

DOI: 10.13652/j.issn.1003-5788.2021.10.021

# 基于模糊鉴别主成分分析的生菜贮藏时间鉴别

## Discrimination of lettuce storage time based on fuzzy discriminant principal component analysis

侯晓蕾<sup>1</sup> 武小红<sup>2</sup> 武斌<sup>3</sup> 沈嘉棋<sup>1</sup> 汪鑫<sup>1</sup>HOU Xiao-lei<sup>1</sup> WU Xiao-hong<sup>2</sup> WU Bin<sup>3</sup> SHEN Jia-qi<sup>1</sup> WANG Xin<sup>1</sup>

(1. 江苏大学卓越学院, 江苏 镇江 212013; 2. 江苏大学电气信息工程学院,

江苏 镇江 212013; 3. 滁州职业技术学院信息工程学院, 安徽 滁州 239000)

(1. *Institute of Talented Engineering Students, Jiangsu University, Zhenjiang, Jiangsu 212013, China;**2. School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China;**3. School of Information Engineering, Chuzhou Polytechnic, Chuzhou, Anhui 239000, China)*

**摘要:**目的:设计一种较为准确的生菜贮藏时间鉴别模型。方法:以新鲜生菜为试验对象,分别采用主成分分析、鉴别主成分分析、模糊鉴别主成分分析3种特征提取方法对预处理后的生菜近红外光谱数据进行特征提取,选出具有较高准确度的贮藏时间鉴别算法,并建立基于模糊鉴别主成分分析的生菜贮藏时间鉴别模型。结果:利用模糊鉴别主成分分析进行特征提取后的鉴别准确度显著增加。主成分分析、鉴别主成分分析和模糊鉴别主成分分析的最高准确度分别为46.67%, 86.67%, 93.33%。结论:采用近红外光谱分析和模糊鉴别主成分分析的鉴别模型具有较高的准确性和优越性。

**关键词:**生菜;贮藏时间;近红外光谱;模糊鉴别主成分分析

**Abstract:** Objective: This study focuses on designing a more accurate identification model of lettuce storage time. Methods: The near-infrared spectral data from extraction feature of the preprocessed lettuce was obtained by principal component analysis (PCA), discriminant principal component analysis (DPCA) and fuzzy discriminant principal component analysis (FDPCA) respectively. An algorithm of higher accuracy in storage time discrimination was explored, and then a lettuce storage time dis-

criminant model based on FDPCA was established. Results: The identification accuracy raised dramatically after FDPCA was used to extract feature. When employing PCA, DPCA and FDPCA algorithms, the highest accuracies achieved were 46.67%, 86.67% and 93.33% respectively. Conclusion: This discrimination model of employing near-infrared spectroscopy and FDPCA was characterized by high accuracy and superiority.

**Keywords:** lettuce; storage time; near infrared spectra; fuzzy discriminant principal component analysis

蔬菜作为日常生活中必不可少的食物,可为人体提供所需的膳食纤维和矿物质<sup>[1]</sup>,也可有效防治各种疾病的发生<sup>[2-3]</sup>。然而,受限于当前的保鲜技术贮藏时间过长,生菜中的营养物质会受到不同程度的影响,降低其营养价值<sup>[4]</sup>。因此借助仪器和算法设计一种鉴别生菜贮藏时间的模型具有较高的研究价值。

徐晓霞等<sup>[5]</sup>通过对不同贮藏温度下的鲜切生菜进行腐败细菌分离,能较准确地判断各种细菌的腐败性及生菜的腐败程度。刘勇等<sup>[6]</sup>利用偏最小二乘回归系数法对波长进行特征提取,对比分析了特征波长和全波长下香肠中亚硝酸盐含量预测模型的检测精度。胡亚南等<sup>[7]</sup>针对食品包装袋膜张力较难控制问题,设计了一种模糊分数阶PID袋膜张力控制系统,该控制系统能够满足包装机袋膜张力控制精度和稳定性需求;王敏等<sup>[8]</sup>利用电子鼻获取食品的气味变化,该技术可以实现无损、低成本、实时地检测冰箱中食品的新鲜度;张桂君等<sup>[9]</sup>对保鲜技术进行了整理归纳与优点判断,并对此研究方向的未来进行了展望。

