

面向追踪湿地松养分胁迫响应的高通量表型系统研究

张慧春^{1,2} 周子阳^{1,3} 边黎明^{4,5} 高琦⁴ 于皓¹ 郭宇明¹ 周磊^{1,2}

(1.南京林业大学机械电子工程学院,南京 210037; 2.南京林业大学林业资源高效加工利用协同创新中心,南京 210037;
3.南京地铁运营有限责任公司,南京 210000; 4.南京林业大学林草学院,南京 210037;
5.南方现代林业协同创新中心,南京 210037)

摘要: 植物表型分析是制约农林现代化发展的关键瓶颈之一。传统表型分析方法存在效率低下、操作复杂等局限,难以实现大规模、动态监测植物在环境胁迫下的生理响应。随着高通量表型技术的快速发展,多源传感器数据融合已成为研究植物健康与胁迫适应的重要手段。然而,现有系统难以应对植株高度变化不一、不同生长阶段表型特征变异大的现象,导致采集设备适应性差、作业效率有限,制约动态生理响应的精准捕捉。为此,本研究以湿地松为研究对象开展梯度养分胁迫试验(正常、轻度、重度),设计并构建了一种自走式高通量表型监测系统,该系统集成可见光、多光谱等多源成像传感器,可根据株高动态变化自动调节传感器空间位置,实现对360株样本的植物表型信息高效采集。在算法层面,系统引入了一种基于遗传算法的递归特征消除交叉验证方法(Genetic algorithm-recursive feature elimination with cross-validation, GA-RFECV),用于筛选与养分胁迫高度相关的敏感特征,并结合机器学习模型构建湿地松养分胁迫响应的分类框架。试验结果表明GA-RFECV方法提高了模型监测精度,其中随机森林(Random forest, RF)模型在验证集上的准确率、精确率、召回率和F1分数分别达到0.694、0.695、0.694、0.685。在进一步结合超参数优化后,差分进化算法(Differential evolution, DE)优化的极端梯度提升(Extreme gradient boosting, XGBoost)模型在验证集上的综合性能最优,相比于其他模型表现较好,准确率、精确率、召回率和F1分数分别提升至0.759、0.770、0.759、0.756,验证了混合特征选择与超参数优化策略在植物养分胁迫分类中的有效性。本研究提出并构建的自走式高通量表型监测系统在植物养分胁迫的精准高效追踪方面展现出较大优势,为精准施肥、抗逆品种选育、林木养分大规模监测提供了可靠的技术支撑与研究方法。

关键词: 植物表型监测平台; 养分胁迫; 超参数优化; 高通量表型; 混合特征选择

中图分类号: TP391.1 文献标识码: A 文章编号: 1000-1298(2026)01-0149-10 **OSID:** 

High-throughput Phenotyping Systems for Tracking Nutrient Stress Response of *Pinus Elliottii*

ZHANG Huichun^{1,2} ZHOU Ziyang^{1,3} BIAN Liming^{4,5} GAO Qi⁴ YU Hao¹ GUO Yuming¹ ZHOU Lei^{1,2}

(1. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

2. Co-innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University, Nanjing 210037, China

3. Nanjing Metro Operation Co., Ltd., Nanjing 210000, China

4. College of Forestry and Grassland, Nanjing Forestry University, Nanjing 210037, China

5. Co-innovation Center for Sustainable Forestry in Southern, Nanjing 210037, China)

Abstract: Plant phenotyping is one of the key bottlenecks restricting the modernization of agriculture and forestry. Traditional phenotyping methods suffer from limitations such as low efficiency and complex operation, making it difficult to achieve large-scale, dynamic monitoring of plant physiological responses under environmental stress. With the rapid development of high-throughput phenotyping technology, multi-source sensor data fusion has become an important means of studying plant health and stress adaptation. However, existing systems are unable to cope with the phenomenon of varying plant height

收稿日期: 2025-09-02 修回日期: 2025-10-16

基金项目: 农业生物育种国家科技重大专项(2023ZD0405805)、国家自然科学基金项目(32171790、32171818)、江苏省农业科技自主创新资金项目(CX(23)3126)、江苏省333高层次人才培养工程项目(20242-154)和江苏省研究生科研与实践创新计划项目(SJCX24_0365)

作者简介: 张慧春(1978—),女,教授,博士,主要从事农林信息技术与装备研究,E-mail: njzhanghc@hotmail.com

and large phenotypic variations at different growth stages, resulting in poor adaptability of data acquisition equipment and limited operational efficiency, which restricted the accurate capture of dynamic physiological responses. To address this, a gradient nutrient stress experiment (normal, mild, and severe) on slash pine was conducted, designing and constructing a self-propelled high-throughput phenotyping monitoring system. This system integrated multi-source imaging sensors such as visible light and multispectral sensors, and can automatically adjust the spatial position of the sensors according to dynamic changes in plant height, achieving efficient collection of plant phenotypic information from 360 samples. At the algorithmic level, the system introduced a genetic algorithm-recursive feature elimination with cross-validation (GA - RFECV) method to screen sensitive features highly correlated with nutrient stress, and combined this with a machine learning model to construct a classification framework for the nutrient stress response of slash pine. Experimental results showed that the GA - RFECV method improved the model's monitoring accuracy, with the random forest (RF) model achieving accuracy, precision, recall, and F1 score of 0.694, 0.695, 0.694, and 0.685 on the validation set, respectively. After further hyperparameter optimization, the extreme gradient boosting (XGBoost) model optimized by differential evolution (DE) achieved the best overall performance on the validation set, outperforming other models. Accuracy, precision, recall, and F1 score improved to 0.759, 0.770, 0.759, and 0.756, respectively, validating the effectiveness of the hybrid feature selection and hyperparameter optimization strategy in plant nutrient stress classification. The self-propelled high-throughput phenotypic monitoring system proposed and constructed demonstrated significant advantages in the accurate and efficient tracking of plant nutrient stress, providing reliable technical support and research methods for precision fertilization, stress-resistant variety breeding, and large-scale forest nutrient monitoring.

