

基于机器学习的植物乳杆菌LP1Z高密度发酵培养基优化

吴家琳^{1,2}, 李滢², 王涓¹, 陈玲², 黄士轩^{1,2}, 周润^{1,2}, 黄惠书^{1,2}, 张菊梅^{1,2}, 高鹤^{2,3}, 吴磊²,
赵辉², 陈谋通^{2*}, 吴清平^{2*}

(1. 华南农业大学食品学院, 广东广州 510642) (2. 广东省科学院微生物研究所, 华南应用微生物国家重点实验室, 广东省微生物安全与健康重点实验室, 国家卫健委微生物食品营养与安全科技创新平台, 广东广州 510070) (3. 广东科环生物科技有限公司, 广东广州 510663)

摘要: 目前常用的培养基优化方法是单因素分析结合响应面法的配方设计, 但该方法预测变量有限且较难反映菌株复杂的代谢网络, 导致配方优化效能有限。该研究以高效拮抗幽门螺杆菌的专利菌株 LP1Z 作为研究对象, 对菌株发酵培养基各组分在单因素优化基础上, 利用三种算法 (随机森林、深度神经网络和梯度提升决策树 LightGBM) 分别构建高密度发酵培养基成分优化预测模型, 对比不同算法所获得预测模型的准确性及其与响应面法的差异。结果显示, LP1Z-LightGBM 预测模型的 Pearson 相关系数、 R^2 最接近 1, 且均方根误差 (RMSE) 低至 0.02, 具有最优的预测效能。通过 LP1Z-LightGBM 预测获得的最优方案: 葡萄糖 30.00 g/L, 硫酸镁 0.30 g/L, 硫酸锰 0.03 g/L, 乙酸钠 7.00 g/L, 酵母粉 12.00 g/L, 柠檬酸三铵 2.00 g/L, 磷酸氢二钾 1.00 g/L, 吐温-80 2.00 mL/L, 使 LP1Z 的菌体发酵密度 (OD_{600}) 达 1.61, 明显优于响应面法所获得的菌体密度 ($P < 0.05$)。综上所述, 与传统的响应面法相比, LP1Z-LightGBM 能更好地完成菌株密度的预测。

关键词: 植物乳杆菌; 高密度培养; 培养基优化; 机器学习; 响应面法

文章编号: 1673-9078(2025)04-131-140

DOI: 10.13982/j.mfst.1673-9078.2025.4.0312

Machine Learning-Based Optimization of High-density Fermentation Medium for *Lactiplantibacillus plantarum* LP1Z

WU Jialin^{1,2}, LI Ying², WANG Juan¹, CHEN Ling², HUANG Shixuan^{1,2}, ZHOU Run^{1,2}, HUANG Huishu^{1,2},
ZHANG Jumei^{1,2}, GAO He^{2,3}, WU Lei², ZHAO Hui², CHEN Moutong^{2*}, WU Qingping^{2*}

(1. College of Food Science, South China Agricultural University, Guangzhou 510642, China) (2. Institute of Microbiology, Guangdong Academy of Sciences, State Key Laboratory of Applied Microbiology Southern China, Guangdong Provincial Key Laboratory of Microbial Safety and Health, National Health Commission Science

引文格式:

吴家琳, 李滢, 王涓, 等. 基于机器学习的植物乳杆菌LP1Z高密度发酵培养基优化[J]. 现代食品科技, 2025, 41(4): 131-140.

WU Jialin, LI Ying, WANG Juan, et al. Machine learning-based optimization of high-density fermentation medium for *Lactiplantibacillus plantarum* LP1Z [J]. Modern Food Science and Technology, 2025, 41(4): 131-140.

收稿日期: 2024-03-14

基金项目: 国家重点研发计划项目 (2022YFD2100703); 国家自然科学基金项目 (32222068; 32072326); 广州市科技计划项目 (2024A04J6592)

作者简介: 吴家琳 (1998-), 女, 硕士, 研究方向: 功能微生物开发利用, E-mail: wujl9821@163.com; 共同第一作者: 李滢 (1985-), 女, 博士, 研究方向: 健康功能微生物的挖掘与开发, E-mail: liying@gdim.cn

通讯作者: 吴清平 (1962-), 男, 博士, 研究员, 研究方向: 微生物安全与健康, E-mail: wuqp203@163.com; 共同通讯作者: 陈谋通 (1984-), 男, 博士, 研究员, 研究方向: 食源性致病菌危害形成与靶向控制, E-mail: cmtoon@163.com

and Technology Innovation Platform for Nutrition and Safety of Microbial Food, Guangzhou 510070, China)
(3.Guangdong KEHUAN Biotechnology Co. Ltd., Guangzhou 510663, China)

Abstract: The single-factor analysis combined with the response surface method (RSM) is usually applied for bacterial medium optimization. However, the variables predicted by this method are limited, and it is difficult to reflect the complex metabolic network of strains, limiting its ability to optimize the formulation. This study focuses on the patented strain *Lactiplantibacillus plantarum* LP1Z, known for its high antagonistic efficacy against *Helicobacter pylori*. On the basis of single-factor optimization, random forest, deep neural network, and gradient boosting decision tree (LightGBM) were used to construct the optimal prediction model of high-density fermentation medium components. The accuracy of this prediction model and its performance compared to RSM were evaluated. The results indicated that the Pearson correlation coefficient and R^2 of the LP1Z-LightGBM prediction model were close to 1, and the RMSE was as low as 0.02, with the best prediction efficiency. The optimal formula of high fermentation medium was obtained by LP1Z-LightGBM prediction, which includes glucose 30.00 g/L, MgSO_4 0.30 g/L, MnSO_4 0.03 g/L, CH_3COONa 7.00 g/L, yeast powder 12.00 g/L, $\text{C}_6\text{H}_5\text{O}_7(\text{NH}_4)_3$ 2.00 g/L, K_2HPO_4 1.00 g/L, and Tween-80 2.00 mL/L. The density of LP1Z cultured in the optimal fermentation media reached 1.61 (OD_{600}), which was significantly higher than that obtained by RSM ($P < 0.05$). In conclusion, compared with the traditional RSM, LP1Z-LightGBM offers superior predictive capability for the optimization of high-density fermentation medium.

