

基于机器学习的药物相关性颌骨坏死发病风险和手术疗效的预测模型构建及验证^Δ

吴萌萌^{1*}, 茅昌飞², 张婧¹, 张源¹, 刘小林³, 浦迎秋^{1#} (1. 南京大学医学院附属口腔医院/南京市口腔医院/南京大学口腔医学研究所药学部, 南京 210008; 2. 江苏省肿瘤医院/南京医科大学附属肿瘤医院/江苏省肿瘤防治研究所乳腺外科, 南京 210009; 3. 江苏省肿瘤医院/南京医科大学附属肿瘤医院/江苏省肿瘤防治研究所药学部, 南京 210009)

中图分类号 R969.3; R782.4 文献标志码 A 文章编号 1001-0408(2026)16-2187-08
DOI 10.6039/j.issn.1001-0408.2026.16.19



摘要 **目的** 构建并验证预测药物相关性颌骨坏死(MRONJ)发病风险及手术疗效的可解释机器学习模型, 为MRONJ风险评估与治疗决策提供量化依据。**方法** 构建2个研究队列。其中, 队列A采用回顾性队列巢式病例对照研究设计, 用于MRONJ发病风险预测模型的构建; 队列B源自队列A中接受手术治疗并符合随访条件的MRONJ病例, 采用前瞻性随访队列研究设计, 用于MRONJ手术疗效预测模型的构建。经缺失值处理、变量标准化及编码后, 采用最小绝对收缩和选择算子(LASSO)回归筛选预测特征, 并通过方差膨胀因子(VIF)进行共线性诊断, 基于筛选后的特征分别构建随机森林(RF)、极端梯度提升(XGB)、多层感知机(MLP)、支持向量机(SVM)和高斯朴素贝叶斯(NB)模型。采用重复20次的双层嵌套交叉验证进行超参数优化和内部性能评价, 其中队列B在内层训练集中采用合成少数类过采样技术(SMOTE)处理类别不平衡。通过受试者操作特征(ROC)曲线、精确率-召回率(PR)曲线、校准曲线及决策曲线分析(DCA)综合评价模型性能, 并运用Shapley加性解释(SHAP)方法量化各临床特征对模型预测结果的相对贡献。**结果** 对于队列A, 共保留碱性磷酸酶、牙齿拔除等7个临床特征; 对于队列B, 共保留年龄、用药时间等8个临床特征。队列A中, RF模型的ROC曲线下面积、PR曲线下面积、Brier评分、敏感度、特异度分别为0.94、0.92、0.09、0.92、0.85; 队列B中, RF模型的上述指标分别为0.93、0.93、0.12、0.83、0.93; 校准曲线和DCA结果显示, RF模型具有较好的校准表现和临床净获益, 所有变量的VIF均小于5。SHAP可视化结果显示, 队列A中碱性磷酸酶、牙齿拔除和药物累积剂量对发病风险预测的贡献较大, 队列B中年龄、用药时间和MRONJ分期对手术疗效预测的贡献较大。**结论** 基于RF算法构建的MRONJ发病风险和手术疗效预测模型在内部验证中表现出较好的区分能力、概率预测准确性和临床应用潜力, 可为MRONJ全流程风险管理和个体化临床决策提供辅助支持, 但仍需在多中心、跨区域独立队列中开展外部验证。

关键词 药物相关性颌骨坏死; 发病风险; 手术疗效; 机器学习; 药物风险管理

Development and validation of machine learning-based predictive models for the risk of medication-related osteonecrosis of the jaw and surgical outcomes

WU Mengmeng¹, MAO Changfei², ZHANG Jing¹, ZHANG Yuan¹, LIU Xiaolin³, PU Yingqiu¹ (1. Dept. of Pharmacy, Nanjing Stomatological Hospital, Affiliated Hospital of Medical School, Institute of Stomatology, Nanjing University, Nanjing 210008, China; 2. Dept. of Breast Surgery, Jiangsu Cancer Hospital/The Affiliated Cancer Hospital of Nanjing Medical University/Jiangsu Institution of Cancer Research, Nanjing 210009, China; 3. Dept. of Pharmacy, Jiangsu Cancer Hospital/The Affiliated Cancer Hospital of Nanjing Medical University/Jiangsu Institution of Cancer Research, Nanjing 210009, China)

ABSTRACT **OBJECTIVE** To develop and validate interpretable machine learning models for predicting the risk of medication-related osteonecrosis of the jaw (MRONJ) and surgical outcomes, thereby providing quantitative support for MRONJ risk assessment and treatment decision-making. **METHODS** Two study cohorts were established. Cohort A adopted a nested case-control

study design to develop a predictive model for the risk of MRONJ onset. Cohort B comprised patients with MRONJ from Cohort A who underwent surgical treatment and met the follow-up eligibility criteria, and employed a prospective follow-up cohort study design to develop a predictive model for the efficacy of MRONJ surgery. After missing-data processing, variable standardization, and encoding, predictive

^Δ **基金项目** 国家自然科学基金项目(No.82204524); 江苏省医院协会医院药事管理研究专项课题(No.JSYGY-2-2024-YS45); 南京市口腔医院高水平医院建设科研项目(No.0224C056)

* **第一作者** 副主任药师。研究方向: 肿瘤药理、临床药学。
E-mail: wmm@njmu.edu.cn

通信作者 副主任药师。研究方向: 医院药学。E-mail: puyingqiu2026@163.com

features were selected using the least absolute shrinkage and selection operator (LASSO) regression, and collinearity was assessed using the variance inflation factor (VIF). Random Forest (RF), eXtreme Gradient Boosting (XGB), Multi-layer Perceptron (MLP), Support Vector Machine (SVM), and Gaussian Naive Bayes (NB) models were subsequently developed using the selected features. Hyperparameter optimization and internal performance evaluation were performed using two-level nested cross-validation repeated 20 times. In Cohort B, the synthetic minority over-sampling technique (SMOTE) was applied within the inner training folds to address class imbalance. Model performance was comprehensively evaluated using the receiver operating characteristic (ROC) curve, the precision-recall (PR) curve, calibration curve, and decision curve analysis (DCA). Shapley additive explanations (SHAP) were used to quantify the relative contributions of individual clinical features to model predictions.

