

# 基于机器学习方法构建幽门螺杆菌感染的风险预测模型



刘吉莉<sup>1</sup>, 王凤美<sup>2</sup>, 刘 阳<sup>1</sup>, 张梦华<sup>3</sup>, 张珍珍<sup>1</sup>, 尹 博<sup>1</sup>, 刘钰彤<sup>1</sup>, 焦胜春<sup>4</sup>

1. 石河子大学药学院 (新疆石河子 832002)
2. 石河子大学第一附属医院消化内科 (新疆石河子 832008)
3. 新疆华世丹药业有限公司 (乌鲁木齐 830000)
4. 石河子大学第一附属医院临床教学科 (新疆石河子 832008)

**【摘要】**目的 基于幽门螺杆菌 (Hp) 感染的相关因素, 构建 Hp 感染的风险预测模型, 为临床预防 Hp 感染提供新方法。方法 对石河子大学第一附属医院消化内科门诊行 <sup>13</sup>C、<sup>14</sup>C 尿素呼气试验的检测者, 使用问卷调查本地 Hp 感染现状。按 7:3 随机将数据集分成训练集和测试集, 以是否发生 Hp 感染为结局变量, 采用单因素分析和多因素 Logistic 回归分析筛选出有统计学差异的特征变量。在训练集中, 使用支持向量 (SVM)、K 邻近 (KNN)、逻辑回归 (LR) 模型、随机森林 (RF)、极限梯度提升 (XGB) 和轻量级梯度提升 (LightGBM) 6 种机器学习方法构建 Hp 感染风险预测模型, 并在测试集进行验证与评价, 通过各模型间预测性能对比, 筛选出最佳模型。应用 Shapley 加法解释 (SHAP) 方法对最佳模型进行可解释性分析。结果 本次调查人群 678 人, 其中训练集 475 人、测试集 203 人, XGB 模型准确率为 0.784, 精确率为 0.777, 召回率为 0.783, F1 值为 0.780, 精确率-召回曲线的曲线下面积 (AUPRC) 为 0.875、受试者工作特征曲线下面积 (AUC) 为 0.885、Brier 值为 0.140, 为最佳预测模型。基于 XGB 模型对特征变量进行重要性排序, 依次为 Hp 认知评分、喜食高盐高脂食物、共用牙杯水杯等生活用品、进食腌制食物、进食生大蒜。结论 基于 XGB 机器学习方法构建的 Hp 感染风险预测模型最佳, 有助于临床早期评估和预防 Hp 感染风险的发生。

**【关键词】**幽门螺杆菌; 相关因素; 机器学习; 预测模型; 支持向量; 随机森林

**【中图分类号】** R181.3\*2

**【文献标识码】** A

## Construction of the risk prediction model of *Helicobacter pylori* infection based on Machine learning method

LIU Jili<sup>1</sup>, WANG Fengmei<sup>2</sup>, LIU Yang<sup>1</sup>, ZHANG Menghua<sup>3</sup>, ZHANG Zhenzhen<sup>1</sup>, YIN Bo<sup>1</sup>, LIU Yutong<sup>1</sup>, JIAO Shengchun<sup>4</sup>

1. School of Pharmacy, Shihezi University, Shihezi 832002, Xinjiang Uighur Autonomous Region, China
2. Department of Gastroenterology, The First Affiliated Hospital of Shihezi University, Shihezi 832008, Xinjiang Uighur Autonomous Region, China
3. Xinjiang Hua Shidan Pharmaceutical Limited Company, Urumqi 830000, China
4. Clinical Teaching Division, The First Affiliated Hospital of Shihezi University, Shihezi 832008, Xinjiang Uighur Autonomous Region, China

Corresponding author: JIAO Shengchun, Email: 3095115810@qq.com

DOI: 10.12173/j.issn.2097-4922.202411064

基金项目: 新疆维吾尔自治区科技项目计划 (2021A03002)

通信作者: 焦胜春, 硕士, 副主任药师, Email: 3095115810@qq.com

**【Abstract】Objective** Based on the risk factors of *Helicobacter pylori* (Hp) infection, to construct the risk prediction model of Hp infection, and provide a new method for clinical prevention of Hp infection. **Methods** Subjects who underwent  $^{13}\text{C}$  and  $^{14}\text{C}$  urea breath test in the department of gastroenterology at The First Affiliated Hospital of Shihezi University were investigated to explore the local Hp infection status using a questionnaire. The dataset was randomly divided into a training set and a test set according to the ratio of 7 : 3, with whether Hp infection occurred as the outcome variable. Univariate analysis and multivariate Logistic regression analysis were used to screen out the characteristic variables with statistical differences. In the training set, six machine learning methods, support vector (SVM), k-proximity (KNN), logistic regression (LR) model, random forest (RF), limit gradient lifting (XGB) and light gradient Enhancer (LightGBM), were used to construct a Hp infection risk prediction model, which was verified and evaluated in the test set. The optimal model was selected by comparing the prediction performance among the models. The interpretability of the optimal model was analyzed by Shapley Additive interpretation (SHAP) method. **Results** There were 678 people in the survey, including 475 in the training set and 203 in the test set. The accuracy rate of the XGB model was 0.784, the accuracy rate was 0.777, the recall rate was 0.783, the F1 value was 0.780, the area under the precision recall curve (AUPRC) value was 0.875, the area under characteristic (AUC) was 0.885, and the Brier value was 0.140, which was the best prediction model. Based on the XGB model, the importance of the characteristic variables was ranked as Hp cognitive score, eating high-salt and high-fat food, sharing daily necessities such as dental cups and water cups, eating pickled food, and eating raw garlic. **Conclusion** The Hp infection risk prediction model based on XGB machine learning method is the best, which is helpful for early clinical assessment and prevention of Hp infection risk.

**【Keywords】** *Helicobacter pylori*; Correlative factor; Machine learning; Prediction model; Support vector; Random forest

幽门螺杆菌 (*Helicobacter pylori*, Hp) 是一种革兰氏阴性杆菌, 通常定植于人体胃黏膜, 被认为是导致多种消化系统疾病的主要致病菌。1983年, 澳大利亚学者马歇尔和沃伦首次从胃炎和胃溃疡患者中分离和培养出这种细菌<sup>[1]</sup>。已有研究表明, Hp与胃癌、消化不良、慢性胃炎、消化性溃疡、胃黏膜相关组织淋巴瘤等疾病的发生发展密切相关<sup>[2-3]</sup>, 1994年世界卫生组织将Hp列为I类致癌物<sup>[4]</sup>, 这些疾病不仅影响患者身心健康, 还会造成沉重的经济负担。据调查显示, 我国人群具有较高的感染率, 一般在40%~60%<sup>[5]</sup>。目前, 临床中关于Hp感染风险预测的工具比较匮乏, 难以识别早期患病风险。近年来, 机器学习被越来越多地应用于医疗领域, 与传统的回归模型相比, 机器学习擅长处理复杂变量间的非线性关系, 构建出更加稳健的风险预测模型<sup>[6]</sup>。本研究旨在借助机器学习方法构建多个Hp感染风险预测模型, 并对其预测性能进行评价, 筛选出最佳预测模型, 探讨Hp感染的相关因素, 以期预防Hp感染提供一个有效的预测工具, 降低Hp感染风险。