Key words: plant phenotypic monitoring platform; nutrient stress; hyperparameter optimization; high-throughput phenotyping; hybrid feature selection

0 引言

随着全球气候变化的持续加剧,林木生态系统正面临日益加剧的养分胁迫。养分胁迫不仅直接限制林木的生长与光合效率,还可能导致碳氮代谢紊乱和抗逆能力下降,进而加速树木衰退并提高病虫害易感性^[1]。长期累积的养分胁迫将进一步破坏林木群落结构和功能稳定性,削弱其在生物多样性维系、碳汇调节和水土保持等方面的生态服务功能^[2]。因此,及时监测与动态评估林木养分胁迫状态,对揭示生态响应机制、提升林木经营与精准管理水平,以及实现林木资源的可持续利用和生态系统恢复具有重要意义。

传统的植物养分诊断依赖于破坏性采样与实验室化学分析,存在时空分辨率低、操作成本高等技术缺陷,且在大规模长势监测中无法满足高效性和可持续性的需求^[3]。近年来,高通量表型组学技术的发展提供了新的解决方案,大大提高了表型获取与解析能力,给植物对各种环境胁迫的反应追踪提供了有力的手段。现有近地高通量表型系统包括固定轨道式和移动车载式^[4-5]。固定轨道系统通常依赖传送带或大型成像舱,受制于部署条件和场景灵活性,难以满足复杂受控环境下的精确监测,从而限制了植物在不同养分条件下连续、动态的生理响应过程精准捕捉^[6]。因此,开发一种能够灵活移动、远

程控制,并可根据株高变化进行高频次、多点位数据采集的“自走式车载表型平台”,成为实现动态、高通量表型分析的关键问题。

在数据分析层面,从高通量平台获取的多源传感器数据具有高维、多模态的特性,给精准评估植物健康状况带来了新的挑战,尽管机器学习算法在此领域展现出巨大潜力,但其性能严重依赖于输入特征的质量^[7]。传统特征选择方法大致可分为特征降维和特征筛选,特征降维主要有主成分分析(Principal component analysis, PCA)等,特征筛选方法主要有过滤式、包裹式和嵌入式等^[8]。然而,这些方法在应对高维表型数据时仍面临较大挑战:①高维数据中存在大量冗余特征,且光谱与形态特征间存在复杂的多重共线性和非线性关系,传统线性降维方法难以有效处理,极易遗漏对养分胁迫响应具有关键生物学意义的敏感特征。②单一的过滤式方法评估效率高但忽略模型交互,包裹式方法性能较优但计算成本高昂,嵌入式方法则与特定模型绑定,泛化性不足^[9]。尽管深度表示学习能够自动提取高层次特征,但在样本量有限且要求特征具有明确生物可解释性的林木表型研究中,其应用受到限制^[10]。为此,混合特征选择策略成为突破瓶颈的新路径,通过融合多种特征选择方法,从多维度、多角度挖掘数据中的有效判别性信息,在保留特征子集生物可解释性的同时提升模型的鲁棒性^[11]。在

模型优化层面,超参数调优是进一步提升性能的关键环节。传统的网格搜索和随机搜索计算成本高、效率有限,而贝叶斯优化、差分进化等智能搜索算法在提升模型性能方面展现出更大潜力,联合优化特征选择与超参数可实现模型精度与稳定性的双重提升^[12]。因此,本研究选择结合混合特征选择与超参数优化的机器学习方法,以在性能与可解释性之间取得平衡。同时,在实验设计中,将以常用特征选择和特征降维方法、单一机器学习模型以及未结合联合优化的传统参数调优策略作为对比基线,以验证所提出方法的有效性与优势。

湿地松(*Pinus elliottii*)作为我国亚热带地区速生丰产林的先锋树种,其生长状况与养分利用效率直接影响森林碳汇功能及生态经济效益^[13]。为解决上述两个层面的核心问题,以湿地松为研究对象,开展湿地松养分胁迫试验,提出将自走式平台的硬件创新与混合特征选择的算法创新相结合的研究思路,通过集成自走式表型平台、遗传算法-递归特征消除交叉验证(Genetic algorithm-recursive feature elimination with cross-validation, GA-RFECV)及超参数优化,构建一套植物养分胁迫响应监测系统框架,为实现林木管理的经验驱动向数据驱动转型提供新的技术路径与方法支撑。

1 材料与方法

1.1 研究地概况

研究地点位于浙江省杭州市余杭区长乐林场(30.36°N,119.84°E),该林场是国家湿地松良种基地和国家级重点林木良种繁育基地。该地区属于亚热带季风气候区,年平均气温 16~18℃,年均降水量 1 400~1 600 mm,相对湿度均值(82±5)%,满足湿地松最适生长水热需求,林场土壤以典型红壤为主,土层深厚,透气性良好,具有较强的持水能力,地形起伏度小于 3%。

1.2 试验设计

选用国家林木种质资源库认证的 30 个湿地松(*Pinus elliottii*)半同胞家系(编号 I1-I30),2024 年 4 月,将苗木移栽至 25 cm×26 cm 塑料容器,每盆装填约 7.5 kg 基质,该基质由 90% 黄心土(来源于林场黄心土,pH 值 5.5~6.5,全氮质量比 0.76 g/kg,全磷质量比 0.08 g/kg,有效磷质量比 1.24 mg/kg)与 10% 珍珠岩混合,以改善通气与排水性能。为系统研究湿地松对养分胁迫的响应,本研究以磷胁迫作为养分胁迫的主要实现方式,通过设置不同水平的磷供应来模拟不同程度的养分胁迫环境。试验采用三水平磷胁迫梯度设计,参考现有磷胁迫研究的

处理方式^[14-15],在保持其它营养元素正常供应的条件下,通过添加不同浓度的磷酸二氢钾溶液构建磷胁迫梯度,包括:重度养分胁迫组(Severe nutrient stress group,SD,0 mmol/L KH₂PO₄),轻度养分胁迫组(Mild nutrient stress group,MD,0.1 mmol/L KH₂PO₄),正常生长组(Normal growth group,NG,1 mmol/L KH₂PO₄)。磷胁迫试验流程如图 1 所示。

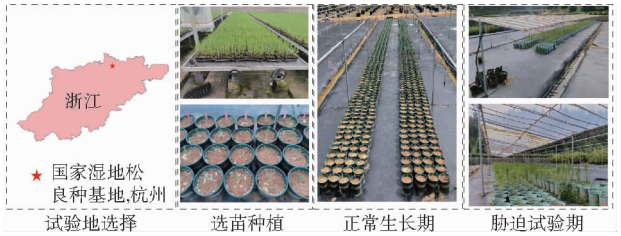


图 1 养分胁迫试验过程

Fig. 1 Nutrient stress test process

各组均同步施用缺磷的改良 Hoagland 全营养液(海博生物,HB8870-3),以确保除磷外必需元素的正常供应,钾离子差异通过 KCl 溶液进行配平。每个家系在每个处理下设 4 次重复,采用完全随机区组设计,共 30 家系×3 处理×4 重复=360 株样本。移栽后设置 52 d 缓苗期,自 2024 年 5 月下旬起架设开合式遮雨棚,并开始实施磷胁迫处理,持续至 9 月中旬,共计约 105 d。期间每 30 d 定量施加对应浓度营养液 1 L/盆(精度±50 mL),采用蠕动泵控制供给量。为避免稀释效应,磷营养液施加与日常灌溉分离进行。灌溉管理由长乐林场配套的物联网智控系统自动调节,维持基质含水量在 50%~80% 之间,以排除水分因子对磷梯度处理的干扰。

