

格氏乳杆菌 GU-G23 高密度培养因子 多维度优化及其验证

廖艳婷¹, 李锐定¹, 黄国宏², 梁晓琳², 周樊¹, 麻志泽¹, 时凤翠¹, 李全阳^{1*}

(1. 广西大学轻工与食品工程学院, 广西南宁 530004)

(2. 广西职业技术学院食品与生物技术学院, 广西南宁 530226)

摘要:为提高具有益生特性格氏乳杆菌 GU-G23 的生物量, 获取其高密度培养因子, 该研究首先通过单因素实验和 Plackett-Burman 实验筛选出该菌株主要的生长营养因子为鱼蛋白胨、胰蛋白胨、柠檬酸三铵。以响应面实验设计组合作为机器模型训练样本, 采用随机森林回归 (Random Forest Regression, RFR) 和径向基神经网络 (Radial Basis Function Neural Network, RBF) 模型对其培养基配方进行预测。以决定系数 (R-squared, R^2)、平均绝对误差 (Mean Absolute Deviation, MAE)、均方误差 (Mean-Square Error, MSE) 和平均绝对百分误差 (Mean Absolute Percentage Error, MAPE) 作为模型评价指标, 比较认为 RBF 在该研究中具有更好的预测性能。随后选择 RBF 神经网络和遗传算法 (Genetic Algorithm, GA) 结合对培养基主要成分进行了优化。最终获得该菌株培养基的优化配方为: 鱼蛋白胨 29.89 g/L, 胰蛋白胨 23.33 g/L, 柠檬酸三铵 4.34 g/L, 蔗糖 15.00 g/L, 低聚果糖 15.00 g/L, 乙酸钠 5.00 g/L 磷酸氢二钾 0.40 g/L, 七水硫酸镁 0.58 g/L, 一水硫酸锰 0.29 g/L, 吐温-80 1.00 g/L。在此培养基条件下, 所得样品活菌数达到 5.21×10^9 CFU/mL, 是未优化前的 4.57 倍。该研究对微生物高密度培养优化预测提供了新的方法。

关键词: 格氏乳杆菌; 培养基优化; 随机森林; 径向基神经网络; 遗传算法

文章编号: 1673-9078(2023)10-79-88

DOI: 10.13982/j.mfst.1673-9078.2023.10.1326

Multi-dimensional Optimization of High-density *Lactobacillus gasseri* GU-G23 Culture

LIAO Yanting¹, LI Ruiding¹, HUANG Guohong², LIANG Xiaolin², ZHOU Fan¹, MA Zhize¹, SHI Fengcui¹,
LI Quanyang^{1*}

(1.College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China)

(2.College of Food and Biotechnology, Guangxi Polytechnic, Nanning 530226, China)

Abstract: To increase the biomass of probiotic *Lactobacillus gasseri* GU-G23 and identify factors associated with a high-density culture, the main nutritional factors required using a single-factor experiment and Plackett-Burman design were established. The three key factors comprised fish peptone, tryptone, and triammonium citrate. The training sample for the machine learning model was generated according to the response surface experimental design. The random forest regression (RFR) and radial basis function neural network (RBF) models were used to predict their culture medium formulas. The coefficient of determination (R-squared, R^2), mean absolute deviation (MAE), mean-square error (MSE), and mean absolute percentage error (MAPE) were adopted as model evaluation metrics. Comparisons revealed that RBF exhibited superior predictive performance. Subsequently, a combination of RBF neural network and genetic algorithm (GA) was selected to optimize the main components of the culture medium. The optimized formula comprised 29.89 g/L fish peptone, 23.33 g/L tryptone, 4.34 g/L triammonium

引文格式:

廖艳婷,李锐定,黄国宏,等.格氏乳杆菌 GU-G23 高密度培养因子多维度优化及其验证[J].现代食品科技,2023,39(10):79-88

LIAO Yanting, LI Ruiding, HUANG Guohong, et al. Multi-dimensional optimization of high-density *Lactobacillus gasseri* GU-G23 culture [J]. Modern Food Science and Technology, 2023, 39(10): 79-88

收稿日期: 2022-10-18

基金项目: 国家自然科学基金面上项目 (31871802)

作者简介: 廖艳婷 (1996-), 女, 在读硕士研究生, 研究方向: 益生菌资源挖掘与应用, E-mail: Elinor829@163.com

通讯作者: 李全阳 (1964-), 男, 博士, 教授, 研究方向: 食品营养与抗衰老机制, E-mail: liquanyang@gxu.edu.cn

citrate, 15.00 g/L sucrose, 15.00 g/L xylo-oligosaccharides, 5.00 g/L sodium acetate, 0.40 g/L dipotassium hydrogen phosphate, 0.58 g/L magnesium sulfate heptahydrate, 0.29 g/L manganese sulfate monohydrate, and 1.00 g/L Tween-80. Under this medium composition, the number of viable bacterial cells reached 5.21×10^9 CFU/mL, which was 4.57 times higher than that before optimization. This study provides a new approach for the optimization of microbial high-density culture medium prediction.

Key words: *Lactobacillus gasseri*; culture medium optimization; random forest; radial basis function neural network; genetic algorithm

格氏乳杆菌 (*Lactobacillus gasseri*) 是革兰氏阳性菌, 兼性厌氧或微需氧, 是广泛存在于人类口腔、胃肠道、阴道和母乳中的内在菌。最早在 1980 年被 Lauer 等^[1]和 Kim 等^[2]发现, 目前在卫生部指定的《可用于食品的菌种名单》中亦包含有格氏乳杆菌, 并且被美国食品药品监督管理局定义为一般公认安全类添加剂状态微生物^[3], 因此该菌株具有较大的发展潜力。格氏乳杆菌 GU-G23 是本课题组从广西长寿核心区分离筛选的长寿老人源益生菌株^[4], 因此有必要对其高密度培养因子进行深入探索为发挥其更大的应用价值提供技术保障。

随着近两年新冠疫情的爆发, 追求能增强免疫、健康的营养食品成为人们迫切关注的问题^[5], 而格氏乳杆菌有多种健康益处, 包括维持肠道及阴道健康、产生抗菌细菌素、增强免疫调节、抗氧化、控制肥胖等功效^[6-11], 能够满足大众对健康的需求。据国际乳业联合会建议, 益生菌对宿主发挥作用的活菌数最低为 10^7 菌落形成单位 (Colony Forming Unit, CFU) 每克或每毫升^[12], 但益生菌自身特殊的生长条件, 对其进入市场的益生菌的成活率有一定的影响, 因此高密度培养是益生菌制品生产环节中的关键技术。在发酵培养基优化研究中, 单一优化方法是多数研究人员采取的常用方法, 其中响应面法 (Response Surface Methodology, RSM) 是较为快速经济的优化方法, 并得到了广泛的应用。随机森林 (Random Forest Regression, RFR) 是一种基于决策树模型, 具有训练速度快、泛化能力强等特性的集成算法^[13]。以往研究表明, 随机森林法具备可处理数据量大、降低计算量、提高数据预测精度等优势^[14,15], 在金融、医学、生物学、遗传学等方面已被广泛运用^[16-18]。径向基神经网络 (Radial Basis Function Neural Network, RBF) 属于局部逼近网络, 具有学习快速的优点, 此外其网络结构简单, 可以完美地逼近任何连续的非线性函数^[19]。并且一些研究中指出 RBF 神经网络模型相对于反向传递 (Back Propagation, BP) 神经网络模型具有较强的实用性和更高的准确性^[20]。目前, RFR 和 RBF 在微生物培养基优化方面的联合应用鲜有研究。微生物培养基是一个多组分的复杂体系, 常规方法建立的数

