

颅脑损伤院内获得性肺炎预测：一项基于机器学习的
PCA-Logistic 回归分析模型

冯金周 刘发健 匡永勤 江 华

【摘要】目的 探讨 PCA-Logistic 回归分析模型在颅脑损伤病人院内获得性肺炎(HAP)预测建模中的应用效果。方法 收集 2011 年 12 月至 2017 年 11 月开颅手术治疗的 108 例颅脑损伤的相关临床数据,建立 PCA-Logistic 回归分析模型,利用受试者工作特征(ROC)曲线评估模型预测效果。结果 PCA-Logistic 回归分析模型发现影响病人 HAP 发生的重要临床指标,经 ROC 曲线评估 PCA-Logistic 回归分析模型预测 HAP 结局具有较高的预测效能,灵敏度为 83.9%,特异度为 94.8%,曲线下面积为 0.949。结论 PCA-Logistic 回归分析模型可以有效的挖掘颅脑损伤后的临床变量,可建立 HAP 的预测模型,不规范的肠外营养支持可能是影响 HAP 发生的重要临床因素。

【关键词】 颅脑损伤;院内获得性肺炎;主成分分析;Logistic 回归分析

【文章编号】 1009-153X(2019)01-0035-04 【文献标志码】 A 【中国图书资料分类号】 R 651.1+5

Prediction of hospital-acquired pneumonia after traumatic brain injury based on the PCA-Logistic regression analysis model

FENG Jin-zhou¹, LIU Fa-jian¹, KUANG Yong-qin², JIANG Hua³. 1. Department of Neurosurgery, Sichuan Provincial People's Hospital, Chengdu 610101, China; 2. Department of Neurosurgery, General Hospital, Western Command, PLA, Chengdu 610083, China; 3. Metabolomics and Multidisciplinary Laboratory for Trauma Research, Sichuan Provincial People's, Chengdu 610101, China

【Abstract】 Objective To explore the application of principal components analysis (PCA)- Logistic regression analysis to prediction of the hospital-acquired pneumonia (HAP) in the patients with traumatic brain injury (TBI), and to find the pathophysiological patterns and important risk factors related to their clinical prognoses. Methods Original dataset was constituted by the data of 108 patients with TBI derived from the Database of the Trauma Center of Sichuan Provincial People's Hospital from 2011 to 2017. The primary outcome was HAP. After the dataset evaluation and cleaning, the PCA-Logistic regression model was built to identify the risk factors related to HAP. Receiver Operating Characteristic (ROC) curve was used to evaluate the PCA-Logistic regression model. Results The PCA-Logistic regression model analysis found the important clinical indicators affecting the patients HAP, and the PCA-Logistic regression model was evaluated by ROC curve. The HAP outcome model had good predictive power (sensitivity, 83.9%; specificity, 94.8%; AUC, 0.949). Conclusions PCA-Logistic regression analysis can effectively mine the clinical variables of the patients with TBI and establish a clinical prognosis prediction model. The abnormal parenteral nutrition support after severe TBI may be an important clinical factor affecting the occurrence of HAP in the patients with TBI.

【Key words】 Traumatic brain injury; Hospital-acquired pneumonia; Principal component analysis; Logistic regression analysis

颅脑损伤病人院内获得性肺炎(hospital-acquired pneumonia, HAP)的发生率可高达 60%^[1]。一旦发生 HAP,会明显延长病人住院时间,尤其是重症监护室住院时间(intensive care unit length of stay, ICU-LOS),以及机械通气时间,甚至病死率也会明显增高^[2,3]。了解颅脑损伤后临床指标与 HAP 的关系,可以预测 HAP 发生风险,减少不良预后^[4]。本文

根据颅脑损伤后临床常见且易获得的指标,采用主成分分析(principal components analysis, PCA) + Logistic 回归分析方法建立 HAP 预测模型,探讨基于机器学习的模式识别技术在临床预测中应用的可行性。

