

机器学习在龋病预测模型中的应用

方伟 王炳蔚 张军 刘景

新疆医科大学第五附属医院口腔科 乌鲁木齐 830011

[摘要] 龋病是全球范围内最常见的口腔疾病之一，给个人健康和社会经济带来了严重影响。随着机器学习技术的快速发展，它在医疗领域中的应用也越来越受到关注，研究者们开始探索其在龋病预测中的应用。本文综述了机器学习在龋病预测模型中的应用现状，包括数据来源、特征选择、模型构建、性能评估及未来发展方向。通过对现有文献的分析，结果发现机器学习方法能够在龋病风险评估、个性化干预方案和公共健康政策制定中发挥重要作用。本综述旨在帮助读者了解机器学习在龋病预测中的潜力和挑战，促进该领域的进一步研究和应用。

[关键词] 龋病；机器学习；风险评估

[中图分类号] R781.1 **[文献标志码]** A **[doi]** 10.7518/gjkq.2026102



本文链接

OSID 码

Application of machine learning in caries prediction models

Fang Wei, Wang Bingwei, Zhang Jun, Liu Jing

Dept. of Stomatology, Fifth Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, China

Correspondence: Liu Jing, Email: 77088180@qq.com

[Abstract] Caries is one of the most common oral diseases worldwide, with serious health and socioeconomic implications. With the rapid development of machine learning technology, its application in the medical field has attracted increasing attention, and researchers have begun to explore its application in caries prediction. This study reviews the application status of machine learning in caries prediction models, including data sources, feature selection, model construction, performance evaluation, and future development directions. By analyzing the existing literature, we found that machine learning methods can play an important role in caries risk assessment, personalized intervention programs, and public health policy making. This review aims to help readers understand the potential and challenges of machine learning in caries prediction and promote further research and application in this field.

[Key words] dental caries; machine learning; risk assessment

龋病是由细菌感染引起的牙齿硬组织破坏，是一种慢性进行性疾病^[1-2]。根据世界卫生组织（World Health Organization, WHO）的统计，龋病的发病率在全球范围内高达60%~90%^[3]。龋病作为一种常见的口腔疾病，其早期预测和干预对于保护口腔健康具有重要意义。传统的龋病预测主要依赖于临床经验和统计方法，而这种方式存在着主观性，特别是在处理复杂的多变量数据时

这些方法存在局限性^[4-7]。近年来，机器学习的兴起为龋病预测提供了新的视角，可以通过分析大量的临床和人口统计数据，识别出潜在的风险因素，从而实现更为精准的预测^[8-12]。

1 机器学习基础

机器学习，作为人工智能领域的核心分支，专注于通过数据驱动的方式训练模型，赋能计算机从经验中汲取知识并做出精准预测^[13]。它主要划分为监督学习、无监督学习和半监督学习三大类别^[14]。

[收稿日期] 2025-01-08；**[修回日期]** 2025-06-12

[第一作者] 方伟，住院医师，硕士，Email: fangwei9911@126.com

[通信作者] 刘景，主任医师，硕士，Email: 77088180@qq.com

监督学习, 依赖于已标注的数据集进行模型训练, 通过输入与输出对的学习, 构建预测模型^[15-16]。常见算法如线性回归, 通过最小化误差寻找最佳拟合; 决策树, 递归选择特征以分类或回归^[17]; 支持向量机 (support vector machine, SVM), 寻求最优决策边界^[18]; 以及神经网络, 模拟人脑结构处理复杂数据^[19]。无监督学习则不依赖标注数据, 仅基于数据本身探索内在结构^[20]。聚类算法如K-Means, 将数据划分为不同簇; 降维技术如主成分分析方法 (principal component analysis, PCA), 简化数据维度以促进理解或加速计算^[21]。半监督学习融合两者, 利用少量标注数据与大量未标注数据, 提升学习效率与模型泛化能力^[15]。此方法在标注成本高昂时尤为有效, 通过自训练、协同训练等策略, 最大化利用有限资源, 推动模型性能迈向新高。

综上所述, 机器学习以其多样化的学习模式, 正深刻改变着数据处理与决策制定的方式, 引领人工智能迈向更加智能化的未来。

2 龋病相关数据类型

在龋病预测中, 数据的多样性和准确性对于构建有效的预测模型至关重要。以下是对临床数据、人口统计数据、生物标志物以及生活方式因素这4类数据的详细描述。

2.1 临床数据

临床数据是龋病预测中最为直接和关键的数据来源, 通常包括以下部分。

1) 牙齿检查结果: 如龋齿的数量、位置、严重程度等。这些数据可以通过口腔检查获得, 是评估龋病现状的基础。牙齿检查结果, 一般可包含视诊结果、光学照片、影像学资料等这几类资料。在用于龋病预测模型的构建时, 需将上述牙齿检查结果转化为以下几种记录结果: ①龋齿数量, 记录每颗龋齿的具体位置 (如右上第一磨牙); ②龋齿位置, 详细标注龋齿在牙面的具体位置 (如咬合面、邻面等); ③严重程度, 根据龋齿深度和范围, 分为浅龋、中龋、深龋, 并记录是否已影响牙髓; ④其他发现, 记录牙齿磨损、裂纹、填充物状况等。

