

基于高光谱成像和多重优化策略的橙子早期腐烂检测

蔡乐天^{1,2} 蔡仲磊^{1,2} 章海亮³ 张译之³ 李江波²

(1. 石河子大学机械电气工程学院, 石河子 832003; 2. 北京市农林科学院智能装备技术研究中心, 北京 100097;

3. 华东交通大学电气与自动化工程学院, 南昌 330013)

摘要: 针对橙子早期腐烂识别中传统方法检测精度低的问题, 提出了一种融合高光谱成像与多重优化策略的检测方法, 旨在实现早期腐烂橙子的有效检测。基于 450 ~ 1 050 nm 波段的高光谱数据, 通过 Borderline - SMOTE 算法合成增强数据 (KL 散度为 0.02, Wasserstein 距离为 3.4), 采用 ReliefF 算法筛选了 24 个关键特征波长, 结合贝叶斯优化构建机器学习及 CNN 分类模型, 并系统地评估了模型分类性能与计算效率。CNN 模型经贝叶斯优化和特征筛选后分类误差降至 0.008, 运行时间由 910.4 s 减少至 177.9 s (降低 80.5%), 测试集准确率达 99.0%。

关键词: 橙子早期腐烂; 贝叶斯优化; 数据合成; 特征选择; 高光谱成像

中图分类号: TP391

文献标识码: A

文章编号: 1000-1298(2026)11-0397-08

OSID:



Early Detection of Orange Decay Based on Hyperspectral Imaging and Multiple Optimization Strategies

CAI Letian^{1,2} CAI Zhonglei^{1,2} ZHANG Hailiang³ ZHANG Yizhi³ LI Jiangbo²

(1. College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, China

2. Intelligent Equipment Research Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, China

3. College of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: A detection method that combined hyperspectral imaging and multiple optimization strategies was proposed to effectively detect early decay of oranges, addressing the issue of low detection accuracy in traditional methods. Traditional detection approaches, such as visual inspection and manual sorting, were highly subjective and fail to identify subtle changes in early decayed oranges, leading to substantial post-harvest losses in the citrus industry. To solve this problem, the hyperspectral data in the 450 ~ 1 050 nm wavelength range were collected by using a professional hyperspectral imaging system, covering the visible and near-infrared regions closely related to fruit internal quality. Then, enhanced data were synthesized by the Borderline - SMOTE algorithm with Kullback - Leibler (KL) divergence of 0.02 and Wasserstein distance of 3.4, which effectively alleviated the sample imbalance problem between healthy and early decayed oranges. Subsequently, totally 24 key characteristic wavelengths were screened out by the ReliefF algorithm to eliminate redundant information and reduce computational complexity. Machine learning based classification models and CNN model were constructed in combination with Bayesian optimization, which optimized key hyper parameters to improve model performance. A systematic evaluation was carried out on their classification performance and computational efficiency. After Bayesian optimization and feature selection, the classification error of the CNN model was reduced to 0.008. The running time was significantly decreased from 910.4 s to 177.9 s, representing a reduction of 80.5%. The accuracy rate on the test set reached 99.0%. The research result can not only provide a reliable technical solution for early decay detection of oranges, but also lay a theoretical foundation for the development of rapid and intelligent detection equipment, which was of great significance for promoting the high-quality development of the citrus industry.

Key words: early decay in oranges; Bayesian optimization; data synthesis; feature selection; hyperspectral imaging

收稿日期: 2025-02-05 修回日期: 2025-02-19

基金项目: 北京市农林科学院杰出科学家培育计划项目 (JKZX202405) 和国家自然科学基金项目 (31972152, 32260622)

作者简介: 蔡乐天 (1999—), 男, 硕士生, 主要从事农产品品质快速无损检测研究, E-mail: cailetianclt@163.com

通信作者: 李江波 (1982—), 男, 研究员, 博士生导师, 主要从事农产品无损检测分级与智能包装研究, E-mail: lijib@necita.org.cn

0 引言

柑橘类水果具有丰富的营养成分和药用价值,除直接食用外,还可以加工成陈皮、罐头、蜜饯、精油、香氛等产品,进一步延伸其产业链,增加产品附加值^[1]。我国柑橘栽培历史悠久,目前,柑橘是国内种植面积最大、产量最高的水果^[2]。然而柑橘在生长、采后运输、贮藏和销售等过程中容易受到真菌侵染,这对鲜食和加工副产品造成了极大的安全隐患^[3]。因此,早期识别并剔除腐烂的柑橘对避免经济损失和保障食品安全具有重要意义。

由于初期受真菌侵染区域的果皮与健康区域相似,因此,目视检查或常规方法无法有效检测早期腐烂的柑橘^[4]。如今,近红外光谱^[5]、荧光成像^[6]和激光后向散射成像^[7]等技术已应用于水果腐烂检测。然而,这些技术均有缺点,例如近红外光谱只能进行点测量、荧光成像对柑橘果的普适性差、激光后向散射成像的空间信息获取能力有限等,阻碍了它们在柑橘早期腐烂检测中的进一步应用^[8-9]。