学模型往往难以准确描述这种非线性关系。随着人工智能算法的出现, 一些机器学习模型在高密度培养基优化中取得了一定的成绩^[21,22]。然而在最近的研究中表明, 采用单一模型泛化性能较差, 凭借一或两个维度的优化方法中往往忽视了特征选择和参数组合对模型的影响, 因此其预测结果具有较大的不确定性^[23,24]。

基于上述情况, 本文采用多维度预测方法, 对影响格氏乳杆菌生长的主要营养因素进行探究, 以期减少误差产生概率, 实现精准预测的目的。尽可能提高培养物的活菌数, 为实现新菌株的高密度培养提供技术支持。

1 材料与方法

1.1 材料与试剂

1.1.1 菌株

格氏乳杆菌 (*Lactobacillus gasseri*) GU-G23 是本团队前期筛分的一株百岁老人源乳杆菌, 经规范筛分检测认为该菌株具有多种益生特性, 由本实验室提供。

1.1.2 培养基

MRS 液体培养基: 葡萄糖 20.00 g, 蛋白胨 10.00 g, 牛肉膏 10.00 g, 酵母浸粉 5.00 g, Tween-80 1.00 mL, 磷酸氢二钾 0.40 g, 七水硫酸镁 0.58 g, 一水硫酸锰 0.29 g, 柠檬酸三铵 2.00 g, 三水合乙酸钠 5.00 g, 加蒸馏水至 1 000 mL, 调节 pH 值为 6.2~6.4。121 °C 高压蒸汽灭菌 20 min。

MRS 固体培养基: 在 MRS 液体培养基的基础上添加 1.5%~2.0% (*m/m*) 琼脂粉, 用于菌株活化及平板计数。

增殖培养基: 按照实验设计配方配制, 121 °C 灭菌 20 min。

1.1.3 试剂

葡萄糖、蔗糖、乳糖、麦芽糖、低聚果糖、低聚木糖和海藻糖, 天津市科密欧化学试剂有限公司; 酸水解酪蛋白、胰蛋白胨、大豆蛋白胨、鱼蛋白胨、酵母浸粉和牛肉膏, 北京奥博星生物技术有限公司; 三水合乙酸钠、柠檬酸三铵、磷酸氢二钾和吐温-80, 天津市致远化学试剂有限公司。

1.2 仪器与设备

DHP-9272 电热恒温培养箱, 上海齐欣科学仪器有限公司; ZWYR-240 恒温培养振荡器, 上海仪智械分析仪器制造有限公司; Scan-1200 全自动微生物计数仪, 法国英特塞恩斯(上海)有限公司; RTS-1 生物生长检测反应器, 达启(中国)有限公司; 全自动立式高压灭菌器, 致微(厦门)仪器有限公司。

1.3 方法

1.3.1 菌株 GU-G23 种子液的活化

格氏乳杆菌 GU-G23 菌种保藏于体积分数 40% 的甘油冻存管中, 实验前将菌种以 2% (V/V) 接种量接种于 MRS 液体培养基中传代培养两次, 再于固体培养基中划线, 生长 24 h 后挑取单菌落至 MRS 液体培养基中, 37 °C 恒温培养箱中培养 15 h 后用于实验。

1.3.2 格氏乳杆菌 GU-G23 活菌数的检测

菌株活菌数的测定方法按照国标 GB 4789.35-2016 《食品安全国家标准食品微生物学检验乳酸菌检验》执行, 并于 37 °C 培养 48 h 后采用全自动微生物计数仪进行计数, 结果以 CFU/mL 表示。

1.3.3 格氏乳杆菌 GU-G23 生长曲线测定

预先准备好干净无菌的 50 mL 离心管, 倒入 10 mL 培养液, 挑取纯化后的格氏乳杆菌 GU-G23 单菌落至培养液中, 反应器温度设定为 37 °C, 并上机连续测试 24 h, 以 MRS 培养基作为优化对照, 绘制出菌株优化前后的生长曲线。

1.3.4 培养基主要成分筛选

1.3.4.1 单因素实验

以 MRS 培养基为基础培养基, 固定除碳源以外的因素, 选择葡萄糖 (Glucose, Glu)、蔗糖 (Sucrose, Suc)、乳糖 (Lactose, Lac)、麦芽糖 (Maltose, Mal)、低聚果糖 (Fructo-oligosaccharide, Fos)、低聚木糖 (Xylo-oligosaccharides, Xos)、海藻糖 (Trehalose, Tre) 7 种碳源分别以 2% 的添加量加入培养基中, 培养后测定活菌数。分析不同碳源对菌株生长的影响, 并根据单因素实验结果进一步对碳源种类进行复配、添加比例优化以及总添加量的确定。同理, 确定氮源: 酸水解酪蛋白 (Acid Hydrolyzed Casein, AHC)、胰蛋白胨 (Tryptone, TP)、大豆蛋白胨 (Soybean Peptone, SP)、鱼蛋白胨 (Fish Peptone, FP)、酵母浸粉 (Yeast Extract, YE)、牛肉膏 (Beef Paste, BP) 种类、复配种类、复配添加比例及总添加量。

1.3.4.2 Plackett-Burman 实验设计

为了获得能够显著影响格氏乳杆菌 GU-G23 生长

的因素。将确定的培养基成分: 蔗糖 (A)、低聚果糖 (B)、鱼蛋白胨 (C)、胰蛋白胨 (D)、三水合乙酸钠 (E)、柠檬酸三铵 (F) 和磷酸氢二钾 (G) 等 7 个因素作为考察对象。采用 Plackett-Burman 实验设计, 以格氏乳杆菌的活菌数为响应值, 筛选出影响因素, 实验设计见表 1。

表 1 Plackett-Burman 实验设计筛选各因素水平

Table 1 Screening factors and levels of plackett-burman design

变量值编码	变量/(g/L)	水平		
		-1	0	+1
A	蔗糖	10	15	20
B	低聚果糖	10	15	20
C	鱼蛋白胨	15	20	25
D	胰蛋白胨	15	20	25
E	三水合乙酸钠	2	5	8
F	柠檬酸三铵	1	2	3
G	磷酸氢二钾	0.2	0.4	0.6

1.3.4.3 最陡爬坡实验

根据 Plackett-Burman 实验设计得到的回归系数, 来决定显著因素的最陡爬坡实验的方向和梯度, 从而确定实验因素的中心点^[25], 以进行响应面实验设计。