然而关于生菜贮藏时间和新鲜度的相关研究相对较

**基金项目:**国家自然科学基金资助项目(编号:31471413);安徽省教育厅高校自然科学研究重点项目(编号:KJ2019A1129);滁州职业技术学院校级自科重点项目(编号:YJZ-2020-12);滁州职业技术学院院级人才项目(编号:YG2019026, YG2019024);江苏大学大学生创新训练计划项目(编号:202010299246Y)

**作者简介:**侯晓蕾,女,江苏大学在读本科生。

**通信作者:**武小红(1971—),男,江苏大学教授,博士。

E-mail: wxh\_www@163.com

**收稿日期:**2021-06-02

少。祝红等<sup>[10]</sup>通过对电子鼻的传感器响应值进行 PCA 方法的分析,明显区分出不同贮藏条件下的鲜湿米粉样品,但其经济成本仍较高。汪天宇等<sup>[11]</sup>利用模糊聚类算法计算出相应对象数据集合的聚类中心与试验样本对相应聚类中心的隶属度,按照最大隶属度原则实现对山核桃外壳、内隔(隔、壳内壁等)和核桃仁的合理分类;试验拟分别采用主成分分析(PCA)、鉴别主成分分析(DPCA),以及将模糊集理论与 DPCA 结合的模糊鉴别主成分分析(FDPCA),对数据进行特征提取,并使用 K-最近邻分类算法(KNN)建立生菜贮藏时间鉴别模型,以期寻找更为准确生菜贮藏时间鉴别模型,为生菜的品质检测提供依据。

## 1 材料与方 法

### 1.1 材料与仪器

#### 1.1.1 材料与试剂

生菜:北山 3 号,挑选无腐烂虫害、大小一致的生菜,市售。

#### 1.1.2 主要仪器设备

低温恒温保鲜柜:MIR-554-PC 型,日本三洋电机株式会社;

傅里叶变换近红外光谱仪:Antaris II 型,美国赛默飞世尔仪器公司。

### 1.2 方 法

#### 1.2.1 近红外光谱采集

(1) 材料准备:采集新鲜生菜样本 60 个,放入贴有标签的保鲜袋中,随后放入 4℃ 保鲜柜中贮藏备用。

(2) 仪器预热:为减小试验误差,保证数据的可靠性,采集数据前将近红外光谱仪开机预热 1 h,采集时保持实验室温度为 20~25℃,相对湿度为 70%。

(3) 光谱扫描:利用反射积分球模式采集茶叶近红外光谱,扫描次数 32 次,光谱波数 4 000~10 000 cm<sup>-1</sup>,扫描间隔 3.856 cm<sup>-1</sup>。每隔 12 h 取出所有样本进行近红外光谱检测,共检测 3 次,并分别记为贮藏时间 1、贮藏时间 2 和贮藏时间 3,共获得 180 组近红外光谱数据。其中训练样本 60 个(每类 20 个),测试样本 120 个(每类 40 个)。

1.2.2 模糊鉴别主成分分析(FDPCA) FDPCA<sup>[12]</sup>是一种基于 DPCA<sup>[13]</sup>和模糊集合理论<sup>[14]</sup>的特征提取算法。该算法包括以下步骤:

(1) 聚类中心计算出每类训练样本的均值,将其作为聚类中心  $V$ 。

$$V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} x_{\text{mean}1} \\ x_{\text{mean}2} \\ x_{\text{mean}3} \end{bmatrix}, \quad (1)$$

式中:

$v_i$ ——第  $i$  类训练样本的类中心值(其值与第  $i$  类训

练样本的均值  $x_{\text{mean}i}$  相等,即  $v_i = x_{\text{mean}i}$ )。

(2) 计算模糊隶属度,若  $k^{\text{th}}$  样本  $x_k$  的模糊隶属度值  $u_{ik}^{\text{mf}}$  在第  $i$  类别中最大,则  $x_k$  属于  $i^{\text{th}}$  类。

$$u_{ik}^{\text{mf}} = \left[ \sum_{j=1}^n \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{\text{mf}-1}} \right]^{-1}, \forall i, k, \quad (2)$$

