

基于机器学习的苏州地区9岁儿童 第一恒磨牙龋病预测模型研究

陈灵芝¹ 王霞琴¹ 朱凯飞² 任坤³ 吴桢¹

1. 苏州市吴中人民医院口腔科, 苏州 215128; 2. 苏州市吴中区疾控中心, 苏州 215128;

3. 苏州科技大学工程学院重点实验室, 苏州 215000

[摘要] **目的** 利用机器学习算法构建苏州地区9岁儿童第一恒磨牙龋病预测模型, 筛选危险因素。**方法** 采用随机分层整群抽样的方法, 在吴中区14个乡镇、街道的38所小学中随机抽取9岁在校学生进行口腔检查和问卷调查。采用Logistic多因素回归分析龋齿的危险因素。将数据集按8:2随机分为训练集及验证集, 使用R 4.3.1构建随机森林、决策树、极端梯度提升(XGBoost)、Logistic回归、轻量级梯度提升(LightGBM)5种机器学习算法, 应用受试者工作特征曲线下面积(AUC)评估5种模型的预测效果。通过沙普利加和解释(SHAP)量化特征对龋齿预测模型的边际贡献。**结果** 研究纳入符合标准的样本7 225例, 其中第一恒磨牙患龋率为54.96%, 多因素Logistic回归分析显示, 甜饮料、甜点和糖果、零食频率、刷牙后睡前零食等与第一恒磨牙龋齿的发生存在关联($P<0.05$)。决策树、Logistic回归、轻量级梯度提升、随机森林、极端梯度提升这5种预测模型的AUC值分别为75.5%、83.9%、88.6%、88.9%、90.1%。对比独热编码后的变量, 高频甜食(如甜点糖果每天 ≥ 2 次、母亲含糖饮食每天 ≥ 2 次)与不良口腔卫生习惯(如刷牙后睡前常吃零食、刷牙不规律)的SHAP值为正。**结论** 基于极端梯度提升算法构建苏州地区9岁儿童第一恒磨牙龋病的预测模型, 具有较好的预测效果。高频甜食和不良口腔卫生习惯对第一恒磨牙患龋有强正向影响, 是关键的驱动因素, 可用于针对性干预措施的制定。

[关键词] 第一恒磨牙; 机器学习; 影响因素; 预测模型

[中图分类号] 781.1 **[文献标志码]** A **[doi]** 10.7518/hxkq.2025.2025135



本文链接 开放科学标识码

Machine learning-based prediction model for caries in the first molars of 9-year-old children in Suzhou

Chen Lingzhi¹, Wang Xiaqin¹, Zhu Kaifei², Ren Kun³, Wu Zhen¹

1. Dept. of Dentistry, Suzhou Wuzhong People's Hospital, Suzhou 215128, China; 2. Suzhou Wuzhong Center for Disease Control and Prevention, Suzhou 215128, China; 3. The Key Laboratory of Engineering of Suzhou University of Science and Technology, Suzhou 215000, China

Supported by: Suzhou Wuzhong District Science and Technology Plan Youth Project (WZYW2021017)

Correspondence: Wu Zhen, E-mail: baikashui7707@126.com

[Abstract] **Objective** This study aimed to use machine learning algorithms to build a prediction model of the first permanent molar caries of 9-year-old children in Suzhou and screen out risk factors. **Methods** Random stratified whole group sampling was applied to randomly select 9-year-old students from 38 primary schools in 14 town-

ships and streets in Wuzhong District for oral examination and questionnaire survey. Multifactor Logistics regression was used to analyze the risk factors of tooth decay. The data set was randomly divided into training sets and verification sets according to 8:2, and R 4.3.1 was used to build five machine learning algorithms: random forest, decision tree, extreme gradient boosting

[收稿日期] 2025-04-10; **[修回日期]** 2025-11-01

[基金项目] 苏州市吴中区科技计划项目(医疗卫生领域)青年项目(WZYW2021017)

[第一作者] 陈灵芝, 主管护师, 学士, E-mail: clz823173371@sina.com

[通信作者] 吴桢, 副主任医师, 硕士, E-mail: baikashui7707@126.com

(XGBoost), Logistics regression, and lightweight gradient enhancement (LightGBM). The predictive effect of these five models was evaluated using the area under the characteristic curve (AUC). The marginal contribution of quantitative characteristics to the caries prediction model was determined through Shapley additive explanations (SHAP). **Results** This study included 7 225 samples that met the standard. The caries rate of the first permanent molar was 54.96%. Multifactor Logistic regression analysis showed that sweet drinks, dessert and candy, snack frequency, and snacks before going to bed after brushing teeth were correlated with the occurrence of first permanent molar caries ($P<0.05$). The AUC values of decision tree, Logistic regression, LightGBM, random forest, and XGBoost were 75.5%, 83.9%, 88.6%, 88.9%, and 90.1%, respectively. Compared with the variables after single heat coding, the SHAP value of high-frequency sweets (such as dessert candy ≥ 2 times a day, mother's sugary diet ≥ 2 times a day) and bad oral hygiene habits (such as frequent snacks before going to bed after brushing teeth and irregular brushing teeth) exhibited the highest positive. **Conclusion** XGBoost algorithm has a good prediction effect for first permanent molar caries in 9-year-old children. High-frequency sweet factors and bad oral hygiene habits have a strong positive impact on the risk of first permanent molar caries and are key drivers that can be used in the formulation of targeted interventions.