RESULTS Seven clinical features, including alkaline phosphatase and tooth extraction, were retained for Cohort A; whereas eight clinical features, including age and duration of medication use, were retained for Cohort B. In Cohort A, the area under the ROC curve, the area under the PR curve, Brier score, sensitivity, and specificity of the RF model were 0.94, 0.92, 0.09, 0.92, and 0.85, respectively; the corresponding values in Cohort B were 0.93, 0.93, 0.12, 0.83, and 0.93, respectively. The results of calibration curves and DCA indicated that the RF models achieved favorable calibration and clinical net benefit, with VIF values for all variables below 5. SHAP visualization results showed that alkaline phosphatase, tooth extraction, and cumulative drug dose made substantial contributions to MRONJ risk prediction in Cohort A, whereas age, duration of medication use, and MRONJ stage contributed substantially to surgical outcome prediction in Cohort B.

CONCLUSIONS The RF-based models for predicting MRONJ risk and surgical outcomes demonstrated good discrimination, accuracy of probabilistic predictions, and potential clinical utility during internal validation. These models may provide decision support for comprehensive MRONJ risk management and individualized clinical decision-making. However, external validation in independent, multicenter, and geographically diverse cohorts is still required.

KEYWORDS medication-related osteonecrosis of the jaw; risk of onset; surgical outcomes; machine learning; drug risk management

2003年,Marx^[1]首次报道了双膦酸盐相关颌骨坏死病例,随后美国FDA于2005年在相关药品说明书中增加了警示。2007年,美国口腔颌面外科医师学会(American Association of Oral and Maxillofacial Surgeons, AAOMS)正式提出“双膦酸盐相关性颌骨坏死(bisphosphonate-related osteonecrosis of the jaw, BRONJ)”的定义^[2],并在后续更新中逐步完善诊断与管理。2014年,AAOMS将该疾病统一命名为“药物相关性颌骨坏死(medication-related osteonecrosis of the jaw, MRONJ)”^[3],并沿用至今。

MRONJ是一种严重的颌骨不良反应,由治疗恶性肿瘤骨转移或骨质疏松等疾病的抗骨吸收药物、抗血管生成类药物或激素类药物所致,以面部软组织瘘管经久不愈、骨外露和病理性骨折为主要表现,并伴有红肿热痛、咀嚼障碍,严重影响患者的生活质量^[4]。随着临床对MRONJ认识的不断深入,其预防策略与早期诊断已成为相关研究重点之一,但MRONJ发生发展受药物、局部创伤及患者全身状况等多种因素的共同影响^[5],导致肿瘤科及内分泌科医生在制订用药方案时,难以对个体患者的发病风险进行精准评估。相关指南指出,MRONJ早期应采取保守治疗的方式;对于病变持续或具有手术指征的患者,可考虑手术治疗^[6]。然而,手术疗效存在明显的个体差异,其疗效预测亦成为口腔颌面外科医生临床决策中的关键难点。近年来,机器学习算法在精准医疗领域中的应用不断拓展,在药源性疾病的风险预测和治疗决策等方面也具有强大潜能^[7-8]。考虑到临床须在用药前、用药中、围手术期、用药后及随访期进行MRONJ全周期风险评估,为应对上述挑战,本研究拟针对

对MRONJ的发病风险及手术疗效两个关键临床问题,开发机器学习预测模型并进行验证,以期为该疾病的临床风险评估与治疗决策提供量化依据。

1 资料与方法

1.1 研究设计与对象

队列A采用基于回顾性队列的巢式病例对照研究设计,纳入2015年1月—2022年8月于江苏省肿瘤医院就诊并首次使用MRONJ相关风险药物的患者,用于MRONJ发病风险预测模型的构建。该研究方案经该院伦理委员会批准(批准号2019-022),研究数据经脱敏处理后进行回顾性分析。该队列的纳入标准为:(1)接受唑来膦酸或地舒单抗等MRONJ相关风险药物治疗;(2)用药、口腔检查及主要临床资料完整。该队列的排除标准为:(1)既往有头颈部放射治疗史者;(2)存在颌骨原发肿瘤或恶性肿瘤颌骨转移者;(3)无法明确风险药物暴露时间或剂量者。在队列内构建病例组和对照组。其中,病例组为在江苏省肿瘤医院使用相关风险药物且经口腔专科检查确诊为MRONJ的患者;在病例组患者确诊MRONJ的当日,从同一源队列中采用危险密度抽样法随机选取处于风险状态但尚未发生MRONJ的患者,并按原发疾病进行匹配,构建与病例组样本量均衡的对照组。纳入患者的所有特征变量均为确诊日期前已发生或已记录的信息。前述风险药物包括注射用唑来膦酸浓溶液(规格5 mL:4 mg,国药准字HJ20150049,意大利Fisiopharma S.R.L.)、注射用唑来膦酸浓溶液[规格5 mL:4 mg(按C₅H₁₀N₂O₇P₂计),国药准字H20223183,华润双鹤药业股份有限公司]、地舒单抗注射液[规格120 mg(1.7 mL)/瓶,国药准字SJ20190025,美国Amgen

Manufacturing Limited LLC]。在发病风险预测模型的构建过程中,研究团队将发生MRONJ定义为阳性结局,未发生MRONJ定义为阴性结局。

队列B采用前瞻性随访队列研究设计,纳入2022年8月—2025年6月在南京大学医学院附属口腔医院接受手术治疗且来源于队列A的MRONJ患者,用于手术疗效预测模型的构建。该研究方案经南京大学医学院附属口腔医院伦理委员会批准(批准号KY-2022NL-046),所有纳入患者均签署知情同意书。该队列的纳入标准包括:(1)符合2022年AAOMS指南诊断标准的1~3期MRONJ^[4];(2)经多学科评估后确定具有手术指征,手术包括清创术和颌骨部分切除术;(3)术后随访时间至少6个月。该队列的排除标准包括:(1)既往有头颈部放射治疗史者;(2)存在恶性肿瘤转移至颌面头颈部者;(3)存在全身症状,不耐受手术治疗者;(4)合并血液系统疾病者。纳入患者均接受规范化随访(截至2025年12月),术后每6个月复查1次。根据《药物相关性颌骨坏死临床诊疗专家共识》疗效评价标准^[6],于术后6个月进行手术疗效评价,将完成全程随访者中“治愈”“改善”“稳定”的作为治疗有效组,“进展”的作为治疗无效组。在手术疗效预测模型的构建过程中,研究团队将治疗无效定义为阳性结局;治疗有效定义为阴性结局。

