PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2023,50(6):1489~1496
www.pibb.ac.cn

基于拉曼技术的单细胞生长检测方法*

李新立　张欣雨　杨　强　李肃义**
（吉林大学仪器科学与电气工程学院，长春 130061）

摘要　目的　单细胞生长检测可以更加科学地揭示微生物代谢变化的规律，为后期微生物工程应用提供指导。针对微生物生长应用于食品安全期和最佳食用期的精准检测问题，本文提出一种基于拉曼技术的单细胞生长检测方法。**方法**　首先，通过同步培养实验采集了枯草芽孢杆菌两个批次共900个单细胞拉曼光谱（SCRS）数据，其中600个用于训练和测试，另一批次300个用于模型验证。其次，基于主成分分析的特征关系矩阵，提出CP-SP特征评估方法以筛选SCRS特征用于模型检测。再基于XGBoost构建检测模型，并应用网格搜索和交叉验证对检测模型进行调优。最后，应用混淆矩阵、ROC曲线评估模型对细胞滞后期、对数期和稳定期的检测准确率、敏感性和特异性。**结果**　选用CP-SP筛选的第一、第二和第四主成分较特征贡献率前3个主成分的分类性能提高了3.1%，调优后的细胞生长检测模型测试准确率为96.0%，验证准确率为92.3%。**结论**　基于拉曼技术的单细胞生长检测方法能准确识别单细胞生长状态且具有较高的泛化能力，可为食品安全和保鲜制定精准调控机制提供科学指导。

关键词　单细胞拉曼光谱，细胞生长，食品安全和保鲜，特征关系矩阵，XGBoost
中图分类号　Q26，R857.3，TP391.41　　**DOI**：10.16476/j.pibb.2022.0311

单细胞生长检测可以更加科学地揭示微生物代谢变化的规律，在食品污染物检测中，细胞生长状态是重要的监控指标[1]。细胞生长主要分为滞后期（lag phase）、对数期（log phase）、稳定期（stationary phase）和凋亡期（apoptosis phase）4个时期[2]，不同的生长时期表现出不同的代谢和生产能力。能够在生长的不同时间点识别和检测特定的细菌是非常重要的，抑制食品污染一方面可以通过抑制食品中腐败菌（或致病菌）分裂增殖使其停留在滞后期，另一方面，通过诱导微生物源保鲜剂（如枯草芽孢杆菌（*Bacillus subtilis*）[3]等）进入对数期来抑制腐败菌生长，以保持食品良好的感官品质和理化特性，有效延长食品货架期[4]。

在微生物生长检测中，细胞形态学[5]和吸光度检测[6]已经成为检测的金标准，但检测依赖于群体细胞培养，且检测周期较长，可能错过抑制腐败菌增殖的最佳时期。基因芯片[7]和聚合酶链式反应（PCR）[8]等分子生物学检测技术虽然灵敏度高，但检测需要破坏细胞结构，且无法检测低丰度微生物，而食品污染通常是由很少的细菌或菌落引起的[9-10]。单细胞拉曼光谱（SCRS）技术具有快速、灵敏和原位非侵入的检测优势，已经被用于在菌株水平上识别细菌[11-12]，进一步报道称，SCRS具有为胞内蛋白质、核酸、脂类等提供生物分析的化学组成和结构信息的优势，可以从单细胞水平上检测食品中微生物生命周期各个阶段[13-14]，相较研究者应用的随机森林方法对不同生长时期SCRS数据91.2%识别准确率[15]，极限梯度提升（eXtreme gradient boosting，XGBoost）模型继承了集成学习、树形结构的高可靠性和特征识别能力，可以更有效地识别SCRS所蕴含的细胞“指纹”信息。

因此，本文提出一种基于拉曼技术的单细胞生长检测方法，用于解决食品工程中食品安全期和最佳食用期的精准检测问题。首先，采集枯草芽孢杆菌两个批次共900个SCRS数据，分别作为模型训

* 国家重点研发计划（2022YFC2807904）和吉林大学研究生创新基金（2022059）资助项目。
** 通讯联系人。
Tel：13756887890，E-mail：lsy@jlu.edu.cn
收稿日期：2022-07-05，接受日期：2022-08-17

练、测试和验证数据。其次，分别基于聚合度和耦合度评估方法，提出基于聚合度（compactness，CP）和耦合度（separation，SP）联合（CP-SP）的特征评估方法，筛选最优的SCAS检测特征。然后，基于检测特征构建XGBoost细胞生长检测模型，应用网格搜索和交叉验证对检测模型进行调优，并应用混淆矩阵、受试者操作特征（receiver operating characteristic curve，ROC）曲线评估模型对细胞生长检测准确率、敏感性和特异性。最后，应用检测模型检测另一批次SCRS数据生长状态，验证模型泛化能力。