[Key words] first permanent molar; machine learning; influencing factor; prediction model

儿童龋病是国内外口腔医学长期关注、重点防治的慢性进行性破坏性疾病，至今仍呈现逐年增长的趋势，特别是在儿童替牙期^[1]。第四次全国口腔健康流行病学调查结果^[2]显示，儿童患龋率较10年前大幅度上升，严重影响儿童身心健康，同时增加家庭和社会经济负担^[3]；儿童的龋患风险因个体患龋经历、膳食糖的消耗、饮食习惯、口腔卫生预防行为、家庭及社会背景等因素的差异表现不同水平，患龋风险随着风险因素的变化出现动态改变^[4-5]。吴中区位于苏州市南部，属吴文化的发源地，是糖文化饮食代表地区，日常饮食制作工艺以精细及久煮为特点，风味调配中蜜、糖的使用频率高且范围广。同时，米面、糕点类食物极为盛行，整体饮食风貌以显著的精细、甜腻、黏糯著称^[6]，特殊的饮食偏好加重了该地区儿童罹患龋病的风险和负担。

伴随大数据应用的持续迭代，机器学习广泛实践于医学领域，比如开发疾病预测模型^[7-8]。通过数据训练模型，发现规律，预测趋势，机器学习展现了良好的预测精准度^[9]。风险控制与龋病的防治存在明确的正相关关系，采集分析患龋风险信息，预测患龋风险人群及高风险行为，针对性防龋治龋，是龋病防治的最新趋势^[10]。研究基于苏州地区的儿童饮食偏好及相关性行为的测量，采用随机森林、决策树、极端梯度提升等算法建立机器学习模型，综合比较出苏州地区儿童更为适配的龋病风险预测模型，推进针对性干预措施的制定。

1 材料和方法

1.1 研究对象及抽样

研究样本来自2024年苏州市吴中区儿童口腔综合干预项目的常住人口，根据苏州地区开展儿童口腔健康检查的政策措施，采用随机分层整群抽样进行调查，选取该地区7个街道（长桥街道、城南街道、越溪街道、横泾街道、郭巷街道、香山街道、太湖街道）和7个镇（角直镇、光福镇、临湖镇、木渎镇、金庭镇、东山镇、胥口镇）内38所公办小学的9岁儿童，各城乡的样本数量等额抽取，最终实际纳入儿童共计7 225名，收回完整调查研究7 225份。本研究经苏州市吴中人民医院伦理委员会审核批准（KY2021013），研究对象的信息采集和整理过程均告知家长，获得同意后开展。

纳入标准：1) 调查地9岁常住在校学生；2) 家长知情同意参与调查；3) 具备参与的意愿及能力。排除标准：1) 患有特殊疾病或服用药物的儿童；2) 非双亲家庭；3) 因主、客观因素无法参与的家庭。

1.2 数据采集

1.2.1 口腔检查

参照《第四次全国口腔流行病学抽样调查方案》及世界卫生组织最新版口腔流行病学调查方法^[11]。使用一次性口腔镜和牙周指数（community periodontal index, CPI）探针，在统一人工光源下，以视诊结合探诊的方式进行儿童口腔健康检

查。以班级为单位,检查前30 min常规刷牙,检查中必要情况时借助棉签拭去隐蔽处软垢,对第一恒磨牙进行评估,配专人记录。

1.2.2 问卷调查程序

参照《第四次全国口腔流行病学抽样调查方案》及调查表,包含内容如下。1) 社会人口学特征:性别、家庭籍贯、父母的最高学历等;2) 饮食偏好:含糖食品、饮品的发生频率和发生节点等;3) 口腔护理相关行为:餐后漱口、刷牙方法、刷牙频率、刷牙时长及刷牙后睡前进食等。开发线上程序,系统对输入数据进行加权分析,提示风险等级并罗列相关风险因素。在调查开展前,由各学校、班级向家长发放线上调查端口、纸质知情同意书,并附流程解释说明,由家长与儿童共同完成问卷调查。

1.3 标准定义

1.3.1 龋齿

参照《口腔健康调查检查方法》(WS/T472-2015)标准^[12],将存在龋洞、因龋缺失、因龋充填的牙齿均定义为龋齿。

1.3.2 户籍所在地

以儿童生活主要照顾者为调查对象,分为苏州本地与迁入两类,非苏州户籍均归纳为迁入。

1.4 质量控制

1.4.1 口腔检查

检查者由苏州市吴中人民医院6名5年以上临床经验并具备执业医师证书的口腔医生组成,登记员由3名5年以上临床经验的口腔科护士组成,均经项目办理论与实践的考核合格后进行,检查过程中,接受另一位检查者以调查对象5%的比率复查,标准一致性检验 kappa 值 ≥ 0.85 。此外,不定时接受质量控制小组的监督及可靠性检查,发现偏离标准时即刻校准。

1.4.2 线上程序问卷调查

综合儿童龋病风险评估因素及指标的优缺点,制定适合非专业人员使用的评估机制, Cronbach's α 系数0.82, 凯泽-迈耶-奥金度量(Kaiser-Meyer-Olkin measure, KMO)值0.91,信效度检验表现较好。以问答形式开发线上信息采集程序,通过手机端输入信息,设置逻辑自检机制,出现逻辑相悖或中途结束未完成等情况时跳出提醒框。程序开发人员加强与项目负责人沟通,深化项目认识,明确程序目的,完善程序的功能性并优化使用体验。同时建立用户端家长的使用反馈机制,数据采集完成后,按照标准及要求核查和验收资