两个队列的研究流程见图1。

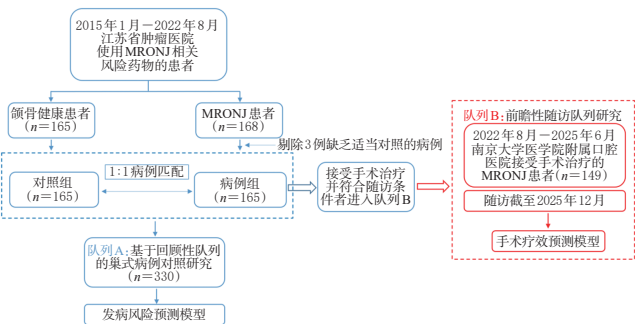


图1 两个队列的研究流程图

1.2 研究方法

1.2.1 数据采集与临床特征确定

研究数据来源于南京大学医学院附属口腔医院和江苏省肿瘤医院电子病历数据库。借助MS SQL Server和Oracle开发工具编写结构化查询语言,连接数据源,完成临床信息提取并构建临床数据库。根据既往文献及临床共识,采集与MRONJ发生及手术疗效有关的临床变量。最终纳入分析的临床特征涵盖全身因素、口腔相关性因素、药物相关因素、手术因素4个方面,具体变量构成见表1。

1.2.2 数据预处理与缺失值处理

保留缺失比例小于30%的变量,作为候选特征纳入分析。采用多重插补法对缺失数据进行处理:基于马尔科夫链蒙特卡洛方法生成5个完整数据集,并整合插补结果用于后续建模分析,以降低缺失数据引入的潜在偏倚^[9]。

表1 临床特征变量

分类	临床特征	变量类型	分类	临床特征	变量类型
全身因素	年龄	连续变量	口腔相关因素	牙齿拔除	二分类变量
	体重指数	连续变量		牙周病	二分类变量
	性别	二分类变量		口腔义齿	二分类变量
	吸烟	二分类变量		口腔感染史	二分类变量
	饮酒	二分类变量		口腔外科治疗史	二分类变量
	糖尿病	二分类变量	药物相关因素	用药时间	连续变量
	高血压	二分类变量		药物累积剂量	连续变量
	恶性肿瘤	二分类变量		用药间隔	连续变量
	碱性磷酸酶	连续变量		药物类型	二分类变量
	C反应蛋白	连续变量		给药途径	二分类变量
手术因素	D-二聚体	连续变量	手术因素	MRONJ分期	多分类变量
	使用类固醇药物	二分类变量		手术方式	二分类变量
	化疗	二分类变量			

注:累积剂量以药品实际总使用量与限定日剂量的比值[用药频度(DDDs)]进行换算。

采用Z-score对连续变量进行标准化处理,使其转换成均值为0、标准差为1的标准化变量,以消除特征间量纲差异对模型性能的影响;采用独热编码方法对分类变量赋值,以适配机器学习模型的输入要求^[10]。

1.2.3 临床特征降维处理与选择

为筛选最具预测价值的特征并防止过拟合,本研究采用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)回归对候选特征进行降维处理与筛选^[11]:将L1正则化项引入模型,通过调节正则化参数λ,在模型拟合优度和模型稀疏简洁度(特征数量)之间寻求最优平衡;λ通过交叉验证确定,最终保留回归系数非零的特征用于模型构建;同时,根据交叉验证结果,重点关注λ_min(最小验证误差)及λ_1se(最小误差1个标准差内)对应的模型。

1.2.4 自变量共线性诊断

运用方差膨胀因子(variance inflation factor, VIF)检验法评估降维后自变量间的共线性,以VIF<5为无显著共线性的判断标准^[12]。

1.2.5 模型训练与模型性能评价

为降低模型评估偏倚并获得稳健的泛化性能估计,本研究采用双层嵌套交叉验证进行机器学习模型的超参数优化和模型评估^[13]。外层采用分层5折交叉验证,每次以1折为外层测试集,其余4折为外层训练集;内层在外层训练集中采用分层3折交叉验证,根据内层验证结果确定各算法的最优超参数。随后,在完整外层训练集上重新拟合模型,并对外层测试集上评价预测性能。上述过程重复20次,最终汇总各次外层测试结果并计算模型性能指标的均值及离散程度,作为模型性能评估的依据。

队列A采用1:1匹配的巢式病例对照研究设计,结局分布相对均衡,可直接用于模型训练。队列B中,治疗有效病例存在明显的类别不平衡,为避免模型偏向多数类,本研究采用合成少数类过采样技术(synthetic mi-

nority over-sampling technique, SMOTE)算法对治疗无效组进行过采样^[14]。SMOTE采样比例作为内层交叉验证中的待优化参数之一,与模型超参数共同进行选择,以提升模型对少数类结局的识别能力。SMOTE算法仅在数据集划分后应用于内层训练集,以防止数据泄露到外层测试集中,以保证模型评估的可靠性。

基于预先筛选的临床特征,本研究分别构建随机森林(Random Forest, RF)、极端梯度提升(eXtreme Gradient Boosting, XGB)、多层感知机(Multi-layer Perceptron, MLP)、支持向量机(Support Vector Machine, SVM)和高斯朴素贝叶斯(Gaussian Naive Bayes, NB)发病风险及手术疗效预测模型。各模型算法的超参数均在内层交叉验证中进行优化,其预测性能则在外层测试集中进行评价。通过受试者操作特征(receiver operating characteristic, ROC)曲线、精确率-召回率(precision-recall, PR)曲线、校准曲线和决策曲线分析(decision curve analysis, DCA)对模型性能进行综合评价,主要指标包括ROC曲线下面积(ROC-AUC)、PR曲线下面积(PR-AUC)、Brier评分、敏感度、特异度、准确率、F1值。其中,ROC-AUC用于评价模型的区分能力,其值越大表示区分能力越好;PR-AUC主要反映模型对阳性结局的识别能力,尤其适用于类别不平衡数据;Brier评分为预测概率与实际结局之间的均方误差,其值越小表示概率预测总体误差越小;敏感度为模型正确识别阳性病例的比例;特异度为模型正确识别阴性病例的比例;准确率为全部样本中预测正确病例的比例;F1值为阳性预测值与敏感度的调和平均数^[15]。最终,选择ROC-AUC高、校准性能良好、临床净获益较高的模型;若多个模型性能相近,则优先选择结构更简单、可解释性更好的模型^[15]。

