

DOI: 10. 12138/j. issn. 1671-9638. 20233331

论著·医院感染监测专题

常规检验数据挖掘对急性缺血性脑卒中并发卒中相关肺炎的预测价值

曾瑞璜, 陈智熙, 王小林, 秦昌宏

(叙永县中医医院检验科, 四川 叙永 646400)

[摘要] **目的** 通过对常规检验数据的挖掘分析, 充分发挥其剩余价值, 为急性缺血性脑卒中并发卒中相关性肺炎(SAP)提供辅助预测价值。**方法** 回顾性分析某院 2019 年 6 月—2021 年 6 月收治的急性缺血性脑卒中病例, 根据是否并发 SAP 分为 SAP 组和非 SAP 组, 以 7:3 比例分为训练集和测试集。收集 2021 年 7 月—2022 年 6 月收治的急性缺血性脑卒中病例作为验证集。通过最小绝对收缩和选择算子(LASSO)筛选出与 SAP 有关的检验参数, 使用训练集、测试集和验证集构建、验证联合预测列线图模型。预测模型分别采用受试者工作特征(ROC)曲线评估区分度, 校准曲线评估校准度, 临床决策曲线(DCA)评估临床实用性。将列线图模型布置成网页计算器, 提高临床实用价值。**结果** 共纳入 379 例急性缺血性脑卒中患者作为研究的基本人群, 其中 SAP 42 例, 发病率为 11.08%; 以 7:3 分配方式, 分为训练集 265 例, 测试集 114 例。2021 年 7 月—2022 年 6 月收治的 157 例急性缺血性脑卒中病例作为验证集, 其中 SAP 24 例, 发病率为 15.29%。LASSO 筛选出 5 个与 SAP 有关的检验参数分别为中性粒细胞(NEU)、淋巴细胞(LYM)、前清蛋白(PA)、纤维蛋白原(Fib)、D-二聚体; 校准曲线显示训练集和测试集的预测概率和实际概率相一致, 具有较好的校准度; DCA 曲线显示训练集高风险阈值为 0~0.75, 净获益为 0~0.11。测试集高风险阈值为 0~0.65, 净获益为 0~0.11, 具有较好的临床实用性。ROC 曲线显示全数据集联合预测曲线下面积(AUC)为 0.924, 灵敏度为 83.33%, 特异度为 87.24%。训练集联合预测 AUC 为 0.922, 灵敏度为 79.31%, 特异度为 91.95%。测试集联合预测 AUC 为 0.919, 灵敏度为 84.62%, 特异度为 86.14%, 均具有良好的区分性能。验证集联合预测 AUC 为 0.850, 灵敏度为 66.67%, 特异度为 89.47%, 模型具有较好的外部适用性。网页计算器布置在 <https://ww-rstudiomn.shinyapps.io/SAP-nomogram/> 上, 可通过二维码访问, 经过测试, 性能稳定。**结论** 通过常规检验数据的挖掘, 对急性缺血性脑卒中并发 SAP 的预测提供了一定临床价值, 为早期的治疗和干预提供依据。

[关键词] 数据挖掘; 卒中相关肺炎; 急性缺血性脑卒中; 列线图; 预测

[中图分类号] R181.3⁺2

Predictive value of routine testing data mining for acute ischemic stroke complicated with stroke-associated pneumonia

ZENG Rui-huang, CHEN Zhi-xi, WANG Xiao-lin, QIN Chang-hong (Department of Laboratory Medicine, Xuyong Hospital of Chinese Medicine, Xuyong 646400, China)

[Abstract] **Objective** Through mining and analyzing the routine test data to develop its residual value and provide auxiliary predictive value for acute ischemic stroke (AIS) complicated with stroke-associated pneumonia (SAP). **Methods** AIS patients admitted to a hospital from June 2019 to June 2021 were retrospectively analyzed, divided into SAP group and non-SAP group according to whether they were complicated with SAP, and subdivided into training set and testing set at a ratio of 7:3. AIS patients admitted to hospital from July 2021 to June 2022 were collected as the validation set. SAP-related test parameters were screened by the least absolute shrinkage and selection operator (LASSO). Nomogram model of the combined prediction was constructed and validated with training set, testing set and verification set. Discrimination and calibration of prediction model were assessed by receiver operating chara-

