

基于高光谱技术的酸奶中常见致病菌的快速鉴别及计数

石吉勇, 吴胜斌, 邹小波*, 赵 号, 胡雪桃, 张 芳

江苏大学食品与生物工程学院, 江苏 镇江 212013

摘 要 酸奶是一种发酵型乳制品饮料, 因其特殊的功能性和良好的口感而广受欢迎。但由于商业链的不正当运行, 如奶源非法获取、灭菌不充分等原因, 导致酸奶中致病菌大量滋生, 酸奶中毒事件频繁发生。酸奶中常见的致病菌主要有大肠杆菌、金黄色葡萄球菌和沙门氏菌, 这三种致病菌由人体摄入并达到一定的数量时会产生腹痛、腹泻等严重的消化道疾病, 并且会破坏人体肠道内的正常菌群平衡, 因此国标对奶制品中这三种致病菌的数量已有明确的限量规定。由于酸奶的主要消费对象为老人和小孩, 故其潜在危害不容小觑。传统菌落检测方法虽具有简单, 灵敏、可操作性强等优点, 但当不同菌落混杂在一起时无法同时进行定性定量的检测, 且具有试剂成本高, 检测周期长, 人为因素影响较大等缺点。因此开发一种快速、简单、准确的混合鉴定计数方法为避免致病菌对酸奶的潜在危害提供了有效的途径。高光谱技术同时包含样本的光谱信息与图像信息, 既能够根据化学组分的微小变化进行精确识别(光谱信息), 又能够反映出菌株在外部多层次的变化(图像信息)。因此该研究尝试对比高光谱图像技术和光谱技术, 采用模式识别的方法, 对比不同的模型识别结果, 优选出最佳识别率的识别模型作为计数模型, 最后通过最佳鉴别计数模型的识别分类结果来达到对酸奶中常见致病菌鉴定计数的目的。首先, 购买酸奶中常见的乳酸菌种(保加利亚乳杆菌、嗜热链球菌、嗜酸乳杆菌、干酪乳杆菌、植物乳杆菌)和潜在污染的致病菌种(金黄色葡萄球菌、大肠杆菌、沙门氏菌)等标准菌株进行培养, 提取经过 48 h 培养后的菌落图像信息和光谱信息。采用几种不同的预处理方式(SNV, MC, MSC, 1st DER, 2nd DER)对所提取的光谱数据进行预处理, 并应用遗传算法筛选光谱数据中冗余的波段, 保留有效波段。利用图像处理技术对图像信息中的菌株与培养基背景进行去除, 然后采用主成分分析法从每幅图中优选出 3 个特征波长, 并运用图像处理技术从特征波长所对应图像中提取菌株的 18 个基于 GLCM 的纹理特征信息。挑选合适的主成分分别建立不同的鉴别模型(LDA, KNN, BP-ANN, LS-SVM), 通过其最终的鉴别模型的识别率来确定最佳鉴别计数模型。最后从标准菌株中分别挑选出 30 株进行计数测试, 通过比较模式识别的分类数量结果与菌株的实际数量来验证模式识别效果的准确率。研究表明, 运用 SNV 预处理后光谱数据在提高信噪比效果上明显优于其他几种预处理方式。745.790 8, 773.098 4 和 779.207 0 nm 为图像信息中方差贡献率最大的三个波长, 运用从特征波长所对应的图像中所提取的纹理特征信息建立图像识别模型。通过对比图像信息和光谱信息的模式识别结果发现, 光谱特征鉴别模型普遍优于图像纹理特征鉴别模型, 且当主成分数为 9 时, 运用光谱特征所建立的 LS-SVM 模型的校正集识别率为 96.25%, 预测集的识别率为 91.88%, 为最优模型。采用优选的最优模型对菌株进行识别计数, 大肠杆菌计数的相对误差为 3.33%, 金黄色葡萄球菌和沙门氏菌计数的相对误差均为 0, 验证了高光谱技术应用于酸奶中常见致病菌的鉴别计数的可行性。

关键词 酸奶; 致病菌; 高光谱; 计数; 灰度共生矩阵

中图分类号: TS252.54 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)04-1186-06

收稿日期: 2018-03-10, 修订日期: 2018-07-18

基金项目: 国家自然科学基金项目(31772073, 31671844), 江苏省重点研发计划(BE2016306), 江苏省六大人才高峰(GDZB-016)资助