式中:

$u_{ik}^{\text{mf}}$ ——第  $k$  个样本  $x_k$  隶属于第  $i$  类样本的模糊隶属度;

$c$ ——类别数;

$v_i$ ——第  $i$  类训练样本的类中心值;

$n$ ——样本数;

$m_i$ ——权重系数,  $m_i \in (1, +\infty)$ 。

(3) 计算模糊类间离散度矩阵和模糊总体离散度矩阵。

$$S_{\text{FB}} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^{\text{mf}} (v_i - \bar{x})(v_i - \bar{x})^T, \quad (3)$$

$$S_{\text{FT}} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^{\text{mf}} (x_k - \bar{x})(x_k - \bar{x})^T, \quad (4)$$

式中:

$S_{\text{FB}}$ ——模糊类间离散度矩阵;

$S_{\text{FT}}$ ——模糊总体离散度矩阵;

$\bar{x}$ ——样本的总均值,  $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$ 。

(4) 计算最大特征值和特征向量,选出最大特征值  $\lambda_1$  和与之相对应的特征向量  $\psi_1$ ,并将  $\psi_1$  作为最优鉴别向量集的第一个向量。

$$S_{\text{FT}}^{-1} S_{\text{FB}} \psi = \lambda \psi, \quad (5)$$

式中:

$S_{\text{FT}}^{-1}$ ——模糊总体离散度矩阵的逆矩阵;

$\lambda$ ——所求特征值集合;

$\psi$ ——所求特征向量集合。

(5) 计算最优鉴别向量集,根据前  $r$  个最优鉴别向量  $\psi_1, \psi_2, \dots, \psi_r$  ( $r \geq 1$ ) 计算得到第  $(r+1)$  个最优鉴别向量  $\psi_{r+1}$ ,进而获得  $p$  ( $p > r$ ) 个最优鉴别向量组成最优鉴别向量集  $\phi = \{\psi_1, \psi_2, \dots, \psi_p\}$ 。

$$P S_{\text{FB}} S_{\text{FT}}^{-1} \psi_{r+1} = \beta \psi_{r+1}, \quad (6)$$

$$P = I - S_{\text{FT}}^{-1} \phi Q^{-1} \phi^T, \quad (7)$$

$$Q = \phi^T S_{\text{FT}}^{-1} \phi, \quad (8)$$

$$\phi = [\psi_1, \psi_2, \dots, \psi_r]^T, \quad (9)$$

式中:

$\psi_{r+1}$ ——第  $(r+1)$  个最优鉴别向量;

$\beta$ ——第  $(r+1)$  个最优鉴别向量所对应的特征值;

$I$ ——单位矩阵;

$\phi$ ——最优鉴别向量集。

(6) 测试样本的线性转换。

$$Y = [\psi_1, \psi_2, \dots, \psi_p]^T X, \quad (10)$$

式中:

$[\psi_1, \psi_2, \dots, \psi_p]^T$ ——最优鉴别向量集的转置矩阵;

$X$ ——测试样本;

$Y$ ——测试样本投影到最优鉴别向量集上得到的投影样本。

1.2.3 准确度比较 为了比较 PCA、DPCA、FDPCA 的准确度差异,控制以下两个参数变量,计算并绘制准确度变化图:权重指数  $m_i$  为 2,  $K$  近邻分类器的参数  $K$  分别取 1, 3, 5, 7, 9, 11, 13, 15, 17;  $K$  近邻分类器的参数  $K$  为 5, 权重指数  $m_i$  分别取 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 3.0, 3.5, 4.0, 4.5。

1.2.4 数据处理 利用多元散射校正技术(MSC)对未分类的原始光谱数据进行预处理,旨在削弱温度、湿度等其他因素对数据采集的影响;分别利用 PCA、DPCA 和 FDPCA 对校正后的数据进行特征提取,以实现数据降维;利用  $K$ -最近邻分类算法分类,进而建立生菜贮藏时间鉴别模型。