2) 病史: 患者的口腔健康历史, 包括以往的龋病发生情况、治疗记录等。这些信息有助于了解患者的龋病风险趋势。口腔相关病史的收集内

容主要包括主诉、现病史、既往口腔病史、全身病史、家族史和生活习惯等方面。收集方法包括面对面访谈、问卷调查、病历查阅和家属补充, 确保信息全面准确。

3) 治疗记录: 包括患者接受的预防和治疗措施, 如涂氟、窝沟封闭、充填治疗等。这些记录可以反映患者的口腔健康维护状况。利用表格标准化记录相关内容, 涂氟相关资料数据: ①涂氟剂类型, 记录使用的氟化物类型 (如氟化钠凝胶、氟保护漆等); ②涂氟部位, 记录涂氟的牙齿位置 (如上颌前牙、下颌磨牙等); ③涂氟频率, 记录涂氟的间隔时间 (如每3个月、每6个月)。窝沟封闭相关资料数据: ①记录封闭的牙齿位置 (如右上第一磨牙、左下第二磨牙等); ②封闭剂类型, 记录使用的封闭剂材料 (如树脂基封闭剂、玻璃离子等)。充填治疗相关资料数据: ①记录充填的牙齿位置 (如左上第一前磨牙、右下第二磨牙等); ②充填材料, 记录使用的充填材料 (如复合树脂、玻璃离子、银汞合金等)。

将以上收集到的资料信息录入电子病历系统或相应调查表格中, 采用标准化格式记录, 以便后续风险评估。

2.2 人口统计数据

人口统计数据提供了患者的社会背景信息, 对于理解龋病在不同人群中的分布和影响因素具有重要意义。它通常包括以下部分。1) 年龄: 不同年龄段的人群在龋病发病率和易感性上存在差异。例如, 儿童和老年人是龋病的高发人群^[6]。2) 性别: 性别也可能影响龋病的发生风险。研究^[22]表明, 男性在某些年龄段可能具有较高的龋病风险。3) 社会经济状态: 包括教育程度、收入水平、职业等。这些因素与口腔健康意识和资源获取能力相关, 从而间接影响龋病的发生^[23]。

2.3 生物标志物

生物标志物是龋病预测中的新兴领域, 它们提供了关于口腔微生态和生理状态的信息。常见的生物标志物包括以下部分。1) 唾液中的细菌种类: 如变异链球菌 (*Streptococcus mutans*, *S. mutans*)、乳酸杆菌 (lactic acid bacteria, LAB) 等致龋菌的数量和比例。这些细菌与龋病的发生密切相关, 可以作为预测龋病风险的指标^[24-25]。2) 生化指标: 如唾液中的pH值、氟离子浓度等。这些指标反映了口腔环境的酸碱度和防龋能力, 对于评估龋病风险具有重要价值^[26]。此类收集与整理

主要依靠所测生物学标本所得出的具体数据进行整理,在整理此类数据时需要确保所有数据的单位一致。将所收集的数据及时录入电子表格或数据库。

2.4 生活方式因素

生活方式因素是影响龋病发生的重要外部条件。它们通常包括以下部分。1) 饮食习惯:如糖的摄入量、食物种类和频率等。高糖饮食和频繁进食是龋病的重要风险因素^[27-28]。2) 口腔卫生习惯:如刷牙频率、使用牙线等。良好的口腔卫生习惯可以降低龋病的发生风险^[29-30]。

综上所述,这4类数据在龋病预测中各自扮演着重要的角色。通过综合分析这些数据,可以构建出更加准确和全面的龋病预测模型,为口腔健康管理和预防策略的制定提供有力支持。

3 龋病预测模型中的机器学习应用

3.1 数据收集与预处理

机器学习模型需要大量的数据进行训练和测试。数据的质量直接影响模型的性能^[31]。在龋病预测中,数据的收集和预处理是非常关键的一步。数据的收集可以通过临床检查、问卷调查、影像扫描等方式进行。而数据的预处理则包括数据清洗、缺失值处理、异常值检测等步骤,以确保数据的准确性和完整性^[32]。

