

基于改进LightGBM的冷鲜羊肉新鲜度无损检测分类

鄂计祥^{1,2}, 姜新华^{1,2*}, 徐子洋^{1,2}, 白洁^{1,2}, 李靖^{1,2}, 吴国栋², 翟成珺³

(1. 内蒙古农业大学计算机与信息工程学院, 内蒙古呼和浩特 010018)

(2. 内蒙古自治区农牧业大数据研究与应用重点实验室, 内蒙古呼和浩特 010018)

(3. 内蒙古自治区教育考试院, 内蒙古呼和浩特 010011)

摘要: 为了实现羊肉新鲜度的快速、无损检测分类。该研究首先通过结合多种新鲜度指标与高光谱数据构建了冷鲜羊肉新鲜度的原始数据集, 并采用不同的预处理方法对其建立支持向量回归 (Support Vector Regression, SVR), 确定了最佳的预处理方法。然后利用基于随机森林的递归特征消除和自编码器混合的特征选择方法对经预处理后的数据集进行波段选择, 对选择后的特征数据集建立轻量梯度提升机 (Light Gradient Boosting Machine, LightGBM), 以实现冷鲜羊肉新鲜度检测。由于传统的 LightGBM 模型在处理高维、稀疏的光谱数据时, 分类性能不高且缺少泛化能力。故该文提出一种基于混合蝙蝠算法和星鸦算法的多策略优化方法对 LightGBM 进行改进, 将改进后的 LightGBM 与改进前 LightGBM、XGBoost、支持向量机 (Support Vector Machine, SVM) 和随机森林 (Random Forest, RF) 等分类方法进行对比分析, 并在不同规模的数据集上进行验证。结果表明: 多策略改进后 LightGBM 的新鲜度等级分类精度达到了 0.988 1, 相比于未改进的模型提升了 3.57%, 查准率、召回率和 F1 分值分别为 0.988 6、0.988 2 和 0.988 3, 均优于其他分类方法。且模型在检测不同规模数据集的新鲜度等级时, 其分类准确率稳定在 0.984 0~0.989 0, 说明其具有良好的泛化性和鲁棒性。综上证明了该文提出的改进后 LightGBM 模型的有效性。为实现快速、无损的冷鲜羊肉新鲜度检测分类提供了一种新方法。

关键词: 羊肉新鲜度; 高光谱; 特征选择; 多指标; 轻量梯度提升机

文章编号: 1673-9078(2026)1-310-320

DOI: 10.13982/j.mfst.1673-9078.2026.1.1439

A Non-destructive Detection Classification Method for Chilled Mutton's Freshness Based on Improved LightGBM Model

E Jixiang^{1,2}, JIANG Xinhua^{1,2*}, XU Ziyang^{1,2}, BAI Jie^{1,2}, LI Jing^{1,2}, WU Guodong², ZHAI Chengjun³

(1.College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

(2.Key Laboratory of Big Data Research and Application of Agriculture and Animal Husbandry, Hohhot 010018, China)

(3.Education Examination Authority of Inner Mongolia Autonomous Region, Hohhot 010011, China)

Abstract: In order to achieve rapid and non-destructive detection and classification of the freshness of mutton, in this work, a raw dataset of chilled mutton's freshness was first constructed through integrating multiple freshness indicators with hyperspectral data. A variety of preprocessing techniques were applied to establish a Support Vector Regression model, through which the optimal preprocessing method was determined. Subsequently, a feature selection process was initiated via Random Forest-based Recursive

引文格式:

鄂计祥,姜新华,徐子洋,等.基于改进LightGBM的冷鲜羊肉新鲜度无损检测分类[J].现代食品科技,2026,42(1):310-320.

E Jixiang, JIANG Xinhua, XU Ziyang, et al. A non-destructive detection classification method for chilled mutton's freshness based on improved lightGBM model [J]. Modern Food Science and Technology, 2026, 42(1): 310-320.

收稿日期: 2024-09-26; 修回日期: 2024-12-14; 接受日期: 2024-12-23

基金项目: 国家自然科学基金项目 (62061037; 31960494); 内蒙古自然科学基金项目 (2023LHMS06017); 内蒙古自治区科技攻关计划项目 (0200GG0169)

作者简介: 鄂计祥 (2000-), 男, 硕士研究生, 研究方向: 模式识别和智能信息处理, E-mail: ejx@emails.imau.edu.cn

通讯作者: 姜新华 (1977-), 男, 博士, 教授, 研究方向: 图像处理、计算机视觉和数据科学与智能信息处理, E-mail: jiangxh@imau.edu.cn

Feature Elimination and an autoencoder for band selection on the basis of the preprocessed datasets. A Light Gradient Boosting Machine (LightGBM) was subsequently built on the selected feature dataset to detect the freshness of chilled mutton. Given the limitations of traditional LightGBM in handling high-dimensional and sparse spectral data, particularly the low classification performance and lack of generalization capability, a multi-strategy optimization method based on the Hybrid Bat Algorithm and Nutcracker Optimization Algorithm was introduced to improve LightGBM. The improved LightGBM was compared with classification methods such as the original LightGBM, XGBoost, Support Vector Machine (SVM), and Random Forest (RF), followed by validation on datasets of different scales. The results demonstrated that the multi-strategy improved LightGBM achieved a freshness grade classification accuracy of 0.988 1, resulting in a 3.57% improvement compared with the unimproved version, with weighted precision, weighted recall, and weighted F1 scores reaching 0.988 6, 0.988 2, and 0.988 3, respectively, all outperforming other classification methods. Furthermore, when detecting freshness grades across datasets of different scales, the model exhibited a consistent accuracy within the range of 0.984 0 to 0.989 0, indicating its good generalization capability and robustness. In summary, the effectiveness of the proposed improved LightGBM model has been validated in this work, providing a novel approach for rapid and non-destructive detection and classification of chilled mutton's freshness.

Key words: mutton freshness; hyperspectral; feature selection; multi-indicator; lightGBM

羊肉具有独特的风味和高营养价值,在全球各地的饮食文化中享有重要位置^[1]。然而其食用性和安全性易受到新鲜度的影响^[2]。冷鲜保存是维持羊肉新鲜度的关键技术^[3],但是在此条件下,肉类的化学属性和微生物活性仍会发生变化,影响羊肉的品质和消费安全,从而降低消费者的接受度。因此,建立一种快速、无损的检测方法,以监测冷鲜羊肉在储存过程中的新鲜度变化,对于保证其品质至关重要。

羊肉新鲜度的传统检测方法包括感官评价和实验室理化分析^[4],虽然能够获得较为准确的结果,但前者具有主观性,极易受到经验影响;后者通常需要耗费大量时间、人力,并且破坏样本,难以实现快速和批量检测。随着高光谱成像(Hyperspectral Imaging, HSI)技术的发展,其在食品检测中的巨大优势越加受到重视^[4],高光谱数据能够反映样本中化学键特征及微小变化的详细光谱信息,为无损检测提供了强有力的技术支持。Cheng等^[6]提出了一种结合二维相关光谱分析的轻量级三维卷积神经网络模型,构建了冷冻猪肉脂质氧化的定量预测模型;徐子洋等^[7]利用高光谱技术结合改进的深度随机森林算法,构建了羊肉新鲜度的多标记分类模型,其分类效率优于传统多标记分类算法;Yu等^[8]结合低成本高光谱成像技术与多种分类器,精准区分了冷冻牛肉的状态变化,并实现了新鲜度的可视化;Wan等^[9]通过光谱图像融合技术结合LSSVM模型,成功预测了与羊肉新鲜度相关的肌红蛋白含量。除肉类检测的其他检测领域中,高光谱技术和机器学习算法的结合也进行了应用,如豆腐质量的检测^[10]、白酒掺假^[11]与山核桃种子质量检测和品种分类^[12]等。然而,随着随机森林和支持向量机等传