1.3.5 模型训练样本构建

以响应面实验优化设计 (Box-Behnken Design, BBD) 数据作为训练样本, 将单因素实验、Plackett-Burman 实验和最陡爬坡实验确定的鱼蛋白胨、胰蛋白胨和柠檬酸三铵进行优化实验, 以格氏乳杆菌 GU-G23 的活菌数为响应值, 构建训练样本, 在 1.3.4.3 实验结果上设计实验因素及水平, 如表 2 所示。

表 2 Box-Behnken Design 因素水平及编码

Table 2 Factor levels and coding of Box-Behnken design

因素/(g/L)	水平		
	+1	0	-1
A 鱼蛋白胨	30	25	20
B 胰蛋白胨	30	25	20
C 柠檬酸三铵	5	3	1

1.3.6 RFR 随机森林模型

在培养基优化的模型中, 预测值为连续变量, 为求解上述三个显著影响因素和目标函数之间的关系, 本文采用 RFR 随机森林回归算法研究其函数关系, 具体数学模型如式 (1):

$$\begin{cases} \text{RFR} = (\text{ntree}, \text{mtry}) \\ d = F(x_1, x_2, x_3, \text{RFR}) \end{cases} \quad (1)$$

式中:

ntree——模型中回归树的棵树;

mtry——随机特征数目;
d——活菌数的预测;
F——随机森林回归算法表达的目标与变量之间的不确定函数;
 x_i ——自变量。

如图 1 所示,基于 R 语言平台运用 RandomForest 程序包建立随机森林回归的格氏乳杆菌高密度培养基优化模型,通过随机将 BBD 响应面设计的实验组合随机按照 7:3 的比例分为训练集和测试集。设置 mtry=10, ntree=180, 构建模型。

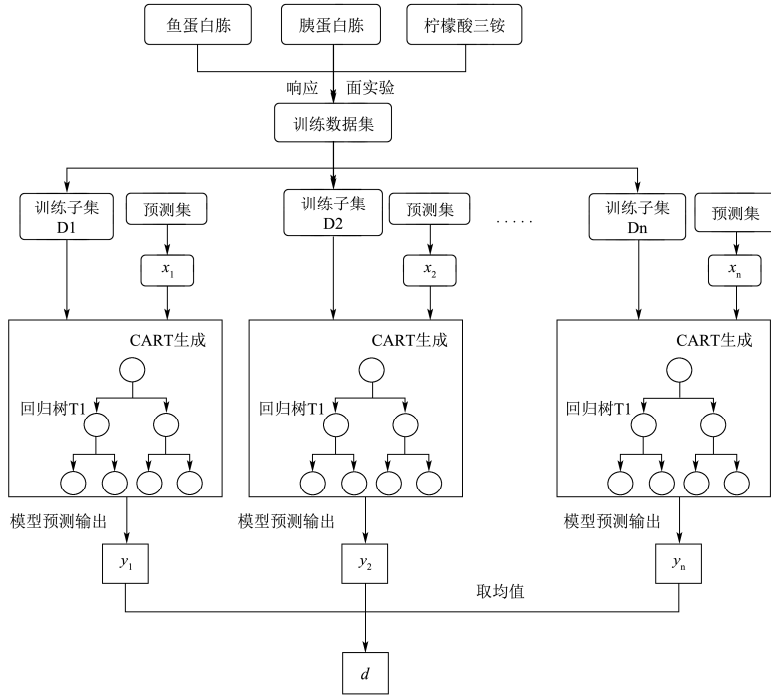


图 1 随机森林建模过程

Fig.1 Random forest modeling process

1.3.7 RBF 神经网络建模

径向基函数神经网络(RBF 网络)是基于函数逼近理论而构造的一类前馈型网络,这类网络的学习等价于在多维空间中寻找训练数据的最佳拟合平面,模型结构如图 2 所示,它包含 3 层前馈网络,即输入层、隐层和输出层^[26,27]。RBF 网络以径向基函数作为隐单元的前提,构成隐含层空间。隐含层对输入向量进行变换,将降维空间的模式变换到高维空间内,使得降维空间内的线性不可分问题在高维空间内线性可分^[28]。根据本文单因素检测结果确定响应面设计的 3 个因素鱼蛋白胨、胰蛋白胨、柠檬酸三铵添加量作为 RBF 神经网络的输入层节点,菌株的活菌数为输出值,隐含层神经元数量可根据经验公式(2)求出,具体模型结构如图 2 所示。随后选择 BBD 响应面实验组合中 70%实验结果作为模型的训练集,剩余 30%作为测试集,以完成 RBF 神经网络的自我学习训练和预测。

$$N_h = \frac{N_s}{\alpha(N_i + N_0)}$$

(2)

式中:
 N_s ——训练集的样本数;
 α ——通常取值范围为[2, 10];

N_i ——输入层神经元个数;
 N_0 ——输出层的神经元个数。

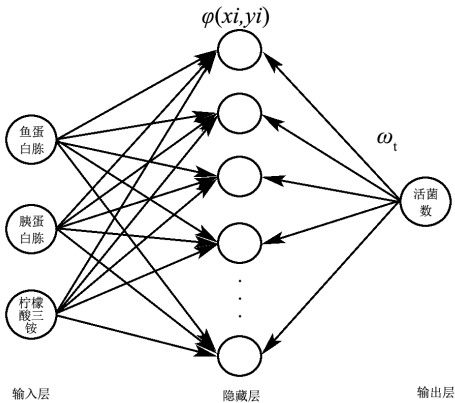


图 2 RBF 神经网络结构图

Fig.2 Structure of the RBF neural network

1.3.8 模型预测结果比较

为评估模型准确性及可靠性,选择解释方差 R^2 、平均绝对误差(MAE)、均方误差(MSE)、平均绝对百分误差(MAPE)作为评价指标,比较 RFR 及 RBF 模型的预测能力。其中 R^2 代表模型的可解释性,取值范围为[0, 1]; MAE、MSE、MAPE 反映样本数据的离散程度,值越低说明精度越高。

1.3.9 GA 遗传算法寻优

基于 1.3.6.1 和 1.3.6.2 所构建的 RFR 随机森林模型和 RBF 神经网络进行格氏乳杆菌 GU-G23 高密度培养增殖培养基配方优化, 将拟合度最佳的模型作为 GA 遗传算法的适应度函数, 结合构建混合模型对菌株的培养基组成寻求最优解^[29]。设置最大遗传代数为 500, 变量的二进制位数 8, 变异概率为 0.01, 交叉概率为 0.8, 代沟 0.25, 运行 Matlap R 2019 程序, 得到每代种群平均适应度变化结果, 以此寻找函数的全局最优值。

1.4 数据处理与分析

每组实验均进行 3 次平行测定, Plackett-Burman 和 Box-Behnken 响应面设计采用 Design-Expert 12.0 软件进行设计与分析; 模型预测及其优化采用 Matlap R 2019 和 R 软件 (版本 4.0.1) 执行, 采用 SPSS Statistics 26.0 数据软件和 Graph Pad Prism 8.0 进行数据分析和作图。