Key words: *Lactiplantibacillus plantarum*; high-density culture; medium optimization; machine learning; response surface method

益生菌是一种能为宿主带来健康益处的活性微生物^[1], 笔者所在的研究团队已建成了相关的健康功能微生物菌种资源库(库容>2万株)及基因数据库并进行功能挖掘^[2-4]。益生菌的活菌数与其功效密切相关, 产品中活菌数至少在 $10^6 \sim 10^7$ CFU/mL 时才会对人体健康产生影响^[5]。高密度培养是菌粉生产中提高益生菌活菌数的重要手段, 但由于不同菌株间存在基因及代谢差异, 导致它们对营养物质、增殖环境的需求不同, 因此需要个性化的培养方案以实现菌株的高密度培养^[6]。其中, 培养基优化是实现高密度培养的重要手段之一^[7], 目前常用优化培养基的方法是对培养基成分进行单因素实验结合响应面法的配方设计。

响应面法因其模型简单、可以分析多变量间相互作用等优点被广泛用于培养基设计^[8]。然而, 响应面法不能同时预测多个变量的互作规律, 难以反映不同成分在菌株代谢过程中复杂交错的变化规律^[9]。这些局限限制了响应面法在微生物培养基设计中的应用, 因此, 需要引入能囊括更多变量及更好解析变量间相互关系的模型来提升益生菌发酵培养基的优化效率。

机器学习(Machine Learning, ML)是人工智能的一个分支, 由于 ML 可以逼近不同的复杂函数、处理许多传统方法无法预测的数据集, 因此它可能比响应面法更适用于益生菌培养基的优化^[10]。Arab

等^[11]使用人工神经网络和遗传算法相结合的 ML 优化 $G \times 15$ 砧木体外增殖矿物培养基, 发现与传统优化方案相比, ML 优化获得培养基的促增殖作用更加明显。王强等^[12]使用反向传播神经网络和遗传算法建立番茄红素发酵培养基的优化模型, 使番茄红素产量较传统方案提高了 31.6%。这些研究均证实 ML 能克服响应面法的局限, 可用于高效、精准地预测益生菌高密度培养基的最优成分配比。

算法是 ML 的核心, 对于不同的数据类型和数据特征, 不同算法的选择将影响 ML 的效率、准确性、可解释性及扩展性^[13]。ML 算法主要有朴素贝叶斯、随机森林(Random Forest, RF)、神经网络、决策树等^[14]。目前研究者们大多利用神经网络算法构建培养基成分的预测模型^[11,12,15], 而对于 RF 及决策树算法是否适用于构建培养基组分预测模型则未有定论。为探讨构建益生菌培养基成分预测模型的最优算法, 本研究采用高效拮抗幽门螺杆菌的专利菌株植物乳杆菌 LP1Z^[16,17]作为研究对象, 使用 RF、深度神经网络(Deep-Learning Neural Network, DNN)及梯度提升决策树 LightGBM 三种算法建立不同培养基成分质量浓度与菌体密度之间的函数关系, 对比基于不同 ML 算法所构建预测模型的效能。本研究有望为后续研究者基于 ML 构建培养基优化模型的算法选择提供理论依据。

1 材料与方法

1.1 材料、试剂与仪器

1.1.1 菌种

植物乳杆菌 (*Lactiplantibacillus plantarum*) LP1Z, 保藏编号为 GDMCC 62256, 于 2022 年 2 月 21 日保藏于广东省微生物菌种保藏中心 (GDMCC)^[16]。

1.1.2 基础培养基制备

在未特殊标注时, 本研究中植物乳杆菌 LP1Z 采用改良 MRS 培养基进行活化、增殖培养及活菌计数, 其中改良 MRS 液体培养基组分共计 8 种, 具体成分及使用量分别为: 葡萄糖 20.00 g, 酵母粉 18.00 g, 硫酸镁 0.20 g, 硫酸锰 0.005 g, 乙酸钠 5.00 g, 磷酸氢二钾 2.00 g, 柠檬酸三铵 2.00 g, 吐温-80 1.00 mL, 用蒸馏水定容到 1 000 mL, 调整 pH 值 5.7 ± 0.2 , 121 °C 灭菌 15 min。改良 MRS 固体培养基配方在改良 MRS 液体培养基配方上添加 15.00 g/L 的琼脂粉。

1.1.3 仪器与设备

SHP-250 型生化培养箱, 广东环凯微生物科技有限公司; 雷磁 PHS-3C 型 pH 计, 上海仪电科技; A90 高压灭菌锅, 德国爱安姆; JJ200Y 型电子天平, 常熟市双杰测试仪器厂; BSC 系列生物安全柜, 北京东联哈尔仪器制造有限公司; Epoch 2 全波长吸收光酶标仪, 美国安捷伦。

1.2 试验方法

1.2.1 菌种活化及种子液的制备

将甘油管保藏的植物乳杆菌 LP1Z 菌株, 按体积分数 2% 接种于 5 mL 改良 MRS 液体培养基中, 37 °C 培养 24 h 作为第一代菌种, 后进行传代培养 2 次作为活化好的菌种。将活化好的菌种转接至改良 MRS 液体培养基中, 置于 37 °C 培养箱中静置培养 12 h 作为种子液。

1.2.2 植物乳杆菌 LP1Z 发酵培养基的单因素优化

采用单因素优化法探究改良 MRS 培养基中 8 种成分 (葡萄糖、酵母粉、硫酸镁、硫酸锰、乙酸钠、磷酸氢二钾、柠檬酸三铵、吐温-80) 的质量浓度对植物乳杆菌 LP1Z 菌体密度和活菌数的影响。对上述任意一种成分开展单因素优化实验时, 剩余 7 种成分的质量浓度同 1.1.2。单因素实验中, 葡萄糖的质量浓度分别设置为 0.00、10.00、20.00、30.00、

