

Logistic 回归和 BP 神经网络及支持向量机模型对老年 COPD 患者 30 d 内急性加重再入院风险的预测性能比较

张瑞¹ 常艳¹ 张晓娜¹ 路露² 丁玲³ 芦鸿雁¹

¹宁夏医科大学总医院护理部,银川 750004; ²宁夏医科大学总医院普胸外科,银川 750004;

³宁夏医科大学总医院呼吸内科,银川 750004

通信作者:芦鸿雁, Email: hyalu@hotmail.com

【摘要】 目的 比较 Logistic 回归、BP 神经网络及支持向量机 3 种模型对老年慢性阻塞性肺疾病 (COPD) 患者 30 d 内急性加重再入院风险的预测性能,以期对再入院高风险患者筛查及预防提供科学依据。方法 制作 COPD 患者调查问卷表,包括一般资料调查表、改良英国医学研究委员会呼吸困难量表 (mMRC)、日常生活能力评估量表 (ADL)、老年抑郁量表、微型营养评定量表 (MNA-SF) 及 COPD 评估量表 (CAT),采用便利抽样法,选择宁夏地区 13 所综合医院 2019 年 4 月至 2020 年 8 月呼吸内科治疗的老年 COPD 患者作为调查对象,并随访至出院后 30 d。探讨患者再入院影响因素,基于影响因素构建 Logistic 回归模型、BP 神经网络模型和支持向量机模型,按照训练集与测试集 7:3 的比例分为训练集样本和测试集样本,分别通过查准率、召回率、正确率、F1 指数及受试者工作特征曲线下面积 (AUC) 对模型预测效能进行比较。结果 共调查 1 120 例患者,其中非再入院患者 879 例,再入院患者 241 例。单因素分析显示,非再入院患者与再入院患者的年龄、文化程度、吸烟情况、糖尿病和冠心病比例、过去 1 年因 COPD 急性加重住院次数、季节因素及长期家庭氧疗、规律用药、康复锻炼比例、病程、ADL、抑郁状况、mMRC、营养状况比较差异均有统计学意义。二元 Logistic 回归分析显示,文化程度、吸烟情况、冠心病史、过去 1 年因 COPD 急性加重住院次数、季节因素、是否长期家庭氧疗、是否规律用药、营养状况是老年 COPD 患者 30 d 急性加重再入院的影响因素。训练集显示,Logistic 回归模型、BP 神经网络模型及支持向量机模型的查准率分别为 70.95%、76.51%、84.78%,召回率分别为 79.55%、86.36%、88.64%,正确率分别为 87.81%、90.81%、93.82%,F1 指数分别为 0.75、0.81、0.87, AUC 分别为 0.850、0.893、0.921。测试集显示,Logistic 回归模型、BP 神经网络模型及支持向量机模型的查准率分别为 78.38%、80.65%、88.57%,召回率分别为 70.73%、60.98%、75.61%,正确率分别为 85.82%、84.40%、90.07%,F1 指数分别为 0.74、0.69、0.82, AUC 分别为 0.814、0.775、0.858。结论 与 Logistic 回归模型和 BP 神经网络模型相比,支持向量机模型的预测效果更好,可以有效预测老年 COPD 患者 30 d 内急性加重再入院风险。

【关键词】 老年; 慢性阻塞性肺疾病; Logistic 回归; BP 神经网络; 支持向量机

基金项目:宁夏回族自治区重点研发计划项目 (2021BEG03116)

DOI: 10.3760/cma.j.cn121430-20220303-00198

Comparison of the predictive performance of Logistic regression, BP neural network and support vector machine model for the risk of acute exacerbation of readmission in elderly patients with chronic obstructive pulmonary disease within 30 days

Zhang Rui¹, Chang Yan¹, Zhang Xiaona¹, Lu Lu², Ding Ling³, Lu Hongyan¹

¹Department of Nursing, the General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia Hui Autonomous Region, China; ²Department of General Thoracic Surgery, the General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia Hui Autonomous Region, China; ³Department of Respiratory Medicine, the General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia Hui Autonomous Region, China