作者简介: 石吉勇, 1984年, 江苏大学食品与生物工程学院副教授 e-mail: shi_jiyong@ujs.edu.cn

* 通讯联系人 e-mail: zou_xiaobo@ujs.edu.cn

引言

酸奶是一种传统的发酵乳制品,人体摄入酸奶后会产生饱腹感,因此有利于在自然生活条件下控制食欲和体重^[1]。酸奶属于发酵型乳制品,其中添加的乳酸菌能在生长繁殖过程中赋予酸奶酸性,因此人们普遍认为酸奶具有较高的安全性。然而随着生产规模的扩大,产业链的不正常运行、奶源的非法获取等原因,造成酸奶中致病菌的大量滋生,酸奶中毒事件频繁发生。据调查统计,酸奶中常见的致病菌主要是大肠杆菌、金黄色葡萄球菌和沙门氏菌,人体摄入一定量的这三种菌均能引起严重的消化道疾病,国标对此已有明确的限量规定。因此开发一种快速、简单易行、准确的鉴别方法,为避免致病菌对酸奶的潜在危害提供有效的手段。

传统的菌落检测及计数方法主要有国家标准检测法^[2]、基于DNA聚合酶链式反应(PCR)检测法^[3]、机器视觉检测法。此类方法具有准确、灵敏、可操作性强等优点,但当不同种类的菌落混合在一起时,需要将其单个菌落挑出依次鉴定计数,无法对其同时进行鉴定及计数操作。随着近些年来光谱学技术的不断发展,国内外越来越多的研究学者开始将光谱学技术推广应用于微生物检测的各个领域,如Siri-patrawan等采用近红外光谱技术对两种不同的大肠杆菌悬液进行了鉴别^[4],石吉勇、胡雪桃等采用近红外光谱技术实现了对醋醅中常见杆菌的快速鉴别研究^[5]等。这种纯光谱分析方法具有无损、快速、分析稳定性高等优点,但不能将光谱与特定菌落相对应,因此无法同时达到鉴别和计数的效果^[6]。

高光谱技术同时包含了样本的图像信息及光谱信息^[7],能够获取感兴趣区域像素点的特征信息,并结合样本的空间信息可同时达到鉴别和计数的效果。这种鉴别计数方法省去了复杂的前处理、缩短培养周期和计数误差。而近些年来,鲜有将其应用在酸奶中致病菌的检测上。由于菌落在生长过程中其外部(纹理、颜色等)和内部组成(化学组分、代谢产物等)会发生多层次的变化,因此本研究通过提取菌落的图像纹理信息和光谱信息,对比菌落纹理模型和光谱模型,优选出最佳鉴别计数模型来进行酸奶中致病菌的鉴别计数。

1 实验部分

1.1 菌株与培养基

实验所用菌株保加利亚乳杆菌(*Lactobacillus bulgaricus*)、干酪乳杆菌(*Lactobacillus casei*)、嗜热链球菌(*Streptococcus thermophiles*)、嗜酸乳杆菌(*Lactobacillus acidophilus*)、植物乳杆菌(*Lactobacillus plantanum*)、沙门氏菌(*Salmonella*)购买于中国工业微生物菌种保藏中心(China Center of Industrial Culture Collection, CICC),大肠杆菌(*Escherichia coli*)、金黄色葡萄球菌(*Staphylococcus aureus*)由江苏大学食品加工与贮藏实验室提供。本实验中所用的培养基为MRS乳酸菌专用培养基,其主要成分为:水1 000 mL,蛋白胨10 g,牛肉膏5 g,酵母粉4 g,葡萄糖20 g,吐

温80 mL, $K_2HP_4 \cdot 7H_2O$ 2 g, $CH_3COONa \cdot 3H_2O$ 5 g, $C_6H_5O_7(NH_4)_3$ 2 g, $MgSO_4 \cdot 7H_2O$ 0.2 g, $MnSO_4 \cdot 4H_2O$ 0.05 g,琼脂粉15 g, pH 6.2~7.0。