1 资料与方法

1.1 数据来源 病例纳入标准:①年龄 14~65 岁;②有颅脑损伤史;③有开颅手术史。排除标准:①妊娠;②院外已经接受过紧急颅脑手术;③住院时间不足 24 h。收集 2011 年 12 月至 2017 年 11 月四川省医学科学院·四川省人民医院创伤中心数据库中开颅手术治疗的颅脑损伤的临床数据,符合纳入标准的病

doi:10.13798/j.issn.1009-153X.2019.01.010
基金项目:四川省卫生厅科研资助项目(110212,16PJ447)
作者单位:610110 成都,四川省医学科学院·四川省人民医院(东院) 神经外科(冯金周、刘发健),创伤与代谢组多学科实验室(江 华); 610083 成都,中国人民解放军西部战区总医院神经外科(匡永勤)
通讯作者:江 华,E-mail:feng720930@126.com

人共 108 例,其中男 95 例,女 13 例;年龄(41.80 ± 12.60)岁。开放性颅脑损伤 47 例。多发伤 53 例。交通事故伤 63 例,跌倒伤 22 例,高处坠落伤 9 例,打击伤 10 例,刀砍伤 2 例,爆炸伤 1 例,挤压伤 1 例。脑挫裂伤 90 例,原发性脑干损伤 20 例,脑震荡 6 例,对冲伤 41 例,硬膜下血肿 56 例,硬膜外血肿 54 例,脑内血肿 24 例,术前脑疝 40 例。入院时:GCS 评分(8.50 ± 4.00)分,创伤严重程度评分(95.10 ± 6.90)分,血糖(8.20 ± 3.90) mmol/L,白细胞数[(15.70 ± 6.40) $\times 10^9$ /L],体温(38.30 ± 1.00) $^{\circ}\text{C}$,血氧饱和度(95.10 ± 6.90)%。

1.2 数据集的构建 根据颅脑损伤后早期容易获得的临床指标,建立数据集。这些指标包括人口统计学指标,例如年龄、性别等;临床严重程度指标,例如受伤原因,GCS 评分,创伤严重程度评分、瞳孔反应、多发伤、血肿量、血肿部位;继发性损害,例如缺氧、低血压、低体温、酸中毒、凝血功能障碍、血压(收缩压或舒张压)异常;住院处治情况,例如伤后距手术时间、气管切开术、机械通气时间、糖皮质激素使用、鼻胃管使用、肠外营养时间、单纯脂肪乳剂使用;颅脑 CT 特征和各种血糖、白细胞等。并发症:感染性并发症(除肺部感染)、住院期间其他并发症。主要结局:临床死亡与 HAP。次要结局:住院时间、ICU-LOS、机械通气时间、并发症。

1.3 HAP 诊断标准^[5,6] HAP 是指病人入院时未处于肺炎潜伏期,且在入院时不用气管插管,入院 ≥ 48 h 后发生的肺炎,包括呼吸机相关性肺炎以及气管插管 48~72 h 后发生的肺炎。肺炎临床诊断标准^[7]:进行性肺部浸润性病变,合并 2 个以下临床症状:发热 ≥ 38 $^{\circ}\text{C}$;新出现的咳嗽、咳痰,并出现脓性分泌物,伴有或不伴有胸痛;肺实变体征,和(或)湿啰音;外周血白细胞 $\geq 10 \times 10^9$ /L 或 $\leq 4 \times 10^9$ /L,伴有或不伴有核左移。

1.4 统计学分析 采用 Mac OS X 10.9.2 系统进行分析;符合正态分布的变量采用 $\bar{x} \pm s$ 描述;非正态分布变量采用中位数及四分间距描述。分析步骤:①采用 AMAP 程序包,对所有临床指标进行 PCA;②以累积贡献度大于绝对多数(2/3)提取主成分;③以临床结局为应变量,以主成分为因变量进行 Logistic 回归分析,建立回归预测模型,并计算模型中临床变量的相应系数。采用受试者工作特征(receiver operating characteristic, ROC)曲线计算灵敏度、特异度以及曲线下面积(area under curve, AUC)。检验水准 $\alpha = 0.05$ 。

2 结果

2.1 入院后治疗情况 受伤至手术时间(60.70 ± 28.80)h。行气管切开术 13 例,亚低温治疗 18 例,使用糖皮质激素 61 例,鼻饲营养 37 例,肠外营养 15 例,仅用脂肪乳肠外营养支持 16 例。肠内营养时间(2.10 ± 1.90)d。发生误吸 41 例,酸中毒 21 例,凝血功能障碍 24 例,营养后遗症 10 例。

