

基于遗传算法和深度神经网络的近红外高光谱检测猪肉新鲜度

谢安国, 纪思媛, 李月玲, 王满生, 张 玉

Detection of Pork Freshness Using NIR Hyperspectral Imaging Based on Genetic Algorithm and Deep Neural Network

XIE Anguo, JI Siyuan, LI Yueling, WANG Mansheng, and ZHANG Yu

在线阅读 View online: <https://doi.org/10.13386/j.issn1002-0306.2023120096>

您可能感兴趣的其他文章

Articles you may be interested in

基于遗传算法-神经网络的豆渣酶解工艺优化及其动力学研究

Optimization of Enzymatic Hydrolysis of Soybean Dregs Based on Genetic Algorithm-Neural Network and Its Kinetics

食品工业科技. 2021, 42(16): 213-220 <https://doi.org/10.13386/j.issn1002-0306.2020120261>

基于人工神经网络耦合遗传算法(BP-GA)优化茶氨酸-葡萄糖美拉德反应的条件

Optimization of Theanine-Glucose Maillard Reaction Conditions Based on Artificial Neural Network Coupled Genetic Algorithm (BP-GA)

食品工业科技. 2023, 44(24): 183-192 <https://doi.org/10.13386/j.issn1002-0306.2023020165>

基于主成分分析和人工神经网络的近红外光谱大豆产地识别

Soybean Origin Identification Based by Near-Infrared Spectrum Based on Principal Component Analysis and Artificial Neural Network Model

食品工业科技. 2021, 42(9): 270-274 <https://doi.org/10.13386/j.issn1002-0306.2020060271>

基于人工神经网络耦联遗传算法(BP-GA)优化干酪乳杆菌LTL1361冻干保护剂配方

Optimization of Lyophilized Protective Agent Formulation of *Lactobacillus casei* LTL1361 Based on Artificial Neural Network Coupled Genetic Algorithm (BP-GA)

食品工业科技. 2022, 43(21): 175-184 <https://doi.org/10.13386/j.issn1002-0306.2022010143>

NIR高光谱成像技术联用SPA算法快速检测五花肉的过氧化值

NIR Hyperspectral Imaging Combined with SPA Algorithm for the Rapid Detection of Peroxidation Value of Pork Belly

食品工业科技. 2020, 41(8): 236-241 <https://doi.org/10.13386/j.issn1002-0306.2020.08.037>

基于近红外高光谱成像技术的花生冻伤检测方法研究

Peanut Frostbite Detection Method Based on Near Infrared Hyperspectral Imaging Technology

食品工业科技. 2024, 45(6): 226-233 <https://doi.org/10.13386/j.issn1002-0306.2023030252>



关注微信公众号, 获得更多资讯信息

谢安国, 纪思媛, 李月玲, 等. 基于遗传算法和深度神经网络的近红外高光谱检测猪肉新鲜度 [J]. 食品工业科技, 2024, 45(17): 345–351. doi: 10.13386/j.issn1002-0306.2023120096

XIE Anguo, JI Siyuan, LI Yueling, et al. Detection of Pork Freshness Using NIR Hyperspectral Imaging Based on Genetic Algorithm and Deep Neural Network[J]. Science and Technology of Food Industry, 2024, 45(17): 345–351. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2023120096

· 分析检测 ·

基于遗传算法和深度神经网络的近红外高光谱检测猪肉新鲜度

谢安国¹, 纪思媛¹, 李月玲¹, 王满生^{2,*}, 张 玉³

(1. 南阳理工学院张仲景国医国药学院, 河南南阳 473004;

2. 中国农业科学院麻类研究所, 湖南长沙 410205;

3. 南阳理工学院智能制造学院, 河南南阳 473004)

摘要: 为系统评估基于深度学习的智能辅助高光谱成像系统在猪肉新鲜度指标检测中的效果, 采集了猪肉在 4 °C 冷藏 12 d 中挥发性盐基氮 (volatile basic nitrogen, TVB-N)、菌落总数 (total viable count, TVC) 以及 900~2500 nm 近红外光谱数据。基于 Python 的 TensorFlow 和 Keras 平台, 对高光谱数据进行处理, 建立了深度神经网络的定量检测模型。并利用遗传算法 (GA) 选择与猪肉新鲜度相关的特征光谱波段。结果表明, 遗传算法选取波段对光谱模型的性能有明显提升。当光谱波段数达到 35 和 50 时, GA+ANN 模型预测精度高于全波段的线性回归模型。TVC 为预测指标的预测性能优于 TVB-N, TVC 测试集最佳 R_p^2 为 0.877, RMSEP 为 0.575; 预测 TVB-N 的最佳 R_p^2 为 0.826, RMSEP 为 1.01。此外, 通过研究还发现, 遗传算法优选的近红外光谱波段与肉品的 O-H, N-H, C=O 等分子振动吸收带具有较高的吻合度。本研究为处理近红外和高光谱数据提供了新的方法, 也为猪肉及其他肉品新鲜度快速无损检测提供了技术参考。

关键词: 猪肉品质, 新鲜度, 高光谱成像 (HSI), 近红外光谱 (NIR), TensorFlow, 遗传算法, 神经网络

中图分类号: TS251.7

文献标识码: A

文章编号: 1002-0306(2024)17-0345-07

DOI: 10.13386/j.issn1002-0306.2023120096

本文网刊:



Detection of Pork Freshness Using NIR Hyperspectral Imaging Based on Genetic Algorithm and Deep Neural Network

XIE Anguo¹, JI Siyuan¹, LI Yueling¹, WANG Mansheng^{2,*}, ZHANG Yu³

(1. Zhang Zhongjing School of Chinese Medicine, Nanyang Institute of Technology, Nanyang 473004, China;

2. Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205, China;

3. School of Intelligent Manufacturing, Nanyang Institute of Technology, Nanyang 473004, China)

Abstract: To evaluate the effectiveness of a deep learning which is based intelligent assisted hyperspectral imaging system on the detection of pork freshness indicators, volatile basic nitrogen (TVB-N), total viable count (TVC), and 900~2500 nm near-infrared spectral data were collected from pork which were refrigerated at 4 °C for 12 days. Based on Python's TensorFlow and Keras platform, hyperspectral data was processed and a quantitative detection model of deep neural network was also established. And the characteristic spectral bands related to pork freshness were selected by genetic algorithm (GA). The results showed that the performance of the spectral model could be improved significantly by selecting the band of genetic algorithm. When the number of spectral bands reached 35 and 50, the prediction accuracy of GA+ANN model was higher than that of full-band linear regression model. The predictive performance of TVC was better than that of TVB-N, and the best R_p^2 and RMSEP of TVC were 0.877 and 0.575, respectively. The best R_p^2 and RMSEP for TVB-N

收稿日期: 2023-12-11

基金项目: 河南省自然科学基金 (202300410131); 中国农业科学院科技创新工程 (CAAS-ASTIP-IBFC); 南阳理工学院科研启动经费 (510171)。

作者简介: 谢安国 (1985-), 男, 博士, 讲师, 研究方向: 食品检测, E-mail: anguo@nyist.edu.cn。

* 通信作者: 王满生 (1987-), 男, 博士, 副研究员, 研究方向: 食品安全与保鲜加工, E-mail: wms198704@163.com。

were 0.826 and 1.01, respectively. In addition, it was also found that the NIR band selected by genetic algorithm had a high coincidence with the molecular vibration absorption bands of meat, such as O-H, N-H, C=O and so on. This study provides a new method which can be used for processing the near-infrared and hyperspectral data, and also provides a technical reference for rapid nondestructive testing of pork and other meat freshness.

Key words: pork quality; freshness; hyperspectral image (HSI); near infrared (NIR); TensorFlow; genetic algorithm; neural network

猪肉几乎包含了人体所需的碳水化合物、蛋白质、脂肪、各种无机盐和维生素等重要营养素,是人们获取营养的重要来源^[1],也是我国日常食用量最大的肉品。随着人们食品安全意识的不断增强,肉类食品品质问题受到广泛关注。猪肉新鲜度是衡量猪肉品质优劣的重要指标^[2],传统猪肉新鲜度检测方法有人工感官评定、化学方法测量和微生物检验等^[3],而传统检测方法操作繁琐、耗时长、成本高、难以满足大批量样本的快速检测^[4]。目前,得益于机器视觉^[5]、高光谱技术^[6-8]、近红外光谱技术^[9]、电子鼻^[10]及多传感器信息融合^[11]等技术的运用,猪肉新鲜度检测方法也有了突破性的进展。

食品无损检测技术中必不可少的环节是信号预处理和训练模型,这些复杂的数学计算需要在专业的软件中进行。目前,我国乃至全球高光谱领域的数据处理主要依靠 MATLAB 软件。MATLAB 又名矩阵实验室,包含数十个功能强大的工具箱,是科研人员从事数学计算、建模仿真、算法开发等研究的重要工具。另一种编程工具 Python 是一种开源免费、跨平台的计算机程序设计语言,语言简洁、易读和可扩展^[12]。尤其自 2016 年 DeepMind 的 AlphaGo 战胜了人类围棋世界冠军以来,深度学习为基础的人工智能(AI)在科技领域掀起了一股热潮。从自动驾驶到医疗诊断,从人脸识别到生成自然语言的 ChatGPT,深度学习技术在各个领域的突破性应用一直是全球民众的热门话题。Python 平台包含了用于深度网络训练的专业应用程序编程接口(API)如 TensorFlow、Keras 和 PyTorch 等,开发者可以调用这些接口快速搭建所需的深度学习模型,省去了编写大量基础性代码的工作。因此 Python 在人工智能领域受到越来越多开发者的青睐。此外,还有众多第三方开源的科学计算库都提供了 Python 接口,供用户调用和开发扩展。其中遗传算法工具箱 Geatpy 由华南理工大学、华南农业大学等高校学生联合团队开发,是一个高性能遗传算法库及开放式进化算法框架^[13]。Geatpy 是一个功能强大的进化算法工具箱,提供了耦合度很低的进化算法框架,没有过于抽象的复杂封装,提供多种格式的编码方式以及丰富的选择、交叉和变异等算子。

综上,本文基于 Python 的 TensorFlow 和 Keras 程序构建深度神经网络,处理高光谱数据用于预测猪肉新鲜度变化(以 TVB-N 和 TVC 作为评价指标)。并进一步利用遗传算法筛选高光谱特征波段,以此实

