

糖尿病足溃疡复发风险预测模型的构建：基于 Logistic 回归和支持向量机及 BP 神经网络模型



扫描二维码
查看原文

张娟, 李海芬, 李小曼, 姚苗, 马惠珍, 马强*

【摘要】 背景 全球范围内糖尿病足溃疡(DFUs)首次复发与再次复发率逐年上升,且早期复发风险高于远期风险。导致DFUs复发的风险因素较多,目前缺乏系统的筛选,因此需要探索DFUs复发的危险因素,以便早期识别复发高危人群。目的 探讨Logistic回归、支持向量机(SVM)和BP神经网络(BPNN)模型在DFUs复发风险中的预测价值。方法 选取2020年1月—2021年10月在宁夏医科大学总医院烧伤整形美容科、内分泌科和伤口造口门诊就诊的DFUs患者390例作为开发模型的研究对象。根据患者出院后1年内DFUs是否复发分为复发组116例(29.7%)和非复发组274例(70.3%)。收集两组患者的一般资料包括社会人口学特征、病史评估和临床病例资料并进行比较,采用糖尿病足部自我管理行为量表(DFSBS)评估患者糖尿病足部自我管理行为,采用慢性病风险感知问卷评估患者DFUs风险感知水平。采用多因素Logistic回归分析探讨DFUs患者出院后1年内DFUs复发的影响因素;将患者按照7:3划分为训练集和测试集,运用Logistic回归变量筛选策略,分别建立Logistic回归、SVM和BPNN模型;绘制各模型预测DFUs复发风险的受试者工作特征(ROC)曲线。结果 两组DFUs患者BMI、独居、糖尿病病程、吸烟史、饮酒史、受累足趾截肢史、足溃疡分级、踝肱指数、糖化血红蛋白、溃疡位置在脚底、足趾受累、足部存在行走障碍、骨髓炎、多重耐药菌感染、糖尿病周围神经病变、下肢动脉粥样硬化、足部自我管理行为、DFUs风险感知水平比较,差异均有统计学意义($P<0.05$)。多因素Logistic回归分析结果显示,BMI[OR=0.394, 95%CI(0.285, 0.546), $P<0.001$]、糖尿病病程[OR=1.635, 95%CI(1.303, 2.051), $P<0.001$]、吸烟史[OR=0.186, 95%CI(0.080, 0.434), $P<0.001$]、足溃疡分级[OR=2.139, 95%CI(1.133, 4.038), $P=0.019$]、糖化血红蛋白[OR=2.289, 95%CI(1.485, 3.528), $P<0.001$]、溃疡位置在脚底[OR=3.148, 95%CI(1.344, 7.373), $P=0.008$]、足部自我管理行为[OR=0.744, 95%CI(0.673, 0.822), $P<0.001$]和DFUs风险感知水平[OR=0.892, 95%CI(0.845, 0.942), $P<0.001$]是DFUs患者1年内DFUs复发的影响因素。Logistic回归、SVM和BPNN模型在测试集中预测DFUs复发风险的正确率分别82.43%、94.87%、87.17%,ROC曲线下面积(AUC)分别为0.843、0.937、0.820。Logistic回归、SVM和BPNN模型预测DFUs复发风险的ROC曲线AUC比较,差异有统计学意义($Z=2.741, P<0.05$);SVM模型预测DFUs复发风险的ROC曲线AUC高于Logistic回归和BPNN模型($Z=5.937, P=0.013; Z=3.946, P<0.001$)。结论 SVM模型预测DFUs患者出院后1年内DFUs复发风险的正确率、灵敏度、特异度、AUC等指标均较好,为相对最优的模型,建议进一步推广应用以验证预测模型的效能。

【关键词】 糖尿病;足溃疡;糖尿病足;复发;Logistic模型;支持向量机模型;BP神经网络模型;影响因素分析

【中图分类号】 R 587.1 **【文献标识码】** A DOI: 10.12114/j.issn.1007-9572.2023.0175

【引用本文】 张娟,李海芬,李小曼,等.糖尿病足溃疡复发风险预测模型的构建:基于Logistic回归和支持向量机及BP神经网络模型[J].中国全科医学,2023,26(32):4013-4019. DOI: 10.12114/j.issn.1007-9572.2023.0175. [www.chinagp.net]

ZHANG J, LI H F, LI X M, et al. Construction of recurrence risk prediction model for diabetic foot ulcer on the basis of Logistic regression, support vector machine and BP neural network model [J]. Chinese General Practice, 2023, 26 (32) : 4013-4019.

Construction of Recurrence Risk Prediction Model for Diabetic Foot Ulcer on the Basis of Logistic Regression, Support Vector Machine and BP Neural Network Model ZHANG Juan, LI Haifen, LI Xiaoman, YAO Miao, MA Huizhen, MA Qiang*

Department of Burn Plastic Surgery, General Hospital of Ningxia Medical University, Yinchuan 750001, China

基金项目:宁夏自然科学基金项目(2022AAC03515);宁夏回族自治区重点研发计划项目(2020BEG03025)
750001宁夏回族自治区银川市,宁夏医科大学总医院烧伤整形美容科

*通信作者:马强,副主任医师;E-mail:10901969@qq.com

本文数字出版日期:2023-05-31

*Corresponding author: MA Qiang, Associate chief physician; E-mail: 10901969@qq.com