Corresponding author: Lu Hongyan, Email: hyalu@hotmail.com

【Abstract】 Objective To compare the effectiveness of Logistic regression, BP neural network and support vector machine models in the prediction of 30-day risk of readmission in elderly patients with an exacerbation of chronic obstructive pulmonary disease (COPD) and to provide a scientific basis for the screening and prevention of high-risk patients with readmission. **Methods** The COPD patient survey questionnaire was made, including the general data questionnaire, modified Medical Research Council dyspnea scale (mMRC), activities of daily living (ADL), the geriatric depression scale, the mini nutritional assessment-short form (MNA-SF) and COPD assessment test (CAT). Elderly COPD patients were selected from the department of respiratory medicine of 13 general hospitals in Ningxia from April 2019 to August 2020 by convenience sampling method, and they were followed up 30 days after discharge. To explore the risk factors of patient readmission, Logistic regression model, BP neural network model and support vector machine models were constructed based on the risk factors. According to the ratio of the training set to the testing set of 7:3, the model

was divided into the training set sample and the testing set sample. The prediction efficiency of the model was compared by the precision rate, recall rate and accuracy rate, F1 index and the area under the receiver operator characteristic curve (AUC). **Results** A total of 1 120 patients were investigated, including 879 non-readmission patients and 241 readmission patients. Univariate regression analysis showed that there were statistically significant differences in age, education level, smoking status, proportion of diabetes and coronary heart disease, hospitalization times of acute exacerbation of COPD in the past 1 year, seasonal factors and long-term home oxygen therapy, regular medication, proportion of rehabilitation exercise, course of disease, ADL, depression status, mMRC, nutritional status between non-readmission patients and readmission patients. Binary Logistic regression analysis showed that education level, smoking status, coronary heart disease, hospitalization times of acute exacerbation of COPD in the past 1 year, seasonal factors, whether long-term home oxygen therapy, whether regular medication, nutritional status were the risk factors for 30-day acute exacerbation of readmission in elderly patients with COPD. The training set showed that the accuracy rate of Logistic regression model, BP neural network model and support vector machine models were 70.95%, 76.51% and 84.78%, respectively. The recall rates were 79.55%, 86.36% and 88.64%, respectively. The accuracy rates were 87.81%, 90.81% and 93.82%, respectively. F1 indexes were 0.75, 0.81 and 0.87, respectively. The AUC were 0.850, 0.893 and 0.921, respectively. The testing set showed that the precision rate of Logistic regression model, BP neural network model and support vector machine model were 78.38%, 80.65% and 88.57%, respectively. The recall rates were 70.73%, 60.98% and 75.61%, respectively. The accuracy rates were 85.82%, 84.40% and 90.07%, respectively. F1 indexes were 0.74, 0.69 and 0.82, respectively. The AUC were 0.814, 0.775 and 0.858, respectively. **Conclusion** Comparing with Logistic regression and BP neural network, support vector machine model has better prediction effect, and can effectively predict the risk of acute exacerbation of readmission in elderly patients with COPD within 30 days.

【Key words】 Old age; Chronic obstructive pulmonary disease; Logistic regression; BP neural network; Support vector machine

Fund program: Key Research and Development Program of Ningxia Hui Autonomous Region (2021BEG03116)

DOI: 10.3760/cma.j.cn121430-20220303-00198

《慢性阻塞性肺疾病全球倡议(2021年)》提到,慢性阻塞性肺疾病(chronic obstructive pulmonary disease, COPD)是全球范围内发病率和病死率较高的慢性疾病之一,已成为公共健康的重大挑战^[1],而 COPD 急性加重是其疾病管理过程中的重要事件^[2]。有研究显示, COPD 患者 30 d 内急性加重再入院率为 6.70%~7.54%^[3-5]。COPD 30 d 内急性加重再入院者因其急性加重周期短,不仅严重损害肺功能,增加死亡风险,还占据了大量的医疗资源^[6]。相关研究表明,预测模型对 COPD 患者 30 d 内急性加重再入院高风险人群进行早期识别具有良好的预测效能,但哪种方式构建的模型更适合再入院风险筛查尚无定论^[7]。目前,国内外学者主要基于 Logistic 回归对 COPD 患者的再入院风险进行预测^[8-10], Goto 等^[11]虽就 Logistic 回归模型与决策树算法模型、深度神经网络模型进行比较,但仍缺乏特异性的预测工具。BP 神经网络模型和支持向量机作为新兴的模型构建方法,是数据挖掘分类中的两项重要技术,其预测性能已逐渐被医学领域认可,目前尚缺乏两种模型在 COPD 患者再入院风险预测中的相关报道^[12]。

本研究利用经典 Logistic 回归与 BP 神经网络、支持向量机构建老年 COPD 患者 30 d 内急性加重再入院风险预测模型,并比较 3 种模型的风险预测

性能,以期筛选出合适的模型,为早期识别再入院高风险患者提供依据。

1 资料与方法

1.1 研究对象:于 2019 年 4 月至 2020 年 8 月选择宁夏地区 13 所综合医院呼吸内科 COPD 患者进行问卷调查,2019 年 5 月至 2020 年 9 月对患者进行随访,随访内容包括出院后 30 d 内是否因急性加重再入院、是否规律用药及是否进行康复锻炼等,并记录随访季节,统计患者在春、夏、秋、冬 4 个季节再入院与非再入院人数。根据出院后 30 d 内患者是否因急性加重再次入院分为非再入院组和再入院组,最终纳入研究对象 1 120 例。

