leaves, severely impacting the healthy growth of apple trees and the quality of their fruit. An early detection and diagnosis method for brown spot disease in apple leaves was presented by integrating ultraviolet - visible - near-infrared spectroscopy with machine learning algorithms, demonstrating promising diagnostic performance. Firstly, hyperspectral data of healthy, grade 1, and grade 3 diseased apple leaves were collected by using a Hitachi UH4150 spectrophotometer (wavelength 200 ~ 3 050 nm). Then, the raw spectral data were preprocessed using standard normal variate (SNV), Savitzky - Golay (SG), Z-score standardization (ZS), multiplicative scatter correction (MSC), and first derivative (FD). Subsequently, principal component analysis (PCA) and recursive feature elimination (RFE) algorithms were employed to extract feature wavelengths related to brown spot in the apple leaves. Finally, discriminant models based on the random forest (RF) algorithm and convolutional neural network (CNN) were established to evaluate the capability of using feature wavelengths for early disease diagnosis. The results showed that the SG - RFE -

收稿日期: 2025 - 03 - 06 修回日期: 2025 - 04 - 11

基金项目: 鲁渝科技协作联合攻关项目、财政部和农业农村部:国家现代农业(苹果)产业技术体系项目(CARS - 27)、山东省青年科技人才托举工程项目(SDAST2024QTA050)和山东省高等学校“青创团队计划”项目(2023KJ160)

作者简介: 张宏建(1991—),男,副教授,博士生导师,主要从事智能农机装备研究,E-mail: zhanghongji_an@163.com

通信作者: 王金星(1970—),男,教授,博士生导师,主要从事智能农机装备研究,E-mail: jinxingw@163.com

RF model achieved classification accuracies of 96.03% on the modeling set and 94.34% on the prediction set, demonstrating the best performance. The proposed method showed strong robustness and generalization ability, indicating its potential for practical applications in orchard disease monitoring. Compared with traditional inspection methods, this approach can provide a rapid, non-destructive, and objective solution for early disease detection. The research result can offer valuable technical support for the development of intelligent detection systems and precision agriculture practices for apple production.

Key words: apple leaves; brown spot disease; ultraviolet – visible – near-infrared spectroscopy; machine learning; early diagnosis

0 引言

苹果褐斑病是苹果叶部常见的真菌性病害,主要由非专性生物营养病原体双壳菌引起。该菌能在植物上寄生,导致叶片早期枯萎并脱落,从而影响光合作用,显著降低苹果果实的质量和产量^[1-3]。当叶片早期遭受双壳菌感染后,植物的新陈代谢和组织结构会发生变化。这为使用酶联免疫吸附测定(ELISA)和聚合酶链反应(PCR)等分子技术进行诊断提供了可能性^[4-5]。然而,尽管这些技术具有较高的准确性,但这些方法存在一定局限性。ELISA方法依赖特异性抗体,检测结果受抗原抗体反应影响较大,同时实验流程复杂,对温度、湿度等环境因素要求严格。PCR检测则需要进行DNA/RNA提取和扩增,前处理过程繁琐,并依赖高精度仪器及严格的实验条件。此外,这2种方法均为破坏性检测,需要破坏样本组织进行提取,不适用于活体检测,且检测成本较高,难以满足大规模田间实时监测的需求。机器视觉技术作为一种快速且非破坏性的检测方法,能够高效捕捉病变图像,尤其在病害后期表现突出。然而,在早期感染阶段,由于植物的生理变化尚不明显或病斑较小,机器视觉技术难以准确识别微小的病变,从而影响其早期检测的准确性^[6-8]。因此,开发一种适用于苹果叶片褐斑病早期识别与诊断方法,对苹果病害的早期防治及保障果实健康生长具有重要意义。

近年来,光谱技术凭借实时、无损和快速的技术优势,在果树叶片病害的识别诊断方面取得了显著进展^[9-11]。尽管已有研究^[12-18]表明光谱技术在作物叶片病害监测方面取得了显著的研究成果,然而,利用紫外-可见-近红外光谱技术开展苹果叶片褐斑病早期识别和诊断的研究鲜有报道。

本文以红富士苹果叶片为研究对象,采集健康和早期感染褐斑病叶片的高光谱数据,采用主成分分析(Principal component analysis, PCA)和递归特征消除(Recursive feature elimination, RFE)算法提取与褐斑病叶片相关的特征波长,并结合随机森林(Random forest, RF)和卷积神经网络

(Convolutional neural network, CNN)算法构建褐斑病叶片分类模型,以实现苹果叶片褐斑病的早期识别和诊断。

1 材料与方法

1.1 苹果叶片样本采集与分级

实验样本来源于山东省级农业高科技园区,苹果品种为红富士。样本采集时间为2023年10—11月,共计采集10次,采样间隔为3 d,覆盖不同阶段自然感染病害的叶片。采样时间依据田间病害的自然发生与发展情况选择,涵盖了病害初发期及扩展期,以确保数据能反映叶片光谱特征在病害进程中的变化。在经验丰富的农业专家指导下,随机选择不同感染程度的苹果叶片进行采集。所有叶片在采集后均用泡沫网套进行包裹,并存放于恒温恒湿保鲜盒中,随后运输至山东农业大学机械与电子工程学院智能农业与果蔬加工实验室进行进一步分析。按照病斑面积占比对所收集的叶片进行了挑选和分级。具体分级如下:健康(无病斑)叶片60片,1级病害(病斑面积占整片叶面积10%以下)叶片58片,3级病害(病斑面积占整片叶面积11%~25%)叶片61片,叶片采集与分级结果如图1所示。