40.00 g/L; 酵母粉的质量浓度分别设置为 0.00、6.00、12.00、18.00、24.00 g/L; 硫酸镁的质量浓度分别设置为 0.00、0.10、0.20、0.30、0.40 g/L; 硫酸锰的质量浓度分别设置为 0.00、0.025、0.05、0.075、0.10 g/L; 乙酸钠的质量浓度分别设置为 0.00、2.50、5.00、7.50、10.00 g/L; 磷酸氢二钾的质量浓度分别设置为 0.00、1.00、2.00、3.00、4.00 g/L; 柠檬酸三铵的质量浓度分别设置为 0.00、1.00、2.00、3.00、4.00 g/L; 吐温-80 的质量浓度分别设置为 0.00、0.50、1.00、1.50、2.00 mL/L。将 LP1Z 种子液按体积分数 2% 接入液体培养基中, 在 37 °C 条件下培养 24 h, 测定菌体密度 (OD_{600})。参照陈百莹等^[18]的方案采用菌落平板计数法测定活菌数。

1.2.3 基于 ML 预测模型的培养基优化设计

通过单因素实验获得影响 LP1Z 菌体密度及菌落计数的四个最主要影响因子, 分别为葡萄糖、硫酸镁、硫酸锰和乙酸钠, 参照王强等^[12]的方法设计 49 种配方 (表 1)。将 LP1Z 种子液按 1.2.2 方法接种于相应配方中, 测定发酵液的菌体密度。

1.2.4 基于响应面法的培养基优化设计

通过单因素实验获得影响 LP1Z 菌体密度及菌落计数的三个最主要的影响因子 (葡萄糖、硫酸锰和乙酸钠), 以其进行响应面实验设计, 应用 Design-Expert Version 13 进行 Box-Behnken 实验设计, 共获得 17 种配方 (表 3)。将 LP1Z 种子液按 1.2.2 方法接种于相应配方中, 测定发酵液的菌体密度。综合配方成分、质量浓度及相应菌体密度进行预测模型的构建。

1.3 数据统计与分析

1.3.1 ML 数据集构建

本实验以发酵培养基组成的四个主要影响成分 (葡萄糖、硫酸镁、硫酸锰和乙酸钠) 的质量浓度作为自变量, 以植物乳杆菌 LP1Z 的菌体密度作为因变量 (表 1)。测试集: 训练集=6:4。数据集划分通过随机选取其中 30 组作为训练集, 另外 19 组作为测试集。将所构建的数据集分别使用 RF、DNN 及 LightGBM 三种机器学习算法构建模型。

1.3.2 基于 ML 的培养基优化预测模型构建

LP1Z-RF 培养基优化预测模型由 python 语言 sklearn 集成的 RF 回归算法构建, 具体参数为: max_depth=10, min_samples_split=2, min_

samples_leaf=3。LP1Z-DNN 培养基优化预测模型由 python 语言 tensorflow 平台的 keras 构建, 其中 DNN 网络由一个输入层、两个隐藏层和一个输出层组成, 其中两个隐藏层分别由 8 个和 4 个神经元组成, 使用 PReLU 作为激活函数, 最后一个隐藏层使用 dropout (dropout=0.2) 方法忘记一些节点, 防止过拟合^[9]; 训练过程中具体参数为: batch_size=3, validation_split=0.2, epochs=1 000, 使用 Adam 作监督者, 学习率为 0.001。LP1Z-LightGBM 培养基优化预测模型由 python 语言 LightGBM 平台 gbd 算法构建, 具体参数为: boosting_type: gbd, feature_fraction=0.8, bagging_fraction=0.8, learning_rate=0.001, min_data_in_leaf=3, max_depth=6, num_leaves=48, reg_alpha=0.001, reg_lambda=0.03。

1.3.3 模型评价指标

ML 所构建的培养基优化预测模型通过 Pearson 相关性系数及决定系数 R^2 反映模型的预测性能: 当 Pearson 相关系数 >0.90 且 $R^2 >0.75$, 模型的预测效能被认为是可以接受的, 当 Pearson 相关系数绝对值及 R^2 越接近 1 时, 模型预测效能越好。研究采用均方根误差 (Root Mean Square Error, RMSE) 反映真实值和预测值的偏差, RMSE 越小模型的预测精度越高; 使用残差图反映模型预测值与真实值之间的误差大小, 使用 SHapley 加性解释值 (Shapley Additive exPlanations, SHAP) 的绝对值 |SHAP| 评价四个主要影响成分 (葡萄糖、乙酸钠、硫酸锰和硫酸镁) 对模型影响的大小。

1.3.4 统计分析

使用 SPSS 26 软件 (IBM, NY, USA) 通过方差分析 (ANOVA) 和 Tukey 检验对数据进行分析, $P < 0.05$ 被认为差异存在统计学意义。使用 GraphPad 9.3 (Insightful Science, CA, USA) 软件对数据作图, 所有实验均重复 3 次, 实验数据采用均值 $\pm$ 标准差表示。

2 结果与讨论

2.1 发酵培养基的单因素优化结果

单因素分析发现, 在改良 MRS 培养基的 8 种成分中, 葡萄糖、酵母粉、硫酸锰、硫酸镁及乙酸钠对菌株的发酵菌体密度有重要作用, 其中葡萄糖、硫酸锰、硫酸镁及乙酸钠在过低或者过高质量浓度

时, 均引起菌体密度的下降, 而酵母粉的过高质量浓度则不会引起菌体密度的下降 (图 1)。

葡萄糖是最常见的碳源类物质, 具有为乳酸菌生长提供能量、加快细胞合成和新陈代谢及促进乳酸菌增殖等作用^[20]。如图 1a 所示, 当葡萄糖质量浓度为 10.00 g/L 时, 菌体密度为 1.28, 活菌数为 1.50×10^8 CFU/mL; 当葡萄糖质量浓度增加至 30.00 g/L 时, 菌体密度增加至 1.46, 活菌数增加至 2.83×10^9 CFU/mL ($P < 0.05$ 和 $P < 0.01$), 说明葡萄糖对植物乳杆菌 LP1Z 的生长有重要影响。而继续增加至 40.00 g/L 时会导致菌体密度及活菌数下降 ($P < 0.01$ 和 $P > 0.05$)。过高质量浓度呈现抑制效果, 造成菌株生长迟缓的现象可能是 pH 快速下降导致不利于菌株生长的逆性环境^[21]。