1.1.1 纳入标准:①年龄≥60岁;②符合《慢性阻塞性肺疾病全球倡议(2019版)》中诊断标准,且病情处于稳定期的患者。

1.1.2 排除标准:存在严重视听障碍而无法交流者。

1.1.3 伦理学:本研究符合医学伦理学标准,获得宁夏医科大学总医院伦理委员会批准(审批号:2020-643),所有研究对象均知情同意并自愿参与本研究。

1.2 研究方法

1.2.1 研究工具

1.2.1.1 一般资料调查表:根据研究目的及内容,由研究者通过回顾文献、预调查等方式自行设计一般

资料调查表,包括性别、年龄、共病情况、过去1年因COPD急性加重住院次数等。

1.2.1.2 改良英国医学研究委员会呼吸困难量表(modified Medical Research Council dyspnea scale, mMRC): mMRC量表用于评估COPD患者的呼吸困难严重程度^[13]。共分为5级,仅剧烈活动时感到呼吸困难为0级;平地快步行走或步行小坡时出现气促为1级;因气促,平地行走慢于同龄人或需停下休息为2级;平地行走约100 m或数分钟后需停下喘气为3级;因严重呼吸困难不能离家,或在穿脱衣服时出现呼吸困难为4级^[13]。

1.2.1.3 日常生活能力评估量表(activities of daily living, ADL): ADL量表用于评价患者日常生活活动能力,包括进食、修饰、上下楼梯等10项内容。根据是否需要帮助及帮助的程度将其分为15、10、5、0分共4个等级。总分100分,81~100分为生活完全自理;61~80分为轻度功能障碍,能够独立完成日常活动;41~60分为中度功能障碍,生活需要帮助;≤40分为重度功能障碍或完全依赖,多数日常生活活动不能完成或需他人照料^[14]。

1.2.1.4 老年抑郁量表(geriatric depression scale, GDS): GDS量表用于评价老年人近1周抑郁状态,包括生活满意情况、记忆力、心情等30个条目。回答“是”得1分,“否”得0分,部分条目反向计分。总分30分,0~10分为正常,即无抑郁;11~20分为可能有抑郁症状;21~30分为抑郁^[15]。

1.2.1.5 微型营养评定简表(mini-nutritional assessment short-form, MNA-SF): MNA-SF量表用于评价营养状态,包括体质量变化、活动能力、体质量指数(body mass index, BMI)等6个问题。总分14分,12~14分为营养状况正常,8~11分为有营养不良风险,0~7分为营养不良^[16]。

1.2.1.6 COPD评估量表(COPD assessment test, CAT): CAT量表用于评估COPD严重程度,包括咳嗽、咳痰、胸闷等8个问题。总分40分,>10分为病情轻微,10~20分为病情中等,20~30分为病情严重,>30分为病情非常严重^[13]。

1.2.2 质量控制: 研究者对48例老年COPD患者进行预调查以完善问卷。统一培训调查员,包括研究目的、问卷填写要求等,由培训合格的调查员逐条陈述并根据患者的回答填写问卷。检查有无漏项,核对无误后收回。研究数据采用双人录入,剔除有错误、漏项超过20%及完全雷同的问卷。

1.3 统计学方法: 使用Epidata 3.1录入数据,采用SPSS 25.0及MATLAB R2020b软件进行统计分析。计量资料不符合正态分布,以中位数(四分位数) $[M(Q_L, Q_U)]$ 表示,采用Mann-Whitney U 检验;计数资料以频数和百分比表示,采用 χ^2 检验;等级资料比较采用Wilcoxon W 检验。采用SPSS 25.0软件构建Logistic回归风险预测模型, MATLAB R2020b软件构建BP神经网络模型及支持向量机模型。绘制受试者工作特征曲线(receiver operator characteristic curve, ROC曲线),用ROC曲线下面积(area under the ROC curve, AUC)、正确率、查准率、召回率及F1指数评价模型的预测能力。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 一般资料(表1): 共调查1120例老年COPD患者,其中男性683例(61.0%),女性437例(39.0%);年龄60~96岁;非再入院患者879例,再入院患者241例。

2.2 老年COPD患者30 d内急性加重再入院风险预测模型的构建

2.2.1 风险预测模型输入变量筛选: 单因素分析结果显示,非再入院患者与再入院患者的年龄、康复锻炼比例、病程、ADL等比较差异均有统计学意义(均 $P<0.05$;表1)。以单因素分析中有统计学意义的因素为自变量,是否再入院为因变量,纳入二元Logistic回归分析,结果表明,是否长期家庭氧疗、是否规律用药及季节因素等8个因素是老年COPD患者30 d内急性加重再入院的影响因素(均 $P<0.05$;表2)。将Logistic回归分析有意义的8个因素作为预测模型的输入变量。