统学习方法在这些研究中被广泛使用,发现其在处理多类别和高维特征时较为复杂,缺少扩展性,同时需要大量的内存和计算资源。因此,为应对新鲜度光谱数据的复杂性和多类别分类的挑战,需要更高效、更具适应性的检测方法。

轻量梯度提升机(Light Gradient Boosting Machine, LightGBM)基于梯度提升决策树^[13],其在解决分类和回归问题上的表现出色,尤其在多分类问题上体现出独特优势。大部分LightGBM模型的优化算法能实现对模型参数的自适应调整,但其缺少泛化能力,收敛速度不高。传统的LightGBM模型的优化方法包括麻雀算法、粒子群算法等。高兵等^[14]提出基于麻雀搜索算法的改进粒子群算法优化LightGBM模型,提高了分类准确率,但是其在数据关系的捕捉上仍不够充分,参数寻优的智能化程度也有待提高。Xie等^[15]提出基于分段混沌映射和萤火虫扰动策略优化粒子群算法,利用优化后的算法改进LightGBM算法,以实现对物联网网络安全态势进行高精度的评估,但是在不同环境中的泛化能力还有待验证。

为解决上述问题,该文提出了一种基于混合蝙蝠算法(Hybrid Bat Algorithm, HBA)和星鸦算法(Nutcracker Optimization Algorithm, NOA)的多策略优化方法对LightGBM模型进行改进,该方法结合了生物启发式算法的优势,能够更智能、更高效的计算LightGBM模型参数的最优配置,同时可以用于探索更大的参数空间,使其避免陷入局部最优解,找到真正意义上的全局最优解。综上,该研究基于高光谱技术结合多策略改进后的LightGBM模型,构建了冷鲜羊肉新鲜度多指标分类模型,以实现不同新鲜度羊肉的准确检测。

1 材料与方法

1.1 材料与仪器

1.1.1 试验材料

试验材料选用由内蒙古呼和浩特市回民区食品有限公司采购的新鲜羊肉。样本均取自屠宰后经过 0~4 ℃ 预冷成熟处理 24 h 后的羊肉里脊部位，因其组织均匀性最好，使样本初始条件接近。为减少空气湿度及微生物等因素对新鲜度的影响^[16]，对羊肉及其表皮和脂肪进行统一包装，放置于冷藏箱内尽快运回实验室。然后，在实验室标准操作台上，去除羊肉样本表面脂肪、筋膜以及结缔组织。随机选取里脊肉三个不同部位分割成 6 cm×6 cm×1 cm 的肉块，分成五组（一组高光谱数据采集、四组新鲜度指标测定）进行真空包装后，置于 4 ℃ 的冰箱连续存放十四天，每天取样，最终得到 3×5×14=210 个羊肉样本。

在此期间每日测定挥发性盐基氮（Total Volatile Basic Nitrogen, TVB-N）、酸碱度值（Potential of Hydrogen Value, pH）的理化指标含量，以及菌落总数（Total Aerobic Count, TAC）和大肠菌群近似数（Most Probable Number, MPN）的微生物指标含量。

1.1.2 新鲜度指标

TVB-N 是评估肉类新鲜度的关键化学指标，可反映微生物分解蛋白质过程中产生的挥发性碱性物质含量。pH 的变化通常与肉类腐败进程密切相关，较高的 pH 指示了蛋白质分解产物的积累，表明肉类的新鲜度下降。TAC 是肉类样本中需氧菌的总数量，与微生物的生长情况直接相关。MPN 用于估算大肠菌群的数量，其繁殖不仅会加剧肉类的腐败过程，还可能与病原菌的存在相关^[2]。测定 14 日内样本的新鲜度指标：TVB-N 的含量测定^[17]通过 GB 5009.228-2016《食品中挥发性盐基氮的测定》、pH 测定^[18]根据 GB 5009.237-2016《食品 PH 值的测定》、TAC 按照 GB 4789.2-2022《菌落总数测定》规定检测^[19]、MPN 依据 GB 4789.3-2016《大肠杆菌计数》进行计数^[20]。

试验样本的新鲜度分为三个等级：新鲜（Fresh）、次新鲜（Sub Fresh, SFresh）和不新鲜（Not Fresh, NFresh）。根据国家标准 GB 2707-2016《鲜（冻）畜、禽产品》规定^[21]，仅规定新鲜肉类 TVB-N 含量≤15 mg/100 g。为进一步细化新鲜度等级，该研究将次新鲜肉类的 TVB-N 含量划分为 15 mg/100 g 至 25 mg/100 g，并设定不新鲜肉类 TVB-N 含量>25 mg/100 g。由于 pH、TAC 和 MPN 指标尚未明确

制定有关肉类新鲜度的标准。因此，为了准确评价新鲜度等级，将 TVB-N 结合其他指标的变化趋势进行综合分析。该文的指标变化情况可作为肉类品质监控的参考依据，为后续新鲜度评估和研究提供数据支持。

1.1.3 高光谱系统和数据采集

该研究的高光谱系统由美国 Headwall Photonics 公司生产的 Hyperspec VNIRN-Series 高光谱成像仪、扫描平台及两盏 100 W 的可调节卤素灯组成。成像仪覆盖的光谱范围为 400~1 000 nm，分辨率为 2.8 nm，共包含 750 个光谱通道。高光谱数据通过高光谱系统和数据采集软件共同完成，具体流程如图 1 所示。

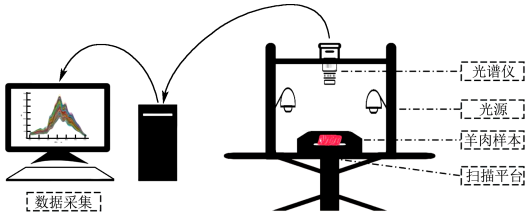


图 1 高光谱数据采集流程

Fig.1 Hyperspectral data acquisition process

采集时，提前 30 min 预热光谱仪，并调节最大通光量为 F2.0，将冷鲜羊肉样本置于扫描平台上，距扫描镜头约 30 cm 处。设置像元混合次数为 6 次，并调整曝光时间，使 DN 值小于 8 500，点击采集按钮，获取到样本的高光谱图像。受照明不均和机箱电流影响，需利用公式（1）对光谱图像进行黑白校正。

$$G = \frac{I-B}{W-B} \tag{1}$$

式中：

B——黑校正图像；

W——白校正图像；

G——校正后图像。

1.2 实验方法

1.2.1 研究框架

该研究的关键流程如图 2 所示。研究的前期任务主要是分析冷鲜羊肉在连续 14 d 内的高光谱数据与试验室测定的化学指标含量变化之间的关系。通过构建回归模型进行预处理方法优选和 RFFEA 特征选择，以探索冷鲜羊肉在储存过程中新鲜度的变化规律，并得到最佳的新鲜度特征数据集。后期任务主要为改进后 LightGBM 分类模型的构造与训练。基于前期提取的特征数据集，对样本的新鲜度等级进行分类，以满足冷鲜羊肉的新鲜度检测。