2 结果与分析

2.1 培养基主要成分筛选

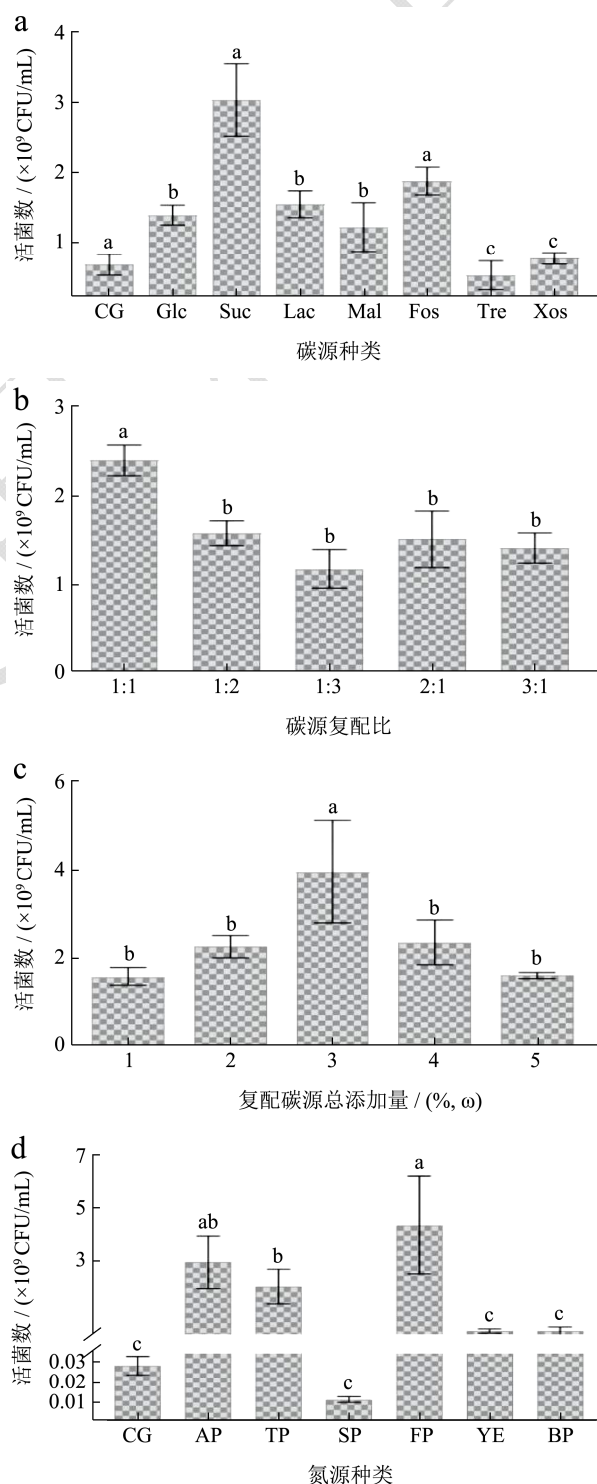
2.1.1 碳源优化

不同种类的碳源对格氏乳杆菌 GU-G23 活菌数的影响如图 3 所示。其中图 3a 说明蔗糖、低聚果糖对菌株的促生长作用最为显著, 活菌数最高达到 2.38×10^9 CFU/mL, 葡萄糖、乳糖和麦芽糖次之, 为筛选出最优碳源, 选择蔗糖和低聚果糖进行碳源复配实验, 结果如图 3b 所示, 当蔗糖与低聚果糖复配比为 1:1 时菌株的活菌数达到最大值, 为 2.40×10^9 CFU/mL, 并相对其他复配比例具有显著性差异 ($P < 0.05$); 随着其余 4 组碳源复合比例的提高, 活菌数呈现下降的趋势。在确定最优的碳源种类及复配比例后又对碳源的最适添加量进行了探究, 结果如图 3c 所示, 随着碳源添加量的递增, 活菌数呈现先增后降的趋势, 当碳源添加量为 3% 时, 菌株的活菌数达到最大值, 为 3.32×10^9 CFU/mL。因此本实验将蔗糖、低聚果糖作为碳源, 复配比例及添加量分别为 1:1、3%。

2.1.2 氮源优化

氮源是影响微生物生长的另一关键营养因素, 结果如图 3d 表明, 所选取的 6 种氮源中鱼蛋白胨对菌株的促生长作用影响最为显著, 活菌数最高达到 4.38×10^9 CFU/mL, 考虑到使用单一氮源成本较高, 因此选择以鱼蛋白胨为主要氮源, 与其它 5 种氮源进行复配, 结果如图 3e 所示, 从中得到对菌株生长有显著影响的组合为鱼蛋白胨和胰蛋白胨 ($P < 0.05$)。除

了氮源种类外。不同的氮源浓度对菌株的生长、产物和副产物代谢有着不同影响, 需进一步探究两者复配比例以及最佳添加浓度, 实验结果如图 3f、3g。可以看到, 复合氮源为鱼蛋白胨和胰蛋白胨质量比 1:1 和 2:1 时菌株活菌数较高, 经方差分析结果表明两者无显著差异, 综合考虑选择经济成本较低的组合, 即质量比为 1:1 的鱼蛋白胨和胰蛋白胨作为复合氮源, 且复合氮源添加浓度为 4%, 菌株活菌数达到最优值, 为 4.54×10^9 CFU/mL。



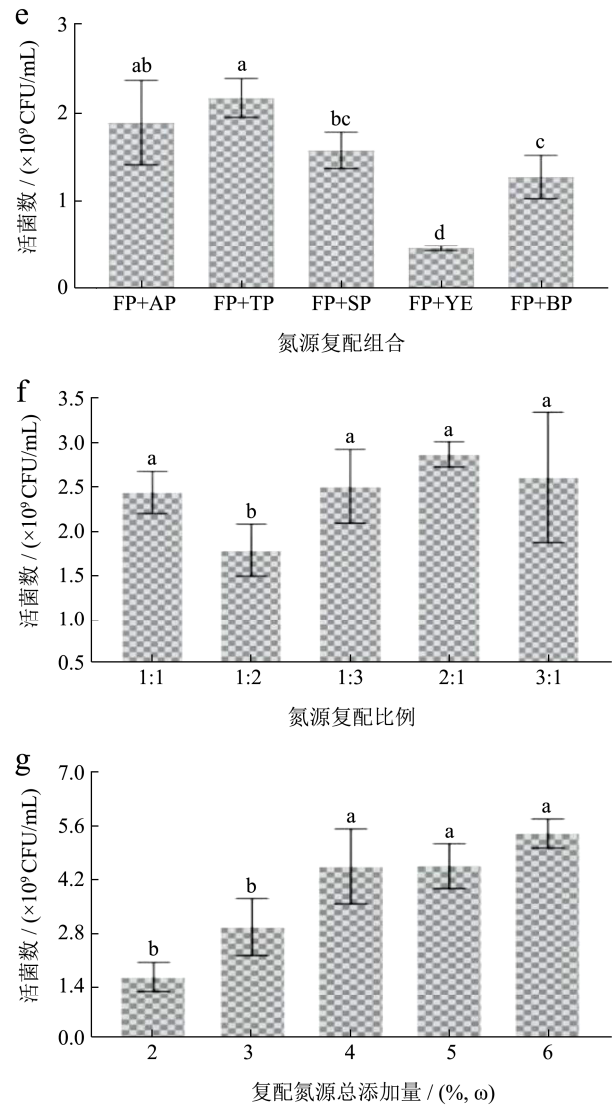


图3 不同碳源和氮源对格氏乳杆菌 GU-G23 菌株生长量的影响

Fig.3 Effect of different carbon and nitrogen sources on the growth of *Lactobacillus gasseri* GU-G23 strain