## 1 资料与方法

### 1.1 研究对象

选取2024年3至8月在石河子大学第一附属医院消化内科门诊行 $^{13}\text{C}$ 、 $^{14}\text{C}$ 尿素呼气试验的检测者进行问卷调查。纳入标准: ①未进行过Hp根除治疗; ②年龄 $\geq 18$ 岁; ③能理解问卷内容并准确回答问题; ④检测前4周内未服用任何抗菌药物、质子泵抑制剂、 $\text{H}_2$ 受体拮抗剂及铋剂等药物。排除标准: ①年龄 $< 18$ 岁; ②孕妇及哺乳期妇女; ③近1周内曾有上消化道出血病史; ④既往行胃部手术史; ⑤有精神病史者; ⑥严重脏器功能衰竭者。本研究已获得患者知情同意, 经石河子大学第一附属医院伦理委员会(伦理审批号: KJ2023-383-01)审批通过。

### 1.2 方法

#### 1.2.1 检测方法

$^{13}\text{C}$ -UBT尿素呼气试验: 受试者应空腹或禁食2h以上, 维持正常呼吸, 吹满集气袋(勿深呼吸, 勿断续吐气), 并立即拧紧气袋盖, 此时收集的为0min呼气样品; 用20mL常温饮用水吞服尿

素  $^{13}\text{C}$  胶囊（深圳市中核海得威生物科技有限公司，规格：75 mg）；安静等候 30 min，期间不要运动、饮食、饮水，受试者按照上述集气方法呼气；然后将 0 min 和 30 min 收集的两管呼气样品在呼气试验测试仪上（海得威，型号：HCBT-01）进行检测。阳性判断值  $\geq 4.0\%$  为 Hp 阳性，反之则为阴性。

$^{14}\text{C}$ -UBT 尿素呼气试验：受试者应空腹或禁食 2 h 以上，用约 20 mL 常温饮用水吞服尿素  $^{14}\text{C}$  胶囊（深圳市中核海得威生物科技有限公司，规格：0.75  $\mu\text{Ci}$ ）；安静等候 25 min，期间不要运动、饮食、饮水；然后开启  $\text{CO}_2$  集气剂一瓶，插入一洁净的有防倒流装置的气体导管，导管下端应浸入集气剂液内，受试者通过导管吹气 3 min；气体样品收集完毕，向样品瓶内加入稀释闪烁液 4.5 mL，加盖密封；溶解摇匀后在呼气试验测试仪上测定 1 min。阳性判断值  $\geq 100$  dpm/mmol 为 Hp 阳性，反之则为阴性。

### 1.2.2 问卷调查

基于《全国幽门螺杆菌研究问卷调查表》，查阅相关文献<sup>[7-9]</sup>，并结合专家意见，设计统一的调查表。在正式调查之前，选取 100 名符合本研究纳排标准的调查对象进行预调查，调查结束后，评估问卷的信效度，对题义模糊、易产生歧义、不合理的条目进行修改，最终形成正式版调查表。开始调查前，向调查对象说明调查目的、过程及保密原则，取得调查对象同意。调查过程中，对于调查对象有疑问的地方，研究者使用统一指导语对调查对象进行解释说明。避免使用暗示和引导性言语，确保结果的真实性、准确性。每位调查对象均由经过培训的研究者指导问卷填写，调查完成后研究者当场进行资料审核，保证内容真实、填写完整、无误后，回收问卷。根据问卷编制原则，样本量为问卷条目数的 5~10 倍<sup>[10]</sup>，本研究中涉及患者一般资料、慢性疾病史、生活饮食习惯、Hp 认知等内容，共计 44 项，考虑 10% 的失访量，计算可知所需样本量为 242~484 例。排除资料不全、描述不清以及回答结果矛盾的问卷，本研究共收集了 678 例样本。

调查内容包括：①一般资料：性别、年龄、身高、体重、居住地、家庭常住人口、婚姻状况、家庭月平均收入、教育程度、工作类型。根据调查对象身高、体重，计算出各自的体重指数（body

mass index, BMI），计算公式为  $\text{BMI} (\text{kg}/\text{m}^2) = \text{体重} / \text{身高的平方}$ 。②慢性疾病：慢性胃炎、消化性溃疡、胃食管返流、胃癌家族史、高血压病、糖尿病、冠心病、其他疾病、家庭成员 Hp 感染。③生活习惯：运动、饭前便后洗手、在外就餐、共用牙杯水杯等生活用品、压力大精神紧张、吸烟、饮酒、睡眠质量。④饮食习惯：饮食规律、饮食偏好、进食生大蒜、进食豆类及其制品、进食奶类及其制品、进食腌制食物、进食烟熏食物。 $\geq 150$  min / 周中等强度的有氧运动定义为有运动习惯；持续吸烟  $\geq 1$  年定义为吸烟；近 3 个月平均每周喝酒  $\geq 1$  次定义为饮酒；频率规定：从不、偶尔（1~2 次 / 周），经常（3~5 次 / 周），每天（ $> 5$  次 / 周）。⑤ Hp 认知水平：Hp 治疗、传播、危害、筛查时机、检测、预防等共计 21 分（0~7 分为认知水平低，8~14 分为认知水平中等，15~21 分为认知水平高）。问卷量表的 Cronbach's  $\alpha$  系数为 0.894，KMO（Kaiser-Meyer-Olkin）值为 0.859，巴特利球形检验显著性为  $P < 0.001$ ，问卷信效度较好。

### 1.2.3 数据预处理

异常值处理：录入时进行溯源修正，若无法修正，删除为缺失值，缺失值  $> 30\%$  的变量进行舍弃。缺失值处理：通过 R 4.3.2 版本 miss Forest 包使用随机森林插补法，对缺失  $< 30\%$  的变量进行处理。对分类变量进行 one-hot 编码转换，使之符合机器学习计算过程；采用合成少数过采样技术（Synthetic Minority Oversampling Technique, SMOTE）解决类别不平衡问题，避免分类器的过度拟合。

### 1.2.4 特征变量筛选

将年龄、BMI 等连续变量转换为分类变量，采用 SPSS 26.0 统计软件进行数据分析及处理，Hp 感染率用百分数表示，组间单因素分析采用 Pearson  $\chi^2$  检验或 Fisher 精确检验，将单因素分析有统计学差异的变量纳入多因素 Logistic 回归分析，采用双向逐步回归法、似然比检验进行自变量特征筛选， $P < 0.05$  为差异有统计学意义。将多因素 Logistic 回归筛选出的相关变量导入 Python 3.9.7 软件后通过 Spearman 相关法分析并绘制各个特征之间的相关系数热力图。

### 1.2.5 模型构建与调优

使用 Python 3.9.7、Anaconda3 集成开发环境

进行模型构建。将数据按 7 : 3 随机分为训练集和测试集, 在训练集进行模型构建, 在测试集进行模型评价。使用 Scikit-learn 包, 基于 6 种常见的机器学习方法进行模型构建, 分别构建支持向量 (support vector machine, SVM) 模型、K 邻近 (k-nearest neighbor, KNN) 模型、逻辑回归 (logistic regression, LR) 模型、随机森林 (random forest, RF) 模型、极限梯度提升 (extreme gradient boosting, XGB) 模型和轻量级梯度提升 (light gradient boosting machine, LightGBM) 模型; 使用 bayes\_opt 包中的贝叶斯优化 (Bayesian optimization, BO) 算法, 获取 6 种机器学习模型的最优超参数组合, 进行模型拟合。

### 1.2.6 模型验证与评价

使用 Scikit-learn 包中的五折交叉验证法评价模型的预测性能并防止模型过拟合, 模型评价指标采用区分度、校准度及决策曲线分析。通过计算准确率 (accuracy)、精确率 (precision)、召回