将需要灭菌处理的材料置于121 °C的高压蒸汽灭菌锅中灭菌20 min后立即转入无菌的超净工作台。分别将经过24 h增菌培养的各菌液用生理盐水10倍梯度进行稀释,以 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7} 梯度稀释的菌液涂布在培养基上进行培养。选择两个合适浓度梯度的菌液(平板上生长的菌落数量在30~300 CFU之间)进行平板涂布,经过多次预实验测试证明 10^{-6} 和 10^{-7} 两个浓度梯度的菌液为合适的梯度。其中每种浓度分别涂布6个平板,每种菌两个浓度一共对应12个平板,八种菌总共96个平板,平行三次一共288个平板。

1.2 高光谱的采集与标定

研究中所采用的高光谱成像系统由江苏大学自主开发研制,由两个部分组成,包括硬件系统和软件系统。硬件系统主要由高光谱摄像机(ImSpector, V10E, Spectral Imaging Ltd, Oulu, 芬兰)、两个对称放置的150 W光纤卤素灯(Fiber-Lter DC950型, 美国)、精密水平电控移动平台(TSA200-A)、电子控制箱(SC300-1A, 中国)和计算机等几个部分组成。软件系统主要是SpectralCube软件(芬兰Spectral Imaging有限公司)。

为了防止基线漂移的影响,在使用高光谱成像系统前,需要提前开启系统。高光谱摄像机的曝光时间设置为50 ms,电控移动平台的水平移动速度为 $1.25 \text{ mm} \cdot \text{s}^{-1}$,工作模式为step模式。将经过48 h培养的菌落取出,将培养皿水平放置于电控移动平台上,采集全波段下的菌落图像。实验所用高光谱摄像机为线阵摄像机,分辨率为 $618 \times 1\ 628$,最终得到菌落的 $618 \times 1\ 628 \times 618$ 的三维像素数据块。在高光谱图像数据采集过程中,由于光照强度分布不均匀以及传感器中暗电流的存在,往往在图像中出现较大噪声,并且不同波长下光谱图像的亮度差异较大,因此采集完成后需要对所采集的图像进行黑白板校正^[8],校正公式^[9]如式(1)

$$R = \frac{I - B}{W - B} \quad (1)$$

式中,I表示原始高光谱图像;B表示用黑板校正后的图像;W表示用白板校正后的图像;R表示校正后的光谱图像。

1.3 光谱的提取

研究中采用ENVI 4.5软件提取菌落的特征光谱。随机挑选培养皿中生长较好、大而厚实的菌落为目标菌落,每种菌种提取60条光谱,8种菌共480条光谱,对应得到了一个 480×618 的原始光谱数据集。

高光谱成像系统所采集到的光谱信息中除了含有样本的信息外,还存在着大量的高频噪声、基线漂移、样本背景信息等噪声。因此需要选择合适的预处理方法对原始光谱数据进行预处理,来提高模型的预测效果。本研究采用标准正态变量变换(standard normal variate transformation, SNV)、一阶导数(first derivatives, 1st DER)、二阶导数(second derivatives, 2nd DER)、多元散射校正(multiplicative scatter correction, MSC)、均值中心化(mean centering, MC)五种光谱预

处理方法对光谱数据进行信息增强^[10], 对比几种预处理对光谱的优化效果来确立最佳预处理方式, 然后运用最佳预处理方式处理后的光谱来进行下一步波段的筛选。

1.4 光谱特征波长的筛选

光谱特征波长的选择标准是依据波段的所包含信息含量, 最大限度的提高模型的预测精度、减少模型计算时间。预处理后所得到的光谱数据含有 618 个波长下的光谱数据, 其数据量冗长, 且相邻的波长所包含的信息存在着重叠, 且相关性较大, 因此需要对全波长进行优化筛选。遗传算法 (genetic algorithms, GA) 的原理主要是借鉴了生物界“优胜劣汰”和遗传机制, 利用选择、交叉和突变等算子的操作, 不间断地进行遗传和迭代, 使得较优的变量被保留, 较差的变量被淘汰, 最终使得整体达到一个最优的结果。因此采用 GA 算法进行波长的筛选, 筛选得到的光谱数据进行建模。