[收稿日期] 2022-09-05

[作者简介] 曾瑞璜(1992-), 男(汉族), 四川省叙永县人, 主管检验师, 主要从事临床检验诊断相关研究。

[通信作者] 秦昌宏 E-mail: 58736753@qq.com

characteristic (ROC) curve and calibration curve respectively, clinical practicability was assessed by decision curve analysis (DCA). Nomograph model was arranged into a web calculator to improve the clinical practical value. **Results** A total of 379 patients with AIS were taken as the basic population of the study, including 42 cases (incidence 11.08%) in SAP group. According to the 7:3 distribution method, 265 cases were divided in training set and 114 in testing set, 157 cases of AIS admitted from July 2021 to June 2022 were used as validation set, including 24 cases (incidence 15.29%) in SAP group. Five test parameters related to SAP were screened out by LASSO, namely neutrophil, lymphocyte, prealbumin, fibrinogen, and D-dimer. The calibration curve showed good calibration that the predicted probability of training set and testing set was consistent with the actual probability. DCA curve showed that the high risk threshold of training set was 0 - 0.75 and the net benefit was 0 - 0.11. The high risk threshold of testing set was 0 - 0.65, the net benefit was 0 - 0.11, with good clinical practicability. ROC curve showed that area under curve (AUC) predicted by full data set was 0.924, the sensitivity and specificity were 83.33% and 87.24% respectively. AUC predicted by training set was 0.922, the sensitivity and specificity were 79.31%, and 91.95% respectively. AUC predicted by testing sets was 0.919, sensitivity and specificity were 84.62% and 86.14% respectively, all of which had good discrimination performance. AUC predict by validation set was 0.850, sensitivity and specificity were 66.67% and 89.47% respectively. The model has good external applicability. The web calculator was arranged at <https://ww-rstudiomn.shinyapps.io/SAP-nomgram/>, which can be accessed via QR code. Test showed that the performance was stable. **Conclusion** The mining of routine test data provides a clinical value for the prediction of AIS complicated with SAP, thus provides a basis for early treatment and intervention.

[Key words] data mining; stroke-associated pneumonia; acute ischemic stroke; nomogram; prediction

心血管疾病是中国居民死亡的首要原因,患病率仍处于持续上升趋势,其中缺血性脑卒中的发病人数居首位,具有较高的致残率和致死率^[1]。肺炎是卒中后最常见的并发症,是院内死亡和卒中后不良结局的主要原因^[2-3]。因此早期诊断急性缺血性脑卒中并发卒中相关肺炎(stroke-associated pneumonia, SAP)具有重要意义。以往相关研究多采用单指标,评分量表及新型生物标志物等方式进行预测^[4-6],在预测效能和临床实用性方面具有一定局限性。而列线图作为一种模型可视化方式,在疾病的诊断和预测方面具有较好的临床实用性^[7-8],近年来随着预测计算器的兴起,使得列线图的使用和实用性得到极大的提高^[9],并且常规检验数据易获取、成本低,具有一定可操作性。因此本研究拟通过对常规检验数据进行挖掘分析,应用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)筛选相关检验参数,构建联合检测的列线图模型,布置网页计算器,充分发挥检验数据的剩余价值,为临床提供经济、方便、准确的预测模型。

1 资料与方法

1.1 研究对象 选取 2019 年 6 月—2021 年 6 月叙永县中医医院确诊为急性缺血性脑卒中的患者

作为研究的基本人群,2021 年 7 月—2022 年 6 月在该院确诊为急性缺血性脑卒中的患者作为外部验证人群。纳入标准:符合 SAP 的诊断标准^[10]。排除标准:(1)脑卒中后遗症期和出血性脑卒中;(2)严重肝、肾功能不全;(3)自身免疫性疾病;(4)血液疾病和肿瘤。

1.2 方法

1.2.1 资料收集 回顾性收集患者的性别、年龄、高血压病史、糖尿病史,及入院 24 h 内的相关实验室检测结果。使用迈瑞 BC-6900 检测血常规参数:白细胞(WBC)、中性粒细胞(NEU)、淋巴细胞(LYM)、血红蛋白(Hb)、红细胞体积分布宽度变异系数(RDW-CV)、血小板(PLT)。使用德国思塔高 STA R MAX 检测凝血相关项目参数:纤维蛋白原(Fib)和 D-二聚体。使用迈瑞 BS-2000M 检测生化相关参数:丙氨酸氨基转移酶(ALT)、天门冬氨酸氨基转移酶(AST)、 γ -谷氨酰转肽酶(γ -GGT)、清蛋白(Alb)、前清蛋白(PA)、尿素(Ur)、肌酐(Cr)、尿酸(UA)、葡萄糖(Glu)、甘油三酯(TG)、总胆固醇(TC)、高密度脂蛋白胆固醇(HDL-C)、低密度脂蛋白胆固醇(LDL-C)、载脂蛋白 A1(APOA1)、载脂蛋白 B(APOB)、同型半胱氨酸(Hcy),脂蛋白 a[Lp(a)]。试剂盒和校准品分别购自深圳迈瑞和德国思塔高公司。