高光谱成像技术将光谱信息与图像信息融为一体,具有计算机图像和光谱技术的双重优势^[10];根据光源和传感器的相对位置,高光谱成像一般有反射和透射两种模式。对于柑橘早期腐烂果的检测,现有涉及高光谱成像的相关研究主要集中在两方面:光谱法,主要利用腐烂组织和健康组织的光谱信息建立相关分类模型^[11];图像法主要是基于降维后的高光谱图像,开发图像处理算法识别果实腐烂区域^[12]。总体而言,这两类方法都能很好地检测出腐烂果,尤其是基于降维后的多光谱图像,腐烂柑橘果实的检测准确率可达98.6%^[13]。但这些研究仍存在一些问题,譬如,使用降维法处理后高光谱图像的颜色或强度总是被随机分配,因此基于图像处理的方法导致模型的稳定性较低^[14],而利用光谱信息构建的算法模型往往精度较低^[15]。因此,基于高光谱成像技术,开发一种更稳定、准确、快速的柑橘早期腐烂检测模型十分必要。

在检测模型构建时,不同的参数选择会直接影响模型的性能,提升或者减弱模型的检测精度。找到最优的参数组合,使得模型在给定的数据集上表现最好一直是各项检测任务所追求的目标^[16]。但参数优化往往是一项困难的任务,因为它需要在众多可能的参数组合中搜索最佳组合。为了解决这个问题,研究者提出了许多不同的参数优化方法,如经验法^[17]、网格搜索法^[18]、遗传算法^[19]、粒子群算法^[20]等。这些常见的参数寻优算法都存在着不同的局限性,并不能很好地满足目前研究的实际需求。

贝叶斯优化是一种基于贝叶斯定理和高斯过程回归的优化算法,其能够利用上一次参数的信息用于调整当前阶段的参数。贝叶斯优化能够在较少的迭代次数内找到全局最优解,避免了传统算法易陷入局部最优的问题。如今,贝叶斯优化算法已经被广泛应用于机器学习、自然语言处理、深度学习等诸多领域并取得了良好的效果^[21]。

本文以橙子为研究对象,获取反射高光谱图像,设计一套模型参数在线稳健调整方法,灵活、准确地对早期腐烂橙子进行检测,提高腐烂果检测准确率。

1 材料与方法

1.1 腐烂样本制备

本研究中所用橙子样本均购自水果市场。实验前,将橙子从冷藏环境中取出,使用纯净水进行清洗,随后在室温(22℃)下放置24 h以达到稳定状态。实验采用人工接种指状青霉菌孢子溶液的方法制备早期腐烂样本。具体操作为:从自然受指状青霉菌侵染后腐烂的橙子中分离霉菌,并溶于蒸馏水中制备成孢子溶液。使用带有钢针的注射器(针头直径1 mm,长度5 mm),在健康橙子样本的随机位置进行接种,将约20 μL的真菌孢子溶液注入橙皮下5 mm深处。接种48 h后采集样本的高光谱图像,此时橙子的腐烂情况无法通过肉眼判别,因此采集完成后切开橙子检查确认是否为腐烂样本。本实验中采集了80个腐烂橙子样本的高光谱图像,作为对照,另挑选了80个表皮无缺陷的健康橙子进行高光谱图像采集。

1.2 高光谱图像采集

本研究所有样本的高光谱图像均采用图1所示的高光谱成像系统采集。系统的光谱成像范围为325~1 100 nm,相邻光谱间隔约为0.778 nm。该系统主要包括成像光谱仪(型号ImSpector V10E-QE)、面阵CCD相机(型号ICL-B1620)、卤素灯光源(Antefore International Co.,中国台湾地区)、样本移动台(型号EZHR17EN)和计算机(Intel(R)Core™i7, RAM 32.00 GB)。图像采集过程中,先将样本放置在样本台上,然后打开高光谱图像采集软件和样本台运动控制软件,当样本台将测试样本(腐烂区域的一面朝向相机)移过相机视场时,成像光谱仪和相机记录样本信息,并将测试样本的图像数据同步存储在计算机上。获取样本的原始高光谱图像后,需在相同的系统参数下采集白参考图像和暗参考图像进行图像校正^[22]。公式为