料,随机抽取8%的比率进行人工电话回访,与线上调查的变量进行比对,保证信息采集的可靠性。

1.4.3 数据库

将口腔检查数据与线上采集信息经过验证后,导入项目数据库。数据库存储在专用设备,由专人负责保管,专用于本研究相关工作。

1.5 统计学方法

使用 Visual Studio Code (VS Code) 1.85.2 结合 R 4.3.1 进行数据整理与统计分析。定性数据使用 R 语言进行统计学分析,采用频数(百分比)的方式进行描述,通过 χ^2 检验对组间进行比较。多因素 Logistic 回归用于评估第一恒磨牙龋齿的相关因素。使用 VS Code 对数据进行建模,数据集按 8:2 的比例随机划分为训练集和验证集,用于模型构建与性能验证。将 Logistic 回归分析中 $P < 0.05$ 的因素纳入模型。依次运用决策树模型探讨各特征对儿童第一恒磨牙龋齿风险的影响;借助随机森林模型,整合多棵决策树,增强预测的稳定性和准确性;利用极端梯度提升模型,处理特征间复杂的非线性关系;依托 Logistic 回归模型,分析各特征对患龋的线性影响;基于极端梯度提升与决策树,构建轻量级梯度提升模型,实现计算速度更快且内存占用降低,以适配不同分析需求。模型评估指标涵盖准确率、精准度、召回率、特异度及受试者工作特征曲线下面积(area under the curve, AUC)值。其中, AUC 值显示较强的模型区分能力。对比后择优模型引入沙普利加和解释(Shapley additive explanations, SHAP)进行分析。 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 基本情况

项目共计纳入符合要求的9岁儿童7 225名,其中第一恒磨牙患龋人数3 971人(54.96%)。患龋儿童中,男性2 074人(占全体男性儿童的53.98%,占患龋儿童总数的52.23%),女性1 897人(占全体女性儿童的56.07%,占患龋儿童总数的47.77%)。迁入籍家庭3 138个,其中患龋儿童家庭1 558个(49.65%)。苏州籍家庭4 087个,其中患龋儿童家庭2 413个(59.04%)。单因素分析显示,家长籍贯、甜饮料、甜点和糖果、零食频率、刷牙后睡前零食、母亲含糖饮食频率等均有统计学意义,详见表1。

表 1 苏州地区儿童第一恒磨牙龋齿相关因素基本情况及单因素分析结果

Tab 1 Basic situation and single-factor analysis results of the relevant factors of first permanent molar caries in children in Suzhou

因素	总例数/ [n (%)]	患龋/ [n (%)]	χ^2 值	P值
性别			3.097	0.078
男	3 842 (53.18)	2 074 (53.98)		
女	3 383 (46.82)	1 897 (56.07)		
籍贯			62.868	2.21E-15
迁入	3 138 (43.43)	1 558 (49.65)		
苏州	4 087 (56.57)	2 413 (59.04)		
甜饮料			381.733	2.56E-80
很少/从不	812 (11.24)	362 (44.58)		
每月1~3次	1 062 (14.70)	463 (43.60)		
每周1次	1 560 (21.59)	718 (46.03)		
每周2~6次	2 101 (29.08)	1 183 (56.31)		
每天1次	1 238 (17.13)	908 (73.34)		
每天≥2次	452 (6.26)	337 (74.56)		
甜点和糖果			424.121	1.87E-89
很少/从不	1 074 (14.87)	472 (43.95)		
每月1~3次	1 228 (17.00)	566 (46.09)		
每周1次	1 427 (19.75)	674 (47.23)		
每周2~6次	1 817 (25.15)	987 (54.32)		
每天1次	775 (10.73)	566 (73.03)		
每天≥2次	904 (12.51)	706 (78.10)		
零食频率			775.092	2.83E-16
很少/从不	1 525 (21.11)	575 (37.70)		
每月1~3次	799 (11.06)	281 (35.17)		
每周1次	907 (12.55)	457 (50.39)		
每周2~6次	2 619 (36.25)	1 525 (58.23)		
每天1次	888 (12.29)	685 (77.14)		
每天≥2次	487 (6.74)	448 (91.99)		
刷牙后睡前零食			560.941	1.56E-12
不吃	3 619 (50.09)	1 671 (46.17)		
偶尔	2 323 (32.15)	1 220 (52.52)		
经常	1 283 (17.76)	1 080 (84.18)		
几岁起刷牙/岁			98.633	1.92E-20
2	1 467 (20.30)	694 (47.31)		
3	2 572 (35.60)	1 338 (52.02)		
4	1 562 (21.62)	913 (58.45)		
5	1 055 (14.60)	683 (64.74)		
6	569 (7.88)	343 (60.28)		
定期牙科诊所			813.410	6.55E-17
否	2 706 (37.45)	903 (33.37)		
是	4 519 (62.55)	3 068 (67.89)		
刷几次牙			220.263	1.48E-4
不是每天刷	611 (8.46)	403 (65.96)		
每天1次	1 954 (27.04)	1 307 (66.89)		
每天2次及以上	4 660 (64.50)	2 261 (48.52)		
刷牙时长/min			4.781	0.091
1	1 872 (25.91)	1 063 (56.78)		
2~3	4 179 (57.84)	2 288 (54.75)		