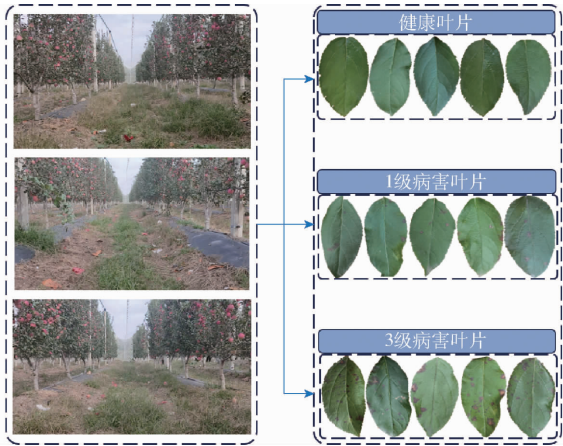


图1 叶片采集与分级

Fig.1 Leaf collection and grading

1.2 光谱信息采集

图2展示了本实验所使用的日立UH4150型分光光度计内部运行过程。该光度计由钨灯、反射镜、

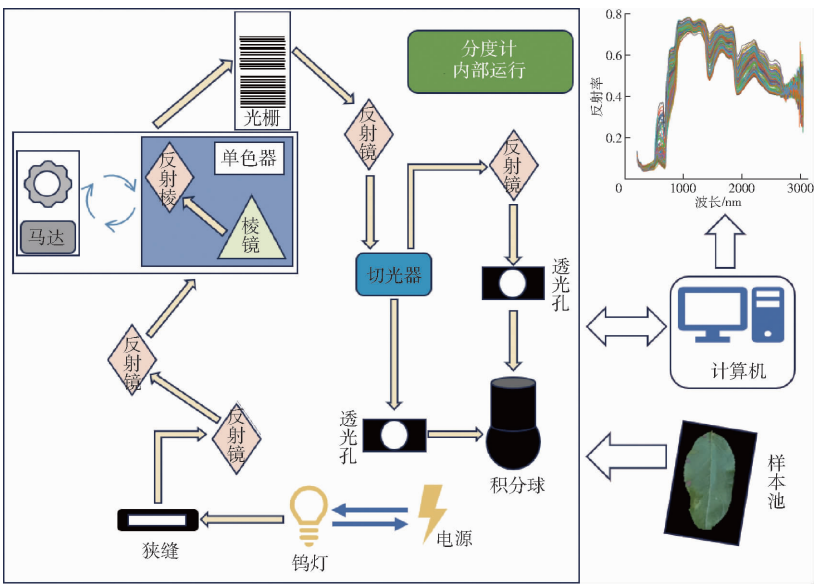


图 2 叶片光谱信息采集流程

Fig. 2 Leaf spectral information acquisition

积分球、单色器、切光器及电源等组件构成。在叶片光谱信息采集过程中，钨灯发出的光束通过反射镜和单色器传递至切光器和积分球。当光束照射到积分球内的样本池时，经过反射后，光束被内部检测器收集，从而完成叶片光谱信息的采集。分光光度计的主要参数设置如下：测量灯自动切换，反射率为 R ，栅缝宽度为 4 nm ，扫描速度为 $2\,400\text{ nm/min}$ ，采样间隔为 5 nm ，波长范围为 $200\sim3\,050\text{ nm}$ 。

1.3 黑白校正

为了消除仪器本底噪声和系统偏差，并提高测量结果的准确性与可靠性，对获取的叶片原始数据进行校正，这一过程通过黑白校正 (White-black correction) 来实现^[19]。黑白校正过程包括以下步骤：首先，测量无光条件下的暗电流信号，以消除仪器的本底噪声；其次，进行白板校正，使用仪器自带的高反射率标准白板进行参考光谱的测量，以校正光源和探测器的系统响应；此外，为了减少仪器漂移对测量结果的影响，每次测量前均重新采集暗电流和白板光谱，并在实验过程中进行 5 次重复测量取平均值，以提高数据的稳定性和可靠性；最后，计算叶片的光谱数据，即

$$I(\lambda) = \frac{S(\lambda) - D(\lambda)}{R(\lambda) - D(\lambda)} \tag{1}$$

式中 $I(\lambda)$ ——校正后的归一化光谱信号
 $S(\lambda)$ ——样本原始光谱信号
 $D(\lambda)$ ——暗电流信号
 $R(\lambda)$ ——白板参考光谱信号

1.4 光谱数据处理

1.4.1 光谱数据预处理

光谱信息的采集容易受到环境因素的干扰，常

见问题包括噪声、幅度差异和基线漂移等，这些因素会影响光谱数据的准确性。因此，数据预处理在光谱分析中显得尤为重要，合理去除这些干扰因素能够显著提高模型的识别精度^[20]。本文采用了 5 种常见的光谱数据预处理方法：标准正态变换 (Standard normal variate, SNV)、多项式平滑滤波 (Savitzky - Golay, SG)、均值标准化 (Z-score standardization, ZS)、多元散射校正 (Multiplicative scatter correction, MSC) 和一阶导数 (First derivative, FD)。

SNV 方法用于消除光谱数据中由散射效应等不确定因素引起的干扰^[21]。SG 方法是基于时域内局部多项式最小二乘法拟合，广泛用于数据平滑和去噪，为平衡去噪能力与信号保真度，本研究通过预实验对比不同平滑窗口 (5、7、9、11、13) 和多项式阶数 (2、3) 的组合，并通过信噪比、均方误差、光谱平滑度及光谱特征保持度等指标，优选去噪效果与光谱平滑度较好的参数组合^[22]。ZS 方法首先对数据进行归一化处理，再进行基线校正，旨在使不同浓度的光谱数据具有可比性，同时消除荧光背景、周围环境及黑体辐射等因素对光谱基线的影响^[23]。MSC 方法通过对光谱数据进行尺度变换，消除样品粒径、密度等因素导致的散射效应，同时校正基线漂移，减少环境噪声影响，提高光谱稳定性和建模精度^[24]。FD 方法通过计算光谱数据的斜率变化，去除基线漂移并增强光谱特征，使得光谱中的微小变化更加明显，提高重叠峰的分辨率和消除背景干扰^[25]。

1.4.2 特征波长提取

鉴于本文所采集的苹果叶片光谱数据包含 571 个波长，存在大量冗余信息，若直接使用所有波长进

行实验,模型的计算量将大幅增加,从而影响计算效率和准确性。因此,必须对光谱数据进行降维处理,以提取关键的光谱特征波长,从而提高分析的效率和精度。为此,采用了 PCA 和 RFE 方法来提取有效的特征波长。

PCA 是一种常用于数据降维和特征提取的统计方法^[26]。其通过将原始数据中的多个相关变量转换为一组互不相关的主成分,达到简化数据结构、去除冗余信息并突出主要特征的目的。其核心原理是对数据的协方差矩阵进行特征值分解,选择方差最大的方向作为主成分,从而保留数据中的主要变异性,并有效降低维度。

RFE 是一种用于特征选择的算法^[27]。该算法的核心思想是通过多轮迭代的方式,在每一轮中评估特征对模型的贡献度,避免冗余信息干扰,根据评估结果逐步删除最不重要的特征,直到剩余的特征数量满足预设的要求或达到优化目标。RFE 能有效降维,提升模型性能与计算效率,并减少过拟合风险。