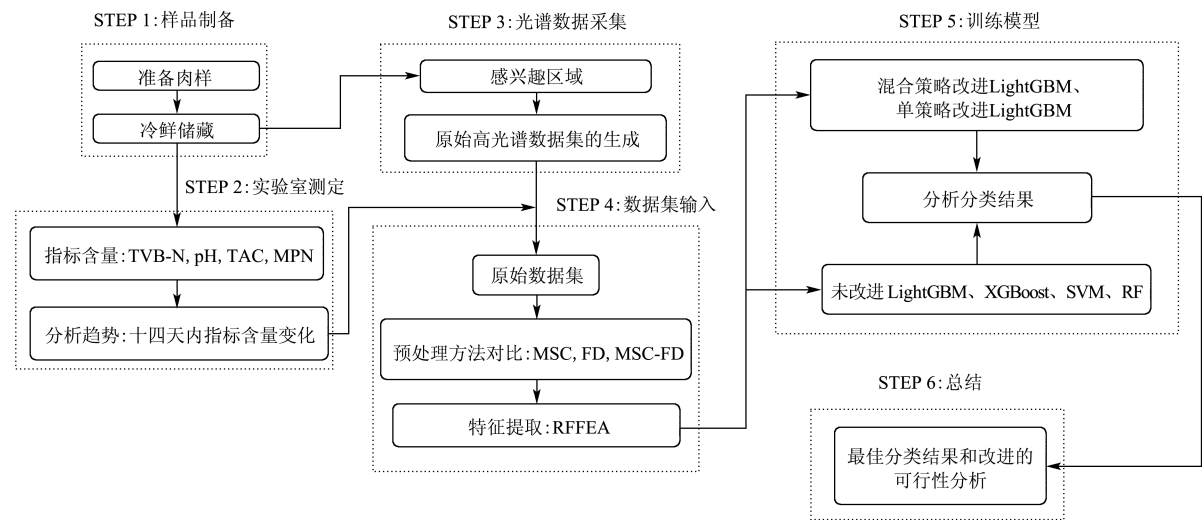


图 2 实验关键步骤的流程

Fig.2 Flow of experimental key steps

1.2.2 预处理

采集每日的高光谱图像后，利用 ENVI 5.6 提取其感兴趣区域（Region of Interest, ROI），得到原始光谱像元亮度（Digital Number, DN）值。为避免少量脂肪或其他结缔组织等影响数据准确度，每个样本仅提取瘦肉区域。从每日羊肉样本的光谱图像中随机选取 20 个大小为 10×10 像素的 ROI，以构建实验数据集 280（样本数）×125（波段数）。在此基础上，将每日光谱图像中 ROI 的选取数量扩大为原来的三倍，以得到验证数据集 840（样本数）×125（波段数）。以样本数命名两个数据集，其新鲜度的组成数量关系如表 1 所示。

表 1 数据集新鲜度组成

Table 1 Freshness composition of the dataset		
新鲜度	280 数据集	840 数据集
Fresh	60	180
SFresh	120	360
NFresh	100	300

部分原始光谱 DN 值图像如图 3 所示。可知光谱波段出现较大的波动，这种波动可能是由噪声引起，来源于仪器的不稳定性 and 环境干扰^[16]。另外，由于羊肉样本的物理性质，如切块样本的不均匀和表面的粗糙程度不可控，也可使部分光谱的基线出现变化。因此，针对数据的预处理具有必要性。常用的预处理方法有 SG 平滑滤波（Savitzky-Golay, SG）、多元散射校正（Multiplicative Scatter Correction, MSC）、一阶导数（First Derivative, FD）等，通过对比分析确定最佳的预处理方法。

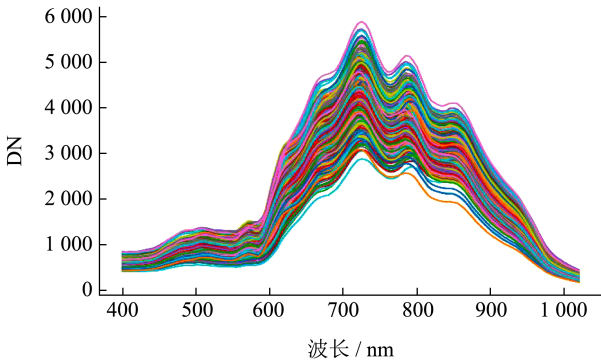


图 3 部分原始光谱反射 DN 值

Fig.3 Original spectral reflection DN image

1.2.3 特征波段选择

为得到最佳的新鲜度特征数据集。该文提出一种基于随机森林的递归特征消除和自编码器混合的特征选择方法（RF-RFE and Autoencoder Hybrid, RFMEA）。其结合了自编码器的非线性特征提取能力与随机森林的递归特征消除。首先，通过自编码器对高光谱数据的潜在结构特征进行提取，强化了对数据内在关联和非线性模式的捕捉。然后，在多轮迭代中利用随机森林的递归特征消除，基于特征重要性逐步去除冗余信息，以保留最具判别力的特征波段。该策略在降低数据维度的同时保留了关键辨识信息，旨在使模型在复杂的光谱数据分析中既具备深层的特征表征能力，又可以实现噪声去除和特征精简。

设定数据集 Z ，样本数为 i ，原始特征波段数为 j 。自编码器通过最小化输入数据与其重建输出之间的误差，学习数据的低维潜在特征表示。接着，其编码器部分对原始数据集 Z 进行编码，生成新的特征波段集

$Z_{encoded}$ 。随后对特征集进行特征重要性评估，通过逐步移除贡献较小的特征波段，将特征数量减少至最佳数量 N ($N_{max}=125$, $N_{min}=2$)。并得到最佳的特征子集 S_N 用于建立新鲜度特征数据集。RFFEA 算法流程如表 2 所示。

另外，采用竞争自适应重加权抽样（Competitive Adaptive Reweighted Sampling, CARS）和连续投影算法（Successive Projections Algorithm, SPA）作为对比，以评估 RFFEA 在光谱特征选择中的有效性。CARS 是常用于光谱分析中的变量选择方法，其通过基于回归系数的权重分配，赋予重要变量更高的权重，以实现 对特征波段的选择。SPA 也适用于光谱数据的分析，它通过将变量投影到多维空间从而选择最有潜力减少 冗余的特征变量，以增强特征集的可解释性。

1.2.4 基于HBA和NOA多策略改进的LightGBM分类模型

该研究提出了一种基于 HBA 和 NOA 的多策略优化方法（生物启发式）对 LightGBM 分类模型进行改进，以增强模型对高光谱数据中复杂特征的识别能力。LightGBM 作为高效的梯度提升框架，采用了基于决策树的学习算法，能够保持对新数据的高效泛化。模型的子决策树选择过程如图 4 所示。