1.2.2 LASSO 筛选检验参数 LASSO 通过构造

一个惩罚函数,控制参数 lambda 对高纬度特征变量进行压缩,将不重要特征变量系数压缩为 0,留下相关性较强的特征变量。参数设置:模型设置为二项式分布(“family-binomial”),衡量模型指标函数为 -2 倍的对数似然值(type. measure = deviance),通过 10 折交叉验证选择最优 lambda 参数。

1.2.3 联合检测列线图预测模型构建与验证 采用训练集、测试集和验证集将 LASSO 筛选出的检验项目参数通过回归系数加权后对模型进行可视化分析与验证。每个指标根据相应权重具有不同的评分范围,根据指标得分总和从而获得患者的患病风险概率。构建的列线图模型通过 Hosmer-Lemeshow 检验分析列线图的拟合度,绘制校准曲线分析模型的校准度,采用受试者工作特征(receiver operating characteristic,ROC)曲线判断模型预测效能,评估模型的区分度。通过临床决策曲线(decision curve analysis,DCA)判断模型的净获益,评估模型的临床实用性。

1.2.4 布置网页计算器 通过将传统列线图以网页计算器形式布置到 shinyapps.io 网页上,临床医生通过访问相应网址或者二维码能方便、快捷的使

用该计算器进行风险预测,提高临床实用性。

1.3 统计学方法 应用 SPSS 19.0 和 R 软件,版本 4.2.1(<https://www.r.project.org>)进行统计分析。计数资料采用例数和百分比[n(%)]表示,组间比较采用 χ^2 检验。计量资料满足正态分布的数据,采用均数 ± 标准差($\bar{x} \pm s$)表示,组间比较采用 *t* 检验。不符合正态分布的数据采用中位数(四分位数)[*M*(*P*₂₅, *P*₇₅)]表示,组间比较采用 *Mann-Whitney U* 检验。LASSO 筛选参数使用“glmnet”软件包,列线图构建使用“rms”软件包,校准曲线绘制使用“riskRegression”软件包,DCA 绘制使用“rmda”软件包。网页计算器布置使用“DynNom”软件包。区分度评估采用 ROC 曲线。*P* ≤ 0.05 为差异具有统计学意义。

2 结果

2.1 研究人群基本情况 本次研究共纳入 379 例急性缺血性脑卒中患者作为研究的基本人群,其中 SAP 42 例,发病率为 11.08%。SAP 组与非 SAP 组患者的基本情况见表 1。

表 1 研究人群基本情况

Table 1 Basic conditions of studied population

项目	SAP 组(<i>n</i> = 42)	非 SAP 组(<i>n</i> = 337)	<i>t</i> / <i>Z</i> / χ^2	<i>P</i>
男性[例(%)]	23(54.76)	198(58.75)	0.245	0.621
年龄(岁)	67.55 ± 9.08	65.99 ± 9.43	1.012	0.312
高血压病[例(%)]	31(73.81)	263(78.04)	0.384	0.535
糖尿病史[例(%)]	11(26.19)	71(21.07)	0.578	0.447
WBC(× 10 ⁹ /L)	10.44 ± 4.34	7.42 ± 2.52	4.417	<0.001
NEU(× 10 ⁹ /L)	8.57 ± 4.09	5.22 ± 2.40	5.190	<0.001
LYM(× 10 ⁹ /L)	1.13 ± 0.50	1.58 ± 0.64	4.387	<0.001
Hb(g/L)	126.14 ± 20.44	134.69 ± 17.86	2.878	0.004
RDW-CV(%)	13.35(12.90,14.03)	13.10(12.60,13.70)	2.121	0.034
PLT(× 10 ⁹ /L)	232.69 ± 78.34	225.14 ± 69.93	0.651	0.516
ALT(U/L)	14.00(10.75,22.25)	17.00(13.00,24.00)	1.412	0.158
AST(U/L)	20.00(16.00,26.25)	20.00(17.00,24.00)	0.059	0.953
γ -GGT(U/L)	23.00(16.75,35.25)	28.00(20.00,44.00)	1.422	0.155
Alb(g/L)	37.29 ± 3.80	40.05 ± 3.90	4.333	<0.001
PA(mg/L)	185.17 ± 70.19	241.50 ± 59.54	5.663	<0.001
Ur(mmol/L)	6.76 ± 2.45	5.82 ± 1.89	2.932	0.004
Cr(μ mol/L)	87.45 ± 40.27	78.11 ± 25.29	1.467	0.149
UA(μ mol/L)	315.10 ± 102.70	324.55 ± 100.76	0.573	0.567

续表 1 (Table 1, Continued)