1.5 分类模型构建

在特征波长提取完成后,选择合适的分类模型对提取的特征进行有效分类是研究中的关键步骤。考虑到高维光谱数据的复杂性,本研究采用了 RF 和 CNN 2 种分类模型。

RF 是一种集成学习算法,常用于分类任务^[28]。其通过构建多个决策树并结合投票结果实现高准确性和鲁棒性。在训练中,随机森林通过有放回地抽取子集构建不同的树,为增强模型多样性,每棵树在节点分裂时仅从全部特征中随机选择一个子集进行划分。这种特征随机化策略不仅降低了单棵树的方差,还能避免过拟合,提高泛化能力。此外,RF 具备较强的抗噪性和鲁棒性,能够适应高维数据、非线性关系及部分缺失数据,减少异常值或数据偏差对分类性能的影响。

CNN 在光谱信息分类中具有显著优势^[29]。它的核心原理是通过卷积层提取图像或数据中的局部特征,并通过池化层降低数据维度,从而减少计算量并增强模型的鲁棒性。卷积神经网络通过多个卷积层、池化层和全连接层的组合,能够逐层提取不同层次的特征。在本研究中,选用 ResNet-50 作为主干网络,其深度残差结构能够有效缓解深层网络的梯度消失问题,并提高模型的训练稳定性。相比传统 CNN 结构,ResNet-50 能够在较深的网络中保持有效特征的学习能力,从而提升光谱分类任务的准确性。

分类模型建立后,将建模集的准确率作为分类模型性能的辅助评价指标,将预测集的准确率作为模型性能的主要评价指标,并通过混淆矩阵来评估

分类结果中不同类别的预测与真实分布。

2 结果与分析

2.1 苹果叶片光谱分析

由于光谱曲线的边缘区域噪声较大,故去除受噪声影响较大的首部和尾部两部分波段,保留波长 400 ~ 2 700 nm 内的 461 个波段作为有效光谱波段进行光谱特性分析。如图 3a 所示,不同苹果叶片的光谱曲线变化趋势基本一致。但由于植被反射光谱携带了大量植被内部理化参数的信息,而病菌的感染会对植被理化参数造成一系列影响,因此健康植被和感病植被在光谱表现上会呈现出一定差异。如图 3b 所示,健康和褐斑病感染的苹果叶片在全波段光谱范围(波长 400 ~ 2 700 nm)内表现出显著差异。具体来说,在可见光波段(波长 400 ~ 680 nm)由于褐斑病感染造成色素含量的下降,使得感病叶片反射率升高,在介于可见光和近红外区间之间的红边波段(波长 680 ~ 760 nm),由于对叶内叶绿素含量及叶片结构的高敏感性和对噪声的不敏感,使得反射率呈现出急剧上升的形态,在近红外波段(波长 760 ~ 1 300 nm),由于褐斑病感染造成细胞组织结构的破坏,使得感病叶片反射率降低,在短波红外波段(波长 1 300 ~ 2 700 nm)由于褐斑病感染造成叶片水分含量的降低,使得感病叶片反射率升高。

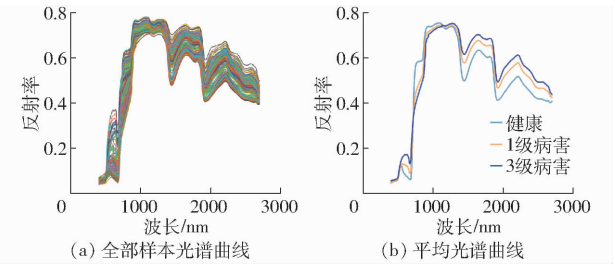


图 3 苹果叶片光谱曲线

Fig. 3 Apple leaf spectral curves

2.2 样本划分

为了构建苹果叶片褐斑病早期分类模型,提取 179 组光谱数据,并采用分层抽样的方法对样本数据集按照比例 7:3 进行了建模集和预测集的划分,划分结果如表 1 所示^[30]。

表 1 样本划分

Tab. 1 Sample classification

分组	健康叶片数量	1 级病害叶片数量	3 级病害叶片数量
建模集	42	41	43
预测集	18	17	18

2.3 特征波长提取

2.3.1 基于 PCA 的特征波长提取

采用 PCA 对原始光谱数据及 5 种不同预处理

方法处理后的数据进行降维分析。通过将光谱样本投影到前 3 个主成分 (PC1、PC2、PC3) 空间中, 生成三维散点图, 直观展示样本在主成分空间中的分布特征, 如图 4 所示。分析散点图可以得出前 3 个主成分的单贡献率及其累计贡献率, 进而评估主成分对数据总方差的解释能力。图 4 显示, PC1 的贡献率最高, 均在 60% 以上, 且经过不同预处理方法后, PC1、PC2 和 PC3 的累计贡献率均超过 83%, 能够有效代表光谱数据的信息。在不同预处理方法中, SG 预处理后的累计贡献率最高, 达到 97%, 其中 PC1 为 81%, PC2 为 12%, PC3 为 4%; 而 FD 预处理后的累计贡献率最低, 为 83%, 其中 PC1 为 60%, PC2 为 15%, PC3 为 8%。这些结果表明, 前 3 个主成分解释了原始数据的绝大部分信息。同时, 散点图揭示了 3 类叶片呈现一定的聚类趋势, 尽管原始数据经过 ZS、MSC 和 FD 预处理的聚类效果较差, 但经过 SG 和 SNV 预处理后, 健康叶片、1 级病害叶片和 3 级病害叶片之间的分类边界更加清晰。然而, 3 类叶片样本仍存在一定的重叠现象, 说明需要进一步处理 PCA 后的光谱数据, 以实现更精确的分类。

主成分载荷是描述主成分与原始光谱波长特征关系的向量, 反映了每个波长特征对主成分的贡献程度, 载荷绝对值反映了主成分与各波长的相关性, 并表征了它们之间的关联强度, 主成分曲线的波峰或波谷位置对应最大权重系数, 这些位置对应的波长即为提取的特征波长, 为不遗漏原始光谱的关键

