

# 基于近红外光谱和 FOA-RF 的猪肉新鲜度检测

张 丽<sup>1,2</sup> 李志辉<sup>3</sup> 马明星<sup>1,2</sup>

(1. 河南工业职业技术学院, 河南 南阳 473000; 2. 河南省柔性制造工程研究中心, 河南 南阳 473000;

3. 郑州大学, 河南 郑州 450001)

**摘要:** [目的] 实现猪肉新鲜度的快速、无损且高精度检测, 解决传统检测方法效率低、破坏性强及单一模型预测精度不足的问题。[方法] 提出一种近红外光谱(NIRS)结合果蝇优化算法(FOA)改进随机森林(RF)的猪肉新鲜度检测模型。以总挥发性盐基氮(TVB-N)质量分数作为猪肉新鲜度评价指标, 首先采集不同贮藏阶段猪肉样品的近红外光谱数据(扫描范围 1 000~1 800 nm), 通过多元散射校正(MSC)与一阶导数结合的预处理方法消除光谱噪声与基线漂移; 采用 FOA 优化 RF 的关键超参数(决策树数量、最小叶子节点样本数、最大特征数), 构建果蝇优化算法改进随机森林(FOA-RF)预测模型。[结果] 在各类预测模型中, FOA-RF 模型对猪肉 TVB-N 质量分数的估算精度最高。该模型在预测集上的均方根误差(RMSEP)仅为 1.582 mg/100 g; 同时, 其预测集相关系数( $R_p$ )为 0.978, 决定系数( $R_p^2$ )为 0.956, 残差预测偏差(RPDp)也高达 4.723, 显著优于其他对比模型。相比之下, 传统偏最小二乘回归(PLSR)、未优化随机森林以及网格搜索优化随机森林(GS-RF)等模型的综合预测性能均不及 FOA-RF 模型。[结论] 该方法高效、精准, 可满足肉类工业现场快速检测需求。

**关键词:** 近红外光谱; 果蝇优化算法; 随机森林; 猪肉新鲜度; 总挥发性盐基氮; 无损检测

## Pork freshness monitoring based on near-infrared spectroscopy and random forest improved by fruit fly optimization algorithm

ZHANG Li<sup>1,2</sup> LI Zhihui<sup>3</sup> MA Mingxing<sup>1,2</sup>

(1. Henan Polytechnic Institute, Nanyang, Henan 473000, China; 2. Henan Engineering Research Center of Flexible Manufacturing, Nanyang, Henan 473000, China; 3. Zhengzhou University, Zhengzhou, Henan 450001, China)

**Abstract:** [Objective] To achieve rapid, non-destructive, and high-precision monitoring of pork freshness, addressing the low efficiency, high destructiveness, and insufficient prediction accuracy of single models in conventional monitoring. [Methods] A pork freshness monitoring model was proposed based on near-infrared spectroscopy (NIRS) combined with random forest (RF) improved by the fruit fly optimization algorithm (FOA). With the total volatile basic nitrogen (TVB-N) content as the freshness indicator, near-infrared spectral data of pork samples at different storage stages are collected (scanning range: 1 000~1 800 nm). Spectral noise and baseline drift are eliminated via a preprocessing method combining multiplicative scatter correction (MSC) and first-derivative transformation. Then, FOA is employed to optimize key hyperparameters (number of decision trees, minimum leaf node sample size, and maximum number of features) of RF to construct the FOA-RF model. [Results] Among all the prediction models evaluated, the FOA-RF model demonstrates the highest accuracy for predicting pork TVB-N content. The preprocessing method combining MSC and first-derivative transformation effectively enhances the quality of the spectral data. The FOA-RF model achieves a root mean square error of prediction (RMSEP) of only 1.582 mg/100 g, a correlation coefficient of prediction ( $R_p$ ) of 0.978, a coefficient of determination of prediction ( $R_p^2$ ) as high as 0.956, and a residual prediction deviation of prediction (RPDp) of 4.723, significantly outperforming the other comparative models. The overall predictive performance of

基金项目: 河南省科技攻关项目(编号: 252102210224, 252102210008)

通信作者: 张丽(1979—), 女, 河南工业职业技术学院副教授, 硕士。E-mail: bkgfs78@126.com

收稿日期: 2025-09-01 改回日期: 2025-11-22

引用格式: 张丽, 李志辉, 马明星. 基于近红外光谱和 FOA-RF 的猪肉新鲜度检测[J]. 食品与机械, 2025, 41(12): 51-58.

Citation: ZHANG Li, LI Zhihui, MA Mingxing. Pork freshness monitoring based on near-infrared spectroscopy and random forest improved by fruit fly optimization algorithm[J]. Food & Machinery, 2025, 41(12): 51-58.

partial least squares regression (PLSR), the un-optimized RF model, and the grid search-optimized random forest (GS-RF) model is inferior to that of the FOA-RF model. **[Conclusion]** The method proposed in this study provides an efficient and accurate new approach for non-destructive monitoring of pork freshness, meeting the demand for rapid testing in the meat industry.