1 数据和方法

基于拉曼技术的单细胞生长检测方法主要流程如图1所示，步骤包括：a. SCRS数据采集，确定数据样品、采集条件和数据划分；b. 数据预处理，确定SCRS数据预处理方法和参数；c. 特征评估与筛选，提出基于CP-SP的特征评估方法，用来筛选出具有高内聚、低耦合的SCRS特征组合；d. 优化XGBoost超参数和确定模型评估方案，构建细胞生长检测模型并确定模型评估方法；e. 方法应用，调用检测模型，验证模型的泛化能力。

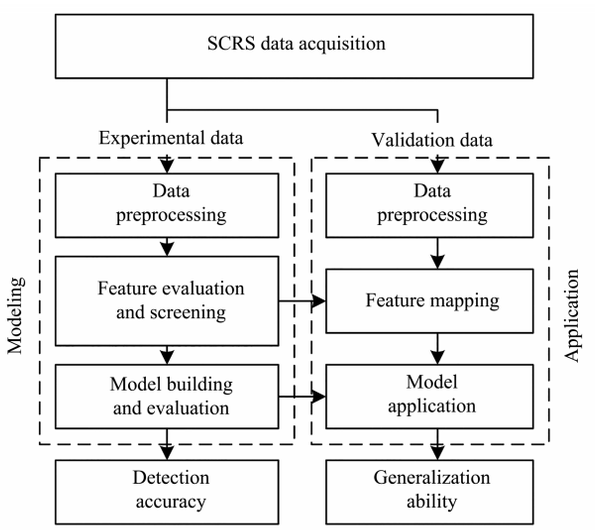


Fig. 1 Experimental procedure of cell growth detection

1.1 SCRS数据采集

1.1.1 分光光度计检测和SCRS检测条件

在细胞培养实验中，从接种时刻起，每隔1 h吸取3 ml菌液，应用紫外分光光度计检测，记录菌液在600 nm处的吸光度（A_{600}），以A_{600}值作为纵坐标，培养时间作为横坐标绘制细胞生长曲线，图2是3组平行实验组重复测量的细胞A_{600}，制作生长曲线，为消除误差所带来的影响，在3组平行实验相同的时间点各吸取中层菌液1 μl，均匀点在拉曼芯片（基底材料硅酸盐玻璃上覆盖对细胞无损伤的镀层金属Al）上风干，在Hooke P300（中国科学院长春光机所和Hooke Instruments研发）拉曼显微镜下观察风干样品。Hooke P300激光光斑约为1 μm（100×物镜，NA为0.8，激发波长532 nm，衍射直径1.22×0.532 μm/0.8），所采集SCRS基本覆盖单细胞（宽度为（0.4±0.1）μm，长度为（1.3±0.5）μm）的大部分信息。

SCRS检测可以获取单个细胞生长过程的实时生化变化，是进行单细胞活体生长检测的实用工具。由于微生物SCRS在600~1 800 cm^{-1}波段具有明显的光谱响应，可以作为其表型指纹区域，Hooke P300光谱仪主要参数设置为：激发波长（excitation wavelength）532 nm，光栅（grating）1 200 g/mm，激发功率（laser power）3 mW，积分时间（integration time）8 s。

1.1.2 微生物样品选择与同步培养

枯草芽孢杆菌是常见的微生物源食品保鲜剂[16]，通过竞争生长空间、资源或产生抗菌化合物（芬芳素等）来抑制腐败菌生长，准确检测并延长食品中腐败菌的滞后期和微生物源保鲜剂的对数可以较好维持食品感官品质和营养价值，延长食品货架期。

本文选用枯草芽孢杆菌作为细胞生长检测的微生物样品并进行同步培养实验，记录微生物的培养时间，根据图2生长曲线显示：培养2 h为滞后期，该时期细胞分裂迟缓，繁殖极少；3~5 h为对数期，微生物在该时期生长迅速，呈现指数生长趋势；培养至8~14 h进入稳定期。本研究对高浓度培养液未做稀释也未建立相关稀释倍数间的回归方程，仍然可以得到大部分菌株的生长曲线。尽管有些菌株生长曲线没有明显衰亡期，但这并不影响生长曲线的走向以及对菌株对数期和稳定期的判断。随着培养时间延长，未发现明显的凋亡期界限，这并不影响生长曲线的趋势以及对单细胞生长时期的识别，这里仅涉及细胞前3个生长时期检测。分两批次采集900个枯草芽孢杆菌SCRS数据，第一批次600个（200个/时期×3时期）作为实验组用于训练和测试，另一批次300个（100个/时期×3时期）作为验证组用于模型验证，检验模型泛化能力。