现快速无损检测猪肉新鲜度。为人工智能算法在食品品质的无损检测领域的应用,提供一定的技术参考。

1 材料与方法

1.1 材料与仪器

实验所用猪肉为 15 头猪胴体的完整背最长肌冷鲜肉,正大集团(洛阳)提供。将肉品分割成 5 cm×3 cm×2.5 cm 的块状样品,放置在 4℃ 的冷库内贮藏待用。

高光谱成像系统由镜头、光谱仪、相机、光源和移动平台等部件组成,整体设备由中国台湾五环光学有限公司组装;其中核心部件 N17E 光谱仪 芬兰 ImSpector 公司,波长 900~2500 nm 内含有 256 个波段;XEVA 992 CCD 相机 比利时 Xenics 公司,像素分辨率 320×300;LRH-250A 生化培养箱 广东省医疗器械厂;MLS-3750 灭菌锅、MIR-554 低温恒温培养箱 日本 SANYO 公司。

1.2 实验方法

1.2.1 猪肉贮藏条件及样品处理 贮藏在 4℃ 冷库中的肉块,从试验的第 0 d 至第 12 d 每天随机抽取 7 块检测,共取 91 块。利用高光谱成像系统获得光谱数据,每块肉正反面共扫描 2 次,总得到 182 个光谱曲线。然后将肉块用剪刀剪碎后拍打成肉泥,立即称取肉泥 15 g 检测菌落总数(TVC),另外准确称取肉泥 10 g 检测挥发性盐基氮(TVB-N)值。每个样品重复 3 次。

1.2.2 冷鲜猪肉高光谱的扫描 高光谱成像系统采用“推扫式”获得光谱反射图像,采用近红外摄像头。高光谱数据经光谱矫正后,选取并提取目标区域的平均光谱。

1.2.3 挥发性盐基氮和菌落总数测定分析 挥发性盐基氮的测定:每份样品中精确称取扫描后的肉泥 10 g,参考 GB 5009.228-2016《食品中挥发性盐基氮的测定》进行 TVB-N 的测定;菌落总数的测定:扫描后的肉块经拍打成肉泥后从中准确称取 15 g,参考 GB 4789.2-2016《食品微生物学检验 菌落总数测定》进行 TVC 的测定。

1.3 Python 程序设计

1.3.1 Python 平台启动与配置 本实验中硬件条件为 Intel(R) Core i5-12490F CPU, 16 G 内存, NVIDIA GeForce RTX 4070 显卡,软件环境为 Windows10 (64 位),编程语言 Python 3.11,开发平台 Spyder。

另外,代码编写所用的工具包为 TensorFlow、Keras、geatpy、math、random、xlrd、matplotlib、numpy、sklearn。

1.3.2 机器学习的设计流程 图 1 为总体算法流程图,包含冷鲜猪肉理化指标的检测、高光谱信息的采集、数据的预处理、遗传算法筛选特征波段、神经网络的设计及模型评价等内容。

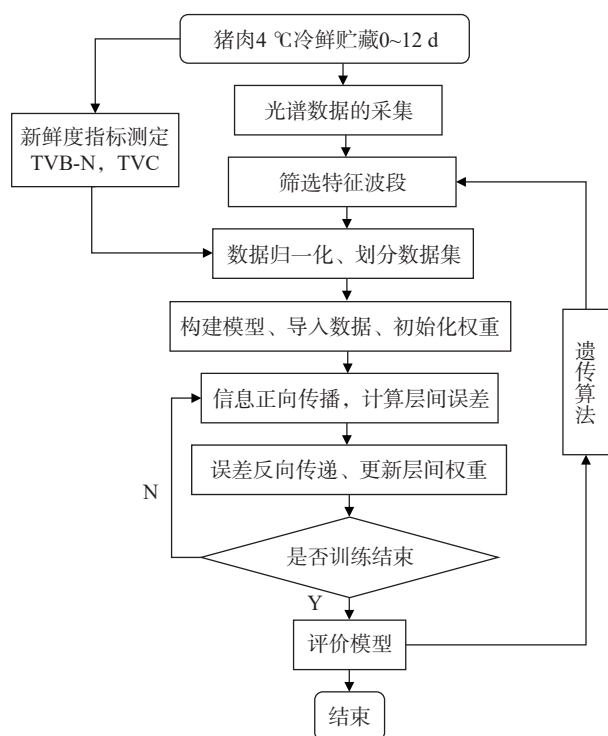


图 1 总体算法的设计图

Fig.1 Design diagram of overall algorithm

1.3.3 神经网络回归模型的建立 神经网络的输入输出数据须归一化处理,将网络的输入、输出数据转换成 [0,1] 区间内。神经网络由 3 个层次组成,即输入层、输出层和隐含层,其中隐含层可为单层或多层结构^[9]。本实验隐含层设置为 4 层,输入层为 50 时,隐含层神经元个数分别为 50, 50, 25, 10。每层的神经元接受输入信息后,通过权重 w 连接互相传递信息,本文中传递函数选择 relu 函数。训练算法选择梯度下降算法;设置优化器为 adam ,损失函数为 $\text{mean_absolute_error}$ 。终止条件为训练达到所设置的最大训练次数或训练时间,或者网络误差下降到期望误差时,网络均会停止学习。

