

• 临床研究 •

应用人工神经网络评价老年重症监护室患者的预后

周伟炜 宋以信 关有彦

【摘要】 目的 比较人工神经网络(ANN)和急性生理学及慢性健康状况评价Ⅱ(APACHEⅡ)评价老年重症监护室(GICU)患者死亡风险的预测能力。方法 回顾分析北京大学第一医院GICU2005年1月至2006年12月收治的177例年龄≥65周岁患者的病例资料。记录患者入GICU当日的资料,计算APACHEⅡ评分和预期死亡风险。随机抽取117例患者的资料作为ANN的训练集,其余60例患者的资料作为ANN的检验集。根据APACHEⅡ的22个变量建立ANN22,根据单个变量对于预后预测的影响大小,去除影响较小的变量,分别利用10个和8个变量建立ANN10和ANN8。应用APACHEⅡ和ANN22、ANN10、ANN8预测检验集患者的死亡风险。应用受试者工作特性曲线下面积(aROC)判断ANN和APACHEⅡ的预测能力。结果 ANN22和APACHEⅡ的aROC分别为0.943和0.949,差异没有统计学意义($P=0.829$)。ANN10和ANN8的aROC分别为0.968和0.926。结论 应用ANN预测GICU患者的死亡率与应用以Logistic回归为基础的APACHEⅡ相比,结果相当。应用较少的变量所建立的ANN预测准确性仍很高。

【关键词】 人工神经网络;急性生理学及慢性健康状况评价Ⅱ;老年;重症监护室;预后

Artificial neural network in the mortality prediction for patients in geriatric intensive care unit

ZHOU Weiwei, SONG Yixin, GUAN Youyan

Geriatric Department, the First Hospital, Peking University, Beijing 100034, China

【Abstract】 Objective To compare the ability of artificial neural networks(ANN) and the acute physiology and chronic health evaluation Ⅱ (APACHE Ⅱ) to predict mortality for patients in geriatric intensive care unit (GICU). Methods The purpose of this retrospective case series was to compare ANN and APACHE Ⅱ in the mortality prediction for a cohort of patients admitted to a seven-bed GICU in a Beijing teaching general hospital. All 177 patients were older than 65 years and consecutively admitted to our GICU from Jan 2005 to Dec 2006. The 22 variables used to obtain APACHE Ⅱ score and risk of death were collected from each patient on admission. All data were randomly allocated to either the training ($n=117$) or validation set ($n=60$). Three ANN models were developed using the data from the training set, namely ANN22 (trained with all the 22 variables), ANN10 (trained with the 10 highest information gain variables) and ANN8 (trained with the 8 highest information gain variables). Three ANN models and APACHE Ⅱ were used to predict mortality in the validation set. The accuracy of ANN and APACHE Ⅱ was assessed by area under the receiver operator characteristics curve (aROC). Results The aROC was 0.943 for ANN22 and 0.949 for APACHE Ⅱ in predicting GICU mortality($P=0.829$). For ANN10 and ANN8, the aROC was 0.968 and 0.926, respectively. Conclusion Both ANN and APACHE Ⅱ have similar performance in predicting GICU outcome. ANN uses fewer variables and yet is comparable to APACHE Ⅱ.

【Key words】 artificial neural network model; acute physiology and chronic health evaluation Ⅱ; elderly; intensive care unit; prognosis

近年来,以 Logistic 回归统计方法为基础的预后评价模型,如急性生理学及慢性健康状况评价Ⅱ (acute physiology and chronic health evaluation Ⅱ, APACHE Ⅱ), 死亡率预测模型(mortality

收稿日期:2008-06-06

作者单位:100034 北京市,北京大学第一医院老年内科(周伟炜、宋以信),泌尿外科(关有彦)

通讯作者:宋以信, E-mail: doctorsongyixin@yahoo.com.cn

probability model),以及简易急性生理学评分(simplified acute physiology score),已成为重症监护室(intensive care unit, ICU)普遍接受和使用的评分系统,被用来预测患者的死亡风险^[1~4]。然而,传统的 Logistic 回归统计学方法要求变量满足独立性的条件,并难以处理变量间复杂的非线性关系,还要求研究者了解哪些变量更能预测特定的结果^[5]。此外,数据丢失和样本量过小会使预测结果不可靠^[6]。而且这些模型都是针对西方人设计的,不能完全代表我国人群的特点,尤其是老年人群。