## 2 结果分析

### 2.1 原始数据采集

利用近红外光谱仪分别采集 3 类贮藏时间生菜样本的近红外光谱数据,FT-NIR 图如图 1 所示。由图 1 可知,随着贮藏时间的推移,生菜样本在同一波数下的吸光度略有降低。除了包含对数据分析有价值的生菜化学成分的光谱吸收数据信息外,图 1 中还掺杂着影响数据分析准确度的光散射信息,因此,需要对原始数据进行预处理,使有用信息得到加强。

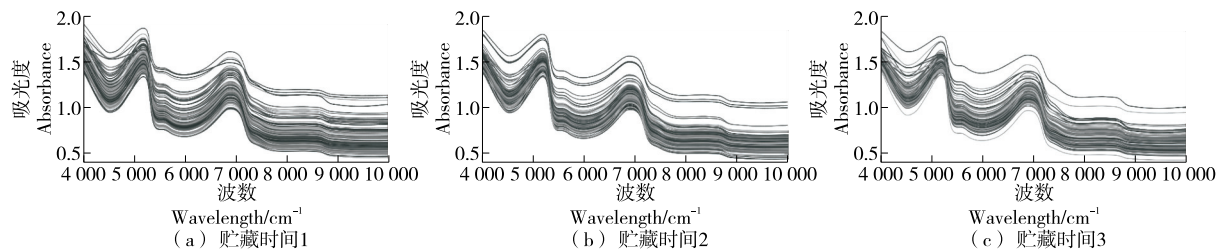


图 1 3 类生菜样本近红外光谱图

Figure 1 FT-NIR spectra of three kinds of lettuce samples

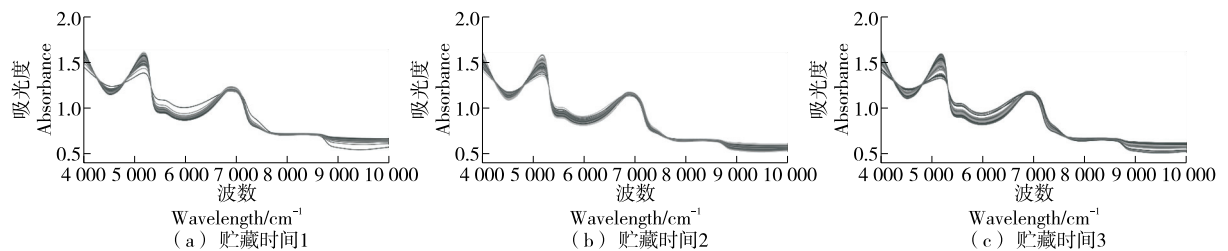


图 2 MSC 预处理后生菜近红外光谱图

Figure 2 FT-NIR spectra pretreated with MSC

### 2.2 数据预处理

近红外光谱仪所获数据不能直接用于生菜贮藏时间的分类<sup>[14]</sup>。当光谱仪长期用于数据采集时,温度和湿度<sup>[15]</sup>等因素会对传感器的灵敏性造成一定干扰,可能存在一定偏差。为此,利用多元散射校正(MSC)对原始数据进行预处理,以削弱温度和湿度等因素对数据的影响,确保数据的真实性。对图 1 进行 MSC 处理后的光谱如图 2 所示。由图 2 可知,MSC 预处理可有效降噪。

### 2.3 DPCA 的分类

结合主成分分析与鉴别分析,DPCA 主要致力于寻找一组最优判别向量进行最优转换,对 MSC 处理后的 1 557 维光谱数据进行降维,获得由 4 个最优判别向量组成的最优判别向量集,结合 KNN 算法对数据分类,测试样本分类情况如图 3 所示。其中,贮藏时间 1 与贮藏时间 2 有 10 处以上重叠,重叠样本越多越不利于提高准确度。

### 2.4 FDPCA 的分类

由图 4 可知,120 个样本在所属类别的模糊隶属度均  $>0.5$ ,在非所属类别的模糊隶属度均  $<0.5$ 。

基于 FDPCA 的样本分类情况如图 5 所示。相较于 DPCA,FDPCA 中相同类型的数据集群仍然紧密地聚合,且贮藏时间 1 与贮藏时间 2 仅有 3 处重叠,说明基于模糊理论的 FDPCA 对分类较为有益,尤其是在集群分布重叠的区域,FDPCA 可以对这些重叠的数据点分配权重,降低数据结构的复杂性,提高分类的准确性。