1.2.6 模型可解释性分析

使用Shapley加性解释(Shapley additive explanations, SHAP)方法对模型进行可解释性分析。SHAP是一种基于合作博弈论、目前较常用的模型解释方法,可量化各临床特征对模型预测结果的相对贡献,用于揭示模型决策过程并提高其临床可解释性^[16],结果以蜂巢图、瀑布图展示。SHAP蜂巢图中,每个点对应1例患者;红色表示特征取值较高或二分类变量编码为“1”,蓝色表示特征取值较低或二分类变量编码为“0”;队列A中,SHAP值>0表示该特征推动模型向“发生MRONJ”方向预测,SHAP值<0表示推动模型向“未发生MRONJ”方向预测;队列B中,SHAP值>0表示推动模型向“治疗无效”方向预测,SHAP值<0表示推动模型向“治疗有效”方向预测。瀑布图中,红色条表示对应特征使模型输出概率升高,蓝色条表示对应特征使模型输出概率降低。

1.2.7 预测模型的网络化部署

为提高所建发病风险预测及手术疗效预测模型的临床可及性与应用便利性,本研究将经验证后的最优模型部署为交互式网络应用程序。该应用程序前端用于采集模型所需的临床特征变量,后端则进行数据预处理、模型调用及个体风险计算,并自动输出MRONJ发生风险及手术疗效的预测结果。系统经测试后,部署于网络服务器并对外发布。

1.2.8 统计学方法

本研究中,临床基线资料的单因素分析采用SPSS 26.0软件完成;基于LASSO回归的预测特征筛选及机器学习预测模型构建(含SHAP模型解释)均在Python 3.11软件中完成。符合正态分布的连续变量以 $\bar{x} \pm s$ 表示,组间比较采用 t 检验;不符合正态分布的以 $M(P_{25}, P_{75})$ 表示,组间比较采用Mann-Whitney U 检验。分类变量以例数或率表示,组间比较采用Pearson χ^2 检验(总样本量 ≥ 40 且单元格频数 ≥ 5 时)或Fisher精确检验(理论频数不满足上述要求时)。所有统计检验均为双侧检验,检验水准 $\alpha=0.05$ 。

2 结果

2.1 临床基本资料

队列A中,对照组和病例组各有患者165例;病例组中有16例不符合手术适应证,故最终进入队列B的有149例(有效组121例、无效组28例)。两个队列的患者临床基础资料见表2。

2.2 特征变量筛选结果

对于队列A, λ_{1se} 对应模型在保持预测性能稳定的同时可显著降低模型的复杂度,故最终保留了7个临床特征,分别是牙齿拔除、碱性磷酸酶、年龄、药物累积剂量、使用类固醇药物、口腔感染史、高血压(图2)。对于队列B, λ_{1se} 附近模型的预测偏差波动较大,为保证模型的稳定性及预测性能,最终选择 λ_{min} 对应模型,最终共筛选出8个临床特征,分别是年龄、用药时间、MRONJ分期、糖尿病、D-二聚体、给药途径、口腔感染史、手术方式(图3)。

2.3 模型性能评价与模型选择结果

队列A和队列B的RF、XGB、MLP、SVM、NB预测模型的ROC曲线、PR曲线、校准曲线及DCA分别见图4和图5,各模型的评价参数见表3。队列A中,RF模型的ROC-AUC、PR-AUC、Brier评分、敏感度、特异度分别为0.94、0.92、0.09、0.92、0.85;队列B中,RF模型的上述指标分别为0.93、0.93、0.12、0.83、0.93;校准曲线和DCA结果显示,RF模型具有较好的校准表现和临床净获益,因此将该模型选为最终的预测模型用于后续分析。此外,对纳入模型的变量进行共线性诊断,所有变量的VIF均小于5,即未见明显多重共线性问题。虽然,RF模型对变量共线性相对不敏感,但上述结果进一步支持了所建模型的稳定性。