【 Abstract 】 Background The rates of first and multiple recurrence of diabetic foot ulcers (DFUs) are increasing annually worldwide, and the risk of early recurrence is higher than the distant recurrence. There are numerous risk factors for DFUs recurrence, and there is a lack of systematic screening. Therefore, there is a need to explore the risk factors for DFUs recurrence in order to identify high-risk population of recurrence at an early stage. **Objective** To explore the predictive value of Logistic regression (LR), support vector machine (SVM), BP neural network model (BPNN) in the recurrence risk of DFUs. **Methods** From January 2020 to October 2021, a total of patients with DFUs attending the Department of Burn Plastic Surgery, Endocrinology and Wound Ostomy Outpatient Department in General Hospital of Ningxia Medical University were selected as the research objects and divided into the recurrence group ($n=116$, 29.7%) and non-recurrence group ($n=274$, 70.3%) according to the recurrence of DFUs within 1 year after discharge. General information was collected and compared between the two groups of patients, including sociodemographic characteristics, medical history assessment and clinical case information. The Diabetes Foot Self-care Behavior Scale (DFSBS) was used to assess the self-management behavior of diabetes foot in patients and chronic diseases risk perception questionnaire was used to assess the risk perception level of DFUs among patients. Multivariable Logistic regression analysis was used to explore the influencing factors of DFUs recurrence in patients within 1 year after discharge. The patients were divided into training and test sets according to the ratio of 7 to 3, the LR, SVM and BPNN recurrence risk prediction models were developed based on Logistic regression variable screening strategy. The receiver operating characteristic (ROC) curves of each model were plotted to predict the recurrence risk of DFUs. **Results** There were significant differences in BMI, living alone, duration of diabetes, history of smoking, history of alcohol consumption, history of involved toe amputation, classification of diabetic foot ulcers, ankle-brachial index, glycosylated hemoglobin, sole ulcer, toe involvement, walking impairment, osteomyelitis, multidrug-resistant bacteria infection, diabetic peripheral neuropathy, lower limb atherosclerosis, self-management behavior of diabetes foot, level of risk perception in both groups of DFUs patients ($P<0.05$). Multivariable Logistic regression analysis showed that BMI [$OR=0.394$, 95% $CI(0.285, 0.546)$, $P<0.001$], duration of diabetes [$OR=1.635$, 95% $CI(1.303, 2.051)$, $P<0.001$], history of smoking [$OR=0.186$, 95% $CI(0.080, 0.434)$, $P<0.001$], classification of diabetic foot ulcers [$OR=2.139$, 95% $CI(1.133, 4.038)$, $P=0.019$], glycosylated hemoglobin [$OR=2.289$, 95% $CI(1.485, 3.528)$, $P<0.001$], sole ulcer [$OR=3.148$, 95% $CI(1.344, 7.373)$, $P=0.008$], self-management behavior of diabetes foot [$OR=0.744$, 95% $CI(0.673, 0.822)$, $P<0.001$] and level of risk perception [$OR=0.892$, 95% $CI(0.845, 0.942)$, $P<0.001$] were influencing factors of the recurrence of DFUs within 1 year ($P<0.05$). The accuracy rates of LR, SVM and BPNN models to predict the recurrence risk of DFUs in the test sets were 82.43%, 94.87% and 87.17%, with AUCs of 0.843, 0.937 and 0.820, respectively. There were significant differences in AUC of ROC curves of LR, SVM and BPNN recurrence risk prediction models of DFUs ($Z=2.741$, $P<0.05$); the AUC of ROC curve of SVM recurrence risk prediction model was higher than the LR and BPNN models ($Z=5.937$, $P=0.013$; $Z=3.946$, $P<0.001$). **Conclusion** SVM model can predict the recurrence risk of DFUs patients within 1 year after discharge with good accuracy rate, sensitivity, specificity, AUC and other indicators, which is the relative optimal model. It is recommended to further promote and apply the prediction model to verify its effectiveness.

【 Key words 】 Diabetes mellitus; Foot ulcer; Diabetic foot; Recurrence; Logistic models; Support vector machine; Back propagation neural network; Root cause analysis

国际糖尿病协会数据显示, 2019 年全球糖尿病患病率为 9.3%, 到 2045 年将升至 10.9%^[1]。与此同时, 全球慢性创面的患病率也在增加, 以糖尿病足溃疡 (DFUs) 和下肢静脉溃疡最为常见^[2]。DFUs 是糖尿病严重的并发症之一, 可导致高死亡率和致残率^[3]。DFUs 患者的终生风险达到 15%~25%^[4], DFUs 患者截肢的年发生率为 5.1%^[1]。有研究报道即使创面愈合后 DFUs 在 1 年内复发率仍高达 40%, 3 年复发率为 50%~60%, 5 年内复发率为 65%^[4]。显然, DFUs 预后不稳定, 易复发且早期复发风险较高。

吕静等^[5]采用 Logistic 回归建立 DFUs 复发风险预测模型, 1 年内 DFUs 复发率为 26.9%, 吸烟、愈伤组

织、足部肤色异常、胼胝质、糖尿病周围神经病变和冠心病是 DFUs 复发的风险因素。CRAWFORD 等^[6]的研究显示 DFUs 史、无法感受到 10 g 单丝和没有任何足踏板脉搏是 DFUs 复发的独立危险因素。最近, AAN DE STEGGE 等^[7]建立了 DFUs 复发风险预测模型, 模型 1 主要针对复发性足底溃疡, 模型 2 主要针对无法识别的重复性压力引起的溃疡。以往研究中 DFUs 复发风险预测模型的建立方法单一, 未对比不同算法的预测效能, 在此情况下本研究运用机器学习 (ML) 中的支持向量机 (SVM) 和 BP 神经网络 (BPNN) 开发出 DFUs 复发风险预测模型, 并对比其预测效能, 以便临床医护人员能够及时识别出有复发性 DFUs 风险的患者, 指导临床