续表1

因素	总例数/ [n (%)]	患龋/ [n (%)]	χ^2 值	P值
≥3	1 174 (16.25)	620 (52.81)		
母亲患龋			20.873	4.91E-06
否	2 952 (40.86)	1 527 (51.73)		
是	4 273 (59.14)	2 444 (57.20)		
母亲最高学历			9.669 6	0.007 9
初中及以下	20 (0.28)	14 (70.00)		
高中、中专及大专	5 315 (73.56)	2 867 (53.94)		
本科及以上	1 890 (26.16)	1 090 (57.67)		
母亲含糖饮食频率			154.057	1.83E-31
很少/从不	648 (8.97)	338 (52.16)		
每月1~3次	477 (6.60)	290 (60.80)		
每周1次	1 409 (19.50)	602 (42.73)		
每周2~6次	3 353 (46.41)	924 (57.38)		
每天1次	1 008 (13.95)	571 (56.65)		
每天≥2次	330 (4.57)	246 (74.55)		
母亲刷牙频率			26.545	1.72E-06
不是每天刷	47 (0.65)	34 (72.34)		
每天1次	1 918 (26.55)	1 137 (59.28)		
每天2次及以上	5 260 (72.80)	2 800 (53.23)		
父亲患龋			13.805	0.000 2
否	3 935 (54.46)	2 084 (52.96)		
是	3 290 (45.54)	1 887 (57.36)		
父亲最高学历			5.25	0.072 4
初中及以下	45 (0.62)	27 (60.00)		
高中、中专及大专	4 998 (69.18)	2 703 (54.08)		
本科及以上	2 182 (30.20)	1 241 (56.87)		
父亲刷牙频率			11.16	0.003 7
不是每天刷	76 (1.05)	37 (48.68)		
每天1次	2 011 (27.83)	1 047 (52.06)		
每天2次及以上	5 138 (71.11)	2 887 (56.19)		
饭后漱口			168.695	1.43E-38
偶尔或从不	3 421 (47.35)	2 155 (62.99)		
漱口	3 804 (52.65)	1 816 (47.74)		

注: $P < 0.05$, 差异有统计学意义。

2.2 多因素 Logistic 回归分析

本研究因变量设定为是否患龋, 多因素 Logistic 回归分析显示, 甜饮料频率、甜点和糖果频

率、零食频率、刷牙后睡前零食等差异均有统计学意义, 详见表2。

表 2 苏州地区9岁儿童第一恒磨牙龋齿相关因素的多因素 Logistic 回归分析

Tab 2 Multifactor Logistic regression analysis of factors related to first permanent molar caries in 9-year-old children in Suzhou

因素	B 值	S.E 值	Wald 值	P 值	OR (95%CI) 值
家长籍贯 (苏州)					1.00 (Reference)
迁入	0.15	0.08	1.84	0.66	1.16 (0.99-1.37)
甜饮料 (很少/从不)					1.00 (Reference)
每月1~3次	-1.44	0.32	-4.47	<0.001	0.24 (0.13-0.45)
每周1次	-0.81	0.23	-3.46	<0.001	0.45 (0.28-0.70)
每周2~6次	0.35	0.25	1.43	0.153	1.42 (0.88-2.30)

续表2

因素	B 值	S.E 值	Wald 值	P 值	OR (95%CI) 值
每天1次	-2.51	0.44	-5.67	<0.001	0.08 (0.03-0.19)
每天≥2次	-1.38	0.41	-3.36	<0.001	0.25 (0.11-0.56)
甜点和糖果 (很少/从不)					1.00 (Reference)
每月1~3次	1.15	0.03	3.87	<0.001	3.17 (1.77-5.68)
每周1次	0.30	0.22	1.40	0.161	1.36 (0.89-2.08)
每周2~6次	-0.07	0.24	-0.31	0.758	0.93 (0.58-1.48)
每天1次	3.52	0.45	7.78	<0.001	33.83 (13.94-82.12)
每天≥2次	3.27	0.42	7.82	<0.001	26.34 (11.60-59.81)
零食频率 (很少/从不)					1.00 (Reference)
每月1~3次	0.25	0.17	1.48	0.140	1.28 (0.92-1.78)
每周1次	1.08	0.14	7.96	<0.001	2.93 (2.25-3.82)
每周2~6次	1.39	0.11	12.51	<0.001	4.0 (3.22-4.96)
每天1次	2.11	0.14	14.95	<0.001	8.23 (6.24-10.85)
每天≥2次	3.58	0.24	15.09	<0.001	35.90 (22.55-57.15)
刷牙后睡前零食 (经常)					1.00 (Reference)
不吃	-1.81	0.11	-15.76	<0.001	0.16 (0.13-0.21)
偶尔	-1.10	0.12	-8.97	<0.001	0.33 (0.26-0.42)
几岁刷牙 (2岁)					1.00 (Reference)
3岁	0.10	0.09	1.11	0.269	1.10 (0.93-1.31)
4岁	0.19	0.10	1.90	0.058	1.21 (0.99-1.47)
5岁	0.32	0.12	2.77	0.006	1.38 (1.10-1.37)
6岁	-0.20	0.14	-1.37	0.170	0.82 (0.62-1.09)
刷几次牙 (每天2次及以上)					1.00 (Reference)
每天1次	0.81	0.09	9.44	<0.001	2.25 (1.90-2.66)
不是每天刷	0.78	0.13	6.19	<0.001	2.18 (1.70-2.79)
刷牙时长 (1 min)					1.00 (Reference)
2~3 min	0.06	0.08	0.81	0.418	1.06 (0.91-1.24)
≥3 min	0.09	0.10	0.83	0.405	1.09 (0.89-1.34)
饭后漱口 (偶尔或从不)					1.00 (Reference)
漱口	-0.67	0.08	-8.45	<0.001	0.51 (0.44-0.60)
母亲最高学历 (初中及以下)					1.00 (Reference)
高中、中专及大专	0.63	0.87	0.72	0.470	1.88 (0.34-10.42)
本科及以上	0.66	0.87	0.76	0.045	1.93 (0.35-10.63)
母亲含糖饮食频率 (很少/从不)					1.00 (Reference)
每月1~3次	0.26	0.22	1.18	0.237	1.29 (0.84-1.98)
每周1次	-0.33	0.14	-2.39	0.017	0.72 (0.55-0.94)
每周2~6次	0.19	0.12	1.55	0.121	1.21 (0.95-1.54)
每天1次	-0.14	0.14	-0.97	0.332	0.87 (0.66-1.15)
每天≥2次	0.94	0.21	4.44	<0.001	2.56 (1.69-3.88)
母亲刷牙频率 (每天2次及以上)					1.00 (Reference)
不是每天刷	0.14	0.51	0.27	0.789	1.15 (0.42-3.11)
每天1次	0.21	0.08	2.43	0.015	1.23 (1.04-1.45)
母亲患龋 (是)					1.00 (Reference)
否	-0.31	0.08	-4.11	<0.001	0.73 (0.63-0.85)
父亲最高学历 (初中及以下)					1.00 (Reference)
高中、中专及大专	-0.71	0.43	-1.64	0.100	0.49 (0.21-1.15)
本科及以上	-0.76	0.44	-1.74	0.042	0.47 (0.20-1.10)
父亲患龋 (是)					1.00 (Reference)
否	0.30	0.08	3.59	<0.001	1.35 (1.15-1.60)