1.3.4 遗传算法筛选波段 本实验中遗传算法筛选特征波段是基于 Python 平台 geatpy 提供的开放的面向对象进化算法框架来实现的,此框架包含:问题类、进化算法模板类、种群类三部分,用框架的方法可以极大地提高编程效率^[14]。本实验 geatpy 遗传算法的目标函数设定为求 R^2 最大值;本实验需要对比筛选 5~50 个特征波段的建模效果,因此依次运行了初始种群个数为 5、10、25、35 和 50 个时的神经网络光谱模型,并返回目标函数值 R^2 。代码中设置最

大遗传代数 1000,设置差分进化的变异缩放因子为 0.1,设置交叉概率为 0.5。

1.4 模型的主要评价指标

本实验中共采集了 182 个高光谱图像,通过随机分配 120 个数据用于校准集(又称为训练集),30 个用于验证集,剩余 32 块归属到测试集(又称为预测集)。评估模型性能的统计标准回归模型相关系数(R^2)和均方误差(RMSE),一个较好的光谱模型应具有较高的 R^2 以及较低的 RMSE。校准集和预测集的相关系数分别记为 R_c^2 和 R_p^2 ;校准集和预测集的均方误差记为 RMSEC 和 RMSEP。

神经网络建立和遗传算法筛选特征波段两部分属于模型训练阶段。本文分别用隔 5 取 1 获得 50 个波段、随机抽取 50 个波段代入人工神经网络(ANN)模型,对比遗传算法筛选波段的效果。并用 256 个全波段光谱构建多元线性回归(multiple linear regression, MLR)光谱模型进行了对照。神经网络、遗传算法和多元线性回归等算法编程,以及数据绘图均在 Python 3 平台完成。

2 结果与分析

2.1 冷鲜猪肉在冷藏期间理化指标及高光谱的变化

挥发性盐基氮(TVB-N)是猪肉蛋白质分解时产生的碱性含氮物质,其含量会随着猪肉腐败程度的加剧而增加,可用于反映猪肉的新鲜度变化,也是评价猪肉新鲜度的有力指标^[15]。菌落总数(TVC)可用来判定食品被细菌污染的程度及卫生质量^[16]。因此,TVB-N 和 TVC 都是我国食品法规衡量肉品鲜度的重要指标,通常 $\text{TVB-N} > 25 \text{ mg/100 g}$ 或者 $\text{TVC} > 7 \log \text{ CFU/g}$ 会被看作不新鲜的肉。冷鲜猪肉在冷藏期间 TVB-N 值和 TVC 变化如图 2 所示,随着冷藏时间的延长,TVB-N 值和 TVC 总体上呈增加趋势。TVC 值在 1 d 后略有下降,这主要是因为部分常温菌在低温条件下生长受到抑制甚至减少导致的;此后逐步上涨,这是因为嗜冷菌开始慢慢增长引起。TVB-N 值在第 4、5 d 大幅上涨,Luo 等^[17]研究表明肉在氧化过程中,蛋白质降解导致大量酶的释放,这反过来又加速了蛋白质的降解、氨基酸的分解和代谢产物的释放。

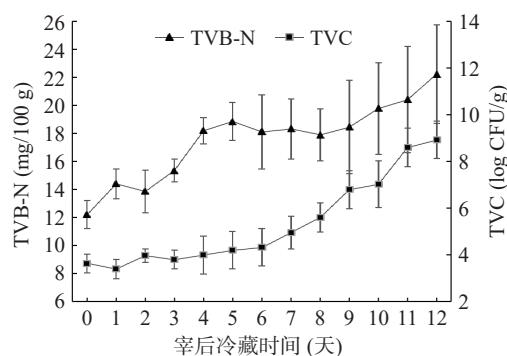


图 2 不同冷藏时间猪肉 TVB-N 和 TVC 值

Fig.2 TVB-N and TVC values of pork at different refrigeration times

图 3 展示了不同冷藏时间猪肉样品的反射率光谱曲线, 在 970、1200 和 1450 nm 附近呈现出明显的吸收特征。970 nm 和 1450 nm 两个谱带通常被认为是 OH 伸缩振动的一级倍频和二级倍频吸收^[18]; 中心在 1200 nm 附近有一个宽的比较弱的组合频吸收, 该宽峰被归属为两个水分子键合到 OH⁻上的吸收^[19]。并且脂质中 C-H 的伸缩振动会影响 920 和 1210 nm 附近峰, 研究^[20] 已知不同类型的脂质, 如胆固醇、胆固醇酯、磷脂和甘油三酯, 在近红外区域具有脂质独特的又彼此相似的吸收光谱。此外, 不同冷藏时间的近红外波段的高光谱曲线形状基本一致, 但是强度不一。随着冷藏时间的延长, 肉品反射值总体呈上升趋势, 即吸收度降低, 这是因为猪肉表面因水分损失而变干^[21]; 肉块因氧化而变色, 且腐败也会使得肉块失去原有的光泽。

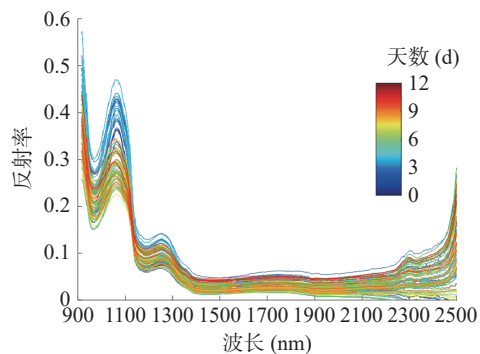


图 3 不同新鲜度的猪肉近红外波段光谱反射图
Fig.3 Near infrared spectral reflectance of pork with different freshness