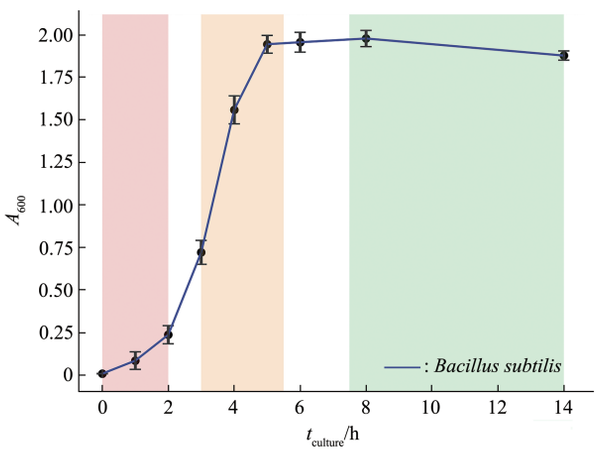


Fig. 2　Growth curve for simultaneous culture of single−cell microorganisms

1.2　数据预处理

SCRS数据预处理是准确鉴定细胞生长时期的前提，通过拉曼光谱仪采集的SCRS数据包含大量的干扰信息，如光谱仪噪声、荧光背景等，干扰信息使得检测模型的识别性能降低，此外，SCRS数据包含1 340个光谱信号像素点，信息较多，为了提高模型检测稳定性，在分析数据之前，需要对SCRS数据进行预处理。本文应用Hooke intP拉曼分析软件（V2.0，Hooke P300配套软件）对SCRS数据预处理，包括：应用Savitzky-Golay算法对SCRS数据进行平滑滤波，窗口宽度为7个像素点，采用3阶多项式拟合；应用AirPLS算法去除SCRS数据背景信号，Lambda=15，最大迭代次数ItermaxAirPls=12；应用Min-Max对SCRS数据归一化处理。

1.3　模型建立与优化

1.3.1　特征提取方法

主成分分析（principal component analysis，PCA）是一种无监督降维方法，利用正交变换把线性相关的高维SCRS变量转换为少数线性无关的特征变量，该线性无关光谱特征也称为主成分。PCA可以将高维SCRS数据转换为少数几个主成分光谱来表征全谱信息，降维后的SCRS特征作为检测模型输入特征在降低计算复杂度、提升数据处理速度时，应尽量减少光谱信息量的损失。

1.3.2　XGBoost模型与优化

XGBoost模型继承了集成学习、树形结构的高可靠性和特征识别能力，能精准识别SCRS所蕴含的细胞指纹信息，实现复杂环境下单细胞生长时期精准检测。本文将特征提取的SCRS光谱特征输入XGBoost检测模型，并通过不断迭代学习SCRS预测值与真实值的残差，确定新的决策树，将树的累加结果逐步逼近真实值进而完成训练，然后以细胞生长时期（滞后期、对数期和稳定期）预测概率作为判别依据，应用XGBoost对单细胞生长时期检测，在损失函数中增加L2正则化项防止模型过拟合。

使用网格搜索和交叉验证（GridSearchCV）对XGBoost检测模型进行参数调优，通过网格搜索（GridSearch）遍历XGBoost模型主要超参数组合，基于先验知识通常设置XGBoost基分类器模型（booster）为树形结构（gbtree）或线性结构（gbliner）；设置分类器个数（n_estimatores）的典型值为80 100和120；设置树深度（max_depth）典型值一般为5、6和7。并应用3折交叉验证（CV）方法优化模型训练性能，降低抽样随机性所带来的预测误差，提高模型的泛化能力。

1.4　模型评估

1.4.1　基于CP-SP的特征评估

传统PCA特征选择往往借助特征累计贡献率来选择光谱特征，但贡献率大的特征作用于分类器性能并不一定大，本文借鉴聚合度和耦合度评估方法，直接从最优分类性能角度出发，通过计算PCA降维后SCRS散点分布的簇内聚合度和簇间耦合度，提出基于CP-SP的特征评估方法，用来筛选出具有高内聚、低耦合的SCRS特征组合，以下为CP-SP核心实现算法：

a. $\overline{CP}$用来表示PCA降维后簇内（某生长时期）SCRS数据点的聚合度[17]，见公式（1）和公式（2）：

$$\overline{CP_i} = \frac{1}{|\Omega_i|}\sum_{x_i \in \Omega_i}\left\| x_i - w_i \right\| \tag{1}$$

$$\overline{CP} = \frac{1}{k}\sum_{k=1}^{k}\overline{CP_k} \tag{2}$$

针对某生长时期SCRS数据集Ω_i，用1−范数用来计算簇内SCRS样本与簇中心点的距离之和，然后计算3个（k = 3，表示3个生长时期）簇紧密性平均值表征细胞生长时期特征关系矩阵的聚合度，$\overline{CP}$值越高意味着相同生长时期SCRS数据聚合度越高。

b. $\overline{SP}$用来表示PCA降维后簇间SCRS数据点的耦合度[18]，见公式（3）：

$$\overline{SP} = \frac{2}{k^2 - k}\sum_{i=1}^{k}\sum_{j=i+1}^{k}\left\| w_i - w_j \right\|_2 \quad (3)$$