1.3 基于自走式表型平台的多源图像采集

利用构建的自走式植物表型平台获取湿地松的多光谱图像和可见光图像,自走式植物表型平台由自走式底盘、高度调节模块和成像传感器组成,其中,自走式底盘可通过远程无线遥控控制平台的前进、后退、转弯等功能,高度调节模块可通过电机驱动实现成像传感器相对植物位置的调整,包括上下高度调节和平仰角度调节,以获取最佳视角的冠层图像,可适应不同植物种植差异和同种植物不同时期生长差异。平台搭载了 RedEdge-MX 多光谱相机(MicaSense, Seattle, 美国)和 D435i 深度相机(Intel, Santa Clara, 美国),在采集过程中,首先将自走式植物表型平台移动至目标植株位置,随后通过高度调节模块将传感器调节至植株冠层上方约 1 m 处,依次完成多光谱图像和可见光图像的采集,多光谱图像通过 Wi-Fi 连接多光谱相机实时采集,包含蓝(B,(475±32) nm)、绿(G,(560±27) nm)、红

(R, (668 ± 16) nm)、近红外 (NIR, (842 ± 57) nm) 和红边 (RE, (717 ± 12) nm) 5 个通道, 图像分辨率为 1 280 像素 × 960 像素, 可见光图像通过 D435i 深度相机实时采集, 分辨率为 1 920 像素 × 1 080 像素, 所有图像采集完成后立即存储于平台终端。为保证反射率定标精度, 在每次采集前后均拍摄标准反射率校正灰度板, 并基于校正板的已知反射率对原始数值进行辐射定标, 将其转换为相对反射率值, 以减少光照条件和传感器响应差异对数据精度的影响, 湿地松图像采集时间为胁迫处理后第 105 天 (2024 年 9 月 14 日), 为确保数据的准确性与一致性, 采集工作统一安排在日常 10:00—13:00 的太阳辐射相对稳定时段进行, 以减小云影和太阳高度变化等因素对数据的影响。本研究共设置 3 个养分胁迫处理水平, 每个处理包含 120 株湿地松, 每株植株在采集时获取 1 幅多光谱图像和 1 幅可见光图像, 最终获得 1 800 幅多光谱图像和 360 幅可见光图像。自走式植物表型平台采集的图像如图 2 所示。

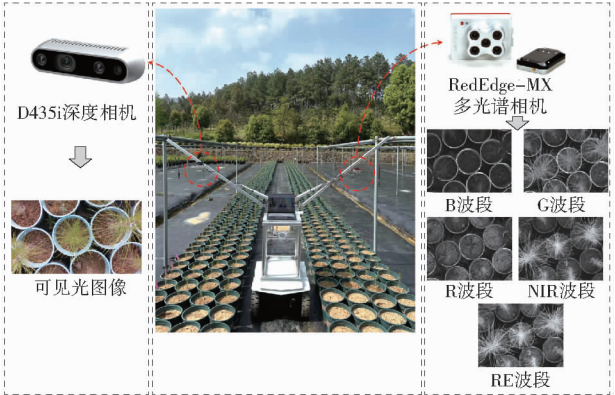


图 2 基于自走式表型平台的多源图像采集
Fig.2 Multi-source image acquisition based on self-propelled phenotyping platform

1.4 混合特征选择与传统特征选择方法

特征选择是机器学习和数据分析中的关键步骤, 其核心目标是从原始数据中选择出对模型预测最具价值的特征, 以降低模型复杂度并提升性能, 传统的特征选择方法主要分为特征筛选和特征降维两类^[16]。其中, 特征选择方法旨在直接从原始特征集中选取出最重要的特征, 以减少冗余, 提高模型的泛化能力, 递归特征消除交叉验证 (RFECV) 是一种常用方法, 通过基于随机森林基础模型的特征重要性分数递归移除贡献度最低的特征, 每次移除固定数量的特征 (步长为 1), 并在每一步结合 3 折交叉验证准确率评估特征子集的性能, 以确定最优特征子集, 同时防止过拟合^[17]。与特征选择不同, 特征降维方法通过映射或转换原始特征空间来降低数据维度, 在保留关键信息的同时去除冗余或无关特征,

PCA 是其中一种经典的特征降维技术, 其原始数据投影到一组新的正交特征 (即主成分) 上来减少维度, 优先保留方差较大的主成分以最大程度地保持数据主要信息, 根据现有研究, 本文 PCA 方法在降维过程中选择累计方差贡献率超过 95% 的主成分数, 从而保留数据中大部分信息^[18]。

然而, 随着数据维度持续攀升, 单一的特征选择方法难以充分处理高维数据中的冗余特征和复杂的非线性关系^[19]。为克服传统特征筛选方法的局限性, 本文提出了一种混合特征选择方法, 即 GA - RFECV。该方法的创新点在于将 GA 的全局搜索能力与 RFECV 的局部精炼相结合: GA 通过模拟自然选择过程高效探索高维特征空间^[20], 捕捉非线性特征关系, 避免 PCA 的线性假设或 RFECV 贪婪策略导致的局部最优; RFECV 则对 GA 生成的候选特征子集进行迭代消除, 进一步去除冗余特征, 提升特征选择的准确性和模型泛化能力。该方法的实现流程如下: 首先, GA 模拟自然选择过程搜索最优特征子集, 其步骤包括: 随机生成多个特征子集, 评估每个个体在目标任务上的性能, 在迭代进化过程中, 优选性能卓越的个体进行交叉和变异操作以生成新的特征子集。随后, 对 GA 生成的每个特征子集, 采用 RFECV 进行精简, 并通过交叉验证评估其性能。最终, 该混合策略能够实现特征选择的全局优化。

为验证 GA - RFECV 的有效性, 本文设计了消融实验和对比实验。消融实验: 为评估 GA 和 RFECV 在混合策略中的贡献, 在相同数据集上分别运行以下变体: 仅 RFECV (无 GA 全局搜索); 完整 GA - RFECV。对比实验: 为与传统方法比较, 选择了以下基线算法: PCA (特征降维); 纯 RFECV (特征选择)。所有算法使用相同机器学习模型和默认超参数, 以确保公平性。

1.5 超参数优化与机器学习建模算法

合适的机器学习建模算法选择决定了最终结果的准确性和稳定性。本研究采用了多种先进的机器学习算法, 包括随机森林 (Random forest, RF)、极端梯度提升 (Extreme gradient boosting, XGBoost)、梯度提升决策树 (Gradient boosting decision tree, GBDT)、轻量级梯度提升机 (Light gradient boosting machine, LightGBM) 以及极端随机树 (Extreme randomized trees, ERT), 为有效训练和评估这些算法, 数据集按比例 7:3 划分为训练集和验证集, 确保模型在湿地松养分胁迫响应建模中的性能得到可靠验证。

超参数优化是提升机器学习模型性能的关键步骤, 其选择直接影响模型的预测准确性和泛化能力。本研究采用贝叶斯优化 (Bayesian optimization, BO)、

模拟退火算法 (Simulated annealing, SA) 和差分进化算法 (Differential evolution, DE) 对模型超参数进行调优。在超参数优化过程中, 本文统一以分类准确率 (Accuracy) 作为优化目标函数, 并在训练集上采用 5 折交叉验证计算模型性能, 以避免因数据划分带来的偶然性和信息泄露。所有模型的最终性能均在独立验证集上评估。对于 BO, 最大迭代次数设为 50; 对于 SA, 最大迭代次数同样设为 50, 并在初始温度下降过程中, 若连续若干迭代未提升准确率则提前终止; 对于 DE, 种群规模设置为 10, 最大迭代次数为 50。3 类优化器的搜索空间均涵盖模型的主要超参数。

