网络出版时间:2026-03-13 DOI:10.13207/j.jnwafu.2026.09.007

基于递归特征消除的烟草花叶病高光谱精准识别

盛鑫[1],郝兴安[2],杨照[3],李顺欣[1],刘星[1],冯明晨[1],毛锐[1,4]

(1 西北农林科技大学 信息工程学院,陕西 杨凌 712100;2 西北农林科技大学 植物保护学院,陕西 杨凌 712100;3 陕西省烟草公司咸阳市公司,陕西 咸阳 712000;4 陕西省农业信息智能感知与分析工程技术研究中心,陕西 杨凌 712100)

[摘 要] 【目的】针对烟草花叶病毒(tobacco mosaic virus,TMV)早期侵染引发的微弱光谱变化难以有效捕捉与可靠判别等问题,构建基于生理敏感特征优选与可解释性机器学习的TMV高光谱精准识别方法体系,为TMV的早期识别与精准防控提供技术支撑。【方法】采集健康与TMV早期侵染烟叶的高光谱(400~1 000 nm)反射数据,经白板校正、感兴趣区域提取等预处理后,计算归一化色素叶绿素指数(normalized pigment chlorophyll index,NPCI)、水分指数(water index, WI)等52个候选植被指数;采用递归特征消除(recursive feature elimination,RFE)筛选最优特征子集;在此基础上构建并对比评估K近邻(K-nearest neighbors,KNN)、支持向量机(support vector machine,SVM)等4类机器学习模型的性能,结合随机搜索与网格搜索优化超参数,并利用独立测试集验证模型泛化能力;进一步引入沙普利加性解释(shapley additive explanations,SHAP),对最优模型进行可解释性分析,量化关键特征对分类决策的贡献。【结果】经RFE筛选得到NPCI与WI双特征子集,证明色素与水分变化是TMV早期识别的关键信息,有效解决了高光谱数据的维数灾难问题。基于双特征子集构建的最佳识别模型KNN,其受试者工作特征曲线下面积(receiver operating characteristic-area under curve,ROC-AUC)、5折交叉验证准确率(accuracy,ACC)及F1分数(balanced F score,F1-score)分别为0.987,0.955和0.961;在独立测试集上,ROC-AUC仍达0.960,泛化性能良好。SHAP分析表明,NPCI与WI为驱动模型决策的核心因子,其值升高会正向推动样本被判定为"TMV侵染",其变化与病毒侵染诱导的叶绿素降解及水分胁迫过程相一致。【结论】仅利用NPCI与WI 2个植被指数即可实现对TMV的早期高精度识别,显著降低了模型复杂度。所提出的"RFE特征优选+KNN+SHAP"方法体系在提升识别性能的同时,揭示了"色素变化为主、水分变化为辅"的早期光谱响应机理,为TMV的早期快速识别及低成本、便携式烟草TMV监测装备的研发提供了参考依据。

[关键词] 烟草花叶病毒;早期识别;高光谱成像;植被指数;递归特征消除;SHAP可解释性分析

[中图分类号] S432.4+1;O657.7 [文献标志码] A [文章编号] 1671-9387(2026)09-0067-11

Hyperspectral precise detection of tobacco mosaic disease using recursive feature elimination

SHENG Xin[1],HAO Xing'an[2],YANG Zhao[3],LI Shunxin[1],LIU Xing[1],
FENG Mingchen[1],MAO Rui[1,4]

(1 *College of Information Engineering,Northwest A&F University,Yangling,Shaanxi* 712100,*China*;
2 *College of Plant Protection , Northwest A&F University,Yangling,Shaanxi* 712100,*China*;
3 *Shaanxi Tobacco Company Xianyang City Company,Xianyang,Shaanxi* 712000,*China*;4 *Shaanxi Engineering Research Center of Agricultural Information Intelligent Perception and Analysis,Yangling,Shaanxi* 712100,*China*)

Abstract:【Objective】Aiming to address the challenge of effectively capturing and reliably discriminating

[收稿日期] 2025-07-28 [修回日期] 2025-09-24
[基金项目] 中国烟草总公司陕西省公司2024年度科技项目(KJ-2024-06);陕西省秦创原"科学家+工程师"队伍建设项目(2025QCY-KXJ-070)
[作者简介] 盛鑫(2001—),女,河南濮阳人,在读硕士,主要从事人工智能、病害智能识别模型研究。E-mail:fcp@nwafu.edu.cn
[通信作者] 毛锐(1977—),女,陕西三原人,副教授,博士,硕士生导师,主要从事智慧农业、人工智能研究。E-mail:maorui@nwafu.edu.cn

the subtle spectral changes caused by early infection of tobacco mosaic virus (TMV), a precise hyperspectral recognition framework based on physiologically sensitive feature selection and interpretable machine learning was developed, to provide technical support for early TMV detection and precise control. 【Method】 Hyperspectral (400－1 000 nm) reflectance data of healthy and early TMV infected tobacco leaves were collected. After pretreatment with whiteboard correction and region of interest extraction, fifty-two candidate vegetation indices, including normalized pigment chlorophyll index (NPCI) and water index (WI), were calculated. Recursive feature elimination (RFE) was used to select the optimal feature subset. Based on the selected subset, four categories of machine-learning models, including k-nearest neighbors (KNN) and support vector machine (SVM), were developed and comparatively evaluated. Hyperparameters were optimized using random search and grid search, and model generalization performance was assessed using an independent test set. Furthermore, shapley additive explanations (SHAP) was introduced to interpret the best-performing model by quantifying the contribution of key features to classification decisions. 【Result】 A dual-feature subset consisting of NPCI and WI was obtained through RFE screening, demonstrating that pigment and water-related changes are key information for early TMV identification, which effectively mitigates the curse of dimensionality in hyperspectral data. The KNN model built on the dual-feature subset achieved the best performance, with receiver operating characteristic-area under curve (ROC-AUC) of 0.987, five-fold cross-validation accuracy (ACC) of 0.955 and balanced F score (F1-score) of 0.961. On the independent test set, the ROC-AUC remained as high as 0.960, indicating good generalization performance. SHAP analysis further showed that NPCI and WI were the dominant drivers of model decision-making: an increase in their values positively drove samples to be classified as "TMV-infected", and its variation was consistent with TMV-induced chlorophyll degradation and water-stress processes. 【Conclusion】 Early and precise identification of TMV can be achieved using only two vegetation indices (NPCI and WI), substantially reducing model complexity. The proposed "RFE-based feature selection＋KNN＋SHAP" methodological framework not only improves identification performance but also reveals an early spectral response mechanism dominated by pigment changes and supplemented by water status variations. This provides a reference for the early and rapid identification of TMV, as well as for the development of low-cost, portable TMV monitoring equipment for tobacco.