项目	SAP 组 (n = 42)	非 SAP 组 (n = 337)	t/Z/ χ^2	P
Glu(mmol/L)	5.40(4.67,7.19)	5.30(4.74,6.05)	0.555	0.579
TG(mmol/L)	0.93(0.69,1.29)	1.39(0.98,2.13)	4.023	<0.001
TC(mmol/L)	4.19 ± 0.95	4.57 ± 1.08	2.213	0.028
HDL-C(mmol/L)	1.28 ± 0.38	1.19 ± 0.30	1.393	0.098
LDL-C(mmol/L)	2.42 ± 0.75	2.67 ± 0.81	1.907	0.057
ApoA1(g/L)	1.30 ± 0.32	1.39 ± 0.25	2.188	0.029
ApoB(g/L)	0.82 ± 0.25	0.90 ± 0.26	1.894	0.059
Lp(a)(mg/L)	184.00(122.75,307.25)	145.00(83.00,253.50)	2.149	0.032
Hcy(μ mol/L)	16.00(12.00,20.00)	15.00(12.00,21.00)	0.150	0.881
Fib(g/L)	4.71 ± 1.45	3.42 ± 0.89	5.619	<0.001
D-二聚体(μ g/mL)	0.94(0.48,1.46)	0.36(0.26,0.55)	5.968	<0.001

2.2 建模验证人群拆分 将 2019 年 6 月—2021 年 6 月收治的 379 例急性缺血性脑卒中病例以 7:3 分配方式,分为训练集 265 例,年龄(66.07 ± 9.30)岁;测试集 114 例,年龄(66.38 ± 9.64)岁;两组患者年龄、性别、高血压病史、糖尿病史、NEU、LYM、PA、Fib、D-二聚体等临床特征比较,差异均无统计学意义(均 $P > 0.05$),见表 2。将 2021 年 7 月—2022 年 6 月收治的 157 例急性缺血性脑卒中病例作为验证集,其中 SAP 24 例,发病率为 15.29%。

表 2 训练集和测试集临床数据比较[例(%)]

Table 2 Comparison of clinical data in training set and testing set(No. of cases[%])

项目	训练集 (n = 265)	测试集 (n = 114)	t/ χ^2	P
男性	148(55.85)	73(64.04)	2.197	0.138
高血压病史	211(79.62)	83(72.81)	2.128	0.145
糖尿病史	59(22.26)	23(20.18)	0.205	0.651
NEU>6.67 × 10 ⁹ /L	62(23.40)	29(25.44)	0.182	0.669
LYM>1.31 × 10 ⁹ /L	160(60.38)	64(56.14)	0.592	0.442
PA>171 mg/L	225(84.91)	98(85.96)	0.071	0.790
Fib>3.68 g/L	79(29.81)	42(36.84)	1.813	0.178
D-二聚体>0.51 μ g/mL	71(26.79)	36(31.58)	0.901	0.342

2.3 模型建立 将纳入的 25 个检验参数通过 LASSO 的 10 折交叉验证对特征变量组合进行拟合,采用 -2 倍的对数似然值作为评价标准,将 25 个特征变量通过增加 lambda 值进行压缩组合,见图 1。通过对压缩过程中不同变量数目组合间的性

能评价,剔除不重要的临床特征变量,从而筛选出最优变量及最优 lambda 参数,见图 2。最终选择 lambda.1se 作为最优变量选择参数,共筛选出 5 个检验参数,分别 NEU、LYM、PA、Fib、D-二聚体。