针对3个细胞生长时期，用2-范数来计算簇中心之间的平均距离表征细胞生长时期特征关系矩阵的耦合度，同公式（1）中w表示簇中心，$\overline{SP}$值越高意味不同生长时期SCRS数据耦合度越低。

c. 聚合度和耦合度是对SCRS散点分布独立性度量的两个标准，只有同时保证分布的簇内高内聚和簇间低耦合才能实现最大分类性能。本文设置聚合度与耦合度相同的权重系数，将二者比值计为CP-SP得分（S_{CP-SP}）见公式（4），计算PCA降维后SCRS散点分布的簇内聚合度越大和簇间耦合度越小，CP-SP得分越高，分类性能越好。

$$S_{CP-SP} = \overline{CP} / \overline{SP} \quad (4)$$

1.4.2　检测结果评估

随机选取实验组600个SCRS数据的20%作为测试数据（每个生长时期40个预测样本）用于评估检测细胞生长结果，以某一细胞生长时期（稳定期）为例，使用真阳（true positive，TP）表示样本真实为稳定期，预测为稳定期，真阴（true negative，TN）表示样本真实非稳定期，预测为非稳定期，假阳（false positive，FP）表示样本真实非稳定期，预测为稳定期，假阴（false negative，FN）表示样本真实为稳定期，预测为非稳定期。基于此，本文使用混淆矩阵和ROC曲线评估检测结果。

a. 混淆矩阵[19]是机器学习中常用的多分类结果的可视化工具，应用混淆矩阵评估XGBoost模型检测结果，能直观地表示每个生长时期预测准确率，其横坐标为生长时期真实标签，纵坐标为生长时期预测标签。混淆矩阵单元格数值表示测试样本在3个生长时期的预测占比，对角线的值表示检测模型对该生长时期预测准确率（precision，P）。

$$P = TP/(TP + FP) \times 100\% \quad (5)$$

b. ROC曲线[20]用来评估XGBoost模型敏感性和特异性，在二维坐标轴中，横坐标为假阳性率（false postive rate，FPR），表示特异度，纵坐标为真阳性率（true postive rate，TPR），表示灵敏度。AUC（area under the curve of ROC）值表示ROC曲线下方的面积，是XGBoost模型敏感性和特异性的量化表示方法。

$$FPR = FP/(FP + TN) \quad (6)$$

$$TPR = TP/(TP + FN) \quad (7)$$

2　结果和讨论

2.1　SCRS数据处理和分析结果

应用Hooke intP软件对两批次同步培养的枯草芽孢杆菌SCRS数据进行预处理，用堆叠图（stacked lines by Y offsets）显示实验组滞后期、对数期和稳定期3个生长时期SCRS数据预处理效果，如图3所示，分别以实线和阴影部分表示各生长时期200个SCRS数据平均值和方差，由于微生物生长过程中的异质性较为稳定，表现出3个生长时期光谱具有较低的方差。

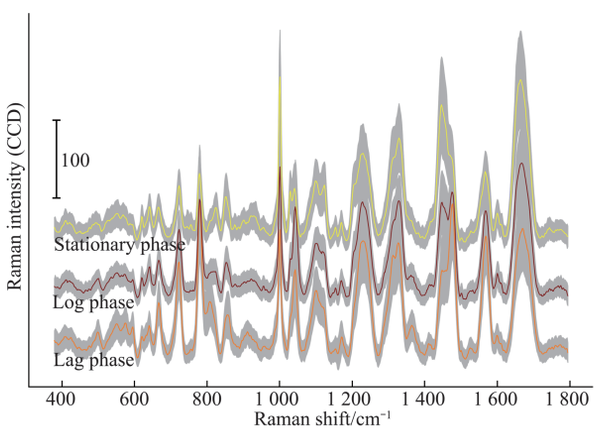


Fig. 3　Pre-processing results of SCRS data for different growth periods

对枯草芽孢杆菌3个生长时期SCRS数据做探索性数据分析（EDA），分别用密度直方图（图4a）和带抖动点的箱线图（图4b）来观测3组数据信噪比（SNR）分布情况，其中滞后期光谱信噪比均值和方差为（6.8±2.6），对数期光谱为（6.6±2.6），稳定期光谱为（6.7±2.7），3个生长时期SCRS数据特征呈现较为稳定的均匀分布，保证了预期检测结果不受光谱SNR影响。

2.2　基于CP-SP的特征评估结果

为了可视化显示效果，在预处理后的SCRS数据集中，每个细胞生长时期随机选择50个光谱数据进行主成分分析，应用Python可视化库Seaborn中Pairplot函数绘制光谱前10个PCA特征关系矩阵，图5显示前4个SCRS特征关系效果，其中主成分PC_m、PC_n的特征关系分布对应于图中［m，n］坐标，对角线（m=n）表示3个生长时期在PC1、PC2、PC3、PC4四个PCA特征上的分布，非对角