Key words: tobacco mosaic virus; early detection; hyperspectral imaging; vegetation indices; recursive feature elimination; shapley additive explanations analysis

烟草(*Nicotiana tabacum* L.)是我国重要的特色经济作物,其全产业链年均税收及财政贡献已突破1.5万亿元,在国民经济中占据重要战略地位。然而,烟草花叶病毒(tobacco mosaic virus,TMV)的频发严重制约了烟叶的产量与品质,重病田块减产可超50%。TMV侵染早期主要引发生理生化层面的微弱变化(如叶绿素降解),而明显的纹理及形态异常往往在侵染中后期显现[1-2]。传统的田间目测法识别TMV的主观性强且效率低下,而酶联免疫吸附(enzyme-linked immunosorbent assay,ELISA)和分子生物学(reverse transcription-polymerase chain reaction,RT-PCR)等实验室检测技术,虽然精度高但具有破坏性且时效性差,难以满足现代智慧农业对病害早期识别的实时需求。因此,构建TMV早期精准快速识别的智能化方法具有重要的实际意义。

基于RGB图像的传统机器视觉方法难以捕捉人眼不可见的早期微弱光谱变化[3],相比之下,高光谱成像技术能够获取可见光至近红外波段的连续光谱信息,对叶片内部水分、色素及细胞结构的细微变化具有极高的敏感度,已被证明在番茄斑萎病毒等病害早期识别中优于传统RGB方法[4-8]。然而,高光谱技术的应用仍面临"数据高维冗余"与"模型可解释性不足"两大挑战。首先,TMV早期侵染引起的光谱变化幅度极小,且数百个连续波段间存在高度共线性,直接利用全波段建模不仅计算负荷大,且极易引发"维数灾难",导致模型过拟合且泛化能力下降[9]。其次,现有研究虽通过连续投影算法等手段使特征波段筛选达到了0.85的分类精度[6],但其往往侧重于数据挖掘层面的显著性特

征统计，缺乏对特征波段背后生理学原理的深入解析。在复杂田间环境下，病害信号常与干旱、营养缺失等非生物胁迫信号重叠[3,10]，若缺乏明确的生理学解析支撑，模型的鲁棒性与可信度将会受到质疑。

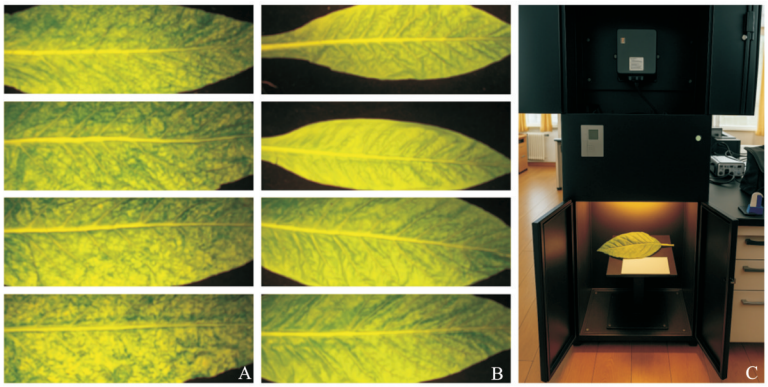
植被指数（vegetation indices，VIs）通过特定波段的数学组合，能有效增强目标信号并抑制背景噪声，是量化植物色素、水分及健康状态的有效手段[11]。相较于全波段数据，基于生理机制构建的植被指数（如叶绿素指数、水分指数）更能直观地反映植物在胁迫下的特异性响应。但是不同植被指数对特定病害的敏感度存在显著差异，针对 TMV 早期侵染的特异性特征，目前尚缺乏系统研究[12-13]。增强机器学习模型的可解释分析，阐明特征与预测结果间的量化关系，是推动该技术落地的关键。

鉴于此，本研究以 TMV 侵染早期烟叶为对象，构建了基于生理敏感特征优选与可解释性机器学习的高光谱精准识别方法体系，通过筛选对 TMV 早期胁迫响应最敏感的特征子集确定最优识别模型，旨在明确 TMV 早期识别的关键光谱特性，为 TMV 的早期识别与精准防控提供理论依据与技术支撑。

1　烟叶高光谱数据采集与预处理

1.1　试验材料

试验烟叶样本采自陕西省咸阳市旬邑县马栏烟区，该区属暖温带半干旱气候，土壤类型为潮土。为确保样本具有代表性，分别于 2024 和 2025 年分批次采集不同田块烟叶，并严格剔除虫害、营养缺失等非病毒因素造成的异常叶片。经美国 Agdia 公司 TMV 检测试纸条确诊，将样本分为 TMV 侵染叶片（早期黄化，微弱斑点，图 1-A）与健康叶片（无病变，图 1-B）两类。以 2024 年采集的 110 个样本（其中 TMV 侵染 63 个，健康 47 个）作为建模集，2025 年采集的 50 个样本（其中 TMV 侵染 27 个，健康 23 个）作为独立测试集，评估模型的泛化性能。



A. TMV 侵染烟叶；B. 健康烟叶；C. 高光谱数据采集系统。
A. Tobacco leaves infected with TMV; B. Uninfected tobacco leaves; C. Hyperspectral data acquisition system.
图 1　烟草叶片 TMV 侵染表型特征及高光谱数据采集平台
Fig. 1　Phenotypic characteristics of TMV-infected tobacco leaves and hyperspectral data acquisition platform

1.2　高光谱数据采集系统

高光谱数据采集使用的推扫式高光谱成像系统（SOC710，美国 Surface Optics 公司）如图 1-C 所示，光谱范围为 400～1 000 nm，波段数 512 个，图像空间分辨率为 1 392 像素×1 040 像素。采集在暗室中进行，利用 12 V/100 W 卤钨灯配合积分球提供均匀漫反射照明。采集参数设置为：物距（62.0±0.5）cm，曝光时间 30 ms。采集过程同步获取标准白板与暗电流数据，用于后续辐射校正。

1.3　高光谱数据预处理

利用 ENVI5.6 软件对原始影像进行高光谱数据预处理，包括白板校正，计算公式为：

$$R(\lambda)=\frac{I_{\mathrm{leaf}(\lambda)}-I_{\mathrm{dark}(\lambda)}}{I_{\mathrm{white}(\lambda)}-I_{\mathrm{dark}(\lambda)}}。\tag{1}$$

式中：$R(\lambda)$ 为叶片在波长 λ 处的反射率，$I_{\mathrm{leaf}(\lambda)}$、$I_{\mathrm{white}(\lambda)}$、$I_{\mathrm{dark}(\lambda)}$ 分别为叶片、标准白板及暗电流的原始像元值。

通过阈值分割结合人工修正提取感兴趣区域（region of interest，ROI），并构建掩膜剔除背景。计算 ROI 内高光谱立方体的空间均值，得到每片烟叶样本在 400～1 000 nm 处的平均光谱反射率曲线，以该曲线作为后续植被指数计算及机器学习建模的核心光谱特征。

1.4 植被指数特征集构建

为解决全波段数据冗余问题,综合已有文献的研究结果[14-19],筛选并计算了52个与叶绿素含量、水分状态及细胞结构高度相关的植被指数,其详细信息如表1所示。据此,构建了52维初始特征向量,涵盖红边、近红外-短波红外及宽光谱结构指数等。对所有特征进行Z分数(Z-score)标准化处理,消除量纲差异,为后续特征筛选与建模奠定数据基础。