续表2

因素	B 值	S.E 值	Wald 值	P 值	OR (95%CI) 值
父亲刷牙频率 (每天2次及以上)					1.00 (Reference)
每天1次	-0.40	0.08	-4.90	<0.001	0.67 (0.57-0.79)
不是每天刷	-0.10	0.32	-0.30	0.765	0.91 (0.48-1.71)
定期口腔牙科诊室看牙 (是)					1.00 (Reference)
否	-1.45	0.07	-21.25	<0.001	0.23 (0.20-0.27)

注: P<0.05, 差异有统计学意义。

2.3 模型的预测效能对比

将 Logistic 回归分析后具有统计学意义的变量纳入 5 模型, 算法分类结果见表 3。5 种模型的准确率、精准度均大于 69.5%, AUC 均大于 75.5%。对比后总体表现最佳的是极端梯度提升模型, AUC 为 90.1%, 准确率为 81.2%, 精准度为 80.9%, 召回率为 84.7%, 特异度为 83.8%, 整体预测能力突出, 随机森林和轻量级梯度提升综合性能也较好, AUC 值为 88.9% 和 88.6%, 决策树和 Logistic 回归次之, AUC 为 75.5% 和 83.9%。5 模型的 ROC 曲线见图 1。极端梯度提升模型显示儿童第一恒磨牙龋病前 6 位影响因素分别为零食频率、甜点心糖果、甜饮料、刷牙后睡前零食、家长籍贯和母亲含糖饮食 (图 2)。

表 3 模型性能评价

模型	准确率	精准度	召回率	特异度	AUC
决策树	71.2	69.5	81.8	74.7	75.5
Logistic 回归	75.6	75.5	80.2	80.5	83.9
轻量级梯度提升	79.6	79.1	83.7	82.5	88.6
随机森林	80.9	80.5	84.5	83.5	88.9
极端梯度提升	81.2	80.9	84.7	83.8	90.1

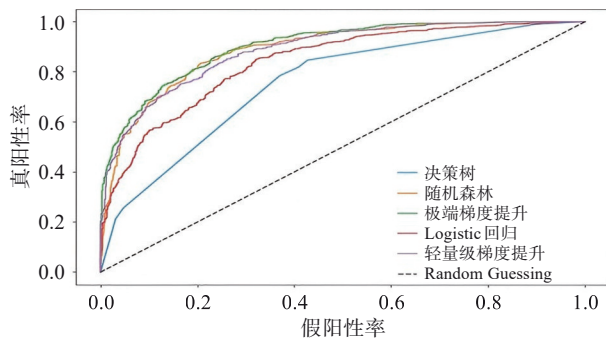


图 1 机器学习算法的 ROC 曲线

Fig 1 ROC curve of machine learning algorithm

2.4 SHAP 分析

通过对比独热编码后的变量, 高频甜食 (如: 甜点和糖果每天 ≥ 2 次, 母亲含糖饮食每天 ≥ 2 次) 与不良口腔习惯 (如刷牙后睡前经常吃零食, 不规律刷牙) 的 SHAP 值显著为正, 对模型输出