2.2.2 风险预测模型的构建: 纳入Logistic回归分析有意义的8个因素,构建Logistic回归、BP神经网络及支持向量机风险预测模型,各模型训练集取70%,即784例样本用于训练;测试集取30%,即336例样本用于预测。

2.3 老年COPD患者30 d内急性加重再入院风险预测模型的预测性能(图1;表3): 在训练集中,支持向量机模型查准率、召回率、正确率、F1指数和AUC均优于BP神经网络模型及Logistic回归模型;在测试集中,支持向量机模型的各项评价指标均优于其他两种模型,Logistic回归模型的召回率、正确率、F1指数及AUC均高于BP神经网络模型,但查准率低于BP神经网络模型。

表1 是否再次入院两组老年 COPD 患者临床特征的单因素分析

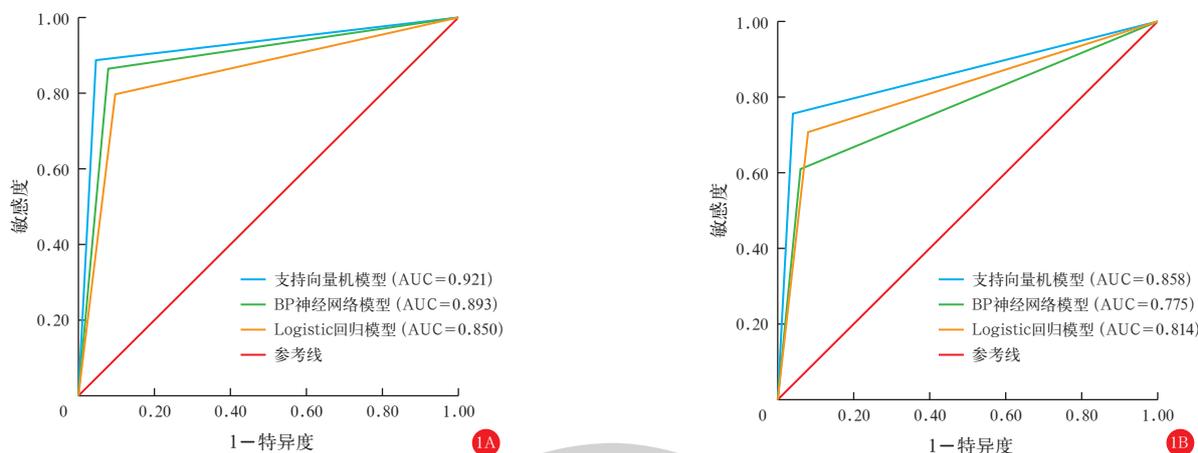
项目	非再入院组 (n=879)	再入院组 (n=241)	χ^2 值	P 值	项目	非再入院组 (n=879)	再入院组 (n=241)	χ^2 / Z 值	P 值
年龄 [例 (%)]			8.868	0.012	季节因素 [例 (%)]			43.608	<0.001
60~69 岁	362 (41.2)	77 (32.0)			春季	173 (19.7)	58 (24.1)		
70~79 岁	370 (42.1)	108 (44.8)			夏季	218 (24.8)	23 (9.5)		
≥80 岁	147 (16.7)	56 (23.2)			秋季	247 (28.0)	52 (21.6)		
性别 [例 (%)]			0.350	0.554	冬季	241 (27.5)	108 (44.8)		
男性	540 (61.4)	143 (59.3)			长期家庭氧疗 [例 (%)]	266 (30.3)	48 (19.9)	10.032	0.002
女性	339 (38.6)	98 (40.7)			规律用药 [例 (%)]	225 (25.6)	35 (14.5)	13.014	<0.001
BMI [例 (%)]			6.350	0.096	康复锻炼 [例 (%)]	291 (33.1)	55 (22.8)	9.370	0.002
<18.5 kg/m ²	150 (17.1)	38 (15.8)			动脉血二氧化碳分压 [mmHg, M(Q _L , Q _U)]	41.15 (36.00, 48.00)	41.90 (36.00, 49.05)	-0.698	0.485
18.5~23.9 kg/m ²	348 (39.6)	97 (40.2)			动脉血氧分压 [mmHg, M(Q _L , Q _U)]	66.35 (58.78, 77.00)	65.90 (59.90, 81.25)	-0.426	0.670
24.0~27.9 kg/m ²	261 (29.7)	59 (24.5)			经皮血氧饱和度 [M(Q _L , Q _U)]	0.92 (0.90, 0.94)	0.92 (0.90, 0.94)	-0.762	0.446
≥28.0 kg/m ²	120 (13.6)	47 (19.5)			呼吸频率 [次/min, M(Q _L , Q _U)]	21 (20, 22)	21 (20, 22)	-0.779	0.436
婚姻状况 [例 (%)]			4.283	0.117	病程 [年, M(Q _L , Q _U)]	4.0 (1.0, 10.0)	6.0 (3.0, 10.5)	-1.774	<0.001