表1 52个植被指数及其计算公式

Table 1 52 vegetation indices and their calculation formulas

简称 Abbreviation	植被指数名称 Name of the vegetation index	计算公式 Calculation
NPCI[14]	归一化色素叶绿素指数 Normalized pigment chlorophyll ratio index	$(R_{680}-R_{430})/(R_{680}+R_{430})$
WI[14]	水分指数 Water index	R_{900}/R_{970}
Lic2[15]	Lichtenthaler 二号指数 Lichtenthaler index 2	R_{440}/R_{690}
BGI2[14]	蓝绿色素指数2 Blue green pigment index	R_{450}/R_{550}
HI1[16]	健康指数1 Health-index 1	$(R_{739}-R_{402})/(R_{739}+R_{402})-0.5R_{403}$
PSNDc[14]	色素特异性归一化差异指数c Pigment specific normalized difference c	$(R_{800}-R_{470})/(R_{800}+R_{470})$
TCARI2[17]	二成分农业遥感指数2 Two-component agriculture remote sensing index 2	$3(R_{700}-R_{670}-0.2(R_{700}-R_{550})R_{700}/R_{670})$
ND[14]	归一化差异指数 Normalized difference index	$(R_{800}-R_{530})/(R_{800}+R_{530})$
ARI[16]	花青素反射指数 Anthocyanin reflectance index	$1/R_{550}-1/R_{700}$
PhRI[16]	生理反射指数 Physiological reflectance index	$(R_{550}-R_{531})/(R_{550}+R_{531})$
GI[16]	绿色指数 Green index	R_{554}/R_{677}
PSRI4[18]	植物衰老反射指数4 Plant senescence reflectance index 4	$(R_{690}-R_{500})/R_{750}$
PSSRc[14]	色素特异性简单比值指数c Pigment specific simple ratio c	R_{800}/R_{470}
NRI[16]	氮反射指数 Nitrogen reflectance index	$(R_{570}-R_{670})/(R_{570}+R_{670})$
mSR[14]	改良红边比值指数 Modified red-edge ratio	$(R_{750}-R_{445})/(R_{705}-R_{445})$
PSRI1[18]	植物衰老反射指数1 Plant senescence reflectance index 1	$(R_{678}-R_{500})/R_{750}$
RENDVI[17]	红边归一化差异指数 Red-edge normalized difference vegetation index	$(R_{750}-R_{705})/(R_{750}+R_{705})$
PRI[17]	光化学指数 Photochemical reflectance index	$(R_{531}-R_{570})/(R_{531}+R_{570})$
PSI[14]	植物衰老反射指数 Plant senescence reflectance index	$(R_{680}-R_{500})/R_{750}$
SR[14]	简单比值指数 Simple ratio	R_{695}/R_{760}
NDVI[14]	归一化差异植被指数 Normalized difference vegetation index	$(R_{800}-R_{670})/(R_{800}+R_{670})$
SIPI[16]	结构不敏感色素指数 Structure insensitive pigment index	$(R_{800}-R_{445})/(R_{800}+R_{680})$
CLS2[17]	叶绿素a+b含量指数2 Chlorophyll a+b content index 2	$(R_{698}-R_{570})/(R_{698}+R_{570})-R_{734}$
PSNDb[14]	色素特异性归一化差异指数b Pigment specific normalized difference b	$(R_{800}-R_{650})/(R_{800}+R_{650})$
PSSRa[14]	色素特异性简单比值指数a Pigment specific simple ratio a	R_{800}/R_{675}
TCARI1[16]	二成分农业遥感指数1 Two-component agriculture remote sensing index 1	$3(R_{717}-R_{668}-0.2(R_{717}-R_{668})R_{717}/R_{668})$
SBRI[17]	土壤调整植被指数 Soil-adjusted vegetation index	$(R_{570}-R_{531})/(R_{570}+R_{531})+0.5R_{704}$
PSNDa[14]	色素特异性归一化差异指数a Pigment specific normalized difference a	$(R_{800}-R_{675})/(R_{800}+R_{675})$
HI3[14]	健康指数3 Health-index 3	$(R_{534}-R_{698})/(R_{534}+R_{698})-0.5R_{704}$
MSR[16]	改良简单比值指数 Modified simple ratio index	$(R_{800}/R_{670}-1)/\sqrt{R_{800}/R_{670}+1}$
PMI2[17]	白粉病指数2 Powdery mildew index 2	$(R_{520}-R_{584})/(R_{520}+R_{584})+R_{724}$
PMI1[16]	白粉病指数1 Powdery mildew index 1	$(R_{515}-R_{698})/(R_{515}+R_{698})-0.5R_{738}$
YRI1[16]	黄锈病指数 Yellow rust index	$(R_{730}-R_{419})/(R_{730}+R_{419})-0.5R_{736}$
CRI1[17]	类胡萝卜素反射率指数1 Carotenoid ratio index 1	$1/R_{510}-1/R_{550}$
CVI[16]	叶绿素植被指数 ChlorophyII vegetation index	$R_{840}R_{668}/R_{560}R_{560}$
CIRE[16]	红边叶绿素指数 ChlorophyII index-rededge	$R_{840}/R_{717}-1$
CIG[16]	绿光叶绿素指数 ChlorophyII index-green	$R_{840}/R_{668}-1$
RVSI[16]	红边植被胁迫指数 Red-edge vegetation stress index	$(R_{712}+R_{752})/2-R_{732}$
Lic1[15]	Lichtenthaler 一号指数 Lichtenthaler index 1	$(R_{800}-R_{680})/(R_{800}+R_{680})$
mNDI[14]	红边归一化指数 Modified red-edge normalized difference vegetation index	$(R_{750}-R_{705})/(R_{750}-R_{705}-2R_{445})$
MCARI4[18]	改良叶绿素吸收比值指数4 Modified chlorophyll absorption ratio index 4	$R_{700}-R_{670}-0.2(R_{700}-R_{550})R_{700}/R_{670}$

表 1（续）　Table 1 (continued)

简称 Abbreviation	植被指数名称 Name of the vegetation index	计算公式 Calculation
SAVI[16]	土壤调整植被指数 Soil adjusted vegetation index	$1.5(R_{840}-R_{668})/(R_{840}+R_{668}+0.5)$
GNDVI[16]	绿色归一化差异植被指数 Green normalized difference vegetation index	$(R_{840}-R_{560})/(R_{840}+R_{560})$
PSSRb[14]	色素特异性简单比值指数 b Pigment specific simple ratio b	R_{800}/R_{650}
EVI[16]	增强型植被指数 Enhanced vegetation index	$2.5(R_{840}-R_{668})/(R_{840}+6R_{668}-7.5R_{475}+1)$
TVI[16]	三角形植被指数 Triangular vegetation index	$0.5(120(R_{750}-R_{550})-200(R_{670}+R_{550}))$
NDVI2[14]	归一化差异植被指数 2 Normalized difference vegetation index 2	$(R_{840}-R_{675})/(R_{840}+R_{675})$
OSAVI2[19]	优化土壤调整植被指数 2 Optimized soil-adjusted vegetation index 2	$1.16(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$
CLS3[14]	叶绿素 a+b 含量指数 3 Chlorophyll a+b content index 3	$(R_{698}-R_{570})/(R_{698}+R_{570})+R_{734}$
MCARI1[16]	改良叶绿素吸收比值指数 1 Modified chlorophyll absorption ratio index 1	$R_{701}-R_{671}-0.2(R_{701}-R_{549})R_{671}/R_{701}$
OSAVI[19]	优化土壤调整植被指数 Optimized soil-adjusted vegetation index	$1.16(R_{840}-R_{668})/(R_{840}+R_{668}+0.16)$
RVI[16]	比值植被指数 Ratio vegetation index	R_{840}/R_{668}

注：表中 R_{680} 表示波长 680 nm 下的光谱反射率。其余以此类推。

Note: In the table, R_{680} represents the spectral reflectance at a wavelength of 680 nm. The rest can be deduced by analogy.