医生和患者进行预防性治疗的联合决策，最终达到降低DFUs复发风险的目的。

1 对象与方法

1.1 研究对象 选取2020年1月—2021年10月在宁夏医科大学总医院烧伤整形美容科、内分泌科和伤口造口门诊就诊的DFUs患者作为开发模型的研究对象。纳入标准：(1)符合2020年美国糖尿病学会糖尿病医学诊疗标准中DFUs的诊断标准^[8]；(2)因初发DFUs而就诊；(3)自愿参与，且签署知情同意书。排除标准：(1)糖尿病合并妊娠或特殊类型糖尿病；(2)合并严重恶性肿瘤等危重疾病；(3)有沟通障碍；(4)有精神疾病；(5)中途失访。本研究已通过宁夏医科大学总医院伦理审查(编号：KYLL-2021-677)。

1.2 样本量计算及分组 本研究共纳入36个候选预测变量，根据多变量预测模型透明报告对样本量的要求^[9]：阳性结局事件数至少是候选预测变量个数的10倍及以上，即本模型开发至少需要360个DFUs结局事件。经数据清洗，本研究共纳入390例DFUs患者为研究对象。

本研究结局变量是DFUs愈合出院后1年内是否复发的情况。DFUs复发诊断标准参照国际糖尿病足溃疡工作组(IWGDF)指南^[10]和《2020IWGDF指南：糖尿病足的诊断》标准^[11]：既往有过DFUs史的患者出现新的足溃疡，不论这次溃疡发生的部位与上次是否相同。随访1年，以初诊DFUs住院诊治且溃疡愈合后的时间为起点，若随访期内同一患者DFUs多次复发，则仅记1次，患者复发信息以第一次复发时收集的信息为参考。根据患者出院后1年内DFUs是否复发分为复发组116例(29.7%)和非复发组274例(70.3%)。

1.3 观察指标

1.3.1 一般资料 根据研究目的及内容，由研究小组通过文献回顾、预调查和专家建议等方式自行设计一般资料收集表，内容包括社会人口学特征、病史评估、临床病例资料。(1)社会人口学特征包括性别、年龄、BMI、独居；(2)病史评估包括糖尿病病程、吸烟史(每天至少吸1支烟，持续或累计半年以上且仍在吸烟^[9])、饮酒史(每周至少饮酒1次，持续或累计饮酒半年以上且仍在饮酒^[9])、受累足趾截肢史、溃疡因创伤引起；(3)临床病例资料包括足溃疡分级、踝肱指数、糖化血红蛋白、溃疡位置在脚底、足部胼胝、足趾受累、足部存在行走障碍、骨髓炎、多重耐药菌感染、下肢动脉粥样硬化、糖尿病周围神经病变。

1.3.2 糖尿病足部自我管理行为量表(DFSBS) 采用DFSBS评估糖尿病足部自我管理行为，DFSBS由CHIN等^[12]研制，共7个条目，量表总分为7~35分，得分越高表示足部自我管理行为越好，量表Cronbach's α 系数为0.835，重测信度为0.916。

1.3.3 慢性病风险感知问卷 采用慢性病风险感知问

卷评估患者DFUs风险感知水平，该问卷由方蕾^[13]编制，用于评估慢性病患者的风险感知水平，包括经济风险、身体诊疗风险及社会心理风险，共3个维度12个条目。该问卷的Cronbach's α 系数为0.833，各维度Cronbach's α 系数为经济风险0.716、身体诊疗风险0.769、社会心理风险0.781，具有良好的信效度。

1.4 质量控制 运用医院信息系统收集临床资料，通过电话随访、门诊随访与院外随访管理系统相结合的方式，记录患者DFUs的复发情况，随访时间为1年，由临床医生、糖尿病专科护士或伤口专科护士判断患者的复发情况及症状的严重程度。收集资料前统一在烧伤整形美容科、伤口造口门诊和内分泌科选派3名糖尿病专科护士和3名伤口专科护士共6名作为资料收集员，由研究者对其进行同质化培训和指导，包括随访资料收集的内容、方法和注意事项等。选取5例DFUs患者进行预调查以完善资料收集表。采用问卷星编制《糖尿病足溃疡问卷调查表》，运用金山文档编制《糖尿病足溃疡在线资料收集表》，通过以上两种途径完成资料收集。由培训合格的资料收集员协助患者作答问卷。电话随访由烧伤整形美容科护士进行，门诊随访在伤口造口门诊完成，院外管理系统随访由内分泌科教育护士完成。

1.5 模型构建

1.5.1 数据初步筛选和预处理 审核数据的变量特征，并按既定规则进行初步筛选：(1)删除每列数据缺失占比>90%的变量；(2)删除每列单个类别比例>90%的变量；(3)删除每列变异系数(CV)<0.05的变量；(4)删除不合逻辑的异常值，连续性变量的缺失值用中位数填补，分类变量的缺失值用众数填补。

1.5.2 特征及算法的选择 本研究的特征选择步骤在训练数据集中，以DFUs1年内是否复发为因变量，其他变量为自变量进行初始逐步Logistic回归分析，利用OR值及其95%CI等指标增加模型的可解释性。将单因素分析中P<0.05的预测因子纳入多因素分析，最后使用逐步回归确定进入模型的预测因子。BPNN对数据资料分布无严格限制，可识别变量间复杂的非线性关系，具有很强的自适应能力，且有高度的容错性。SVM是一种线性和非线性分类方法，具有预测可靠性高、稳定性强和泛化能力强等优势^[14]。