已婚	652 (74.2)	188 (78.0)			ADL [例 (%)]			-2.980	0.003
未婚	66 (7.5)	22 (9.1)			完全独立	150 (17.1)	17 (7.1)		
丧偶	161 (18.3)	31 (12.9)			轻度功能障碍	510 (58.0)	140 (58.1)		
居住方式 [例 (%)]			0.251	0.617	中度功能障碍	146 (16.6)	50 (20.7)		
非独居	793 (90.2)	220 (91.3)			重度功能障碍 或完全依赖	73 (8.3)	34 (14.1)		
独居	86 (9.8)	21 (8.7)			抑郁状况 [例 (%)]			-2.845	0.004
文化程度 [例 (%)]			16.568	0.002	无抑郁	400 (45.5)	81 (33.6)		
文盲	360 (41.0)	124 (51.5)			轻度抑郁	377 (42.9)	101 (41.9)		
小学	259 (29.5)	75 (31.1)			中、重度抑郁	102 (11.6)	59 (24.5)		
初中	162 (18.4)	24 (10.0)			mMRC [例 (%)]			-4.071	<0.001
高中、中专	76 (8.6)	16 (6.6)			0 级	104 (11.8)	9 (3.7)		
大专及以上	22 (2.5)	2 (0.8)			1 级	221 (25.2)	38 (15.8)		
职业 [例 (%)]			0.311	0.577	2 级	315 (35.8)	53 (22.0)		
有污染接触史	28 (3.2)	6 (2.5)			3 级	196 (22.3)	89 (36.9)		
无污染接触史	851 (96.8)	235 (97.5)			4 级	43 (4.9)	52 (21.6)		
吸烟情况 [例 (%)]			51.240	<0.001	营养状况 [例 (%)]			-8.612	<0.001
从不吸烟	520 (59.2)	121 (50.2)			正常营养状态	285 (32.4)	22 (9.1)		
戒烟	319 (36.3)	76 (31.5)			营养不良风险	410 (46.6)	92 (38.2)		
吸烟	40 (4.5)	44 (18.3)			营养不良	184 (21.0)	127 (52.7)		
共病数量 [例 (%)]			0.071	0.790	CAT [例 (%)]			-1.774	0.076
≤2 种	725 (82.5)	197 (81.7)			病情轻微	29 (3.3)	11 (4.6)		
>2 种	154 (17.5)	44 (18.3)			病情中等	305 (34.7)	75 (31.1)		
高血压 [例 (%)]	384 (43.7)	119 (49.4)	2.476	0.116	病情严重	453 (51.5)	133 (55.2)		
糖尿病 [例 (%)]	168 (19.1)	67 (27.8)	8.611	0.003	病情非常严重	92 (10.5)	22 (9.1)		
冠心病 [例 (%)]	220 (25.0)	40 (16.6)	7.542	0.006					
低蛋白血症 [例 (%)]	15 (1.7)	8 (3.3)	2.447	0.118					
过去 1 年因 COPD 急性加重住院次数 [例 (%)]			35.901	<0.001					
<2 次	526 (59.8)	92 (38.2)							
≥2 次	353 (40.2)	149 (61.8)							