**Keywords:** near-infrared spectroscopy; fruit fly optimization algorithm; random forest; pork freshness; total volatile basic nitrogen; non-destructive monitoring

猪肉在中国居民膳食结构中占据重要地位,2024 年猪肉消费量占肉类总消费量的 56.3%,其新鲜度直接关系到消费者健康、食品安全与市场经济效益<sup>[1]</sup>。在屠宰、冷链运输及零售等环节中,猪肉易受微生物繁殖、酶解作用及环境温湿度变化的影响,导致蛋白质逐渐降解产生氨、三甲胺等挥发性含氮化合物,进而引发品质劣变<sup>[2]</sup>。

在传统的猪肉新鲜度检测体系中,感官评定(如观察色泽、嗅闻气味、检测弹性)与化学分析方法长期占据主导地位。然而,前者易受主观判断影响,后者则存在流程繁琐、耗时长、需破坏样本等局限,无法适应当前肉类产业对“实时、无损、批量筛查”的技术要求<sup>[3]</sup>。在此背景下,近红外光谱分析技术因其高效、无创伤及多指标同步检测能力,逐渐发展为食品品质监控领域的有效手段<sup>[4]</sup>。该技术通过检测样品中 C—H、O—H、N—H 等官能团的振动吸收光谱,建立光谱信息与化学指标之间的关联模型,从而实现品质的间接预测<sup>[5]</sup>。在肉类新鲜度检测中,该技术已取得显著进展。例如,Leng 等<sup>[6]</sup>采用近红外光谱结合支持向量回归(SVR)模型预测牛肉挥发性盐基氮(TVB-N)含量,但 SVR 模型对核函数参数敏感,且在小样本数据集中泛化能力不足。王冬等<sup>[7]</sup>进一步指出,光谱预处理方法与特征选择算法是提升模型精度的关键。在猪肉检测方面,田文强等<sup>[8]</sup>采用了多指标融合方法,但未充分考虑算法超参数优化;方瑶等<sup>[9]</sup>对比了不同预处理方法对 TVB-N 含量预测的影响,却未深入优化预测模型结构。常用的建模方法如偏最小二乘回归(PLSR)在处理高维光谱数据时易受非线性干扰且对异常值敏感,预测精度常受限<sup>[10-11]</sup>。随机森林(RF)通过构建多棵决策树来降低过拟合风险,具有抗噪声能力强、适应非线性数据的优势,已被逐步应用于肉类新鲜度预测<sup>[12]</sup>。然而,RF 的预测性能高度依赖超参数(如决策树数量、最小叶子节点样本数)的选择,传统经验调参或网格搜索(GS)方法存在效率低、易陷入局部最优的问题<sup>[13]</sup>。

针对超参数优化难题,各类智能优化算法被广泛引入。网格搜索计算量大,在多参数优化中易出现“维度灾难”<sup>[14]</sup>;粒子群优化(PSO)算法虽收敛速度快,但易陷入局部最优<sup>[15]</sup>。果蝇优化算法是 Pan 于 2011 年提出的一种模拟果蝇觅食行为的全局优化算法,具有原理简单、参数少、收敛速度快的特点,在函数优化和神经网络训练等领域展现出优异性能<sup>[16]</sup>。将果蝇优化算法(FOA)用于 RF

超参数寻优,可有效规避 GS 的维度灾难与局部最优问题,进一步提升模型预测精度与训练效率。Lan 等<sup>[17]</sup>采用鲸鱼优化算法改进随机森林-AdaBoost 集成(RF-AdaBoost)模型,显著提升了预测精度;孙康慧等<sup>[18]</sup>基于 FOA 优化 LightGBM 算法,有效降低了预测误差。然而,目前将 FOA 应用于近红外光谱结合 RF 的猪肉新鲜度检测研究尚未见报道,相关技术体系仍有待完善。因此,研究拟以猪肉 TVB-N 质量分数为核心检测指标,整合近红外光谱技术与果蝇优化算法改进随机森林(FOA-RF)算法,构建猪肉新鲜度无损检测模型,旨在为猪肉品质快速检测提供新方法。

## 1 材料与方法

### 1.1 猪肉样品准备

选用市售新鲜猪背最长肌作为试验样本,原料来源于 6 月龄健康商品猪(杜长大三元杂交品种)。样品采购后于 1 h 内运送至实验室,全程采用 4℃保温箱贮藏。在实验室中,剔除样本表面筋膜与脂肪组织后,沿垂直于肌纤维方向将其切割为直径 40 mm、厚度 10 mm 的圆柱形试样,共计制备 120 个有效样品。所有样品于 4℃冷藏环境中保存,连续贮藏 7 d,每日随机抽取 15~20 个样本,同步采集光谱数据并测定 TVB-N 质量分数,以确保覆盖 4.8~38.2 mg/100 g 范围内不同新鲜度等级的数据分布。

### 1.2 主要试剂与仪器

硼酸、氧化镁、盐酸、甲基红、亚甲基蓝:分析纯,国药集团化学试剂有限公司;

超纯水:电阻率 18.2 MΩ·cm,实验室自制;

近红外光谱仪:SupNIR-2700 型,聚光科技(杭州)股份有限公司;

凯氏定氮仪:K9840 型,海能未来技术集团股份有限公司;

电子天平:GN1324 型,福州民桥精密仪器有限公司;

绞肉机:JRJ300-I 型,九阳股份有限公司;

数显恒温水浴锅:HH-4 型,常州澳华仪器有限公司。

### 1.3 近红外光谱采集

光谱采集前进行仪器预热与校准:开启近红外光谱仪,预热 30 min 以保证光源稳定性;采用标准白板(聚四氟乙烯材质)进行基线校正,消除环境光与仪器漂移干扰。将猪肉样品置于样品池中,保持温度 25℃(室温控