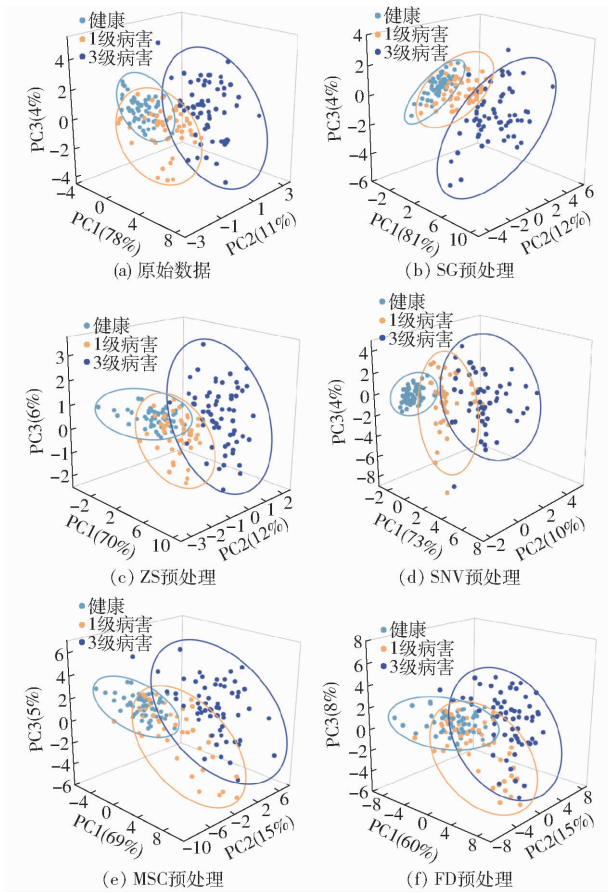


图 4 样本分布散点图

Fig. 4 Scatter plots of sample distribution

信息, 提取前 4 个主成分进行后续研究。载荷曲线如图 5 所示。本文选取载荷绝对值大于 0.05 的波峰以及波谷作为特征波长。基于该方法具体提取的特征波长信息详见表 2。

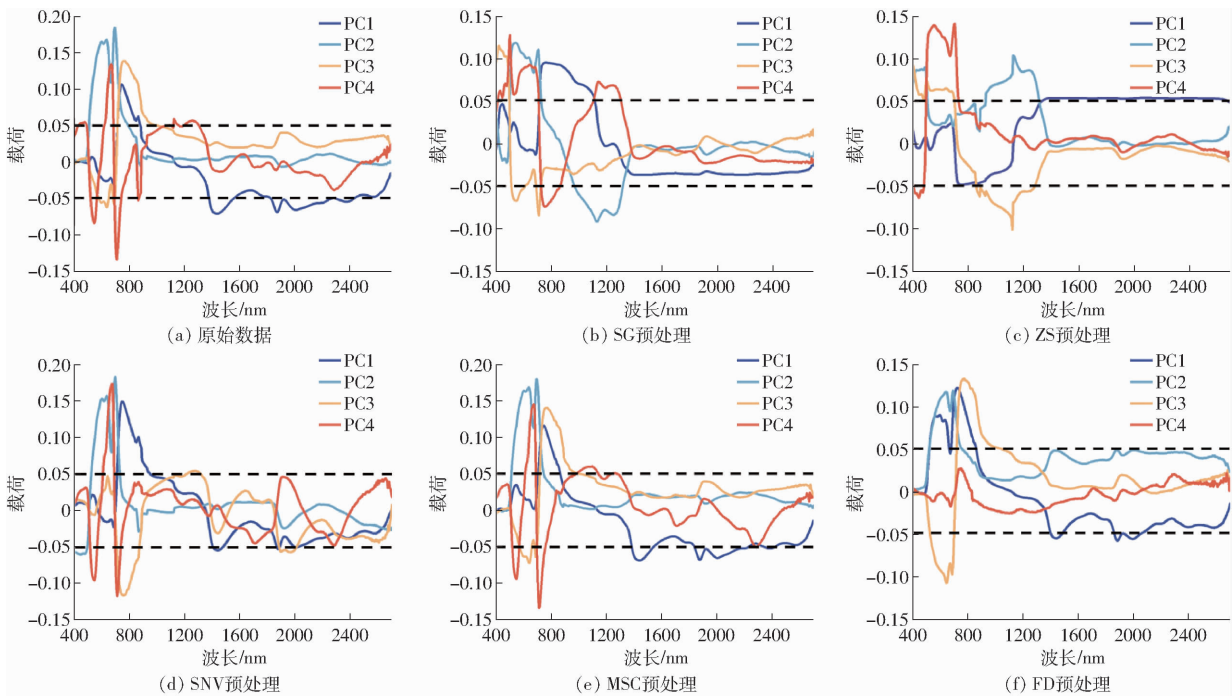


图 5 主成分载荷图

Fig. 5 Principal component loadings plot

表 2 基于 PCA 和 RFE 优选的特征波长

Tab.2 Feature wavelengths optimized based on PCA and RFE

特征选择方法	预处理方法	提取个数	特征波长/nm
PCA	未预处理	14	480,545,635,675,695,710,750,765,865,1 125,1 440,1 875,1 920,2 015
	SG 预处理	9	420,500,545,705,710,750,1 135,1 180,1 250
	ZS 预处理	12	445,470,480,500,560,670,705,850,885,925,1 120,1 135
	SNV 预处理	14	540,600,615,640,675,695,710,750,760,865,1 285,1 440,1 915,1 955
	MSC 预处理	13	405,545,635,675,695,710,750,765,1095,1265,1 440, 1 875,2 015
	FD 预处理	13	555,590,640,670,685,690,720,770,1 435,1 885,1 925, 1 995,2 005
RFE	未预处理	10	405,410,425,1 095,1 925,1 940,1 950,2 345,2 375,2 675
	SG 预处理	10	405,410,415,475,1 465,1 945,2 175,2 345,2 365,2 570
	ZS 预处理	10	405,410,425,455,1 700,1 945,2 490,2 510,2 520,2 670
	SNV 预处理	10	625,630,635,640,645,650,690,695,720,870
	MSC 预处理	10	405,425,455,835,1 345,1 885,2 450,2 485,2 510,2 685
	FD 预处理	10	405,410,430,815,2 110,2 205,2 375,2 480,2 500,2 510