$$R_c = \frac{R_0 - R_{dark}}{R_{white} - R_{dark}}$$

式中 R_c ——校正后的高光谱图像

R_0 ——原始高光谱图像

R_{dark} ——暗参考图像

R_{white} ——白参考图像

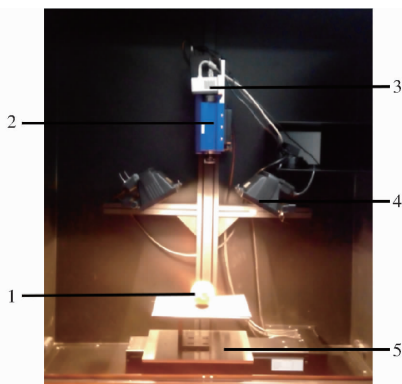


图 1 高光谱图像采集系统

Fig. 1 Hyperspectral image acquisition system

1. 被测样品 2. 光谱仪 3. 面阵 CCD 相机 4. 卤素灯 5. 载物台

1.3 数据合成算法

合成小样本数据有助于更全面地探索数据的潜在分布^[23]。本研究基于原始 160 条光谱数据(包括 80 个腐烂样本和 80 个健康样本),分别使用 ADASYN、SMOTE、Borderline - SMOTE 共 3 种数据合成算法各生成数量一致的光谱数据,即生成 80 条腐烂样本光谱数据和 80 条健康样本光谱数据。采用随机的方式将 320 条数据按训练集与测试集数量比 3:1 的方式划分,共得到 224 个训练样本和 96 个测试样本用于后续分析。ADASYN 算法可以自适应地采集少数类样本数据来合成新的少数类样本。在合成新样本时,通过在少数类样本与其近邻之间进行插值来生成^[24]。SMOTE 算法是一种经典的过采样算法,其基本原理是对于每个少数类样本,从其近邻中随机选择一个近邻,然后在该样本与所选近邻之间随机选取一点作为新合成的样本^[25]。Borderline - SMOTE 算法是对 SMOTE 算法的改进,该算法首先识别出那些处于分类边界上的少数类样本,这些样本被认为是对分类器性能影响较大的关键样本;然后,仅对这些边界样本应用 SMOTE 算法来合成新的少数类样本^[26]。

1.4 贝叶斯优化

贝叶斯优化首先会选择一个合适的概率模型(如高斯过程)作为代理模型,对目标函数进行建模^[27]。高斯过程根据已有的观测数据,对目标函数在整个搜索空间中的取值进行概率分布估计,它能够捕捉到函数值之间的相关性,从而利用已知点信息来推测未知点的情况。贝叶斯优化算法包括以下

几个部分:明确需要优化的目标函数 $f(x)$ 和参数空间 x 的取值范围。根据对参数的先验知识或假设,选择一个合适的概率分布作为参数的先验分布。基于后验分布进行优化以及设定迭代的停止条件,当满足停止条件时,迭代过程结束,得到最终的最优参数估计值。

1.5 ReliefF 特征选择算法

ReliefF 算法是一种基于统计学习的过滤式特征选择方法,其核心思想是通过分析特征在同类及异类样本邻近分布中的差异性来量化特征与类别的相关性。相较于原始 Relief 算法,ReliefF 算法通过引入多类别处理机制、概率估计策略及噪声鲁棒性改进,有效解决了原始方法在处理多分类问题、缺失数据及样本噪声方面的局限性,更适用于高维数据的特征筛选^[28]。在本研究中,使用 ReliefF 算法对橙子样本的 771 个光谱数据特征进行筛选,保留与分类模型最相关的光谱特征。

1.6 基于机器学习的分类模型

多项式支持向量机(SVM)、高斯 SVM、决策树和随机森林作为经典的机器学习分类模型在各研究领域被广泛使用^[29],本研究基于样本高光谱图像,对这 4 种经贝叶斯优化的分类模型性能进行评估。多项式 SVM 是一种基于多项式核函数的分类模型,通过将原始特征空间映射到高维多项式空间,构造最优分离超平面以实现类别划分,其性能由多项式阶数及正则化参数共同调控。高斯 SVM 则采用径向基核函数,通过非线性映射捕捉样本间的局部相似性,适用于复杂分布数据的分类任务,其分类性能受核宽度和正则化参数决定。决策树以信息增益或基尼不纯度作为分裂准则,递归构建树状结构实现特征空间的轴平行划分,具有直观可解释性,其性能由最大深度、最小分裂样本数及叶子节点最小样本数参数决定。随机森林通过 Bagging 集成策略与随机特征子采样构建多棵决策树,利用多数投票机制降低方差,显著提升模型鲁棒性与预测精度,尤其适用于高噪声、高维异构数据集。随机森林模型的性能受树的数量和最大特征参数影响。

1.7 神经网络

卷积神经网络(CNN)是一种专为处理复杂非线性数据设计的深度学习架构,其性能显著依赖于网络层数(深度)与每层神经元(滤波器)数量的优化配置。网络层数决定模型的表征能力:浅层网络可捕获基础边缘与纹理特征,而深层网络通过堆叠卷积模块逐步提取高阶语义信息,但深度过大会引发梯度消失或计算复杂度激增。神经元数量则直接影响特征映射的丰富性与模型容量,过多易导致过

拟合,需结合正则化等策略平衡^[30]。本研究依据贝叶斯优化激活函数(ReLU、Tanh、Sigmoid等)、神经元个数(每层1~10个)、全连接层数(1~3层)、数据是否标准化及正则化强度来构建早期腐烂橙子的最佳检测模型。