2　模型构建与性能评估

2.1　基于 RFE 的特征筛选

采用递归特征消除算法[20]（recursive feature elimination，RFE）剔除冗余变量，解决 52 个植被指数间的多重共线性问题。结合随机森林（random forest，RF）[21]算法，利用平均不纯度减少量（mean decrease of impurity importance，MDI）衡量特征重要性，其表达式为：

$$I_j=\sum_{t\in T_j}p(t)\Delta i(t)。\tag{2}$$

式中：I_j 为第 j 个特征的基尼重要性；T_j 为使用特征 j 进行分裂的节点集合；$p(t)$ 表示节点 t 的样本占比，$\Delta i(t)$ 为节点 t 使用该特征后的基尼不纯度下降量。

通过迭代训练，每轮迭代均按公式（3）规则，移除重要性最低的特征，直至达到设定的最小特征数 $d_{\min}$。结合 5 折交叉验证的平均准确率变化曲线，确定能够表征 TMV 早期特征的最优特征子集。特征的重要性越高，说明其在病害样本识别中提供的信息增益越大。

$$F_{K+1}=F_K\setminus\{\arg\min_{f\in F_K}I_j(f)\}。\tag{3}$$

式中：F_K、F_{K+1} 分别为第 K、$K+1$ 轮剩余特征集合；“\”为集合差（移除元素）；f 为候选特征，$I_j(f)$ 表示以特征 f 训练当前 RF 后得到的整体基尼指标，$\arg\min_{f\in F_K}I_j(f)$ 为本轮使基尼最小的特征。

2.2　机器学习模型优选与参数调优

为确定 TMV 识别的最优模型，本研究系统评估了 4 大类 13 种机器学习模型，其中包括线性与概率类的逻辑回归[22]（logistic regression，LR）、线性判别分析[23]（linear discriminant analysis，LDA）、二次判别分析[24]（quadratic discriminant analysis，QDA）、朴素贝叶斯[25]（gaussian naive bayes，NB），核方法与实例类的支持向量机[26]（support vector machine，SVM）、K 近邻[27]（k-nearest neighbors，KNN），树模型与集成学习类的决策树[28]（decision tree，DT）、RF、自适应提升[29]（adaptive boosting，AdaBoost）、梯度提升决策树[30]（gradient boosting decision tree，GBDT）、极端梯度提升[31]（extreme gradient boosting，XGBoost）以及轻量级梯度提升机[32]（light gradient boosting machine，LightGBM），神经网络类的多层感知器[33]（multilayer perceptron，MLP）。模型训练基于 2.1 节筛选的最优特征子集。采用“粗调+精调”策略进行超参数优化，首先采用随机搜索确定参数的大致范围，随后利用网格搜索在局部范围内以 5 折交叉验证准确率为目标函数锁定最优参数。

13 种模型的最优参数配置见表 2。此处以 KNN 模型为例说明超参数调优过程。首先，采用随机搜索，叶节点在[10，60]随机选择，邻居数在区间[1，50]随机选取，距离参数（p）从曼哈顿距离（$p=1$）和欧氏距离（$p=2$）中选择，权重参数在均匀加权和距离加权之间随机抽样。该过程共执行 50 次随机选择，最后获得性能最佳的参数组合为：邻居数为 15，距离参数（p）为 2，叶节点规模为 25，权重参数采用距离加权。之后采用网格搜索，即在第一阶段最佳参数附近进行精细搜索。邻居数以最优值 15 为中心，在区间[10，20]以步长 1 搜索；叶节点规模以最优值 25 为中心，在区间[15，35]以步长 5 搜索；距离参数固定为欧氏距离（$p=2$），权重模式固定为距离加权。最终确定的模型最优参数为：邻居数为 10，距离参数 p 为 2，叶节点规模为 15，权重参数采用距离加权。

表2 13个机器学习模型的最佳参数配置

Table 2 Optimal parameter configurations for 13 machine learning models

模型 Model	参数 Parameter	模型 Model	参数 Parameter
LR	惩罚系数为7.426,正则化选择参数为L_2正则,迭代上限为10^3 Penalty coefficient is 7.426, regularization selection parameter is L_2 regularization, and the upper limit of iterations is 10^3.	RF	树数为70,最大深度为3 The number of trees is 70, and the maximum depth is 3
LDA	收缩系数为0.05 Shrinkage coefficient is 0.05	AdaBoost	学习率为0.25,弱分类器数为75 Learning rate is 0.25, and the number of weak classifiers is 75
QDA	正则化参数为0 Regularization parameter is 0	GBDT	学习率为0.15,迭代次数为150 Learning rate is 0.15, and the number of iterations is 150
NB	方差平滑参数为10^{-9} Variance smoothing parameter is 10^{-9}	XGBoost	学习率为0.05,最大深度为2 Learning rate is 0.05, and the maximum depth is 2
SVM	惩罚系数为1.0,核函数为RBF Penalty coefficient is 1.0, and the kernel function is RBF	LightGBM	学习率为0.05,叶节点数为16 Learning rate is 0.05, and the number of leaf nodes is 16
KNN	邻居数为10,距离参数为2,叶节点规模为15,权重参数为距离加权 The number of neighbors is 10, the distance parameter is 2, the scale of leaf nodes is 15, and the weight parameter is distance weighted	MLP	正则化系数为10^{-5} Regularization coefficient is 10^{-5}
DT	最大深度为5 The maximum depth is 5		

2.3 模型性能评估

采用受试者工作特征曲线下面积(receiver operating characteristic-area under curve, ROC-AUC)、准确率(accuracy, ACC)及F1分数(balanced F score, F1-score)作为核心评价指标,分别评估模型的综合判别能力、整体预测精度及正负样本的平衡性。

ROC-AUC是基于真阳性率(T_{TPR})和假阳性率(F_{FPR})在不同分类阈值下的变化关系,计算公式为:

$$R=\int_0^1 T_{TPR}(F_{FPR})\,\mathrm{d}(F_{FPR})。\tag{4}$$

式中:R表示ROC-AUC值;T_{TPR}与召回率等价,反映模型对正类样本的识别能力;F_{FPR}是负类样本被错误识别为正类样本的比例。

ACC用于衡量模型对全部样本的整体预测正确性,其计算公式为:

$$A=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}}。\tag{5}$$

式中:A表示ACC值,N_{TP}表示被正确识别为病毒侵染的病叶数,N_{TN}表示被模型正确识别为健康的叶片数,N_{FP}表示被模型错误识别为病毒侵染的健康叶片数,N_{FN}表示被模型识别为健康的病叶数。