2.3.2 基于 RFE 的特征波长提取

在特征波长提取过程中,选择前 10 个最重要的特征波长作为最终保留的特征。使用支持向量回归 (SVR) 作为基础模型进行训练,并通过贝叶斯优化与 K 折交叉验证相结合的方式,优化 SVR 模型的惩罚参数 c 和核函数 γ 。经过多轮实验,最终确定了不同预处理条件下的最优参数配置:原始光谱数据下 c 为 8, γ 为 0.24;经 SG 预处理数据下 c 为 16, γ 为 0.856;经 ZS 预处理数据下 c 为 8, γ 为 0.526;经 SNV 预处理数据下 c 为 2, γ 为

0.036;经 MSC 预处理数据下 c 为 4, γ 为 0.384;经 FD 预处理数据下 c 为 4, γ 为 0.262。图 6 展示了不同预处理方法下提取的特征波长,图中红色标记表示通过 RFE 方法提取出的特征波长所在的位置。由图 6 可知,原始数据、SG 预处理、ZS 预处理、MSC 预处理和 FD 预处理下提取的特征波长主要集中在波长 400 ~ 500 nm 以及波长 1 900 ~ 2 600 nm 的区间,而经 SNV 预处理下提取的特征波长则主要集中在 600 ~ 900 nm。具体提取的特征波长信息详见表 2。

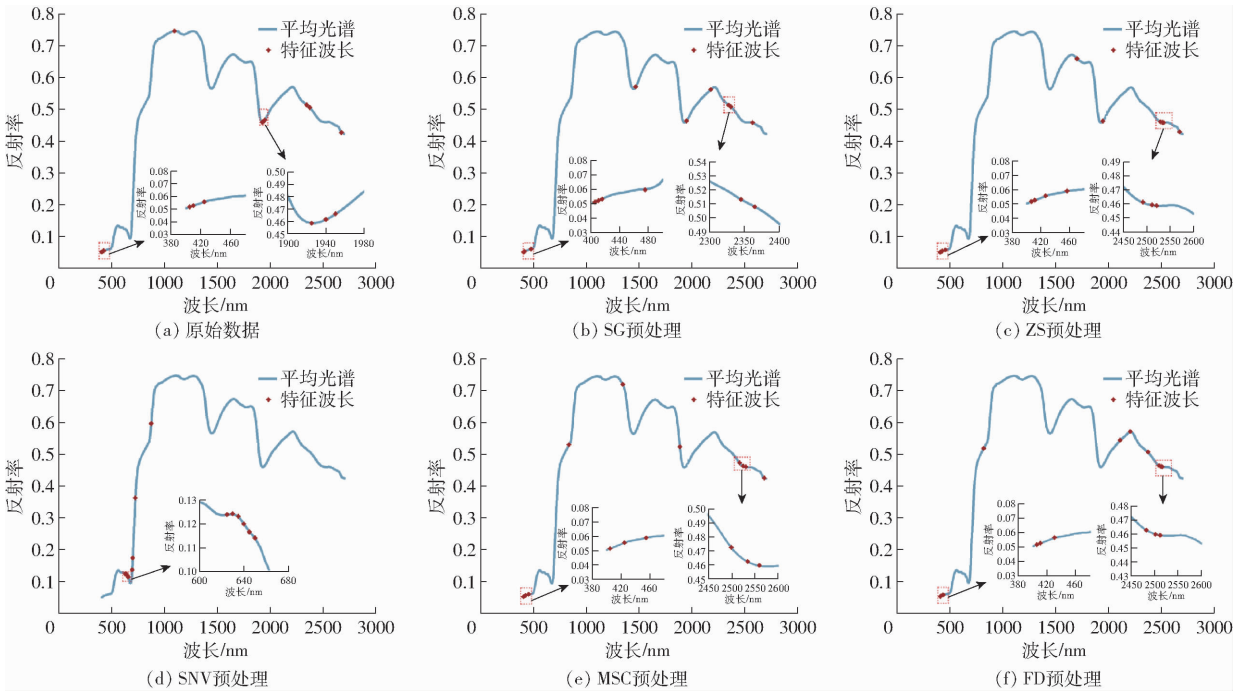


图 6 提取的特征波长

Fig.6 Feature wavelength extraction

2.3.3 基于 PCA 和 RFE 提取的特征波长

表 2 展示了基于 PCA 和 RFE 提取的与叶片褐斑病相关的所有特征波长,对表中数据观察发现,多

数波长集中在 400 ~ 500 nm、1 100 ~ 1 500 nm、1 800 ~ 2 000 nm 和 2 300 ~ 2 500 nm 范围内。其中,400 ~ 500 nm 属于可见光波段,其反射率变化与叶绿素 a/b

含量密切相关,当病原体侵染叶片后会导致细胞叶绿体损伤,显著降低该波段吸收使反射率升高^[31]。1 100~1 500 nm、1 800~2 000 nm 和 2 300~2 500 nm 波段则响应病斑扩展过程中的细胞失水现象,其反射率上升与纤维损失和自由水减少直接相关^[32-33]。

2.4 苹果叶片褐斑病分类模型构建

2.4.1 基于 PCA 提取的特征波长建模

对经过 PCA 提取的特征波长进行 RF 和 CNN 建模,在 RF 模型构建中,为优化模型性能,采用交叉验证对超参数进行优化,最终选择 300 棵决策树,并基于验证结果确定最大树深度为 10,以在泛化能力和计算复杂度之间取得最佳平衡。同时,采用预剪枝策略,将最小样本分裂数设为 5,以减少过深分裂导致的过拟合风险,最小样本节点数设为 5,以进一步简化模型结构,提高模型的稳健性。在 CNN 模型构建中,采用 ResNet-50 作为主干网络,并使用 Adam 优化器进行训练。最大训练次数为 500 次,初始学习率为 0.001,并引入 L2 正则化以抑制过拟合,正则化参数设置为 0.000 1。为提升训练稳定性和收敛效果,学习率采用分段式调度策略,每经过 200 个训练周期,学习率衰减至原来的 0.1 倍,从而在训练后期实现更精细的权重更新,提高模型的泛化能力。建模结果如表 3 所示。从表 3 可以看出,所建模型在建模集上的准确率均超过 90%,在预测集上的准确率均超过 84%。通过对比不同模型的训练结果,RF 模型的建模效果整体优于 CNN 模型。具体来说,RF 模型在建模集上的准确率均超过 90%,预测集的准确率均超过 86%,其中 SG 预处理后的建模效果最佳,建模集和预测集的准确率分别为 93.65% 和 92.45%。而 CNN 模型在建模集上的