2.2 临床结局 108 例中,死亡 13 例(12.04%);发生 HAP 31 例(28.70%),发生并发症 10 例(9.26%)。住院时间(32.80 ± 27.90)d,ICU-LOS(14.50 ± 5.80)d,机械通气时间(3.70 ± 2.60)d。

2.3 PCA-Logistic 回归分析结果 以 36 个临床指标作为变量,进行 PCA,提取前 12 个主成分,并以 HAP 作为因变量进行 Logistic 回归分析,发现 5 个变量有统计学差异,见表 1。

计算 36 个临床指标系数= $\text{sum}(\text{sqrt}((\exp(\beta) \times \text{loading value})^2))$,以第 75 百分位数作为截点,提取出对临床结局具有重要影响力的变量,绘制 HAP 为结局的临床指标系数分布图,发现大于第 75 百分位数的指标包括性别(gender)、肠外营养(PN)、脑内血肿(intracerebral-hematoma)、受伤原因(cause)、受伤距手术时间(time-surgery)、脑震荡(concussion)、GCS 评分、白细胞(WBC)、糖皮质激素(glucosedialstolic)。见图 1。

2.4 ROC 曲线分析结果 ROC 曲线显示灵敏度为 83.9%,特异度为 94.8%,AUC=0.949。见图 2。

3 讨论

对于早期预测颅脑损伤的预后,因其病情的复杂与多变性导致预测的准确性与伦理等问题而备受争议。随着计算机科学和应用数学与临床医学的结合日益紧密,越来越多的分析和计算工具应用到了临床研究之中,以解决临床数据分析中的各种问题。如对于重型颅脑损伤死亡的早期预测模型经大宗病例验证,具有很高的准确度而对临床医生具有预测指导意义^[8]。对于复杂临床数据的处理,多元 Logistic 回归方程已不适用,无法建立预测模型,需要寻找一种新的分析方法。本文针对临床常见的颅脑损伤病人,在满足一定筛选条件下,提取复杂的数据指标,采用 PCA,结合 Logistic 回归分析,建立以 HAP 为结局的临床预测模型,具有较高的辨识度和特异度。

临床数据的复杂性一直以来都是研究者们面临

表 1 以院内获得性肺炎为结局的主成分 Logistic 回归分析结果

变量	回归系数	标准误	z 值	P 值
解释变量	-1.54	0.38	-4.06	<0.001
主成分 1	-1.02	0.24	-4.25	<0.001
主成分 3	-0.48	0.22	-2.19	0.03
主成分 6	1.13	0.37	3.08	<0.001
主成分 10	-0.91	0.38	-2.37	0.02
主成分 11	1.39	0.48	2.90	<0.001