3.1.1 数据收集

在龋病预测的研究与实践中,数据收集是一个综合性的过程,它融合了多种方法和技术以确保预测模型的准确性和可靠性^[32]。

龋病的诊断和风险评估依赖于多维度数据的综合采集。临床检查是核心环节,医生通过专业工具记录每颗牙齿的龋洞数量、位置(牙冠、牙颈或牙根)及严重程度(浅龋至深龋),为初步诊断提供客观依据。同时,问卷调查补充患者的行为数据,包括高糖饮食频率、口腔卫生习惯(如刷牙和牙线使用情况),以分析潜在的龋病风险因素。此外,影像学检查(如X线和CT扫描)进一步提供牙齿内部结构的详细信息,X线用于早期龋损的检测和评估,而CT则通过三维成像辅助复杂病例的精准诊断和治疗规划。

龋病研究数据收集面临隐私顾虑导致信息失真、文化差异影响标准化、数据真实性受主观影响、工具缺陷降低质量以及样本偏差限制代表性

等主要难点,需要通过匿名收集并说明数据用途来减少隐私担忧,优化问卷设计并简化语言及分级敏感问题,采用社区和电子问卷等多渠道方式扩大样本覆盖,同时结合临床与问卷数据进行交叉验证以确保信息真实可靠,从而有效提升数据质量并支撑研究科学性。

综上所述,临床检查、问卷调查和影像扫描技术在龋病预测的数据收集中各自扮演着不可或缺的角色。它们相互补充,共同构成了一个全面、深入且多维度的数据集,为构建准确、可靠的龋病预测模型提供了坚实的基础。

3.1.2 数据预处理

在数据预处理的关键环节中,数据清洗与转换、标准化共同构成了确保数据质量和适应性的重要步骤,对于后续的龋病预测模型训练至关重要^[33]。

数据清洗是消除数据不准确性和不一致性的关键步骤,主要包括3个核心环节:首先需筛查并去除重复数据,避免其对模型训练造成干扰;其次要妥善处理缺失值,根据具体情况选择删除或填补等方法;最后需通过统计方法识别异常值,并采取相应处理措施,确保数据质量。

数据转换与标准化是数据预处理的另一重要组成部分。数据转换通过将分类变量数值化等方式,使数据更适配算法处理;标准化则通过归一化等方法消除特征间的量纲差异,防止某些特征对模型产生过度影响。

这些预处理步骤共同构成了高质量数据分析的基础,不仅显著提升龋病预测模型的性能,更为研究结果的可靠性提供了坚实保障。

3.2 特征选择与提取

在构建龋病预测模型的过程中,特征选择与提取是关乎模型性能优劣的关键步骤^[34]。这一过程旨在从众多可用的特征中,精心挑选出那些对龋病预测最具影响力的特征,从而有效降低数据维度,提升模型的预测精度和运算效率。特征选择常用的有相关性分析、PCA和信息增益。

相关性分析通过计算特征与目标变量(如龋病发生)的相关系数评估线性关系,方法简单直观,适用于线性模型的初步特征筛选。但其仅能识别线性关联,无法捕捉非线性关系或特征交互作用,可能导致复杂模型中遗漏重要特征^[35]。

PCA通过正交变换将高维数据降维,保留主要变异信息,适用于处理图像等高维数据,能提

升模型泛化能力并降低过拟合风险。但其主成分解释性较差,难以对应原始特征,且对非线性关系处理有限^[36]。

信息增益基于信息论评估特征对目标变量的贡献,擅长处理分类问题和非线性关系,能提升分类模型准确性。但对连续特征效果有限,且易偏向多取值特征^[37]。

特征选择的目标非常明确,那就是从海量的特征中筛选出关键特征变量,这些精选的特征应该具备高度的预测能力,同时又能保持模型的简洁性^[38-39]。综合来看,相关性分析适合线性模型的快速筛选,PCA适合高维数据的降维和去噪,而信息增益则更适合分类问题中的非线性特征选择。根据数据特性和模型需求选择合适的方法,可以在提升模型性能的同时,平衡计算效率、可解释性和泛化能力。

3.3 模型训练与评估

3.3.1 模型的选择

在完成特征选择这一关键步骤后,接下来的重要任务便是挑选恰当的机器学习算法,以进行模型的构建、训练与评估^[40]。这一过程对于龋病预测模型的最终成效具有决定性意义。在众多的机器学习算法中,SVM、随机森林(random forest, RF)以及深度学习(deep learning, DL)等都是备受青睐的选择。

SVM基于统计学习理论,通过最大化间隔寻找最优分类超平面,在高维数据(如基因或临床指标)中表现优异,尤其适合类别界限清晰的场景^[41]。但其计算复杂度较高,对噪声和不平衡数据敏感,且训练速度较慢^[42]。决策树通过递归分割构建树形结构,模型直观但容易过拟合^[43]。RF通过集成多棵决策树提升泛化能力,对噪声数据鲁棒性强,擅长处理混合类型特征(如临床、问卷数据),但在高维稀疏数据(如影像数据)上表现有限^[44]。DL通过神经网络自动提取高层特征,在处理大规模图像、序列等复杂数据时表现突出。其优势在于强大的非线性建模能力,但依赖海量数据和计算资源,且需要精细调参^[45]。