F1-score是精确率(P)与召回率(R)的调和平均值,用于平衡模型对正类样本的"查准率"与"查全率",其计算公式为:

$$F=2\frac{P\times R}{P+R}。\tag{6}$$

式中:F表示F1-score值,P是衡量正类预测结果中实际为正类的比例,R是衡量实际正类样本中被正确识别的比例。

研究采用ROC-AUC、ACC、F1-score 3个指标协同评估,比较4大类机器学习模型在最优特征子集上的性能,以筛选TMV识别性能优异的模型。为验证模型性能差异的显著性,对表现优异的模型执行10次重复的5折交叉验证,并进行配对t检验($P<0.05$),筛选鲁棒性最优的识别模型。

2.4 SHAP可解释性分析

针对筛选出的最优模型,引入沙普利加性解释[34](shapley additive explanations, SHAP)方法解析模型的决策机制。SHAP基于博弈论思想,通过计算每个特征对模型输出的边际贡献来量化其重要性。对于输入样本x,SHAP值定义为:

$$\phi_j(f,x)=\sum_{S\subseteq F\setminus\{j\}}\frac{|S|!(|F|-|S|-1)!}{|F|!}\left[f_{S\cup\{j\}}(x)-f_S(x)\right]。\tag{7}$$

式中:$\phi_j(f,x)$表示特征j对输入样本x在模型函数f中的贡献;f为模型函数,表示用于预测的机器学习模型;F表示特征全集;$S\subseteq F\setminus\{j\}$为不包含第$j$个特征的特征子集;$|F|$为特征全集$F$中特征的总数;$|S|$为集合$S$中元素的个数;!表示相应数值的阶乘;$f_{S\cup\{j\}}(x)$为在子集$S$中添加特征$j$后,对输入$x$进行预测的输出;$f_S(x)$表示仅使用特征子集$S$训练的模型在样本$x$上的输出;$\left[f_{S\cup\{j\}}(x)-f_S(x)\right]$为加入特

征j之后的预测改变量。

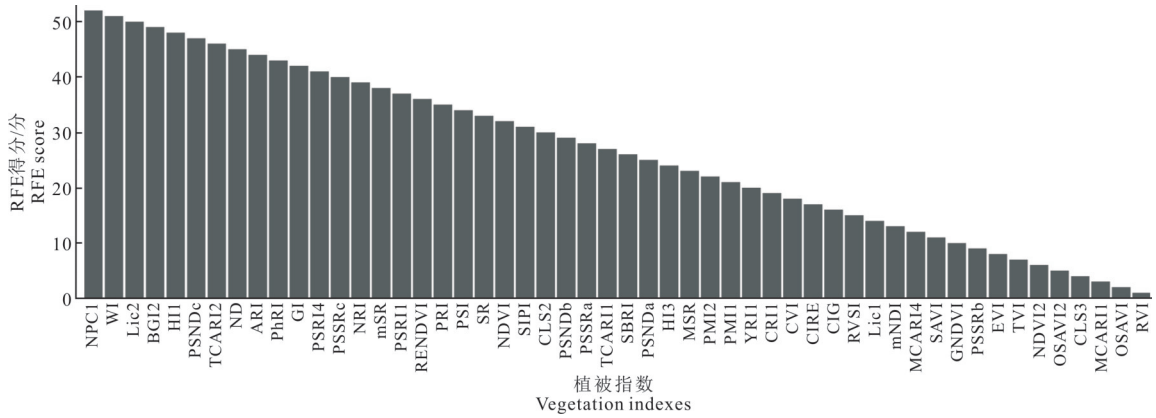
解释分析统一基于标准化后的特征输入进行,以消除特征尺度差异对边际贡献评估的影响。采用平均特征贡献条形图与样本特征方向影响点图分别揭示各特征在整体模型决策中的重要程度以及展示特征在不同样本上的方向性作用机制。

3　结果与分析

3.1　基于RFE的最优特征子集筛选

利用RFE算法,结合RF对52个候选植被指数进行重要性排序,结果如图2所示。其中,NPCI与WI的重要性得分分别为52.3与50.8,明显高于其他特征,表明叶绿素降解引起的色素变化及细胞膜受损导致的水分状态改变是区分健康与TMV侵染烟叶的最关键生理因子。

特征维度对模型分类精度的影响如图3所示。图3表明,随着特征数量增加,模型准确率呈先升后降趋势。仅保留NPCI与WI这2个特征时,模型5折交叉验证的平均准确率达到峰值(0.918)。此后,随着特征维度的持续增加,模型性能反而出现波动下滑。这表明NPCI和WI已经充分表征了TMV早期侵染的光谱差异,引入额外特征并未提供增量信息,反而引入噪声导致信息冗余,进而降低了模型的性能。



植被指数的详细信息同表1。The detailed information of the vegetation index is shown in Table 1.

图2　基于RFE特征重要度得分的条形图

Fig. 2　Bar chart of RFE-based feature importance scores

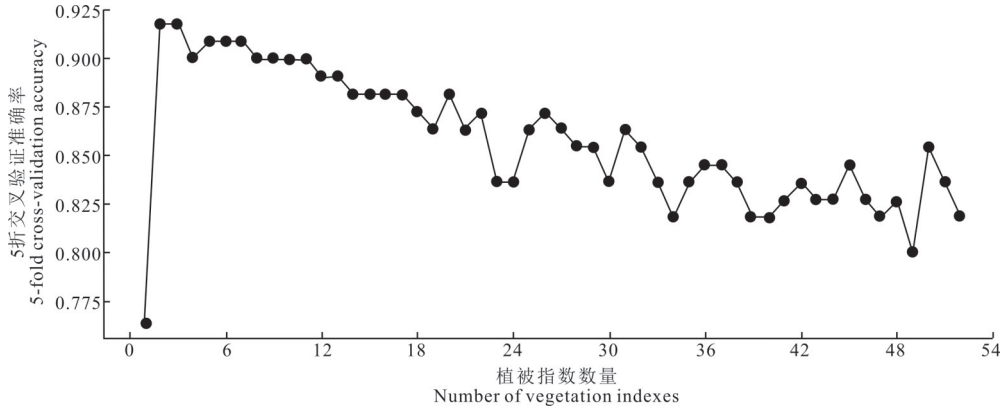


图3　不同植被指数数量下模型准确率折线图

Fig. 3　Line chart of model accuracy under different numbers of vegetation indexes

NPCI对叶绿素与类胡萝卜素的比率极为敏感。TMV侵染诱导叶绿体解体,导致红光波段(680 nm)吸收减弱、反射率升高,进而改变NPCI值。WI则主要响应叶片水分状态,病毒侵染会破坏细胞膜通透性并影响气孔导度,导致近红外波段(900和970 nm)吸收特性发生微弱变化。因此,优选出的NPCI与WI双特征子集实现了从52维到2维的高效降维,在大幅度降低计算复杂度的同时保留了核心病理信息。