(是/否患龋结果) 有强正向影响, 是关键驱动因素; 而防护行为 (如规律漱口) 的 SHAP 值偏向负或 0, 提示其可降低患龋风险。“刷牙后睡前零食-经常” 的 SHAP 值显著为正, 正向提升风险; 而家长特征部分, “籍贯-苏州” 的 SHAP 值分布比较分散, 部分为正, “籍贯-迁入” 的 SHAP 值接近 0, 表明对模型输出有一定影响, 但需要结合其他变量综合判断; “父、母患龋-否” 的 SHAP 值部分为正, 而 “父、母患龋-是” 的 SHAP 值接近 0, 表明需结合其他变量综合判断对模型输出的影响。这些分析为揭示变量与模型的关联提供了直观的证据, 见图 3。

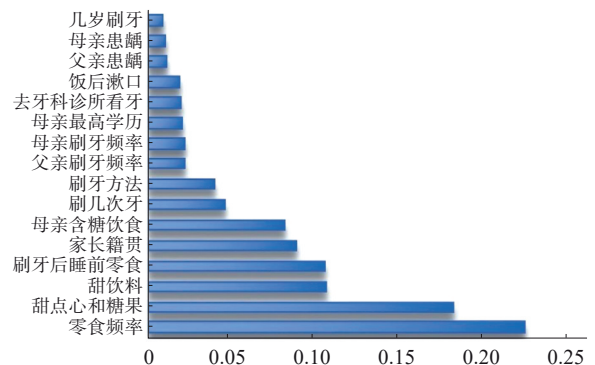


图 2 极端梯度提升模型中重要特征排名

Fig 2 Ranking of important feature in extreme gradient boosting models

3 讨论

龋病是细菌发酵膳食中碳水化合物产生酸性副产物对局部硬组织的慢性破坏^[13], 替牙初期的第一恒磨牙更易患龋, 至今处于上升趋势^[14]。地处饮食偏好甜及发酵类碳水化合物地区的儿童, 龋病发生风险不断增加^[15]; 且随着年龄的增加还将升高^[16]。当前, 已远高于《中国防治慢性病中长期规划 (2017—2025 年)》^[17]中提出 “12 岁儿童患龋率控制在 30% 以内” 的目标。随着龋病防治的推进, 龋病从广泛流行转变为在特定群体中流行, 75% 的龋风险负担倾向于集中在 25%~40%

的人口^[18]。综上所述，龋病在儿童群体中呈现特定流行态势，作为具有突出易患特点的疾病，通过龋病风险评估获得易患信息，是现代龋病防治的基础。采集、分析风险信息，预测风险人群，特别是高风险人群，针对性实施无创、微创、有创的递进式龋防治技术已成为临床治疗的趋势^[19]。同时，通过个体风险识别形成个性化干预，可达成医疗资源与经济效益的合理分配^[20]。

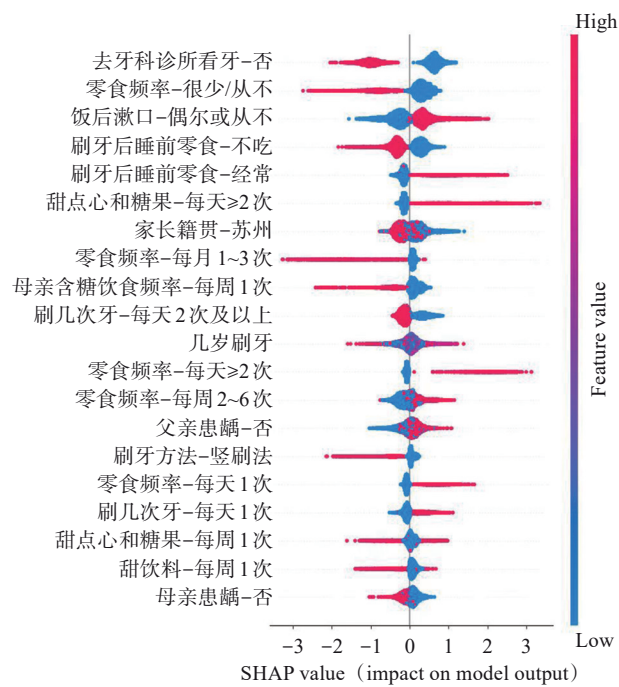


图3 SHAP值

Fig 3 SHAP value

研究借助机器学习构建苏州地区9岁儿童第一恒磨牙龋齿预测模型，筛选出极端梯度提升模型成为最优模型。该模型显示零食频率、甜点和糖果、甜饮料、刷牙后睡前零食、家长籍贯和母亲含糖饮食频率是该地区儿童第一恒磨牙龋病的前6位影响因素，与严鑫森等^[21]研究基本一致。研究^[22-23]表明，频繁食用易酵解含糖食物与饮料的儿童更易患龋；马婷等^[24]研究也表明，刷牙后睡前进食是儿童患龋的独立因素。与以往研究不同的是，苏州籍家庭的儿童龋齿发生率高于迁入籍家庭，成为重要的影响因素，符合其他研究^[25-26]基于不同文化背景下饮食的差异，导致微生物结构影响龋病差异化的报道。其次，母亲实际含糖饮食行为的负向影响超越母亲受教育水平的正向影响^[27-28]。表明在高等教育普及的经济发达地区，母亲受教育水平差距不断缩小，在口腔健康认知水平基本持平的情况下，母亲实际含糖饮食行为则成为影响儿童龋病更为突出的因素。究其原因，