酵母粉是常用的发酵培养基氮源, 其含有丰富的蛋白质、氨基酸、多肽和核苷酸等营养物质, 对发酵乳酸菌具有较好的效果^[22]。如图 1b 所示, 当酵母粉质量浓度为 0.00 g/L 时, 菌体密度为 0.15, 活菌数为 9.97×10^7 CFU/mL; 当酵母粉质量浓度增加至 12.00 g/L 时, 菌体密度增加至 1.51, 活菌数增加至 1.46×10^9 CFU/mL, 菌体密度及活菌计数均有明显提升 ($P < 0.01$ 和 $P < 0.01$)。而继续增加酵母粉质量浓度对菌体密度和活菌数无明显提升。秦雅莉等^[21]优化发酵乳杆菌 SS-31 培养基时, 使用酵母粉作为氮源增菌效果显著, 活菌数达到 5.87×10^9 CFU/mL, 与本研究结果相似。

研究表明, 微量元素如镁离子、锰离子等可以作为某些活性物质的组成成分或者酶的激活剂发挥作用, 从而促进乳酸菌生长^[23]。由图 1c 可知, 当硫酸镁质量浓度为 0.00 g/L 时, 菌体密度为 1.49, 活菌数为 1.79×10^9 CFU/mL; 当硫酸镁质量浓度增加至 0.30 g/L 时, 菌体密度为 1.48, 活菌数为 2.22×10^9 CFU/mL, 与 0.00 g/L 相比差异存在统计学意义 ($P < 0.05$ 和 $P \geq 0.05$)。而继续增加质量浓度至 0.40 g/L 时会导致菌体密度及活菌数下降 ($P < 0.01$ 和 $P \geq 0.05$)。由图 1d 可知, 当硫酸锰质量浓度为 0.00 g/L 时, 菌体密度为 1.18, 活菌数为 8.27×10^8 CFU/mL; 当硫酸锰质量浓度增加至 0.05 g/L, 菌体密度增加至 1.48, 活菌数增加至 1.72×10^9 CFU/mL, 与 0.00 g/L 相比差异存在统计学意义 ($P < 0.01$ 和 $P < 0.01$)。而继续增加质量浓度至 0.08 g/L 时会导致菌体密度及活菌数下降 ($P > 0.05$ 和 $P > 0.05$)。

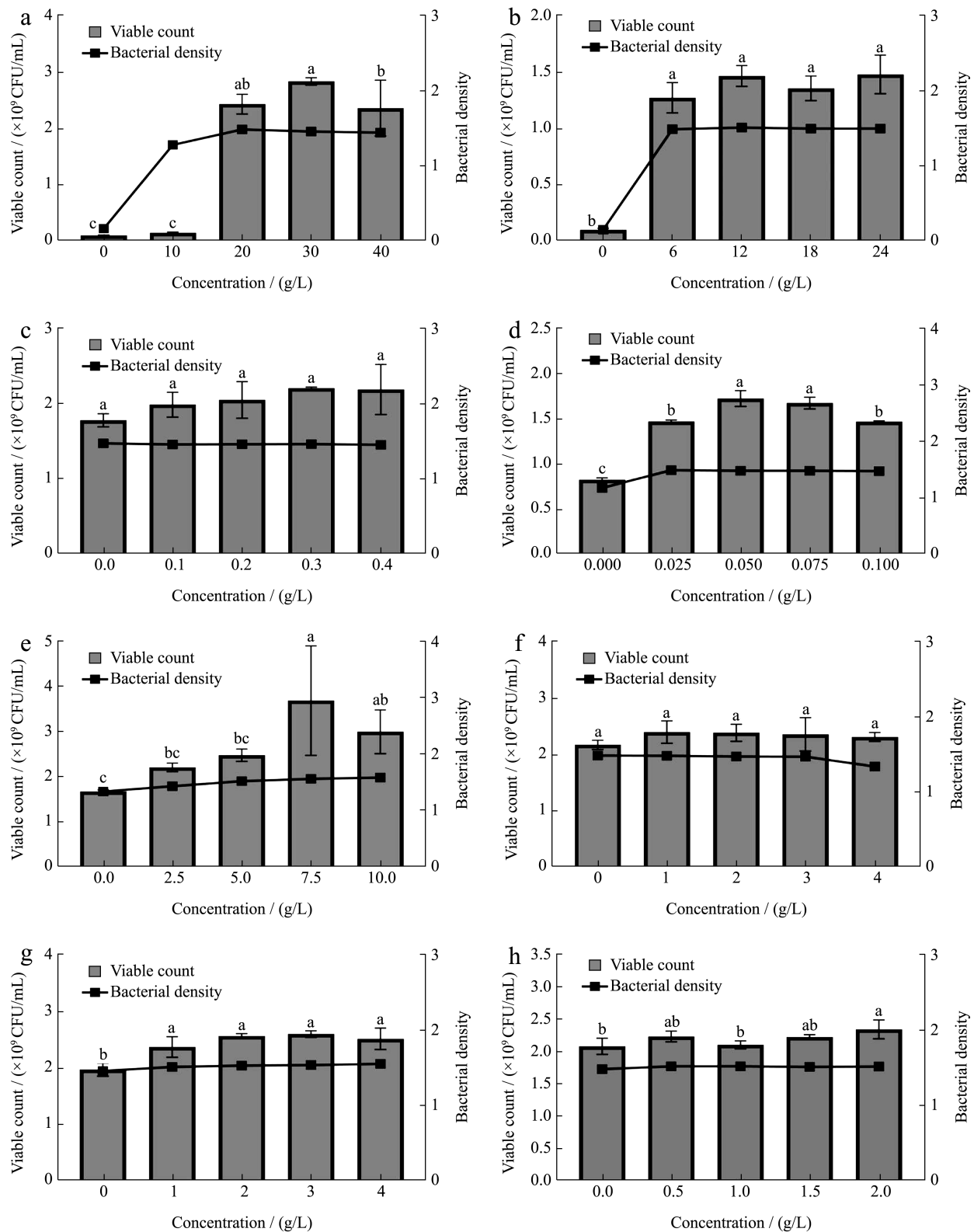


图1 培养基各组分浓度对植物乳杆菌 LP1Z 生长的影响

Fig.1 The effect of each component concentration on *Lactiplantibacillus plantarum* LP1Z growth