### 2.5 分类准确度

设置维数为 6, 9, 12, 15, 18, 20, 寻找分类准确度最大

值时的最优维数,结果如表 1 所示。由表 1 可知,将 1 557 维数据降至 20 维,PCA、DPCA 和 FDPDA 的分类准确度均达到最大值。

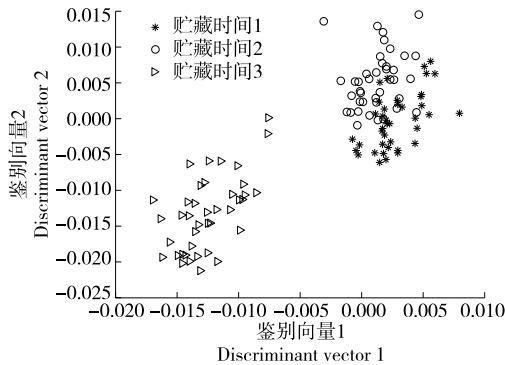


图 3 DPCA 测试样本分类情况

Figure 3 The classification of train samples with DPCA

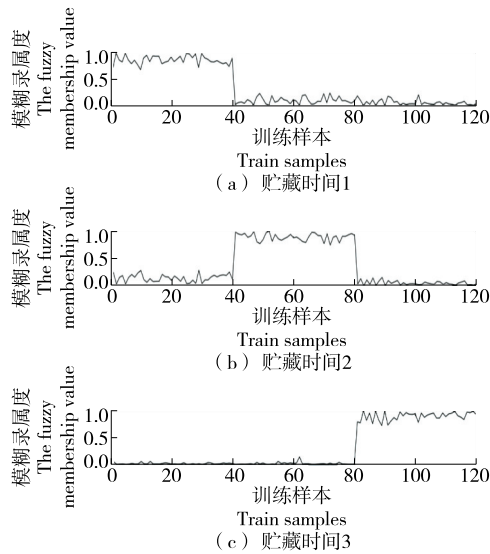


图 4 FDPDA 模糊隶属度

Figure 4 The fuzzy membership values of train samples with FDPDA

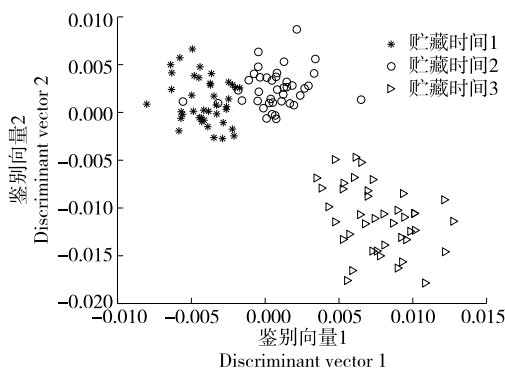


图 5 FDPDA 测试样本分类情况

Figure 5 The classification of train samples with FDPDA

由图 6 和图 7 可知,当 K 近邻分类器的参数 K 与权重指数  $m_f$  取不同值时,FDPCA 的分类准确度均高于 DPCA 和 PCA 的,其中 K 近邻分类器的参数 K 取 5,权重指数  $m_f$  取 2 时,FDPCA、DPCA、PCA 的最高准确度分别为 93.33%,86.67%,46.67%。设置 K 近邻分类器的参数 K 为 5,将预处理后的 1 557 维数据直接进行 KNN 分

表 1 PCA、DPCA 和 FDPDA 在不同维数下的分类准确度

Table 1 Classification accuracies of PCA, DPCA and FDPDA in different dimensions %

维数	PCA	DPCA	FDPDA
6	33.33	50.00	46.67
9	36.67	73.77	80.00
12	36.67	71.67	83.33
15	36.67	75.00	93.33
18	36.67	81.67	91.67
20	36.67	86.67	93.33