1.5.3 模型的建立及评价 纳入Logistic回归中有意义的11个独立预测因素，构建Logistic、BPNN及SVM风险预测模型，各模型训练集和测试集按7:3比例，292例样本用于训练，98例样本用于预测。以受试者工作特征(ROC)曲线下面积(AUC)、查准率、正确率、召回率和F1值作为模型评估的指标，在各指标结果不一致时，以AUC作为主要参考，其中AUC为0.50~0.70提示预测效果较差，AUC为>0.70~0.90提示预测效果中等，AUC>0.90提示预测效果非常好。

1.6 统计学方法 运用 Excel 导出并录入数据, 采用 SPSS 25.0 及 MATLAB R2020b 分析数据。非正态分布的定量资料以 $M(P_{25}, P_{75})$ 表示, 两组间比较采用 Mann-Whitney U 检验; 定性资料以频数及百分比表示, 两组间比较采用 χ^2 检验, 等级资料比较采用 Kruskal-Wallis H 检验。采用 SPSS 25.0 构建 Logistic 回归模型预测 DFUs 复发的影响因素, 采用 MATLAB R2020b 软件构建 BPNN 和 SVM 模型, 模型的拟合优度采用 Hosmer-Lemeshow 检验, 绘制各模型预测 DFUs 复发风险的 ROC 曲线, 采用查准率、正确率、召回率、F1 指数及 AUC 评价各模型的预测效能。运用 DeLong 检验比较各模型的性能及 AUC 间的差异。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 一般资料比较 两组 DFUs 患者性别、年龄、溃疡因创伤引起、足部胼胝比较, 差异均无统计学意义 ($P > 0.05$); 两组 DFUs 患者 BMI、独居、糖尿病病程、吸烟史、饮酒史、受累足趾截肢史、足溃疡分级、踝肱指数、糖化血红蛋白、溃疡位置在脚底、足趾受累、足部存在行走障碍、骨髓炎、多重耐药菌感染、糖尿病周围神经病变、下肢动脉粥样硬化、足部自我管理

行为、DFUs 风险感知水平比较, 差异均有统计学意义 ($P < 0.05$), 见表 1。

2.2 DFUs 患者出院后 1 年内 DFUs 复发影响因素的多因素 Logistic 回归分析 以 DFUs 患者出院后 1 年内 DFUs 是否复发为因变量, 以表 1 中差异有统计学意义的因素: BMI、独居、糖尿病病程、吸烟史、饮酒史、受累足趾截肢史、足溃疡分级、踝肱指数、糖化血红蛋白、溃疡位置在脚底、足趾受累、足部存在行走障碍、骨髓炎、多重耐药菌感染、糖尿病周围神经病变、下肢动脉粥样硬化、足部自我管理行为、DFUs 风险感知水平为自变量进行多因素 Logistic 回归分析 (各变量赋值情况见表 2), 结果显示, BMI、糖尿病病程、吸烟史、足溃疡分级、糖化血红蛋白、溃疡位置在脚底、足部自我管理行为和 DFUs 风险感知水平是 DFUs 复发的影响因素 ($P < 0.05$), 见表 3。

2.3 三种模型的运行结果

2.3.1 Logistic 回归模型运行结果 将多因素 Logistic 回归分析有统计学意义的 8 个影响因素作为预测模型的输入变量。Logistic 回归在训练集上的拟合值与真实值相比, 其预测 DFUs 患者 DFUs 复发的查准率、正确率和 AUC 分别为 85.26%、84.41% 和 0.855, 见表 4。

表 1 两组 DFUs 患者一般资料比较

Table 1 Comparison of general information between the two groups of patients with DFUs

组别	例数	性别 [例 (%)]		年龄 [$M(P_{25}, P_{75})$, 岁]	BMI [$M(P_{25}, P_{75})$, kg/m^2]	独居 [例 (%)]	糖尿病病程 [$M(P_{25}, P_{75})$, 年]	吸烟史 [例 (%)]
		男	女					
复发组	116	55 (47.4)	61 (52.6)	57.0 (53.0, 65.0)	22.5 (21.5, 23.0)	55 (47.4)	11 (10.0, 12.0)	33 (28.4)
未复发组	274	153 (55.8)	121 (44.2)	58.0 (52.0, 65.0)	23.0 (22.0, 24.5)	97 (35.4)	9 (8.0, 11.0)	115 (42.0)
检验统计量值		2.325		-0.247 ^a	-5.660 ^a	4.944	-6.802 ^a	6.329
P 值		0.151		0.805	<0.001	0.026	<0.001	0.012
组别	饮酒史 [例 (%)]	受累足趾截肢史 [例 (%)]	溃疡因创伤引起 [例 (%)]	足溃疡分级 [例 (%)]				
				I 级	II 级	III 级	IV 级	V 级
复发组	35 (30.2)	63 (54.3)	55 (47.4)	4 (3.4)	17 (14.7)	43 (37.0)	35 (30.2)	17 (14.7)
未复发组	127 (46.4)	75 (27.4)	112 (40.9)	62 (22.6)	88 (32.1)	74 (27.0)	43 (15.7)	7 (2.6)
检验统计量值	8.784	25.866	1.423	57.628 ^b				
P 值	0.003	<0.001	0.233	<0.001				
组别	踝肱指数 [例 (%)]		溃疡位置在脚底 [例 (%)]	足部胼胝 [例 (%)]	足趾受累 [例 (%)]	足部存在行走障碍 [例 (%)]	骨髓炎 [例 (%)]	
	差	良						
复发组	48 (41.4)	68 (58.6)	66 (56.9)	37 (31.9)	57 (49.1)	91 (78.4)	55 (47.4)	
未复发组	68 (24.8)	206 (75.2)	92 (33.6)	76 (27.7)	83 (30.3)	157 (57.3)	95 (34.7)	
检验统计量值	10.697		18.390	0.685	12.679	15.744	5.591	
P 值	<0.001		<0.001	0.408	<0.001	<0.001	0.018	
组别	多重耐药菌感染 [例 (%)]	下肢动脉粥样硬化 [例 (%)]	糖尿病周围神经病变 [例 (%)]	糖化血红蛋白 [$M(P_{25}, P_{75})$, %]	足部自我管理行为 [$M(P_{25}, P_{75})$, 分]	DFUs 风险感知水平 [$M(P_{25}, P_{75})$, 分]		
复发组	70 (60.3)	51 (44.0)	57 (49.1)	8.5 (7.6, 9.4)	18.45 (16.31, 22.37)	34.52 (31.31, 39.58)		
未复发组	85 (31.0)	75 (27.4)	92 (33.6)	7.5 (7.2, 8.4)	22.80 (20.33, 25.49)	40.58 (37.72, 43.36)		
检验统计量值	29.261	10.260	8.359	-6.911 ^a	-8.077 ^a	-6.768 ^a		
P 值	<0.001	<0.001	0.004	<0.001	<0.001	<0.001		