准确率均超过 90%,预测集的准确率均超过 84%,其中 SNV 预处理后的建模效果最佳,建模集和预测集的准确率分别为 92.86% 和 90.57%。

2.4.2 基于 RFE 提取的特征波长建模

对经过 RFE 提取的特征波长进行 RF 和 CNN 建模,采用与 PCA 提取的特征波长建模相同的参数。建模结果如表 4 所示。由表 4 可知,所建模型在建模集上的准确率均超过 80%,在预测集上的准确率均超过 77%。对比不同模型的表现,RF 模型的建模效果整体优于 CNN 模型。在 RF 建模中,建模集准确率均超过 82%,预测集准确率均超过 79%,其中 SG 预处理后的效果最佳,建模集和预测集准确率分别为 96.03% 和 94.34%。在 CNN 建模中,建模集准确率均超过 80%,预测集准确率均超过 77%,ZS 预处理后的效果最佳,建模集和预测集准确率分别为 94.44% 和 92.45%。此外,经过 SNV 预处理后的建模准确率显著低于原始数据以及 SG、ZS、MSC 和 FD 预处理后的结果。分析原因,特征波长的提取主要集中在波长 600~900 nm 的区间,该波段较窄,信息较为有限,未能有效涵盖总体波段特征,因此导致建模准确率低于其他预处理方法。

表 4 样本经 RFE 提取特征后建模结果
Tab.4 Modeling results after feature extraction using RFE

模型	预处理方法	准确率/%	
		建模集	预测集
RF	原始数据	91.27	88.68
	SG	96.03	94.34
	ZS	92.86	90.57
	SNV	82.54	79.25
	MSC	92.06	90.57
	FD	91.27	88.68
CNN	原始数据	90.48	86.79
	SG	93.65	90.57
	ZS	94.44	92.45
	SNV	80.95	77.36
	MSC	91.27	88.68
	FD	90.48	86.79

综上所述,基于 PCA 和 RFE 提取的特征波长结合 RF 和 CNN 算法所构建的分类模型均取得了较好的分类效果,整体的分类准确率高 于 89%。在所构建的所有模型中,SG-RFE-RF 模型取得了最好的预测结果,模型建模集和预测集的分类准确率分别为 96.03% 和 94.34%,该结果表明了所选的特征波长结合机器学习算法具备早期识别和诊断叶片褐斑病的能力。图 7 展示了 SG-RFE-RF 模型的分 类结果混淆矩阵,在建模集的混淆矩阵中,3 片 1

表 3 样本经 PCA 提取特征后模型建模结果
Tab.3 Modeling results after feature extraction using PCA

模型	预处理方法	准确率/%	
		建模集	预测集
RF	原始数据	91.27	86.79
	SG	93.65	92.45
	ZS	91.27	88.68
	SNV	92.06	90.57
	MSC	91.27	90.57
	FD	90.48	86.79
CNN	原始数据	90.48	84.91
	SG	92.86	88.68
	ZS	91.27	86.79
	SNV	92.86	90.57
	MSC	91.27	88.68
	FD	90.48	86.79

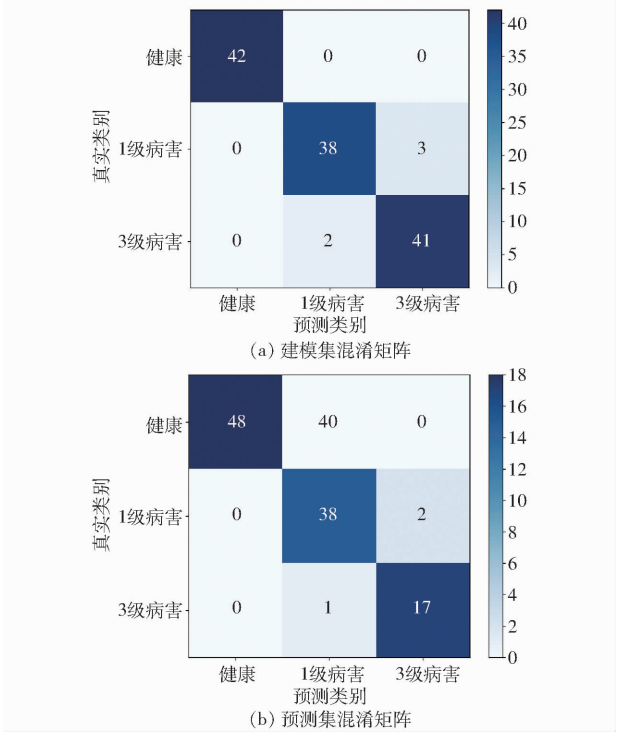


图 7 混淆矩阵

Fig. 7 Confusion matrix

级病害的叶片被错分成了 3 级病害,2 片 3 级病害的叶片被错分成 1 级病害,在预测集的混淆矩阵中,2 片 1 级病害的叶片被错分成了 3 级病害,1 片 3 级

病害的叶片被错分成了 1 级病害,这可能是由于 1 级病害和 3 级病害叶片之间的光谱特征差异不明显,病斑形态相似等原因造成的,但分类结果已达到了苹果叶片褐斑病早期识别和诊断的目的。

3 结论

- (1) 基于紫外-可见-近红外光谱技术对苹果叶片褐斑病的早期识别与诊断进行探索。
- (2) 通过使用日立 UH4150 型分光光度计采集波长 200 ~ 3 050 nm 范围内的光谱数据,并对原始光谱数据进行预处理,比较了 5 种预处理方法 (SG、ZS、SNV、MSC、FD)。实验结果表明,基于 SG 的预处理方法在提高识别精度方面效果最佳。
- (3) 在预处理的基础上,采用 PCA 和 RFE 方法对全波段数据进行了特征降维。降维后,2 种方法均有效剔除与病害信息相关性较低的波段,显著降低了模型的计算复杂度,同时确保了较高的识别精度。
- (4) 使用 RF 和 CNN 2 种分类模型构建了病害识别模型,并对预处理方法、特征提取方法和分类模型的不同组合进行了比较。在所有组合中,SG - RFE - RF 模型表现最佳,建模集和预测集准确率分别为 96. 03% 和 94. 34% 。