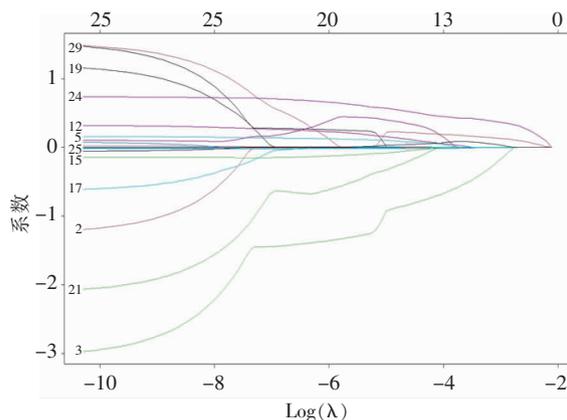


图 1 LASSO 压缩参数

Figure 1 LASSO compressing parameters

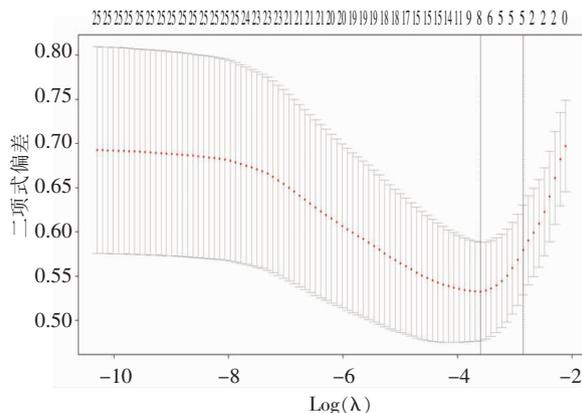


图 2 LASSO 筛选参数

Figure 2 LASSO screening parameters

将筛选出的 5 个检验参数根据最佳截断值进行二分类变量转化,以简化列线图模型和网页计算器,便于临床使用,见表 3。根据 5 个检验参数相应回归系数权重构建列线图模型,通过每个检验参数不同的权重可以获得不同的评分,从而计算出相应的风险概率。

表 3 检验参数最佳截断值

Table 3 Optimal cut-off values of test parameters

项目	约登指数	最佳截断值
NEU	0.452 9	6.67
LYM	0.370 1	1.31
PA	0.476 5	171.00
Fib	0.509 9	3.68
D-二聚体	0.441 4	0.51

2.4 模型的验证 为了防止联合检测出现过拟合,通过对全数据集的拆分,使用训练集构建模型,测试集进行内部验证模型,验证集进行外部验证模型。以确保联合检测模型的准确性及稳定性。区分度评估显示全数据集曲线下面积(Area Under Curve, AUC) = 0.924,灵敏度为 83.33%,特异度为 87.24%,见图 3。训练集 AUC = 0.922,灵敏度为 79.31%,特异性为 91.95%,见图 4。测试集 AUC = 0.919,灵敏度为 84.62%,特异性为 86.14%,见图 5。结果表明模型具有良好的预测效能,全数据集、训练集、测试集 5 个检验参数 AUC 与联合检测相比差异有统计学意义($P < 0.05$),见表 4。校准度评估显示训练集其 Hosmer-Lemeshow 检验拟合优度差异无统计学意义($\chi^2 = 6.565, P = 0.476$),校准曲线显示预测概率和实际概率具有良好的一致性。测试集其 Hosmer-Lemeshow 检验拟合优度差异无统计学意义($\chi^2 = 7.015, P = 0.319$),校准曲线显示预测概率和实际概率较一致,见图 6。临床实用性评估显示训练集 DCA 结果显示高风险阈值在 0~0.75,净获益为 0~0.11。测试集高风险阈值在 0~0.65,净获益为 0~0.11,见图 7。均具有良好的临

床实用性,外部验证采用验证集,其 AUC = 0.850,灵敏度为 66.67%,特异度为 89.47%,阳性似然比为 6.33,阴性似然比为 0.37,阳性预测值为 53.30%,阴性预测值为 93.70%,模型具有良好的外部适用性,见图 8。