表2 临床基本资料表

变量	队列A				队列B			
	对照组 (n=165)	病例组 (n=165)	$t/U\chi^2$	P	有效组 (n=121)	无效组 (n=28)	$t/U\chi^2$	P
年龄($\bar{x}\pm s$)/岁	61.45 $\pm$ 10.87	70.16 $\pm$ 10.25	-7.488	<0.001	69.72 $\pm$ 10.99	70.00 $\pm$ 7.70	-0.159	0.874
男性/例(%)	76(46.1)	85(51.5)	0.982	0.322	67(55.4)	11(39.3)	1.758	0.185
体重指数($\bar{x}\pm s$)/(kg/m ²)	23.44 $\pm$ 3.44	23.84 $\pm$ 3.28	-1.081	0.280	23.69 $\pm$ 3.31	23.99 $\pm$ 3.16	-0.449	0.648
吸烟/例(%)	48(29.1)	93(56.4)	25.076	<0.001	62(51.2)	22(78.6)	5.840	0.016
饮酒/例(%)	48(29.1)	47(28.5)	0.015	0.903	37(30.6)	9(32.1)	0	>0.999
糖尿病/例(%)	31(18.8)	57(34.5)	10.475	0.001	44(36.4)	9(32.1)	0.041	0.840
高血压/例(%)	50(30.3)	57(34.5)	0.678	0.410	45(37.2)	9(32.1)	0.080	0.778
恶性肿瘤/例(%)	143(86.7)	151(91.5)	1.995	0.158	106(87.6)	25(89.3)		>0.999
碱性磷酸酶($\bar{x}\pm s$)(U/L)	232.32 $\pm$ 98.14	131.91 $\pm$ 81.75	10.098	<0.001	139.23 $\pm$ 90.07	115.45 $\pm$ 72.58	1.489	0.143
C反应蛋白[M(<i>P</i> ₂₅ , <i>P</i> ₇₅)] (mg/L)	5.90(3.00, 9.50)	7.50(4.10, 12.40)	11 019.500	0.003	7.30(3.60, 11.70)	8.00(5.67, 13.15)	1 377.500	0.125
D-二聚体[M(<i>P</i> ₂₅ , <i>P</i> ₇₅)](mg/L)	0.32(0.22, 0.50)	0.34(0.22, 0.50)	12 915.500	0.421	0.34(0.21, 0.48)	0.56(0.30, 3.29)	1 064.000	0.002
使用类固醇药物/例(%)	19(11.5)	81(49.1)	55.153	<0.001	61(50.4)	12(42.9)	0.261	0.609
化疗/例(%)	70(42.4)	65(39.4)	0.313	0.576	51(42.1)	8(28.6)	1.231	0.267
牙齿拔除/例(%)	44(26.7)	139(84.2)	110.711	<0.001	85(70.2)	27(96.4)	7.006	0.008
牙周病/例(%)	63(38.2)	80(48.5)	3.566	0.059	51(42.1)	21(75.0)	8.555	0.003
口腔义齿/例(%)	40(24.2)	39(23.6)	0.017	0.897	28(23.1)	9(32.1)	0.564	0.453
口腔感染史/例(%)	27(16.4)	82(49.7)	41.440	<0.001	59(48.8)	15(53.6)	0.062	0.803
口腔外科治疗史/例(%)	46(27.9)	64(38.8)	4.418	0.036	46(38.0)	9(32.1)	0.132	0.717
用药时间($\bar{x}\pm s$)/月	35.16 $\pm$ 20.71	37.56 $\pm$ 19.76	-1.077	0.282	36.36 $\pm$ 19.50	41.36 $\pm$ 21.31	-1.136	0.263
药物累积剂量($\bar{x}\pm s$)	292.92 $\pm$ 134.37	330.74 $\pm$ 137.83	-2.524	0.012	322.09 $\pm$ 142.70	353.96 $\pm$ 112.50	-1.280	0.207
用药间隔[M(<i>P</i> ₂₅ , <i>P</i> ₇₅)]/月	0(0, 5.00)	2.00(0, 7.00)	12 296.500	0.106	2.00(0, 7.00)	0(0, 5.25)	1 890.500	0.315
药物类型/例(%)			1.774	0.183			0.348	0.555
双膦酸盐	99(60.0)	87(52.7)			59(48.8)	16(57.1)		
地舒单抗	66(40.0)	78(47.3)			62(51.2)	12(42.9)		
给药途径/例(%)			0.788	0.375			0.412	0.521
静脉注射	97(58.8)	89(53.9)			61(50.4)	16(57.1)		
皮下注射	68(41.2)	76(46.1)			60(49.6)	12(42.9)		
MRONJ分期/例(%)							34.533	<0.001
1期					25(20.7)	6(21.4)		
2期					76(62.8)	3(10.7)		
3期					20(16.5)	19(67.9)		
手术方式/例(%)							0.366	0.545
清创术					28(23.1)	8(28.6)		
颌骨部分切除术					93(76.9)	20(71.4)		

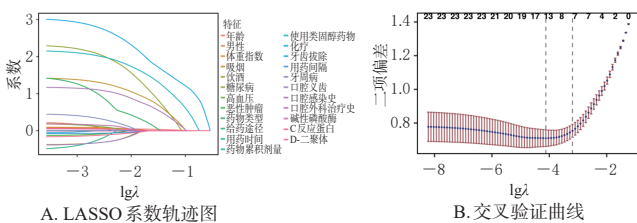


图2 风险预测模型变量特征筛选系数轨迹图和交叉验证图(队列A)

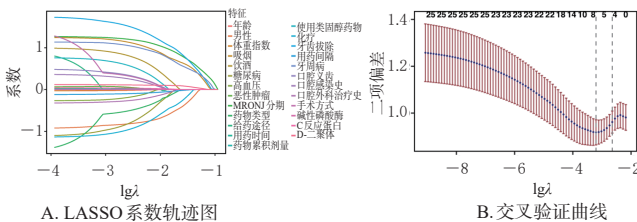


图3 手术疗效预测模型变量特征筛选系数轨迹图和交叉验证图(队列B)

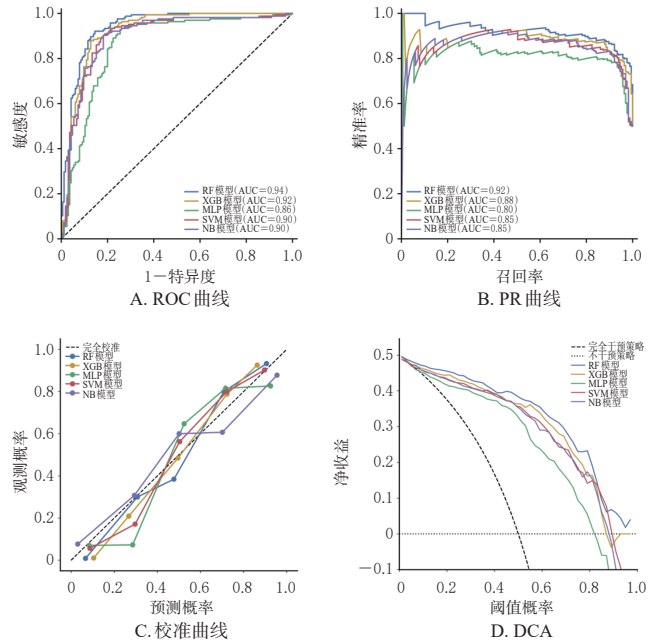


图4 风险预测模型的评估(队列A)