LightGBM 的基本结构包括直方图（Histogram）、基于梯度的一侧采样（Gradient-based One-Side Sampling, GOSS）以及排除零特征的有效负载（Exclusive Feature Bundling, EFB）等算法。旨在提高

不同数据集上的计算效率，尤其在特征维度较高的情况下，可以通过直方图分桶、特征稀疏等处理方式降低复杂度。然而，LightGBM 的复杂算法结构带来了大量可配置的超参数，且默认的参数配置无法充分挖掘出数据中的差异。因此，为了提高冷鲜羊肉新鲜度检测的分类精度和模型鲁棒性，采用该文的多策略优化方法对模型的超参数进行配置。使 LightGBM 模型更好地适应光谱数据的复杂非线性结构和高维稀疏性，以提供更可靠的技术支持。

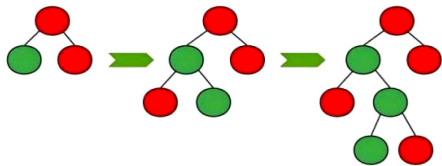


图 4 子决策树选择过程

Fig.4 The process of selecting sub-decision trees

混合蝙蝠算法^[22]是一种基于群体的启发式算法，主要利用了蝙蝠算法（Bat Algorithm, BA）的特性来优化目标模型的性能，目的是结合混合优势来克服单一算法面临的局限，通过模拟蝙蝠的回声定位行为来优化问题，其会根据搜索过程中的反馈信息优化算法中参数的调整策略，以适应不同阶段的搜索需求。

星鸦优化算法^[23]是一种基于种群的元启发式算法，模拟了星鸦在寻找和储存食物的行为，在广泛的区域内搜索食物，并且能够记住食物的位置，反复利用一个食物源，直到耗尽，如果食物短缺，星鸦会随机的探索新的地区，但仍保留旧有食物的位置。

表 2 RFFEA特征选择算法流程

Table 2 RFFEA feature extraction algorithm workflow

Input:	Data $Z(i, j)$, Target variable y , number of features $N_{max}=125$, $N_{min}=2$
1.	Standardized scaling of raw data: $Z(i, j) \rightarrow Z_{scaled}(i, j)$ # 数据标准化，消除特征间尺度差异
2.	Create an autoencoder (optimizer='adam', loss='mse'), then Start training: $\text{Min } PZ_{scaled} - \hat{Z}_{scaled}P^2$ # 创建自编码器并训练
3.	The encoding process maps data to a new feature space: $Z_{encoder}=\text{encoder.predict}(Z_{scaled})$ # 利用编码器生成新特征空间
4.	Feature selection: for N in range ($N_{min}, N_{max} + 1$): $rfe = RFE(\text{estimator} = rf, n_{\text{features_to_select}} = N)$ $rfe.fit(Z_{encoder}, y)$ # 最佳特征数为 N
5.	Root mean square as an evaluation indicator: $mse = \text{mean_squared_error}(rf.predict)$ # 均方误差作为评价指标
6.	Update Best Features: if $mse < best_mse$: $mse = best_mse, n = best_n, S_N = Z.columns[selector.support]$ # 若当前 MSE 为最小值，则更新最佳特征数量和最佳特征集
Output:	mse 、 N 、 S_N

基于 HBA 和 NOA 的多策略优化方法, 结合了两种生物启发式算法的优势, 利用蝙蝠搜索实现速度、位置的更新、并通过模仿星鸦觅食的行为以适应参数数量和灵活性。其中引入随机性提高跳跃能力, 同时添加收敛因子对多策略优化算法的收敛速度和性能分析。在参数寻优过程中, 进行全局搜索并记忆时间线内的最优解, 同时能够便利的重复访问, 这在算法中通过 Lévy 飞行和策略性使用历史最优解进行平衡。此外, 在位置更新中加入变异操作和自适应阈值的动态调整最终实现局部与全局搜索的动态平衡, 最终返回最佳的参数值组合。以对 LightGBM 模型进行自主配置。优化算法模拟对象命名为飞行物, 算法的具体实施流程如图 5 所示。

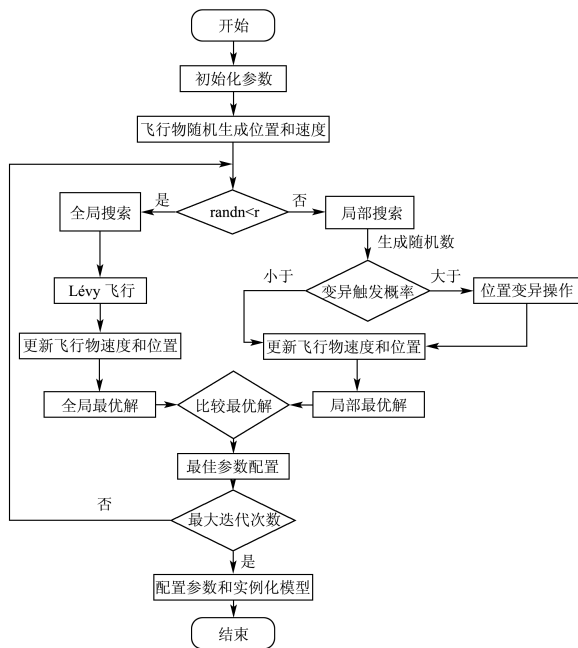


图 5 算法流程

Fig.5 Algorithm workflow

实现 Lévy 飞行时, 执行公式 (2)、(3)、(4):

$$\sigma = \left\{ \frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma[(1+\beta)/2]\beta 2^{(\beta-1)/2}} \right\}^{1/\beta} \quad (2)$$

$$u = \sigma \cdot \text{randn}, v = \text{randn} \quad (3)$$

$$\text{levy}(\beta, d) = u / (|v|^{1/\beta}) \quad (4)$$

式中:

randn——生成标准正态分布的随机数;

u——位置参数 (通常为 0);

σ ——决定了每次跳跃幅度, β 控制了步长分布的尾部厚度;

d——指定飞行中步长向量的大小维度, 即 u 和 v 的每个元素都是在 d 维空间中随机生成。

为量化算法在每次迭代中的收敛程度, 引入收敛因子 C 来衡量个体相对全局最优解的距离, 执行公式 (5):

$$C = \frac{1}{N} \sum_{i=1}^n \frac{f(P_{best}) - f(P_i)}{f(P_{best})} \quad (5)$$

式中:

N——种群大小, 表示飞行物数量;

P_{best} ——当前找到的全局最优位置;

P_i ——第 i 个飞行物的位置;

C——收敛因子衡量了种群中个体相对于全局最优解的偏离程度。

随着 C 值减小, 说明飞行物越来越接近全局最优解, 种群的多样性逐渐减小。

为增强算法的自适应性, 引入自适应阈值 r, 执行公式 (6):

$$r = r_{\min} + (r_{\max} - r_{\min}) \left(1 - \frac{t}{T}\right)^{\gamma} \quad (6)$$

式中:

$r_{\min}$ 、 $r_{\max}$ ——阈值最大、最小值, t 和 T 分别是当前迭代次数和最大迭代次数;

γ ——自定义调整参数, 控制 r 的递减速度。

自适应阈值的调整使得算法能够在迭代初期更多探索, 而在后期更专注于开发局部最优解。

为避免算法陷入局部最优解, 引入位置变异操作为飞行物的搜索增加随机性, 执行公式 (7):

$$P_{i+1}^{mutated} = P_{i+1} + \delta \cdot \text{randn}(d) \quad (7)$$

式中:

δ ——变异强度系数控制了变异的幅度。

每只飞行物根据其速度和与全局最优解两者间的相对位置来更新自己的位置, 引入随机数 Q, 用来控制搜索的广度和深度。算法进行速度和位置更新执行公式 (8)、(9):

$$V_{i+1} = V_i + (P_i - P_{best}) \times Q \quad (8)$$

$$P_{i+1} = P_i + V_{i+1} \quad (9)$$

通过随机选择是否使用 Lévy 飞行的方式进行位置更新, 执行公式 (10):

$$\text{if } \text{randn} < r: P_{i+1} = \text{levy}(\beta, d) \quad (10)$$

由于融入了随机性, 以增强全局搜索能力并避免陷入局部最优解, 使得算法能够在全局最优解和局部最优解两者间进行有效的平衡。

1.2.5 模型的评价指标和数据分析

该研究设置分类结果的评价指标包括准确率 (Accuracy)、加权查准率 (Weighted Precision)、加权召回率 (Weighted Recall)、加权 F1 分值 (Weighted F1 Score), 对冷鲜羊肉新鲜度多指标分类模型的性能进行评估。这些评价指标的选择基于研究数据中存在的类别不平衡以及对新鲜度分类结果的精确性和可靠性的要求, 旨在确保模型在实际应用中的有效性。

设定 n 是类别数, 若 N_i 是第 i 类的样本数, N 为总样本数。在样本分类时: 真正类 (True Positives, TP) 为正确预测为正的样本数; 假负类 (False Negatives, FN) 为错误预测为负的正样本数; 假正类 (False Positives, FP) 为错误预测为正的负样本数; 真负类 (True Negatives, TN) 为正确预测为负的样本数。

准确率计算公式 (11):

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (11)$$

加权召回率计算公式 (12):

$$Weighted Recall = \sum_{i=1}^n \left(\frac{N_i}{N} \times \frac{TP_i}{TP_i + FN_i} \right) \quad (12)$$

加权查准率计算公式 (13):

$$Weighted Precision = \sum_{i=1}^n \left(\frac{N_i}{N} \times \frac{TP_i}{TP_i + FP_i} \right) \quad (13)$$

加权 F1 分值计算公式 (14):

$$Weighted F1 Score = 2 \times \frac{Weighted Precision \times Weighted Recall}{Weighted Precision + Weighted Recall} \quad (14)$$

加权查准率能够衡量模型对所有类别预测的准确性, 反映模型对各个新鲜度类别的有效识别; 加权召回率则重点考察模型对各类别实际样本的识别能力, 确保模型在检测所有新鲜度类别时的全面性; 加权 F1 分值作为加权查准率和加权召回率的调和平均, 能够综合评估模型在保持精度和全面性方面的整体性能。此外, 引入混淆矩阵, 定量地展示 LightGBM 模型对不同新鲜度类别的分类效果, 并揭示其在特定类别上的不足之处。

2 结果与讨论

2.1 试验室测定结果

在 14 d 内羊肉新鲜度指标 TVB-N、pH、TAC 以及 MPN 的含量随时间的变化趋势, 如图 6 所示。TVB-N 含量和 MPN 在整个过程中的呈现不同的阶段性增长, 而 TAC 和 pH 的增加趋势相对较为平稳。随着储存时间的延长, 需氧菌数量在低温条件下持续缓慢增加, 虽然整体呈现上升趋势, 但并未表现出大幅波动。

在储存初期, 蛋白质分解和挥发性碱性物质的积累较慢, 肉类新鲜度相对稳定。第 4 天、第 10 天时 MPN 和 TVB-N 含量呈现显著增长, 此时微生物的生长和酶促反应开始加速。第 5~6 天和第 7~9 天之间, TVB-N 含量出现了短暂的平稳, 可能是氮化合物的生

产和消耗之间达到了一种暂时的平衡状态, 此外, 由于羊肉中的非蛋白氮源参与组织中复杂的化学反应, 导致了氮化合物的消耗, 在消耗量大于生成量时会造成特定时间段内 TVB-N 含量的短暂下降^[24]。随着储存时间的延长和腐败程度的加深, 微生物代谢产物逐渐累积, TVB-N 和 MPN 在储存后期显著增加, 也表明肉类腐败的加速。因此, TAC 和 pH 的缓慢上升结合 TVB-N 和 MPN 的突变特征, 可以反映低温条件下的抑制效应, 但并不阻止微生物的累积和活性增长。综上可为后续肉类新鲜度的动态监测及分类模型的构建提供重要依据。

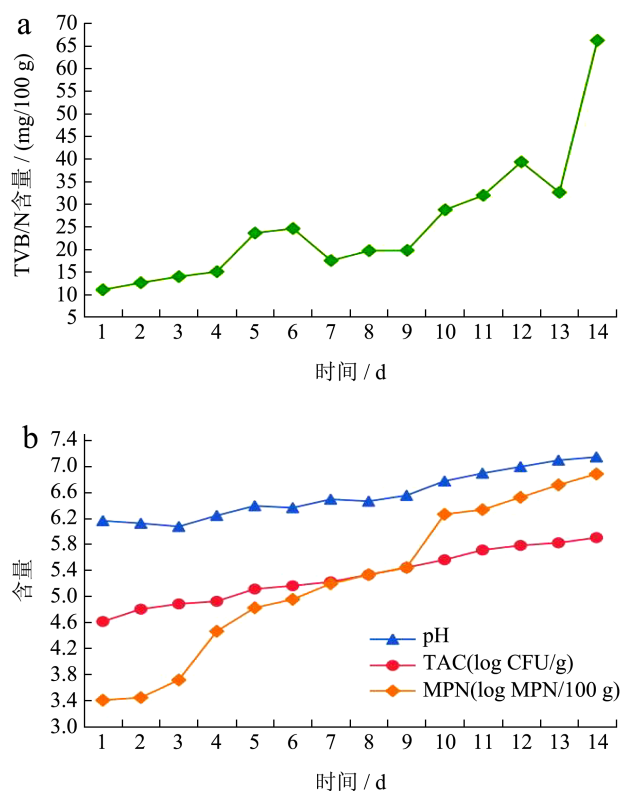


图 6 14 d 内羊肉新鲜度指标的变化

Fig.6 Changes in mutton freshness indicators with in 14 days

注: a 为 TVB-N; b 为 pH、TAC 和 MPN。

2.2 预处理结果与分析

为了得到最佳的预处理方法, 利用三种单一预处理方法及最优的两种预处理方法组合建立 SVR 模型, 设置平均绝对误差 (Mean Absolute Error, MAE)、均方误差 (Mean Squared Error, MSE)、解释方差分数 (Explained Variance Score, EVS) 以及决定系数 (R^2) 等指标综合评价各预处理方法的性能。按照 3:1 划分训练集和测试集。不同预处理方法的评价结果如表 3 所示。

表 3 不同预处理方法建立的模型性能
Table 3 Performance of models built with different preprocessing methods