1.6 评估指标

为了系统全面评估所提出模型的性能, 本研究采用了 4 项常用评估指标: 准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 分数 (F1 score)。具体来说, 准确率能够快速反映模型在整体样本上的正确分类能力, 提供总体全局性能的直观度量; 精确率量化被预测为正例的样本中真正正例的比例, 反映模型判别的准确性; 召回率度量真正正例被正确识别的覆盖率, 体现模型对关键目标的捕获能力; 而 F1 分数则综合了精确率与召回率的调和均值, 是评价模型在平衡误报与漏报方面效果的重要指标。上述指标通过多维度量化模型性能, 共同构建对模型分类器表现的全面评价体系。

2 数据预处理与特征提取

2.1 可见光图像处理

可见光图像主要用于提取湿地松的颜色指数和纹理特征。由于湿地松的针叶在冠层区域高度集中, 而枝条尾部纤细分散且比例较小, 增加了图像分割的难度, 因此本研究首先对可见光图像进行预处理, 选定冠层区域作为感兴趣区域 (Region of interest, ROI), 以排除尾部干扰。随后, 采用改进的超绿指数 (Excess green index, ExG) 对冠层图像进行分割, 生成二值化图像。其中改进超绿指数是通过增大绿色通道的权重系数, 减小红色和蓝色通道的权重系数^[21], 从而优化分割效果, 阈值通过大津法 (Nobuyuki otsu method, OTSU) 自动确定。为了提取颜色特征, 将冠层图像拆分为蓝 (Blue, B)、绿 (Green, G)、红 (Red, R) 3 个通道, 并利用上述二值化图像分别对各通道进行掩膜处理, 获得对应的掩膜图像。进一步计算各掩膜图像中非零像素的数量及像素颜色总值, 从而提取标准化的颜色分量 b 、 g 、 r 。整个颜色特征提取过程如图 3 所示。纹理特征提取基于灰度共生矩阵 (Gray level co-occurrence

matrix, GLCM)。首先, 对冠层图像进行掩膜处理并转换为灰度化图像, 灰度级设为 256, 确保纹理信息的准确提取。GLCM 分析采用 3×3 滑动窗口, 步长为 1, 以精确捕捉局部纹理特征, 为降低方向依赖性并增强特征的稳定性和鲁棒性, 分别计算 0° 、 45° 、 90° 和 135° 4 个方向的 GLCM 特征, 并取其平均值作为最终纹理特征。所有可见光图像的处理与分析均在 Python 环境中完成 (Python 3.10, scikit-image 0.20.0, opencv-python 4.5.5), 以确保数据处理流程的可复现性。

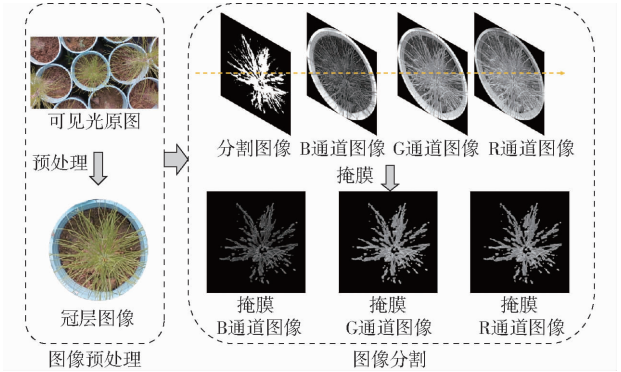


图 3 湿地松可见光图像处理过程

Fig. 3 Process of visible light image processing of *Pinus elliotii*

2.2 多光谱图像处理

多光谱图像主要用于提取湿地松的植被指数信息。由于多光谱图像各波段间存在相位差, 预处理阶段首先使用 Fiji 软件 (2.16.0 版本) 的 Register Virtual Stack Slices 插件进行配准, 采用 Rigid 模式, 配准后的 5 个波段图像将用于后续的图像分割处理。在湿地松的 5 个波段中, NIR 波段呈现最高亮度, B 波段亮度最低。基于此特性, 本研究使用大津分割算法 OTSU 分割 NIR 波段图像, 以获得湿地松二值化图像, OTSU 算法通过 OpenCV 库 (4.5.5 版本) 实现, 自动计算阈值。随后, 利用湿地松二值化图像对配准后的 5 个波段图像进行掩膜处理, 最终得到 5 个波段的湿地松前景掩膜图像。在湿地松前景掩膜图像上框选 ROI, 结合图像中包含的反射率校正灰度板 (Calibrated reflectance panel, CRP, MicaSense 标定板) 和下行光照传感器的记录数据, 对各波段图像进行反射率转换。具体流程包括: ①利用 CRP 在各波段的已知反射率值, 建立像素值 (Digital number, DN) 与实际反射率之间的线性关系, 获得波段校正系数。②结合下行光照传感器记录的入射光强, 进一步对各波段图像进行光照条件校正, 从而消除环境光照变化带来的影响。该流程参考了 MicaSense 官方数据处理指南 (<https://micasense.github.io/imageprocessing/index.html>)。计

算 ROI 内各波段掩膜图像中所有非零像素点对应的反射率值,从而获得湿地松在 B、G、R、NIR 和 RE 波段的反射率信息。反射率转换和光照校正使用 Python 脚本 (3.10 版本) 实现,依赖 numpy 和 Opencv-Python 库,具体脚本基于 MicaSense 官方提供的图像处理工作流程。多光谱图像处理流程如图 4 所示。

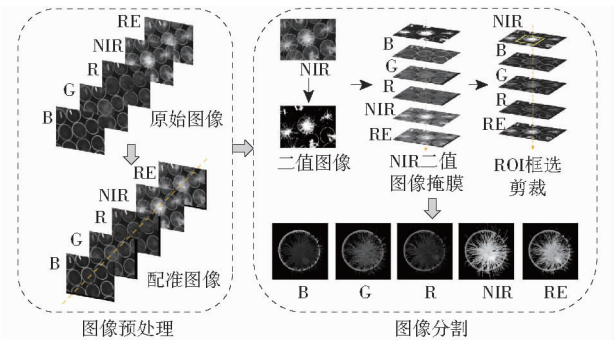


图 4 湿地松多光谱图像处理过程

Fig. 4 Multispectral image processing process of slash pine

2.3 表型特征参数提取

本研究基于可见光和多光谱图像提取了可见光纹理特征、可见光颜色指数和多光谱植被指数。可见光纹理特征通过 GLCM 在 0°、45°、90°、135° 4 个方向计算纹理特征,并取各方向均值作为最终特征向量。具体包括 8 种纹理特征:均值 (Mean, MEA)、方差 (Variance, VAR)、均匀性 (Homogeneity, HOM)、对比度 (Contrast, CON)、相异性 (Dissimilarity, DIS)、熵 (Entropy, ENT)、能量 (Energy, ENE) 和相关性 (Correlation, COR)。