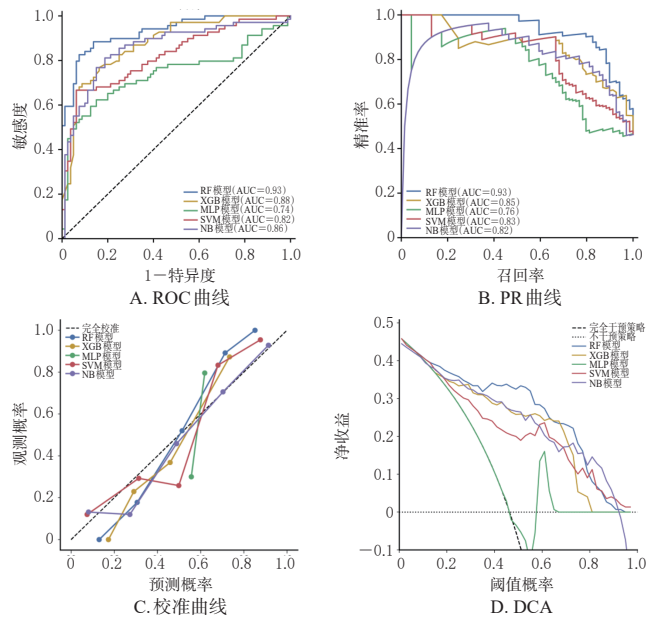


图5 手术疗效预测模型的评估(队列B)

表3 机器学习模型的评价参数

队列	模型	ROC-AUC	PR-AUC	Brier评分	敏感度	特异度	准确率	F1值
队列A	RF	0.94	0.92	0.09	0.92	0.85	0.89	0.89
	XGB	0.92	0.88	0.11	0.88	0.87	0.88	0.88
	MLP	0.86	0.80	0.15	0.92	0.75	0.83	0.85
	SVM	0.90	0.85	0.12	0.90	0.82	0.86	0.87
	NB	0.90	0.85	0.12	0.90	0.82	0.86	0.87
队列B	RF	0.93	0.93	0.12	0.83	0.93	0.88	0.86
	XGB	0.88	0.85	0.15	0.70	0.91	0.81	0.77
	MLP	0.74	0.76	0.25	0.67	0.78	0.72	0.69
	SVM	0.82	0.83	0.17	0.62	0.94	0.79	0.74
	NB	0.86	0.82	0.15	0.77	0.85	0.81	0.79

2.4 模型的可解释性分析结果

SHAP解释结果显示,队列A中,碱性磷酸酶是对MRONJ发病风险预测贡献最大的变量,其次是牙齿拔

除、药物累积剂量、年龄、使用类固醇药物、口腔感染史和高血压。队列B中,年龄对手术疗效预测的贡献最大,其次是用药时间、MRONJ分期、糖尿病、D-二聚体、给药途径、口腔感染史和手术方式(图6)。



注: $E[f(X)]$ 表示模型基线输出, $f(x)$ 表示该个体的最终模型输出。
图6 发病风险和手术疗效预测模型的SHAP分析结果

2.5 预测模型临床应用示例

本研究将构建的MRONJ发病风险预测模型及手术疗效预测模型集成至网络应用程序中,以支持其在临床实践中的应用。该应用程序可通过以下地址访问:<http://39.102.138.68:6002/>。操作时,须输入患者临床特征,分别输出MRONJ发生风险和手术疗效预测结果,并以概率形式展示不同结局发生的可能性。同时,该程序可通过SHAP方法对个体预测结果进行可解释性分析,展示各特征对模型输出结果的贡献大小及作用方向。具体示例(具体图示可扫描本文首页二维码查看“增强出版”板块中的附图1)显示,某患者MRONJ发病风险预测模型输出阴性类别(未发生MRONJ)的概率为67.2%,碱性磷酸酶、牙齿拔除、年龄及累积剂量对该患者预测结果的贡献相对较大;其MRONJ手术疗效预测模型输出阴性类别(手术治疗有效)概率为70.2%,其中MRONJ分期和用药时间是推动预测结果向阴性类别变化的主要因素。

3 讨论

3.1 MRONJ发病风险预测模型的构建与评价

本研究基于巢式病例对照研究和机器学习构建了MRONJ发病风险预测模型,通过1:1匹配结局均衡的数

据集,为模型训练奠定了稳定基础。随后,本研究采用多重插补法对缺失数据进行处理,降低了潜在的系统性偏倚^[11]。然而,数据插补可能影响模型估算,故模型泛化性能仍需在独立的前瞻性队列研究中进一步验证^[17]。在有限的样本量下,本研究采用嵌套交叉验证(外层评估+内层调参)严格隔离测试数据,确保了评估结果的无偏性^[18]。该方法是报告机器学习模型性能的优选方案,适用于本研究这类样本量有限的药源性疾病发生风险预测模型的构建^[19-20]。

本研究通过LASSO回归筛选特征后构建了5种机器学习模型,其中RF模型在区分度、校准度及临床净获益方面表现最优。SHAP分析结果显示,不同临床特征对MRONJ发病风险预测的影响方向并不一致。其中,牙齿拔除、药物累积剂量增加、年龄增加、使用类固醇药物、口腔感染史是推动模型向较高MRONJ发生风险方向预测的主要因素,而较高的碱性磷酸酶水平则是推动模型向较低风险方向预测的主要因素;高血压虽呈弱正向作用,但对模型输出的贡献相对较小。

牙齿拔除是模型中贡献较大的局部危险特征。该操作可造成牙槽骨暴露、局部血供受损及创面修复需求增加,在抗骨吸收药物的作用下,破骨细胞功能和骨重塑将受到抑制,拔牙创面的骨形成与软组织愈合能力下降,从而可能增加局部感染风险^[21]。药物累积剂量增加也表现为正向风险贡献。随着抗骨吸收药物暴露增加,其对骨转换和骨重塑的抑制效应也可能逐渐累积,从而使颌骨在受到微损伤、承受咀嚼应力或发生局部感染后难以完成正常修复^[22]。牙齿拔除和药物累积剂量因素对MRONJ的影响与已有文献观点一致^[21-22]。但需要注意的是,本研究采用巢式病例对照设计,模型中的药物累积剂量反映的是病例确诊前的药物暴露水平,因而更适合解释为较高累积暴露与较高预测风险相关,但不能据此明确量效关系。

年龄增加主要推动模型向高风险方向预测,既往研究亦将高龄列为MRONJ相关因素之一,可能与老年患者骨代谢能力下降、血管修复能力减弱和联合用药增多有关^[22]。使用类固醇药物也趋于风险增加方向,其可能通过抑制免疫反应、影响成骨细胞功能、减少骨形成及延缓创面愈合等途径增加MRONJ的发病风险^[23]。