制),从样品横截面采集光谱,每个样品重复扫描3次,取平均光谱作为最终数据,以吸光度形式保存。光谱采集过程中,每隔10个样品重新校正白板,确保数据一致性。

#### 1.4 TVB-N质量分数测定

按GB 5009.228—2016《食品安全国家标准 食品中挥发性盐基氮的测定》中自动凯氏定氮仪法执行。每个样品平行测定3次,取平均值作为最终TVB-N质量分数。

表1 校正集与预测集猪肉样品TVB-N质量统计

Table 1 TVB-N content statistics in pork samples of calibration and prediction sets

样本集	样本数	最大值/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	最小值/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	平均值/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	标准差/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	变异系数/%
校正集	84	39.25	4.82	18.36	8.15	44.4
预测集	28	36.78	6.25	17.92	7.83	43.7
总集	112	39.25	4.82	18.27	8.07	44.2

由表1可知,校正集与预测集的TVB-N质量分数范围、平均值与标准差均较为接近,变异系数均 $>40\%$ ,表明样本覆盖了猪肉从新鲜( $\text{TVB-N} < 10 \text{ mg}/100 \text{ g}$ )到腐败( $\text{TVB-N} > 30 \text{ mg}/100 \text{ g}$ )的全阶段,数据分布合理,可用于模型构建与验证。

#### 1.6 光谱预处理

近红外光谱在采集过程中易受多种因素干扰,包括样品颗粒度、表面散射及仪器噪声等,因此需通过预处理手段提升建模精度。研究评估多元散射校正(MSC)、标准正态变量变换(SNV)、一阶导数和二阶导数4种常用预处理方法的建模效果,并探讨组合预处理方式的优化潜力。预处理效果通过信噪比(SNR)和校正集均方根误差(RMSEC)两个指标进行量化:SNR反映光谱噪声水平,其值按式(1)计算,值越高代表噪声抑制效果越好;RMSEC体现预处理后数据与参考值的拟合程度,其值越低表明相关性越强。

$$S_{\text{NR}} = 10 \times \lg \left( \frac{\bar{I}^2}{\sigma^2} \right), \quad (1)$$

式中:

$S_{\text{NR}}$ ——信噪比;

$\bar{I}$ ——吸光度平均值;

$\sigma^2$ ——吸光度的方差。

## 2 模型构建与优化

### 2.1 RF

RF是基于Bagging集成策略的机器学习算法,通过以下步骤构建:

(1) 自助采样(bootstrap):从校正集中随机抽取 $n$ 个样本(有放回),构建 $n$ 个决策树的训练集。

### 1.5 样本集划分

采用肯纳德—斯通(KS)算法将112组有效数据按3:1的比例划分为校正集(84个样本)与预测集(28个样本)。KS算法基于样本在光谱空间中的欧氏距离,优先选择具有代表性的样本进入校正集,可确保校正集与预测集的分布一致性,避免因样本分布不均导致的模型过拟合<sup>[19]</sup>。校正集与预测集的TVB-N质量分数统计结果见表1。

(2) 决策树构建:每个决策树采用随机特征选择(每次分裂时仅考虑部分特征),以均方误差(MSE)最小化为分裂准则,不进行剪枝。

(3) 集成预测:对于回归问题,RF的输出为所有决策树预测结果的平均值。

(4) 模型评估:采用袋外数据(OOB,未被采样的样本)计算OOB误差,评估模型泛化能力。

RF的关键超参数包括:决策树数量( $n_{\text{estimators}}$ )、最小叶子节点样本数( $\text{min\_samples\_leaf}$ )、最大特征数( $\text{max\_features}$ )。超参数的选择直接影响模型性能: $n_{\text{estimators}}$ 过少易导致欠拟合,过多则增加计算成本; $\text{min\_samples\_leaf}$ 过小易过拟合,过大则模型欠拟合; $\text{max\_features}$ 影响特征选择的随机性,需合理设置以平衡偏差与方差<sup>[20]</sup>。

### 2.2 FOA优化RF模型

FOA是模拟果蝇觅食行为的智能优化算法,其核心思想是:果蝇通过嗅觉搜索食物(全局搜索),找到食物气味源后通过视觉精确定位(局部搜索),最终实现最优解寻优<sup>[21]</sup>。采用FOA优化RF的3个关键超参数,具体步骤:

Step1:确定优化变量与搜索范围。根据文献调研与预试验<sup>[22]</sup>,对RF模型的3个关键超参数进行了初步范围测试。预试验结果表明,当决策树数量 $<50$ 时模型欠拟合, $>300$ 时计算成本显著增加而性能提升有限;最小叶子节点样本数在1~10范围内模型表现稳定;最大特征数在光谱特征数开根号 $\sqrt{d}$ 到 $d$ 之间可有效平衡模型的偏差与方差。因此,最终确定超参数搜索范围决策树数量 $x_1 \in [50 \ 300]$ ,最小叶子节点样本数 $x_2 \in [1 \ 10]$ ,最大特征数 $x_3 \in [\sqrt{d} \ d]$ , $d$ 为预处理后的光谱特征数(研究中

$d=801$ , 对应 1 000~1 800 nm 波长范围内的 801 个数据点), 故最大特征数  $x_3 \in [29\ 801]$ 。

**Step2: FOA 参数设置。**种群规模 (population size) 30, 最大迭代次数 (max iteration) 100, 初始搜索半径 (initial step) 10, 嗅觉搜索步长衰减系数 0.9 (每迭代 1 次, 步长乘 0.9, 实现全局搜索向局部搜索过渡)。