由于不同指数对植被特性的表征具有互补性,为克服单一指数表征能力不足的局限,本研究在系统梳理文献的基础上,精选了 31 项多光谱植被指数,分别为:归一化植被指数 (Normalized difference vegetation index, NDVI)^[22]、绿色归一化植被指数 (Green normalized difference vegetation index, GNDVI)^[22]、增强型植被指数 (Enhanced vegetation index, EVI)^[22]、归一化红边指数 (Normalized difference red edge, NDRE)^[22]、比值植被指数 (Ratio vegetation index, RVI)^[23]、差值植被指数 (Different influential factors, DVI)^[23]、三角植被指数 (Triangle vegetation index, TVI)^[23]、改进简单比值植被指数 (Modified simple ratio index, MSR)^[23]、红边叶绿素指数 (Red-edge chlorophyll index, RECI)^[23]、绿色叶绿素指数 (Green chlorophyll index, GCI)^[23]、优化土壤调节植被指数 (Optimized soil-adjusted vegetation index, OSAVI)^[23]、改进土壤调节植被指数 (Modified soil-adjusted vegetation index, MSAVI)^[23]、改进叶绿素吸收比指数 (Modified chlorophyll absorption ratio

index, MCARI)^[23]、归一化色素叶绿素比率指数 (Normalized pigment chlorophyll ratio index, NPCI)^[23]、结构不敏感色素指数 (Structure insensitive pigment index, SIPI)^[23]、氮反射指数 (Nitrogen reflectance index, NRI)^[23]、单红边波段叶绿素指数 (Single red-edge band chlorophyll index, SRCI)^[23]、绿色优化土壤调节植被指数 (Green Optimized Soil Adjusted Vegetation index, GOSAVI)^[23]、植物衰老反射指数 (Plant senescence reflectance index, PSRI)^[23]、调整型转换土壤调节植被指数 (Adjusted transformed soil-adjusted vegetable index, ATSA)^[24]、冠层叶绿素含量指数 (Canopy chlorophyll content index, CCCI)^[24]、绿色大气抗扰植被指数 (Green atmospherically resistant index, GARI)^[24]、绿色差值植被指数 (Green difference vegetable index, GDVI)^[24]、绿色叶片指数 (Green leaf index, GLI)^[24]、改进绿红植被指数 (Modified green-red vegetable index, MGR)^[24]、归一化绿-红差值指数 (Normalized green-red difference index, NGRD)^[24]、再归一化植被指数 (Renormalized difference vegetable index, RDVI)^[24]、红-绿-蓝植被指数 (Red green-blue vegetable index, RGB)^[24]、红/绿简单比值植被指数 (Simple ratio red/green, RGR)^[24]、改进型三角植被指数 (Modified triangle vegetation index, MTVI)^[24]、宽动态范围植被指数 (Wide dynamic range vegetable index, WDR)^[24]。

30 项可见光颜色指数分别为:归一化红光指数 (Normalized red index, NR)^[25]、归一化绿光指数 (Normalized green index, NG)^[25]、归一化蓝光指数 (Normalized blue index, NB)^[25]、绿蓝比值指数 (Green-blue ratio, GB)^[25]、红蓝比值指数 (Red-blue ratio, RB)^[25]、绿红比值指数 (Green-red ratio, GR)^[25]、红减蓝指数 (Red minus blue, RminB)^[25]、红加蓝指数 (Red plus blue, RplusB)^[25]、红减绿指数 (Red minus green, RminG)^[25]、红加绿指数 (Red plus green, RplusG)^[25]、绿减蓝指数 (Green minus blue, GminB)^[25]、蓝加绿指数 (Blue plus green, BplusG)^[25]、灰度指数 (Grayness index, GRAY)^[25]、绿叶面积指数 (Green leaf area index, GLAI)^[25]、光谱多边形面积 (Spectral histogram polygon, SHP)^[25]、归一化差值指数 (Normalized difference index, NDI)^[26]、可见光大气阻抗植被指数 (Visible atmospherically resistant index, VARI)^[26]、韦伯指数 (Weber index, WI)^[26]、彩色植被指数 (Color index of vegetation extraction, CIVE)^[26]、绿叶算法 (Green leaf algorithm, GLA)^[26]、河岸植被指数 (Kawashima

index, IKAW)^[27]、改良绿红植被指数 (Modified green red vegetation index, MGRVI)^[27]、红绿蓝植被指数 (Red green blue vegetation index, RGBVI)^[27]、超绿指数 (Excess green, ExG)^[27]、超红指数 (Excess red, ExR)^[27]、超绿超红差值指数 (Excess green-excess red difference, ExGR)^[27]、绿叶指数 (Green leaf index, GLI)^[27]、红光指数 (Redness index, RI)^[28]、归一化绿蓝差值指数 (Normalized green-blue difference index, NGBDI)^[28]、归一化绿红差值指数 (Normalized green-red difference index, NGRDI)^[28], 构建多维特征空间以全面表征植被表型信息。

可见光颜色指数通过前述可见光图像处理获取的 *b*、*g*、*r* 三通道颜色分量, 通过特征组合计算生成, 如超绿指数 $ExG = 2g - b - r$ 等。多光谱植被指数则基于多光谱图像处理获取的 5 个波段 (*B*、*G*、*R*、*NIR*、*RE*) 的 DN 值转换的反射率, 经组合运算构建植被指数, 如归一化植被指数 $NDVI = (NIR - R) / (NIR + R)$, 后文中多光谱植被指数用 MS - * 代指, * 代表多光谱植被指数名称, 可见光植被指数用

VIS - * 代指, * 代指可见光颜色指数和纹理特征。

3 结果与分析

3.1 混合特征选择与传统特征选择算法消融实验结果分析

为获得对模型预测性能影响较大的最优特征子集, 本研究采用了混合特征选择算法 (GA - RFECV) 与两种传统特征选择算法 (PCA 和 RFECV) 对原始数据进行降维和特征筛选。对 3 种特征选择方法 (PCA、RFECV 和 GA - RFECV) 的最终筛选结果和模型性能进行了综合比较与分析。表 1 列出了各方法最终选择的特征变量, 图 5 则直观展示了不同方法在特征选择过程中的结果对比。图 5c 中交叉验证评分指标为准确率。结合表 1 和图 5 可以看出, 特征降维算法 PCA 通过计算各主成分的方差贡献率, 最终将原始特征空间压缩筛选出 9 个主成分, 累计方差贡献率大于 95%, 特征筛选算法 RFECV 最终筛选出 41 个最优特征, 获得的最高交叉验证评分为 60.3%, 混合筛选算法 GA - RFECV

表 1 PCA、RFECV、GA - RFECV 最终用于建模的特征
Tab.1 PCA, RFECV, and GA - RFECV ultimately used as features for modeling