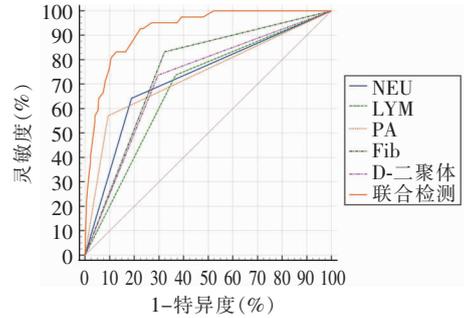


图 3 全数据集 ROC 曲线

Figure 3 ROC curve of full data set

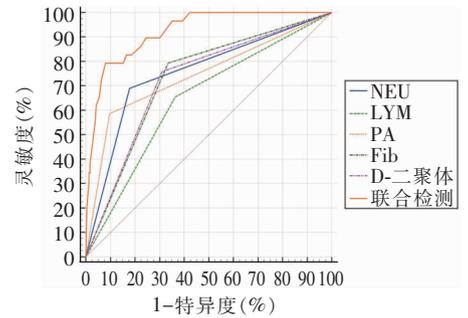


图 4 训练集 ROC 曲线

Figure 4 ROC curve of training set

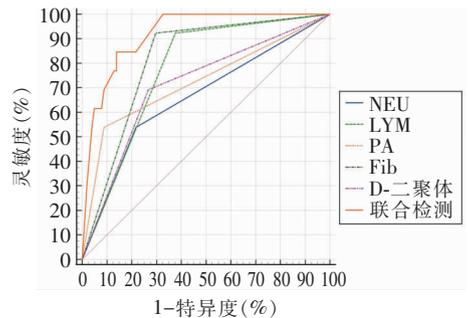


图 5 测试集 ROC 曲线

Figure 5 ROC curve of testing set

表 4 预测效能比较

Table 4 Comparison of prediction performance

项目	AUC	灵敏度(%)	特异度(%)	阳性似然比	阴性似然比	阳性预测值	阴性预测值	P
训练集								
NEU	0.756	68.97	82.20	3.88	0.38	32.30	95.60	<0.000 1*
LYM	0.645	65.52	63.56	1.80	0.54	18.10	93.70	0.002 2*
PA	0.744	58.62	90.25	6.01	0.45	42.50	94.70	<0.000 1*
Fib	0.729	79.31	66.53	2.37	0.31	22.50	96.30	<0.000 1*
D-二聚体	0.725	75.86	69.07	2.45	0.35	23.20	95.90	<0.000 1*
联合预测模型	0.922	79.31	91.95	9.85	0.23	54.80	97.30	<0.000 1
测试集								
NEU	0.66	53.85	78.22	2.47	0.59	24.10	92.90	0.032 2*
LYM	0.773	92.31	62.38	2.45	0.12	24.00	98.40	<0.000 1*
PA	0.725	53.85	91.09	6.04	0.51	43.80	93.90	0.002 2*
Fib	0.813	92.31	70.30	3.11	0.11	28.60	98.60	<0.000 1*
D-二聚体	0.712	69.23	73.27	2.59	0.42	25.00	94.90	0.002 5*
联合预测模型	0.917	84.62	86.14	6.10	0.18	44.00	97.80	<0.000 1

注：* 表示与同数据集联合预测模型相比，差异有统计学意义(P<0.05)。

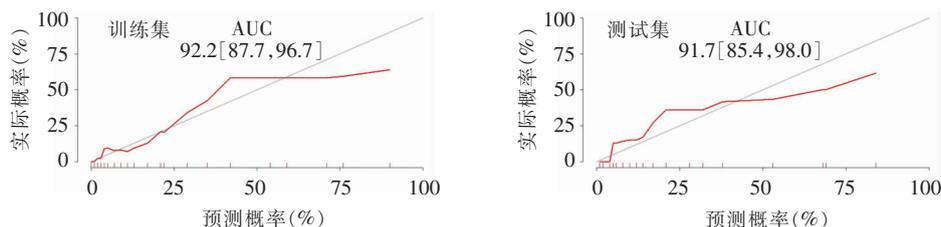


图 6 训练集和测试集校准曲线

Figure 6 Calibration curve of training set and testing set

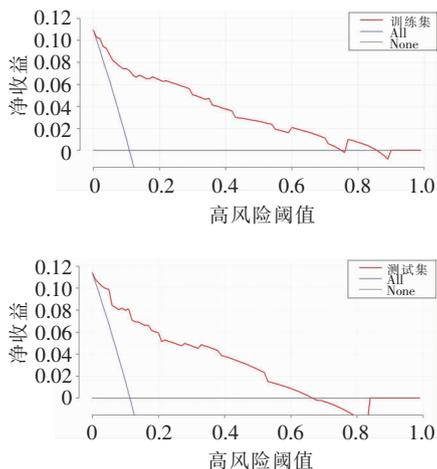


图 7 训练集和测试集 DCA 曲线

Figure 7 DCA curve of training set and testing set

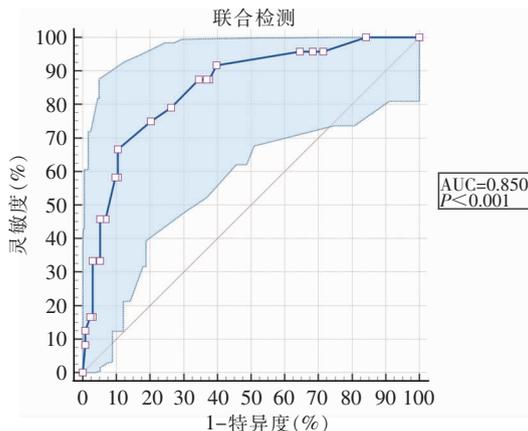


图 8 验证集 ROC 曲线

Figure 8 ROC curve of validation set

2.5 模型呈现 将构建好的联合检测列线图预测模型,见图 9。通过 R 语言的“DynNom”软件包布置到 <https://ww-rstudiomn.shinyapps.io/SAP->