人工神经网络(artificial neural network, ANN)是近年来迅速发展起来的一门集神经科学、计算机科学、信息科学、工程科学为一体的边缘交叉学科。Koholen 曾给出 ANN 的定义:人工神经网络是由具有适应性的简单单元组成的广泛并行互联的网络,它的组织能够模拟生物神经系统对真实世界物体做出交互反应。它采用物理可实现的系统来模仿人脑神经细胞的结构和功能,由于它具有自适应性,能够分析错综复杂的非线性数据,受到了临床工作者的特别关注。ANN 与人脑的工作方式相似,它的主要优势是可以帮助研究者在大型数据库中发现变量之间的相关性^[7];即使输入信息不完全、不准确或模糊不清,只要输入的模式接近于训练样本,ANN 仍能给出正确的推理预测。近年来,ANN 以其良好的预测能力越来越多地被应用在医学领域中。

综上,以 Logistic 回归统计方法为基础的预后评价模型所能提供的信息是有限的,在预测的最终目的是决定患者治疗的情况下更为明显。中国的医疗保险制度尚不完善,随着中国老龄人口的日趋增加,探索能准确预测老年危重患者预后模型的研究显得越来越重要。一方面临床医师不能仅凭年龄、急性生理改变,或者经验等因素武断地终止治疗;另一方面我国医疗卫生的人力、财力、物力等资源有限,所以需要预测模型尽可能提供准确的预后评价,帮助临床医师根据预测结果及时甄别危重患者。目前在 ICU 中,APACHE II 是应用最广泛的评分系统。众多研究表明,应用 ANN 预测 ICU 患者的死亡率与应用 Logistic 回归预测模型相比,结果相当,甚至更好^[8~12];并且利用人工神经网络进行预测可以减少所需变量的个数,其准确性仍较高^[8~10]。因此建立针对老年危重患者的人工神经网络预后评价模型,对于提高老年重症监护室(geriatric intensive care unit, GICU)救治和管理水平具有更重要的意义。本研究根据 APACHE II 的变量建立 ANN,探讨其对老

年患者死亡风险预测能力的优势。

1 对象和方法

1.1 研究对象 2005年1月至2006年12月北京大学第一医院 GICU 收治的年龄 ≥ 65 周岁的患者,重复入院者只记录第一次入院资料,观察期结束时仍留院的患者亦不计在内。

1.2 资料收集和数据处理 (1)一般临床资料记录患者姓名、年龄、性别、既往病史。(2)记录患者入 GICU 当日体温、血压、心率、呼吸频率、动脉血氧分压、动脉血 pH 值、血钠、血钾、血肌酐、血细胞比容、白细胞计数和 Glasgow 昏迷评分、原有慢性疾病、入监护室主要原因等指标。所有生理指标均取入 GICU 前后 24h 内的最差值。根据 APACHE II 评分标准^[13]计算相应得分并计算出住院死亡预期危险 $\ln(R/1-R) = -3.517 + (\text{APACHE II 得分} \times 0.146) + 0.603$ (仅限于急诊手术后患者)+患者入 ICU 的主要疾病得分。(3)ANN 的结构和参数:ANN 采用前向型网络设计和误差反向传播学习算法,包括输入层、隐含层和输出层。输入层包括 m 个输入神经元;隐含层为 1 个,包括 n 个神经元;输出层包括 1 个输出神经元。根据 Kolmogorov 定理,隐含层神经元的个数取 $n = 2 \times m + 1$ 。利用 Levenberg-Marquardt 规则进行训练,网络的误差平方和 $< 10^{-3}$ 时训练收敛结束,权重与偏值随机初始化。输入层到隐含层的传递函数采用 S 型正切函数 tansig、隐含层到输出层的传递函数采用 S 型对数函数 logsig。为了让网络达到更好的性能,对输入数据进行了归一化处理,对于两分类指标分别赋值 0 和 1,多分类资料转化为多个 0-1 变量,用各指标值除以该指标的最大值,即用 $x' = x/\max(x)$ 进行归一化处理。利用 Matlab7.0 软件编程实现 ANN 的训练和检验。

将患者随机分为两个样本集,训练集 117 例患者(2/3)和检验集 60 例患者(1/3)。用训练集使 ANN"学习"如何预测患者的死亡风险,在训练阶段并不应用检验集,因此,由检验集得出的预测结果可以用来评价 ANN 在实际应用中的价值。

利用训练集,并采用 APACHE II 预测死亡率所需的 22 个变量建立 ANN22,不同的是 APACHE II 系统在评价慢性健康状况评分时纳入所有出现异常的器官系统,而笔者在建立人工神经网络模型时只考虑呼吸、心血管、肝脏、肾脏和免疫系统慢性疾病。