新鲜度指标	预处理方法	MSE	MAE	EVS	R ²
TVB-N	None	0.305 5	0.380 6	0.698 5	0.694 5
	MSC (best)	0.112 4	0.202 1	0.921 2	0.877 6
	FD	0.123 2	0.212 9	0.915 1	0.896 9
	SG	0.167 3	0.278 7	0.846 3	0.832 7
	MSC-FD	0.068 6	0.187 9	0.932 7	0.931 3
pH	None	0.237 5	0.364 1	0.764 9	0.762 5
	MSC (best)	0.034 4	0.129 9	0.969 5	0.965 6
	FD	0.041 7	0.154 4	0.963 5	0.958 3
	SG	0.079 6	0.200 9	0.925 1	0.920 4
	MSC-FD	0.058 7	0.184 6	0.942 5	0.941 2
TAC	None	0.161 9	0.276 3	0.838 5	0.838 1
	MSC (best)	0.021 5	0.105 6	0.981 2	0.978 5
	FD	0.032 1	0.135 6	0.974 5	0.967 8
	SG	0.074 2	0.191 6	0.928 2	0.925 7
	MSC-FD	0.041 1	0.154 1	0.961 1	0.951 1
MPN	None	0.180 6	0.309 3	0.819 6	0.819 4
	MSC (best)	0.040 6	0.139 6	0.963 8	0.959 4
	FD	0.041 1	0.150 8	0.962 6	0.958 9
	SG	0.077 5	0.197 7	0.925 8	0.922 4
	MSC-FD	0.055 3	0.173 5	0.946 1	0.944 6

多元散射校正正在对数据预处理后各个指标均有最出色的表现，其次为一阶导数。组合最优和次优的预处理方法，结果出现精度下降，可能是在通过多元散射校正去除散射效应来调整基线后使数据的均匀性提高^[7]，影响了一阶导数处理突出变化部分的能力，处理过程中导致某些关键特征被失真。综上，将多元散射校正作为该研究的主要预处理手段，预处理后的光谱反射 DN 值图像如图 7 所示。

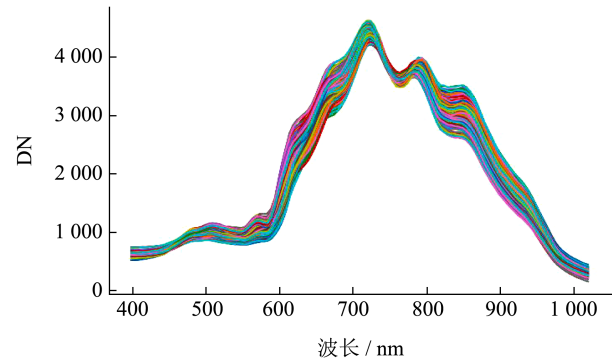


图 7 光谱预处理后 DN 值

Fig.7 The DN values after pretreatment of the spectrum

2.3 特征选择结果与分析

新鲜度数据集经过 MSC 预处理后，分别使用 RFEEA、CRAS 和 SPA 三种特征提取方法对其进行特征波段选择。以 TVB-N、pH、TAC 和 MPN 的含量作为目标变量，利用支持向量回归模型计算均方误差 (MSE) 和决定系数 (R^2)，以评估特征选择方法的效果。MSE 越小，表示模型预测的精度越高； R^2 值越接近 1，表示模型对数据的解释能力越强。不同特征选择的方法对比结果如表 4 所示。

表 4 特征选择对比结果表
Table 4 Comparison results of feature selection

特征选择方法	波段数	MSE _c	R _c ²	MSE _p	R _p ²
RFEEA	35	2.626 3	0.986 9	2.639 4	0.985 4
CRAS	43	5.380 0	0.967 0	5.424 8	0.970 1
SPA	50	11.245 8	0.946 4	13.072 4	0.920 8

RFEEA 更好的识别出了新鲜度光谱数据的特征相关。在特征数量方面，RFEEA 选择了 35 个特征波段，而 CRAS 和 SPA 分别选择了 43 个和 50 个特征波段。综合评价指标，RFEEA 在均方误差和决定系数方面表现均优于 CRAS 和 SPA。具体的，RFEEA 在选择更少的特征波段时，其选择的特征具有较强的代表性和较多的信息密度。相比之下，CRAS 和 SPA 选择了更多的特征波段，但是波段中包含了较多冗余信息，使得其信息密度的降低，导致模型的复杂度增加。特别是 CARS 通过逐步调整特征权重处理光谱数据的连续特征时^[25]，可能因权重调整策略而忽略相关特征。而 SPA 在选择时通过逐步投影的方式来减少特征间的共线性，但易使选择的特征具有冗余，从而导致特征集中包含不必要的波段^[26]。由此，RFEEA 在对冷鲜羊肉新鲜度的光谱特征波段选择中表现出较高的效率。

选择误差值最小的特征数 N 并确定 S_N 中特征波段的组成，其分布如图 8 所示。

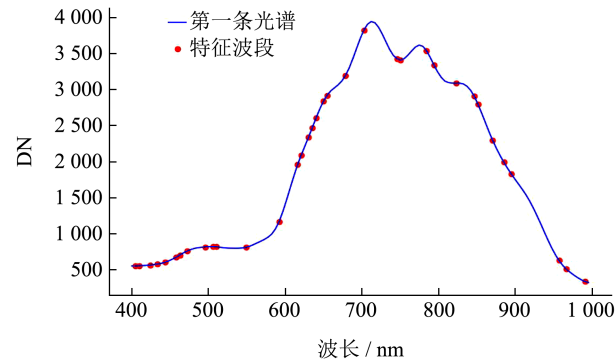


图 8 特征波段

Fig.8 Feature bands

利用 RFEEA 进行特征波段的选择：在 400~500 nm 波段的特征波段为：405、410、424、434、443、458、462、472、496，这些波段可能与样品的光学特性密切相关，其变化可以指示肉类的氧化程度和新鲜度，其中 472.47、496.48 nm 可能与肌红蛋白吸收有关^[27]；500~600 nm 波段的特征波段为：506、510、549、592，其中 549.29、592.51 nm 两个波段可能与脱氧肌红蛋白相关^[28]，脱氧肌红蛋白的增多通常表明肉类的新鲜度降低；600~700 nm 波段的特征波段为：616、621、630、635、640、650、654、678，以及在 700~800 nm 波段的特征波段为：702、746、750、784、794，其中 750.96、784.57 nm 可能与 C-H 键振动的第三泛音有关。最后在 800~1 000 nm 波段的特征波段为：822、846、851、871、885、895、957、967、991，其中 991.04 nm 可能与 O-H 键振动的第二泛音相关^[29]，能够反映样本中水分含量的变化。

综上，RFEEA 能够更有效地处理光谱数据中的复杂特征，且生成了具有更强的代表性和稳定性的特征数据集，有助于提升检测模型在冷鲜羊肉新鲜度分类中的准确性和鲁棒性。

2.4 新鲜度检测分类结果与分析

为了验证多策略改进后 LightGBM 的模型性能。该研究对新鲜度特征数据集分别构建了 XGBoost、SVM、RF、LightGBM 以及单策略（HBA）改进的 LightGBM 等模型。