注: COPD 为慢性阻塞性肺疾病, BMI 为体质指数, ADL 为日常生活能力评估量表, mMRC 为改良英国医学研究委员会呼吸困难量表, CAT 为慢性阻塞性肺疾病评估量表; 1 mmHg ≈ 0.133 kPa

表2 老年 COPD 患者 30 d 内急性加重再入院影响因素的二元 Logistic 回归分析

影响因素	β 值	s_e	χ^2 值	P 值	OR 值	95%CI	影响因素	β 值	s_e	χ^2 值	P 值	OR 值	95%CI
常量	-8.939	1.069	69.916	0.000	0.000		家庭氧疗	-1.033	0.325	10.077	0.002	0.356	0.188 ~ 0.674
年龄	0.193	0.176	1.197	0.274	1.213	0.859 ~ 1.713	规律用药	-1.182	0.355	11.095	0.001	0.307	0.153 ~ 0.615
文化程度	-0.379	0.138	7.518	0.006	0.684	0.522 ~ 0.897	康复锻炼	-0.432	0.294	2.166	0.141	0.649	0.365 ~ 1.154
吸烟情况	1.008	0.194	26.902	0.000	2.740	1.872 ~ 4.010	病程	0.006	0.015	0.186	0.666	1.006	0.978 ~ 1.036
过去 1 年因 COPD 急性加重住院次数	1.308	0.264	24.615	0.000	3.700	2.207 ~ 6.204	ADL	-0.221	0.190	1.361	0.243	0.801	0.552 ~ 1.162
季节因素			16.660	0.001			抑郁状况	0.095	0.187	0.255	0.613	1.099	0.761 ~ 1.587
春季	1.319	0.425	9.654	0.002	3.740	1.627 ~ 8.597	mMRC	0.026	0.132	0.038	0.846	1.026	0.792 ~ 1.329
秋季	1.167	0.421	7.668	0.006	3.212	1.406 ~ 7.336	营养状况	1.506	0.218	47.573	0.000	4.508	2.938 ~ 6.915
冬季	1.635	0.404	16.403	0.000	5.130	2.325 ~ 11.317	冠心病	0.929	0.348	7.138	0.008	2.533	1.281 ~ 5.009
							糖尿病	0.568	0.409	1.930	0.165	1.764	0.792 ~ 3.930

注: 季节因素中夏季为参考分类; COPD 为慢性阻塞性肺疾病, ADL 为日常生活能力评估量表, mMRC 为改良英国医学研究委员会呼吸困难量表, OR 为优势比, 95%CI 为 95% 可信区间; 空白代表无此项



注：ROC 曲线为受试者工作特征曲线，AUC 为 ROC 曲线下面积

图 1 Logistic 回归模型、BP 神经网络模型和支持向量机模型训练集(A)和测试集(B)的 ROC 曲线

表 3 Logistic 回归、BP 神经网络、支持向量机 3 种模型预测老年 COPD 患者 30 d 内急性加重再入院的预测性能

评价指标	训练集模型			测试集模型		
	Logistic 回归	BP 神经网络	支持向量机	Logistic 回归	BP 神经网络	支持向量机
查准率 (%)	70.95	76.51	84.78	78.38	80.65	88.57
召回率 (%)	79.55	86.36	88.64	70.73	60.98	75.61
正确率 (%)	87.81	90.81	93.82	85.82	84.40	90.07
F1 指数	0.75	0.81	0.87	0.74	0.69	0.82
AUC	0.850	0.893	0.921	0.814	0.775	0.858

注：COPD 为慢性阻塞性肺疾病，AUC 为受试者工作特征曲线下面积

3 讨论

3.1 老年 COPD 患者 30 d 内急性加重再入院现状及其影响因素：本研究显示，老年 COPD 患者 30 d 内急性加重再入院率为 21.52%，低于年龄 < 60 岁 COPD 患者的再入院率 (6.70% ~ 7.54%)^[3-4]。随着年龄的增长，机体免疫力低下、咳嗽反射减弱及营养不良等是 COPD 患者急性加重的常见诱因，使得老年 COPD 群体急性加重再入院率明显升高^[17]。通过探究老年 COPD 患者 30 d 内急性加重再入院的相关因素，发现文化程度、吸烟情况、冠心病史、过去 1 年因 COPD 急性加重住院次数、是否家庭氧疗、是否规律用药、营养状况及季节因素等 8 个因素会对患者再入院情况产生影响，这与相关研究结论^[18-22]一致。为降低患者再入院率，Lau 等学者^[3]在美国、澳大利亚等国家分别采用快速动脉闭塞量表 (rapid arterial occlusion evaluation, RACE)、LACE 指数对 COPD 患者再入院情况进行预测，并对两者的有效性进行检测；Donzć 等^[23]使用 HOSPITAL 评分量表对美国乃至国际多中心人群再入院情况进

行预测。此外，也有学者基于加速计的活动监测、预测量表及相关模型对 COPD 患者 30 d 内急性加重再入院风险进行预测并比较，研究结果均表明，基于模型进行的风险预测结果更可靠，但哪种方式构建的模型更适合再入院风险筛查尚无定论^[7-10]。Goto 等^[5]虽就 Logistic 回归模型与决策树算法模型、深度学习模型进行比较，但各预测模型存在不同的优缺点，目前尚无专门的风险预测模型用于识别 COPD 患者再入院风险。

3.2 支持向量机模型的预测性能更好：本研究预测模型构建指标筛选共纳入 28 个变量，最终将 Logistic 回归分析有意义的 8 个因素纳入模型构建，分别采用 Logistic 回归、BP 神经网络及支持向量机构建老年 COPD 患者 30 d 内急性加重再入院风险预测模型，通过比较 3 种模型的预测性能发现，支持向量机模型风险预测性能优于其他两种模型。在训练集中支持向量机模型的各项评价指标均高于 Logistic 回归模型和 BP 神经网络模型，其预测结果为支持向量机模型 > BP 神经网络模型 > Logistic 回归模型。在测试集中支持向量机模型的各项评价指标均优于其他两种模型，但 Logistic 回归模型的召回率高于 BP 神经网络模型，查准率低于 BP 神经网络模型。由于查准率和召回率是一组此消彼长的评价指标，仅用单个指标无法对算法的效果进行总体评价，为了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^[24]，其结果为 Logistic 回归模型 > BP 神经网络模型。这可能与研究的数据有较大的关系，研究数据的数据量、变量个数均有可