参 考 文 献

[1] 董向丽,高月娥,李保华,等. 苹果褐斑病在山东半岛中部的周年流行动态[J]. 中国农业科学, 2015, 48(3): 479 - 487.
Dong Xiangli, Gao Yuee, Li Baohua, et al. Epidemic dynamics of apple marssonina leaf blotch over whole growth season in the central area of Shandong peninsula[J]. Scientia Agricultura Sinica, 2015, 48(3): 479 - 487. (in Chinese)

[2] 霍学喜,刘天军,刘军弟,等. 2020 年度中国苹果产业发展报告(精简版)[J]. 中国果菜, 2022, 42(2): 1 - 6.
Huo Xuexi, Liu Tianjun, Liu Jundi, et al. China apple industry development report in 2020[J]. China Fruit and Vegetable, 2022, 42(2): 1 - 6. (in Chinese)

[3] 王树桐,王亚南,曹克强. 近年我国重要苹果病害发生概况及研究进展[J]. 植物保护, 2018, 44(5): 13 - 25, 50.
Wang Shutong, Wang Ya'nan, Cao Keqiang. Occurrence of and research progress in important apple diseases in China in recent years[J]. Plant Protection, 2018, 44(5): 13 - 25, 50. (in Chinese)

[4] Onozuka N, Ohki T, Oka N, et al. One-step real-time multiplex reverse transcription-polymerase chain reaction assay with melt curve analysis for detection of potato leafroll virus, potato virus S, potato virus X, and potato virus Y[J]. Virology Journal, 2021, 18: 1 - 10.

[5] Mbewe W, Mtonga A, Chiipanthenga M, et al. Incidence and distribution of sweetpotato viruses and their implication on sweetpotato seed system in Malawi[J]. Journal of Plant Pathology, 2021, 103(3): 961 - 968.

[6] 杨宁,程巍,张钊源,等. 基于 FPGA 加速的 Mask R - CNN 稻瘟病高通量自适应识别模型研究[J]. 农业机械学报, 2024, 55(7): 298 - 304, 314.
Yang Ning, Cheng Wei, Zhang Zhaoyuan, et al. Research on high-througput adaptive recognition Mask R - CNN model for rice blast disease based on FPGA acceleration[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 298 - 304, 314. (in Chinese)

[7] 袁杰,谢霖伟,郭旭,等. 基于改进 YOLO v7 的苹果叶片病害检测方法[J]. 农业机械学报, 2024, 55(11): 68 - 74.
Yuan Jie, Xie Linwei, Guo Xu, et al. Apple leaf disease detection method based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 68 - 74. (in Chinese)

[8] 李云红,张蕾涛,李丽敏,等. 基于 CycleGAN - IA 方法和 M - ConvNext 网络的苹果叶片病害图像识别[J]. 农业机械学报, 2024, 55(4): 204 - 212.
Li Yunhong, Zhang Leitao, Li Limin, et al. Image recognition of apple leaf disease based on CycleGAN - IA method and M - ConvNext network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 204 - 212. (in Chinese)

[9] Terentev A, Dolzhenko V, Fedotov A, et al. Current state of hyperspectral remote sensing for early plant disease detection: a

review[J]. *Sensors*, 2022, 22(3): 757.

[10] Appeltans S, Pieters J G, Mouazen A M. Detection of leek white tip disease under field conditions using hyperspectral proximal sensing and supervised machine learning[J]. *Computers and Electronics in Agriculture*, 2021, 190: 106453.

[11] Knauer U, Warnemunde S, Menz P, et al. Detection of apple proliferation disease using hyperspectral imaging and machine learning techniques[J]. *Sensors*, 2024, 24(23): 7774.

[12] 吴叶兰,陈怡宇,廉小亲,等. 高光谱成像的柑橘病虫害叶片识别方法[J]. *光谱学与光谱分析*, 2021, 41(12): 3837 – 3843.
Wu Yelan, Chen Yiyu, Lian Xiaoqin, et al. Study on the identification method of citrus leaves based on hyperspectral imaging technique[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(12): 3837 – 3843. (in Chinese)

[13] 秦立峰,张熹,张晓茜. 基于高光谱病害特征提取的温室黄瓜霜霉病早期检测[J]. *农业机械学报*, 2020, 51(11): 212 – 220.
Qin Lifeng, Zhang Xi, Zhang Xiaoqian. Early detection of cucumber downy mildew in greenhouse by hyperspectral disease differential feature extraction[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(11): 212 – 220. (in Chinese)

[14] 刘莉,陶红燕,方静,等. 基于近红外高光谱的梨叶片炭疽病与黑斑病识别[J]. *农业机械学报*, 2022, 53(2): 221 – 230.
Liu Li, Tao Hongyan, Fang Jing, et al. Identifying anthracnose and black spot of pear leaves on near-infrared hyperspectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(2): 221 – 230. (in Chinese)

[15] 邱鸿霖,刘天元,孔丽丽,等. 基于可见光谱特征波长提取和分类算法的柑橘黄龙病快检研究[J]. *光谱学与光谱分析*, 2024, 44(6): 1518 – 1525.
Qiu Honglin, Liu Tianyuan, Kong Lili, et al. Rapid detection of citrus huanglongbing based on extraction of characteristic wavelength of visible spectrum and classification algorithm[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(6): 1518 – 1525. (in Chinese)

[16] Nguyen C, Sagan V, Maimaitiyiming M, et al. Early detection of plant viral disease using hyperspectral imaging and deep learning[J]. *Sensors*, 2021, 21(3): 742.

[17] Bagheri N, Mohamadi-Monavar H, Azizi A, et al. Detection of Fire Blight disease in pear trees by hyperspectral data[J]. *European Journal of Remote Sensing*, 2018, 51(1): 1 – 10.

[18] Terentev A, Badenko V, Shaydayuk E, et al. Hyperspectral remote sensing for early detection of wheat leaf rust caused by *puccinia triticina*[J]. *Agriculture*, 2023, 13(6): 1186.