可能是在糖文化饮食的大环境下，家长对致龋性饮食高度包容并接受。此外，睡前进食、频繁食用高糖食物、口腔健康教育及氟化物暴露不足等也都是致龋的关键因素^[29]。

机器学习是一种利用数据训练模型进行高维数据拟合，强调准确性的技术。相较传统的统计分析方法，机器学习擅长对非结构化复杂数据降维、提取特征及模式识别，能够适应动态变化并持续自我优化，挖掘人为设定规则以外的隐藏规律，让预测更具精准性。本研究通过对多种模型开展数据训练，发现极端梯度提升模型在准确性、精准度、灵敏度、特异度、AUC均表现最优，该模型不仅具备可解释性，还能学习复杂模式与关系，降低预测误差，成为表现最为突出的预测模型。引入SHAP分析对预测模型验证，以明确模型的合理性，揭示龋病发生的关键风险因素及其交互作用，清晰地解释模型在不同数据点的决策依据，为临床决策与预防策略制定提供支撑，推动理论研究向临床应用转化。

本研究基于极端梯度提升构建苏州地区9岁儿童第一恒磨牙龋病预测模型，用于龋病风险预测。创新点在于：构建融合特殊地区饮食文化背景、人口学特征、饮食偏好、口腔护理行为等因素的机器学习预测模型，实现对大数据样本精准且稳健地预测。通过寻找高危风险因素，识别高危风险儿童，实现个体化精准干预。然而，本研究属单中心研究，尽管样本虽然具备了风俗文化的典型代表性，但仍存在以偏概全的可能，需要更大范围内的数据样本及更具前瞻性的研究予以验证。本课题将持续采集儿童龋病风险相关数据，拓展研究视角，构建可与本地更大样本适配的龋病预测模型，推动早期识别与精准干预措施落地。

利益冲突声明：作者声明本文无利益冲突。

[参考文献]

[1] 马军,陶茂萱. 儿童青少年健康知识手册[M]. 北京: 北京大学医学出版社, 2018.
Ma J, Tao MX. Handbook of health knowledge for children and adolescents[M]. Beijing: Peking University Medical Press, 2018.

[2] 王兴. 第四次全国口腔健康流行病学调查报告[M]. 北京: 人民卫生出版社, 2018.
Wang X. Report on the fourth national epidemiological survey of oral health[M]. Beijing: People's Medical Pub-

- lishing House, 2018.
- [3] 刘亚琪, 张琼, 王艳, 等. 重度低龄儿童龋儿童全身麻醉与常规门诊下牙病治疗的疗效及卫生经济学评价[J]. 华西口腔医学杂志, 2021, 39(6): 703-708.
Liu YQ, Zhang Q, Wang Y, et al. Evaluation of therapeutic effect and health economic of general anesthesia and routine outpatient dental treatment in children with severe early child caries[J]. West China J Stomatol, 2021, 39(6): 703-708.
- [4] Selwitz RH, Ismail AI, Pitts NB. Dental caries[J]. Lancet, 2007, 369(9555): 51-59.
- [5] 张琼, 汪俊, 夏斌, 等. 低龄儿童龋的临床管理专家共识[J]. 华西口腔医学杂志, 2022, 40(5): 495-503.
Zhang Q, Wang J, Xia B, et al. Expert consensus on clinical management for early children caries[J]. West China J Stomatol, 2022, 40(5): 495-503.
- [6] 陈婷. 苏州糖文化研究[J]. 农业考古, 2015(3): 233-236.
Chen T. Study on Suzhou sugar culture[J]. Agric Res, 2015(3): 233-236.
- [7] Battineni G, Sagaro GG, Chinatalapudi N, et al. Applications of machine learning predictive models in the chronic disease diagnosis[J]. J Pers Med, 2020, 10(2): 21.
- [8] Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry—A systematic review[J]. J Dent Sci, 2021, 16(1): 508-522.
- [9] Sadegh-Zadeh SA, Rahmani Qeranqayeh A, Benkhalifa E, et al. Dental caries risk assessment in children 5 years old and under via machine learning[J]. Dent J (Basel), 2022, 10(9): 164.
- [10] Hänsel Petersson G, Fure S, Bratthall D. Evaluation of a computer-based caries risk assessment program in an elderly group of individuals[J]. Acta Odontol Scand, 2003, 61(3): 164-171.
- [11] World Health Organization. Oral health surveys: basic methods[M]. 5th ed. Geneva: World Health Organization, 2013.
- [12] 中华人民共和国国家卫生和计划生育委员会. 口腔健康调查检查方法: WS/T472-2015[S]. 北京: 中国标准出版社, 2015.
National Health and Family Planning Commission of the People's Republic of China. Oral health survey-Examination methods: WS/T 472-2015[S]. Beijing: Standards Press of China, 2015.
- [13] Zeng Y, Fadaak A, Alomeir N, et al. *Lactobacillus plan-*
tarum disrupts *S. mutans-C. albicans* cross-kingdom biofilms[J]. Front Cell Infect Microbiol, 2022, 12: 872012.
- [14] Hu Z, Yan XJ, Song Y, et al. Trends of dental caries in permanent teeth among 12-year-old Chinese children: evidence from five consecutive national surveys between 1995 and 2014[J]. BMC Oral Health, 2021, 21(1): 467.
- [15] Zhu F, Chen Y, Yu Y, et al. Caries prevalence of the first permanent molars in 6-8 years old children[J]. PLoS One, 2021, 16(1): e0245345.
- [16] Uribe SE, Innes N, Maldupa I. The global prevalence of early childhood caries: a systematic review with meta-analysis using the WHO diagnostic criteria[J]. Int J Paediatr Dent, 2021, 31(6): 817-830.
- [17] 国务院办公厅. 国务院办公厅关于印发中国防治慢性病中长期规划(2017—2025年)的通知[R/OL]. 2017-02-14[2025-04-10]. https://www.gov.cn/zhengce/content/2017-02/14/content_5167886.htm.
General Office of the State Council. Notice of the general office of the state council on printing and distributing the long-term plan for the prevention and treatment of chronic diseases in China[R/OL]. 2017-02-14[2025-04-10]. https://www.gov.cn/zhengce/content/2017-02/14/content_5167886.htm.
- [18] Macek MD, Heller KE, Selwitz RH, et al. Is 75 percent of dental caries really found in 25 percent of the population[J]. J Public Health Dent, 2004, 64(1): 20-25.
- [19] 周学东, 凌均荣, 梁景平, 等. 龋病临床治疗难度因素及处理[J]. 华西口腔医学杂志, 2017, 35(1): 1-7.
Zhou XD, Ling JQ, Liang JP, et al. Difficulty influence factors of dental caries clinical treatment[J]. West China J Stomatol, 2017, 35(1): 1-7.
- [20] Chaffee BW, Cheng J, Featherstone JD. Baseline caries risk assessment as a predictor of caries incidence[J]. J Dent, 2015, 43(5): 518-524.
- [21] 严鑫森, 孙桃兰, 卢雨航, 等. 基于机器学习的四川省12岁儿童龋齿预测模型[J]. 华西口腔医学杂志, 2023, 41(6): 686-693.
Yan XM, Sun TL, Lu YH, et al. Prediction model of dental caries in 12-year-old children in Sichuan Province based on machine learning[J]. West China J Stomatol, 2023, 41(6): 686-693.
- [22] Hong J, Whelton H, Douglas G, et al. Consumption frequency of added sugars and UK children's dental caries[J]. Community Dent Oral Epidemiol, 2018, 46(5): 457-