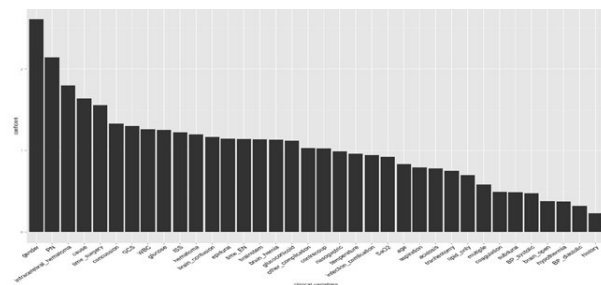


图 1 以院内获得性肺炎为结局的临床指标系数分布图

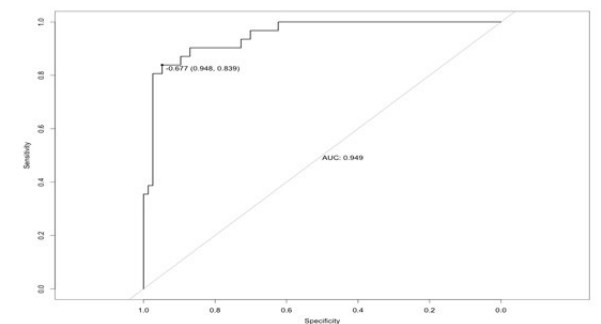


图 2 以院内获得性肺炎结局的 ROC 曲线

的挑战:一方面数据需要符合分析模型的限制条件;另一方面又需要尽可能保留数据特征,发现数据背后反映的临床状况。机器学习算法中,PCA 是最为常用的降维方法,在保留数据特征的同时,将高维空间的数据映射到低维空间中,建立对应的线性方程组,即相应的主成分^[9]。本文采用 PCA 方法将颅脑损伤病人的 36 个临床指标所构成的高维数据空间映射到了与之对应的主成分上,按照各成分的贡献度截取累计贡献度超过总体绝对多数(2/3)的 12 个主成分,从而获得了可以表征数据信息的一组特征向量;然后,利用二分类变量的临床结局,进一步建立 Logistic 回归分析模型,筛选特征向量。PCA-Logistic 回归分析模型是结合了 PCA 降维和 Logistic 回归分析的有监督机器学习模式识别方法^[10]。通过计算机进行 PCA 原始变量降维,从而得到了临床数据在高维空间中映射的多个特征向量;再通过

Logistic 回归分析模型以临床结局筛选出这些特征向量,从而得到了研究所需要的、以临床结局为判断依据的一组主成分^[11]。本文以 HAP 为结局的 PCA-Logistic 回归分析模型中,有 5 个主成分显著影响 HAP 是否发生的判别;再利用 ROC 曲线对 PCA-Logistic 回归分析模型的判别效果进行评估,发现模型具有较好的判别效果(灵敏度 83.9%,特异度 94.8%,AUC=0.949)。PCA-Logistic 回归分析模型能够对颅脑损伤病人在特定的临床结局下所对应的主成分进行识别,采集病人的临床数据,对颅脑损伤病人的是否发生 HAP 的风险进行预测,有助于提升颅脑损伤的救治水平。

每一个主成分在模型中都有对应的回归系数,由此可以得到该主成分在对应结局情况下的发生比。每一个主成分又由 36 个临床指标所构成,结合上这些变量所对应的载荷值,最终获得临床指标在 PCA-Logistic 回归分析模型中对应的影响系数。这些系数可以表征临床指标对特定结局(如 HAP)的影响力大小;讨论影响力相对较大的临床指标,进而可以发现影响病人结局的临床模式和重要因素。本文以 HAP 为结局的 PCA-Logistic 回归分析模型中临床指标系数的分布较为接近,提示 HAP 的发生是受综合性的临床状况和干预措施影响。其中,临床系数相对较大(大于第 75 百分位数)的指标包括性别、肠外营养、脑内血肿、受伤原因、受伤距离手术时间、脑震荡、GCS 评分、白细胞以及糖皮质激素。我们认为,一方面,HAP 发生与颅脑损伤和昏迷严重程度有关;另一方面,重型颅脑损伤后肠外营养的使用也对 HAP 的发生产生影响。循证医学证据提示肠外营养和肠内营养的选择对颅脑损伤病人的临床结局并没有显著的影响,也不是不良临床结局的风险因素^[12-14]。事实上,临床中营养支持的具体操作方法,可能才是影响病人 HAP 发生的重要风险因素^[15]。不规范的肠外营养使用,例如已证实的脂肪乳单独输注、未使用中心静脉或经外周静脉的中心静脉置管、临床营养师或药剂师未参与营养支持方案的制定等,可能会对病人肠外营养支持效果产生影响进而造成不良结局。因此肠外营养的规范化使用不仅需要制定国家级的指南,同时临床营养规范化培训需要在护理和临床治疗方面进一步加强,以确保肠外营养规范化操作的执行,避免临床不良预后的发生。

除上述影响预后的病理生理指标外,仍然不可忽视临床采取的干预措施对病人预后的影响。气管切开、糖皮质激素的同时使用,也是死亡的高危因

素;颅脑损伤后昏迷不能自主排痰,颅内压增高导致呕吐误吸等导致呼吸功能障碍,采取气管切呼吸机管控仍然是最为有效的办法^[16,17]。但由于气管切开增加了呼吸道与外界细菌的接触,也是增加医院获得性感染的主要原因之一。我们也发现,HAP是颅脑损伤病人中极易发生的呼吸系统并发症,一旦发生,将严重影响病人的呼吸功能。这同样也反映了HAP是影响病人生存的重要风险因素。