1.6 特征波长图像的选取

经过校正后的高光谱图像包含 618 个波长下的图像, 相邻波段的图像可能存在着信息重叠, 反映出的图像信息具有较强的相关性, 使数据在一定程度上产生重叠, 这种信息的冗余严重影响着后期数据处理速度和效率。实验中运用多元图像主成分分析法来选择特征波长图像, 运用 ENVI4.5 软件对校正后的高光谱图像进行主成分分析, 并根据各个主成分方差贡献率的大小, 从中挑选出能代表菌落样本大多数信息的主成分图。采用主成分分析得到的主成分图像由各个波长的图像与其所对应的权重系数以线性组合的方式组合而成, 权重系数越大代表该波长所对应的图像对主成分图的贡献率越高, 此波长所对应的图像即为最佳特征图像。

1.7 鉴别计数模型的建立

在进行模型选择的过程中, 一个更简洁、变量更少的模型往往更受青睐, 因为这类模型不但具有更高的稳定性和鲁棒性^[11], 而且还能大大缩短运算时间和提高运算效率。本研究通过对比线性判别分析 (linear discriminant analysis, LDA)、K-最近邻法 (K-nearest neighbors, KNN)、误差反向传播神经网络 (back propagation-artificial neural network, BP-ANN)、最小二乘支持向量机 (least squares support vector machines, LS-SVM) 等四种模型的识别率来判断最佳的鉴别计数模型。由于 LDA 和 KNN 两种模型的识别结果取决于所选择的主成分数, 因此本研究选取前 10 个主成分和 10 个 K 值对这两种模型进行优化。BP-ANN 模型的输出层单元设为 4 个 (菌株类别), 选择双曲线正切函数为模型的传递函数, 初始权重设定为 0.95, 动量因子和学习因子均设置为 0.1, 模型的收敛误差设置为 0.000 2, 训练迭代次数为 1 500 次。采用交叉验证和网格搜索法对 LS-SVM 模型的最佳参数进行选择, 以定标集交互验证均方根误差最小为指标确定最优的 (γ, σ^2) (径向基核函数正则化参数, 基于 REF 核函数的参数)。经过优选得到, 最佳的主成分为 9, 最佳参数 $\gamma = 2.321 5$, $\sigma^2 = 2.153 2$ 。

2 结果与讨论

2.1 光谱预处理

通过反复比较 SNV, 1st DER, 2nd DER, MSC, MC 五种预处理方法对全波段的光谱处理效果发现, SNV 在减少基线漂移和其他噪声方面的能力较强, 因此后续采用 SNV 作为光谱预处理方法。图 1(a) 为表示菌落的原始光谱图, 图 1(b) 是经过 SNV 处理后的菌落光谱图, 两幅图的横坐标均表示高光谱的波长 (nm), 纵坐标表示反射率 (R 为反射强度)。通过对比原始光谱图和经过 SNV 预处理后的光谱图, 可以看出几种菌落在全波长范围内具有相似的反射率趋势。而在相同波长下, 几种菌落又具有不同的反射率表明几种菌落可以通过建模来做进一步的区分。

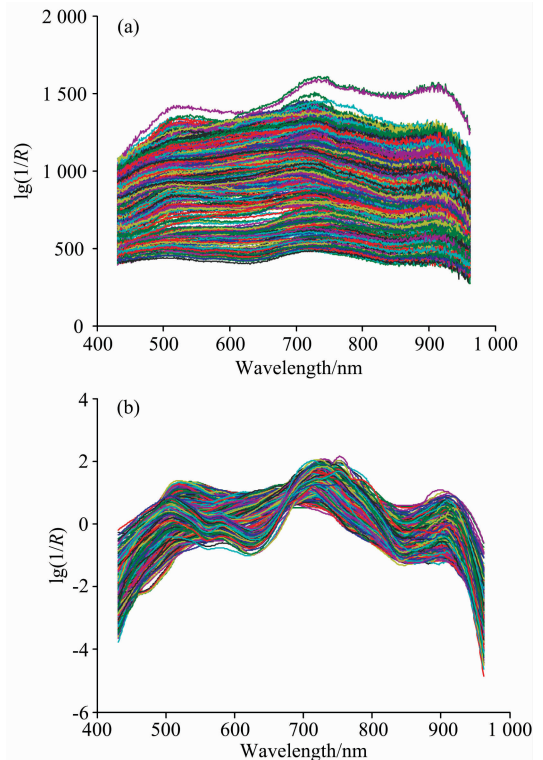


图 1 未经过处理的原始光谱 (a) 和经过 SNV 预处理的光谱 (b)