特征选择算法	特征个数	最终选择的特征
PCA	9	PC1、PC2、PC3、PC4、PC5、PC6、PC7、PC8、PC9 MS - B、MS - G、MS - R、MS - GNDVI、MS - TVI、MS - GCI、MS - NRI、MS - GOSAVI、MS - GARI、MS - GDVI、MS - GLI、MS - MGR、MS - NGRD、MS - RGB、MS - RGR、MS - MTVI、VIS - b、VIS - g、VIS - r、VIS - NG、VIS - NB、VIS - RminB、VIS - RplusB、VIS - RplusG、VIS - GminB、VIS - BplusG、VIS - RGBVI、VIS - NDI、VIS - GLA、VIS - ExR、VIS - GLI、VIS - GRAY、VIS - SHP、VIS - RI、VIS - mean、VIS - variance、VIS - homogeneity、VIS - contrast、VIS - dissimilarity、VIS - entropy、VIS - energy
RFECV	41	
GA - RFECV	14	MS - G、MS - R、MS - NIR、MS - NDVI、MS - DVI、MS - TVI、MS - RECI、MS - GLI、MS - RGB、VIS - NB、VIS - BplusG、VIS - NDI、VIS - GRAY、VIS - variance

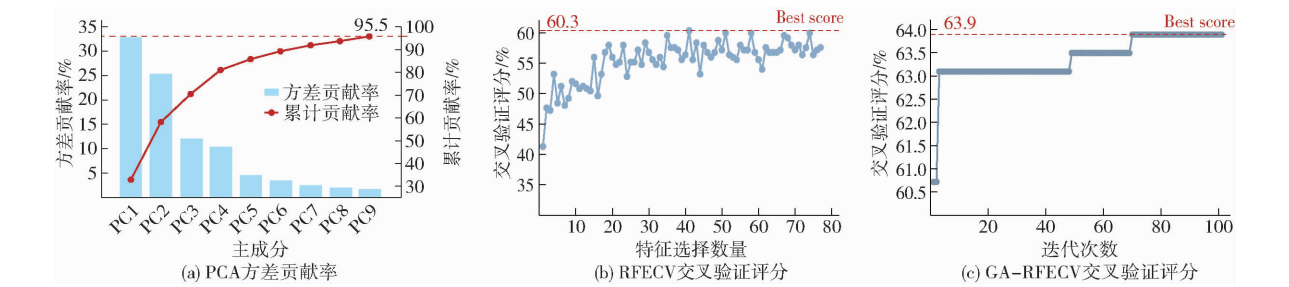


图 5 PCA、RFECV、GA - RFECV 特征选择结果
Fig. 5 Feature selection results of PCA, RFECV and GA - RFECV

结合了遗传算法的全局搜索能力与 RFECV 的监督筛选策略, 经过多次迭代, GA - RFECV 最终筛选出 14 个特征, 其迭代过程中的最高交叉验证评分达到 63.9%, 高于 RFECV 的 60.3%, 表明相比于传统的特征筛选算法 RFECV, 遗传算法的引入优化了特征子集, 在特征选择过程中引入遗传算法的全局搜索能够有效跳出局部最优, 从而获得更高的预测精度,

使其更契合湿地松养分状态评估任务的需求。

3.2 基于不同特征选择算法的湿地松养分胁迫响应模型对比实验结果分析

采用 3 种特征选择方法——PCA、RFECV、GA - RFECV 筛选最优特征子集。具体筛选结果为: PCA 获得 9 个最优特征, RFECV 获得 41 个最优特征, GA - RFECV 获得 14 个最优特征。基于筛选出的

这些最优特征子集,分别应用 RF、XGBoost、GBDT、LightGBM 和 ERT 共 5 种机器学习算法构建湿地松养分胁迫响应模型,验证集的评估结果见表 2。

表 2 PCA、RFECV、GA – RFECV 下湿地松养分胁迫响应模型结果

Tab.2 Results of nutrient stress response model of <i>Pinus elliottii</i> under PCA, RFECV and GA – RFECV					
特征选择 算法	机器学习 算法	准确率	精确率	召回率	F1 分数
PCA	RF	0.611	0.615	0.612	0.610
	XGBoost	0.602	0.601	0.602	0.600
	GBDT	0.583	0.585	0.584	0.584
	LightGBM	0.639	0.643	0.639	0.632
	ERT	0.602	0.604	0.604	0.599
RFECV	RF	0.639	0.636	0.640	0.633
	XGBoost	0.648	0.643	0.649	0.644
	GBDT	0.630	0.624	0.631	0.618
	LightGBM	0.657	0.653	0.658	0.652
	ERT	0.630	0.629	0.630	0.626
GA – RFECV	RF	0.694	0.695	0.694	0.685
	XGBoost	0.657	0.655	0.658	0.655
	GBDT	0.685	0.690	0.685	0.682
	LightGBM	0.648	0.654	0.649	0.644
	ERT	0.676	0.675	0.677	0.670

从表 2 可以看出,在默认参数设置下,不同特征选择方法下各机器学习模型的表现存在差异。具体来说,使用 PCA 特征时,LightGBM 模型在验证集上的表现相对较优,其准确率、精确率、召回率和 F1 分数分别为 0.639、0.643、0.639、0.632;使用 RFECV 特征时,LightGBM 模型表现相对较优,准确率、精确率、召回率和 F1 分数分别为 0.657、0.653、0.658、0.652;使用 GA – RFECV 特征时,RF 模型表现相对较优,准确率、精确率、召回率和 F1 分数分别为 0.694、0.695、0.694、0.685。与 PCA 相比,GA – RFECV 在验证集上的准确率、精确率、召回率和 F1 分数分别绝对提升了 5.5%、5.2%、5.5%、5.3%;与 RFECV 相比,分别绝对提升了 3.7%、4.2%、3.6%、3.3%。这表明,在湿地松养分胁迫响应中,GA – RFECV 混合特征选择方法能够更有效地捕捉与湿地松磷胁迫程度相关的关键信息,从而提高模型区分重度(SD)、轻度(MD)和正常(NG)3 类生长状态的准确性。

3.3 混合特征选择算法下基于超参数优化的湿地松养分胁迫响应分析

为进一步提升湿地松养分胁迫响应模型的性能,本研究在 GA – RFECV 特征选择的基础上,结合交叉验证,采用 BO、SA、DE 对 RF、XGBoost、ERT、GBDT、LightGBM 5 种机器学习模型的关键超参数

进行优化,旨在发现更优的超参数组合以提升模型 的分类精度。表 3 展示了超参数优化后各模型在验证集上的性能表现。从表 3 可见,相较于其他模型,采用 DE 算法优化的 XGBoost 模型表现相对更优,其准确率、精确率、召回率和 F1 分数分别达到 0.759、0.770、0.759、0.756。BO 和 SA 优化的 XGBoost 模型也表现出较高的精度,但略低于 DE – XGBoost。