注: ^a 为 Z 值, ^b 为 H 值, 余检验统计量值为 χ^2 值; DFUs= 糖尿病足溃疡。

2.3.2 SVM 模型运行结果 以 DFUs 是否复发为因变量 (赋值: 否 =0, 是 =1), 多因素 Logistic 回归分析筛选出的变量为自变量 (赋值同表 2), 在训练集上建立 SVM, 利用 tune.svm () 函数计算, 得出 10 折交叉验证错误率最低时的最优参数为 C=100、 $\gamma=0.01$, 此时训练集 SVM 预测 DFUs 患者复发的查准率、正确率、AUC 分别为 95.30%、96.53% 和 0.943, 见表 4。

2.3.3 BPNN 模型运行结果 以 DFUs 是否复发为因变量 (赋值: 否 =0, 是 =1), 以多因素 Logistic 回归分析中有统计学意义的变量为自变量 (赋值同表 2), 计量资料采用标准化公式, 将输入层协变量全部归一化处理为 (0, 1) (区间的变量) 输入 BPNN 模型。对测试集预测的正确率为 96.06%, 训练集预测的正确率为 88.36%, 见表 4。

2.4 三种模型的预测价值比较 三种模型预测患者 DFUs 复发的 ROC 曲线均具有较高的正确率, 在训练集和测试集中, 三种模型预测 DFUs 复发风险的 ROC 曲线 AUC 比较结果为: SVM>Logistic 回归>BPNN。由 ROC 曲线可知, SVM 在训练集和测试集中预测患者 DFUs 复发的 AUC 均最高, 测试集中的 AUC 为 0.937 [95%CI (0.916, 0.955)], 且正确率最高为 94.87%, 灵敏度为 0.99, 特异度为 0.87, SVM 模型整体表现较另外两种预测模型预测效能更优。经 DeLong 检验, Logistic 回归、SVM 和 BPNN 模型预测 DFUs 复发风险的 ROC 曲线

AUC 比较, 差异有统计学意义 ($Z=2.741, P<0.05$); SVM 模型预测 DFUs 复发风险的 ROC 曲线 AUC 大于 Logistic 回归和 BPNN 模型, 差异有统计学意义 ($Z=5.937, P=0.013; Z=3.946, P<0.001$); Logistic 回归模型预测 DFUs 复发风险的 ROC 曲线 AUC 大于 BPNN 模型, 差异有统计学意义 ($Z=4.302, P=0.022$), 见图 1、表 4。

3 讨论

DFUs 复发是糖尿病患者不良预后的重要标志^[15]。本研究关注引起 DFUs 复发的主客观变量, 自行研制

表 3 DFUs 复发影响因素的多因素 Logistic 回归分析

Table 3 Multivariable Logistic regression analysis of influencing factors of DFUs recurrence

自变量	β	SE	Wald χ^2 值	P 值	OR 值	95%CI
常量	18.018	4.174	18.636	<0.001	—	—
BMI	-0.931	0.166	31.337	<0.001	0.394	(0.285, 0.546)
独居	-0.562	0.412	1.861	0.173	0.570	(0.255, 1.278)
糖尿病病程	0.492	0.116	18.050	<0.001	1.635	(1.303, 2.051)
吸烟史	-1.679	0.431	15.209	<0.001	0.186	(0.080, 0.434)
饮酒史	-0.489	0.401	1.490	0.222	0.613	(0.279, 1.345)
受累足趾截肢史	0.402	0.483	0.693	0.405	1.495	(0.580, 3.853)
足溃疡分级	0.760	0.324	5.497	0.019	2.139	(1.133, 4.038)
踝肱指数	-0.093	0.519	0.032	0.858	0.911	(0.329, 2.521)
糖化血红蛋白	0.828	0.221	14.081	<0.001	2.289	(1.485, 3.528)
溃疡位置在脚底	1.147	0.434	6.972	0.008	3.148	(1.344, 7.373)
足趾受累	0.032	0.470	0.005	0.946	1.032	(0.411, 2.596)
足部存在行走障碍	-0.408	0.541	0.568	0.451	0.665	(0.231, 1.920)
骨髓炎	-0.938	0.587	2.555	0.110	0.391	(0.124, 1.236)
多重耐药菌感染	-0.119	0.174	0.464	0.496	0.888	(0.632, 1.249)
下肢动脉粥样硬化	0.110	0.452	0.059	0.808	1.116	(0.460, 2.705)
糖尿病周围神经病变	-0.311	0.600	0.270	0.604	0.732	(0.226, 2.372)
足部自我管理行为	-0.296	0.051	33.509	<0.001	0.744	(0.673, 0.822)
DFUs 风险感知水平	-0.114	0.028	16.686	<0.001	0.892	(0.845, 0.942)