率 (recall)、F1 分数 (F1 score)、受试者工作特征 (receiver operating characteristic, ROC) 曲线下面积 (area under characteristic, AUC)、精准-召回曲线的曲线下面积 (area under the precision recall curve, AUPRC) 评价模型的区分度; 通过绘制校准曲线及计算布里尔评分 (Brier score) 评价模型的校准度; 通过绘制决策曲线分析 (decision curve analysis, DCA) 评价模型在临床决策中的效用性。

## 2 结果

### 2.1 一般情况

本次调查人群 678 人 (Hp 阳性 313 人, 阴性 365 人), 其中训练集 475 人, 测试集 203 人。研究结果显示, Hp 感染率为 46.16%。Hp 阳性与 Hp 阴性组 Hp 认知评分、共用牙刷水杯等生活用品、无特殊饮食偏好、进食腌制食物、喜食高盐高脂食物、喜食热烫食物、进食生大蒜差异有统计学意义 ( $P < 0.05$ )。具体见表 1。

表1 Hp感染的单因素分析 [n (%)]  
Table 1. Univariate analysis of Hp infection [n (%)]

因素	例数 (n=678)	Hp阴性 (n=365)	Hp阳性 (n=313)	感染率 (%)	$\chi^2$	P
基本情况						
性别					1.935	0.164
男	299 (44.10)	152 (41.64)	147 (46.96)	49.16		
女	379 (55.90)	213 (58.36)	166 (53.04)	43.80		
年龄 (岁)					4.598	0.100
18~40	266 (39.23)	150 (41.10)	116 (37.06)	43.61		
41~60	386 (56.93)	206 (56.44)	180 (57.51)	46.63		
>60	26 (3.83)	9 (2.47)	17 (5.43)	65.38		
民族					1.337	0.248
汉族	626 (92.33)	341 (93.42)	285 (91.05)	45.53		
其他民族	52 (7.67)	24 (6.58)	28 (8.95)	53.85		
BMI (kg/m <sup>2</sup> )					4.493	0.106
<18.5	44 (6.49)	30 (8.22)	14 (4.47)	31.82		
18.5~24	329 (48.53)	169 (46.30)	160 (51.12)	48.63		
>24	305 (44.99)	166 (45.48)	139 (44.41)	45.57		
居住地					3.480	0.062
城镇	462 (68.14)	260 (71.23)	202 (64.54)	43.72		
农村	216 (31.86)	105 (28.77)	111 (35.46)	51.39		
家庭常住人口 (人)					0.134	0.935
1	34 (5.01)	19 (5.21)	15 (4.79)	44.12		
2~4	592 (87.32)	319 (87.40)	273 (87.22)	46.11		
≥5	52 (7.67)	27 (7.40)	25 (7.99)	48.08		
婚姻状况					3.722	0.155
未婚	114 (16.81)	62 (16.99)	52 (16.61)	45.61		
已婚	519 (76.55)	285 (78.08)	234 (74.76)	45.09		
其他	45 (6.64)	18 (4.93)	27 (8.63)	60.00		

续表1

因素	例数 (n=678)	Hp阴性 (n=365)	Hp阳性 (n=313)	感染率 (%)	$\chi^2$	P
家庭月平均收入 (元)					1.139	0.566
<3 000	135 (19.91)	73 (20.00)	62 (19.81)	45.93		
3 000~5 000	260 (38.35)	146 (40.00)	114 (36.42)	43.85		
>5 000	283 (41.74)	146 (40.00)	137 (43.77)	48.41		
教育程度					1.526	0.466
初中及以下	180 (26.55)	90 (24.66)	90 (28.75)	50.00		
高中或中专	164 (24.19)	92 (25.21)	72 (23.00)	43.90		
大专以上	334 (49.26)	183 (50.14)	151 (48.24)	45.21		
工作类型					2.145	0.342
体力劳动	193 (28.47)	112 (30.68)	81 (25.88)	41.97		
脑力劳动	271 (39.97)	144 (39.45)	127 (40.58)	46.86		
无工作 (退休)	214 (31.56)	109 (29.86)	105 (33.55)	49.07		
HP认知评分 (分)					98.643	<0.001
0~7	191 (28.17)	48 (13.15)	143 (45.69)	74.87		
8~14	382 (56.34)	234 (64.11)	148 (47.28)	38.74		
15~21	105 (15.49)	83 (22.74)	22 (7.03)	20.95		
慢性疾病						
慢性胃炎					3.058	0.080
否	348 (51.33)	176 (48.22)	172 (54.95)	49.43		
是	330 (48.67)	189 (51.78)	141 (45.05)	42.73		
消化性溃疡					0.008	0.928
否	615 (90.71)	331 (90.68)	284 (90.73)	46.18		
是	63 (9.29)	34 (9.32)	29 (9.27)	46.03		
胃食管反流					0.485	0.486
否	569 (83.92)	303 (83.01)	266 (84.98)	46.75		
是	109 (16.08)	62 (16.99)	47 (15.02)	43.12		
胃癌家族史					0.394	0.530
否	660 (97.35)	354 (96.99)	306 (97.76)	46.36		
是	18 (2.65)	11 (3.01)	7 (2.24)	38.89		
高血压病					2.451	0.117
否	605 (89.23)	332 (90.96)	273 (87.22)	45.12		
是	73 (10.77)	33 (9.04)	40 (12.78)	54.79		
糖尿病					0.827	0.363
否	649 (95.72)	347 (95.07)	302 (96.49)	46.53		
是	29 (4.28)	18 (4.93)	11 (3.51)	37.93		
冠心病					0.109	0.742
否	639 (94.25)	345 (94.52)	294 (93.93)	46.01		
是	39 (5.75)	20 (5.48)	19 (6.07)	48.72		
其他疾病					2.492	0.114
否	621 (91.59)	340 (93.15)	281 (89.78)	45.25		
是	57 (8.41)	25 (6.85)	32 (10.22)	56.14		
家庭成员Hp感染					2.516	0.113
否	507 (74.78)	264 (72.33)	243 (77.64)	47.93		
是	171 (25.22)	101 (27.67)	70 (22.36)	40.94		
生活习惯						
饭前便后洗手					0.507	0.776
偶尔	76 (11.21)	38 (50.00)	38 (50.00)	50.00		
经常	291 (42.92)	133 (45.70)	158 (54.30)	45.70		
每次	311 (45.87)	142 (45.66)	169 (54.34)	45.66		
在外就餐					4.187	0.123
从不	35 (5.16)	17 (4.66)	18 (5.75)	51.43		