2.2 神经网络模型效果

2.2.1 超参数的选择对模型结果影响 学习速率和动量因子是神经网络模型中的一个重要的超参数^[22], 如何调整超参数是训练出好模型的关键要素之一^[23]。以单个指标菌落总数(TVC)的训练集数据为对象, 研究光谱建模过程中各参数对模型结果的影响。神经网络输入层神经元个数为 50(以隔 5 取 1 的方式选择光谱), 固定训练次数设为 200, 学习速率设为 0.01。如表 1 所示, 仅变动动量因子时, 当动量因子由 0.01 变为 0.001 时, R_c^2 和 RMSEC 变化明显; 动量因子改变为 0.0001 或更小值时, R_c^2 和 RMSEC 变化不明显。因此继续缩小动量因子没有意义, 动量因子取为 0.0001 模型的性能比较好。

表 1 设置不同动量因子对建模结果影响
Table 1 Influence of setting different momentum factors on modeling results

| 学习速率 | 动量因子 | R_c^2 | RMSEC |
|------|--------|---------|-------|
| 0.01 | 0.01 | 0.664 | 0.976 |
| 0.01 | 0.001 | 0.748 | 0.844 |
| 0.01 | 0.0001 | 0.749 | 0.843 |

学习速率对模型效果的影响也尤为显著。学习速率太大容易出现超调现象^[24], 损失函数剧烈振荡,

没有减小的趋势。学习速率越小, 无法快速地找到好的损失梯度下降方向, 且下降的速度慢, 收敛长^[25]。在本实验中, 动量因子确定为 0.0001 后, 改变学习速率, 由表 2 可知, 改变学习速率, 模型的性能指标变化幅度较小, 当学习速率设置为 0.001 时, R_c^2 最大, RMSEC 最小, 模型的性能较好。

表 2 设置不同学习速率对建模结果影响
Table 2 Influence of setting different learning rates on modeling results

| 学习速率 | 动量因子 | R_c^2 | RMSEC |
|--------|--------|---------|-------|
| 0.01 | 0.0001 | 0.749 | 0.843 |
| 0.001 | 0.0001 | 0.771 | 0.804 |
| 0.0001 | 0.0001 | 0.741 | 0.856 |

2.2.2 训练次数对模型结果的影响 确定了模型的学习速率为 0.001, 动量因子为 0.0001, 改变训练次数, 由表 3 可知, R_c^2 随着训练次数的增加而增大, RMSEC 随着训练次数的增加而减小, 模型的性能渐趋良好; 但随着训练次数的增加, 运行时间逐渐增加, R^2 增加的幅度越来越小。

表 3 神经网络训练次数对建模结果影响
Table 3 Influence of neural network learning times on modeling results

| 训练次数 | R_c^2 | RMSEC | 运行时间(s) |
|------|---------|-------|---------|
| 200 | 0.771 | 0.804 | 27.135 |
| 500 | 0.814 | 0.725 | 42.588 |
| 1000 | 0.839 | 0.676 | 80.076 |
| 3000 | 0.851 | 0.651 | 138.104 |
| 5000 | 0.859 | 0.645 | 186.151 |

当训练次数为 5000 时, 其误差已基本收敛, 若再增加训练次数, 不仅对结果影响不大, 而且增加模型运行时间。故综合考虑模型的性能指标及网络运行时间, 将神经网络的训练次数确定为 5000。

2.3 遗传进化筛选波段的效果

用遗传算法选取 5~50 个不同特征波段以得到适应度函数值, 当选取 5 个特征波段时, 运算结果较差且不稳定, 不足以支撑后续的建模运算。当特征波段为 10 个、25 个、35 个和 50 个时运算结果如图 4 所示。选取 10 个特征波段时, 其种群个体平均目标函数值(图中简写为个体平均), 即回归指标 R_c^2 所收敛的值只有 0.53。当增加选取的特征波段个数至 35 时, 种群个体平均目标函数值达到 0.85, 其中一部分最优个体组合的目标函数值(图中简写为最优个体)达到 0.95。特征波段越多, 模型训练效果越好, 这是光谱信息逐渐丰富的缘故。当遗传算法选取 50 个特征波段, 进化代数 1000 时, 种群个体平均目标值收敛至 0.94 附近, 此时遗传算法已寻得相当合理的个体组合, 已无需筛选更多的光谱波段。