1.8 评价指标

1.8.1 分类模型评价指标

分类模型评价指标为准确率。较高的ACC(准确率)表示模型分类性能优异。TP(真正例)、TN(真负例)、FP(假正例)、FN(假负例)共同构成了混淆矩阵,从而全面反映模型的分类能力与误差分布特性;此外,还使用精确率和召回率评估分类模型性能。

1.8.2 数据相似度评价指标

本研究使用KL(Kullback-Leibler)散度和Wasserstein距离这两个指标来定量衡量合成数据与原始数据的相似程度,从而评判不同数据合成算法的性能。KL散度是一种非对称性度量,用于量化两个概率 P 和 Q 之间的差异,常用于评估合成数据与原始数据在概率密度层面的相似性。Wasserstein距离(Earth Mover's distance, EMD)基于最优运输理论,衡量将一个分布 G 转换为另一个分布 H 所需的最小“成本”。

2 结果与分析

2.1 光谱分析

在高光谱图像中,考虑到单像素的光谱曲线不具代表性,因此,每个样本选取一个包含80~100个像素的区域作为感兴趣区域(Region of interest, ROI),计算所选区域内所有像素的平均光谱作为ROI的代表光谱。所有样本典型ROI的光谱曲线如图2所示。图中蓝色曲线代表橙子腐烂组织的光谱,红色曲线代表橙子健康组织的光谱。从图中可以看出,两种组织的ROI光谱曲线在325~1100 nm区域内整体变化趋势相似,在波长约970 nm处有明显的吸收。从图2中还观察到,腐烂组织光谱的散射特性比健康组织光谱更明显,这可能是由于提取的腐烂组织可能位于样本的不同位置,导致不同ROI之间的光谱强度差异较大。由波长局部放大图可以看出,光谱两端存在较大的噪声,说明相关波长的信噪比较低,因此,本研究采用450~1050 nm区域的光谱数据(共771个波长)进行分析。

2.2 光谱数据合成

基于80个腐烂样本和80个健康样本的原始光谱数据,分别采用ADASYN、SMOTE和Borderline-SMOTE共3种算法合成了等量的光谱数据,并通过

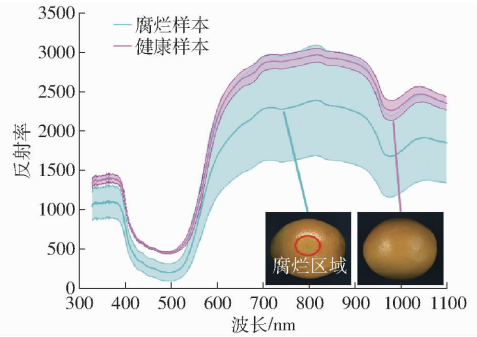


图2 橙子腐烂组织和健康组织的光谱曲线

Fig.2 Spectral curves of decayed and healthy tissues of orange samples

KL散度、Wasserstein距离及PCA(主成分分析)对合成数据质量进行定量和定性的评估。如表1所示,Borderline-SMOTE算法在两项分布相似性指标上均表现最优(KL散度为0.02,Wasserstein距离为3.4),显著优于ADASYN算法(KL散度为0.04,Wasserstein距离为9.1)和SMOTE算法(KL散度为0.04,Wasserstein距离为12.6)。KL散度与Wasserstein距离的联合分析表明,Borderline-SMOTE算法生成的合成数据与原始数据的概率分布差异最小,且在整体分布结构的匹配上更具优势,这可能归因于其聚焦于边界样本的生成策略,有效降低了噪声对数据分布的影响。ADASYN算法与SMOTE算法的KL散度相同,但Wasserstein距离差异显著,说明ADASYN算法在分布结构的全局对齐性上优于SMOTE算法。结合PCA降维结果(图3)进一步验证,Borderline-SMOTE算法的合成数据在低维空间的不同尺度上与原始数据的重叠度最高,表明其能更好地保留原始光谱特征的主成分信息;而ADASYN算法和SMOTE算法的合成数据在靠近中心点的位置上与原始样本重合度高,在偏离中心点的位置上合成数据与原始样本重合度低(图3红框内所示),这可能与生成过程中对少数类样本的过度偏移有关。

表1 不同数据合成算法性能评估

Tab.1 Performance evaluation of different data synthesis algorithms

数据合成算法	KL散度	Wasserstein距离
ADASYN	0.04	9.1
SMOTE	0.04	12.6
Borderline-SMOTE	0.02	3.4

2.3 机器学习分类模型性能评估

基于原始光谱数据和合成后的光谱数据,采用贝叶斯优化方法对多项式SVM、高斯SVM、决策树及随机森林4种分类模型的超参数进行迭代调优(共30次),并通过测试数据的混淆矩阵与最小观