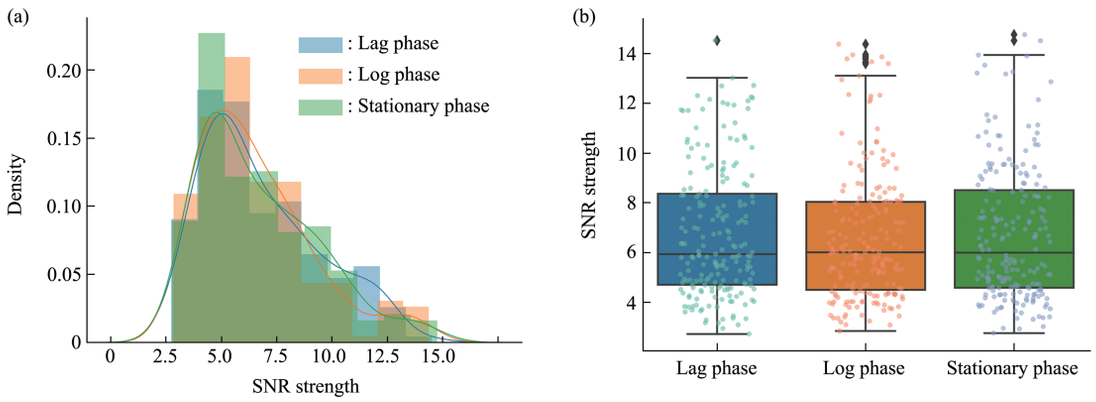


Fig. 4　Results of SCRS data preprocessing for different growth periods

(a) Density histogram of SNR. (b) Boxplot with jittered points.

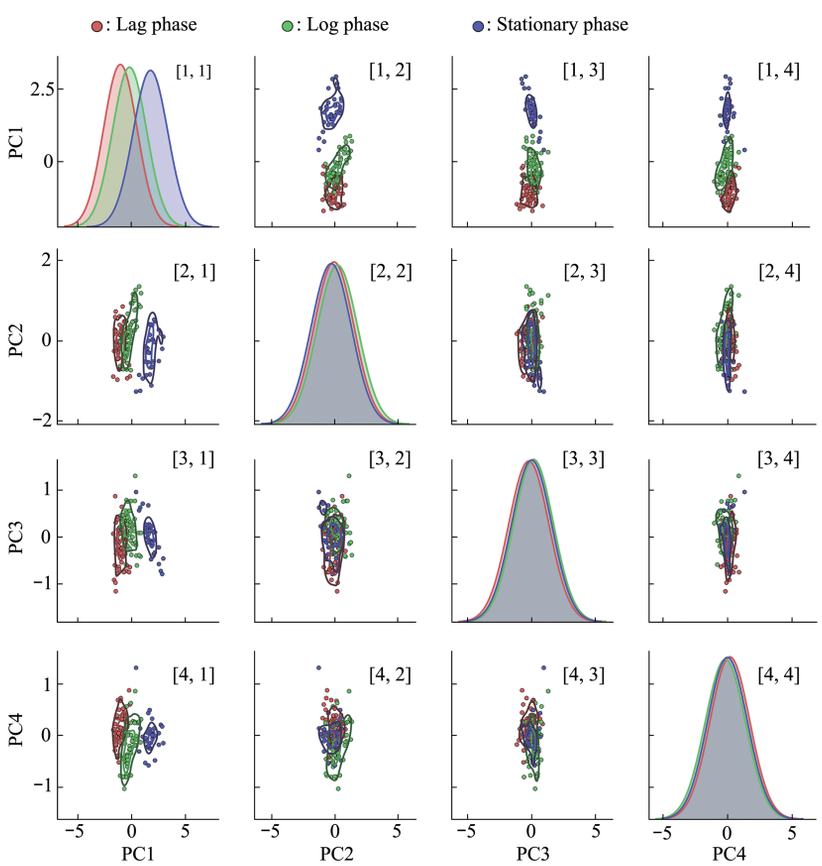


Fig. 5　PCA features relationship matrix of SCRS data

线（$m \neq n$）用核函数密度估计图（Kdeplot）表示两个不同特征之间的相关图。观察SCRS数据特征分布情况，发现SCRS的PC1特征不论是从对角线上的分布图还是与其他特征构成的散点图，都能按生长时期标签表现出明显的区分，同时，PC1与PC2、PC1与PC3、PC1与PC4表现出3个清晰可分离的、簇密度图。

如何量化评估图5 SCRS数据的聚合度和耦合度，本文提出基于CP-SP的特征评估方法来量化SCRS数据的1 340维特征的簇内聚合度和簇间耦

合度，图6是根据公式（4）计算的CP-SP得分和特征贡献率组成的PCA前10个光谱特征评估结果，其中左下角为PCA特征贡献率，右上角热图对应显示PCA特征关系矩阵中$m<n$坐标位置CP-SP得分，分别计算每个PCA特征与其余9个特征CP-SP得分均值和方差。为减小检测模型输入特征复杂度，对比图6中贡献率大于3%的4个特征中，PC4（3.00±1.79）较PC3（2.01±1.83）具有更高的CP-SP得分，对比PC1（5.64±1.26）、PC2（2.79±1.79）分别与PC3和PC4组成检测模型输入特征，PC1、PC2、PC4较PC1、PC2、PC3作为输入特征的XGBoost检测准确率（P）提升3.1%，CP-SP评估筛选的检测特征以更少的光谱信息损失实现更高的SCRS数据生长时期分类性能。