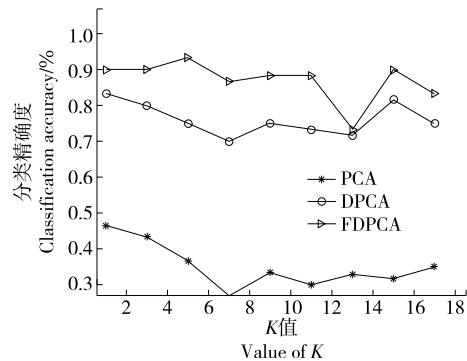


图 6 权重指数  $m_f=2$ ,分类准确度随 K 近邻分类器的参数 K 的变化

Figure 6 Weight=2, the change of classification accuracies with the K value

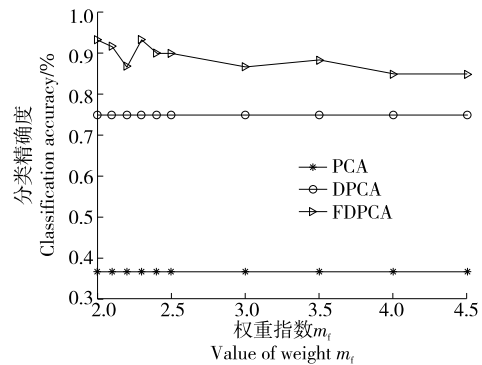


图 7 K 近邻分类器的参数 K=2,分类准确度随权重指数的变化

Figure 7 K=2, the change of classification accuracies with weight  $m_f$

类,其准确度仅为 35%,可能是高维数据无法排除无关甚至有干扰性的特征。因此,基于 FDPCA 的生菜贮藏时间鉴别模型能够有效提取数据特征,提高分类准确度。

### 3 结论

经过对比主成分分析、鉴别主成分分析,以及将模糊集理论与鉴别主成分分析结合的模糊鉴别主成分分析 3 种算法的鉴别准确度,确定了一种基于近红外光谱分析的生菜贮藏时间鉴别系统。结果表明:当 K 近邻分类器的参数  $K=5$ ,权重指数为 2 时,模糊鉴别主成分分析的准确度最高为 93.33%;当 K 近邻分类器的参数  $K=1$ ,权重指数为 2 时,鉴别主成分分析和主成分分析的最高准确度分别为 86.67%,46.67%。通过比较,不同参数设定下利用模糊鉴别主成分分析进行特征提取后的鉴别准确度均显著高于鉴别主成分分析和主成分分析,且与未降维原始数据分类准确度 35%相比,模糊鉴别主成分分析在提高鉴别准确度方面发挥重要作用,即结合模糊鉴别主成分分析和 K 近邻算法分类器是一种有效鉴别生菜贮藏时间的方法。但该模型不能忽略复杂贮藏条件对最终分类结果的影响,后续可将贮藏条件量化,作为变量引入降维后的变量列表中,确定权重后再分类,以此解决复杂贮藏条件引起的误差。

### 参考文献

[1] 邓梦雅,朱丽,吴东慧,等.蔬菜中矿物质含量测定、营养评价及风险评估[J].食品研究与开发,2018,39(9):97-102.  
DENG Meng-ya, ZHU Li, WU Dong-hui, et al. Mineral content and nutritional value evaluation and risk assessment in vegetables[J]. Food Research and Development, 2018, 39(9): 97-102.

[2] 柳青,黄广学,刘继伟,等.六种特菜的营养成分分析与品质评价[J].农产品加工,2019(16):41-44,48.  
LUI Qing, HUANG Guang-xue, LIU Ji-wei, et al. Nutritional composition analysis and quality evaluation of six special dishes[J]. Farm Products Processing, 2019(16): 41-44, 48.

[4] 吴超群,李希,路甲鹏,等.中国居民心血管疾病危险因素分布报告[J].中国循环杂志,2021,36(1):4-13.  
WU Chao-qun, LI Xi, LU Jia-peng, et al. Report on geographical disparity of cardiovascular risk factors in China[J]. Chinese Circulation Journal, 2021, 36(1): 4-13.