注: —表示无此项数据。

表 4 Logistic 回归、SVM 和 BPNN 模型对 DFUs 复发风险的预测性能比较

Table 4 Comparison of predictive performance of Logistic regression, SVM and BPNN recurrence risk prediction models for DFUs

评价指标	训练集			测试集		
	Logistic 回归模型	SVM 模型	BPNN 模型	Logistic 回归模型	SVM 模型	BPNN 模型
查准率 (%)	85.26	95.30	88.36	85.42	97.06	90.06
召回率 (%)	82.74	89.79	68.40	82.94	88.19	65.83
正确率 (%)	84.41	96.53	90.64	82.43	94.87	87.17
F1 指数	0.85	0.97	0.94	0.84	0.96	0.91
AUC	0.855	0.943	0.837	0.843	0.937	0.820
灵敏度	0.97	0.99	0.98	0.97	0.99	0.99
特异度	0.84	0.92	0.78	0.81	0.87	0.75

注: AUC=受试者工作特征曲线下面积, SVM=支持向量机, BPNN=BP 神经网络。

表 2 DFUs 复发影响因素的多因素 Logistic 回归分析赋值表

Table 2 Assignment table for multivariable Logistic regression analysis of influencing factors of DFUs recurrence

变量	赋值
DFUs 复发	是 =1, 否 =0
BMI	实测值
独居	是 =1, 否 =0
糖尿病病程	实测值
吸烟史	有 =1, 无 =0
饮酒史	有 =1, 无 =0
受累足趾截肢史	有 =1, 无 =0
足溃疡分级	I 级 =1, II 级 =2, III 级 =3, IV 级 =4, V 级 =5
踝肱指数	良 =0, 差 =1
糖化血红蛋白	实测值
溃疡位置在脚底	是 =1, 否 =0
足趾受累	是 =1, 否 =0
足部存在行走障碍	是 =1, 否 =0
骨髓炎	有 =1, 无 =0
多重耐药菌感染	有 =1, 无 =0
下肢动脉粥样硬化	有 =1, 无 =0
糖尿病周围神经病变	有 =1, 无 =0
足部自我管理行为	实测值
DFUs 风险感知水平	实测值

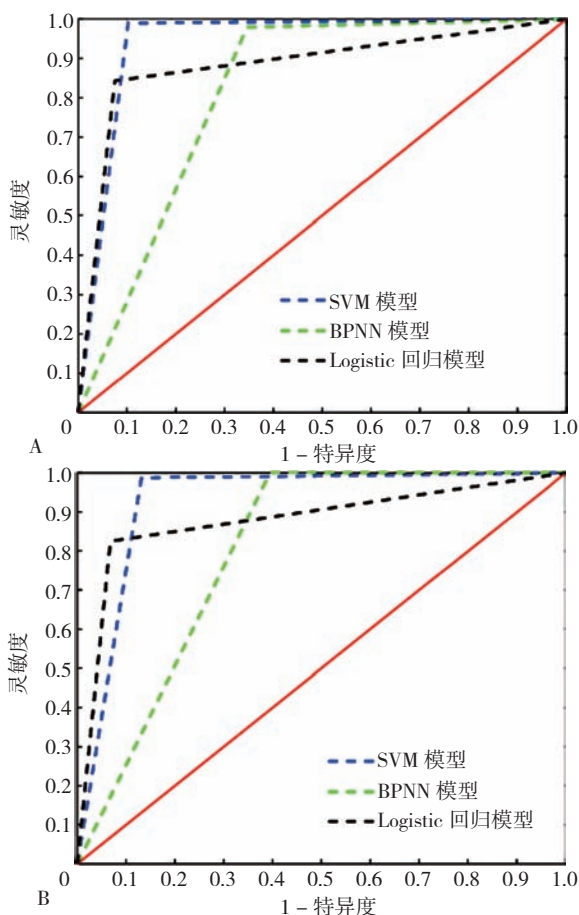
DFUs 复发高危因素在线筛查表, 短期随访(出院后 1 年) DFUs 患者的复发状况, 运用机器学习算法探讨 DFUs 复发的危险因素, 构建出复发风险预测模型以便实现快速筛查, 旨在为糖尿病高危足 DFUs 复发提供全面的精准干预。本研究 DFUs 患者出院后 1 年内的复发率为 29.74%, 高于吕静等^[5] DFUs 1 年复发率为 26.9% 的研究结果, 低于 WANG 等^[16] 报道的治愈后 DFUs 1 年内再发病率为 31.6%, 且低于 HICKS 等^[17] 的研究结果。目前各区域性医疗机构有关 DFUs 复发率的报道有所不同^[18], 究其原因与医疗卫生技术、居民生活经济水平、区域性糖尿病防治教育普及程度、糖尿病高危足筛查及精准诊疗等因素有关。提前对高危足进行筛查和防治对预防 DFUs 复发尤为重要^[19]。通过数据挖掘技术形成的预测模型有助于在院外尽早识别 DFUs 复发的危险因素, 可及早发现 DFUs 好发的患病高危人群, 对高危人群实施干预措施, 对于有严重复杂并发症的 DFUs, 可启动快速转诊路径, 送往专业的足病治疗中心, 对提高人群健康水平意义重大。