通过结合混合蝙蝠算法和星鸦优化算法等多策略改进方法来优化 LightGBM 的超参数配置。训练集和测试集分别占实验数据集的 0.7 和 0.3。维度设置为 2，用于优化参数树深和学习率。设置初始种群大小 20、最大迭代次数 50、飞行物的声波调节参数 $Q_{\min}=0$ 、 $Q_{\max}=2$ 、飞行参数 $r=0.5$ （用于决定是否应用 Lévy 飞行扰动的概率）。单策略改进的 LightGBM 模型仅使用 HBA 进行优化，其他参数保持默认设置。多策略改进的 LightGBM 模型与其他模型在冷鲜羊肉新鲜度检测任务中的性能对比结果如表 5 所示。

表 5 基于不同模型的分类性能比较

Table 5 Classification performance comparison of different models

模型	Accuracy	Weighted Precision	Weighted Recall	Weighted F1 Score
多策略改进 LightGBM	0.988 1	0.988 6	0.988 2	0.988 3
单策略改进 LightGBM	0.964 3	0.966 8	0.964 3	0.963 9
LightGBM	0.952 4	0.953 0	0.952 1	0.952 0
XGBoost	0.940 5	0.946 9	0.940 4	0.938 9
SVM	0.928 6	0.935 5	0.928 6	0.925 9
RF	0.916 7	0.928 8	0.916 7	0.913 2

在食品安全的背景下，需要最大程度识别出所有可能不新鲜的羊肉样本，以防止不适宜使用的肉类流入市场，因此较高的召回率至关重要。结合整体的准确率可以直观的反映了不同模型多分类的综合性能。多策略改进的 LightGBM 和单策略改进的 LightGBM，评价指标均优于传统的 LightGBM、XGBoost、SVM 和 RF 分类方法。

对于现有研究^[7]，多策略改进的 LightGBM 模型对羊肉样本进行新鲜度分类时，准确率和召回率都达到了最高的水平，分类精度达到 0.988 1、加权查准率和加权召回率分别为 0.988 6 和 0.988 2，表明改进后模型在正类的样本的识别中表现优异。此外，其加权 F1 分数为 0.988 3，反映出改进后 LightGBM 在精确率和召回率之间达到了良好的平衡。改进后模型准确率相比较于未改进前 LightGBM 模型，提升了 3.57%，加权查准率和加权召回率分别从改进前的 0.944 1 和 0.940 7 提升至 0.988 6 和 0.988 2，加权 F1 分值也提升了 3.63%。而原始 LightGBM 模型分类效果优于 XGBoost、SVM 等基线模型，说明其作为该文的分类模型具有可行性。

从模型的结构特点分析，多策略改进方法在参数寻优过程中能够高效的实现全局和局部的平衡，这使得经其改进后的 LightGBM 更加适应处理高维、稀疏的新鲜度光谱数据。由表 5 所示，改进后的模型有效的抑制了冗余信息和噪声干扰，并增强了对非线性和动态特征的适应能力。因此，其在面对冷鲜羊肉的高光谱数据时，能够深层挖掘新鲜度特征中的关键信息，同时保持较低的误差水平。而现有基线模型在数据处理和参数配置方面缺乏针对性优化手段，易受到光谱数据结构和噪声特征的干扰，因此在面对新鲜度特征数据集时不能达到相同的效果。综上，多策略改进方法显著提升了 LightGBM 模型对羊肉新鲜度的分类性能，证明了该方法的有效性。

表 6 多策略改进的LightGBM在不同规模数据集上的分类性能

Table 6 Classification performance of multi-strategy improved lightGBM on datasets of different scales

数据集	Accuracy	Weighted Precision	Weighted Recall	Weighted F1 Score
280 数据集	0.988 1	0.988 6	0.988 2	0.988 3
840 数据集	0.984 1	0.984 4	0.984 1	0.984 2

此外，为验证方法的泛化性，即其在不同样本集上的稳定性和鲁棒性。使用 840 数据集进行检验，分类性能如表 6 所示，得到两个数据集在多策略改进后模型分类的混淆矩阵如图 9 所示。在实验数据集和检验数据集上，模型对冷鲜羊肉新鲜度等级的各分类指

标稳定在 98.40%~98.90% 之间，仅有少数的新鲜羊肉样本出现分类错误。尽管数据规模明显增加，但是模型的性能仅发生轻微变化。此外，模型的召回率接近 0.99，即对不同新鲜度等级的判别能力较高。相比于现有研究^[16]，改进后的 LightGBM 模型在面对不同数据集规模时，表现出较强的泛化性和鲁棒性。

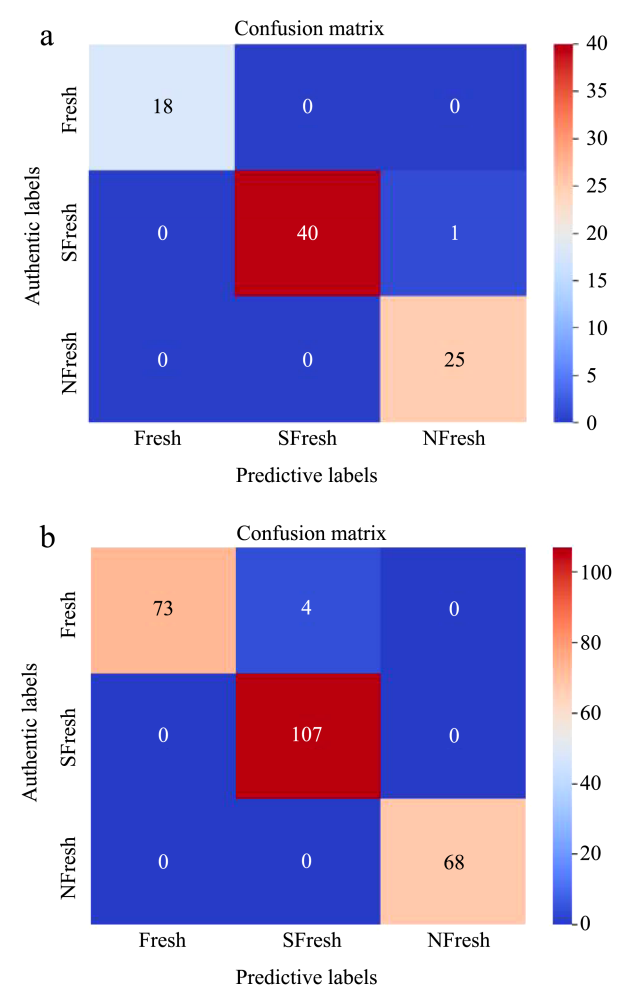


图 9 数据集混淆矩阵

Fig.9 Confusion matrix of dataset

注：a 为实验数据集；b 为检验数据集。

从混淆矩阵 a 上可以看到，在实验数据集中，真实标签为次新鲜的一个样本在分类时被错误的分类为不新鲜，分析其原因，可能是由于冷鲜羊肉样本在新鲜度等级更迭的过程中腐败程度不均匀，导致次新鲜羊肉样本在提取 ROI 时受到不纯净数据污染。部分处于次新鲜末期的样本区域腐败程度已经超过了次新鲜等级，由此被模型错误分类为不新鲜。因此，如何更高效的进行 ROI 提取，在未来研究中需要更多的关注。