注: a 为 Glucose, b 为 Yeast exact, c 为 MgSO₄, d 为 MnSO₄, e 为 CH₃COONa, f 为 K₂HPO₄, g 为 C₆H₅O₇ (NH₄)₃, h 为 Tween 80。有相同字母表示不同质量浓度间差异不显著, 无相同字母表示不同质量浓度间差异显著 ($P < 0.05$)。

表 1 基于ML预测模型的培养基配方设计及对应培养菌体密度

Table 1 The design of medium formulation based on ML prediction model and corresponding bacterial density

样本	葡萄糖 /(g/L)	硫酸镁 /(g/L)	硫酸锰 /(g/L)	乙酸钠 /(g/L)	菌体密度 /(OD ₆₀₀)
1	10.00	0.05	0.01	1.00	1.33 ± 0.01
2	10.00	0.10	0.02	2.00	1.36 ± 0.02
3	10.00	0.15	0.03	3.00	1.38 ± 0.02
4	10.00	0.20	0.04	4.00	1.38 ± 0.02
5	10.00	0.25	0.05	5.00	1.37 ± 0.01
6	10.00	0.30	0.06	6.00	1.34 ± 0.02
7	10.00	0.35	0.07	7.00	1.34 ± 0.02
8	15.00	0.05	0.02	3.00	1.48 ± 0.02
9	15.00	0.10	0.03	4.00	1.48 ± 0.02
10	15.00	0.15	0.04	5.00	1.50 ± 0.02
11	15.00	0.20	0.05	6.00	1.50 ± 0.02
12	15.00	0.25	0.06	7.00	1.50 ± 0.02
13	15.00	0.30	0.07	1.00	1.42 ± 0.01
14	15.00	0.35	0.01	2.00	1.45 ± 0.01
15	20.00	0.05	0.03	5.00	1.59 ± 0.02
16	20.00	0.10	0.04	6.00	1.60 ± 0.02
17	20.00	0.15	0.05	7.00	1.60 ± 0.01
18	20.00	0.20	0.06	1.00	1.53 ± 0.02
19	20.00	0.25	0.07	2.00	1.55 ± 0.02
20	20.00	0.30	0.01	3.00	1.54 ± 0.02
21	20.00	0.35	0.02	4.00	1.56 ± 0.02
22	25.00	0.05	0.04	7.00	1.59 ± 0.01
23	25.00	0.10	0.05	1.00	1.46 ± 0.01
24	25.00	0.15	0.06	2.00	1.50 ± 0.01
25	25.00	0.20	0.07	3.00	1.53 ± 0.02
26	25.00	0.25	0.01	4.00	1.53 ± 0.02
27	25.00	0.30	0.02	5.00	1.56 ± 0.01
28	25.00	0.35	0.03	6.00	1.59 ± 0.01
29	30.00	0.05	0.05	2.00	1.49 ± 0.02
30	30.00	0.10	0.06	3.00	1.53 ± 0.02
31	30.00	0.15	0.07	4.00	1.54 ± 0.02
32	30.00	0.20	0.01	5.00	1.55 ± 0.02
33	30.00	0.25	0.02	6.00	1.58 ± 0.02
34	30.00	0.30	0.03	7.00	1.61 ± 0.02
35	30.00	0.35	0.04	1.00	1.46 ± 0.02
36	35.00	0.05	0.06	4.00	1.55 ± 0.01
37	35.00	0.10	0.07	5.00	1.57 ± 0.01
38	35.00	0.15	0.01	6.00	1.58 ± 0.02
39	35.00	0.20	0.02	7.00	1.60 ± 0.02
40	35.00	0.25	0.03	1.00	1.46 ± 0.01
41	35.00	0.30	0.04	2.00	1.49 ± 0.01
42	35.00	0.35	0.05	3.00	1.52 ± 0.01
43	40.00	0.05	0.07	6.00	1.56 ± 0.02
44	40.00	0.10	0.01	7.00	1.55 ± 0.03
45	40.00	0.15	0.02	1.00	1.41 ± 0.02
46	40.00	0.20	0.03	2.00	1.46 ± 0.01
47	40.00	0.25	0.04	3.00	1.49 ± 0.02
48	40.00	0.30	0.05	4.00	1.52 ± 0.02
49	40.00	0.35	0.06	5.00	1.54 ± 0.01

在缓冲体系（乙酸钠、磷酸氢二钾和柠檬酸三铵）的单因素优化实验中，当乙酸钠质量浓度为 2.50 g/L 时，菌体密度为 1.44，活菌数为 2.22×10^9 CFU/mL。当乙酸钠质量浓度增加至 7.50 g/L，菌体密度增加至 1.57，活菌数增加至 3.70×10^9 CFU/mL，与 2.50 g/L 相比差异存在统计学意义 ($P<0.01$ 和 $P>0.05$)；而继续增加质量浓度至 10.00 g/L 时会导致菌体密度及活菌数下降 ($P<0.01$ 和 $P>0.05$)（图 1e）。而磷酸氢二钾（图 1f）和柠檬酸三铵（图 1g）两种成分的单因素结果提示不同质量浓度对菌体密度的影响均不存在统计学意义。左梦楠等^[24]在优化发酵乳杆菌 BLHN3 的培养基时，选择柠檬酸铵 / 乙酸钠 / 磷酸氢二钾组合的体系作为最佳缓冲盐，菌株发酵密度显著提高，活菌数可达 6.05×10^9 CFU/mL。

吐温 -80 既是表面活性剂，也是菌体生长的有利因子，添加适量的吐温可以促进乳酸菌的增殖^[18]。如图 1h 所示，当吐温 -80 质量浓度为 0.50 mL/L 时，菌体密度为 1.52，活菌数为 2.24×10^9 CFU/mL；当吐温 -80 质量浓度增加至 1.00 mL/L 时，菌体密度为 1.52，活菌数为 2.11×10^9 CFU/mL，较 0.50 mL/L 未上升 ($P>0.05$ 和 $P>0.05$)；而进一步增加吐温 -80 质量浓度至 1.50 mL/L，菌体密度为 1.51，活菌数为 2.23×10^9 CFU/mL，与 0.50 mL/L 相比差异亦不存在统计学意义 ($P>0.05$ 和 $P>0.05$)。