表 3 BO、SA、DE 超参数优化下的湿地松养分胁迫响应模型结果

Tab.3 Results of <i>Pinus elliottii</i> nutrient stress response model under BO, SA, and DE hyperparameter optimization					
超参数优化 算法	机器学习 算法	准确率	精确率	召回率	F1 分数
BO	RF	0.639	0.634	0.639	0.625
	XGBoost	0.685	0.689	0.685	0.682
	GBDT	0.639	0.632	0.639	0.633
	LightGBM	0.676	0.681	0.676	0.667
	ERT	0.648	0.647	0.648	0.640
SA	RF	0.667	0.666	0.667	0.659
	XGBoost	0.685	0.680	0.685	0.678
	GBDT	0.648	0.648	0.648	0.642
	LightGBM	0.648	0.650	0.648	0.640
	ERT	0.620	0.614	0.620	0.608
DE	RF	0.667	0.665	0.667	0.669
	XGBoost	0.759	0.770	0.759	0.756
	GBDT	0.731	0.740	0.731	0.726
	LightGBM	0.713	0.714	0.713	0.709
	ERT	0.694	0.707	0.694	0.685

为深入分析各优化算法下最优模型 的分类性能,本研究展示了这些模型在验证集上的混淆矩阵和分类精度对比雷达图(图 6,NG 为正常生长组,SD 为重度养分胁迫组,MD 为轻度养分胁迫组)。图 6a ~ 6c 为基于气泡形式的验证集混淆矩阵,其中行表示真实类别,列表示预测类别。气泡内的数字代表该类别中的样本数量,蓝色气泡表示被正确分类的样本,红色气泡表示被错误分类的样本,气泡的面积与样本数成正比,面积越大表明对应类别的样本数越多,右侧的图例通过嵌套的透明气泡展示了样本数量的相对大小关系,用于辅助说明气泡面积的对比尺度,整体上,蓝色气泡面积越大,说明模型在该类别上的预测准确性越高。图 6d 所示雷达图中形状覆盖面积较大的多边形通常代表整体性能更优的模型。从混淆矩阵气泡图中可以看出,BO – XGBoost 算法在 NG 与 MD 的识别上保持了较高的准确性,但在 MD 与 SD 之间存在混淆,表明其在处理类别边界模糊问题时存在不足,SA – XGBoost 算法在 NG 的判别上表现突出,但在 MD 和 SD 上误判

较多,尤其 SD 样本经常被错误归入其他类别,导致类别间识别不均衡,相比之下,DE - XGBoost 算法的整体表现更为优越,其在 3 类上的正确分类数量均高于 BO 与 SA,尤其在 MD 的识别中展现出极高的准确性和稳定性,仅有极少量误判,显示出较强的泛

化能力,而从图 6d 雷达图中可以明显看出,相较于 SA - XGBoost 和 BO - XGBoost,DE - XGBoost 在各项分类评价指标均展现出更优越的综合性能,总体而言,DE - XGBoost 算法在湿地松养分胁迫响应中的表现最优。

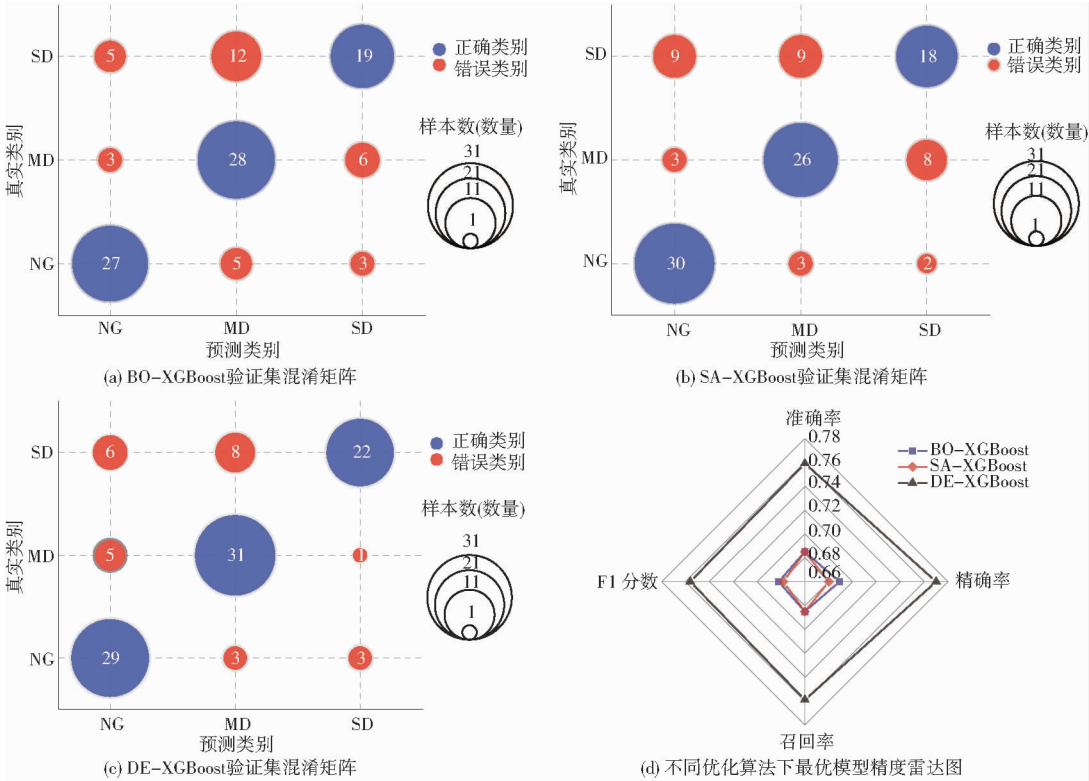


图 6 不同优化算法下最优模型验证集混淆矩阵及精度对比

Fig. 6 Confusion matrix and accuracy comparison of optimal model validation set under different optimization algorithms

4 结论

(1)构建了用于追踪湿地松养分胁迫响应的自走式表型平台,并提出 GA - RFECV 混合特征选择方法,实现了关键表型特征的高效筛选,为养分状态评估提供了可靠数据基础。

(2)在 GA - RFECV 特征支持下,构建的机器学习模型表现优异,其中 RF 模型准确率、精确率、召

回率和 F1 分数分别达到了 0.694、0.695、0.694、0.685,提升了模型的分类精度、稳定性和泛化能力。

(3)进一步结合 BO、SA 和 DE 对模型超参数进行优化,DE - XGBoost 模型性能最佳,其准确率、精确率、召回率和 F1 分数分别提升至 0.759、0.770、0.759、0.756,表明了特征选择与超参数优化联合作用在提升湿地松养分状态评估精度方面的有效性。

参 考 文 献

[1] WONG M Y, WURZBURGER N, HALL J S, et al. Trees adjust nutrient acquisition strategies across tropical forest secondary succession[J]. New Phytologist, 2024,243(1):132 - 144.

[2] YANG J P, JIANG J, FU Z P, et al. Integrating phenology information with UAV multispectral data for rice nitrogen nutrition diagnosis[J]. European Journal of Agronomy, 2025,169:127696.

[3] 陈耀晖,李家一,鲍泽韩,等.自主导航柑橘表型巡检机器人设计与试验[J].农业机械学报,2025,56(3):49 - 57.

CHEN Yaohui, LI Jiayi, BAO Zehan, et al. Design and experimentation of autonomous navigated citrus phenotype inspection robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 49 - 57. (in Chinese)

[4] 宋鹏,李正达,杨蒙,等.作物表型机器人研究现状与展望[J].农业机械学报,2025,56(3):1 - 17.