Fig. 1 Raw spectral (a) of the eight bacteria and SNV preprocessed spectral (b)

2.2 高光谱特征图像的筛选

运用 ENVI 4.5 软件对校正后的高光谱图像进行多元主成分分析。结果表明, 前八个主成分的累计方差贡献率高达 92.77%, 能够有效的代表样本信息, 因此选取了前八个主成分图像作为研究对象。高光谱图像中不仅含有菌落信息还含有大量的培养基等背景信息, 若未及时处理将会影响后期建模效果。因此本实验采取阈值分割法, 将目标菌落与培养基相分离, 再对原图像进行二值化处理形成掩模模板, 将制作的掩模模板依次掩模图像达到分离菌株与培养基背景的目的, 效果如图 2 所示。图 2 中, 菌 A, B, C, D, E, F, G 和 H 分别对应着保加利亚乳杆菌、大肠杆菌、干酪乳杆菌、金黄色葡萄球菌、沙门氏菌、嗜热链球菌、噬酸乳杆菌、植物乳杆菌。

通过比较前八个主成分的方差贡献率可以看出, 各菌株

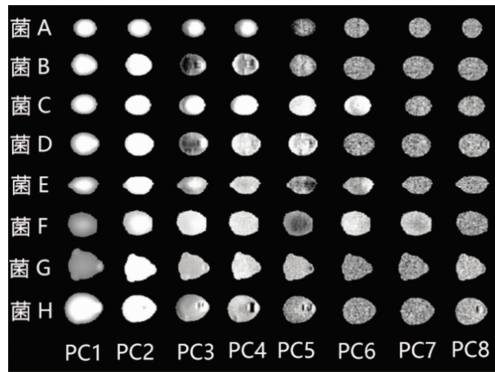


图 2 菌株掩模后的前 8 主成分图像
Fig. 2 The first 8 principal component images after strain masking

的第一主成分(PC1)的方差贡献率均达到 85.00% 以上,能够精确地反映出各菌株图像的绝大部分特征信息,因此本研究从 PC1 图像中寻找特征波长图像。PC1 图像由 618 个波段所对应的图像经过线性组合而成,每个波段分别对应一个权重系数,权重系数越大表明该波长对应的图像对主成分图的贡献率越大,即为特征波长下的图像。通过比较 618 个波长所对应的权重系数,权重系数最大的三个波长分别为 754.790 8, 773.098 4 和 779.207 0 nm。

2.3 高光谱图像纹理特征的提取

纹理特征反映的是图像的灰度像素在空间里的分布,是图像信息中最重要的特征之一。菌落在生长繁殖过程中,其纹理特征随着生长的时间而发生改变。因此本实验每种菌随机挑选 40 样本,分别从每个特征波长图像下的 0°, 45°, 90° 和 135° 几个方向分别提取基于灰度共生矩阵(GLCM)^[12]的能量、相关性、逆差矩、自相关、熵、各向异性等 18 个纹理特征变量。

2.4 光谱信息与纹理特征信息 PCA 比较

由于从菌落信息中所提取的各纹理特征变量、光谱特征变量之间存在着一定的相关性,因此有必要对其进行主成分分析。如图 3 所示,(a)表示纹理特征主成分分析,(b)表示光谱信息主成分分析。从图 3(a)中可以明显的观察到,仅有 3 种菌(嗜酸乳杆菌、植物乳杆菌、沙门氏菌)能够达到较好的区分效果,其他菌种则严重的重叠在一起,并不能达到有效区分效果。相比纹理特征主成分分析,图 3(b)中八种菌落则有较好的聚类效应,仅有少数的重叠在一起。导致两种方式主成分分析效果不同的原因可能是内部识别原理的不同,纹理特征依赖于菌落的外部变化和图像质量,而光谱信号则是通过菌株内部化学成分含氢基团(如 C—H 和 N—H 等)合频和倍频吸收的不同来呈现。