2.2 三种基于ML建立的培养基优化预测模型性能评估

如图 1 所示，葡萄糖、硫酸锰、硫酸镁及乙酸钠的质量浓度对 LP1Z 的菌体密度具有明显影响，且质量浓度过高或过低均影响菌体密度，因此本研究针对这 4 种成分设计了 49 种配方，菌体密度测量结果见表 1。

研究通过 RF、DNN 和 LightGBM 三种算法分别构建了 LP1Z 培养基优化预测模型 LP1Z-RF、LP1Z-DNN 和 LP1Z-LightGBM，具体模型预测效能的评估指标如表 2 所示，三个模型的 Pearson 相关性系数均 >0.90 且 $R^2>0.75$ ，证实应用三种不同 ML 算法所构建的培养基优化预测模型均合理，三种模型均可用于模拟预测植物乳杆菌 LP1Z 培养基组成与菌体密度之间的关系。对比不同模型的预测效能发现，由 LP1Z-LightGBM 预测准确度最高，Pearson 系数为 0.96， R^2 为 0.92，而预测偏差最小，RMSE 为 0.02，表明基于 LP1Z-LightGBM 的预测效能最好。

表 2 不同ML模型评价指标

模型	算法	Pearson 相关系数	R^2	RMSE
LP1Z-RF	RF	0.94	0.88	0.03
LP1Z-DNN	DNN	0.96	0.89	0.03
LP1Z-LightGBM	LightGBM	0.96	0.92	0.02

图 2 展示了不同算法构建的模型预测值与实际值的比较，如图 2 所示，进一步分析发现，LP1Z-RF 模型中真实值与模型预测值残差最大，范围在 -0.04 至 0.04 之间；LP1Z-DNN 模型中真实值与模型预测值残差次之，范围在 -0.04 至 0.02 之间；LP1Z-LightGBM 模型中真实值与模型预测值残差最小，范围在 -0.02 至 0.02 之间。由此可知，利用 LP1Z-LightGBM 模型所得预测值与真实值之间的误差最小。

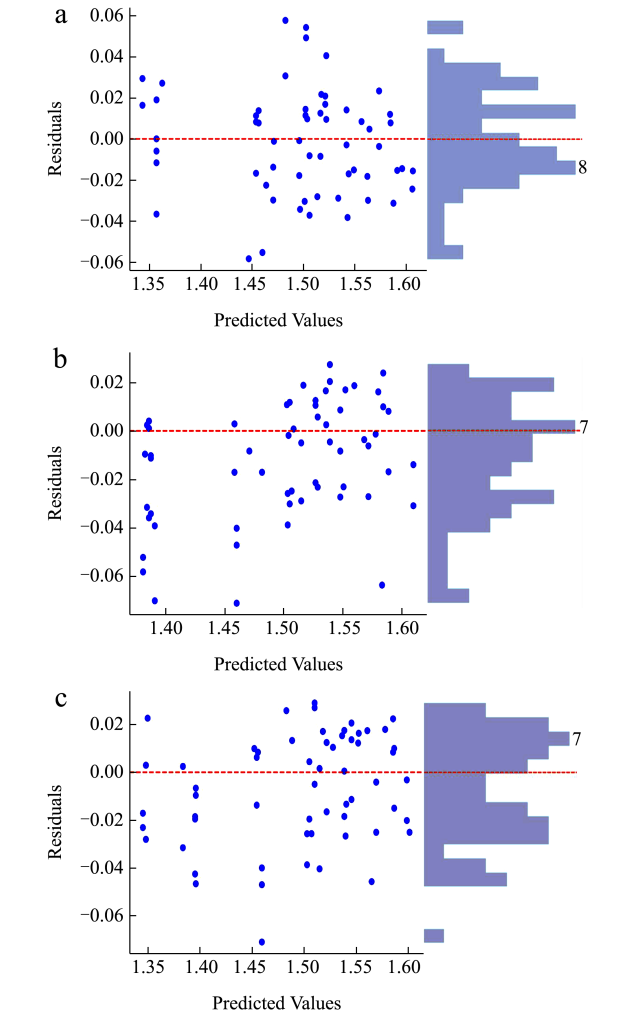


图 2 不同算法构建的模型预测值与实际值比较
Fig.2 Comparison between the predicted value and the actual value of the model constructed by different algorithms

注：a 为 LP1Z-RF，b 为 LP1Z-DNN，c 为 LP1Z-LightGBM。
柱状图为左图的统计结果，虚线表示残差为 0。

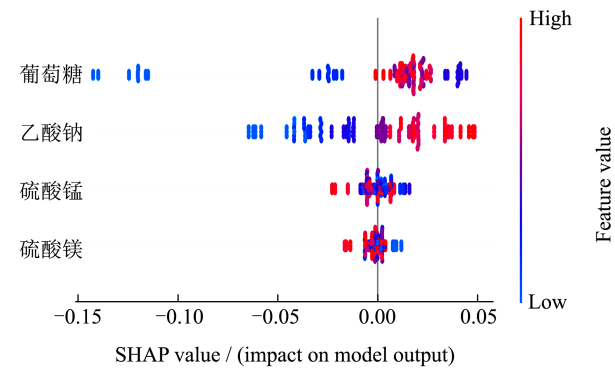


图 3 植物乳杆菌 LP1Z 培养基成分显著性分析

Fig.3 The significant analysis of *Lactiplantibacillus plantarum* LP1Z culture-medium components

表 3 Box-Behnken实验设计

Table 3 The design of Box-Behnken experiment

组别	葡萄糖 /(g/L)	硫酸锰 /(g/L)	乙酸钠 /(g/L)	菌体密度 /(OD ₆₀₀)
1	25.00	0.10	2.00	1.45 ± 0.03
2	25.00	0.06	6.00	1.54 ± 0.02
3	25.00	0.01	2.00	1.43 ± 0.03
4	40.00	0.10	6.00	1.52 ± 0.03
5	10.00	0.01	6.00	1.29 ± 0.03
6	10.00	0.10	6.00	1.30 ± 0.03
7	40.00	0.06	2.00	1.42 ± 0.02
8	25.00	0.10	10.00	1.57 ± 0.02
9	25.00	0.06	6.00	1.51 ± 0.02
10	25.00	0.06	6.00	1.52 ± 0.01
11	25.00	0.06	6.00	1.52 ± 0.00
12	25.00	0.06	6.00	1.52 ± 0.02
13	40.00	0.06	10.00	1.53 ± 0.01
14	40.00	0.01	6.00	1.45 ± 0.02
15	10.00	0.06	10.00	1.25 ± 0.01
16	10.00	0.06	2.00	1.26 ± 0.02
17	25.00	0.01	10.00	1.52 ± 0.01