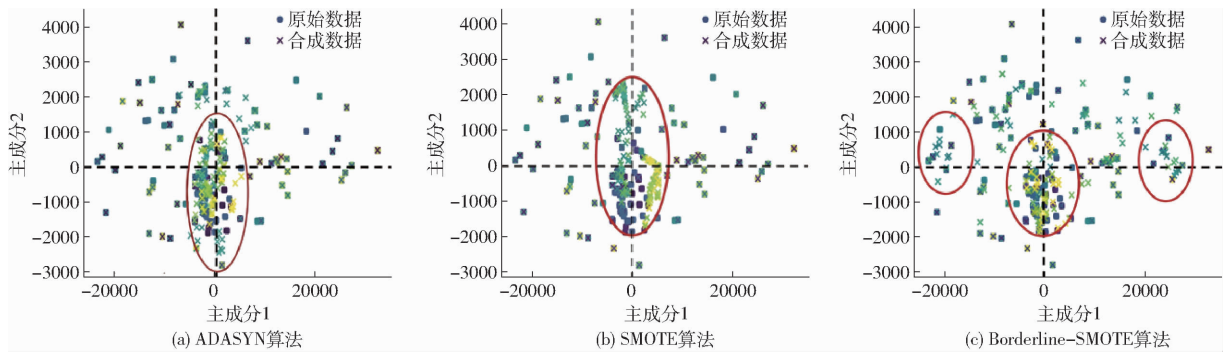


图 3 PCA 降维后原始数据与合成数据的分布

Fig. 3 Distribution of raw data and synthesis data after PCA dimensionality reduction

测值(目标函数值)评估模型性能。从图 4 可以看出, 高斯 SVM 在贝叶斯优化过程中取得了最小的观测值(0.017), 显著优于其他模型(多项式 SVM: 0.021, 随机森林: 0.063, 决策树: 0.077), 表明其参数组合对当前分类任务具有更强的适应性。结合混淆矩阵(图 5)进一步分析发现, 高斯 SVM 在正类(腐烂样本)检测中表现尤为突出, 其召回率(95.8%)和精确率(92.0%)均高于其他模型, 综合 ACC 达到 93.8%, 显著优于多项式 SVM (ACC 为 91.7%) 与随机森林 (ACC 为 89.6%)。这一优势可能源于高斯核函数对光谱数据非线性特征的强捕捉能力, 通过优化后的核宽度参数有效平衡了局部敏感性与全局泛化性, 从而降低了复杂光谱特征下的误分类风险。此外, 从贝叶斯优化的迭代过程可知, 高斯 SVM 和多项式 SVM 的收敛速度最快(约 13 次迭代后趋于稳定), 而决策树和作为集成模型的随机森林需更多迭代次数才能逼近最优参数, 说明 SVM 模型的超参数空间可能更易被高效搜索。然而, 所有模型对正类的错误分类率(高斯 SVM: 6.2%, 多项式 SVM: 8.3%; 随机森林: 10.4%, 决策树: 15.6%)仍不可忽视, 因此, 需进一步构建高适应性模型。

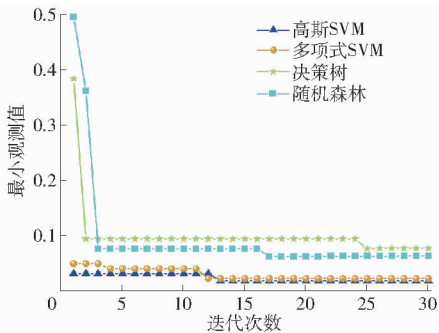


图 4 基于贝叶斯优化的 4 种分类模型超参数收敛曲线

Fig. 4 Hyperparameter convergence curves of four classification models based on Bayesian optimization

2.4 特征波长挑选

为了进一步提高模型性能, 基于 450 ~ 1 050 nm

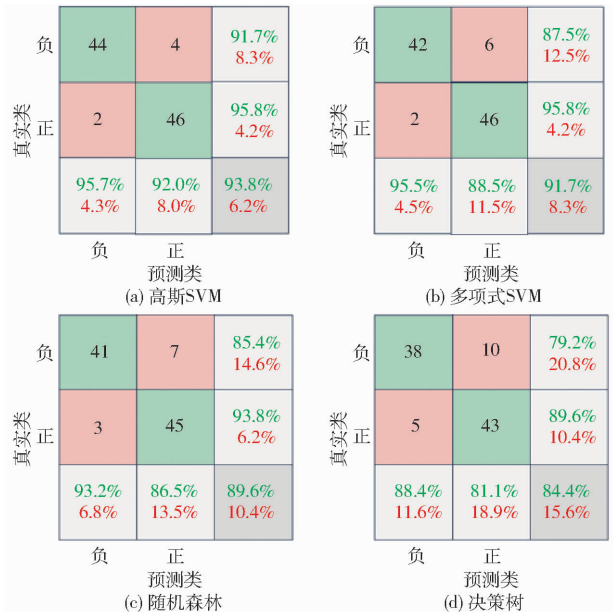


图 5 4 种分类模型在测试集上的混淆矩阵

Fig. 5 Confusion matrix of four classification models on test set

范围内的光谱数据, 使用 ReliefF 算法进行特征波长筛选。各个波段的重要性分数分布如图 6 所示, 本研究设定重要性分数阈值为 0.01, 挑选了大于该阈值的 24 个特征波长。从图 7 可以看出, 挑选的特征波长主要集中在 530 ~ 600 nm 内, 这可能与腐烂导致的橙子组分变化有关。总体来说, ReliefF 算法通过近邻样本的权重迭代机制, 有效捕捉了局部特征的交互作用, 避免了高维光谱数据中冗余噪声的干扰, 从而实现了关键波段的精准定位。