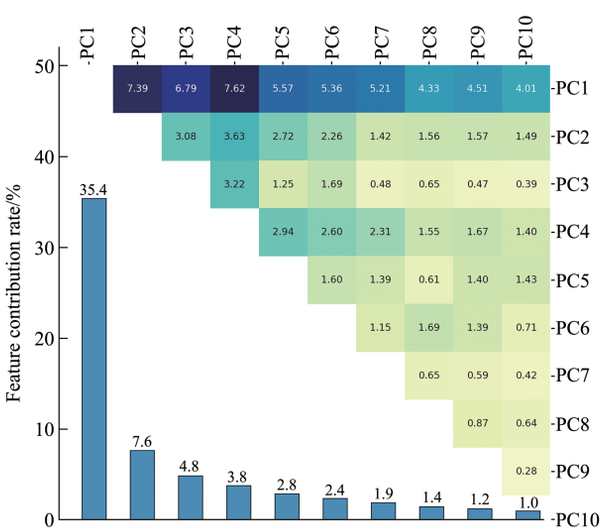


Fig. 6　CP−SP scores and feature contribution rates for the top 10 PCA features

2.3　模型训练与测试结果

选择SCRS数据的PC1、PC2、PC4作为检测模型输入特征，通过网格搜索和3折交叉验证优选主要超参数。基分类器模型：gbtree；基分类器个数：120；树深度：6。图7显示随着CART树个数（Num_round）增加，使用GridSearchCV优化后的超参数训练和测试收敛于第10棵CART树，对比默认超参数需要在第40棵CART树加入后收敛，优选后的模型收敛精度（Log_loss）和收敛速度都有明显提升。

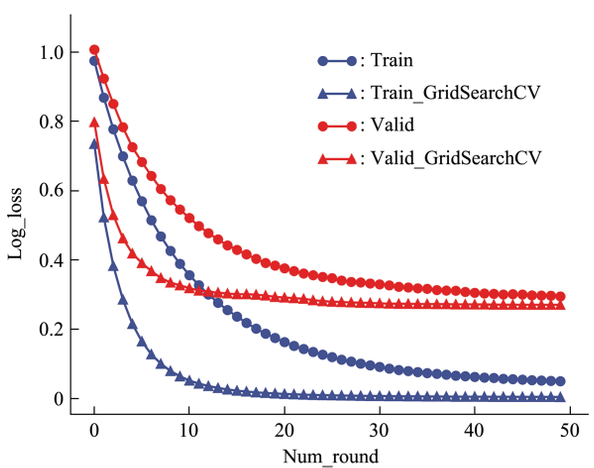


Fig. 7　Logarithmic loss functions for training and validing

图8a为实验组20%测试数据生长时期识别结果的混淆矩阵，3个生长时期平均测试准确率为96.0%，仅用PC1、PC2、PC4三个光谱特征较应用全谱的随机森林检测准确率提高5.3%[15]。进一步应用ROC曲线来评估模型敏感性和特异性（图8b），从标签维度和样本维度的ROC快速逼近左上

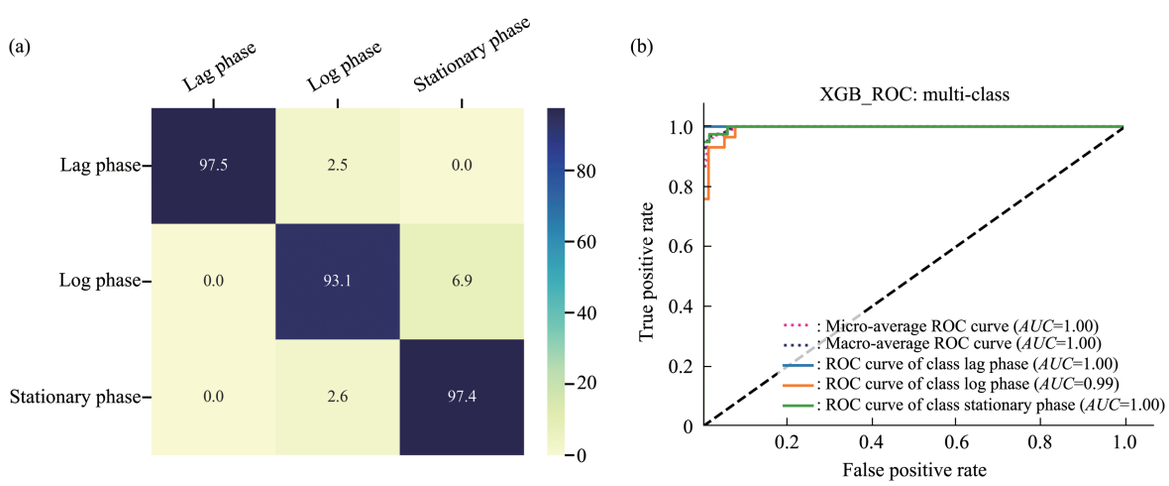


Fig. 8　XGBoost−based cell growth assay results and model evaluation

(a) Confusion matrix. (b) ROC curve.