除上述几种主要的机器学习方法外,还存在其他一些方法。如线性回归通过最小化误差平方和拟合直线,适合线性关系但对异常值敏感。逻辑回归利用Sigmoid函数处理分类问题,简单易懂但仅限于线性可分数据^[46]。K近邻(K-nearest neighbors, KNN)基于邻近样本投票,简单但对

高维数据效果差^[47]。梯度提升机(gradient boosting machine, GBM)逐步优化残差,精度高但易过拟合^[48]。时间序列模型捕捉时间依赖性,适合趋势预测但对非线性处理有限。贝叶斯方法^[49]基于概率推理,适合小数据但对先验敏感。选择模型时需结合数据特性、问题类型和计算资源。

因此,在机器学习实践中,模型的选择应基于数据特性、问题类型和计算资源的综合考量。针对不同场景,如数据量、维度、非线性关系、时间依赖性等,需选择适当的模型训练方法,以确保模型的性能、解释性和计算效率达到最优平衡。对于二分类或多分类任务(如预测是否患龋病或龋病严重程度),SVM和RF是首选;而对于特征提取或复杂模式识别任务(如从影像数据中提取龋病特征),DL更为合适。

3.3.2 模型的训练

模型训练是构建龋病预测模型的核心环节,这一过程精细而系统,旨在通过一系列步骤打造出性能优异的预测工具^[50]。首先,数据划分作为训练的先决条件,其重要性不言而喻。通常,会将整个数据集精心分割为3个部分:训练集、验证集和测试集。这三者各司其职,共同支撑着模型的训练与优化。

在龋病预测模型的开发过程中,训练集、验证集和测试集发挥着关键作用。训练集作为模型的学习资料库,包含了丰富的患者数据(如口腔检查记录、饮食习惯、氟化物使用情况等临床指标),让模型能够深入挖掘龋病发生的潜在规律。验证集则像一位严格的教练,通过评估模型在不同参数下的预测准确性(如对患龋风险的判断能力),指导研究人员选择最优的模型配置。最终的测试集则模拟真实临床场景,用完全独立的患者数据(包括各类口腔检查指标和实际龋病发生情况)来客观验证模型的泛化性能,确保其在实际医疗环境中能够可靠地预测龋病风险。这种三阶段评估体系既保证了模型的学习效果,又避免了过拟合,使预测结果更具临床参考价值。

由于龋病样本的获取存在一定的难度,样本量往往有限,这会对模型的性能造成限制。在样本量有限的情况下,选择合适的验证方法对模型训练至关重要。K折交叉验证^[51](如5折或10折)是最常用的方法,能充分利用数据并提高评估稳定性;若样本极少,可采用留一法交叉验证^[52]或自助法^[53]。对于分类问题,分层K折交叉验证能保

持类别平衡;时间序列数据则适合时间序列交叉验证,避免未来信息泄露。此外,结合数据增强^[54-55](如图像旋转、文本替换)、迁移学习^[56](利用预训练模型)和正则化^[57](如L1/L2正则化),可进一步提升模型泛化能力,减少过拟合风险。通过这些方法,可以在有限数据下实现更可靠的模型训练与评估。

综上所述,数据划分、模型训练和模型调优是模型训练过程中不可或缺的3个步骤。它们相互协作,共同推动着模型从初步构建到性能优化的全过程。在龋病预测模型的构建中,这些步骤的严格执行将为研究者打造出一个既准确又可靠的预测工具。

3.3.3 模型的评估

在构建和优化龋病预测模型的过程中,模型的评估是一个至关重要的环节^[58]。为了全面且准确地衡量模型的性能,需要借助一系列评估指标。这些指标不仅能够帮助了解模型在预测龋病方面的准确性,还能揭示模型在处理不同类别数据时的表现差异。

首先,准确率是一个直观且常用的评估指标。它表示模型正确预测样本占总样本的比例,是衡量模型整体预测能力的基础指标。然而,在龋病预测等不平衡分类问题中,单纯依赖准确率可能无法全面反映模型性能,因为模型可能在多数类上表现良好,而在少数类上表现不佳。

为了更细致地评估模型在不同类别上的表现,引入了召回率和F1值。召回率,又称灵敏度或真阳性率,反映了模型能够正确识别出的龋病患者占有所有实际龋病患者的比例。这一指标对于龋病预测尤为重要,因为希望模型能够尽可能多地识别出真正的龋病患者,以便及时采取干预措施。而F1值则是准确率和召回率的调和平均,它综合考虑了模型在两类上的表现,提供了一个更为全面的性能评估。