2.5 神经网络分类模型性能评估

基于 771 个特征波长和挑选后的 24 个特征波长数据, 本研究通过贝叶斯优化相关神经网络参数来构建早期腐烂橙子的最佳检测模型。表 2 所示为经过贝叶斯优化后的不同特征波长下的神经网络模型参数, 基于全特征波长构建的模型需要更复杂的网络结构(第 1 层 2 个神经元、第 2 层 5 个神经元), 而基于 24 个特征波长构建的模型在保持性能的同

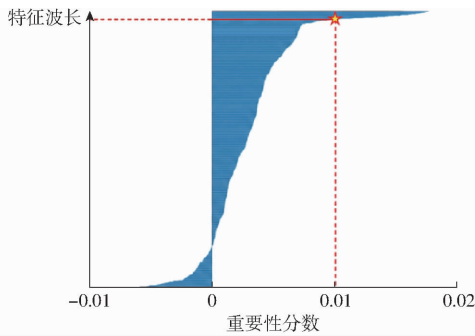


图6 特征波长重要性分数分布

Fig. 6 Importance score distribution of feature wavelengths

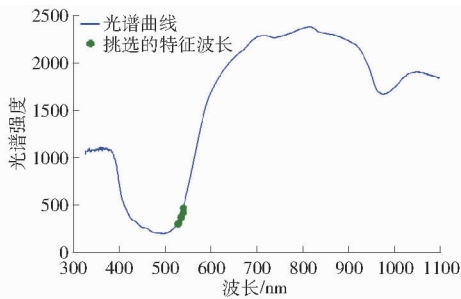


图7 基于 ReliefF 算法特征波长筛选结果

Fig. 7 Feature wavelengths selection results by ReliefF

时显著简化了结构(第1层1个神经元、第2层4个神经元)。尽管两者的正则化强度相近(约 10^{-8} 量级),但简化后的模型通过降低参数量减少了过拟合风险。

表2 经过贝叶斯优化后不同特征波长下的神经网络模型参数

Tab. 2 CNN parameters under different characteristic wavelengths after Bayesian optimization

优化参数	771个特征波长	24个特征波长
输入数据预处理	标准化	标准化
激活函数	ReLU	ReLU
神经元个数		
第1层	2	1
第2层	5	4
第3层	5	5
全连接层个数	3	3
正则化强度	4.6901×10^{-8}	6.7187×10^{-8}

表3 基于CNN的全特征与筛选特征波长分类性能及计算效率对比

Tab. 3 Comparison of classification performance and computational efficiency for CNN models with full and selected feature wavelengths

波长数量	样本类型	训练集			测试集			运行时间/s
		样本数量	误分类数量	准确率/%	样本数量	误分类数量	准确率/%	
771	腐烂	112	2	98.2	48	0	100.0	910.4
	健康	112	5	95.5	48	2	93.4	
	总数	224	7	96.8	96	2	96.9	
24	腐烂	112	1	99.1	48	0	100.0	177.9
	健康	112	3	97.3	48	1	98.0	
	总数	224	4	98.2	96	1	99.0	

如图8所示,基于771个全特征波长构建的CNN的估计最小分类误差为0.012,而经过特征筛选后基于24个特征波长的CNN降低至0.008,且收敛速度更快。这表明,冗余波长引入的高维噪声可能干扰模型训练,导致收敛缓慢和误差累积。ReliefF算法筛选的24个特征波长聚焦于腐烂相关的关键光谱响应,有效提升了模型对核心特征的捕捉能力。此外,观测误差与估计误差曲线的贴近程度表明,贝叶斯优化在参数搜索中具有较高的可靠性,尤其在低维特征空间下表现更优。

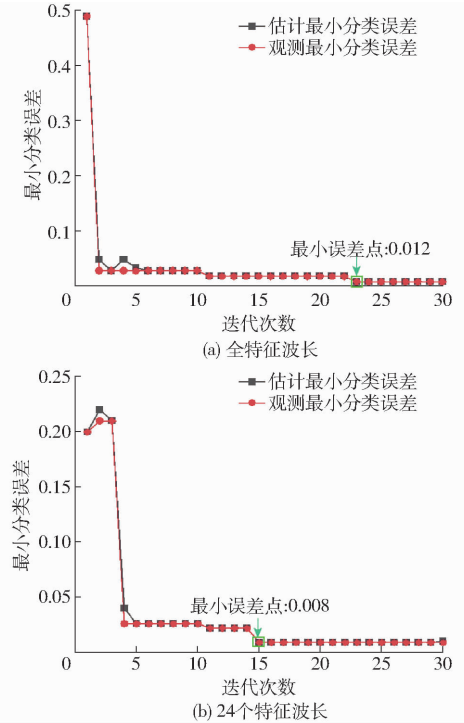


图8 基于贝叶斯优化的全特征与筛选特征波长下CNN分类误差收敛曲线对比

Fig. 8 Comparison of CNN classification error convergence curves under full and selected feature wavelengths based on Bayesian optimization