[19] 郭志明,赵春江,黄文倩,等. 苹果糖度高光谱图像可视化预测的光强度校正方法[J]. *农业机械学报*, 2015, 46(7): 227 – 232.
Guo Zhiming, Zhao Chunjiang, Huang Wenqian, et al. Intensity correction of visualized prediction for sugar content in apple using hyperspectral imaging[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(7): 227 – 232. (in Chinese)

[20] 高升,王巧华,施行,等. 便携式红提葡萄多品质可见/近红外检测仪设计与试验[J]. *农业机械学报*, 2021, 52(2): 308 – 319.
Gao Sheng, Wang Qiaohua, Shi Hang, et al. Design and test of portable red globe grape extraction multi-quality visible near infrared detector[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(2): 308 – 319. (in Chinese)

[21] 王东,孙欣,张岳阳,等. 辣椒叶片水分胁迫高光谱特性研究[J]. *农业机械学报*, 2024, 55(7): 336 – 344.
Wang Dong, Sun Xin, Zhang Yueyang, et al. Spectral characteristics of water stress in chili pepper leaves[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(7): 336 – 344. (in Chinese)

[22] 黄亦其,刘祥焕,黄震宇,等. 基于高光谱成像的甘蔗叶片早期轮斑病与锈病识别技术[J]. *农业机械学报*, 2023, 54(4): 259 – 267.
Huang Yiqi, Liu Xianghuan, Huang Zhenyu, et al. Identification of early wheel spot and rust on sugarcane leaves based on spectral analysis[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(4): 259 – 267. (in Chinese)

[23] 孙嘉豪,张伟,施鉴苓,等. 光谱数据预处理策略选择及应用[J]. *计量学报*, 2023, 44(8): 1284 – 1292.
Sun Jiahao, Zhang Wei, Shi Jianqin, et al. Selection and application of spectral data preprocessing strategy[J]. *Acta Metrologica Sinica*, 2023, 44(8): 1284 – 1292. (in Chinese)

[24] 刘文政,周雪健,平凤娇,等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. *农业机械学报*, 2024, 55(2): 372 – 383.
Liu Wenzheng, Zhou Xuejian, Ping Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible-near-infrared spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(2): 372 – 383. (in Chinese)

[25] 张楠楠,张晓,王城坤,等. 基于高光谱和连续投影算法的棉花叶面积指数估测[J]. *农业机械学报*, 2022, 53(增刊 1): 257 – 262.
Zhang Nannan, Zhang Xiao, Wang Chengkun, et al. Cotton LAI estimation based on hyperspectral and successive projection algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(S1): 257 – 262. (in Chinese)

械学报, 2024,55(8): 391–400.

Liu Bo, Xu Tao, Xu Qiangqiang, et al. Estimation of gross primary production of paddy field in Jiangxi Province based on remote sensing measured sun-induced chlorophyll fluorescence and its correlation with meteorological factors[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(8): 391–400. (in Chinese)

[40] Luyssaert S, Inglima I, Jung M, et al. CO₂ balance of boreal, temperate, and tropical forests derived from a global database [J]. Global Change Biology, 2007,13(12): 2509–2537.

[41] Johnston A S A, Meade A, Ardö J, et al. Temperature thresholds of ecosystem respiration at a global scale[J]. Nature Ecology & Evolution, 2021,5(4): 487–494.

[42] Garbulsky M, Penuelas J, Papale D, et al. Patterns and control of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems[J]. Global Ecology and Biogeography, 2010,19: 253–267.

[43] Curiel Yuste J, Baldocchi D, Misson L, et al. Microbial soil respiration and its dependency on carbon inputs, soil temperature and moisture in two contrasting ecosystems[J]. AGU Fall Meeting Abstracts, 2006,13(9):2018–2035.

[44] Chen X, Mo X, Hu S, et al. Contributions of climate change and human activities to ET and GPP trends over North China Plain from 2000 to 2014[J]. Journal of Geographical Sciences, 2017,27(6): 661–680.

[45] 陈治仲, 陈上, 线子仪, 等. 黄淮海平原植被日光诱导叶绿素荧光对骤旱的响应[J]. 农业机械学报, 2024, 55(10): 327–338.

Chen Zhizhong, Chen Shang, Xian Ziyi, et al. Response of solar-induced chlorophyll fluorescence to flash drought in Huang–Huai–Hai Plain[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10): 327–338. (in Chinese)

[46] Liang L, Geng D, Yan J, et al. Remote sensing estimation and spatiotemporal pattern analysis of terrestrial net ecosystem productivity in China[J]. Remote Sensing, 2022,14(8): 1902.

~~~~~

(上接第 295 页)

[26] 田有文, 吴伟, 林磊, 等. 基于深度学习与高光谱成像的蓝莓果蝇虫害无损检测[J]. 农业机械学报, 2023, 54(1): 393–401.

Tian Youwen, Wu Wei, Lin Lei, et al. Nondestructive detection of blueberry fruit fly pests based on deep learning and hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1): 393–401. (in Chinese)

[27] 董建江, 田野, 张建兴, 等. 基于随机森林算法的底栖动物高光谱数据分类方法研究[J]. 光谱学与光谱分析, 2023, 43(10): 3015–3022.

Dong Jianjiang, Tian Ye, Zhang Jianxing, et al. Research on the classification method of benthic fauna based on hyperspectral data and random forest algorithm[J]. Spectroscopy and Spectral Analysis, 2023, 43(10): 3015–3022. (in Chinese)

[28] 祝清震, 朱艳秋, 王爱臣, 等. 基于无人机多光谱植被指数与纹理特征的水稻叶绿素含量反演[J]. 农业机械学报, 2024, 55(12): 287–293.

Zhu Qingzhen, Zhu Yanqiu, Wang Aichen, et al. Accurate inversion of rice chlorophyll content by integrating multispectral and texture features derived from UAV multispectral imagery [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 287–293. (in Chinese)

[29] Bari B S, Islam M N, Rashid M, et al. A real-time approach of diagnosing rice leaf disease using deep learning-based Faster R–CNN framework[J]. PeerJ Computer Science, 2021, 7: e432.

[30] 赵娟, 沈懋生, 浦育歌, 等. 基于近红外光谱与多品质指标的苹果出库评价模型研究[J]. 农业机械学报, 2023, 54(2): 386–395.

Zhao Juan, Shen Maosheng, Pu Yuge, et al. Out-of-warehouse evaluation and prediction model of apple based on near-infrared spectroscopy combined with multiple quality indexes [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 386–395. (in Chinese)

[31] Feng Z H, Wang L Y, Yang Z Q, et al. Hyperspectral monitoring of powdery mildew disease severity in wheat based on machine learning[J]. Frontiers in Plant Science, 2022, 13: 828454.

[32] Zhang X, Vinatzer B A, Li S. Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease[J]. Scientific Reports, 2024, 14(1): 27666.

[33] Gold K M, Townsend P A, Chlus A, et al. Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato[J]. Remote Sensing, 2020, 12(2): 286.