本研究从简单个人水平和复杂临床水平分别筛选出与 DFUs 复发相关的影响因素, 运用机器学习算法建立最优预测模型, 最后进行模型评价。多因素 Logistic 回归分析结果显示, BMI、糖尿病病程、吸烟史、足溃疡分级、糖化血红蛋白、溃疡位置在脚底、足部自我管理行为和 DFUs 风险感知水平是 DFUs 患者 1 年内 DFUs 复发的影响因素, 与大多数研究结果相符^[5, 8, 10], 为 DFUs 复发的早期筛查与防治提供了可行干预方案。

目前, 许多关于 DFUs 复发风险的研究常使用传统或常见的机器学习算法建立模型, 很少采用多种机器学习算法进行多样化建模来比较模型预测性能。本研究预测模型指标筛选共纳入 22 个变量, 将多因素 Logistic 回归分析中有统计学意义的 8 个因素纳入模型构建, 分别建立 Logistic 回归、SVM 与 BPNN 风险预测模型, 并且在测试集中评估了三种模型的预测效能, 研究结果显示, 三种预测模型 AUC 均 >0.70, 正确率均 >80%, 表示具有较高的准确性和良好的预测能力^[20]。训练集与测试集中 SVM 模型的各项评价指标均高于 Logistic 回归和 BPNN 模型, BPNN 相比 Logistic 模型的召回率较低, 但查准率更准确。仅用单个指标无法对算法的效果进行总体评价, 为调和查准率与召回率的差异, 更加准确地评价模型的综合性能, 本研究使用 F1 指数对各模型预测效果进行综合评价, 其结果为 BPNN>Logistic 回归, 可能与样本量和变量数影响预测模型的效能有关系, 本研究中测试集的预测数据量有限, 限制了 BPNN 算法的优势。综合上述指标, 基于 SVM 模型构建的预测模型具有良好诊断效能且更为稳定, 在预测 DFUs 复发方面的正确率、灵敏度和 AUC 等性能指标上优于其他两种模型, 提示通过结合临床病例资料和患者报告结局资料所建立的预测模型可作为有效预测 DFUs 复发的辅助决策工具, 具有良好的应用前景。

机器学习算法在获取数据体征与分析复杂数据方面具有显著优势, 从早期经典的逻辑回归统计、决策树分析等传统方法, 到新兴的深度学习神经网络学习算法, 机器学习已被广泛应用于各个领域^[21]。不同模型具有各自的优势, 与 SVM 模型相比, BPNN 模型存在局部极小值和收敛速度慢等不足, 与 Logistic 回归模型相比, SVM 模型不直接依赖于数据分布, 而对于多维、非线性医疗数据, SVM 在分类性能、泛化能力、建模计算量等方面具有明显优势^[22]; Logistic 回归模型中的回归系数虽可解释各个影响因素的流行病学意义, 但是却无法对存在严重共线性的变量做出明确解释^[23]。

综上所述, 基于 SVM 的 DFUs 复发风险预测模型效果最优, 运用于 DFUs 临床评估将有助于提高易复发人群筛查率, 可以提前给予复发风险预警, 利于降低疾病负担且节约医疗成本, 为 DFUs 防治提供一定的临床



注: A 为训练集受试者工作特征 (ROC) 曲线, B 为测试集 ROC 曲线; SVM=支持向量机, BPNN=BP 神经网络。

图 1 Logistic 回归、SVM 和 BPNN 模型训练集和测试集预测 DFUs 复发风险的 ROC 曲线

Figure 1 ROC curves of Logistic regression, SVM and BPNN recurrence risk prediction models for DFUs in the the training and test sets

指导和帮助。除此之外，在临床科研方面，通过预测复发风险实现对患者的分组，从而控制组间的可比性。由于本研究仅关注了 DFUs 的短期预后结局，今后可扩充样本量，纳入不同地域的人群，开展多中心和大样本的队列研究，探索新的输入变量，优化模型准确度，进一步验证预测模型的效能，并采用多种机器学习算法，如决策树和随机森林等进行比较，联合预测以综合判断模型的预测价值，构建更权威的模型算法提升研究的准确性和推广性。

作者贡献：张娟提出研究思路，设计研究方案，研究的实施与可行性分析，包括复发风险预测模型建模思路，数据收集，文章的构思和论文撰写；张娟、李海芬、李小曼、姚苗、马惠珍负责筛选研究对象，数据采集和整理，构建预测模型统计学处理，结果的分析与解释，图表的设计和制作；马强负责论文的修订，负责文章的质量控制及审校，提供研究经费及材料支持，对文章整体负责，监督管理。