- 464.
- [23] Zeng L, Zeng Y, Zhou Y, et al. Diet and lifestyle habits associated with caries in deciduous teeth among 3- to 5-year-old preschool children in Jiangxi province, China [J]. BMC Oral Health, 2018, 18(1): 224.
- [24] 马婷, 孟凡琦, 韩婧, 等. 乌鲁木齐市3~5岁儿童龋病相关因素的主要成分分析[J]. 口腔疾病防治, 2023, 31(3): 179-185.
- Ma T, Meng FQ, Han J, et al. Principal component analysis of dental caries related factors among 3-to 5-year-old children in Urumqi City[J]. J Prev Treat Stomatol Dis, 2023, 31(3): 179-185.
- [25] 孙雁斐, 卢洁, 杨加震, 等. 基于微生物组新颖指数构建龋病菌群诊断模型[J]. 华西口腔医学杂志, 2023, 41(2): 208-217.
- Sun YF, Lu J, Yang JZ, et al. Construction of a caries diagnosis model based on microbiome novelty score[J]. West China J Stomatol, 2023, 41(2): 208-217.
- [26] Hemadi AS, Huang R, Zhou Y, et al. Salivary proteins and microbiota as biomarkers for early childhood caries risk assessment[J]. Int J Oral Sci, 2017, 9(11): e1.
- [27] 秦丹, 姜浩丰, 沈露, 等. 重庆市10~12岁儿童第一恒磨牙患龋情况及相关因素分析[J]. 华西口腔医学杂志, 2019, 37(6): 608-614.
- Qin D, Jiang HF, Shen L, et al. Prevalence of dental caries and associated factors among 10-12-year-old students in Chongqing[J]. West China J Stomatol, 2019, 37(6): 608-614.
- [28] Chen C, Zhang F, Wang R. Dental caries experience and related risk indicators of 12-year-old students in Jilin, China[J]. Medicine (Baltimore), 2020, 99(28): e20988.
- [29] Thang Le VN, Kim JG, Yang YM, et al. Risk factors for early childhood caries: an umbrella review[J]. Pediatr Dent, 2021, 43(3): 176-194.

(本文编辑 杜冰)

《单颗后牙种植修复决策路径及操作实践》出版发行

书籍名称：单颗后牙种植修复决策路径及操作实践

主编：陈泽涛，黄宝鑫

出版社：人民卫生出版社

内容简介：本书基于口腔种植学的基础与临床研究，全面地阐述了口腔种植的基础理论与临床实践。在基础理论方面，重点强调了与种植临床密切相关的种植影像学、种植影像学、种植材料学、种植生物力学基础、种植药理学、种植美学研究成果与进展。在临床实践方面，详细阐述了口腔种植的临床程序，以临床诊疗过程为主线，囊括了口腔种植时机、口腔种植外科、软硬组织增量、种植负荷、种植修复、牙周与种植、数字化种植，并将其融会贯通。本书通过三条主线分为三个版块向读者全面展现了口腔临床诊疗思维与实践过程，通过口腔种植治疗基础理论为初学者建立贴合临床的系统而全面的认知，通过口腔种植临床决策程序过程对口腔种植治疗程序和技术进行详细论述，以单颗后牙种植为例展现了种植决策治疗所获得的理想功能和美学效果并详细叙述治疗过程。