注：图中小写字母为各组间对照，(P<0.05)。a 为碳源种类优化；b 为碳源复配比例优化；c 为复合碳源添加量优化；d 为氮源种类优化；e 为复合氮源种类优化；f 为氮源复配比例优化；g 为复合氮源添加量优化。

2.1.3 Plackett-Burman 实验设计

以上述碳氮源优化结果选定的培养基主要成分以及 MRS 缓冲盐共 7 个因素进行 Plackett-Burman 实验设计，实验结果见表 3。

由表 3 得知，不同实验设计水平下格氏乳杆菌 GU-G23 的活菌数有一定差异。利用 Design Expert 软件对表 3 中的实验结果进行方差分析，结果如表 4 所示。

由表 4 可知，在促进菌株生长上，鱼蛋白胨、胰蛋白胨、柠檬酸三铵的贡献度最高，分别为 6.28%、34.68%、11.14%。因此选择鱼蛋白胨、胰蛋白胨、柠檬酸三铵为主要研究对象进行后续实验。

表 3 Plackett-Burman 实验设计因素对格氏乳杆菌 GU-G23 菌株活菌数的影响

Table 3 Effect of Plackett-Burman experimental design factors on the active count of <i>Lactobacillus gasseri</i> GU-G23 strain								
序号	变量编码值							活菌数/ (×10 ⁹ CFU/mL)
	A	B	C	D	E	F	G	
1	-1	-1	-1	1	-1	1	1	4.66±0.11 ^f
2	-1	1	1	-1	1	1	1	2.15±0.07 ^g
3	-1	1	1	1	-1	-1	-1	4.67±0.08 ^f
4	-1	-1	1	-1	1	1	-1	2.53±0.12 ^e
5	1	-1	1	1	-1	1	1	1.72±0.43 ^c
6	1	-1	1	1	1	-1	-1	5.44±0.11 ^a
7	-1	1	-1	1	1	-1	1	3.16±0.06 ^c
8	1	-1	-1	-1	1	-1	1	2.19±0.07 ^e
9	1	1	-1	1	1	1	-1	2.69±0.06 ^b
10	-1	-1	-1	-1	-1	-1	-1	1.71±0.40 ^f
11	1	1	-1	-1	-1	1	-1	1.28±0.13 ^f
12	1	1	1	-1	-1	-1	1	3.15±0.49 ^d

注：表中小写字母为各组间对照，P<0.05。表 5、6 同。

表 4 格氏乳杆菌 GU-G23 菌株 Plackett-Burman 实验设计分析表

Table 4 Plackett-Burman experimental design analysis table for <i>Lactobacillus gasseri</i> GU-G23 strain			
Term	Stdized Effect	Sum of Squares	Contribution/%
蔗糖	-4.01	48.36	2.31
低聚果糖	-1.91	11.00	0.53
鱼蛋白胨	6.61	131.25	6.28
胰蛋白胨	15.5	725.21	34.68
三水合乙酸钠	1.61	7.82	0.37
柠檬酸三铵	-8.81	233.17	11.14
磷酸氢二钾	-2.15	13.88	0.66

2.1.4 最陡爬坡实验

根据上述实验结果，保持鱼蛋白胨、胰蛋白胨、柠檬酸三铵以外的水平不变，对这三种成分进行最陡爬坡实验，实验设计和结果见表 5。

表 5 格氏乳杆菌 GU-G23 菌株最陡爬坡实验设计及结果

Table 5 Experimental design and results of the steepest ascent of *Lactobacillus gasseri* GU-G23 strain

序号	鱼蛋白胨 (g/L)	胰蛋白胨 (g/L)	柠檬酸三铵 (g/L)	活菌数 (×10 ⁹ CFU/mL)
1	20	20	1	0.83±0.11 ^d
2	25	25	2	2.02±0.86 ^b
3	30	30	3	4.68±0.52 ^a
4	35	35	4	1.56±0.38 ^c
5	40	40	5	1.12±0.10 ^d
6	45	45	6	1.06±0.05 ^d

由表 5 可知,采用最陡爬坡实验可以快速逼近最佳区域,由三者的稳定效应系数可知,其中鱼蛋白胨、胰蛋白胨呈正相关,与柠檬酸三铵呈负相关,因此随着三者添加量的增加,活菌数呈先增后减的变化趋势。在实验条件 3 (鱼蛋白胨 30 g/L,胰蛋白胨 30 g/L,柠檬酸三铵 3 g/L) 下,活菌数达最大值 4.68×10^9 CFU/mL,故以此种组合作为响应面设计的中心点。

2.2 构建模型训练样本

依据实验 2.1.3 和 2.1.4 确定的显著影响因素和实验因素中心点,利用 Design Expert 软件进行三因素三水平的响应面设计,设计及结果见表 6。

表 6 格氏乳杆菌 GU-G23 菌株模型训练样本实验设计表

Table 6 Model training sample design table of *Lactobacillus gasseri* GU-G23 strain

序号	A 鱼蛋白胨	B 胰蛋白胨	C 柠檬酸三铵	活菌数 ($\times 10^9$ CFU/mL)
1	-1	0	-1	3.46 ± 0.40^{ef}
2	0	0	0	5.89 ± 0.31^{ab}
3	0	0	0	6.62 ± 0.87^a
4	0	+1	+1	4.64 ± 0.51^{cd}
5	0	-1	+1	4.28 ± 0.87^{cdef}
6	-1	+1	0	3.25 ± 0.78^{ef}
7	0	0	0	6.56 ± 0.76^a
8	-1	0	+1	3.46 ± 0.48^{ef}
9	+1	0	-1	4.52 ± 0.81^{cde}
10	0	0	0	5.88 ± 0.79^{ab}
11	+1	0	+1	4.16 ± 0.51^{cdef}
12	0	-1	-1	4.04 ± 0.24^{cdef}
13	+1	+1	0	4.80 ± 0.87^{bc}
14	0	0	0	6.10 ± 0.52^a
15	-1	-1	0	3.19 ± 0.61^{ef}
16	0	+1	-1	4.49 ± 0.33^{cde}
17	+1	-1	0	3.04 ± 0.50^f

2.3 RFR 及 RBF 模型的拟合结果

图 4 反应了 RFR 及 RBF 模型对训练样本的训练情况。实际值和预测值重合度越高说明训练效果越好。在 RFR 模型中 (图 4a、4b), R^2 为 0.068, 说明该模型拟合度不佳,其预测优势不明显;但在 RBF 模型中 (图 4c、4d), R^2 为 0.98, 说明该模型的拟合度较高,整体预测趋势良好,故以此进行菌株活菌数的后续模拟。