续表1

因素	例数 (n=678)	Hp阴性 (n=365)	Hp阳性 (n=313)	感染率 (%)	$\chi^2$	P
偶尔	420 (61.95)	239 (65.48)	181 (57.83)	43.10		
经常	223 (32.89)	109 (29.86)	114 (36.42)	51.12		
共用牙杯水杯等生活用品					59.739	<0.001
从不	355 (52.36)	238 (65.21)	117 (37.38)	32.96		
偶尔	196 (28.91)	89 (24.38)	107 (34.19)	54.59		
经常	127 (18.73)	38 (10.41)	89 (28.43)	70.08		
压力大、精神紧张					3.299	0.192
从不	105 (15.49)	48 (13.15)	57 (18.21)	54.29		
偶尔	370 (54.57)	205 (56.16)	165 (52.72)	44.59		
经常	203 (29.94)	112 (30.68)	91 (29.07)	44.83		
运动习惯					0.733	0.392
否	263 (38.79)	147 (40.27)	116 (37.06)	44.11		
是	415 (61.21)	218 (59.73)	197 (62.94)	47.47		
吸烟					0.592	0.442
否	550 (81.12)	300 (82.19)	250 (79.87)	45.45		
是	128 (18.88)	65 (17.81)	63 (20.13)	49.22		
饮酒					2.654	0.103
否	501 (73.89)	279 (76.44)	222 (70.93)	44.31		
是	177 (26.11)	86 (23.56)	91 (29.07)	51.41		
睡眠质量					2.577	0.276
好	155 (22.86)	85 (23.29)	70 (22.36)	45.16		
一般	418 (61.65)	231 (63.29)	187 (59.74)	44.74		
差	105 (15.49)	49 (13.42)	56 (17.89)	53.33		
饮食习惯						
饮食规律					0.058	0.971
规律	226 (33.33)	123 (33.70)	103 (32.91)	45.58		
基本规律	352 (51.92)	188 (51.51)	164 (52.40)	46.59		
不规律	100 (14.75)	54 (14.79)	46 (14.70)	46.00		
饮食无特殊偏好					4.022	0.045
否	388 (57.23)	196 (53.70)	192 (61.34)	49.48		
是	290 (42.77)	169 (46.30)	121 (38.66)	41.72		
轻淡饮食					0.955	0.329
否	477 (70.35)	251 (68.77)	226 (72.20)	47.38		
是	201 (29.65)	114 (31.23)	87 (27.80)	43.28		
喜食甜食					0.371	0.543
否	561 (82.74)	305 (83.56)	256 (81.79)	45.63		
是	117 (17.26)	60 (16.44)	57 (18.21)	48.72		
喜食高盐高脂食物					98.544	<0.001
否	519 (76.55)	334 (91.51)	185 (59.11)	35.65		
是	159 (23.45)	31 (8.49)	128 (40.89)	80.50		
喜食辛辣刺激食物					3.748	0.053
否	488 (71.98)	274 (75.07)	214 (68.37)	43.85		
是	190 (28.02)	91 (24.93)	99 (31.63)	52.11		
喜食热烫饮食					4.454	0.035
否	617 (91.00)	340 (93.15)	277 (88.50)	44.89		
是	61 (9.00)	25 (6.85)	36 (11.50)	59.02		
进食生大蒜					18.322	<0.001
从不	154 (22.71)	72 (19.73)	82 (26.20)	53.25		
偶尔	427 (62.98)	222 (60.82)	205 (65.50)	48.01		
经常	97 (14.31)	71 (19.45)	26 (8.31)	26.80		
进食豆类及其制品					1.333	0.513

续表1

因素	例数 (n=678)	Hp阴性 (n=365)	Hp阳性 (n=313)	感染率 (%)	$\chi^2$	P
从不	14 (2.06)	8 (2.19)	6 (1.92)	42.86		
偶尔	457 (67.40)	239 (65.48)	218 (69.65)	47.70		
经常	207 (30.53)	118 (32.33)	89 (28.43)	43.00		
进食奶类及其制品					2.493	0.288
从不	23 (3.39)	13 (3.56)	10 (3.19)	43.48		
偶尔	405 (59.73)	208 (56.99)	197 (62.94)	48.64		
经常	250 (36.87)	144 (39.45)	106 (33.87)	42.40		
进食腌制食物					83.930	<0.001
从不	72 (10.62)	65 (17.81)	7 (2.24)	9.72		
偶尔	507 (74.78)	280 (76.71)	227 (72.52)	44.77		
经常	99 (14.60)	20 (5.48)	79 (25.24)	79.80		
进食烟熏食物					0.479	0.787
从不	112 (16.52)	63 (17.26)	49 (15.65)	43.75		
偶尔	532 (78.47)	285 (78.08)	247 (78.91)	46.43		
经常	34 (5.01)	17 (4.66)	17 (5.43)	50.00		

### 2.2 Hp感染多因素Logistic回归分析

以是否发生 Hp 感染为因变量（阴性=0，阳性=1），将单因素分析有统计学差异的因素为自变量（赋值见表2），进行多因素 Logistic 回归分析。结果显示 Hp 认知评分、进食生大蒜、进食腌制食物、共用牙杯水杯等生活用品、喜食高盐高脂食物（ $P<0.05$ ）差异有统计学意义。具体见表3。

### 2.3 模型构建与评价

最优特征的相关性热图显示，5个最优特征变量之间相关性不强，均可用于模型构建，具体见图1。使用SVM、KNN、LR、RF、XGB、

LightGBM 机器学习方法对训练集数据进行模型构建。6种机器学习模型预测性能比较，XGB 准确率为0.784，精确率为0.777，召回率为0.783，F1值为0.780，AUPRC为0.875、AUC为0.885，提示XGB模型的区分度最好，具体见表4、图2和图3。6种机器学习模型的校准曲线对比，XGB的Brier值为0.140，提示XGB模型的校准度最好，具体见表3和图4。在决策曲线分析对比中XGB模型有助于临床决策，具体见图5。在对6种机器学习模型进行综合比较后，选择预测能力最好的XGB模型作为最佳预测模型。

表2 多因素Logistic回归分析变量赋值

Table 2. Variable assignment for multivariate Logistic regression analysis

变量	赋值		
Hp认知评分	0~7分=1	8~14分=2	15~21分=3
进食生大蒜	从不=1	偶尔=2	经常=3
进食腌制食物	从不=1	偶尔=2	经常=3
共用牙杯水杯等生活用品	从不=1	偶尔=2	经常=3
喜食高盐高脂食物	否=0	是=1	

表3 Hp感染的多因素Logistic回归分析

Table 3. Multivariate Logistic regression analysis of Hp infection

因素	$\beta$	SE	Wald $\chi^2$	P	OR	95%CI
Hp认知评分 (分)						
0~7					1.000	
8~14	-1.397	0.235	-5.942	<0.001	0.247	(0.156, 0.392)
15~21	-2.578	0.362	-7.112	<0.001	0.076	(0.037, 0.155)
进食生大蒜						
从不					1.000	
偶尔	-0.316	0.255	-1.237	0.216	0.729	(0.442, 1.203)
经常	-1.423	0.367	-3.875	<0.001	0.241	(0.117, 0.495)
进食腌制食物						
从不					1.000	
偶尔	2.225	0.490	4.540	<0.001	9.253	(3.541, 24.183)

续表3

因素	$\beta$	SE	Wald $\chi^2$	P	OR	95%CI
经常 共用牙杯水杯等生活用品	3.917	0.568	6.894	<0.001	50.259	( 16.503, 153.064 )
从不					1.000	
偶尔	1.245	0.236	5.277	<0.001	3.474	( 2.188, 5.517 )
经常	1.547	0.278	5.568	<0.001	4.698	( 2.725, 8.098 )
喜食高盐高脂食物						
否					1.000	
是	2.169	0.269	8.060	<0.001	8.749	( 5.163, 14.825 )