2.3 LP1Z-LightGBM模型解析

根据 LP1Z-LightGBM 模型结果分析可知，葡萄糖对模型的平均 |SHAP| 为 0.04，乙酸钠对模型的平均 |SHAP| 为 0.03，硫酸锰对模型的平均 |SHAP| 为 0.01，硫酸镁对模型的平均 |SHAP| 为 0.004，因此可知在 LP1Z-LightGBM 模型中，葡萄糖的影响

最大，而硫酸镁的影响最小（图 3）。研究指出，在研究的质量浓度范围内，葡萄糖和乙酸钠对模型输出的影响基本呈正反馈，即随着两者质量浓度增加，LP1Z 的菌体密度呈上升趋势；而硫酸锰和硫酸镁对模型预测结果呈负反馈，两者质量浓度过高会导致 LP1Z 的菌体密度下降。

2.4 基于响应面法的培养基优化结果

如图 1 所示，葡萄糖、硫酸锰及乙酸钠的质量浓度对 LP1Z 的菌体密度具有明显影响，且质量浓度过高或过低均影响菌体密度，因此本研究对这 3 种成分进行 Box-Behnken 实验设计获得 17 种配方，菌体密度测量结果见表 3。

2.5 基于LP1Z-LightGBM及响应面法建立的培养基培养效能评价

采用 LP1Z-lightGBM 作为预测模型，通过深度限制为 6 的 Leaf-wise 生长策略，防止模型在训练过程中过度拟合，并进行多次迭代训练得到植物乳杆菌 LP1Z 培养基成分最优组合，各组分质量浓度如下：葡萄糖 30.00 g/L，硫酸镁 0.30 g/L，硫酸锰 0.03 g/L，乙酸钠 7.00 g/L，酵母粉 12.00 g/L，柠檬酸三铵 2.00 g/L，磷酸氢二钾 1.00 g/L，吐温 -80 2.00 mL/L，预测发酵所获得的 LP1Z 菌体密度为 1.62，而实际实验 LP1Z 的菌体密度为 1.61，与预测值的相对误差为 0.13% ($P>0.05$)。由响应面法优化的植物乳杆菌 LP1Z 培养基成分最优组合：葡萄糖 35.00 g/L，硫酸镁 0.30 g/L，硫酸锰 0.10 g/L，乙酸钠 8.00 g/L，酵母粉 12.00 g/L，柠檬酸三铵 2.00 g/L，磷酸氢二钾 1.00 g/L，吐温 -80 2.00 mL/L，预测发酵所获得的 LP1Z 菌体密度为 1.57，而实际实验 LP1Z 的菌体密度为 1.59，与预测值的相对误差为 1.19% ($P<0.05$)（表 4）。对比两种模型的配方，发现使用 LP1Z-LightGBM 所预测的配方能在培养中获得更高的菌体密度 ($P<0.05$)。

ML 是一种可以逼近不同复杂函数、还原数据内部复杂规律的方法。已有多项研究指出 ML 可以更好地还原培养基不同成分间复杂的互作关系，更准确地预测培养基的组分^[11,25]。如赖长龙等^[26]构建 ANN-GA 模型优化植物乳杆菌的发酵培养基，通过对模型预测的最优配方进行发酵，植物乳杆菌的总生物量显著增加至 10.86 g/L，模型预测误差 < 2%。此外，Pandey 等^[27]利用响应面结合 ANN-GA 的方法优化乳酸克鲁维酵母细胞产人干扰素培养基，构

建 ANN-GA 模型的 R^2 为 0.95，模型预测人干扰素最大产量 2.175 mg/L，实验验证结果为 2.015 mg/L，预测值与真实值相近。本研究结果表明利用培养基

优化预测系统 LP1Z-LightGBM 获得的培养基配方促增殖效果明显优于响应面法预测结果，且预测值与实验值的误差更小。

表 4 LP1Z-LightGBM与响应面法预测的最优配方及其实际菌体密度

Table 4 The optimal formulation of prediction by LP1Z-LightGBM and response surface method and actual bacterial density

优化模型	最优条件/(g/L)				菌体密度	
	葡萄糖	硫酸镁	硫酸锰	乙酸钠	预测值	实际值
LP1Z-LightGBM	30.00	0.30	0.03	7.00	1.62 ^A	1.61 ± 0.05 ^{Aa}
响应面法	35.00	0.30	0.10	8.00	1.57 ^B	1.59 ± 0.04 ^{Ab}

注：A 和 A 表示 LP1Z-LightGBM 模型的预测值和实际值对比差异不显著；A 和 B 表示响应面模型的预测值和实际值对比差异显著；a 和 b 表示两个模型实际值对比差异显著 ($P<0.05$)。

然而，本研究也发现并非所有基于 ML 的系统均能准确还原培养基成分对菌体密度的影响。算法是 ML 的核心，目前常用的算法包括随机森林、神经网络和决策树等。在既往研究中，研究者大多采用神经网络进行培养基优化预测模型的构建，而目前暂未有比对不同算法对培养基优化预测模型准确性的研究^[11, 12, 15]。本研究发现，利用梯度提升决策树 LightGBM 算法所构建的模型较目前常用的 DNN 网络在优化培养基成分上具有更好的预测效能。由于 LightGBM 算法采用了 Leaf-wise 方法、基于梯度的单侧采样和独占特征捆绑，比 DNN 更适用于数据集较小的模型^[28]。因此，在后续研究中，建议研究者根据菌株的特征及前期的成分设计，应用不同算法进行培养基预测模型的构建，以更准确还原培养基成分与菌体密度的关系。

3 结论

本研究以植物乳杆菌 LP1Z 为研究对象，建立了由不同 ML 算法构建的培养基优化模型，旨在突破益生菌的高密度培养。结果证实 ML 比响应面法更适用于培养基的优化，而 LightGBM 算法能更精准地还原培养基成分与菌体密度的互作规律，研究结果可作为后续培养基成分优化模型的参考，为益生菌高密度发酵培养基优化提供潜在的优化策略。

参考文献

[1] FERREIRA T H, MAXIMIANO P, URETA M, et al. Molecular Simulation: a remarkable tool to study mechanisms of cell membrane preservation in probiotic bacteria [J]. Current Opinion in Food Science, 2023, 50: 100985.