除了上述指标外,受试者工作特征曲线(receiver operating characteristic curve, ROC)和曲线下面积(area under the curve, AUC)也是评估模型分类性能的重要工具。ROC曲线通过绘制真阳性率(召回率)与假阳性率之间的关系图,直观地展示了模型在不同阈值下的分类能力。一个优秀的模型应该能够在保持较低假阳性率的同时,实现较高的真阳性率。AUC量化了模型的整体分类性能。AUC值越接近1,说明模型的分类性能越

好;反之,若AUC值接近0.5,则模型的分类性能较差。

在使用这些指标进行评估模型性能时,需要注意该预测模型所使用数据分布是否平衡,当预测模型数据分布不平衡时,准确率可能无法反映模型的实际表现。对于龋病预测模型而言,正类识别至关重要,应优先关注召回率,确保模型能有效识别正类。F1分数是精确率和召回率的调和平均,适合在不平衡数据中评估模型。当需要平衡精确率和召回率时,F1分数是一个有效的综合指标。AUC适合评估模型在不平衡数据中的整体表现,尤其是龋病预测模型而言这种注重正负类识别能力上的模型。对于不平衡的数据,应当对其进行适合的方法进行数据的调整,同时选择适合的评价指标进行评价。

综上所述,通过综合运用准确率、召回率、F1值以及ROC曲线和AUC等指标,可以全面且准确地评估龋病预测模型的性能。这些指标不仅帮助研究者了解模型在整体和各类别上的表现,还为其提供了优化模型的方向和依据。

4 研究进展与挑战

目前,机器学习在龋病预测中已经取得了一些进展。一些研究表明,机器学习模型在龋病预测中具有较高的准确性和敏感性。例如,在严鑫淼等^[59]基于四川省4 439名12岁儿童的横断面调查数据中,对比了随机森林、决策树、XGBoost和Logistic回归4种模型的预测效能。研究发现,随机森林模型表现最优(AUC=0.840),其灵敏度和特异度(0.854和0.706)均优于传统Logistic回归(AUC=0.799),相较于传统方法,机器学习模型通过特征重要性排序(如“刷牙后睡前进食”贡献度最高)进一步明确了关键风险因素,为精准预防提供了数据驱动依据。Yang等^[60]的研究表明,通过多元线性回归和4种机器学习算法预测12岁儿童的龋失补指数及龋齿风险组,结果显示:其中一种机器学习方法预测准确率达43.27%,远高于多元线性回归的15.24%;另一种机器学习方法预测准确率为后者的2.84倍。Ogwo等^[61]利用机器学习技术对年轻成年人的龋齿结果进行预测,并识别关键预测因素。通过应用随机森林、支持向量机等算法,机器学习模型展现出比传统统计方法更高的预测精度。Qu等^[62]开发了一种基于非生物

学因素的幼儿龋齿风险预测模型,采用机器学习算法对行为因素进行分析。研究纳入了481名12至60个月无龋齿史的儿童,结果显示:随机森林模型的AUC显著优于传统统计学方法。以上研究表明机器学习在处理复杂非线性数据时具有更高的预测准确性和鲁棒性,为社区筛查提供了便捷工具,并有助于改善家庭口腔护理水平。

尽管机器学习在龋病预测领域取得了显著进展,但仍面临若干关键挑战。首先,数据不平衡问题突出,即健康牙齿样本(负例)远多于龋齿样本(正例),正例(龋病患率)通常仅占10%~20%,导致模型偏向预测健康牙齿而降低龋病识别灵敏度,现有方法如合成少数类过采样技术(synthetic minority over-sampling technique, SMO-TE)虽可部分缓解但可能引发过拟合^[63]。其次,高质量样本获取困难,受限于专业标注成本和隐私问题,现有数据集多不足千例且多为单中心数据,而深度学习模型通常需要数万甚至百万级样本才能达到理想性能。小样本训练易导致模型泛化能力不足且难以覆盖不同人群的龋病特征差异。更关键的是,深度学习模型的黑箱特性阻碍临床落地。例如,深度学习模型[如卷积神经网络(convolutional neural network, CNN)]虽能自动提取特征并实现高准确率,但其决策过程难以直观理解^[64]。

5 结论

机器学习在龋病预测模型中具有广阔的应用前景。通过对数据的收集与预处理、特征选择与提取、模型训练与评估等步骤的综述,可以看到机器学习在龋病预测中的潜力和挑战。未来的研究应注重解决数据不平衡、样本量不足等问题,并探索更加高效和准确的机器学习算法和模型,以提高龋病预测的准确性和可靠性。此外,还需加强对模型的解释性研究,以增强对模型预测结果的理解和信任。机器学习在龋病预测中的应用有望为口腔健康的改善和龋病的预防提供有效的工具和方法。