nomgram/或通过二维码进行访问,见图 10。经过测试网页计算器性能,测试结果稳定,见图 11。

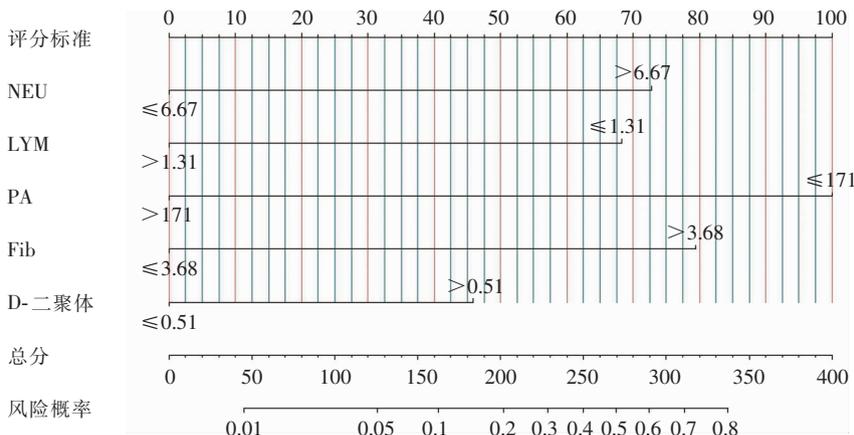


图 9 训练集联合预测列线图模型

Figure 9 Training set of combined prediction nomogram model



图 10 二维码

Figure 10 QR code

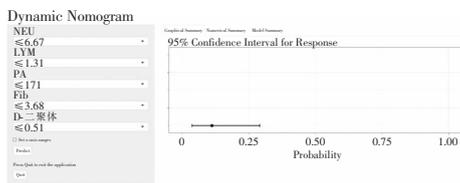


图 11 网页计算器

Figure 11 Web calculator

3 讨论

本研究通过常规检验数据的挖掘分析,构建了急性缺血性脑卒中并发 SAP 的预测模型,模型具有较好的预测效能,且通过布置网页计算器提高了其临床实用性,助力临床进行早期诊断及治疗。近年来列线图在临床医学中广泛使用,其相对于传统 logistic 回归模型,其可视化得到了极大的提高,具有一定的临床实用性^[11-12]。列线图的临床实用性虽然有所提高,但是临床医生的使用率并不高,其使用过

程仍不够简便。因此本研究为了进一步提高临床实用性,对构建的列线图进行优化,将连续变量转变为二分类变量,并采用网页列线图形式,使参数指标更加直观的判断,并且使用性能方面得到较大提升,对于辅助临床医生进行诊断更具有临床实用性^[13]。当然检验参数指标的选择也尤为重要,以往对于 SAP 的预测多采用的是单个指标,评分量表及新的生物标志物,这些方式可能存在一定局限性,例如可能会增加患者检测费用,预测效能不高,不利于临床推广,不便于临床医生使用。因此本研究采用既常规检验数据的挖掘,通过 LASSO 筛选出具有价值的检验参数,充分利用检验数据的剩余价值,开发出新的联合检测预测模型,为临床提供辅助诊断。Li 等^[14]研究显示通过机器学习开发的缺血性脑卒中并发 SAP 的模型中 XGBoost 模型具有较高的诊断效能,其 AUC 为 0.841,灵敏度为 81.00%,特异度为 73.30%。张娜等^[15]研究比较了 6 种预测缺血性脑卒中并发医院获得性肺炎的评分量表,其中 ISAN 量表预测效能最高,AUC 为 0.778。本研究通过 NEU、LYM、PA、Fib、D-二聚体等 5 个检验参数构建的列线图网页计算器,其训练集 AUC 为 0.922、测试集 AUC 为 0.919,验证集 AUC 为 0.850,预测效能均较高,并且临床实用性较高,对于临床辅助诊断具有一定价值。

急性缺血性脑卒中后容易诱发免疫抑制综合征(SIDS),过度、持续的炎症反应可能耗尽免疫系统,最终导致系统免疫的抑制,削弱人体对病原体的抵

抗力并导致肺部感染^[16]。本文研究筛选出的检验参数中,NEU、LYM、Fib、D-二聚体反映机体的炎症及免疫系统状态,研究结果显示 SAP 组中 NEU、Fib、D-二聚体均高于非 SAP 组,LYM 低于非 SAP 组,表明机体可能存在持续的炎症状态,存在免疫抑制风险。相关研究^[17]表明,营养不良会增加感染风险,营养风险评分越低,SAP 发病率越高。PA 作为营养不良的敏感指标,可能是确定临床状态和预后的有效指标,本文研究显示,SAP 组 PA 水平低于非 SAP 组,并且具有良好的预测效能,研究结果与文献^[17]一致。当然本研究也存在一定局限性,由于是单中心小样本量研究,外部验证仅采用了本中心数据,模型外部适用性和准确性可能存在一定偏倚。

综上所述,本研究利用检验相关参数构建的 SAP 预测模型在临床相关性及临床实用性方面都具有一定价值,充分实现了检验项目的剩余价值,为临床早期的预防性抗菌药物治疗、抗炎和免疫调节治疗以及营养支持提供了依据。

利益冲突:所有作者均声明不存在利益冲突。

[参考文献]

[1] Liu SW, Li YC, Zeng XY, et al. burden of cardiovascular diseases in China, 1990 - 2016: findings from the 2016 global burden of disease study[J]. *JAMA Cardiol*, 2019, 4(4): 342 - 352.