口腔感染史同样是推动模型向高风险方向预测的主要因素,提示牙周或牙源性感染形成的持续炎症微环境,可能通过增加局部组织损伤、细菌负荷及骨吸收需求,进一步加重受抑制的骨重塑过程^[22]。因此,在风险药物治疗前完成口腔评估、控制感染并妥善处理潜在拔牙需求,可能是降低MRONJ发生风险的重要环节。

值得注意的是,较高的碱性磷酸酶水平主要推动模型向较低的MRONJ发生风险方向预测,病例组患者的该指标水平显著低于对照组。碱性磷酸酶是反映骨形成和肝胆代谢的非特异性指标,其水平受原发肿瘤类型、骨转移负荷、肝功能、营养状态及药物治疗等多种因

素的影响^[24]。因此,本研究观察到的负向关系尚不能简单解释为碱性磷酸酶升高具有保护作用,尚需结合骨转换标志物予以验证。本研究结果还显示,高血压对风险预测模型的贡献较弱,其独立临床意义仍需进一步评估。

总体而言,上述变量在模型中共同表述了药物暴露、局部口腔损伤和感染、机体修复能力及全身健康状况对MRONJ发生风险的综合影响。本研究通过SHAP方法量化了各特征对个体预测结果的相对贡献及作用方向,为临床风险分层管理提供了可解释的依据。但SHAP值反映的是模型预测层面的关联,变量重要性和方向可能受到变量间交互作用及样本构成的影响,故不宜将SHAP贡献度直接解释为变量的独立危险效应或因果效应,临床应用时仍需结合患者的具体情况进行综合判断。

3.2 MRONJ手术疗效预测模型的构建与评价

手术治疗是MRONJ的重要治疗方式,其疗效差异一直是临床决策的难点。本研究基于前瞻性随访队列研究构建MRONJ手术疗效预测模型,以系统评估患者基础状况、疾病特征及围手术期因素等对手术疗效的影响。

由于手术无效组的样本量相对较小,且存在明显的类别不平衡,故本研究在训练阶段采用了SMOTE算法对少数类样本进行插值处理,以提升模型对治疗无效这一结局的识别能力^[14]。尽管该方法有助于缓解类别偏倚,但合成样本并不代表真实患者,因此模型结果仍需在更大样本量的真实分布人群中进行验证。

本研究采用LASSO回归筛选MRONJ手术疗效预测模型的特征变量,并基于所得特征构建了5种机器学习模型;通过性能比较后,对优选出的RF模型进行SHAP可解释性分析。结果显示,年龄对手术疗效预测的贡献最大,其次是用药时间、MRONJ分期、糖尿病、D-二聚体、给药途径、口腔感染史和手术方式。

年龄增加是推动模型向治疗无效方向预测的主要因素,可能与老年患者组织血供及再生能力下降、全身合并症和联合用药增多,以及感染控制和创面修复能力减弱有关^[25]。用药时间和糖尿病在SHAP分析中并未推动模型向治疗无效方向预测,与通常的临床预期不完全相符。用药时间不仅可反映药物暴露程度,而且受患者基础疾病、生存时间及治疗方案等因素的影响;糖尿病对手术疗效的影响则可能与血糖控制水平、病程及并发症情况有关,亦有文献报道糖尿病不是预测MRONJ手术疗效的主要因素^[26]。本研究未进一步对上述因素进行分层分析,加之治疗无效病例数有限,因此不能据此认为较长用药时间或糖尿病有利于手术疗效的提升,其预测方向可能反映变量间的复杂交互及样本构成特征。

研究显示,MRONJ分期是临床意义较为明确的手术疗效预测特征。较高的分期通常提示患者的坏死骨累及范围更广、局部感染及软组织损伤更严重,从而增

加坏死骨完整清除、手术边界确定和创面有效闭合的难度。

D-二聚体升高可反映患者凝血-纤溶系统激活,并可作为全身炎症状态、感染负荷或基础疾病严重程度的间接指标,可能具有评估MRONJ手术疗效的潜在价值^[27]。但SHAP结果显示,该指标对模型预测结果的影响存在一定重叠和非线性。此外,D-二聚体还可能受恶性肿瘤、血栓事件、感染及抗凝治疗等因素的影响^[28],故目前更适合将其视为综合性预后标志物,尚不能直接解释为独立致病因素。

本研究结果显示,给药途径亦对手术疗效预测具有一定贡献,但其可能同时反映药物类型、治疗强度及基础疾病等差异,因此其独立预测意义仍需进一步验证。

需注意的是,手术方式对疗效的影响可能受适应证混杂的干扰。广泛切除术虽利于坏死骨的彻底清除,但其应用通常指向病变更严重的患者。因此,在解释手术方式与疗效的关系时,应充分考虑病变范围及适应证背景,而非单纯依赖模型贡献度进行术式间的横向比较。

和MRONJ发病风险预测模型类似,口腔感染史同样主要推动模型向治疗无效方向预测。持续的牙源性或牙周感染可增加术区细菌感染负荷和局部炎症反应,影响骨组织及软组织修复^[29]。因此,术前控制活动性感染、术中充分清除感染及坏死组织并加强术后口腔卫生管理,可能有助于降低术后疾病持续或进展的风险。

综上,这些因素在一定程度上反映了患者的整体健康状况、疾病严重程度及围手术期风险,对识别术后高风险人群具有潜在价值。

3.3 基于预测模型的MRONJ全流程风险管理展望

本研究分别针对MRONJ发病风险评估和手术疗效预测两个关键临床节点,构建了可解释的预测模型,尝试为MRONJ的全流程风险管理提供量化参考。在临床实践中,发病风险预测模型可辅助临床医生在高风险药物使用前或使用过程中识别潜在高危患者,从而优化口腔评估频率并规避高风险操作(如择期拔牙);而手术疗效预测模型则有助于术前风险沟通及术后随访策略的制定(如密切监测D-二聚体),有助于实现更为精准化与个体化的治疗管理。尽管本研究结果显示出一定的应用前景,但仍需在更大样本量及跨区域独立人群中进一步验证;同时,预测模型在临床实践中的落地应用仍面临数据质量控制、跨机构数据共享及与现有医疗工作流程融合等挑战。未来,有必要通过多中心合作与数据共享,构建更大规模的临床数据库,以持续优化模型性能并推动其在MRONJ风险管理及药源性疾病精准治疗中的实际应用。

利益冲突 本文所有作者声明不存在利益冲突

参考文献

- [1] Marx R E. Pamidronate(Aredia)and zoledronate(Zometa) induced avascular necrosis of the jaws: a growing epidemic[J]. J Oral Maxillofac Surg, 2003, 61(9): 1115-1117.