[5] 徐晓霞,陈安均,桑伟娜,等.不同温度贮藏鲜切生菜腐败细菌的分离及鉴定[J].食品与发酵工业,2016,42(1):53-58.  
XU Xiao-xia, CHEN An-jun, SANG Wei-na, et al. Isolation and identification of spoilage bacteria of fresh-cut lettuce at different temperatures[J]. Food and Fermentation Industries, 2016, 42(1): 53-58.

[6] 刘峥,殷勇.基于高光谱技术的香肠亚硝酸盐快速检测方法[J].食品与机械,2019,35(5):78-82.  
LIU Zheng, YIN Yong. Rapid detection method of sausage nitrite based on hyperspectral technology[J]. Food & Machinery, 2019, 35

(5): 78-82.

[7] 胡亚南,李鑫,霍蛟飞,等.基于模糊分数阶 PID 的食品包装机张力控制系统设计[J].食品与机械,2020,36(12):79-84.  
HU Ya-nan, LI Xin, HUO Jiao-fei, et al. Design of tension control system of food packaging machine based on fuzzy fractional-order PID[J]. Food & Machinery, 2020, 36(12): 79-84.

[8] 王敏,高凡,张钧煜,等.基于智能电子鼻的冰箱冷藏食品新鲜度原位检测技术[J].传感技术学报,2019,32(2):161-166.  
WANG Min, GAO Fan, ZHANG Jun-yu, et al. In-Situ refrigerator food freshness detection based on smart electronic nose[J]. Chinese Journal of Sensors and Actuators, 2019, 32(2): 161-166.

[9] 张桂君,周茜,王清,等.生菜采后保鲜技术研究现状[J].保鲜与加工,2021,21(1):135-140.  
ZHANG Gui-jun, ZHOU Qian, WANG Qin, et al. Research status of preservation technology on postharvest lettuce [J]. Storage and Process, 2021, 21(1): 135-140.

[10] 祝红,郭宇波,易翠平,等.基于电子鼻的鲜湿米粉新鲜度检测[J].食品与机械,2018,34(7):65-68.  
ZHU Hong, GUO Yu-bo, YI Cui-ping, et al. Freshness detection of fresh rice noodles based on electronic nose[J]. Food & Machinery, 2018, 34(7): 65-68.

[11] 汪天宇,曹成茂,谢承健,等.基于模糊聚类算法的山核桃壳仁分选系统设计[J].食品与机械,2018,34(6):110-114,157.  
WANG Tian-yu, CAO Cheng-mao, XIE Cheng-jian, et al. Design of hickory nuts's shell and kernel sorting system based on fuzzy clustering algorithm[J]. Food & Machinery, 2018, 34(6): 110-114, 157.

[12] 武斌,周树斌,武小红,等. Gath-Geva 联合模糊聚类的生菜近红外光谱聚类分析[J].光谱学与光谱分析,2021,41(3):932-935.  
WU Bin, ZHOU Shu-bin, WU Xiao-hong, et al. Gath-Geva allied fuzzy c-means clustering analysis of NIR spectra of lettuce[J]. Spectroscopy and Spectral Analysis, 2021, 41(3): 932-935.

[13] 赵科文,陈实,蒋浩,等.基于近红外光谱技术的烟丝掺配均匀度测定[J].食品与机械,2020,36(11):183-188.  
ZHAO Ke-wen, CHEN Shi, JIANG Hao, et al. Determination of blending uniformity of cut tobacco based on near infrared spectroscopy[J]. Food & Machinery, 2020, 36(11): 183-188.

[14] 曹念念,刘强,彭菁,等.基于近红外光谱技术的黄桃脆片可溶性固形物和硬度定量检测方法[J].食品与机械,2021,37(3):51-57.  
CAO Nian-nian, LIU Qiang, PENG Jing, et al. Study on quantitative detection of soluble solids and firmness of yellow peach chips by near-infrared spectroscopy[J]. Food & Machinery, 2021, 37(3): 51-57.

[15] 丁倩,石玉,甄润英.不同授粉品种绿宝苹果主要营养品质及主成分分析[J].食品与机械,2018,34(4):54-57.  
DING Qian, SHI Yu, ZHEN Run-yin. Main nutritional quality and principal component analysis of bramley apples with different pollination varieties[J]. Food & Machinery, 2018, 34(4): 54-57.