通过对比不同特征波长下CNN的样本分类结果及运行时间(表3)可以看出,基于771个全特征波长模型在测试集上取得了96.9%的健康样本准

确率,但其运行时间高达 910.4 s;而 24 个特征波长模型在测试集上的检测准确率上升至 99.0%,运行时间减至 177.9 s,缩短了 80.5%。这一结果表明,特征筛选在没有损失分类精度的前提下,显著提升了计算效率。

3 结论

(1)在光谱数据合成算法的对比分析中,相较于 ADASYN 和 SMOTE 算法,Borderline-SMOTE 算法展现出独特优势。通过 KL 散度与 Wasserstein 距离的量化评估,Borderline-SMOTE 合成的数据与原始数据分布最为接近(KL 散度为 0.02, Wasserstein 距离为 3.4)。这主要归因于其边界样本生成策略,该策略有效增强了模型对不同类别重叠区域的适应能力,从而提升了模型在复杂数据分布场景下的泛化性能。

(2)借助 ReliefF 算法从 771 个特征波长中筛选

出的 24 个特征波长,对模型性能优化成效显著。一方面,显著降低了模型的复杂度,避免了因特征过多导致的过拟合风险;另一方面,在保持高分类精度的前提下,大幅缩短了模型运行时间。与特征筛选前的 CNN 模型相比,运行时间减少了 80.5%,充分证明了特征选择在提升模型效率方面的关键作用。

(3)在机器学习分类模型性能评估中,高斯 SVM 经过贝叶斯优化后,展现出卓越的分类性能。通过贝叶斯优化对其核函数参数进行精细调优,有效平衡了模型的局部敏感性与全局泛化性,获得了最小观测误差(0.017),测试集分类准确率达到 93.8%。

(4)基于贝叶斯优化和特征波长筛选的 CNN 模型,在腐烂橙子检测任务中取得了突破性成果。该模型成功实现了高精度(ACC 为 99.0%)与高效率(运行时间为 177.9 s)的良好平衡,为橙子早期腐烂无损检测提供了一种可靠的解决方案。

参 考 文 献

- [1] MANTHEI A, ELEZ-MARTINEZ P, SOLIVA-FORTUNY R, et al. Prebiotic potential of pectin and cello-oligosaccharides from apple bagasse and orange peel produced by high-pressure homogenization and enzymatic hydrolysis[J]. Food Chemistry, 2024, 5(1):435.
- [2] 唐琪,李恒凯,周艳兵,等.基于 Sentinel-2 与时序 Sentinel-1 SAR 特征的赣南柑橘种植区识别方法[J].农业机械学报, 2024,55(3):193-202.
TANG Qi, LI Hengkai, ZHOU Yanbing, et al. Identification of Gannan citrus planting area based on Sentinel-2 and temporal Sentinel-1 SAR features[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 193-202. (in Chinese)
- [3] LI J, LU Y, LU R. Identification of early decayed oranges using structured-illumination reflectance imaging coupled with fast demodulation and improved image processing algorithms[J]. Postharvest Biology and Technology, 2024, 207: 112627.
- [4] ZHANG H, ZHANG S, DONG W, et al. Detection of common defects on mandarins by using visible and near infrared hyperspectral imaging[J]. Infrared Physics & Technology, 2020, 108: 103341.
- [5] LORENTE D, ESCANDELL-MONTERO P, CUBERO S, et al. Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on citrus fruit[J]. Journal of Food Engineering, 2015, 163: 17-24.
- [6] KURITA M, KONDO N, SHIMIZU H, et al. A double image acquisition system with visible and UV LEDs for citrus fruit[J]. J. Robotics Mechatronics, 2009, 21(4): 533-540.
- [7] LORENTE D, ZUDE M, IDIER C, et al. Laser-light backscattering imaging for early decay detection in citrus fruit using both a statistical and a physical model[J]. Journal of Food Engineering, 2015, 154: 76-85.
- [8] SLAUGHTER D, OBENLAND D, THOMPSON J, et al. Non-destructive freeze damage detection in oranges using machine vision and ultraviolet fluorescence[J]. Postharvest Biology and Technology, 2008, 48(3): 341-346.
- [9] OBENLAND D, MARGOSAN D, SMILANICK J, et al. Ultraviolet fluorescence to identify navel oranges with poor peel quality and decay[J]. Hort Technology, 2010, 20(6): 991-995.
- [10] 李斌,张烽,殷海,等.基于高光谱图谱融合技术的黄桃损伤程度判别研究[J].光谱学与光谱分析,2023,43(2):435-441.
LI Bin, ZHANG Feng, YIN Hai, et al. Discrimination of damage degree of yellow peaches based on hyperspectral image-spectrum fusion technology[J]. Spectroscopy and Spectral Analysis, 2023, 43(2): 435-441. (in Chinese)
- [11] FOLCH-FORTUNY A, PRATS-MONTALBÁN J M, CUBERO S, et al. VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits[J]. Chemometrics and Intelligent Laboratory Systems, 2016,156:241-248.
- [12] LUO W, FAN G, TIAN P, et al. Spectrum classification of citrus tissues infected by fungi and multispectral image identification of early rotten oranges[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 279: 121412.
- [13] LI J, HUANG W, TIAN X, et al. Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral

- imaging[J]. Computers & Electronics in Agriculture, 2016, 127:582 – 592.
- [14] ZHANG B, LIU L, GU B, et al. From hyperspectral imaging to multispectral imaging: portability and stability of HIS – MIS algorithms for common defect detection[J]. Postharvest Biology and Technology, 2018, 137:95 – 105.
- [15] GHOSHKHANEH N G, MOLLAZADE K. Optical techniques for fungal disease detection in citrus fruit: a review[J]. Food and Bioprocess Technology, 2023, 16(8): 1668 – 1689.
- [16] WANG B C, HE Y B, LIU J, et al. Fast parameter identification of lithium-ion batteries via classification model-assisted Bayesian optimization[J]. Energy, 2024, 288: 129667.
- [17] LEKOUAGHET B, MERROUCHE W, BOUGUENNA E, et al. Advanced parameter estimation for lithium-ion battery model using the information sharing group teaching optimization algorithm[J]. Journal of Power Sources, 2024, 624: 235615.
- [18] CHEN X, XU Z, WU Y, et al. Heuristic algorithms for reliability estimation based on breadth-first search of a grid tree[J]. Reliability Engineering & System Safety, 2023, 232: 109083.
- [19] 张芳旭,王军,韩宇平,等. 基于遗传算法的二维土壤水与作物生长耦合模拟模型构建和参数优化[J]. 农业机械学报, 2024,55(12):392 – 403.
ZHANG Fangxu, WANG Jun, HAN Yuping, et al. Developing and parameter optimization of two-dimensional soil water transport and crop growth coupling model based on genetic algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 392 – 403. (in Chinese)
- [20] DENG X, HE D, QU L. A novel hybrid algorithm based on arithmetic optimization algorithm and particle swarm optimization for global optimization problems[J]. Journal of Supercomputing, 2024, 80(7):8857 – 8897.
- [21] WANI P, JAVIDI B. 3D integral imaging depth estimation of partially occluded objects using mutual information and Bayesian optimization[J]. Optics Express, 2023, 31(14): 22863.
- [22] LU B, DAO P D, LIU J, et al. Recent advances of hyperspectral imaging technology and applications in agriculture[J]. Remote Sensing, 2020, 12(16):2659.
- [23] ZHANG Y, DENG L, HUANG H, et al. An improved SMOTE based on center offset factor and synthesis strategy for imbalanced data classification[J]. The Journal of Supercomputing, 2024, 80(15):22479 – 22519.
- [24] STANOSHECK J A, CASTELL-PEREZ M E, MOREIRA R G, et al. Oversampling methods for machine learning model data training to improve model capabilities to predict the presence of *Escherichia coli* MG1655 in spinach wash water[J]. Journal of Food Science, 2024, 89(1): 150 – 173.
- [25] 朱潘雨,黄敏,赵鑫. 基于 SMOTE – UVE – SVM 的小麦种子纯度高光谱图像检测[J]. 激光技术, 2024, 48(2):281 – 287.
ZHU Panyu, HUANG Min, ZHAO Xin. Hyperspectral image detection of wheat seed purity based on SMOTE – UVE – SVM [J]. Laser Technology, 2024, 48(2): 281 – 287. (in Chinese)
- [26] WANG H, MENG Y, XU H, et al. Prediction of flood risk levels of urban flooded points though using machine learning with unbalanced data[J]. Journal of Hydrology, 2024, 630: 130742.
- [27] JU S, SHIGA T, FENG L, et al. Designing nanostructures for phonon transport via Bayesian optimization[J]. Physical Review X, 2017, 7(2):021024.
- [28] YONG X, GAO Y L. Improved firefly algorithm for feature selection with the ReliefF-based initialization and the weighted voting mechanism[J]. Neural Computing and Applications, 2023, 35(1):275 – 301.
- [29] DAI W, JIN H, ZHOU L, et al. Testing machine learning algorithms on a binary classification phenological model[J]. Global Ecology and Biogeography, 2022, 32(1): 178 – 190.
- [30] ZHAO X, WANG L, ZHANG Y, et al. A review of convolutional neural networks in computer vision [J]. Artificial Intelligence Review, 2024, 57(4):99.