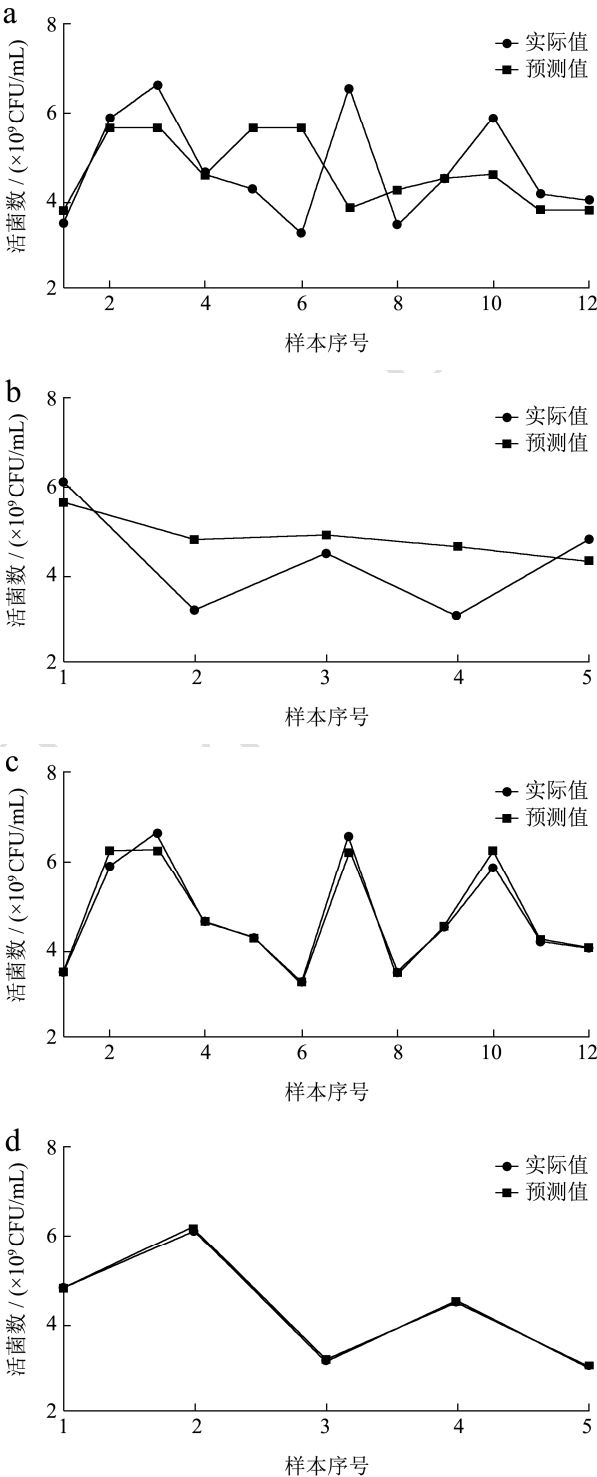


图 4 格氏乳杆菌 GU-G23 菌株培养 RFR 和 RBF 模型训练及测试集与实验结果的拟合效果

Fig.4 Fitting effect of training and testing set of RFR and RBF models with experimental results for *Lactobacillus gasseri* GU-G23 strain culture

注: a、b 为 RFR 模型训练和测试集实测值和预测值对比; c、d 为 RBF 模型训练和测试集实测值和预测值对比。

2.4 RFR 及 RBF 模型的数学评估

对模型进行误差估计是建模过程的重要步骤，能够有效判断所建模型的准确性和可靠性，根据 RFR 及 RBF 模型分别预测训练集和测试集上各组的活菌数，通过计算均方误差(MSE)、平均绝对误差(MAE)、平均绝对百分误差 (MAPE) 来评估所建立的 RFR 和 RBF 模型的准确性，结果如表 7 所示；图 5 是 RFR 和 RBF 模型的回归结果对比，图 5a 说明 RFR 模型实测值与预测值的离散程度较高，回归精度不佳；在模型评估指标中，表现为数值较大，图 5b 则说明了 RBF 模型回归精度最优，结合模型评估指标说明该模型的预测可行性。

表 7 格氏乳杆菌 GU-G23 菌株培养基筛选模型评估指标对比表

Table 7 Comparison table of evaluation indicators of <i>Lactobacillus gasseri</i> GU-G23 strain media screening model				
机器学习模型	数据属性	MSE	MAE	MAPE
RFR	测试集	0.075 3	0.863 2	0.188 9
	预测集	0.144 2	1.005 3	0.296 0
RBF	测试集	0.042 2	0.118 3	0.019 0
	预测集	0.002 1	0.020 8	0.003 4

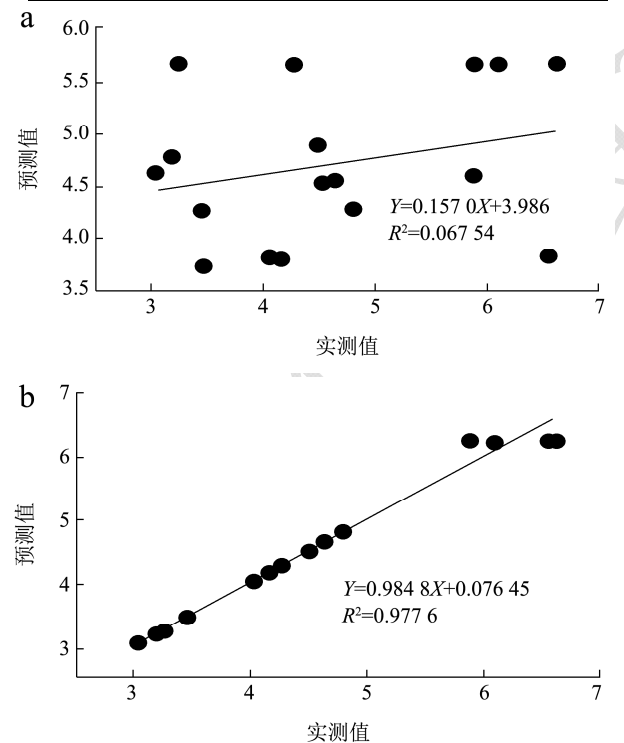


图 5 格氏乳杆菌 GU-G23 菌株培养基筛选的 RFR 和 RBF 模型预测结果散点图

Fig.5 Scatter plot of RFR and RBF model prediction results for *Lactobacillusgasseri* GU-G23 strain media screening

注：a: RFR 模型；b: RBF 模型。

2.5 GA 遗传算法寻优

在上述分析的基础上，利用 RBF 神经网络调节函数的适应度，并以格氏乳杆菌 GU-G23 的活菌数作为输出值，将 RBF 神经网络与遗传算法进行结合对培养基配方进行了全局寻优，依据实验方法 1.3.8 中的参数设置进行寻优，结果如图 6 所示，图中最有价值为活菌数($\times 10^9$ CFU/mL)的倒数，经过 500 次迭代，确定了格氏乳杆菌活菌数的最佳输入条件，分别是鱼蛋白胨 29.89 g/L、胰蛋白胨 23.33 g/L、柠檬酸三铵 4.34 g/L，输出值为 5.30×10^9 CFU/mL。