根据单个变量对模型预后预测的影响大小(图 1)^[10]及临床经验去除影响较小的变量,分别利用 10

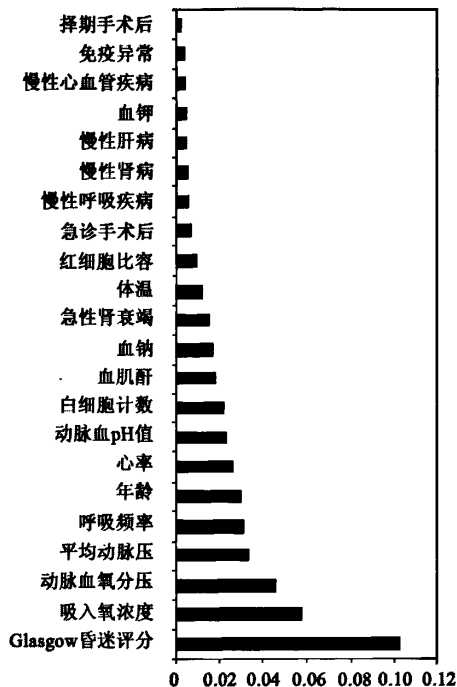


图1 APACHE II的22个变量及其信息增益

个和8个变量建立 ANN10 和 ANN8。信息增益是通过比较变量使分类资料不确定性降低的程度来衡量变量的重要性,变量的信息增益越大,其对模型预测能力的影响越大^[14]。

分别应用 APACHE II 和 ANN 22、ANN10、ANN8 预测检验集患者的死亡风险。

1.3 统计学处理 用 SPSS13.0 软件,ROCKIT (Version 0.91 Beta) 软件和 STATA9.0 软件进行统计分析。计量资料数据以 $\bar{x} \pm s$ 表示,差异统计学意义比较用 t 检验,计数资料比较用 χ^2 检验。采用受试者工作特征曲线下面积(aROC)比较各模型的预测能力。

2 结果

2.1 一般临床资料 共入选 177 名患者,其中男 144 人,女 33 人。观察期间患者死亡 36 例,死亡率为 20.3%。存活患者的平均年龄为 (77.91 ± 6.62) 岁,低于死亡患者的 (81.78 ± 7.99) 岁 ($P < 0.01$)。存活患者住 ICU 时间为 (7.76 ± 10.91) d,明显短于死亡患者的 (53.31 ± 79.72) d ($P < 0.01$)。

2.2 ANN 训练集和检验集患者资料 见表 1,两个样本集间没有显著性差异。

2.3 ANN 和 APACHE II 的预测能力 采用 ROCKIT (Version 0.91 Bata) 软件对各预测模型的预测结果进行 ROC 分析,使用微软 Excel 2003 绘制 ROC 曲线图(图 2),用 aROC 评价预测模型的预测能力,应用 STATA9.0 软件对 ROC 进行比较分析。根据 Youden 指数/正确指数 [Youden 指数 = 灵敏度 - (1 - 特异度)] 的最大值得出诊断临界点,分别计算出 APACHE II 和 ANN 22、ANN10、ANN8 的特异度、灵敏度,详见表 2。

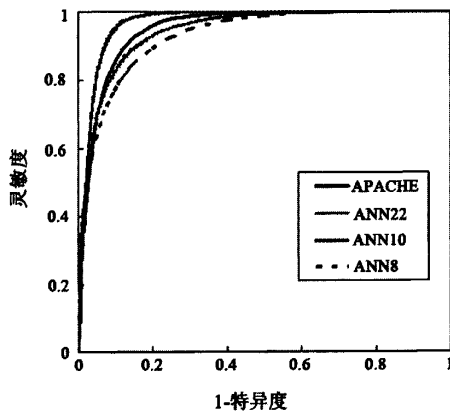


图2 APACHE II 和 ANN ROC 的比较

表1 训练集和检验集患者资料的比较

样本集	年龄(岁)	男性[例(%)]	女性[例(%)]	GICU 死亡率	住 GICU 时间(d)	APACHE II 评分
训练集	78.04	94(80.3%)	23(19.7%)	24(20.5%)	18.35	16.35
检验集	79.97	50(83.3%)	10(16.7%)	12(20%)	10.60	17.93
P 值	0.087		0.539	0.936	0.189	0.075