角，计算的*AUC*值所表示的灵敏度和特异度之和接近1，模型有效且性能良好。

2.4　模型验证结果

为验证检测模型对另一批次枯草芽孢杆菌细胞生长检测的泛化能力，通过调用XGBoost训练模型检测验证组300个SCRS数据生长时期，首先使用**1.2**中数据预处理方法对验证组SCRS数据预处理。其次使用实验数据同样的PCA特征转换方法，选取验证组SCRS数据的PC1、PC2和PC4光谱特征。最后统计训练模型（XGBoostModel.pkl）对验证组细胞生长检测准确率为92.3%，模型具有良好的泛化能力。

3　结　　论

针对微生物生长应用于食品安全期和最佳食用期的精准检测问题，本文提出了基于拉曼技术的单细胞生长检测方法，主要包括：基于光谱PCA特征关系矩阵提出CP-SP特征评估方法，筛选出具有高内聚、低耦合的SCRS特征组合，输入XGBoost检测模型，并使用网格搜索和交叉验证优化检测模型，对同步培养的两批次微生物源保鲜剂滞后期、对数期和稳定期检测识别。实验表明，基于拉曼技术的单细胞生长检测方法以96.0%的细胞生长测试准确率，并以92.3%的细胞生长检测泛化能力，可以从单细胞水平上检测微生物生长，对保持食品良好的感官品质、延长食品货架期有重要作用。

本文提出的特征优选与检测模型方法具有突出的特征识别能力，有利于细胞生长标志物的发现和抑菌机理的研究，为食品安全和保鲜提供技术支撑。

参　考　文　献

[1] 郭娟, 张进, 王佳敏, 等. 天然抗菌剂在食品包装中的研究进展. 食品科学, 2021,**42**(9): 336-346
Guo J, Zhang J, Wang J M, *et al*. Food Sci, 2021, **42**(9): 336-346

[2] Mukherjee R, Verma T, Nandi D, *et al*. Understanding the effects of culture conditions in bacterial growth: a biochemical perspective using Raman microscopy. J Biophotonics, 2020, **13**(1): e201900233

[3] Oregel-Zamudio E, Angoa-Pérez M V, Oyoque-Salcedo G, *et al*. Effect of candelilla wax edible coatings combined with biocontrol bacteria on strawberry quality during the shelf-life. Sci Hortic, 2017, **214**(2017):273-279

[4] 朱亚珠, 夏率博, 陈琳, 等. 一株贝莱斯芽孢杆菌的生长特性及抑菌活性研究. 食品科学技术学报, 2022, **40**(1): 85-92
Zhu Y Z, Xia S B, Chen L, *et al*. J Food Sci Tech, 2022, **40**(1):85-92

[5] 方涵书, 郎明非, 孙晶. 细胞周期分析新方法. 分析化学, 2019, **47**(9): 1293-1301
Fang H S, Lang M F, Sun J. Chin J Anal Chem, 2019, **47**(9): 1293-1301

[6] Zavřel T, Faizi M, Loureiro C, *et al*. Quantitative insights into the cyanobacterial cell economy. Elife, 2019, **8**: e42508

[7] 何景, 程楠, 许文涛. 食品微生物新型快速筛查技术研究进展. 食品科学, 2015, **36**(13): 288-293
He J, Cheng N, Xu W T. Food Sci, 2015, **36**(13):288-293

[8] Lin W, Tian T, Jiang Y, *et al*. A CRISPR/Cas9 eraser strategy for contamination-free PCR end-point detection. Biotechnol Bioeng, 2021, **118**(5):2053-2066

[9] 赵莹彤, 浑婷婷, 詹悦维, 等. 基于微流控的真菌单细胞捕获和培养. 微生物学通报, 2019,**46**(3): 522-530
Zhao Y T, Hun T T, Zhan Y W, *et al*. Chin Microbiol, 2019, **46**(3): 522-530

[10] Xu Y Z, Metris A, Stasinopoulos D M, *et al*. Effect of heat shock and recovery temperature on variability of single cell lag time of *Cronobacter turicensis*. Food Microbiol, 2015, **45**: 195-204

[11] Li R, Dhankhar D, Chen J, *et al*. Identification of live and dead bacteria: a Raman spectroscopic study. IEEE Access, 2019, **7**: 23549-23559

[12] Lu W, Chen X, Wang L, *et al*. Combination of an artificial intelligence approach and laser tweezers Raman spectroscopy for microbial identification. Anal Chem, 2020, **92**(9): 6288-6296

[13] Wang D, He P, Wang Z, *et al*. Advances in single cell Raman spectroscopy technologies for biological and environmental applications. Curr Opin Biotechnol, 2020, **64**: 218-229

[14] Mukherjee R, Verma T, Nandi D, *et al*. Identification of a resonance Raman marker for cytochrome to monitor stress responses in *Escherichia coli*. Anal Bioanal Chem, 2020, **412**(22): 5379-5388

[15] Croxatto A, Marcelpoil R, Orny C, *et al*. Towards automated detection, semi-quantification and identification of microbial growth in clinical bacteriology: a proof of concept. Biomed J, 2017, **40**(6): 317-328