[2] Patel UK, Kodumuri N, Dave M, et al. Stroke-associated pneumonia: a retrospective study of risk factors and outcomes [J]. *Neurologist*, 2020, 25(3): 39 - 48.

[3] Cugy E, Sibon I. Stroke-associated pneumonia risk score: validity in a French stroke unit[J]. *J Stroke Cerebrovasc Dis*, 2017, 26(1): 225 - 229.

[4] Nam KW, Kim TJ, Lee JS, et al. High neutrophil-to-lymphocyte ratio predicts stroke-associated pneumonia[J]. *Stroke*, 2018, 49(8): 1886 - 1892.

[5] Gong SY, Zhou ZW, Zhou MK, et al. Validation of risk scoring models for predicting stroke-associated pneumonia in patients with ischaemic stroke[J]. *Stroke Vasc Neurol*, 2016, 1 (3): 122 - 126.

[6] Hotter B, Hoffmann S, Ulm L, et al. Inflammatory and stress markers predicting pneumonia, outcome, and etiology in patients with stroke: biomarkers for predicting pneumonia, functional outcome, and death after stroke[J]. *Neurol Neuroimmunol Neuroinflamm*, 2020, 7(3): e692.

[7] Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye[J]. *Lancet Oncol*, 2015, 16(4): e173 - e180.

[8] Semenkovich TR, Yan Y, Subramanian M, et al. A clinical nomogram for predicting node-positive disease in esophageal cancer[J]. *Ann Surg*, 2021, 273(6): e214 - e221.

[9] Lo SN, Ma JW, Scolyer RA, et al. Improved risk prediction calculator for sentinel node positivity in patients with melanoma: the Melanoma Institute Australia Nomogram[J]. *J Clin Oncol*, 2020, 38(24): 2719 - 2727.

[10] Smith CJ, Kishore AK, Vail A, et al. Diagnosis of stroke-associated pneumonia: recommendations from the pneumonia in stroke consensus group[J]. *Stroke*, 2015, 46(8): 2335 - 2340.

[11] Maurichi A, Miceli R, Eriksson H, et al. Factors affecting sentinel node metastasis in thin (T1) cutaneous melanomas: development and external validation of a predictive nomogram [J]. *J Clin Oncol*, 2020, 38(14): 1591 - 1601.

[12] Vaidya P, Bera K, Gupta A, et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multi-cohort study for outcome prediction[J]. *Lancet Digit Health*, 2020, 2(3): e116 - e128.

[13] Levine LD, Downes KL, Parry S, et al. A validated calculator to estimate risk of cesarean after an induction of labor with an unfavorable cervix[J]. *Am J Obstet Gynecol*, 2018, 218(2): 254. e1 - 254. e7.

[14] Li X, Wu M, Sun C, et al. Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients[J]. *Eur J Neurol*, 2020, 27(8): 1656 - 1663.

[15] 张娜, 刘红, 杨新颜, 等. 脑卒中相关性肺炎护理风险评价指标体系的构建[J]. *中国感染控制杂志*, 2019, 18(3): 220 - 224.

Zhang N, Liu H, Yang XY, et al. Establishment of nursing risk assessment index system for stroke-associated pneumonia [J]. *Chinese Journal of Infection Control*, 2019, 18(3): 220 - 224.

[16] Liu DD, Chu SF, Chen C, et al. Research progress in stroke-induced immunodepression syndrome (SIDS) and stroke-associated pneumonia (SAP)[J]. *Neurochem Int*, 2018, 114: 42 - 54.

[17] Dai CJ, Yan D, Xu MJ, et al. Geriatric nutritional risk index is related to the risk of stroke-associated pneumonia[J]. *Brain Behav*, 2022, 12(8): e2718.

(本文编辑:陈玉华)

本文引用格式:曾瑞璜,陈智熙,王小林,等.常规检验数据挖掘对急性缺血性脑卒中并发卒中相关肺炎的预测价值[J].*中国感染控制杂志*, 2023, 22(2): 142 - 149. DOI: 10. 12138/j. issn. 1671 - 9638. 20233331.

Cite this article as: ZENG Rui-huang, CHEN Zhi-xi, WANG Xiao-lin, et al. Predictive value of routine testing data mining for acute ischemic stroke complicated with stroke-associated pneumonia[J]. *Chin J Infect Control*, 2023, 22(2): 142 - 149. DOI: 10. 12138/j. issn. 1671 - 9638. 20233331.