2.4 不同波段筛选方法和建模方法预测效果对比

当神经网络模型结构及特征波段选定后, 利用

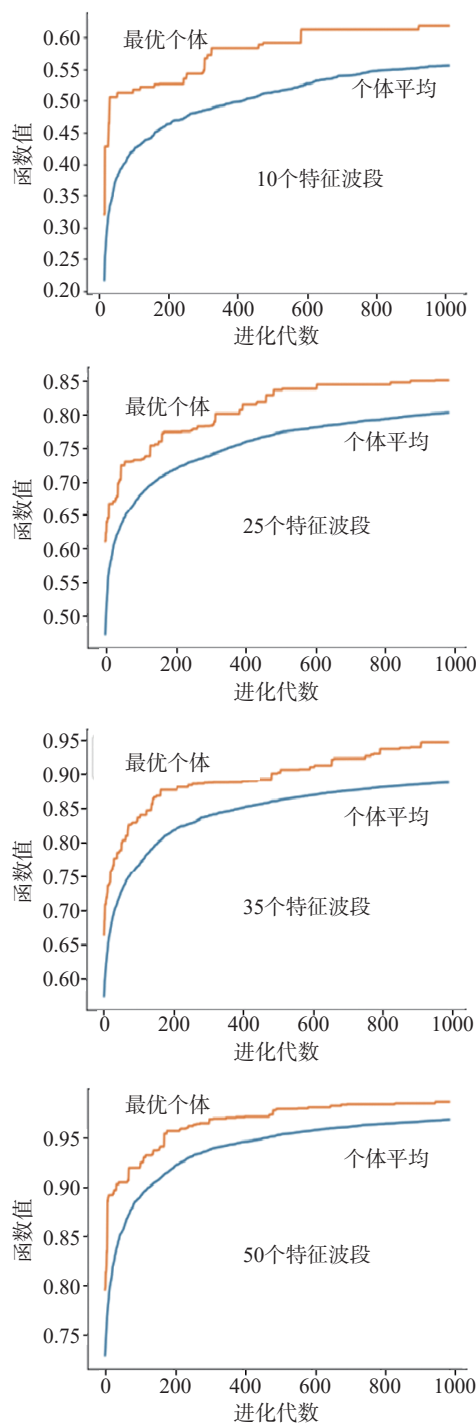


图 4 遗传算法筛选 10~50 个波段适应度函数图

Fig.4 Genetic algorithm screened 10~50 bands fitness function graphs

训练集中 120 个数据和验证集中 30 个数据,训练深度神经网络模型。确定了模型参数后,用 32 个预测集数据作为神经网络的输入,预测新鲜度指标 TVC 和 TVB-N 值。为对比不同的方法筛选波段对建模结果的影响,采用的方法有:从 256 个光谱中均匀地抽取 50 个波段(隔 5 取 1)、随机函数抽取 50 个波段、遗传算法筛选 25、35 和 50 个波段,这几种方法配合 ANN 建模。同时用 256 个全波段光谱构建多元线性回归模型作为预测效果的对照。不同建模结果如表 4 所示。

表 4 不同波段筛选方法和建模方法对新鲜度检测的效果对比

Table 4 Comparison of freshness detection effects of different band optimization methods and modeling methods

| 指标 | 光谱波段数 | 筛选方法 | 建模方法 | 校准集 | | 预测集 | |
|-------|-------|--------|------|---------|-------|---------|-------|
| | | | | R_c^2 | RMSEC | R_p^2 | RMSEP |
| TVC | 50 | 隔5取1 | ANN | 0.859 | 0.645 | 0.792 | 0.747 |
| | 50 | 随机 | ANN | 0.769 | 0.788 | 0.671 | 0.995 |
| | 256 | 全波段 | 线性回归 | 0.914 | 0.482 | 0.839 | 0.658 |
| | 25 | Geatpy | ANN | 0.835 | 0.665 | 0.763 | 0.798 |
| | 35 | Geatpy | ANN | 0.924 | 0.451 | 0.877 | 0.575 |
| | 50 | Geatpy | ANN | 0.931 | 0.432 | 0.874 | 0.582 |
| TVB-N | 50 | 隔5取1 | ANN | 0.787 | 1.12 | 0.694 | 1.344 |
| | 50 | 随机 | ANN | 0.742 | 1.234 | 0.686 | 1.361 |
| | 256 | 全波段 | 线性回归 | 0.865 | 0.892 | 0.789 | 1.114 |
| | 25 | Geatpy | ANN | 0.816 | 1.042 | 0.686 | 1.361 |
| | 35 | Geatpy | ANN | 0.888 | 0.814 | 0.804 | 1.074 |
| | 50 | Geatpy | ANN | 0.894 | 0.791 | 0.826 | 1.01 |

结果表明,随机函数抽取的 50 个波段训练的神经网络模型效果最差, R_p^2 值最低为 0.671。通常情况下,在一定的范围内,神经网络输入层神经元个数越多,预测结果越精确,模型的性能越多。采用遗传算法筛选波段数从 25 上升到 35 及 50 时, R_c^2 值不断上升,RMSEC 进一步减小,模型性能变优。用遗传算法筛选的 35 和 50 波段建模,预测效果都优于均匀抽取或随机抽取的波段组合,甚至超过全波段多元线性回归模型,说明用遗传算法筛选波段可显著提高模型的精度及性能。

对比 TVB-N 和 TVC 不同指标的检测效果,结果表明以 TVC 作为模型预测指标比以 TVB-N 为指标的模型性能更好、预测结果更精确。其原因是在猪肉冷藏的后期 TVB-N 测量值的方差较大,数据规律性没有菌落总数的高。预测 TVB-N 的最佳算法组合是遗传算法筛选 50 波段结合 ANN,最大 R_p^2 值为 0.826, RMSEP 为 1.01。表 4 中特例情况是, TVC 指标模型 50 波段的 R_c^2 值高于 35 波段,但在测试集中 35 波段的 R_p^2 值最高为 0.877。50 波段的 R_p^2 值稍低的原因可能是 50 波段的神经网络有一定程度的过度拟合,35 波段的光谱建模已经可以取得很好的结果。