利益冲突声明:作者声明本文无利益冲突。

6 参考文献

- [1] Selwitz RH, Ismail AI, Pitts NB. Dental caries[J]. *Lancet*, 2007, 369(9555): 51-59.
- [2] 高岩. 口腔组织病理学[M]. 8版. 北京:人民卫生出版社, 2020: 126-127.
- [3] Gao Y. Oral histopathology[M]. 8th ed. Beijing: People's Medical Publishing House, 2020: 126-127.
- [4] Maklennan A, Borg-Bartolo R, Wierichs RJ, et al. A systematic review and meta-analysis on early-childhood-carries global data[J]. *BMC Oral Health*, 2024, 24(1): 835.
- [5] Al-Khalifa KS, Ahmed WM, Azhari AA, et al. The use of artificial intelligence in caries detection: a review[J]. *Bioengineering (Basel)*, 2024, 11(9): 936.
- [6] Dey P, Ogwo C, Tellez M. Comparison of traditional regression modeling vs. AI modeling for the prediction of dental caries: a secondary data analysis[J]. *Front Oral Health*, 2024, 5: 1322733.
- [7] 黄港, 曹桂莹, 刘民. 1990—2019年中国龋病的疾病负担现状及变化趋势研究[J]. *中国全科医学*, 2024, 27(14): 1735-1741.
- [8] Huang G, Cao GY, Liu M. Current situation and trends in the disease burden of dental caries in China, 1990–2019[J]. *Chin Gen Pract*, 2024, 27(14): 1735-1741.
- [9] 郑欣, 程磊, 周学东. 龋病研究的前沿与进展[J]. *中华口腔医学杂志*, 2024, 59(1): 14-22.
- [10] Zheng X, Cheng L, Zhou XD. Frontiers and advances in dental caries research[J]. *Chin J Stomatol*, 2024, 59(1): 14-22.
- [11] Reyes LT, Knorst JK, Ortiz FR, et al. Machine learning in the diagnosis and prognostic prediction of dental caries: a systematic review[J]. *Caries Res*, 2022, 56(3): 161-170.
- [12] Çiftçi BT, Aşantoğrol F. Utilization of machine learning models in predicting caries risk groups and oral health-related risk factors in adults[J]. *BMC Oral Health*, 2024, 24(1): 430.
- [13] Sadeh-Zadeh SA, Bagheri M, Saadat M. Decoding children dental health risks: a machine learning approach to identifying key influencing factors[J]. *Front Artif Intell*, 2024, 7: 1392597.
- [14] Kang IA, Ngnamsie Njimbouom S, Lee KO, et al. DCP: prediction of dental caries using machine learning in personalized medicine[J]. *Appl Sci*, 2022, 12(6): 3043.