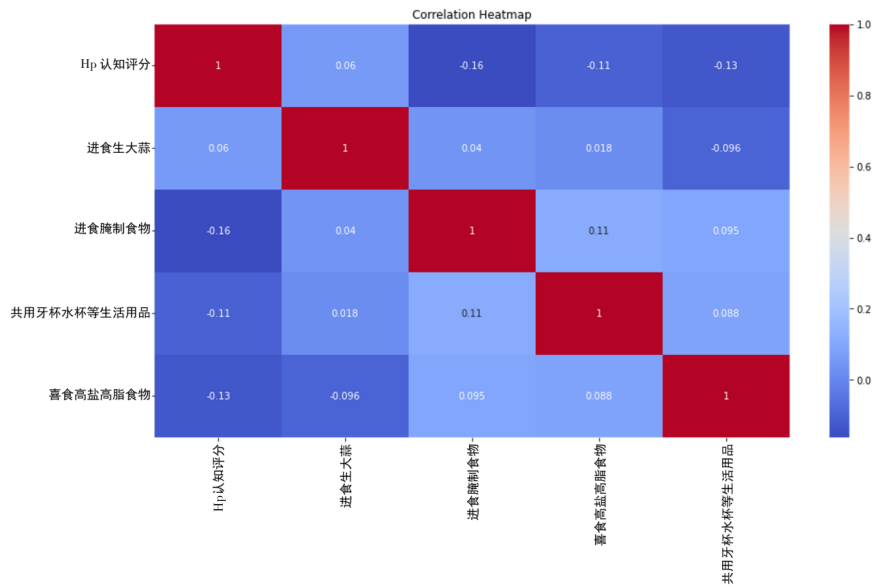


图1 最优特征的相关性热图

Figure 1. Correlation heat map of optimal features

表4 6种机器学习模型预测性能比较

Table 4. Comparison of prediction performance of six machine learning models

模型	准确率	精确率	召回率	F1值	Brier值	AUPRC	AUC
SVM	0.765	0.787	0.670	0.724	0.145	0.869	0.877
KNN	0.745	0.750	0.670	0.708	0.165	0.827	0.842
LR	0.769	0.797	0.670	0.728	0.151	0.865	0.879
RF	0.770	0.724	0.809	0.764	0.152	0.838	0.862
XGB	0.784	0.777	0.783	0.780	0.140	0.875	0.885
LightGBM	0.775	0.730	0.809	0.768	0.146	0.857	0.873

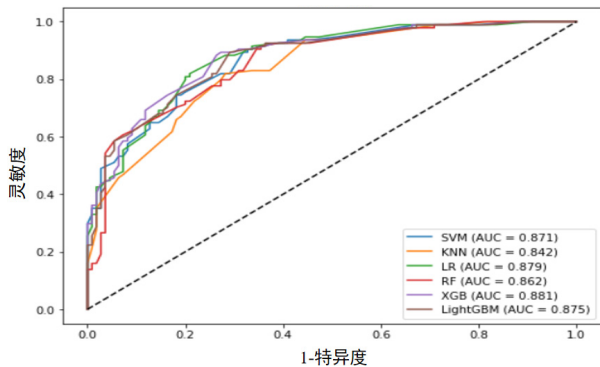


图2 6种机器学习模型的ROC曲线图

Figure 2. ROC curves of six machine learning models

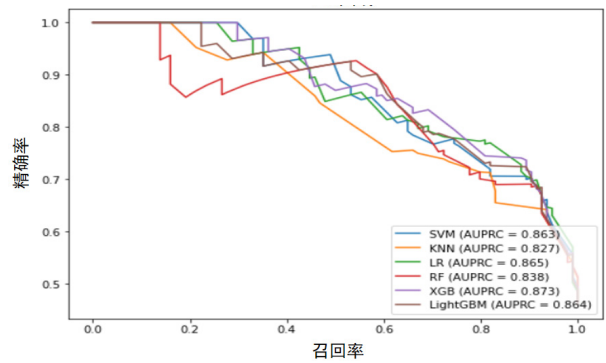


图3 6种机器学习模型的P-R曲线图

Figure 3. P-R curves of six machine learning models



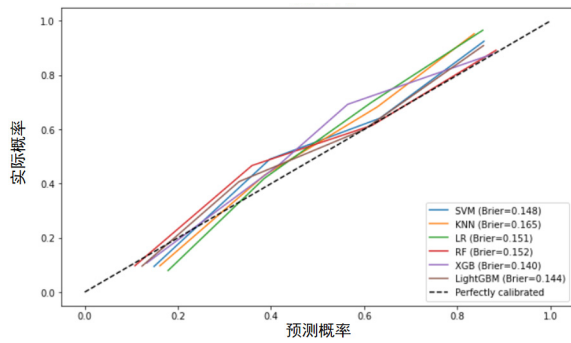


图4 6种机器学习模型的校准曲线图

Figure 4. Calibration curves of six machine learning models

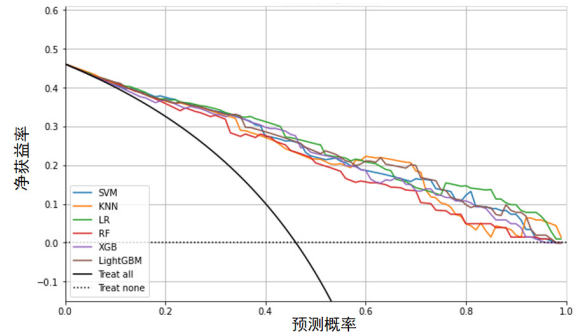


图5 6种机器学习模型的临床决策曲线图

Figure 5. Clinical decision curves of six machine learning models

### 2.4 模型可解释性分析

使用 Shapley 加法解释 (Shapley Additive exPlanation, SHAP) 方法对最佳预测模型 XGB 进行可解释性分析。SHAP 特征重要性矩阵图可以反映每个特征变量对于模型预测结果的贡献程度,从 SHAP 特征重要性矩阵图可知,5 个特征变量重要性从高到低依次是:Hp 认知评分、喜食高盐高脂食物、共用牙杯水杯等生活用品、进食腌制食物、进食生大蒜,具体见图 6。Shapley 加法解释图可以反映特征变量与结局变量之间的效应方向,从 Shapley 加法解释图可知, Hp 认知评分、进食生大蒜与结局变量呈负相关;进食腌制食物、共用牙杯水杯等生活用品、喜食高盐高脂食物与结局变量呈正相关,具体见图 7。

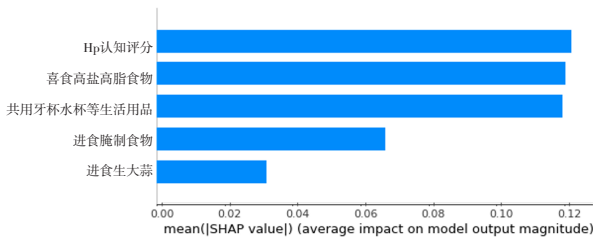


图6 SHAP特征重要性矩阵图

Figure 6. SHAP feature importance matrix

注: 横坐标表示贡献值, 值越大表示特征变量对预测模型的贡献程度越大。

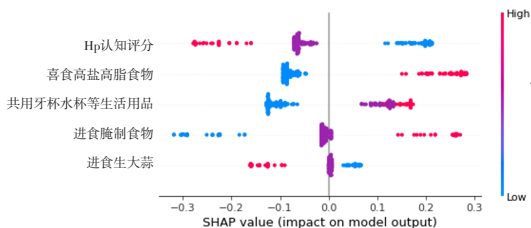


图7 Shapley加法解释图

Figure 7. Shapley addition interpretation diagram

注: 图中每个点代表数据集中的一个样本, 颜色代表数值的大小, 颜色越红代表数值越大, 越蓝代表越小; 横轴表示Shap值, 正值对模型的预测有反向影响 (增加Hp感染的风险), 反之亦然。