**Step3: 适应度函数设计。**以 RF 模型的 RMSEC 最小化为目标, 适应度函数如式 (2) 所示。

$$f_{\text{RMSEC}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i,\text{meas}} - y_{i,\text{pred}})^2}, \quad (2)$$

式中:

$f_{\text{RMSEC}}$ ——适应度函数;

$n$ ——训练集样本数量;

$y_{i,\text{meas}}$ ——第  $i$  个样本的 TVB-N 含量实际值;

$y_{i,\text{pred}}$ ——第  $i$  个样本的 TVB-N 含量预测值。

FOA 优化 RF 模型的算法流程:

**步骤 1:** 初始化果蝇种群位置 (每个个体对应一组 RF 超参数), 随机生成  $N$  组超参数 ( $x_1, x_2, x_3$ ), 确保其在搜索范围内。

**步骤 2:** 计算每个个体的适应度值 (RMSEC), 筛选出当前最优个体 (适应度值最小) 及其位置。

**步骤 3:** 基于当前最优个体位置, 进行嗅觉搜索。新个体位置 = 当前最优位置 + 随机步长  $\times$  搜索半径。

**步骤 4:** 计算新个体的适应度值, 若新个体适应度优于当前最优, 则更新最优个体。

**步骤 5:** 减小搜索半径, 重复步骤 3~4, 直至达到最大迭代次数。

**步骤 6:** 输出最优超参数组合, 用于构建 FOA-RF 模型。

### 3 结果与分析

为验证 FOA-RF 模型的优越性, 构建 6 种对比模型:

**偏最小二乘回归 (PLSR) 模型:** 传统线性建模方法, 采用交叉验证 (留一法) 确定最优主成分数 (最优主成分数为 8)。

**未优化 RF 模型:** 超参数采用 MATLAB 库默认值—— $n_{\text{estimators}}$  为 100,  $\min_{\text{samples\_leaf}}$  为 1,  $\max_{\text{features}}$  为  $\sqrt{d}$  (约 29)。

**网格搜索优化 RF (GS-RF) 模型:** 采用网格搜索优化 RF 超参数, 搜索范围与 FOA 一致; 网格步长:  $n_{\text{estimators}}$  为 50 (50, 100,  $\dots$ , 300),  $\min_{\text{samples\_leaf}}$  为 1 (1, 2,  $\dots$ , 10),  $\max_{\text{features}}$  为 30 (29, 59,  $\dots$ , 801); 交叉验证折数 5。

**贝叶斯优化 (BO) RF 模型:** 初始点数量 30, 最大迭代次数 100。

**鲸鱼优化算法 (WOA) 优化 RF 模型:** 种群规模 30, 最

大迭代次数 100, 收敛常数  $a$  从 2 线性递减至 0。

**1D-CNN 模型:** 采用一维卷积神经网络, 包含 2 个卷积层 (卷积核大小分别为 5 和 3, 滤波器数量分别为 64 和 32)、最大池化层 (池化大小为 2)、全连接层 (128 个神经元) 和输出层。使用 Adam 优化器, 学习率 0.001, 批处理大小 16, 训练轮数 200。FOA-RF 与 BO-RF 和 WOA-RF 优化过程收敛曲线对比图如图 1 所示。

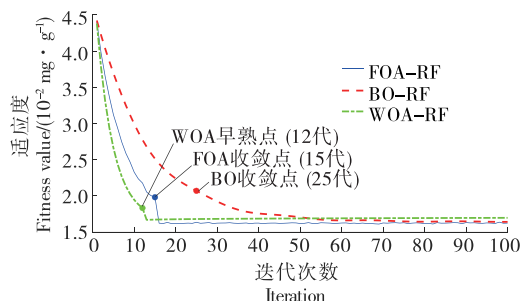


图 1 收敛曲线对比图

Figure 1 Comparison of convergence curves

通过对 3 种优化算法的收敛曲线进行综合分析可知: 在收敛速度与效率方面, FOA-RF 表现最为出色, 它在迭代 15 代后便迅速稳定于最优适应度值 (1.623 mg/100 g), 展现出高效的全局搜索能力; 相比之下, BO-RF 收敛最为缓慢, 直至 25 代才趋于稳定, 表明其在高维参数空间中搜索效率较低; 而 WOA-RF 虽在前期 (前 12 代) 收敛迅速, 但随后陷入平台期, 表现出典型的早熟收敛现象。在全局寻优能力上, FOA-RF 获得了三者中的最佳结果, 证明了其强大的全局优化能力和有效规避局部最优的性能; WOA-RF 因后期开发能力不足而最终陷于局部最优; BO-RF 则受限于其概率代理模型, 在有限迭代内未找到全局最优解。此外, 在算法稳定性方面, FOA-RF 在收敛后曲线最为平稳, 波动极小; BO-RF 则在整个过程中出现多次小幅振荡, 显示出一定的搜索随机性。综上所述, FOA 在收敛速度、全局寻优能力和稳定性方面均具备明显优势。不同预处理方法的效果对比见表 2。

由表 2 可知, MSC+一阶导数组组合预处理方法的 SNR 最高 (35.2), RMSEC 最低 (1.892 mg/100 g), 校正集  $R_p^2$  最高 (0.913), 其效果显著优于单一预处理方法。因此, 选择 MSC+一阶导数作为最终光谱预处理方法。