2.5 遗传算法优选光谱波段的分布特点

遗传算法优选 35 个特征波段在 900~2500 nm 范围内并不是均匀分布的,这些特征波段在图 5 用竖线表示。特征波段在 940~1200、1450~1600 和 1900~2100 nm 三个区间内密集分布,而水中 O-H 伸缩振动产生的倍频与合频吸收带(980、1200、1450 和 1930 nm 附近)恰好好处在这三个区间。此外,与肉品蛋白质结构相关的 N-H 对称振动与反对称振动(1470、1480 和 1500 nm), N-H 伸缩振动(2060 nm), 酰胺基团振动(2040、2080 nm)以及 C=O 伸缩振动 2 倍频(2180 nm)等吸收峰^[26] 也恰好在特征波段附近。遗传算法作为一种基于数学算法的优化方法,其

选定的波段与化学分析领域中测定的分子振动吸收峰具有较高的吻合度。这一有趣的现象背后所蕴含的机理,还有待于今后进一步深入探索。

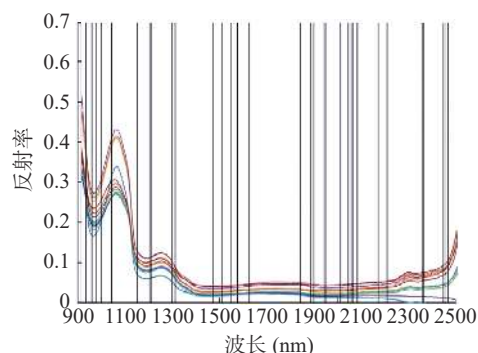


图5 遗传算法优选 35 个特征波段分布图

Fig.5 Distribution of 35 characteristic bands selected by genetic algorithm

在过去的研究中,多种优化算法如连续投影算法(SPA)^[27]、逐步回归算法和偏最小二乘法回归的 β 系数法等^[28]被用于筛选波段并建立光谱模型。结果证实了这些算法能够减少光谱数据冗余,提高了模型的稳健性和计算速率^[29],但一直未检测筛选波段与官能团吸收带的吻合度^[30]。本实验遗传算法筛选的波段反映出一定的可解释性,遗传算法在光谱学上挖掘分子结构信息方面或许具有独特潜力和优势。

3 结论

基于 Python 平台处理高光谱数据并建立深度神经网络,通过遗传算法优化了光谱波段筛选,分别对猪肉新鲜度进行定量预测。此外遗传算法优选的近红外光谱波段与肉品中半结合水的 O-H 键、蛋白质中的 N-H、C=O 和酰胺分子基团吸收带有较高的吻合度。本研究展现了 Python 平台处理 NIR 高光谱数据的可靠性及灵活性,利用深度学习算法可以训练出很好的食品检测模型。Python 作为一种开源的免费编程软件应用广泛,每年都有大量新算法在该平台涌现。可预见在本领域今后会有越来越多的科研工作者尝试 Python 人工智能算法进行光谱分析。

© The Author(s) 2024. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

参考文献

[1] WÓJCIAK K M, HALAGARDA M, ROHN S, et al. Selected nutrients determining the quality of different cuts of organic and conventional pork[J]. *European Food Research and Technology*, 2021, 247: 1389–1400.

[2] KUCHAR C T, NGADI M O. Rapid assessment of pork freshness using miniaturized NIR spectroscopy[J]. *Journal of Food Measurement and Characterization*, 2020, 14: 1105–1115.

[3] SOHAIB M, ANJUM F M, ARSHAD M S, et al. Postharvest intervention technologies for safety enhancement of meat and meat based products; a critical review[J]. *Journal of Food Science and*

Technology, 2016, 53: 19–30.

[4] XIE A G, SUN J, WANG T, et al. Visualized detection of quality change of cooked beef with condiments by hyperspectral imaging technique[J]. *Food Science and Biotechnology*, 2022, 31(10): 1257–1266.

[5] ZAREEF M, CHEN Q, HASSAN M M, et al. An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis[J]. *Food Engineering Reviews*, 2020, 12: 173–190.

[6] XIE A G, SUN D W, ZHU Z, et al. Nondestructive measurements of freezing parameters of frozen porcine meat by NIR hyperspectral imaging[J]. *Food and Bioprocess Technology*, 2016, 9: 1444–1454.

[7] 王彩霞, 王松磊, 贺晓光, 等. 高光谱图谱融合检测羊肉中饱和脂肪酸含量[J]. *光谱学与光谱分析*, 2020, 409(2): 595–601.

[WANG C X, WANG S L, HE X G, et al. Detection of saturated fatty acid content in mutton by using the fusion of hyperspectral spectrum and image information[J]. *Spectroscopy and Spectral Analysis*, 2020, 409(2): 595–601.]

[8] ZHANG J, LIU G, LI Y, et al. Rapid identification of lamb freshness grades using visible and near-infrared spectroscopy (Vis-NIR)[J]. *Journal of Food Composition and Analysis*, 2022, 111: 104590.