3.2　不同机器学习模型的性能评估及最优模型确定

基于优选的NPCI-WI双特征子集,系统评估了13种机器学习模型的分类性能。结果(图4)显示,在ROC-AUC指标上,KNN模型表现最优(0.987);

在ACC方面,KNN与RF并列第一(0.955);在F1-score上,RF与KNN差异甚微(RF为0.962,KNN为0.961)。这表明在低维特征空间中,基于实例学习的KNN与基于集成决策的RF均能有效构建样本间的决策边界。

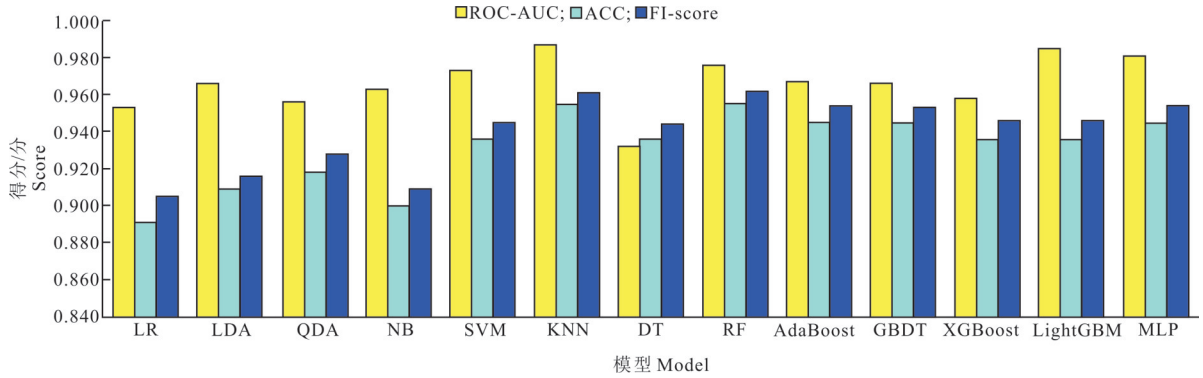


图4 基于最优双特征子集的13种模型性能的对比

Fig. 4 Comparison of performance metrics of 13 models based on optimal dual-feature subset

为进一步验证模型稳健性,对表现突出的KNN与RF执行10次重复5折交叉验证。结果(表3)显示,KNN的平均准确率为0.944±0.042,略高于RF的0.939±0.043。尽管配对t检验显示两者差异未达到统计显著水平(t=1.23,P=0.224>0.05),但KNN具有更窄的95%置信区间([0.932,0.956]),且作为惰性学习算法,其超参数更少、计算开销更低。综合考虑精度、稳定性与模型复杂度,本研究确定KNN为TMV早期识别的最优模型。

此外,为了验证所构建模型在全新数据分布下的泛化性能和预测能力,本研究采用2025年独立采集的烟草样本集开展跨数据集验证试验。结果(表4)表明,在ROC-AUC、ACC及F1-score 3个核心评价指标上,KNN模型均显著优于RF模型。

表3 KNN与RF模型在最优特征子集上的10次重复5折交叉验证性能对比

Table 3 Comparison of KNN and RF model performance on the optimal feature subset using 10 repeated 5-fold cross-validation

模型 Model	平均准确率 Mean accuracy	标准差 Standard deviation	95%置信区间 95% confidence interval
KNN	0.944	0.042	[0.932,0.956]
RF	0.939	0.043	[0.926,0.951]

表4 KNN与RF模型在独立测试集上的性能指标对比

Table 4 Comparison of performance indicators of KNN and RF models on the independent test set

模型 Model	受试者工作特征曲线下面积 Area under the receiver operating characteristic curve	准确率 Accuracy	F1分数 F1-score
KNN	0.960	0.850	0.846
RF	0.939	0.800	0.800

3.3 基于SHAP的模型可解释性分析

利用SHAP算法进一步解析了KNN模型的决策机制,全局重要性分析表明,NPCI的平均SHAP绝对值约为0.30,表示其对模型输出的平均影响幅度最大,可导致预测为TMV侵染的概率变化约30个百分点。WI次之(平均SHAP绝对值约0.17)。这一结果定量证实了在TMV早期侵染引起的光谱响应中,色素组分变化起主导作用,水分变化起协同作用。局部依赖性分析结果(图5)进一步揭示了特征值与预测结果(TMV侵染概率)的方向性关系。由图5可知,NPCI特征值与预测结果呈显著正相关,图中高值点(红色)主要分布在SHAP值的正向区间(0~0.5),表明随着NPCI值升高,模型判定样本为“TMV侵染”的概率显著增加。该结果与生理学机理相符,即TMV导致叶片褪绿黄化,使得NPCI值因红光反射率升高而增大。同理,WI特征值与预测结果亦呈正相关,高WI值倾向于驱动模型做出阳性判断,表明与水分胁迫相关的光谱变异也是判别病害的重要依据。SHAP分析不仅验证了模型决策逻辑的合理性,更从数据挖掘层面证实了TMV侵染早期“色素降解主导、水分胁迫协同”的光谱响应机制。

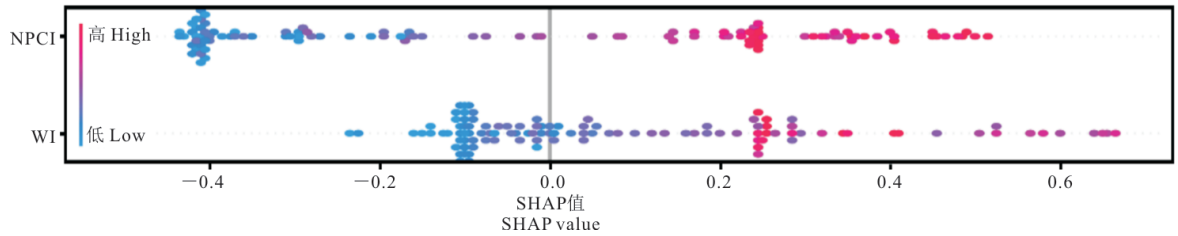


图 5　NPCI与WI特征的SHAP方向性点图

Fig. 5　SHAP directional dot plot for NPCI and WI

4　讨　论

利用RFE算法从52个候选植被指数特征中筛选出NPCI与WI作为识别TMV的最优特征子集。这一结果与HAAGSMA[7]及GRIMMER等[35]的研究结论相佐证，即病毒侵染引发的叶绿素降解与水分代谢紊乱是早期病害识别的关键生理指标。NPCI与WI的高权重表明，TMV识别无需依赖全波段高光谱数据，仅利用特定敏感波段（430，680，900和970 nm）即可实现高精度判别。SHAP分析更从数据挖掘角度量化了这2个特征的边际贡献，进一步从模型层面证实了植物生理状态变化与光谱响应之间的内在关联，验证了“病理敏感指数筛选策略”的有效性，也为未来基于多光谱传感器的低成本、便携式检测装备的研发提供了理论依据[36]。