[16] Zhang S, Zheng Q, Xu B, *et al*. Identification of the fungal pathogens of postharvest disease on peach fruits and the control mechanisms of bacillus subtilis JK-14. Toxins, 2019, **11**(6): 322

[17] Lengyel A, Botta-Dukát Z. Silhouette width using generalized mean-a flexible method for assessing clustering efficiency. Ecol Evol, 2019, **9**(23):13231-13243

[18] Karim M R, Beyan O, Zappa A, *et al*. Deep learning-based clustering approaches for bioinformatics. Briefings Bioinf, 2021, **22**(1):393-415

[19] Deng X, Liu Q, Deng Y, *et al*. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inform Sci, 2016, **340**: 250-261

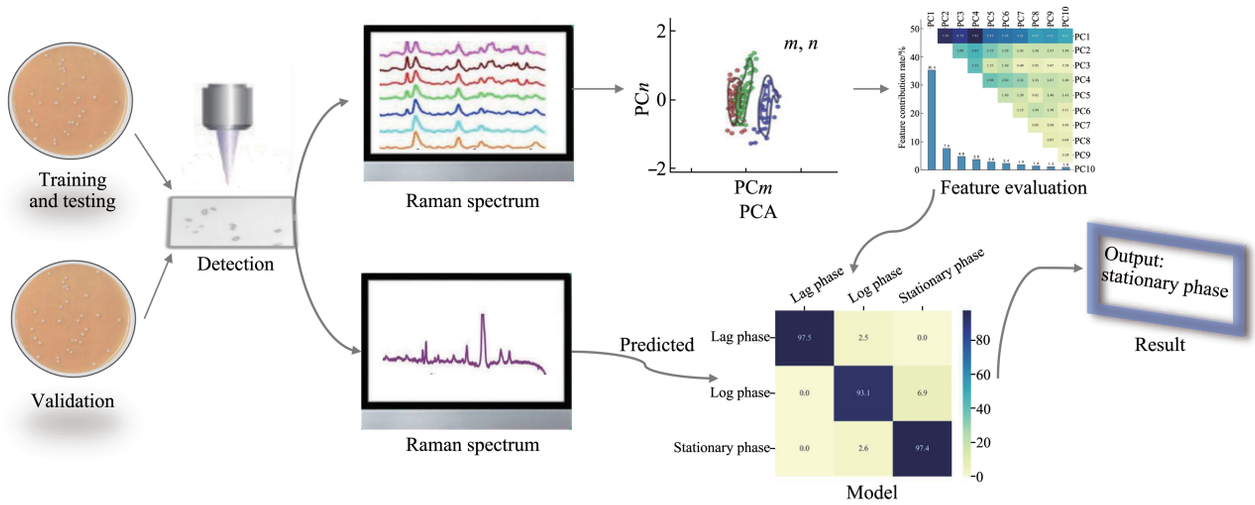
[20] Sachs M C. plotROC: a tool for plotting ROC curves. J Stat Softw, 2017, **79**: 2

Single-cell Growth Detection Based on Raman Technology*

LI Xin-Li, ZHANG Xin-Yu, YANG Qiang, LI Su-Yi**

(*College of Instrumentation and Electrical Engineering, Jilin University, Changchun* 130061, China)

Graphical abstract



Abstract　**Objective**　Single-cell growth detection can more scientifically reveal the rules of microbial metabolic changes and guide later microbial engineering applications. To study the accurate detection of microbial growth during the food safety period and optimal edible period, a single-cell growth detection method based on Raman technology is proposed in this paper. **Methods**　First, a total of 900 single-cell Raman spectroscopy (SCRS) data were collected from two batches of *Bacillus subtilis* through a simultaneous culture experiment, of which 600 were used for training and testing and the other 300 for model validation. Secondly, based on the feature relationship matrix of principal component analysis, CP-SP feature evaluation method was proposed to screen SCRS features for model detection. Then, a detection model based on XGBoost was built, and grid search and cross-validation were applied to optimize the detection model. Finally, confusion matrix and ROC curve were used to evaluate the detection accuracy, sensitivity and specificity of the model for cell lag phase, log phase and stationary phase. **Results**　The experiment found that the classification performance of the first, second, and fourth principal components screened by CP-SP was improved by 3.1% compared with the first three principal components of the feature contribution rate. The test accuracy of the optimized cell growth detection model was 96.0%, and the verification accuracy was 92.3%. **Conclusion**　The results show that the single-cell growth detection method based on Raman technology can accurately identify the single-cell growth state and has a high generalization ability, which can provide scientific guidance for the formulation of precise regulatory mechanisms for food safety and preservation.

Key words　single-cell Raman spectroscopy, cell growth, food safety and preservation, feature relationship matrix, XGBoost

DOI：10.16476/j.pibb.2022.0311

* This work was supported by grants from the National Key Research and Development Program (2022YFC2807904) and the Graduate Innovation Fund Project of Jilin University (2022059).

** Corresponding author.

Tel：86-13756887890，E-mail：lsy@jlu.edu.cn

Received：July 5，2022　Accepted：August 17，2022