能影响预测模型的效能,本研究中测试集预测数据量有限,可能在一定程度上限制了 BP 神经网络算法的优势。综合上述指标,支持向量机模型的预测性能最优。

近年来,支持向量机在疾病风险预测方面应用较多,如余鑫等^[25]研究表明,支持向量机在宫颈癌筛查中具有较高的诊断价值;纪鹏慧等^[26]发现,支持向量机对肺癌的预测效果更佳,应用于临床诊断更有价值;杨磊等^[27]通过比较决策树、随机森林、Logistic 回归及支持向量机 4 种模型在非酒精性肝病中的预测效果,发现基于支持向量机算法的预测模型更具优势,可以有效地预测非酒精性肝病的患病风险。支持向量机作为机器学习中模式识别经典算法之一,是一种以结构风险最小原理建立的对小样本数据有较好预测结果的简化智能模型,是基于统计学习理论的解决小样本、高维度数据的算法,在医学研究中值得进一步推广应用。

综上所述,基于 Logistic 回归、BP 神经网络及支持向量机构建的风险预测模型中,支持向量机模型风险预测性能更好,可为今后再入院的早期预防提供参考资料。另外,构建预测模型所需要的 8 个变量在老年 COPD 患者随访中容易获得,由此构建的模型是预测患者 30 d 内急性加重再入院风险较为方便的工具,可以帮助医护人员有效评估和管理患者急性加重的风险,延缓疾病进程,降低各种并发症的发病率。但本研究未进行外部验证,今后需在不同人群、医院及地区进行验证,以期进一步验证模型的预测效能。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 李思其,高兴林.《慢性阻塞性肺疾病全球倡议》2021 年版更新解读[J]. 临床药物治疗杂志, 2021, 19 (5): 36-42. DOI: 10.3969/j.issn.1672-3384.2021.05.008.

[2] 陈晓灿,霍建民,胡古月. 预测慢性阻塞性肺疾病急性加重再入院风险的研究进展[J]. 临床肺科杂志, 2020, 25 (2): 293-297. DOI: 10.3969/j.issn.1009-6663.2020.02.031.

[3] Lau CS, Siracuse BL, Chamberlain RS. Readmission after COPD exacerbation scale: determining 30-day readmission risk for COPD patients[J]. Int J Chron Obstruct Pulmon Dis, 2017, 12: 1891-1902. DOI: 10.2147/COPD.S136768.

[4] Furumoto A, Ohkusa Y, Chen M, et al. Additive effect of pneumococcal vaccine and influenza vaccine on acute exacerbation in patients with chronic lung disease[J]. Vaccine, 2008, 26 (33): 4284-4289. DOI: 10.1016/j.vaccine.2008.05.037.

[5] Goto T, Faridi MK, Camargo CA Jr, et al. Time-varying readmission diagnoses during 30 days after hospitalization for COPD exacerbation[J]. Med Care, 2018, 56 (8): 673-678. DOI: 10.1097/MLR.0000000000000940.

[6] Perera PN, Armstrong EP, Sherrill DL, et al. Acute exacerbations of COPD in the United States: inpatient burden and predictors of costs and mortality[J]. COPD, 2012, 9 (2): 131-141. DOI: 10.3109/15412555.2011.650239.

[7] Hakim MA, Garden FL, Jennings MD, et al. Performance of the

LACE index to predict 30-day hospital readmissions in patients with chronic obstructive pulmonary disease[J]. Clin Epidemiol, 2017, 10: 51-59. DOI: 10.2147/CLEP.S149574.

[8] Annavarapu S, Goldfarb S, Gelb M, et al. Development and validation of a predictive model to identify patients at risk of severe COPD exacerbations using administrative claims data[J]. Int J Chron Obstruct Pulmon Dis, 2018, 13: 2121-2130. DOI: 10.2147/COPD.S155773.

[9] 张瑞,吴珍珍,常艳,等. 老年慢性阻塞性肺疾病患者 30 天内急性加重再入院风险预测模型的构建与验证[J]. 中国呼吸与危重监护杂志, 2021, 20 (7): 457-464. DOI: 10.7507/1671-6205.202011079.

[10] 明航,蔡坤坤,施丁莉,等. 基于回归分析法建立慢性阻塞性肺疾病急性加重风险预测模型与验证[J]. 中华危重病急救医学, 2021, 33 (1): 64-68. DOI: 10.3760/cma.j.cn121430-20200720-00534.