SONG Peng, LI Zhengda, YANG Meng, et al. Current status and prospects of crop phenotyping robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 1 - 17. (in Chinese)

[5] 张慧春,周宏平,郑加强,等.植物表型平台与图像分析技术研究进展与展望[J].农业机械学报,2020,51(3):1 - 17.

ZHANG Huichun, ZHOU Hongping, ZHENG Jiaqiang, et al. Research progress and prospect in plant phenotyping platform and image analysis technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(3): 1 - 17. (in

Chinese)

[6] 李道亮, 赵晔, 杜壮壮. 农业领域多模态融合技术方法与应用研究进展[J]. 农业机械学报, 2025, 56(1): 1–15.
LI Daoliang, ZHAO Ye, DU Zhuangzhuang. Advances in multi-modal fusion techniques and applications in agricultural field [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(1): 1–15. (in Chinese)

[7] LIU Q, WANG C, JIANG J, et al. Multi-source data fusion improved the potential of proximal fluorescence sensors in predicting nitrogen nutrition status across winter wheat growth stages[J]. Computers and Electronics in Agriculture, 2024, 219: 108786.

[8] 陈挺, 刘香君, 臧璇, 等. 深度特征选择方法研究综述[J]. 计算机应用与软件, 2025, 42(7): 1–11.
CHEN Ting, LIU Xiangjun, ZANG Xuan, et al. Review research on deep feature selection methods[J]. Computer Applications and Software, 2025, 42(7): 1–11. (in Chinese)

[9] DUAN X, WU X, GE J, et al. A novel hierarchical clustering sequential forward feature selection method for paddy rice agriculture mapping based on time-series images[J]. Agriculture, 2024, 14(9): 1468.

[10] ZHANG H, SU H, SHEN T, et al. Integrating PROSPECT–D physics and adversarial domain adaptation resnet for robust cross-ecosystem plant traits estimation[J]. Frontiers in Plant Science, 2025, 16: 1612430.

[11] TIWARI A, CHATURVEDI A. A hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification[J]. Expert Systems with Applications, 2022, 196: 116621.

[12] AILEMOBAYO J, DURODOLA O, ALADE O, et al. Hyperparameter tuning in machine learning: a comprehensive review [J]. Journal of Engineering Research and Reports, 2024, 26(6): 388–395.

[13] 吴亚获, 刁姝, 丁显印, 等. 基于 51K 液相探针构建国外松良种 DNA 指纹图谱[J]. 林业科学, 2025, 61(6): 109–119.
WU Yadi, DIAO Shu, DING Xianyin, et al. Construction of a DNA fingerprinting of elite varieties of introduced exotic pines in china based on 51K liquid-phased probes[J]. Scientia Silvae Sinicae, 2025, 61(6): 109–119. (in Chinese)

[14] LINGER E, ASEFA M, ZHANG T, et al. Phosphorus stress disrupts leaf-root trait relationships and reduces biomass in a tropical cloud forest: evidence from a two-year phosphorus addition experiment[J]. Forest Ecology and Management, 2025, 586: 122726.

[15] JIANG S, TANG Y, FAN R, et al. Response of *Carex breviculmis* to phosphorus deficiency and drought stress[J]. Frontiers in Plant Science, 2023, 14: 1203924.

[16] THENG D, BHOYAR K K. Feature selection techniques for machine learning: a survey of more than two decades of research [J]. Knowledge and Information Systems, 2024, 66(3): 1575–1637.

[17] 杨伟博, 李映雪, 吴芳, 等. 联合多种衍生光谱特征的冬小麦叶绿素含量估算[J]. 农业工程学报, 2025, 41(15): 165–173.
YANG WeiBo, LI Yingxue, WU Fang, et al. Estimating chlorophyll content of winter wheat using multiple derivative canopy spectra[J]. Transactions of the CSAE, 2025, 41(15): 165–173. (in Chinese)

[18] 漆海霞, 黄荟良, 罗锡文, 等. 基于多项式回归和堆叠模型的花生产量预测[J]. 农业工程学报, 2025, 41(8): 165–174.
QI Haixia, HUANG Huiliang, LUO Xiwen, et al. Peanut yield prediction based on polynomial regression and stacked model [J]. Transactions of the CSAE, 2025, 41(8): 165–174. (in Chinese)

[19] 苏红军. 高光谱遥感影像降维: 进展、挑战与展望[J]. 遥感学报, 2022, 26(8): 1504–1529.
SU Hongjun. Dimensionality reduction for hyperspectral remote sensing: advances, challenges, and prospects[J]. National Remote Sensing Bulletin, 2022, 26(8): 1504–1529. (in Chinese)

[20] 范学星, 张慧春, 邹义萍, 等. 基于多光谱成像与机器学习的植物叶绿素含量反演[J]. 林业科学, 2023, 59(7): 78–88.
FAN Xuexing, ZHANG Huichun, ZOU Yiping, et al. Inversion of plant chlorophyll content based on multispectral imaging and machine learning[J]. Scientia Silvae Sinicae, 2023, 59(7): 78–88. (in Chinese)

[21] ZHOU Z, ZHANG H, BIAN L, et al. Integrating sensor fusion with machine learning for comprehensive assessment of phenotypic traits and drought response in poplar species[J]. Plant Biotechnology Journal, 2025, 23(7): 2464–2481.

[22] HUSSAIN S, TESHOME F T, TULU B B, et al. Leaf area index (LAI) prediction using machine learning and UAV based vegetation indices[J]. European Journal of Agronomy, 2025, 168: 127557.

[23] HAN W, KETTERINGS Q M, ZHANG F. A double-layer model for improving the estimation of wheat canopy nitrogen content from unmanned aerial vehicle multispectral imagery[J]. Journal of Integrative Agriculture, 2023, 22(7): 2248–2270.

[24] JU C X, CHEN C, LI R, et al. Remote sensing monitoring of wheat leaf rust based on UAV multispectral imagery and the BPNN method[J]. Food and Energy Security, 2023, 12(4): e477.

[25] MOMTANARO G, PETROZZA A, RUSTIONI L, et al. Phenotyping key fruit quality traits in olive using RGB images and back propagation neural networks[J]. Plant Phenomics, 2023, 5: 61.

[26] MA Y R, MA L L, ZHANG Q, et al. Cotton yield estimation based on vegetation indices and texture features derived from RGB image[J]. Frontiers in Plant Science, 2022, 13: 925986.

[27] ŠTRONER M, URBAN R, SUK T. Filtering green vegetation out from colored point clouds of rocky terrains based on various vegetation indices: comparison of simple statistical methods, support vector machine, and neural network [J]. Remote Sensing, 2023, 15(13): 3254.

[28] ZHOU R, YANG C, LI E, et al. Aboveground biomass estimation of wetland vegetation at the species level using unoccupied aerial vehicle RGB imagery[J]. Frontiers in Plant Science, 2023, 14: 1181887.