2.5 光谱模型与图像模型比较

高光谱原始波段一共含有 618 个波长,采用 GA 算法进行筛选特征变量后剩余 44 个波长。将 480 个光谱数据和图像纹理数据按照 K-S(Kennard-Stone)^[13]方法分成 320 个校正集和 160 个预测集,分别建立 LDA, KNN, BP-ANN 和 LS-SVM 模型,模型识别结果如表 1 所示。从表 1 可以看出,

采用光谱特征信息所建立的模型识别结果明显优于图像纹理特征信息所建立模型的识别结果,其中采用光谱特征信息所建立的 LS-SVM 模型的校正集和预测集的识别效果均达到了 90% 以上,优于其他光谱模型。其中 LDA 和 KNN 模型的预测效果比 BP-ANN 和 LS-SVM 预测效果差的原因可能是由于 LDA 模型属于线性判别模型,其在解决不同种属的样本分类问题效果较好,而对于同种属,不同类间的效果较差。而 BP-ANN 和 LS-SVM 模型既能解决线性问题又能解决非线性问题具有较好的容错性,另外 LS-SVM 算法不存在困扰神经网络的局部最小值问题,识别效果最优,因此选择 LS-SVM 作为鉴别计数的最佳模型。

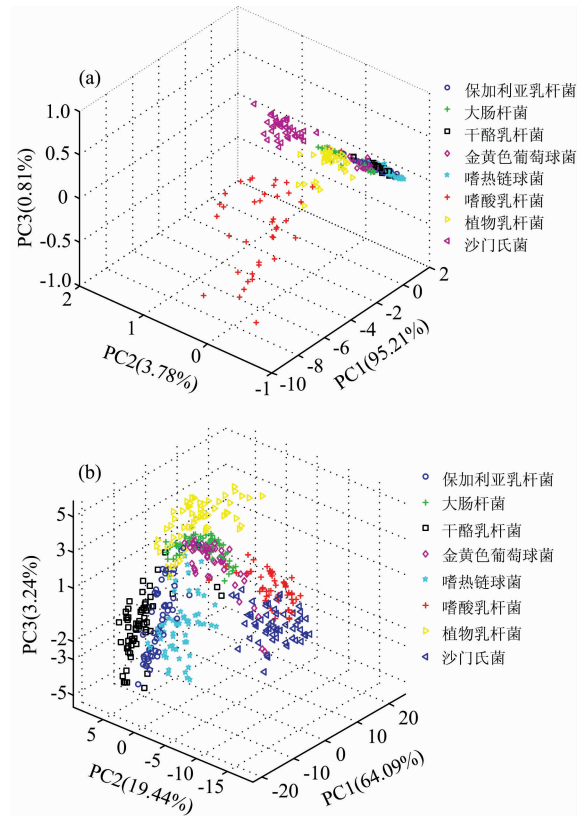


图 3 纹理特征主成分分析 PCA(a) 和光谱特征主成分分析 PCA(b)

Fig. 3 The texture feature (a) and spectral (b) principal component analysis

表 1 光谱信息与图像信息的模型识别率比较
Table 1 Comparison of model recognition rate between spectral and image information

模型	PCs	图像信息识别率/%		PCs	光谱信息识别率/%	
		校正集	预测集		校正集	预测集
LDA	7	78.00	48.33	5	77.81	66.88
KNN	7	77.00	76.67	7	87.50	83.44
BP-ANN	10	85.00	51.67	9	84.06	57.50
LS-SVM	9	93.30	82.86	9	96.25	91.88

2.6 最优模型的计数结果

为了验证模型的计数效果,对八种标准菌株进行了重新

单独培养。将所有菌株分为 4 类, 致病菌大肠杆菌、金黄色葡萄球菌和沙门氏菌分别编号为 1, 2 和 3 类, 其余菌种(保加利亚乳杆菌、干酪乳杆菌、嗜热链球菌、嗜酸乳杆菌、植物乳杆菌)归为第 4 类。每一类菌种选择生长较大、较好的菌落各 30 株, 其中第 4 类中每种菌落各 6 株, 对应的五种菌落总共 30 株。这样 4 类一共有 120 株, 提取每种标准菌光谱信息, 代入最优模型进行验证。根据模型最后识别编号来确定分类计数结果, 其计数结果如表 2 所示。由表 2 可知, 4 类菌种的模型识别计数效果均达到了较好的效果, 其中沙门氏菌和金黄色葡萄球菌计数的相对误差为 0。大肠杆菌计数的相对误差为 3.33%, 分析其原因, 可能是大肠杆菌与保加利亚乳杆菌、干酪乳杆菌、嗜酸乳杆菌、植物乳杆菌等几种菌均属于杆菌, 其内部分泌物或部分结构相似, 模式识别将其误判为其他杆菌所致。