- [12] 刘红艳, 韦曦, 凌均荣. 人工智能在龋病诊疗中的应用与探索[J]. 中华口腔医学杂志, 2024, 59(1): 37-44.
- Liu HY, Wei X, Ling JQ. Application and exploration of artificial intelligence for caries management [J]. Chin J Stomatol, 2024, 59(1): 37-44.
- [13] Didier AJ, Nigro A, Noori Z, et al. Application of machine learning for lung cancer survival prognostication—a systematic review and meta-analysis[J]. Front Artif Intell, 2024, 7: 1365777.
- [14] Nwanosike EM, Conway BR, Merchant HA, et al. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review[J]. Int J Med Inform, 2022, 159: 104679.
- [15] 李小娟, 韩萌, 王乐, 等. 监督与半监督学习下的数据流集成分类综述[J]. 计算机应用研究, 2021, 38(7): 1921-1929.
- Li XJ, Han M, Wang L, et al. Overview of data stream ensemble classification based on supervised and semi-supervised learning[J]. Appl Res Comput, 2021, 38(7): 1921-1929.
- [16] Zaadnoordijk L, Besold TR, Cusack R. Lessons from infant learning for unsupervised machine learning[J]. Nat Mach Intell, 2022, 4(6): 510-520.
- [17] Bala B, Behal S. AI techniques for IoT-based DDoS attack detection: taxonomies, comprehensive review and research challenges[J]. Comput Sci Rev, 2024, 52: 100631.
- [18] Ancajas CMF, Oyedele AS, Butt CM, et al. Advances, opportunities, and challenges in methods for interrogating the structure activity relationships of natural products[J]. Nat Prod Rep, 2024, 41(10): 1543-1578.
- [19] Binetti MS, Massarelli C, Uricchio VF. Machine learning in geosciences: a review of complex environmental monitoring applications[J]. Mach Learn Knowl Extr, 2024, 6(2): 1263-1280.
- [20] 殷瑞刚, 魏帅, 李晗, 等. 深度学习中的无监督学习方法综述[J]. 计算机系统应用, 2016, 25(8): 1-7.
- Yin RG, Wei S, Li H, et al. Introduction of unsupervised learning methods in deep learning[J]. Comput Sys Appl, 2016, 25(8): 1-7.
- [21] Carreras J, Yukie Kikuti Y, Miyaoka M, et al. Artificial intelligence analysis and reverse engineering of molecular subtypes of diffuse large B-cell lymphoma using gene expression data[J]. BioMedInformatics, 2024, 4(1): 295-320.
- [22] Ferraro M, Vieira AR. Explaining gender differences in caries: a multifactorial approach to a multifactorial disease[J]. Int J Dent, 2010, 2010: 649643.
- [23] 邹聚芬, 毕小琴. 成都市社区成年居民口腔健康现状调查及影响因素分析[J]. 中华现代护理杂志, 2022, 28(2): 177-183.
- Zou JF, Bi XQ. Oral health status and influencing factors of adult residents in Chengdu communities [J]. Chin J Mod Nurs, 2022, 28(2): 177-183.
- [24] Li YX, Xiang YG, Ren HX, et al. Association between periodontitis and dental caries: a systematic review and meta-analysis[J]. Clin Oral Investig, 2024, 28(6): 306.
- [25] Ng TC, Luo BW, Lam WY, et al. Updates on caries risk assessment—a literature review[J]. Dent J (Basel), 2024, 12(10): 312.
- [26] 楼云飞, 俞长浩, 戴金保. 唾液 pH 值与年龄及口腔环境的多因素关系[J]. 中华口腔医学杂志, 1986, 21(3): 189-189.
- Lou YF, Yu CH, Dai JB. The multifactorial relationship between salivary pH value, age, and oral environment[J]. Chin J Stomatol, 1986, 21(3): 189-189.
- [27] Johansson I, Holgersson PL, Kressin NR, et al. Snacking habits and caries in young children[J]. Caries Res, 2010, 44(5): 421-430.
- [28] Mobley C, Marshall TA, Milgrom P, et al. The contribution of dietary factors to dental caries and disparities in caries[J]. Acad Pediatr, 2009, 9(6): 410-414.
- [29] Kumar S, Tadakamadla J, Johnson NW. Effect of toothbrushing frequency on incidence and increment of dental caries: a systematic review and meta-analysis[J]. J Dent Res, 2016, 95(11): 1230-1236.
- [30] de Oliveira KMH, Nemezio MA, Romualdo PC, et al. Dental flossing and proximal caries in the primary dentition: a systematic review[J]. Oral Health Prev Dent, 2017, 15(5): 427-434.
- [31] Fenza G, Gallo M, Loia V, et al. Data set quality in machine learning: consistency measure based on group decision making[J]. Appl Soft Comput, 2021,