### 3 讨论

目前构建风险预测模型的方法主要是 Logistic 回归, 但 Logistic 回归无法拟合变量间的非线性关系, 在预测的精准度上仍有较大的提升空间。随着大数据时代的来临, 人工智能广泛应用医学领域, 其中机器学习以其高效的学习能力、较强的可塑性、灵活性, 以及较强的非线性关系处理能力, 已经成为了一种重要的预测分析手段<sup>[11]</sup>。现今, 基于机器学习方法建立的模型已成功用于预测药物不良反应及疾病的发生, 并表现出良好的预测能力<sup>[12]</sup>。

我国 Hp 感染率约为 43.70%, 其中东西部地区患病率较高, 中部地区较低<sup>[13]</sup>。本研究共收集 678 例样本, Hp 感染率为 46.16%, 高于全国平均水平, 可能与研究在门诊进行有关。门诊行 <sup>13</sup>C、<sup>14</sup>C 尿素呼气试验的检测者中, 部分人群已明确诊断患有慢性胃炎、消化性溃疡等疾病, 这也印证了 Hp 感染与胃肠道疾病间的密切联系。其中男性 Hp 感染率为 49.16%, 女性 Hp 感染率为 43.80%, 男性感染率高于女性, 性别与 Hp 感染差异无统计学意义, 与既往研究一致<sup>[14]</sup>。然而, 关于不同性别与 Hp 感染之间的相关性仍存在争议, 部分研究表明男性是 Hp 感染的危险因素<sup>[15]</sup>, 也有研究报道女性是 Hp 感染的危险因素<sup>[16]</sup>。本研究发现年龄与 Hp 感染呈正相关, 原因可能是随着年龄增长, 人体的免疫力逐渐下降, 胃黏膜防御能力减弱, 易受到 Hp 侵袭。本研究未发现吸烟、饮酒与 Hp 感染之间的关联性, 但既往有研究指出吸烟、饮酒会增加 Hp 感染的风险, 可能是烟、酒的长期摄入破坏胃黏膜屏障, 直接损伤胃黏膜, 使胃黏膜防御功能减弱, 从而增加了 Hp 定植概率<sup>[17-19]</sup>。

本研究多因素分析结果显示中高等 Hp 认知评分、经常进食生大蒜是 Hp 感染的保护因素；进食腌制食物、共用牙杯水杯等生活用品、喜食高盐高脂食物是 Hp 感染的危险因素。本研究大部分人群 Hp 认知评分处于中低等水平，Hp 认知评分越低感染率高，这与其他研究结果相似<sup>[20-21]</sup>。具有较高认知水平的人群对 Hp 感染有更深入的了解，可能更倾向于采取预防措施，越不容易感染 Hp，因此 Hp 认知评分高是 Hp 感染的保护因素。Hp 感染与生活饮食习惯密切相关，不良生活习惯如共用牙杯水杯等生活用品会增加 Hp 感染风险<sup>[22-23]</sup>。Hp 主要通过口-口、粪-口途径在人与人之间传播，共用牙杯水杯等生活用品属于亲密接触，可能会使唾液等体液沾染在这些物品上，从而传播给另一个人。建议家庭成员之间尽量避免共用餐具、杯具等个人生活用品，勤洗手，定期消毒餐具。不良饮食习惯，如喜食热烫食物、高盐高脂食物、腌制食物也会增加 Hp 感染的风险，有研究显示高盐高脂食物可损伤胃黏膜屏障，导致 Hp 在胃内的增殖能力增强，毒性因子 Cag A 表达增强，胃黏液黏度改变，故喜食高盐高脂食物更易感染 Hp<sup>[24]</sup>。腌制食物与 Hp 感染相关，可能是由于腌制食物在加工的过程中产生了大量的杂环胺类、N-亚硝基化合物等致癌物质，这些物质可能破坏胃黏膜屏障，增加 Hp 入侵并定植的概率，同时 Hp 感染产生尿素酶和过氧化氢酶，并分泌细胞毒素、诱导上皮细胞释放白介素，使胃黏膜进一步受损<sup>[25]</sup>。医务人员可多向民众宣传健康卫生知识，使其重视不良饮食习惯对身体的危害。亦有研究发现经常进食生大蒜 Hp 感染率低<sup>[26-27]</sup>。大蒜素是大蒜的关键活性成分，其抗菌作用主要归因于其分子结构中的巯基可抑制巯基酶的活性，从而抑制 Hp 的生长和繁殖，产生抗 Hp 作用<sup>[28]</sup>。然而也有研究显示进食生大蒜对 Hp 感染没有影响<sup>[29-30]</sup>。由此可见，进食生大蒜对 Hp 是否有杀菌作用尚未明确，还需要更多随机对照试验提供有力的证据。

目前 Hp 感染风险预测模型的相关研究多采用 Logistic 回归的方法。陈虎等<sup>[31]</sup>通过多因素 Logistic 回归结果构建列线图模型，预测 Hp 感染风险，模型 AUC 为 0.777；袁一鸣等<sup>[32]</sup>采用机器学习中的 Logistic 回归、决策树及添加交互项的 Logistic 回归，建立 Hp 感染的预测模型，交互项

Logistic 回归的预测性能更好，AUC 为 0.739；杜结玲等<sup>[33]</sup>根据向前逐步回归法构建 Hp 感染的风险预测回归方程，训练集为 AUC 为 0.7334。与传统的逻辑回归分析相比，机器学习方法具备出色的学习效率，可塑性强且灵活度高，擅长处理非线性关系<sup>[34]</sup>。

刘丁玮等<sup>[35]</sup>采用 12 种机器学习方法构建模型，识别 Hp 感染高危人群，其中 CatBoost 为最佳模型，AUC 为 0.744；Tran 等<sup>[36]</sup>采用 KNN、SVM、XGB、RF、LR、NB 机器学习方法预测学龄前儿童 Hp 感染状况并识别危险因素，研究表明 XGB 模型与 Logistic 回归的准确率相当；Liu 等<sup>[37]</sup>采用卡方检验、互信息（mutual Information, MI）、Relieff 和信息增益（information gain, IG）4 种方法，进行特征筛选，对排名前 10 和前 15 的特征分别进行 KNN、XGB、RF、LR、NB 模型的构建，其中 XGB 模型最佳，AUC 为 0.861。本研究采用单因素分析、多因素 Logistic 回归筛选相关因素，对有统计学差异的因素分别使用 SVM、KNN、LR、RF、XGB、LightGBM 6 种机器学习方法进行模型构建，选择最佳模型作为 Hp 感染的风险预测模型。对几个模型的拟合效果进行评估后，亦发现 XGB 模型在测试集中的 AUC 值高达 0.885，具有较好的灵敏度、特异度及稳定性，校准曲线和临床决策曲线也证实了 XGB 模型具有良好的校准效果，可为临床决策提供更有力的支持。

本研究存在一定的局限性：第一，本研究是单中心的横断面研究，仅纳入了石河子大学第一附属医院的患者，样本量较少，所观察的患者主要局限于石河子及其周边，可能存在偏倚；第二，本研究为横断面研究，无法进行因果推断，未来可进行多中心、大样本的队列研究，进一步探索预测因素与 Hp 感染之间的因果关系；第三，受研究规模所限，仅进行了内部验证，未对本预测模型进行外部验证；第四，仅纳入部分生活习惯等因素用于模型建立，未来可进一步探索更多相关因素对 Hp 感染风险的评估价值，不断优化模型；第五，仅对比了 6 种常用机器学习方法构建的预测模型性能，其他机器学习方法的效果有待进一步验证。未来可以纳入更广泛的预测模型和多元化的数据，来提高模型的普适性和预测的精确度。