[9] ERNA K H, ROVINA K, MANTIHAN S. Current detection techniques for monitoring the freshness of meat-based products: A review[J]. *Journal of Packaging Technology and Research*, 2021, 5(3): 127–141.

[10] GÓRSKA-HORCZYCHAK E, HORCZYCHAK M, GUZEK D, et al. Chromatographic fingerprints supported by artificial neural network for differentiation of fresh and frozen pork[J]. *Food Control*, 2017, 73: 237–244.

[11] HASSOUN A, AÏT-KADDOUR A, SAHAR A, et al. Monitoring thermal treatments applied to meat using traditional methods and spectroscopic techniques: A review of advances over the last decade[J]. *Food and Bioprocess Technology*, 2021, 14: 195–208.

[12] NGUYEN G, DLUGOLINSKY S, BOBÁK M, et al. Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey[J]. *Artificial Intelligence Review*, 2019, 52: 77–124.

[13] Geatpy—The genetic and evolutionary algorithm toolbox for Python with high performance[EB/OL]. <http://geatpy.com/index.php/home/>.

[14] MAHMUD M, KAISER M S, MCGINNITY T M, et al. Deep learning in mining biological data[J]. *Cognitive Computation*, 2021, 13: 1–33.

[15] DONG K, GUAN Y, WANG Q, et al. Non-destructive prediction of yak meat freshness indicator by hyperspectral techniques in the oxidation process[J]. *Food Chemistry*, 2023, 17: 100541.

[16] CASABURI A, PIOMBINO P, NYCHAS G J, et al. Bacterial populations and the volatile associated to meat spoilage[J]. *Food Microbiology*, 2015, 45: 83–102.

[17] LUO X, DONG K, LIU L, et al. Proteins associated with quality deterioration of prepared chicken breast based on differential proteomics during refrigerated storage[J]. *Journal of the Science of Food and Agriculture*, 2021, 101: 3489–3499.

[18] BÜNING-PFAUE H. Analysis of water in food by near infrared spectroscopy[J]. *Food Chemistry*, 2003, 82(1): 107–115.

[19] HEIMAN A, LICHT S. Fundamental baseline variations in aqueous near-infrared analysis[J]. *Analytica Chimica Acta*, 1999,

394(2): 135–147.

[20] WANG P, WANG P, WANG H W, et al. Mapping lipid and collagen by multispectral photoacoustic imaging of chemical bond vibration[J]. *Journal of Biomedical Optics*, 2012, 17(9): 96010–96011.

[21] 谢安国, 王满生, 石晓微, 等. 牛肉加热过程中低场核磁共振信号与品质特征的动态分析[J]. *食品与机械*, 2020, 36(7): 23–27, 71. [XIE A G, WANG M S, SHI X W, et al. Dynamic analysis of LF-NMR relaxation signals and quality characteristics during heating beef[J]. *Food & Machinery*, 2020, 36(7): 23–27, 71.]

[22] 胡黄水, 赵思远, 刘清雪, 等. 基于动量因子优化学习率的 BP 神经网络 PID 参数整定算法[J]. *吉林大学学报(理学版)*, 2020, 58(6): 1415–1420. [HU H S, ZHAO S Y, LIU Q X, et al. BP neural network PID parameter tuning algorithm based on momentum factor optimized learning rate[J]. *Journal of Jilin University(Science Edition)*, 2020, 58(6): 1415–1420.]

[23] SUN H, SHEN L, ZHONG Q, et al. AdaSAM: Boosting sharpness-aware minimization with adaptive learning rate and momentum for training deep neural networks[J]. *Neural Networks*, 2024, 169: 506–519.

[24] CHANDRA B, SHARMA R K. Deep learning with adaptive learning rate using laplacian score[J]. *Expert Systems with Applications*, 2016, 63: 1–7.

[25] ZHENG S, GUO W, LI C, et al. Application of machine learning and deep learning methods for hydrated electron rate constant prediction[J]. *Environmental Research*, 2023, 231: 115996.

[26] WORKMAN J, WEYER L. Practical guide to interpretive near-infrared spectroscopy[M]. CRC Press, 2007, 47: 4620–4629.

[27] 何鸿举, 王洋洋, 王魏, 等. NIR 高光谱成像技术联用 SPA 算法快速检测五花肉的过氧化值[J]. *食品工业科技*, 2020, 41(8): 236–241. [HE H J, WANG Y Y, WANG W, et al. NIR hyperspectral imaging combined with SPA algorithm for the rapid detection of peroxidation value of pork belly[J]. *Science and Technology of Food Industry*, 2020, 41(8): 236–241.]

[28] XU J L, ESQUERRE C, SUN D W. Methods for performing dimensionality reduction in hyperspectral image classification[J]. *Journal of Near Infrared Spectroscopy*, 2018, 26(1): 61–75.

[29] 李艳坤, 董汝南, 张进, 等. 光谱数据解析中的变量筛选方法[J]. *光谱学与光谱分析*, 2021, 41(11): 3331–3338. [LI Y K, DONG R N, ZHANG J, et al. Variable selection methods in spectral data analysis[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3331.]

[30] CHENG W W, SUN D W, PU H B, et al. Characterization of myofibrils cold structural deformation degrees of frozen pork using hyperspectral imaging coupled with spectral angle mapping algorithm[J]. *Food Chemistry*, 2018, 239: 1001–1008.