3 结论

该研究利用高光谱技术和多策略（HBA 和 NOA）

改进的 LightGBM 模型对冷鲜羊肉的新鲜度进行检测。实验采集了 14 日内冷鲜储存的羊肉样本高光谱数据（400~1 000 nm），并测定期间 4 个新鲜度指标。对比多种预处理方法后，采用多元散射校正作为光谱数据的预处理手段。然后，利用 RFFEA 方法对预处理后的新鲜度数据集进行特征选择，以深入研究羊肉在储存时间内新鲜度变化趋势与高光谱波段的相关性，经过验证后得到包含 35 个特征波段的新鲜度特征数据集。最后，将其输入基于 HBA 和 NOA 多策略改进的 LightGBM 模型中，以检测不同新鲜度样本，实现高精度分类。结果显示，多策略改进的 LightGBM 模型能够有效的将冷鲜羊肉按照新鲜度分成 3 类，在准确率、加权精度、加权召回率和加权 F1 分值上较其他方法均表现最佳，分别达到了 0.988 1、0.988 6、0.988 2 和 0.988 3，显著优于单策略（HBA）改进的 LightGBM 及其它分类模型。此外，通过增加采集每日样本的高光谱图像，获取到更大规模的检验数据集，模型在不同数据集的分类准确率稳定在 98.40%~98.90%，表现出良好的泛化能力和鲁棒性。

未来研究可探索结合更多群智能优化算法，以提升模型在多维特征选择中的表现。此外，考虑将改进后的 LightGBM 方法嵌入自动化检测设备中，以实现冷链运输过程中羊肉新鲜度的实时监控，推动其工业化应用。

参考文献

[1] WANG C, WANG S, HE X, et al. Combination of spectra and texture data of hyperspectral imaging for prediction and visualization of palmitic acid and oleic acid contents in lamb meat [J]. Meat Science, 2020, 169: 108194.

[2] CHENG L, LIU G, HE J, et al. Development of a novel quantitative function between spectral value and metmyoglobin content in Tan mutton [J]. Food Chemistry, 2021, 342: 128351.

[3] TOOMIK E, ROOD L, BOWMAN J P, et al. Microbial spoilage mechanisms of vacuum-packed lamb meat: A review [J]. International Journal of Food Microbiology, 2023, 387: 110056.

[4] LU Y Z, SAEYS W, KIM M, et al. Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress [J]. Postharvest Biology and Technology, 2020, 170: 111318.

[5] CHENG J, SUN J, SHI L, et al. An effective method fusing electronic nose and fluorescence hyperspectral imaging for the detection of pork freshness [J]. Food Bioscience, 2024, 59: 103880.

[6] CHENG J H, SUN J, YAO K S, et al. Generalized and hetero two-dimensional correlation analysis of hyperspectral imaging combined with three-dimensional convolutional neural network for evaluating lipid oxidation in pork [J]. Food Control, 2023, 153: 109940.

[7] 徐子洋,姜新华,白洁,等.基于多标记深度森林算法的冷鲜羊肉新

- 鲜度无损检测方法[J].光谱学与光谱分析,2024,44(2):580-587.
- [8] YU Y W, CHEN W L, ZHANG H W, et al. Discrimination among fresh, frozen-stored and frozen-thawed beef cuts by hyperspectral imaging [J]. Foods, 2024, 13(7): 973.
- [9] WAN G L, FAN S X, LIU G S, et al. Fusion of spectra and texture data of hyperspectral imaging for prediction of myoglobin content in nitrite-cured mutton [J]. Food Control, 2023, 144: 109332.
- [10] MALIK A, RAM B, ARUMUGAM D, et al. Predicting gypsum tofu quality from soybean seeds using hyperspectral imaging and machine learning [J]. Food Control, 2024, 160: 110357.
- [11] WU Y L, LI X L, XU L J, et al. Counterfeit detection of bulk Baijiu based on fluorescence hyperspectral technology and machine learning [J]. Journal of Food Measurement and Characterization, 2024, 18(4): 3032-3041.
- [12] CHEN B Y, SHI B L, GONG J R, et al. Quality detection and variety classification of pecan seeds using hyperspectral imaging technology combined with machine learning [J]. Journal of Food Composition and Analysis, 2024, 131: 106248.
- [13] MAHSA H, ABBAS M, REZA G, et al. A novel scheme for mapping of MVT-Type Pb-Zn prospectivity: lightGBM, a highly efficient gradient boosting decision tree machine learning algorithm [J]. Natural Resources Research, 2023, 32(6): 2417-2438.
- [14] 高兵,郑雅,秦静,等.基于麻雀搜索算法和改进粒子群优化算法的网络入侵检测算法[J].计算机应用,2022,42(4):1201-1206.
- [15] XIE B S, LI F, LI H, et al. Enhanced internet of things security situation assessment model with feature optimization and improved SSA-LightGBM [J]. Mathematics, 2023, 11, 3617.
- [16] XINHUA J, HERU X, LINA Z, et al. Nondestructive detection of chilled mutton freshness based on multi-label information fusion and adaptive BP neural network [J]. Computers and Electronics in Agriculture, 2018, 155: 371-377.
- [17] GB/5009.228-2016,食品安全国家标准 食品中挥发性盐基氮的测定[S].
- [18] GB/5009.237-2016,食品安全国家标准 食品PH值的测定[S].
- [19] GB/4789.2-2022,食品安全国家标准 食品微生物学检验 菌落总数测定[S].
- [20] GB/4789.3-2016,食品安全国家标准 食品微生物学检验 大肠杆菌计数[S].
- [21] GB/2707-2016,食品安全国家标准 鲜(冻)畜、禽产品[S].
- [22] GAN C, CAO W H, LIU K Z, et al. A new hybrid bat algorithm and its application to the ROP optimization in drilling processes [J]. IEEE Transactions on Industrial Informatics, 2019, 16(12): 7338.
- [23] MOHAMED A, REDA M, MOHAMMED J, et al. Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems [J]. Knowledge-Based Systems, 2023, 262: 110248.
- [24] FAN N Y, MA X, LIU G S, et al. Rapid determination of TBARS content by hyperspectral imaging for evaluating lipid oxidation in mutton [J]. Journal of Food Composition and Analysis, 2021, 103: 104110.
- [25] ZHONG L, GUO X, DING M, et al. SHAP values accurately explain the difference in modeling accuracy of convolution neural network between soil full-spectrum and feature-spectrum [J]. Computers and Electronics in Agriculture, 2024, 217: 108627.
- [26] 张冬妍,马苗源,黄莹,等.基于高光谱图像特征融合的榛子水分含量测定[J].现代食品科技,2024,40(5):274-281.
- [27] CIFUNI F G, CONTÒ M, FAILLA S, et al. Potential use of visible reflectance spectra to predict lipid oxidation of rabbit meat [J]. Journal of Food Engineering, 2016, 169: 85-90.
- [28] MAHMUDUL M H, ARIF M M C, CHYNGYZ E, et al. Application of Vis-NIR and SWIR spectroscopy for the segregation of bison muscles based on their color stability [J]. Meat Science, 2022, 108: 108774.
- [29] PU H, KAMRUZZAMAN M, SUN D, et al. Selection of feature wavelengths for developing multispectral imaging systems for quality, safety and authenticity of muscle foods-a review [J]. Trends in Food Science & Technology, 2015, 45(1): 86-104.