综上所述，本研究探讨了 Hp 感染的相关因

素, 通过机器学习方法建立了相关风险预测模型, 该预测模型表现出较好的预测性能和稳定性, 可以较好的预测 Hp 感染发生的风险, 可为临床鉴定 Hp 易感人群和预防 Hp 感染提供科学依据。基于本研究, 未来可以制作网页版评估工具或应用程序提供个性化的幽门螺杆菌感染风险预测和相关建议。

## 参考文献

- 1 Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration[J]. Lancet, 1984, 11(8390): 1311–1315. DOI: [10.1016/s0140-6736\(84\)91816-6](https://doi.org/10.1016/s0140-6736(84)91816-6).
- 2 Pellicano R, Ianiro G, Fagoonee S, et al. Review: extragastric diseases and *Helicobacter pylori*[J]. Helicobacter, 2020, 25(Suppl 1): e12741. DOI: [10.1111/hel.12741](https://doi.org/10.1111/hel.12741).
- 3 Zebasil M, Endalkachew N, Tamirat H, et al. Prevalence and associated risk factors of *Helicobacter pylori* infection in East Africa: a systematic review and Meta-analysis[J]. Braz J Microbiol, 2023, 55(1): 51–64. DOI: [10.1007/s42770-023-01190-0](https://doi.org/10.1007/s42770-023-01190-0).
- 4 Schistosomes, liver flukes and Helicobacter pylori. IARC working group on the evaluation of carcinogenic risks to humans[J]. IARC Monogr Evall Carcinog Risks Hum, 1994, 61: 1–241. <https://pubmed.ncbi.nlm.nih.gov/7715068/>.
- 5 Mărginean CD, Mărginean CO, Melit LE. *Helicobacter pylori*-related extraintestinal manifestations myth or reality[J]. Children (Basel), 2022, 9(9): 1352. DOI: [10.3390/children9091352](https://doi.org/10.3390/children9091352).
- 6 Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine[J]. J Intern Med, 2018, 284(6): 603–619. DOI: [10.1111/joim.12822](https://doi.org/10.1111/joim.12822).
- 7 姚敏, 李艳梅. 人群中幽门螺杆菌感染现状及危险因素分析[J]. 世界最新医学信息文摘, 2024, 18(55): 85–86. [Yao M, Li YM. Analysis of *Helicobacter pylori* infection status and risk factors in population[J]. World Latest Medical Information Digest, 2024, 18(55): 85–86.] DOI: [10.19613/j.cnki.1671-3141.2018.55.036](https://doi.org/10.19613/j.cnki.1671-3141.2018.55.036).
- 8 王闫. 幽门螺杆菌感染患者管理现状及一般人群认知情况调查研究[D]. 郑州: 郑州大学, 2022. DOI: [10.27466/d.cnki.gzzdu.2022.004340](https://doi.org/10.27466/d.cnki.gzzdu.2022.004340).
- 9 Wang YX, Zou JY, Hu LF, et al. What is the general Chinese public's awareness of and attitudes towards *Helicobacter pylori* screening and associated health behaviours? a cross-sectional study[J]. BMJ Open, 2022, 12(1): e057929. DOI: [10.1136/bmjopen-2021-057929](https://doi.org/10.1136/bmjopen-2021-057929).
- 10 Devellis RF, 著. 席仲恩, 杜珏. 译. 量表编制: 理论与应用[M]. 重庆: 重庆大学出版社, 2016: 75–125.
- 11 Liew BXW, Kovacs FM, Rügamer D, et al. Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain[J]. Eur Spine J, 2022, 31(8): 2082–2091. DOI: [10.1007/s00586-022-07188-w](https://doi.org/10.1007/s00586-022-07188-w).
- 12 杨翀, 李旭东, 吕良福, 等. 基于机器学习预测动脉瘤性蛛网膜下腔出血预后模型的临床研究与应用[J]. 中国医院药学杂志, 2024, 44(3): 257–262. [Yang C, Li XD, Lyu LF, et al. Clinical research and application of machine learning-based prognostic model for aneurysmal subarachnoid hemorrhage[J]. Chinese Journal of Hospital Pharmacy, 2024, 44(3): 257–262.] DOI: [10.13286/j.1001-5213.2024.03.03](https://doi.org/10.13286/j.1001-5213.2024.03.03).
- 13 Lu TL, Zhang JM, Li SR, et al. Spatial-temporal distribution and influencing factors of *Helicobacter pylori* infection in Chinese mainland, 2001–2020: a systematic review and meta-analysis[J]. J Clin Gastroenterol, 2022, 56(5): e273–e282. DOI: [10.1097/MCG.0000000000001691](https://doi.org/10.1097/MCG.0000000000001691).
- 14 张晓冬, 张大涯, 陈世铜, 等. 海口市幽门螺杆菌感染现状与相关危险因素调查[J]. 现代消化及介入诊疗, 2024, 29(4): 393–397. [Zhang XD, Zhang DY, Chen SR, et al. Investigation of *Helicobacter pylori* infection status and related risk factors in Haikou city[J]. Modern Digestive and Interventional Diagnosis and Treatment, 2024, 29(4): 393–397.] DOI: [10.3969/j.issn.1672-2159.2024.04.001](https://doi.org/10.3969/j.issn.1672-2159.2024.04.001).
- 15 俞菊英, 尹强龙, 陈勇, 等. 嘉兴市秀洲区体检和门诊人群幽门螺杆菌感染状况调查[J]. 胃肠病学, 2018, 23(6): 363–365. [Yu JY, Yin QL, Chen Y, et al. Investigation on *Helicobacter pylori* infection in physical examination and outpatient population in Xiuzhou district of Jiaxing City[J]. Gastroenterology, 2018, 23(6): 363–365.] DOI: [10.3969/j.issn.1008-7125.2018.06.009](https://doi.org/10.3969/j.issn.1008-7125.2018.06.009).
- 16 田宏扬, 严华芳, 乔春萍, 等. 上海市浦东新区南部体检人群幽门螺杆菌感染情况及相关因素分析[J]. 复旦学报(医学版), 2022, 49(5): 720–725, 732. [Tian HY, Yan HF, Qiao CP, et al. Analysis of *Helicobacter pylori* infection and its related factors among physical examination population in southern Pudong New Area of Shanghai[J]. Journal of Fudan University, 2022, 49(5): 720–725, 732.] DOI: [10.3969/j.issn.1672-8467.2022.05.012](https://doi.org/10.3969/j.issn.1672-8467.2022.05.012).
- 17 吴美璇, 胡泽林, 贾圆露, 等. 幽门螺杆菌感染危险因素的 Meta 分析[J]. 吉林医学, 2024, 45(12): 2997–3000. [Wu MX, Hu ZL, Jia YL, et al. Meta-analysis of risk factors for *Helicobacter pylori* infection[J]. Jilin Medicine, 2024, 45(12): 2997–3000.] DOI: [10.3969/j.issn.1004-0412.2024.12.035](https://doi.org/10.3969/j.issn.1004-0412.2024.12.035).
- 18 马贞, 李曼玲, 唐璐, 等. 幽门螺杆菌感染危险因素 Meta 分析[J]. 现代消化及介入诊疗, 2020, 25(9): 1212–1216. [Ma Z, Li ML, Tang L, et al. Meta-analysis of risk factors for *Helicobacter pylori* infection[J]. Modern Digestive and Interventional Therapy, 2020, 25(9): 1212–1216.] DOI: [10.3969/j.issn.1672-2159.2020.09.018](https://doi.org/10.3969/j.issn.1672-2159.2020.09.018).
- 19 Khan MY, Aslam A, Mihali AB, et al. Effectiveness of *Helicobacter pylori* eradication in preventing metachronous gastric cancer and preneoplastic lesions: a systematic review and meta-analysis[J]. Eur J Gastroenterol Hepatol, 2020, 32(6): 686–694. DOI: [10.1097/MEG.0000000000001740](https://doi.org/10.1097/MEG.0000000000001740).
- 20 赵文芳, 刘鑫洋, 徐灿霞. 普通人群幽门螺杆菌感染认知度和治疗意愿[J]. 临床与病理杂志, 2024, 44(1): 63–70. [Zhao WF, Liu XY, Xu CX. Awareness and willingness to treat *Helicobacter pylori* infection in general population[J]. Chinese Journal of Clinical & Pathology, 2024, 44(1): 63–70.] DOI: [10.11817/](https://doi.org/10.11817/)