相较于同类研究，本研究提出的“RFE特征优选＋KNN分类＋SHAP可解释”框架展现出显著优势。在花叶类型病害研究中，本研究构建的KNN模型的准确率达0.955，优于GRIFFEL等[37]基于SVM方法（0.936）及ZHU等[8]基于LDA方法（0.909）构建模型的准确率。这可能是因为经RFE降维后，特征空间中各类样本呈现出较好的局部聚类特性，使得基于距离度量的KNN算法较寻找全局超平面的算法更具适应性。此外，本研究引入SHAP解释器，填补了病害识别研究中“重精度、轻机理”的空白[38]。通过可视化特征对决策的正负向驱动作用，提升了“黑箱”模型的透明度，使其决策逻辑更符合农学专家的认知体系。

尽管确定的模型性能优异，但本研究仍存在一定局限性。当前模型本质为“健康/TMV”二分类器，由于未纳入黄瓜花叶病毒（cucumber mosaic virus，CMV）、马铃薯Y病毒（potato virus Y，PVY）等具有相似花叶症状的样本，模型在面对多病害并发时可能产生误判。未来需构建涵盖多种病毒的多分类数据集，深入挖掘不同病毒在纹理、叶脉透明度等细微特征上的光谱差异。此外，本研究数据采集环境较为理想，未考虑田间复杂光照、阴影及背景干扰。后续研究将开展大田原位识别试验，结合图像分割与光照校正算法，提升模型在开放环境下的鲁棒性与普适性。

5　结　论

针对TMV早期精准识别难题，构建了基于生理敏感特征优选与可解释性机器学习的高光谱精准识别方法体系，主要结论如下：

1）明确了TMV早期识别的关键光谱特征。利用RFE算法从52个植被指数中筛选出NPCI与WI为最优特征子集，这两者分别表征了病毒侵染导致的色素降解与水分胁迫，具有明确的生理学指征意义，且仅凭这2个特征即可实现0.918的基础分类精度，有效解决了高光谱数据的维度灾难问题。

2）构建了高性能、轻量化的识别模型。经13种机器学习算法对比，KNN模型在最优双特征子集上表现最佳，准确率达0.955，ROC-AUC为0.987，且稳定性优于RF模型。该模型结构简单，计算开销低，具备在资源受限的嵌入式设备上部署的应用潜力。

3）揭示了KNN模型的决策机理。基于SHAP的可解释性分析量化了特征贡献，证实NPCI对判别结果起主导作用（贡献率约0.30），WI起协同作用（贡献率约0.17）。该结果与植物病理学常识高度一致，即“色素变化为主、水分变化为辅”是TMV早期光谱响应的核心规律，增强了模型的可信度。

综上，本研究不仅实现了TMV的高精度早期识别，还阐明了光谱特征背后的农学机理，为研发低成本、便携式烟草病毒病检测装备提供了核心算法与理论支撑。

[参考文献]

[1] CULVER J N，ALWYN G C，LINDBECK G C，et al. Virus-host interactions：induction of chlorotic and necrotic responses in plants by tobamoviruses[J]. Annual Review of Phytopathology，1991，29：193-217.

[2] BALACHANDRAN S，HURRY V M，KELLEY S E，et al.

Concepts of plant biotic stress: some insights into the stress physiology of virus-infected plants, from the perspective of photosynthesis[J]. Physiologia Plantarum, 1997, 100(2): 203-213.

[3] MAHLEIN A K, OERKE E C, STEINER U, et al. Recent advances in sensing plant diseases for precision crop protection[J]. European Journal of Plant Pathology, 2012, 133(1): 197-209.

[4] LAWAL B Y. 基于高光谱成像技术的烟叶病害识别方法研究[D]. 杭州:浙江大学, 2012.
LAWAL B Y. Study on tobacco leaf disease recognition methods based on hyperspectral imaging technology[D]. Hangzhou: Zhejiang University, 2012.

[5] 徐冬云,李新举,杨永花,等. 基于遥感技术的烟草花叶病监测研究[J]. 中国烟草学报, 2016, 22(1): 76-83.
XU D Y, LI X J, YANG Y H, et al. Study on monitoring mosaic virus infected tobacco based on remote sensing technology [J]. Acta Tabacaria Sinica, 2016, 22(1): 76-83.

[6] GU Q, SHENG L, ZHANG T, et al. Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique and machine learning algorithms[J]. Computers and Electronics in Agriculture, 2019, 167: 105066.

[7] HAAGSMA M, HAGERTY C H, KROESE D R, et al. Detection of soil-borne wheat mosaic virus using hyperspectral imaging: from lab to field scans and from hyperspectral to multispectral data[J]. Precision Agriculture, 2023, 24(3): 1030-1048.

[8] ZHU H, CEN H, ZHANG C, et al. Early detection and classification of tobacco leaves inoculated with tobacco mosaic virus based on hyperspectral imaging technique[C]//2016 ASABE Annual International Meeting. Orlando, Florida: American Society of Agricultural and Biological Engineers, 2016: 1-7.

[9] XU B, LI X, HOU W, et al. A similarity-based ranking method for hyperspectral band selection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9585-9599.

[10] YAN K, GAO S, YAN G, et al. A global systematic review of the remote sensing vegetation indices[J]. International Journal of Applied Earth Observation and Geoinformation, 2025, 139: 104560.

[11] KAMENOVA I, DIMITROV P. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria[J]. European Journal of Remote Sensing, 2021, 54(1): 89-108.

[12] LOWE A, HARRISON N, FRENCH A P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress[J]. Plant Methods, 2017, 13(1): 80.

[13] MAHLEIN A K, RUMPF T, WELKE P, et al. Development of spectral indices for detecting and identifying plant diseases [J]. Remote Sensing of Environment, 2013, 128: 21-30.

[14] MUSTAFA G, ZHENG H, LIU Y, et al. Leveraging machine learning to discriminate wheat scab infection levels through hyperspectral reflectance and feature selection methods[J]. European Journal of Agronomy, 2024, 161: 127372.

[15] WATT M S, BUDDENBAUM H, LEONARDO E M C, et al. Using hyperspectral plant traits linked to photosynthetic efficiency to assess N and P partition[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 169: 406-420.

[16] JOSHI P, SANDHU K S, DHILLON G S, et al. Detection and monitoring wheat diseases using unmanned aerial vehicles (UAVs)[J]. Computers and Electronics in Agriculture, 2024, 224: 109158.

[17] 张凝,杨贵军,赵春江,等. 作物病虫害高光谱遥感进展与展望[J]. 遥感学报, 2021, 25(1): 403-422.
ZHANG N, YANG G J, ZHAO C J, et al. Progress and prospects of hyperspectral remote sensing technology for crop diseases and pests[J]. National Remote Sensing Bulletin, 2021, 25(1): 403-422.

[18] ZHANG H, ZHAO J, HUANG L, et al. Development of new indices and use of CARS-Ridge algorithm for wheat fusarium head blight detection using in-situ hyperspectral data[J]. Biosystems Engineering, 2024, 237: 13-25.

[19] MORELL-MONZÓ S, SEBASTIÁ-FRASQUET M T, ESTORNELL J, et al. Detecting abandoned citrus crops using Sentinel-2 time series: a case study in the comunitat Valenciana region (Spain)[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 201: 54-66.