表 2 模型鉴别计数结果

Table 2 The counting results of model discrimination

	标准集	识别集	相对误差/%
1	30	29	3.33
2	30	30	0
3	30	30	0
4	30	30	3.33

对酸奶中常见致病菌的快速鉴别与计数。采集经过 48h 培养的菌落光谱数据及图像数据, 运用 SNV, 1st DER, 2nd DER, MSC 和 MC 五种不同的预处理方式对光谱数据进行预处理。运用多元主成分分析法对高光谱图像进行特征主成分的选取, 根据各波长图像在特征主成分图中所占权重来确定特征波长, 并在特征波长所对应的菌落图像中提取纹理特征信息。对光谱信息和图像纹理信息进行主成分分析去除冗余信息, 采用 GA 算法对光谱数据进行波段的筛选。分别对菌落的光谱信息与图像纹理信息建立 LDA, KNN, BP-ANN 和 LS-SVM 等几种模型, 通过比较模型的识别率来优选最佳的计数模型。结果表明, 经过 SNV 处理的光谱的建模效果较好, 图像信息中方差贡献率最大的 PC1 对应的三个特征波长分别为 754.790 8, 773.098 4 和 779.207 0 nm。对比采用图像纹理信息和光谱信息所建立模型的识别效果可知, 光谱信息所建模型的识别率普遍优于图像纹理信息所建模型的识别率, 其中光谱模型中 LS-SVM 识别率最高, 当主成分为 9 时, 其校正集识别率为 96.25%, 预测集识别率为 91.88%。采用最优模型进行鉴别计数, 均达到了较好的效果。其中金黄色葡萄球菌和沙门氏菌的计数相对误差为 0, 大肠杆菌的计数相对误差为 3.33%, 验证了高光谱技术应用于酸奶中常见致病菌鉴别计数的可行性。

3 结 论

将高光谱技术应用到酸奶中常见致病菌的检测, 实现了

References

- [1] Tremblay A, Doyon C, Sanchez M. Nutrition Reviews, 2015, 73(suppl 1): 24.
- [2] QIN Guang-li, XIAO Wen-jing(秦广利, 肖文静). Farm Products Processing(农产品加工), 2017, 17(9): 65.
- [3] Huang L, Zhao J, Chen Q, et al. Food Research International, 2013, 54(1): 822.
- [4] Siripatrawan U, Makino Y, Kawagoe Y, et al. Sensors & Actuators B Chemical, 2010, 148(2): 367.
- [5] SHI Ji-yong, HU Xue-tao, ZHU Yao-di, et al(石吉勇, 胡雪桃, 朱瑶迪, 等). Transactions of the Chinese Society for Agriculture Machinery(农业机械学报), 2016, 47(2): 245.
- [6] ZHANG Chu, LIU Fei, KONG Wen-wen, et al(张初, 刘飞, 孔汶汶, 等). Transaction of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(20): 271.
- [7] ZOU Xiao-bo, SHEN Ting-ting, ZHU Yao-di, et al(邹小波, 申婷婷, 朱瑶迪, 等). Transaction of the Chinese Society of Agricultural Machinery(农业机械学报), 2016, 47(1): 216.
- [8] Polder G, Gerie V D H, Keizer L, et al. Journal of Near Infrared Spectroscopy, 2003, 11(3): 194.
- [9] Huang M, Wan X, Zhang M, et al. Journal of Food Engineering, 2013, 116(1): 46.
- [10] Huang M, Wang Q, Zhang M, et al. Journal of Food Engineering, 2014, 128(1): 25.
- [11] ZHAO Jie-wen, LIN Hao(赵杰文, 林灏). Date Processing and Analysis Methods in Food and Agricultural Product Testing(食品农产品检测中的数据分析和处理方法). Beijing: Science Press(北京: 科学出版社), 2012.
- [12] ZOU Xiu-guo, DING Wei-min, CHEN Cai-rong, et al(邹修国, 丁为民, 陈彩蓉, 等). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2014, 30(10): 139.
- [13] PAN Guo-feng(潘国锋). Chinese Journal of Spectroscopy Laboratory(光谱实验室), 2011, 28(5): 2701.