#### 3.1 不同模型的预测性能对比

4 种模型 (PLSR、未优化 RF、GS-RF、FOA-RF、1D-CNN、BO-RF 和 WOA-RF) 的预测性能指标见表 3, 预测值与实测值的拟合图如图 2 所示。由表 3 和图 2 可知: ① PLSR 模型性能最差。其预测集均方根误差 (RMSEP) 高达 2.863 mg/100 g,  $R_p^2$  仅为 0.726, 相对分析误差 (RPDP)



表 2 不同光谱预处理方法效果对比

Table 2 Comparison of different spectral preprocessing methods

预处理方法	SNR	RMSEC/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	校正集决定 系数( $R_p^2$ )
原始光谱	28.3	2.568	0.852
MSC	32.6	2.145	0.887
SNV	31.8	2.213	0.879
一阶导数	30.5	2.326	0.868
二阶导数	29.7	2.415	0.859
MSC+一阶导数	35.2	1.892	0.913
SNV+一阶导数	34.1	1.957	0.905

为 2.741 ( $<3.0$ )。说明线性模型难以捕捉猪肉光谱数据与 TVB-N 质量分数间的复杂非线性关系,泛化能力不足。

② 未优化 RF 模型性能提升显著。相较于 PLSR, RF 模型的 RMSEP 降低至 2.215 mg/100 g (降低了 22.6%),  $R_p^2$  提升至 0.812 (提升了 11.8%), RPDp 达到 3.535 ( $>3.0$ )。说明 RF 的集成学习特性可有效处理非线性数据,但默认超参数未达到最优,模型仍有优化空间。③ GS-RF 模型性能进一步优化。通过 GS 优化超参数, GS-RF 的 RMSEP 降至 1.826 mg/100 g,  $R_p^2$  提升至 0.893, RPDp 达到 4.288。说明超参数优化对 RF 模型性能至关重要,但 GS 的“暴力搜索”特性导致其在高维参数空间中易陷入局部最优,优化效果受限。④ ID-CNN 模型表现中等。ID-CNN 的 RMSEP 为 1.945 mg/100 g,  $R_p^2$  为 0.861, RPDp 为 4.025。其性能优于未优化 RF 和 PLSR,但略低于优化后的 RF 变体,表明卷积神经网络具有一定的非线性建模能力,但可能受限于模型结构或训练策略。⑤ BO-RF 和 WOA-RF 模型性能优异。BO-RF 的 RMSEP 为 1.735 mg/100 g,  $R_p^2$  为 0.925, RPDp 为 4.512; WOA-RF 的 RMSEP 为 1.769 mg/100 g,  $R_p^2$  为 0.918, RPDp 为 4.425。两者均通过智能优化算法提升了 RF 模型的性能,其中 BO-RF 的  $R_p^2$  和

RPDp 略高于 WOA-RF, 显示贝叶斯优化在超参数调优中的有效性。⑥ FOA-RF 模型性能最优: 其 RMSEP 最低 (1.582 mg/100 g),  $R_p$  和  $R_p^2$  最高 (分别为 0.978 和 0.956), RPDp 也最高 (4.723)。相较于 GS-RF, FOA-RF 的 RMSEP 降低 13.4%,  $R_p^2$  提升 7.1%, RPDp 提升 10.2%。这是因为果蝇优化算法 (FOA) 通过全局搜索与局部搜索相结合的策略, 可高效找到 RF 超参数的全局最优解, 从而最大化模型预测精度与泛化能力。

### 3.2 FOA-RF 模型的稳定性验证

3.2.1 交叉验证 对校正集进行 5 折交叉验证, 计算每次交叉验证的 RMSEC 与  $R_c^2$ , 结果见表 4。5 折交叉验证的 RMSEC 平均值为 1.626 mg/100 g, 标准差仅为 0.012,  $R_c^2$  平均值为 0.887, 标准差为 0.003, 表明模型对校正集数据的拟合稳定性优良。

3.2.2 重复性试验 重复进行 10 次模型训练 (每次训练时重新划分校正集与预测集, 比例仍为 3:1), 计算 RMSEP 与  $R_p^2$  的平均值与标准差, 结果见表 5。RMSEP 的平均值为 1.593 mg/100 g, 变异系数仅为 1.76%,  $R_p^2$  的变异系数为 0.52%, 表明模型在不同样本划分下的预测性能稳定, 受随机因素影响小。

### 3.3 关键光谱特征分析

基于袋外数据均方误差减少量 (mean decrease MSE) 计算的 FOA-RF 模型特征重要性得分如图 3 所示。

根据图 3 所示, 特征重要性得分较高的关键波段主要集中于 4 个区域:

(1) 1 000~1 180 nm 波段: 该区域主要对应 C—H 键的三级泛音振动以及 O—H 键的伸缩振动, 反映了猪肉中水分和脂肪的初期变化。随着贮藏时间延长, 水分蒸发和脂肪氧化导致该区域光谱响应发生变化, 使其成为 TVB-N 质量分数预测的重要特征区域。

(2) 1 180~1 230 nm 波段: 该区域对应 C—H 伸缩振动的第二泛音, 主要来源于猪肉中脂肪与蛋白质所含的

表 3 猪肉 TVB-N 质量分数预测性能对比

Table 3 Comparison of predictive performance for pork TVB-N content

模型	校正集			预测集			
	RMSEC/( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	相关系数( $R_c^2$ )	决定系数( $R_c^2$ )	RMSEP/( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	$R_p$	$R_p^2$	RPDp
PLSR	1.925	0.908	0.824	2.863	0.852	0.726	2.741
RF	1.783	0.925	0.855	2.215	0.901	0.812	3.535
GS-RF	1.658	0.938	0.880	1.826	0.945	0.893	4.288
ID-CNN	1.701	0.931	0.867	1.945	0.928	0.861	4.025
BO-RF	1.642	0.941	0.885	1.735	0.962	0.925	4.512
WOA-RF	1.658	0.938	0.880	1.769	0.958	0.918	4.425
FOA-RF	1.623	0.942	0.887	1.582	0.978	0.956	4.723