- 106: 107366.
- [32] 郝爽, 李国良, 冯建华, 等. 结构化数据清洗技术综述[J]. 清华大学学报(自然科学版), 2018, 58(12): 1037-1050.
- Hao S, Li GL, Feng JH, et al. Survey of structured data cleaning methods[J]. *J Tsinghua Univ (Sci Technol)*, 2018, 58(12): 1037-1050.
- [33] 卢洁. 基于微生物组大数据的龋病菌群分析及诊断模型构建[D]. 辽宁: 大连医科大学, 2021.
- Lu J. Caries microbiota analysis and diagnostic model construction based on microbiome big data [D]. Liaoning: Dalian Medical University, 2021.
- [34] 胡智翔. 基于机器学习的慢性病预测关键技术研究[D]. 成都: 电子科技大学, 2022.
- Hu ZX. Research on key technologies of chronic disease prediction based on machine learning[D]. Chengdu: University of Electronic Science and Technology of China, 2022.
- [35] Ebrahimi Warkiani M, Moattar MH. A comprehensive survey on recent feature selection methods for mixed data: challenges, solutions and future directions[J]. *Neurocomputing*, 2025, 623: 129372.
- [36] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments[J]. *Philos Trans Roy Soc A Math Phys Eng Sci*, 2016, 374(2065): 20150202.
- [37] Lei S. A feature selection method based on information gain and genetic algorithm[C]//2012 international conference on computer science and electronics engineering. IEEE, 2012: 355-358.
- [38] Chandrashekar G, Sahin F. A survey on feature selection methods[J]. *Comput Electr Eng*, 2014, 40(1): 16-28.
- [39] Li JD, Cheng KW, Wang SH, et al. Feature selection [J]. *ACM Comput Surv*, 2018, 50(6): 1-45.
- [40] Cai J, Luo JW, Wang SL, et al. Feature selection in machine learning: a new perspective[J]. *Neurocomputing*, 2018, 300: 70-79.
- [41] Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: a review[J]. *ISPRS J Photogramm Rem Sens*, 2011, 66(3): 247-259.
- [42] Chang CC, Lin CJ. LIBSVM: a library for support vector machines[J]. *ACM Trans Int Sys Technol*, 2011, 2(3): 27.
- [43] Loh WY. Classification and regression trees[J]. *WIRES Data Min & Knowl*, 2011, 1(1): 14-23.
- [44] Quist J, Taylor L, Staaf J, et al. Random forest modelling of high-dimensional mixed-type data for breast cancer classification[J]. *Cancers (Basel)*, 2021, 13(5): 991.
- [45] Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview[M]. Singapore: Springer Singapore, 2020: 599-608.
- [46] Shi Y, Zhu Q, Guo F, et al. Easing embedding learning by comprehensive transcription of heterogeneous information networks[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, London United Kingdom. ACM, 2018: 2190-2199.
- [47] Zhang SC, Li XL, Zong M, et al. Learning k for kNN classification[J]. *ACM Trans Intell Syst Technol*, 2017, 8(3): 1-19.
- [48] Ke GL, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//31st conference on neural information processing systems. NIPS, 2017: 30.
- [49] Yang SP, Berdine G. Bayesian data analysis[J]. *Southwest Resp Crit Care Chron*, 2020, 8(36): 74-77.
- [50] Mavrogiorgou A, Kiourtis A, Kleftakis S, et al. A catalogue of machine learning algorithms for health-care risk predictions[J]. *Sensors (Basel)*, 2022, 22(22): 8615.
- [51] Berrar D. Cross-validation[M]. 2nd ed. Amsterdam: Elsevier, 2025: 638-644.
- [52] Stone M. Cross-validated choice and assessment of statistical predictions[J]. *J Roy Stat Soc*, 1974, 36(2): 111-133.
- [53] Seiffert C, Khoshgoftaar TM, Van Hulse J, et al. RUSBoost: a hybrid approach to alleviating class imbalance[J]. *IEEE Trans Syst Man Cybern Part A Syst Hum*, 2010, 40(1): 185-197.
- [54] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning[J]. *J Big Data*, 2019, 6(1): 60.
- [55] Wei J, Zou K. EDA: easy data augmentation techniques for boosting performance on text classification tasks[J]. *arXiv*, 2019: arXiv:1901.11196.

- [56] Ribani R, Marengoni M. A survey of transfer learning for convolutional neural networks[C]//2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T). IEEE, 2019: 47-57.
- [57] Zou H, Hastie T. Addendum: regularization and variable selection via the elastic net[J]. J Roy Stat Soc, 2005, 67(5): 768.
- [58] Binuya ME, Engelhardt EG, Schats W, et al. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review[J]. BMC Med Res Methodol, 2022, 22(1): 316.
- [59] 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.
- [60] Yang YH, Kim JS, Jeong SH. Prediction of dental caries in 12-year-old children using machine-learning algorithms[J]. J Korean Acad Oral Health, 2020, 44(1): 55-63.
- [61] Ogwo C, Brown G, Warren J, et al. Predicting dental caries outcomes in young adults using machine learning approach[J]. BMC Oral Health, 2024, 24(1): 529.
- [62] Qu X, Zhang C, Houser SH, et al. Prediction model for early childhood caries risk based on behavioral determinants using a machine learning algorithm[J]. Comput Methods Programs Biomed, 2022, 227: 107221.
- [63] Kumari M, Subbarao N. A hybrid resampling algorithms SMOTE and ENN based deep learning models for identification of Marburg virus inhibitors[J]. Future Med Chem, 2022, 14(10): 701-715.
- [64] Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models[J]. arXiv, 2017: arXiv:1708.08296.

(本文编辑 王姝)

《牙周松牙固定术操作详解》出版发行

书籍名称: 牙周松牙固定术操作详解

主编: 申玉芹

出版社: 人民卫生出版社

内容简介: 本书详细介绍了牙周松牙固定术的操作流程和注意事项, 包括患者的评估、器械的选择、操作步骤的规范操作、维护和康复指导等。针对牙周松牙固定术每一步的详细操作、所需工具、可能遇到的困难以及解决方案进行了详尽的介绍。结合大量的临床案例, 帮助读者更好地理解 and 掌握牙周松牙固定术的应用技巧。本书兼顾理论性和实用性, 既提供了系统的基础知识和实用的操作技巧, 有助于初学者快速掌握牙周松牙固定术的核心技能; 同时提供了深入的理论探讨和丰富的临床案例, 可帮助口腔医学专业人士更好地掌握牙周松牙固定术的操作技巧, 为患者提供更为精准、个性化的牙周治疗服务。