本文无利益冲突。

参考文献

- [1] SAEEDI P, PETERSOHN I, SALPEA P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition [J]. *Diabetes Res Clin Pract*, 2019, 157: 107843. DOI: 10.1016/j.diabres.2019.107843.
- [2] TENTOLOURIS N, EDMONDS M E, JUDE E B, et al. Editorial: understanding diabetic foot disease: current status and emerging treatment approaches [J]. *Front Endocrinol (Lausanne)*, 2021, 12: 753181. DOI: 10.3389/fendo.2021.753181.
- [3] ZHOU X Q, LI M, XIAO M F, et al. ER β accelerates diabetic wound healing by ameliorating hyperglycemia-induced persistent oxidative stress [J]. *Front Endocrinol (Lausanne)*, 2019, 10: 499. DOI: 10.3389/fendo.2019.00499.
- [4] ARMSTRONG D G, BOULTON A J M, BUS S A. Diabetic foot ulcers and their recurrence [J]. *N Engl J Med*, 2017, 376 (24): 2367-2375. DOI: 10.1056/NEJMra1615439.
- [5] 吕静, 袁丽, 李饶, 等. 糖尿病足溃疡复发风险预测模型的构建 [J]. *护理研究*, 2022, 36 (6): 993-998. DOI: 10.12102/j.issn.1009-6493.2022.06.009.
- [6] CRAWFORD F, CEZARD G, CHAPPELL F M, et al. The development and validation of a multivariable prognostic model to predict foot ulceration in diabetes using a systematic review and individual patient data meta-analyses [J]. *Diabet Med*, 2018, 35 (11): 1480-1493. DOI: 10.1111/dme.13797.
- [7] AAN DE STEGGE W B, SCHUT M C, ABU-HANNA A, et al. Development of a prediction model for foot ulcer recurrence in people with diabetes using easy-to-obtain clinical variables [J]. *BMJ Open Diabetes Res Care*, 2021, 9 (1): e002257. DOI: 10.1136/bmjdr-2021-002257.
- [8] American Diabetes Association. Addendum. 9. pharmacologic approaches to glycemic treatment: Standards of medical care in diabetes-2020. *diabetes care* 2020; 43 (suppl. 1): S98-110 [J]. *Diabetes Care*, 2020, 43 (8): 1979. DOI: 10.2337/dc20-ad08a.
- [9] 李柏樟. 白癜风共病糖尿病风险预测模型的构建及验证 [D]. 西安: 空军军医大学, 2022.
- [10] BUS S A, LAVERY L A, MONTEIRO-SOARES M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update) [J]. *Diabetes Metab Res Rev*, 2020, 36 (Suppl 1): e3269. DOI: 10.1002/dmrr.3269.
- [11] HINCHLIFFE R J, FORSYTHE R O, APELQVIST J, et al. Guidelines of the international writing group on the diabetic foot on diagnosis [J]. *Diabetes/Metabolism Research and Reviews*, 2020, e3276. DOI: 10.1002/dmrr.3276.
- [12] CHIN Y F, HUANG T T. Development and validation of a diabetes foot self-care behavior scale [J]. *J Nurs Res*, 2013, 21 (1): 19-25. DOI: 10.1097/jnr.0b013e3182828e59.
- [13] 方蕾. 慢性病病人风险感知问卷的编制及其影响因素研究 [D]. 西安: 第四军医大学, 2015.
- [14] 马倩倩, 孙东旭, 石金铭, 等. 基于支持向量机与 XGboost 的成年人肿瘤患病风险预测研究 [J]. *中国全科医学*, 2020, 23 (12): 1486-1491. DOI: 10.12114/j.issn.1007-9572.2020.00.066.
- [15] PEIKES D, CHEN A, SCHORE J, et al. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials [J]. *JAMA*, 2009, 301 (6): 603-618. DOI: 10.1001/jama.2009.126.
- [16] WANG L M, GAO P, ZHANG M, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013 [J]. *JAMA*, 2017, 317 (24): 2515-2523. DOI: 10.1001/jama.2017.7596.
- [17] HICKS C W, CANNER J K, MATHIOUDAKIS N, et al. Incidence and risk factors associated with ulcer recurrence among patients with diabetic foot ulcers treated in a multidisciplinary setting [J]. *J Surg Res*, 2020, 246: 243-250. DOI: 10.1016/j.jss.2019.09.025.
- [18] FU X L, DING H, MIAO W W, et al. Global recurrence rates in diabetic foot ulcers: a systematic review and meta-analysis [J]. *Diabetes Metab Res Rev*, 2019, 35 (6): e3160. DOI: 10.1002/dmrr.3160.
- [19] 夏磊, 庄若, 吴玲, 等. 复发性糖尿病足溃疡相关危险因素的 Meta 分析 [J]. *中华现代护理杂志*, 2022, 28 (9): 1143-1148. DOI: 10.3760/cma.j.cn115682-20210802-03422.
- [20] 周罗晶, 刘昕, 高巨. 无痛结肠镜检查术患者术后疲劳综合征人工神经网络预测模型的建立 [J]. *中华麻醉学杂志*, 2021, 41 (4): 397-400. DOI: 10.3760/cma.j.cn131073.20201202.00403.
- [21] 吴秋硕, 陆宗庆, 刘瑜, 等. 2010 至 2020 年机器学习方法在重症医学领域应用进展的可视化分析 [J]. *中国急救医学*, 2021, 41 (11): 959-965. DOI: 10.3969/j.issn.1002-1949.2021.11.008.
- [22] ZHAN J B, WANG Q, CHENG J, et al. Current status of severe fever with thrombocytopenia syndrome in China [J]. *Virol Sin*, 2017, 32 (1): 51-62. DOI: 10.1007/s12250-016-3931-1.
- [23] 汪奇, 刘尚全. 2 型糖尿病患者亚临床动脉粥样硬化的多层人工神经网络分类预测模型的构建 [J]. *中国全科医学*, 2021, 24 (36): 4612-4617. DOI: 10.12114/j.issn.1007-9572.2021.00.537.

(收稿日期: 2023-03-22; 修回日期: 2023-05-23)

(本文编辑: 康艳辉)