在高密度培养技术的优化应用中，响应面法、人工神经网络耦合遗传算法等应用广泛，如潘海博等^[30]采用 RBF 神经网络结合遗传算法对三种关键营养成分进行优化，显著提高了罗伊氏乳杆菌 LTR1318 的活菌数；张丹丹等^[31]利用前反馈神经网络结合遗传算法对酿酒酵母 (*Accharomyces cerevisiae*) L9 的富硒能力进行优化，在优化营养条件培养下酵母富硒量提高了 57%。魏泉增等^[32]利用反向传播神经网络结合遗传算法优化链霉菌发酵条件，结果得到代谢产物葡萄糖异构酶的产量比初始发酵产量提高了 41.1%。邹立飞等^[33]通过结合 BP 神经网络和响应面法优化薏仁米酒产氨基酸态氮最佳的发酵条件，比较得到更好的优化效果。以上皆通过一至两个维度对微生物发酵培养基进行优化。本研究在优化方法的设计上，联合 RFR-RBF-GA 模型进行多维度优化菌株培养基组成，结合不同模型的优化结果，在避免响应面优化等方法的片面性和偶然性方面做了一些探索，这或许能够提高仿真精度，精准表达出菌株活菌数和培养基主要成分之间的内在联系，目前采用多维度优化方法在微生物应用领域还比较少见。

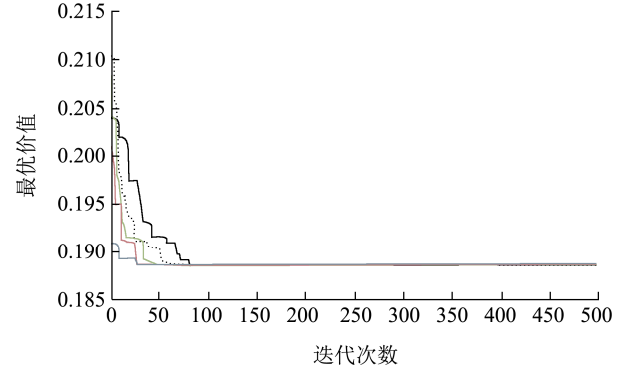


图 6 格氏乳杆菌 GU-G23 培养基筛选的遗传算法迭代寻优结果

Fig.6 Iterative genetic algorithm for *Lactobacillus gasseri* GU-G23 medium screening with superiority-seeking results

2.6 预测模型验证实验

在 GA 寻优条件下的进行活菌数测定, 得到实验值为 5.21×10^9 CFU/mL, 与预测值拟合率为 97.7%。说明实验值与预测值吻合度较高, 可见 RBF-GA 模型有良好的预测和优化性能。同时通过对比优化前后格氏乳杆菌 GU-G23 的生长曲线变化, 亦能反应出优化的效果。如图 7 所示, 菌株在 MRS 培养基中生长对数虽然较长, 为 5~18 h, 但菌株生长速率较低, 生物可及性不高; 而经过培养基优化后, 生长对数期为 5~15 h, 生长速率较大, 具有明显的高活性和高密度特征。本文优化探索出一套格氏乳杆菌 GU-G23 高密度培养优化方法, 经过验证实验检测发现, 该菌株在优化培养基中培养活菌数达到了 5.21×10^9 CFU/mL, 检索发现有关格氏乳杆菌培养物含量的国外报道中, Baltova 等^[34]为了提高格氏乳杆菌 4/13 的免疫调节功能活性, 发现利用含有乳糖、酵母提取物和酪蛋白胨的脱脂奶粉培养基培养, 可将菌株活菌数提高至 3.15×10^9 CFU/mL; Avonts 等^[35]研究 5 种益生菌株在 MRS 和牛奶培养基中的细胞生长和细菌素产生情况, 发现格氏乳杆菌 K7 在仅添加酵母提取物的乳培养基中生长, 活菌数达到 1×10^9 CFU/mL, Terraf 等^[36]评估在不同培养条件下乳酸杆菌形成生物膜能力差异时发现格氏乳杆菌 CRL 1263 添加吐温-80 对菌株的生长至关重要, 最高活菌数达到 1×10^8 CFU/mL。

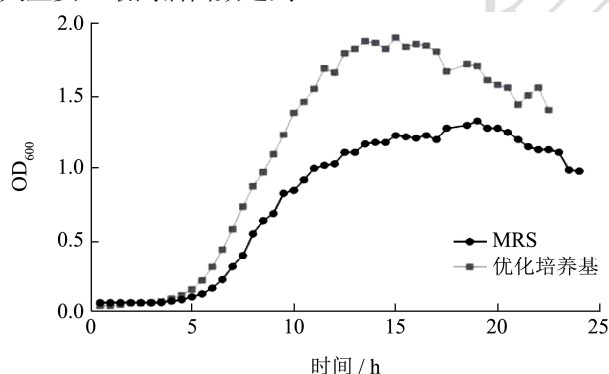


图 7 格氏乳杆菌 GU-G23 培养基优化前后的生长曲线

Fig.7 Growth curve of *Lactobacillus gasseri* GU-G23 before and after media optimization

在国内, 关于格氏乳杆菌高密度培养研究的报道有, 郭艳荣^[37]利用响应面法对格氏乳杆菌 IMAU FB017 的高密度培养物中, 菌体的含量是 3.15×10^9 CFU/mL, 陈齐等^[38]发明了一种格氏乳杆菌 LG-G12 高密度发酵及提高菌粉质量的方法, 其优化培养基得到的菌株活菌数为 3.93×10^9 CFU/mL。而本研究培养物含菌量则为 5.21×10^9 CFU/mL, 因此认为与国内外有关报道相比, 格氏乳杆菌培养物中活菌数为最大值, 这说明本

研究开发的格式乳杆菌培养方法具有一定的先进性。

3 结论

本研究对格氏乳杆菌 GU-G23 菌株高密度培养方法进行了探究, 采用单因素实验和 Plackett-Burman 实验确定显著影响格氏乳杆菌 GU-G23 生长的培养基主要成分为蔗糖、低聚果糖、鱼蛋白胨、胰蛋白胨、柠檬酸三铵。通过比较 RFR 随机森林和 RBF 径向基神经网络模型的预测结果, 确定了运用 RBF-GA 模型对培养基主要成分进行寻优, 最终得到格氏乳杆菌 GU-G23 菌株的最佳培养基为: 鱼蛋白胨的含量 29.89 g/L, 胰蛋白胨 23.33 g/L, 柠檬酸三铵 4.34 g/L, 蔗糖 15.00 g/L, 低聚果糖 15.00 g/L, 三水合乙酸钠 5.00 g/L, 磷酸氢二钾 0.40 g/L, 七水硫酸镁 0.58 g/L, 一水硫酸锰 0.29 g/L, 吐温-80 1.00 g/L, 此条件下该菌株培养物的活菌数达到 5.21×10^9 CFU/mL, 与国内外对比, 该结果是迄今为止有关报道的最高值。