[11] Goto T, Jo T, Matsui H, et al. Machine learning-based prediction models for 30-day readmission after hospitalization for chronic obstructive pulmonary disease[J]. COPD, 2019, 16 (5-6): 338-343. DOI: 10.1080/15412555.2019.1688278.

[12] 林鑫,李晋,刘蕾,等. 二型糖尿病肾病风险预测模型比较[J]. 中华医学图书情报杂志, 2019, 28 (4): 41-45. DOI: 10.3969/j.issn.1671-3982.2019.04.007.

[13] 张庆,徐爱晖. CAT 评分与 mMRC 评分对慢性阻塞性肺疾病患者病情评估比较[J]. 临床肺科杂志, 2018, 23 (12): 2156-2159. DOI: 10.3969/j.issn.1009-6663.2018.12.006.

[14] Fuino M, Rudnytskyi I, Wagner J. On the characteristics of reporting ADL limitations and formal LTC usage across Europe[J]. Eur Actuar J, 2020, 10 (2): 557-597. DOI: 10.1007/s13385-020-00242-1.

[15] 阮顺莉,郭菊红,陈茜,等. 1025 名居家 60 岁以上老年人吞咽障碍现状及其影响因素分析[J]. 护理学报, 2017, 24 (20): 41-44. DOI: 10.16460/j.issn1008-9969.2017.20.041.

[16] 刘锐,陈伟伟,黄坤,等. MNA-SF 和 NRS 2002 评估老年恶性梗阻性黄疸病人营养状况及其影响因素的应用研[J]. 肠外与肠内营养, 2019, 26 (2): 74-78. DOI: 10.16151/j.1007-810x.2019.02.003.

[17] 方佳英,梁彩韵,丁赞. 慢性阻塞性肺疾病急性加重与季节变化及气象因素的相关性研究[J]. 汕头大学医学院学报, 2019, 32 (4): 231-233. DOI: 10.13401/j.cnki.jsunc.2019.04.010.

[18] Bernabeu-Mora R, García-Guillamón G, Valera-Novella E, et al. Frailty is a predictive factor of readmission within 90 days of hospitalization for acute exacerbations of chronic obstructive pulmonary disease: a longitudinal study[J]. Ther Adv Respir Dis, 2017, 11 (10): 383-392. DOI: 10.1177/1753465817726314.

[19] 伍建光,陈正贤,甘兵,等. 慢性阻塞性肺疾病患者 30 日再入院率与体重指数的关系分析[J]. 中国医学创新, 2018, 15 (9): 83-85. DOI: 10.3969/j.issn.1674-4985.2018.09.022.

[20] Kume H. Role of bronchodilators in therapy for COPD—mechanisms of LABA and LAMA on airway smooth muscle[J]. Nihon Rinsho, 2016, 74 (5): 813-819.

[21] Sethi S, Maloney J, Grove L, et al. Airway inflammation and bronchial bacterial colonization in chronic obstructive pulmonary disease[J]. Am J Respir Crit Care Med, 2006, 173 (9): 991-998. DOI: 10.1164/rccm.200509-1525OC.

[22] Collinsworth AW, Brown RM, James CS, et al. The impact of patient education and shared decision making on hospital readmissions for COPD[J]. Int J Chron Obstruct Pulmon Dis, 2018, 13: 1325-1332. DOI: 10.2147/COPD.S154414.

[23] Donzé JD, Williams MV, Robinson EJ, et al. International validity of the HOSPITAL score to predict 30-day potentially avoidable hospital readmissions[J]. JAMA Intern Med, 2016, 176 (4): 496-502. DOI: 10.1001/jamainternmed.2015.8462.

[24] 王洁,杨永忠,郑子薇,等. 石油工人代谢综合征的三种风险预测模型比较[J]. 现代预防医学, 2021, 48 (4): 613-618.

[25] 余鑫,晏沙沙,张曦,等. 常规体检项目构建的支持向量机模型在宫颈癌筛查中的价值研究[J]. 国际检验医学杂志, 2020, 41 (17): 2147-2150. DOI: 10.3969/j.issn.1673-4130.2020.17.022.

[26] 纪鹏慧,段书音,李尊税,等. 基于肿瘤标志群的支持向量机模型在肺癌辅助诊断中的价值[J]. 郑州大学学报(医学版), 2019, 54 (3): 398-402. DOI: 10.13705/j.issn.1671-6825.2018.11.047.

[27] 杨磊,聂艳武,朱凯,等. 不同机器学习算法在乌鲁木齐市 NAFLD 风险预测模型中的应用[J]. 现代预防医学, 2021, 48 (18): 3270-3276.

(收稿日期: 2022-03-03)