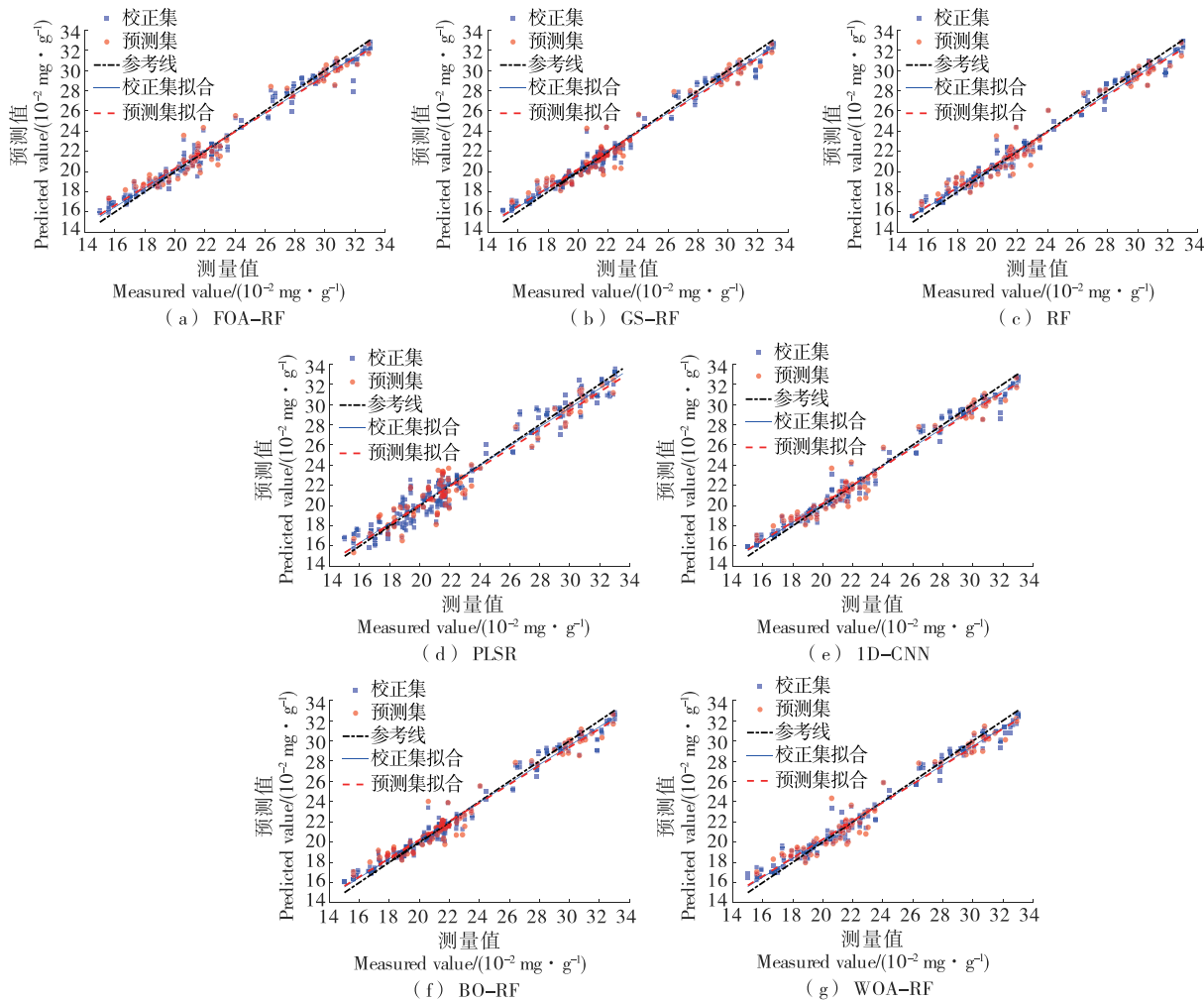


图 2 不同模型预测性能对比

Figure 2 Comparison of predictive performance among different models

表 4 FOA-RF 模型 5 折交叉验证结果

Table 4 Results of 5-fold cross-validation for the FOA-RF model

交叉验证折数	RMSEC/(10 <sup>-2</sup> mg · g <sup>-1</sup> )	R <sub>c</sub> <sup>2</sup>
1	1.635	0.885
2	1.618	0.889
3	1.642	0.883
4	1.609	0.891
5	1.627	0.886
平均值	1.626	0.887
标准差	0.012	0.003

甲基(—CH<sub>3</sub>)和亚甲基(—CH<sub>2</sub>)基团。随着贮藏时间延长,脂肪氧化与蛋白质降解加剧,C—H基团含量发生显著变化,引起该波段吸光度明显波动,从而使其成为预测 TVB-N 质量分数的关键区域。

表 5 FOA-RF 模型重复性试验结果

Table 5 Repeated experiment results of the FOA-RF model

指标	RMSEP/(10 <sup>-2</sup> mg · g <sup>-1</sup> )	R <sub>p</sub> <sup>2</sup>
平均值	1.593	0.954
标准差	0.028	0.005
变异系数/%	1.76	0.52