- j.issn.2095-6959.2024.230407.
- 21 肖婵, 曾小平, 王时波, 等. 慢性胃炎患者幽门螺杆菌检出现状及影响因素研究 [J]. 华南预防医学, 2023, 49(2): 156-159. [Xiao C, Zeng XP, Wang SB, et al. Study on *Helicobacter pylori* detection status and influencing factors in patients with chronic gastritis[J]. South China Preventive Medicine, 2023, 49(2): 156-159.] DOI: 10.12183/j.scjpm.2023.0156.
- 22 陈伟, 周景梅, 顾赟, 等. 东台地区健康体检人群幽门螺杆菌感染状况分析 [J]. 智慧健康, 2024, 10(4): 27-29, 33. [Chen W, Zhou JM, Gu Y, et al. Analysis of *Helicobacter pylori* infection among healthy people in Dongtai area[J]. Smart Health, 2024, 10(4): 27-29, 33.] DOI: 10.19335/j.cnki.2096-1219.2024.04.007.
- 23 达娃卓玛, 李生隆, 土旦格列, 等. 西藏地区体检人群幽门螺杆菌的流行病学调查 [J]. 高原科学研究, 2021, 5(3): 46-54. [Dawa ZM, Li SL, Tu Tangri, et al. Epidemiological investigation of *Helicobacter pylori* in physical examination population in Tibet[J]. Plateau Scientific Research, 2021, 5(3): 46-54.] DOI: 10.16249/j.cnki.2096-4617.2021.03.006.
- 24 Loh JT, Beckett AC, Scholz MB, et al. High-salt conditions alter transcription of *Helicobacter pylori* genes encoding outer membrane proteins[J]. Infect Immun, 2018, 86(3): e00626-17. DOI: 10.1128/IAI.00626-17.
- 25 Diniz LIB, Celino FDMC, Sousa DDJC, et al. Risk factors of *Helicobacter pylori* infection in an urban community in Northeast Brazil and the relationship between the infection and gastric diseases[J]. Rev Soc Bras Med Trop, 2018, 51(2): 183-189. DOI: 10.1590/0037-8682-0412-2016.
- 26 Kim H, Keum N, Giovannucci EL, et al. Garlic intake and gastric cancer risk: results from two large prospective US cohort studies[J]. Int J Cancer, 2018, 143(5): 1047-1053. DOI: 10.1002/ijc.31396.
- 27 罗鹏, 蒲柯, 杨国栋. 南充地区幽门螺杆菌感染现状及相关影响因素分析 [J]. 现代消化及介入诊疗, 2022, 27(8): 951-955. [Luo P, Pu K, Yang GD. Analysis of *Helicobacter pylori* infection status and related influencing factors in Nanchong area[J]. Modern Digestive and Interventional Diagnosis and Treatment, 2022, 27(8): 951-955.] DOI: 10.3969/j.issn.1672-2159.2022.08.004.
- 28 Hagh A, Azimi H, Rahimi R. A comprehensive review on pharmacotherapeutics of three phytochemicals, curcumin, quercetin, and allicin in the treatment of gastric cancer[J]. J Gastrointest Cancer, 2017, 48(4): 314-320. DOI: 10.1007/s12029-017-9997-7.
- 29 张莉莉. 上海宝山地区体检人群幽门螺杆菌感染现状及影响因素调查分析 [D]. 上海: 上海交通大学, 2019. DOI: 10.27307/d.cnki.gsytu.2019.002767.
- 30 Vale FF, Oleastro M. Overview of the phytomedicine approaches against *Helicobacter pylori*[J]. World J Gastroenterol, 2014, 20(19): 5594-5609. DOI: 10.3748/wjg.v20.i19.5594.
- 31 陈虎, 陈光侠, 余思锦, 等. 幽门螺杆菌易感因素分析及感染风险预测模型的构建 [J]. 中国医药导报, 2024, 21(30): 135-140. [Chen H, Chen GX, Yu SJ, et al. Analysis of susceptibility factors and construction of infection risk prediction model for *Helicobacter pylori*[J]. China Medical Review, 2024, 21(30): 135-140.] DOI: 10.20047/j.issn.1673-7210.2024.30.25.
- 32 袁一鸣, 杜结玲, 洪慧斯, 等. 机器学习对 *H.pylori* 感染患者的特征变量及预测模型研究 [J]. 胃肠病学和肝病学杂志, 2024, 33(8): 958-965. [Yuan YM, Du JL, Hong HS, et al. Study on characteristic variables and prediction model of *H.pylori* infection by machine learning[J]. Journal of Gastroenterology and Hepatology, 2024, 33(8): 958-965.] DOI: 10.3969/j.issn.1006-5709.2024.08.002.
- 33 杜结玲, 袁一鸣, 洪慧斯, 等. 数据挖掘构建幽门螺杆菌感染患者的预测模型和防治策略 [J]. 胃肠病学和肝病学杂志, 2022, 31(9): 992-998. [Du JL, Yuan YM, Hong HS, et al. Prediction model and prevention strategy of *Helicobacter pylori* infection by data mining[J]. Journal of Gastroenterology and Hepatology, 2022, 31(9): 992-998.] DOI: 10.3969/j.issn.1006-5709.2022.09.008.
- 34 Liew BXW, Kovacs FM, Rügamer D, Royuela A. Machine learning versus lo-gistic regression for prognostic modelling in individuals with nonspecific neck pain[J]. Eur Spine J, 2022, 31(8): 2082-2091. DOI: 10.1007/s00586-022-07188-w.
- 35 刘丁玮. 基于机器学习识别 HP 感染高危人群及构建根除后复发预测模型 [D]. 南昌: 南昌大学, 2024. DOI: 10.27232/d.cnki.gnchu.2024.003878.
- 36 Tran V, Saad T, Tesfaye M, et al. *Helicobacter pylori* (*H.pylori*) risk factor analysis and prevalence prediction: a machine learning-based approach[J]. BMC Infect Dis, 2022, 22(1): 655. DOI: 10.1186/s12879-022-07625-7.
- 37 Liu M, Liu S, Lu Z, et al. Machine learning based prediction of *Helicobacter pylori* infection study in adults[J]. Med Sci Monit, 2024, 30: e943666. DOI: 10.12659/MSM.943666.

收稿日期: 2024 年 11 月 18 日 修回日期: 2025 年 01 月 21 日  
本文编辑: 李 阳 钟巧妮