- [2] Advisory Task Force on Bisphosphonate-Related Osteonecrosis of the Jaws, American Association of Oral and Maxillofacial Surgeons. American Association of Oral and Maxillofacial Surgeons position paper on bisphosphonate-related osteonecrosis of the jaws[J]. *J Oral Maxillofacial Surg*, 2007, 65(3): 369-376.
- [3] Ruggiero S L, Dodson T B, Fantasia J, et al. American Association of Oral and Maxillofacial Surgeons position paper on medication-related osteonecrosis of the jaw: 2014 update[J]. *J Oral Maxillofacial Surg*, 2014, 72(10): 1938-1956.
- [4] Ruggiero S L, Dodson T B, Aghaloo T, et al. American Association of Oral and Maxillofacial Surgeons' position paper on medication-related osteonecrosis of the jaws: 2022 update[J]. *J Oral Maxillofacial Surg*, 2022, 80(5): 920-943.
- [5] Tani M, Hatano K, Yoshimura A, et al. Cumulative incidence and risk factors for medication-related osteonecrosis of the jaw during long-term prostate cancer management[J]. *Sci Rep*, 2024, 14(1): 13451.
- [6] 何悦,陈珩,安金刚,等. 药物相关性颌骨坏死临床诊疗专家共识[J]. *中国口腔颌面外科杂志*, 2023, 21(4): 313-325.
- [7] Bai C S, Wu L L, Li R J, et al. Machine learning-enabled drug-induced toxicity prediction[J]. *Adv Sci*, 2025, 12(16): 2413405.
- [8] Hu M N, Lin X L, Sun T T, et al. Gut microbiome for predicting immune checkpoint blockade-associated adverse events[J]. *Genome Med*, 2024, 16(1): 16.
- [9] Little R, Rubin D. Statistical analysis with missing data, third edition[M]. New York: John Wiley & Sons, 2019: 223-246.
- [10] James G, Witten D, Hastie T, et al. An introduction to statistical learning: with applications in R[M]. New York: Springer New York, 2013: 83-86, 217-218.
- [11] Efthimiou O, Seo M, Chalkou K, et al. Developing clinical prediction models: a step-by-step guide[J]. *BMJ*, 2024, 386: e078276.
- [12] Kalnins A, Praitis Hill K. Additional caution regarding rules of thumb for variance inflation factors: extending O' Brien to the context of specification error[J]. *Qual Quant*, 2025, 59(1): 291-314.
- [13] Moons K G M, Damen J A A, Kaul T, et al. PROBAST+ AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods[J]. *BMJ*, 2025, 388: e082505.
- [14] Li Y, Yang Y L, Song P H, et al. An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space[J]. *Sci Rep*, 2025, 15: 23521.
- [15] Van Calster B, Collins G S, Vickers A J, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance[J]. *Lancet Digit Heal*, 2025, 7(12): 100916.
- [16] Ponce-Bobadilla A V, Schmitt V, Maier C S, et al. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development[J]. *Clin Transl Sci*, 2024, 17(11): e70056.
- [17] Deforth M, Heinze G, Held U. The performance of prognostic models depended on the choice of missing value imputation algorithm: a simulation study[J]. *J Clin Epidemiol*, 2024, 176: 111539.
- [18] Stahl D. New horizons in prediction modelling using machine learning in older people's healthcare research[J]. *Age Ageing*, 2024, 53(9): afac201.
- [19] Collins G S, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation[J]. *BMJ*, 2024, 384: e074819.
- [20] Riley R D, Archer L, Snell K I E, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study[J]. *BMJ*, 2024, 384: e074820.
- [21] Yokoo S, Kubo S, Yamamoto D, et al. Denosumab dosage and tooth extraction predict medication-related osteonecrosis of the jaw in patients with breast cancer and bone metastases[J]. *Cancers*, 2025, 17(13): 2242.
- [22] Nakamura N, Otsuru M, Miyoshi T, et al. Risk factors for the development of medication-related osteonecrosis of the jaw and effects of tooth extraction with local infection [J]. *J Dent Sci*, 2024, 19(3): 1770-1782.
- [23] Kujanpää M, Vuollo V, Tiisanoja A, et al. Incidence of medication-related osteonecrosis of the jaw and associated antiresorptive drugs in adult Finnish population[J]. *Sci Rep*, 2025, 15(1): 17377.
- [24] Minisola S, Cipriani C, Colangelo L, et al. Diagnostic approach to abnormal alkaline phosphatase value[J]. *Mayo Clinic Proc*, 2025, 100(4): 712-728.
- [25] Kakehashi H, Sakamoto M, Moriyama M, et al. The clinical effectiveness of conservative surgical management in medication-related osteonecrosis of the jaw[J]. *J Oral Maxillofacial Surg Med Pathol*, 2024, 36(4): 471-477.
- [26] Mohr J, Pippich K, Cherdron C, et al. Postoperative healing and recurrence in osteonecrosis of the jaw: influence of risk factors and surgical approach[J]. *Clin Oral Investig*, 2026, 30(4): 156.
- [27] 王睿,何金媛,蒲玉梅,等. D-二聚体对手术治疗药物相关性颌骨坏死患者预后价值的研究[J]. *口腔医学研究*, 2025, 41(7): 568-573.
- [28] Franchini M, Focosi D, Pezzo M P, et al. How we manage a high D-dimer[J]. *Haematologica*, 2024, 109(4): 1035-1045.
- [29] Natsu M, Meuric V, Roginski P, et al. Medication-related osteonecrosis of the jaw: evaluation of a therapeutic strategy in oral surgery[J]. *J Stomatology Oral Maxillofacial Surg*, 2024, 125(5): 101877.

(收稿日期:2026-02-10 修回日期:2026-07-24)
(本文责编:张元媛)