参考文献

- [1] Lauer E, Kandler O. *Lactobacillus gasseri* sp. nov, a new species of the subgenus *Thermobacterium* [J]. Zentralblatt für Bakteriologie: I. Abt. Originale C: Allgemeine, Angewandte und Ökologische Mikrobiologie, 1980, 1(1): 75-78.
- [2] Kim J W, Rajagopal S N. Antibacterial activities of *Lactobacillus crispatus* ATCC 33820 and *Lactobacillus gasseri* ATCC 33323[J]. The Journal of Microbiology, 2001, 39(2): 146-148.
- [3] Stoeker L, Nordone S, Gunderson S, et al. Assessment of *Lactobacillus gasseri* as a candidate oral vaccine vector [J]. Clin Vaccine Immunol, 2011, 18(11): 1834-1844.
- [4] 喻婷, 黄国宏, 梁爱红, 等. 巴马长寿老人源乳酸杆菌的筛选及其益生特性[J]. 现代食品科技, 2017, 33(1): 39-47.
- [5] 吴益芳, 蔡玉荣, 陈金萍. 浅谈新冠病毒疫情对全球功能性营养健康食品的需求和行业发展[J]. 食品安全导刊, 2020, 17: 67.
- [6] Kim J Y, Park Y J, Lee H J, et al. Effect of *Lactobacillus gasseri* BNR17 on irritable bowel syndrome: a randomized, double-blind, placebo-controlled, dose-finding trial [J]. Food Science and Biotechnology, 2018, 27(3): 853-857.
- [7] Phukan N, Brooks A E S, Simoes-Barbosa A. A cell surface aggregation-promoting factor from *Lactobacillus gasseri* contributes to inhibition of trichomonas vaginalis adhesion to human vaginal ectocervical cells [J]. Infection and Immunity, 2018, 86(8): e907-e917.
- [8] Majhenić A Č, Venema K, Allison G E, et al. DNA analysis of the genes encoding acidocin LF221 A and acidocin LF221

- B, two bacteriocins produced by *Lactobacillus gasseri* LF221 [J]. Applied Microbiology and Biotechnology, 2004, 63(6): 705-714.
- [9] Kobatake E, Nakagawa H, Seki T, et al. Protective effects and functional mechanisms of *Lactobacillus gasseri* SBT2055 against oxidative stress [J]. Plos One, 2017, 12(5): e177106.
- [10] Sungur T, Aslim B, Karaaslan C, et al. Impact of exopolysaccharides (EPSs) of *Lactobacillus gasseri* strains isolated from human vagina on cervical tumor cells (HeLa) [J]. Anaerobe, 2017, 47: 137-144.
- [11] 黄元铭,肖玉春,宋利琼,等.加氏乳杆菌 Y20 对高胆固醇大鼠模型的降胆固醇效果及对肠道菌群和肝脏代谢的影响[J].卫生研究,2020,49(4):574-579.
- [12] Qin X, Gao Q, Luo Z. Enhancing the storage and gastrointestinal passage viability of probiotic powder (*Lactobacillus plantarum*) through encapsulation with pickering high internal phase emulsions stabilized with WPI-EGCG covalent conjugate nanoparticles [J]. Food Hydrocolloids, 2021, 116: 106658.
- [13] 李强,衣杨,吴忠道,等.基于机器学习的肠道菌群数据建模与分析研究综述[J].微生物学通报,2021,48(1):180-196.
- [14] 郑利林,徐金英,王晓龙.随机森林算法在湿地研究中的应用[J].湿地科学,2019,17(1):16-24.
- [15] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
- [16] Jurado S, Peralta J, Nebot A, et al. Short-term electric load forecasting using computational intelligence methods [C]// 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2013
- [17] Wang D, Sun Z. Big data analysis and parallel load forecasting of electric power user side [J]. Proceedings of the CSEE, 2015, 35(3): 527-537.
- [18] Wu X, He J, Yip T, et al. A two-stage random forest method for short-term load forecasting [C]// Power & Energy Society General Meeting, 2015
- [19] 王华伟,周鑫,王博,等.基于 RBF 神经网络和 MIGA 的液压锥阀降噪研究[J].机电工程,2022,39(11):1527-1534.
- [20] 李亚蒙,丁军航,孙宝楠,等.BP 和 RBF 神经网络应用于海表温盐短期预测效果对比[J].海洋科学进展,2022,40(2):220-232.
- [21] 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.
- [22] 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.
- [23] 方毅,陈煜之.机器学习与线性模型在资产收益率预测中的比较[J].统计与决策,2022,16:180-183.
- [24] 杜新忠,李叙勇,王慧亮,等.基于贝叶斯模型平均的径流模拟及不确定性分析[J].水文,2014,34(3):6-10.
- [25] 刘栋.罗伊氏乳杆菌 LT018 高密度培养及其冻干技术的研究[D].南宁:广西大学,2017.
- [26] 张玉瑞,陈剑波.基于 RBF 神经网络的时间序列预测[J].计算机工程与应用,2005,11:74-76.
- [27] 严薇荣,施倡元,张惠娟,等.径向基函数神经网络在传染病发病率预测中的应用[J].中华流行病学杂志,2007,28(12): 1219-1222.
- [28] Hamzacebi C. Improving artificial neural networks' performance in seasonal time series forecasting [J]. Information Sciences, 2008, 178(23): 4550-4559.
- [29] Zheng Z Y, Guo X N, Zhu K X, et al. Artificial neural network-Genetic algorithm to optimize wheat germ fermentation condition: Application to the production of two anti-tumor benzoquinones [J]. Food Chemistry, 2017, 227(JUL. 15): 264.
- [30] 潘海博,覃璐琪,梅丽华,等.罗伊氏乳杆菌 LTR1318 的培养工艺优化[J].现代食品科技,2020,36(10):59-67.
- [31] 张丹丹,黄鑫磊,程卫东.基于前反馈神经网络分析优化 *Saccharomyces cerevisiae* L9 富硒条件[J].中国食品添加剂, 2021,32(12):36-42.
- [32] 魏泉增,范江涛,张成丽,等.BP 神经网络结合遗传算法优化链霉菌产葡萄糖异构酶发酵条件[J].食品科技,2021,46(12): 34-41.
- [33] 邹立飞,郑鹏.响应面法、BP 神经网络优化薏仁米酒产氨基酸态氮[J].食品研究与开发,2021,42(9):121-130.
- [34] Baltova K, Dimitrov Z. Probiotic and cultural characteristic of strain *Lactobacillus gasseri* 4/13 of human origin [J]. Biotechnol Biotechnol Equip, 2014, 28(6): 1084-1088.
- [35] Avonts L, Uytven E V, Vuyst L D. Cell growth and bacteriocin production of probiotic *Lactobacillus* strains in different media [J]. International Dairy Journal, 2004, 14(11): 947-955.
- [36] Terraf M C, Juarez T M, Nader-Macias M E, et al. Screening of biofilm formation by beneficial vaginal lactobacilli and influence of culture media components [J]. J Appl Microbiol, 2012, 113(6): 1517-1529.
- [37] 郭艳荣.格氏乳杆菌体外筛选及高密度培养工艺研究[D].呼和浩特:内蒙古农业大学,2020.
- [38] 陈齐,郑建丰,马章献,等.一种格氏乳杆菌 LG-G12 高密度发酵及提高菌粉质量的方法,CN112063547A[P].上海市: 2020-12-11.