(3) 1 430~1 550 nm 波段:此区域为 N—H 伸缩振动(来源于蛋白质氨基)与 O—H 弯曲振动(来源于水分)的叠加吸收区。在猪肉腐败过程中,蛋白质降解产生的 TVB-N(如氨和胺类)中含有大量 N—H 基团,同时水分含量因蒸发和微生物代谢作用发生变化,共同导致该波段的光谱响应与 TVB-N 质量分数高度相关。

(4) 1 680~1 750 nm 波段:该波段对应酰胺 I 带的 C=O 伸缩振动,是蛋白质二级结构(如 α-螺旋、β-折叠)的

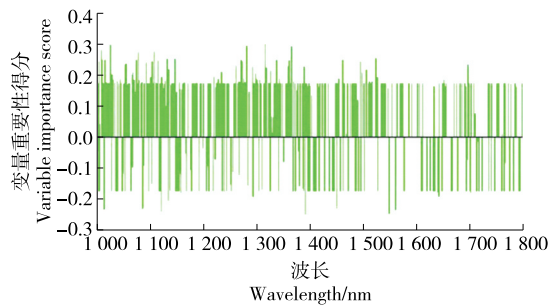


图3 FOA-RF模型特征重要性得分

Figure 3 Feature importance scores of the FOA-RF model

特征吸收区域。蛋白质降解会破坏其原有二级结构,引起C=O基团振动频率和强度改变,因而该波段可间接反映TVB-N质量分数的变化。

上述关键波段的识别不仅从理论上验证了近红外光谱技术用于猪肉TVB-N质量分数预测的可行性,也为后续模型优化(如特征选择与波段筛选)提供了明确方向,有助于简化模型结构、降低计算成本,并进一步提升检测效率。

### 3.4 不同贮藏条件下模型的适用性验证

为验证FOA-RF模型在不同贮藏条件下的适用性,将预测集按贮藏温度分为两组,分别计算模型在两组样品中的预测性能,结果见表6。

表6 FOA-RF模型在不同贮藏条件下的预测性能  
Table 6 Predictive performance of the FOA-RF model under different storage conditions

贮藏温度/℃	样本数	RMSEP/ ( $10^{-2} \text{ mg} \cdot \text{g}^{-1}$ )	$R_p$	$R_p^2$	RPDp
4	14	1.625	0.975	0.953	4.682
0	14	1.538	0.981	0.962	4.765
平均值		1.582	0.978	0.956	4.723

由表6可知,FOA-RF模型在4、0℃贮藏条件下的预测性能均优良,两组的预测性能指标与总预测集接近,差异较小(RMSEP差值 $<0.1 \text{ mg}/100 \text{ g}$ ,  $R_p^2$ 差值 $<0.01$ ),表明FOA-RF模型对贮藏温度变化的鲁棒性较强,可适用于不同冷链条件下的猪肉新鲜度检测,满足肉类工业中多样化的贮藏与运输场景需求。

## 4 结论

该研究提出了一种基于近红外光谱技术与果蝇优化算法改进的随机森林模型,实现了猪肉挥发性盐基氮质量分数的快速、无损和高精度检测。通过多元散射校正与一阶导数组合预处理有效提升了光谱数据质量,结合果蝇优化算法优化随机森林超参数,显著增强了模型预测

性能与泛化能力。研究表明,果蝇优化算法优化随机森林模型在预测集上的均方根误差低至 $1.582 \text{ mg}/100 \text{ g}$ ,决定系数高达0.956,残差预测偏差达4.723,优于传统偏最小二乘回归、未优化随机森林和随机森林以及网格搜索优化随机森林,且在4、0℃贮藏条件下均表现出良好稳定性与适用性。然而,该研究仍存在样本来源单一、贮藏条件有限、光谱范围较窄及模型机理解释不足等局限性。未来研究可围绕多样化样本与贮藏条件、多源信息融合、模型轻量化与嵌入式应用、光谱机理深入解析以及智能优化算法的进一步比较与集成等方面展开,以提升模型的普适性、精度和实用价值。

## 参考文献

- [1] 左晓佳,再努热·吐尔孙. 肉品新鲜度评价及保鲜技术研究进展[J]. 肉类研究, 2023, 37(12): 69-75.
- [2] ZUO X J, TUERXUN Z. Research progress on evaluation of meat freshness and preservation technologies[J]. Meat Research, 2023, 37(12): 69-75.
- [3] WILLIAMS P C, NORRIS K H. Near-infrared technology in the agricultural and food industries[M]. 2nd ed. Boca Raton: CRC Press, 2002.
- [4] 刘瑜明,王巧华,陈远哲,等. 猪肉理化指标的近红外光谱无损检测[J]. 光谱学与光谱分析, 2024, 44(5): 1346-1353.
- [5] LIU Y M, WANG Q H, CHEN Y Z, et al. Non-destructive detection of physicochemical indicators in pork using near-infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2024, 44(5): 1346-1353.
- [6] 周宇坤,陈孝敬,谢忠好,等. 一种近红外光谱数据预处理组合优化策略[J]. 光谱学与光谱分析, 2025, 45(1): 52-58.
- [7] ZHOU Y K, CHEN X J, XIE Z H, et al. A combined optimization strategy for preprocessing near-infrared spectral data[J]. Spectroscopy and Spectral Analysis, 2025, 45(1): 52-58.
- [8] GUO M Q, LIN H, WANG K Q, et al. Data fusion of near-infrared and Raman spectroscopy: an innovative tool for non-destructive prediction of the TVB-N content of salmon samples[J]. Food Research International, 2024, 189: 114-135.
- [9] LENG T, LI F, CHEN Y X, et al. Fast quantification of total volatile basic nitrogen (TVB-N) content in beef and pork by near-infrared spectroscopy: comparison of SVR and PLS model[J]. Meat Science, 2021, 180: 108-129.
- [10] 王冬,栾云霞,王欣然,等. 近红外光谱无损分析肉类品质的研究进展[J]. 肉类研究, 2024, 38(5): 61-70.
- [11] WANG D, LUAN Y X, WANG X R, et al. Research progress in non-destructive analysis of meat quality using near-infrared spectroscopy[J]. Meat Research, 2024, 38(5): 61-70.
- [12] 田文强,王巧华,徐步云,等. 基于近红外光谱的腌制期咸鸭蛋理化指标无损检测[J]. 食品科学, 2023, 44(2): 319-326.
- [13] TIAN W Q, WANG Q H, XU B Y, et al. Non-destructive