综上所述,PCA-Logistic回归分析模型可以成为研究颅脑损伤病人HAP风险预测的工具。通过对临床数据特征的识别,PCA-Logistic回归分析模型可以帮助我们在临床上发现影响不良预后发生的病理生理模式。进一步,由此判断颅脑损伤病人不同病理生理状态的临床预后,协助临床诊疗。本研究局限在于临床样本数少,基于机器学习算法PCA-Logistic回归分析模型还应针对多个主成分建立模式识别,从复杂的临床指标中预测出不同的临床结局,才具有临床可行性。

【参考文献】

- [1] Zygun DA, Zuege DJ, Boiteau PJ, *et al.* Ventilator-associated pneumonia in severe traumatic brain injury[J]. *Neurocrit Care* 2006, 5(2): 108 - 114.
- [2] Asehnoun K, Roquilly A, Sebillé V. Corti-TC trial group, Corticotherapy for traumatic brain-injured patients-- the Corti-TC trial: study protocol for a randomized controlled trial [J]. *Trials*, 2011, 12: 228.
- [3] Yang CC, Shih NC, Chang WC, *et al.* Long-term medical utilization following ventilator-associated pneumonia in acute stroke and traumatic brain injury patients: a case-control study [J]. *BMC Health Services Research*, 2011, 11: 289.
- [4] Safdar N, Dezfulian C, Collard HR, *et al.* Clinical and economic consequences of ventilator-associated pneumonia: a systematic review [J]. *Crit Care Med*, 2005, 33(10): 2184-2193.
- [5] 李 培,施 毅. 医院获得性肺炎及呼吸机相关性肺炎诊治指南解读[J]. *中国循证医学杂志*, 2015, 15(7): 772-776.
- [6] 朱仕超,尹维佳,宗志勇,等. 呼吸机相关性肺炎定义和判断标准研究进展[J]. *中华医院感染学杂志*, 2016, 26(23): 5517-5520.
- [7] 肺炎的诊断标准[J]. *中国社区医师*, 2012, 28(42): 7.
- [8] Gomez PA, de-la-Cruz J, Lora D, *et al.* Validation of a prognostic score for early mortality in severe head injury cases [J]. *J Neurosurg*, 2014, 121(6): 1314-1322.
- [9] Sánchez-Sánchez ML, Belda-Lois JM, Mena-Del Horno S, *et al.* A new methodology based on functional principal component analysis to study postural stability post-stroke [J]. *Clin Biomech (Bristol, Avon)*, 2018, 56: 18-26.
- [10] Begue A, Kowlessur V, Singh U, *et al.* Automatic recognition of medicinal plants using machine learning techniques [J]. *Int J Adv Comp Sci Appl (IJACSA)*, 2017, 8 (4): 166-175.
- [11] 陶 然. Logistic模型多重共线性问题的诊断及改进[J]. *统计与决策*, 2008, (15): 22-25.
- [12] Braga M, Ljungqvist O, Soeters P, *et al.* ESPEN Guidelines on Parenteral Nutrition: surgery [J]. *Clin Nutr*, 2009, 28(4): 378-386.
- [13] Kreyman KG, Berger MM, Deutz NE, *et al.* DGEM (german society for nutritional medicine); ESPEN (european society for parenteral and enteral nutrition). ESPEN Guidelines on Enteral Nutrition: Intensive care [J]. *Clin Nutr*, 2006, 25(2): 210-223.
- [14] Mirtallo J, Canada T, Johnson D, *et al.* Task force for the revision of safe practices for parenteral nutrition. safe practices for parenteral nutrition [J]. *JPEN J Parenter Enteral Nutr*, 2004, 28(6): S39-70.
- [15] 冯金周,曾 俊,孙明伟,等. 早期经皮内镜下胃造口置管行肠内营养对重型颅脑损伤病人临床结局的影响[J]. *肠外与肠内营养*, 2011, 18(6): 331-334.
- [16] Leasure AR, Stirten J, Lu SH. Prevention of ventilator-associated pneumonia through aspiration of subglottic secretions: a systematic review and meta-analysis [J]. *Dimens Crit Care Nurs*. 2012, 31(2): 102-117.
- [17] Blot SI, Poelaert J, Kollef M. A key element for the prevention of ventilator-associated pneumonia in intubated ICU patients [J]. *BMC Infect Dis*, 2014, 28(14): 119.

(2018-08-09收稿,2018-09-29修回)