Rapid Identification and Enumeration of Common Pathogens in Yogurt Using Hyperspectral Imaging

SHI Ji-yong, WU Sheng-bin, ZOU Xiao-bo*, ZHAO Hao, HU Xue-tao, ZHANG Fang
School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract Yogurt is a kind of fermented dairy beverage, and it is celebrated for its special functionality and good taste. However, due to the improper operation of the commercial chain, such as the illegal acquisition of milk sources and so on, the pathogenic bacteria in yogurt are widespread, resulting in frequent occurrence of yogurt poisoning. The main pathogenic in yogurt are *Escherichia coli*, *Staphylococcus aureus* and *Salmonella*. Human consumption of these three kinds of bacteria will cause severe digestive tract diseases and destroy the balance of normal flora in the human intestine after reaching a certain number. Therefore, the Chinese National Standard has a clear limit on the number of the three pathogens in dairy products. Because the main object of yogurt consumption is the old and the children, the potential harm of yogurt should not be underestimated. The traditional colony detection method is simple, sensitive and operable, but when different colonies are mixed together, it can not be qualitatively detected at the same time, and there are shortcomings such as high cost, long detection cycle and human factors. Therefore, it is of great practical significance to develop a fast, simple and accurate mixed identification count method to avoid the potential hazards of pathogenic bacteria in yogurt. Hyperspectral technology integrates the spectral information and spatial location information of the sample. It can not only accurately identify according to the tiny change of chemical components (spectral information), but also reflect the multi-level changes of the strain (image information). Therefore, this study adopts pattern recognition method to compare different models established by the image information and spectral information, and selects the best counting model based on the recognition rate of the model. Finally, the identification and counting of common pathogenic bacteria in yogurt were realized by the classification results of the best model. Firstly, the standard strains of lactic bacteria and potentially contaminated pathogenic bacteria in yogurt were cultured, and the colony image information and spectral information after 48 h, culture were extracted. Then, different pre-processing methods (SNV, MC, MSC, 1stDER, 2ndDER) were used to reduce the spectral data, and the genetic algorithm was used to reduce the excess spectral bands. The image of agar background used image processing technology to mask removal, then 3 characteristic wavelengths were selected from each map by principal component analysis, and 18 texture feature based on gray-level co-occurrence matrix texture information were extracted from the strain of characteristic wavelengths. Different discriminant models (LDA, KNN, BP-ANN, LS-SVM) were established by selecting the appropriate principal component, and the best discriminant model was determined by the recognition rate of the final discriminant model. Finally, 30 strains from each standard strain were selected for counting test, and the accuracy of pattern recognition was verified by comparing the results of classification and quantity of pattern recognition with the actual number of strains. The results showed that the spectral data pretreated by SNV were superior to other pre-treating methods. The 745.790 8, 773.098 4 and 779.207 0 nm were the characteristic wavelengths. Through the contrast of image pattern recognition and spectral information rate results, it was found that the spectral characteristics of the differential model were better than those of the image texture feature identification model, and when the number of principal component was 9, the LS-SVM spectral model was the optimal model, and the recognition rate of the correction set is 96.25%, and the recognition rate of the prediction set is 91.88%. The optimal model was applied to recognize and count the strains. The relative error of *Escherichia coli* count was 3.33%, and the relative error of count of *Staphylococcus aureus* and *Salmonella* was 0, which verified the feasibility of applying hyperspectral technology to identify and count common pathogenic bacteria in yogurt.

Keywords Yogurt; Pathogen; Hyperspectral; Count; Gray-level co-occurrence matrix

(Received Mar. 10, 2018; accepted Jul. 18, 2018)

* Corresponding author