- detection of physicochemical indicators of salted duck eggs during salting using near-infrared spectroscopy[J]. Food Science, 2023, 44(2): 319-326.
- [9] 方瑶, 谢天铎, 郭渭, 等. 基于近红外光谱的金鲳鱼新鲜度快速检测技术[J]. 江苏农业学报, 2021, 37(1): 213-218.
- FANG Y, XIE T H, GUO W, et al. Rapid detection technique for freshness of golden pompano based on near-infrared spectroscopy[J]. Jiangsu Journal of Agricultural Sciences, 2021, 37(1): 213-218.
- [10] 张雷蕾, 李永玉, 彭彦昆, 等. 基于高光谱成像技术的猪肉新鲜度评价[J]. 农业工程学报, 2012, 28(7): 254-259.
- ZHANG L L, LI Y Y, PENG Y K, et al. Determination of pork freshness attributes by hyperspectral imaging technique[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(7): 254-259.
- [11] RONG L, WANG Y R. A fresh-cut papaya freshness prediction model based on partial least squares regression and support vector machine regression[J]. Heliyon, 2024, 10(9): 11-24.
- [12] 任智磊, 赵霄霄, 冯景, 等. 基于近红外光谱结合网格搜索—随机森林—自适应提升算法无损检测牛肉新鲜度[J]. 肉类研究, 2025, 39(11): 1-8.
- REN Z L, ZHAO X X, FENG J, et al. Non-destructive detection of beef freshness based on near-infrared spectroscopy combined with grid search-random forest-adaboost algorithm[J]. Meat Research, 2025, 39(11): 1-8.
- [13] CAI J, CHEN Q F, WAN X H, et al. Determination of total volatile basic nitrogen (TVB-N) content and Warner - Bratzler shear force (WBSF) in pork using Fourier transform near infrared (FT-NIR) spectroscopy[J]. Food Chemistry, 2011, 126(3): 1 354-1 360.
- [14] QU F F, REN D, HE Y, et al. Predicting pork freshness using multi-index statistical information fusion method based on near infrared spectroscopy[J]. Meat Science, 2018, 146: 59-67.
- [15] PROBST P, WRIGHT M N, BOULESTEIX A L. Hyperparameters and tuning strategies for random forest[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9(3): 1 301-1 318.
- [16] PAN W T. A new fruit fly optimization algorithm: taking the financial distress model as an example[J]. Knowledge-Based Systems, 2012, 26: 69-74.
- [17] LAN C F, SONG B W, ZHANG L, et al. State prediction of hydroturbine based on WOA-RF-adaboost[J]. Energy Reports, 2022, 8: 13 129-13 137.
- [18] 孙康慧, 肖安, 夏侯杰. 基于 LightGBM 机器学习算法的江西气温短期预报模型研究[J]. 高原气象, 2024, 43(6): 1 520-1 535.
- SUN K H, XIAO A, XIA H J. Study on short-term temperature forecast model in Jiangxi province based on LightGBM machine learning algorithm[J]. Plateau Meteorology, 2024, 43(6): 1 520-1 535.
- [19] 刘艳群, 肖付刚, 陈彩虹. 基于随机森林回归模型的小麦粉灰分含量快速测定[J]. 食品与机械, 2024, 40(9): 79-83.
- LIU Y Q, XIAO F G, CHEN C H. Study on rapid determination of ash content in wheat flour based on random forest regression model[J]. Food & Machinery, 2024, 40(9): 79-83.
- [20] 张清清. 基于 FOA-RF 模型下煤与瓦斯突出智能预警系统研究[J]. 陕西煤炭, 2024, 43(7): 152-155.
- ZHANG Q Q. Research on intelligent early warning system for coal and gas outburst based on foa-rf model[J]. Shaanxi Coal, 2024, 43(7): 152-155.
- [21] SHENG W, BAO Y. Fruit fly optimization algorithm based fractional order fuzzy-PID controller for electronic throttle [C]// Nonlinear Dynamics, 2013, 73(1/2): 611-619.
- [22] PUJAR P, LAVANYA M. Power factor improvement in IEEE-33 radial distribution systems using fruit fly algorithm compared with bat algorithm by optimizing the DG size[C]// 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS). Pudukkottai, India: [s.n.], 2024: 349-353.