[2] YANG L, XIE X, LI Y, et al. Evaluation of the cholesterol-lowering mechanism of *Enterococcus faecium* strain 132 and *Lactobacillus paracasei* strain 201 in hypercholesterolemia rats [J]. Nutrients, 2021, 13(6): 1982.

[3] JIANG T, LI Y, LI L, et al. *Bifidobacterium longum* 070103 fermented milk improve glucose and lipid metabolism disorders by regulating gut microbiota in mice [J]. Nutrients, 2022, 14(19): 4050.

[4] LIANG T T, XIE X Q, ZHANG J M, et al. Bacterial community and composition of different traditional fermented dairy products in China, South Africa, and Sri Lanka by high-throughput sequencing of 16S rRNA genes [J]. LWT-Food Science and Technology, 2021, 144: 111209.

[5] JAWAN R, ABBASILIASI S, TAN J S, et al. Influence of type and concentration of lyoprotectants, storage temperature and storage duration on cell viability and antibacterial activity of freeze-dried lactic acid bacterium, *Lactococcus lactis* Gh1 [J]. Drying Technology, 2022, 40(9): 1774-1790.

[6] DE ASSIS D A, MACHADO C, MATTE C, et al. High cell density culture of dairy *Propionibacterium* sp. and *Acidipropionibacterium* sp.: a review for food industry applications [J]. Food and Bioprocess Technology, 2022, 15(4): 734-749.

[7] 李玉娥,马玲.乳酸菌高密度培养条件优化研究[J].中国酿造,2021,40(9):150-154.

[8] YOUSEFI M, KHORSHIDIAN N, HOSSEINI H. The ability of probiotic lactobacillus strains in removal of benzo[a]pyrene: a response surface methodology study [J]. Probiotics Antimicrob Proteins, 2022, 14(3): 464-475.

[9] SINGH V, HAQUE S, NIWAS R, et al. Strategies for fermentation medium optimization: an in-depth review [J]. Frontiers in Microbiology, 2016, 7: 2087.

[10] GREENER J G, KANDATHIL S M, MOFFAT L, et al.

- A guide to machine learning for biologists [J]. Nature Reviews Molecular Cell Biology, 2022, 23(1): 40-55.
- [11] ARAB M M, YADOLLAHI A, SHOJAEIYAN A, et al. Artificial neural network genetic algorithm as powerful tool to predict and optimize *in vitro* proliferation mineral medium for G x N15 rootstock [J]. Frontiers in Plant Science, 2016, 7: 1526.
- [12] 王强,冯玲然,余晓斌.基于BP神经网络和遗传算法优化番茄红素发酵培养基[J].食品与生物技术学报,2019, 38(2):111-119.
- [13] RASCHKA S, KAUFMAN B. Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition [J]. Methods, 2020, 180: 89-110.
- [14] SILVA G, FAGUNDES T P, TEIXEIRA B C, et al. Machine learning for hypertension prediction: a systematic review [J]. Current Hypertension Reports, 2022, 24(11): 523-533.
- [15] 郭善才,张瑱,林建原.基于BP神经网络结合响应面法优化香芹菜叶总黄酮提取工艺及其应用[J].中国食品学报, 2023,23(8):263-273.
- [16] 吴清平,商燕燕,李滢,等.一株产溶菌酶并高效拮抗多药耐药幽门螺杆菌的植物乳杆菌LP1Z及其应用: CN114350578B[P].2022-05-27.
- [17] 吴清平,商燕燕,李滢,等.一种源于植物乳杆菌的新型抗菌蛋白LP1Z_01335及其应用:CN117210434A[P].2023-12-12.
- [18] 陈百莹,郑苗,邓泽元,等.植物乳杆菌ZJ316培养基优化和高密度培养的研究[J].中国食品学报,2020,20(7):65-74.
- [19] BUYRUKOGLU S, YILMAZ Y, TOPALCENGİZ Z. Correlation value determined to increase *Salmonella* prediction success of deep neural network for agricultural waters [J]. Environmental Monitoring and Assessment, 2022, 194(5): 373.
- [20] 左梦楠,刘伟,全琦,等.乳酸菌高密度培养技术的研究进展[J].食品工业科技,2022,43(19):436-445.
- [21] 秦雅莉,于福田,赵笑颖,等.发酵乳杆菌SS-31培养基及发酵条件的优化[J].食品与生物技术学报,2022,41(12):48-57.
- [22] 吕佳璐,于冰,张妍,等.干酪乳杆菌YQ336高密度培养条件优化[J].饲料研究,2021,44(12):76-80.
- [23] 杨瑞冬,李伯海,王元弛,等.*Lactobacillus buchneri* IMAU80233高密度发酵工艺优化[J].食品科学,2019, 40(22):147-154.
- [24] 左梦楠,刘伟,张菊华,等.发酵乳杆菌BLHN3的高密度培养优化[J].食品与机械,2022,38(12):181-189.
- [25] KRIEGESKORTE N, GOLAN T. Neural network models and deep learning [J]. Current Biology, 2019, 29(7): R231-R236.
- [26] 赖长龙,曹余,杨玉,等.植物乳杆菌发酵动力学及高密度培养研究[J].食品与发酵工业,2022,48(20):137-144.
- [27] PANDEY R, PRABHU A A, DASU V V. Artificial neural network-genetic algorithm (ANN-GA) based medium optimization for the production of human interferon gamma (hIFN- γ) in *Kluyveromyces lactis* cell factory [J]. Canadian Journal of Chemical Engineering, 2019, 97(4): 843-858.
- [28] MIYATA R, MORIWAKI Y, TERADA T, et al. Prediction and analysis of antifreeze proteins [J]. Heliyon, 2021, 7(9): e07953.