[20] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002, 46(1): 389-422.

[21] ZHAO N, HUANG Y, MEI Y, et al. Random forest-guided decoding of multiplex role of initial brine salinity in shaping dichotomous equilibrium of microbial assembly and flavor compound in liquid phase and solid phase of radish Paocai during fermentation[J]. Food Chemistry, 2025, 502: 147699.

[22] FALAMAKI A, SHAFIEE A H, SHAFIEE A. Assessing liquefaction potential using Arias intensity: a logistic regression approach with an updated database[J]. Soil Dynamics and Earthquake Engineering, 2026, 202: 110028.

[23] PARK M, SOMBORN A, SCHLEHUBER D, et al. Non-destructive quantification of lutein and beta-carotene in spinach by Raman spectroscopy under optimized conditions for linear discriminant analysis[J]. Food Chemistry, 2025, 493(4): 146062.

[24] PRIYA S R K, BALAMBIGA R K, MISHRA P, et al. Sugarcane yield forecast using weather based discriminant analysis [J]. Smart Agricultural Technology, 2023, 3: 100076.

[25] MONDAL D, KOLE D K, ROY K. Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy based binning and Naive Bayes classifier after identification of leaves[J]. Computers and Electronics in Agriculture, 2017, 142: 485-493.

[26] RAY K K, KUMARI A, KUMAR S, et al. Guava leaf disease detection using support vector machine (SVM)[J]. Smart Agricultural Technology, 2025, 12: 101190.

[27] DENG N, XU R, ZHANG Y, et al. Forest biomass carbon stock estimates via a novel approach: K-nearest neighbor-based weighted least squares multiple birth support vector re-

gression coupled with whale optimization algorithm[J]. Computers and Electronics in Agriculture, 2025, 232: 110020.

[28] HANIF M F, VAN DAMME L G W, VANDEN HOLE C, et al. The right fit: a decision tree to select a protocol for assessing the welfare of laying hens[J]. Poultry Science, 2025, 104(12): 105935.

[29] ZHAI X, LIU Y, HONG Y, et al. Improved digital mapping of soil texture using the kernel temperature-vegetation dryness index and adaptive boosting[J]. Ecological Informatics, 2025, 87: 103083.

[30] DU Z, YANG L, ZHANG D, et al. Corn variable-rate seeding decision based on gradient boosting decision tree model[J]. Computers and Electronics in Agriculture, 2022, 198: 107025.

[31] ANDRADE-AMBRIZ Y A, PERÉZ-GARCÍA V, SERRANO-ARELLANO J, et al. Application of eXtreme Gradient Boosting in the performance prediction of a refrigeration system working with alternative refrigerants[J]. International Journal of Refrigeration, 2026, 183: 254-264.

[32] NI C, HUANG H, CUI P, et al. Light Gradient Boosting Machine (LightGBM) to forecasting data and assisting the defrosting strategy design of refrigerators[J]. International Journal of Refrigeration, 2024, 160: 182-196.

[33] YUAN Y, LIN S, DONG X, et al. Construction of a multi-layer perceptron based intelligent platform for dynamic quality monitoring and shelf-life prediction of bivalves[J]. Food Chemistry, 2025, 502: 147717.

[34] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees [J]. Nature Machine Intelligence, 2020, 2(1): 56-67.

[35] GRIMMER M K, JOHN FOULKES M, PAVELEY N D. Foliar pathogenesis and plant water relations: a review [J]. Journal of Experimental Botany, 2012, 63(12): 4321-4331.

[36] DAS S, PRATIHER S, KYAL C, et al. Sparsity regularized deep subspace clustering for multicriterion-based hyperspectral band selection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 4264-4278.

[37] GRIFFEL L M, DELPARTE D, EDWARDS J. Using support vector machines classification to differentiate spectral signatures of potato plants infected with potato virus Y[J]. Computers and Electronics in Agriculture, 2018, 153: 318-324.

[38] ARAÚJO M C U, SALDANHA T C B, GALVAO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.

（责任编辑：温晓平）

（上接第 66 页）

[12] KOTRAPPA N, HEGDE D, KAMALAKANNAN R, et al. Development of comprehensive calibration models for pulp traits by near infrared spectroscopy (NIRS) for pure species and hybrids of *Corymbia* and *Eucalyptus*[J]. Agrociencia Uruguay, 2023, 27(NE2): e1309.

[13] 陆婉珍. 现代近红外光谱分析技术(第2版)[M]. 北京：中国石化出版社. 2007.
LU W Z. Modern near infrared spectroscopy analytical technology (Second Edition)[M]. Beijing: China Petrochemical Press, 2007.

[14] 严衍禄. 近红外光谱分析基础与应用[M]. 北京：中国轻工业出版社，2005.
YAN Y L. Fundamentals and applications of near infrared spectroscopy analysis[M]. Beijing: China Light Industry Press, 2005.

[15] MESQUITA PIMENTA E, DOS SANTOS BRITO E G, RAMALHO F, et al. Validation of models using near-infrared spectroscopy to estimate basic density and chemical composition of *Eucalyptus* wood[J]. iForest-Biogeosciences and Forestry, 2024, 17(6): 338-345.

[16] ÖZDEMIR İ S, FIRAT E Ö, ÖZTURK T, et al. Geographical origin determination of the PDO hazelnut (cv. Giresun Tombul) by chemometric analysis of FT-NIR and Raman spectra acquired from shell and kernel [J]. Journal of Food Science, 2024, 89(8): 4806-4822.

[17] MIRANDA L, VAGNER TAMBARUSSI E, ABJAUD ESTOPA R, et al. Phenotyping wood properties of *Corymbia torelliana*×*Corymbia citriodora* and *Eucalyptus dunnii* based on NIR spectra [J]. Wood Material Science & Engineering, 2024, 19(1): 102-111.

[18] SCHWANNINGER M, RODRIGUES J C, FACKLER K. A review of band assignments in near infrared spectra of wood and wood components[J]. Journal of Near Infrared Spectroscopy, 2011, 19(5): 287-308.

[19] 刘旭，贾仕强，王春英，等. 玉米自交系籽粒的遗传距离与其近红外光谱距离和品种鉴别模型性能的关系研究[J]. 光谱学与光谱分析，2015，35(9)：2510-2515.
LIU X, JIA S Q, WANG C Y, et al. A Study of the relationship among genetic distances, NIR spectra distances, and NIR-based identification model performance of the seeds of maize inbred lines[J]. Spectroscopy and Spectral Analysis, 2015, 35(9): 2510-2515.

[20] 卢万鸿，杨艳，林彦，等. 粗皮桉制浆性能的近红外光谱检测[J]. 分析实验室，2022，41(4)：385-391.
LU W H, YANG Y, LIN Y, et al. Assessment of pulping properties of *Eucalyptus pellita* by near infrared spectroscopy [J]. Chinese Journal of Analysis Laboratory, 2022, 41(4): 385-391.

[21] LUAN Q F, DIAO S, SUN H G, et al. Prediction and comparisons of turpentine content in slash pine at different slope positions using near-infrared spectroscopy[J]. Plants, 2022, 11(7): 914.

（责任编辑：裴阿卫）