表2 APACHE II 和 ANN 的预测能力

评分	aROC	特异度	灵敏度	Youden 指数	诊断临界点	P 值
APACHE II	0.949	89.0%	86.8%	0.758	18.5	
ANN22	0.943	86.5%	87.5%	0.740	0.131	0.829
ANN10	0.968	89.5%	95.6%	0.851	0.687	0.257
ANN8	0.926	87.7%	80.8%	0.685	0.333	0.539

3 讨论

ANN 具有独特的信息存储方式、良好的容错性、大规模的并行处理方式以及强大的自组织、自学习和自适应能力,因此可能替代传统的预后预测模型。Dreiseitl 等^[15]回顾分析了 72 项研究,比较应用 ANN 和 Logistic 回归模型来判断预后的准确性。其中 51% 的研究发现 ANN 预测能力优于 Logistic 回归模型,42% 的研究显示二者之间的预测能力无差异,另外 7% 的研究显示 Logistic 回归模型更好。目前在 ICU 中,APACHE II 是应用最广泛的预后预测模型。存活患者的平均 APACHE II 评分通常为 9~15;而死亡患者的 APACHE II 评分较高,为 19~25。北京大学第一医院 GICU 也证实 APACHE II 能够较准确地判断年龄>75 岁的老年内科重症患者的疾病严重程度和预后^[16]。然而,有研究表明 APACHE II 可能低估患者的死亡风险^[17],也可能高估患者的死亡风险^[18]。因此,笔者比较应用 ANN 和 APACHE II 预测 GICU 患者的死亡率。

通常认为,aROC 值 ≥ 0.8 时模型的预测较为可靠^[19],本研究 4 个模型的 aROC 均 >0.9 。笔者发现 ANN22($P=0.839$),ANN10($P=0.257$)和 ANN8($P=0.539$)的预测能力与 APACHE II 基本相当。笔者之所以建立三个 ANN 模型的原因在于 ANN 的输入变量是待分析系统的内生影响因子,若输入变量过多,可增加 ANN 的系统误差。

本研究最重要的结果在于证实了在预测 GICU 的死亡率方面,ANN 的预测能力与 APACHE II 基本相当,但是 ANN8 和 ANN10 所需的变量明显少于 APACHE II 所需的变量。上述结果与 Wong 等^[12]的研究结果相似,即一些 APACHE II 模型所需的变量是多余的,并不能提高 ANN 的预测能力,因此可以在构建 ANN 时去掉那些对于预测影响较小的变量。值得一提的是,Frize 等^[9]的研究仅用了 APACHE II 中的 6 个重要变量(心率、呼吸频率、吸入氧浓度、动脉血氧分压、动脉血 pH 值和 Glasgow 昏迷评分),其 ANN 的准确性仍较高。显然,在预测死亡风险方面 ANN 比 APACHE II 更有优势,它可以显著减少临床数据收集的时间和成本。

此外,ANN 结构的确定缺乏指导原则,如隐含层层数和神经元个数的选择,网络的训练等。尤其是隐含层神经元个数的选择非常重要,它对建立的 ANN 的性能影响很大,但是目前理论上还没有一

种科学的和普遍的确切方法,一般是根据经验或者通过反复试验确定。因此,本研究还在建立 ANN22 时分别试用 20,30,50 和 60 个隐含层神经元,预测结果均不如根据 Kolmogorov 定理所构建的 ANN。

本研究提示,应用 ANN 预测 GICU 老年患者死亡率与应用以 Logistic 回归为基础的 APACHE II 相比,结果相当,且减少变量的个数不影响 ANN 预测的准确性。因为模型所需变量明显减少,可以大大减少临床工作者收集数据的时间和成本。ANN8 所需的变量中,心率、年龄、呼吸频率、平均动脉压、吸入氧浓度和 Glasgow 昏迷评分在患者入院即刻就可以采集到,动脉血 pH 值和动脉血氧分压两个变量在抽取患者动脉血后的 5min 内就能得到结果;而 ANN10 增加的两个变量白细胞计数和血肌酐的结果亦可以在 30min 内获得。这样,临床医师在短时间内即可完成患者死亡风险的评估,根据预测结果及时甄别危重患者,加强医疗救治的针对性,并可以根据预测结果帮助患者家属参与做出有关治疗的重大决定。

由于北京大学第一医院 GICU 应用 APACHE II 对患者进行死亡风险预测的时间较短,且床位有限,故研究对象的样本量较小,这是本研究的不足。在 ANN 成为公认的预测模型之前,仍需不断地研究与探索,需要更多大样本的多中心临床研究进一步验证 ANN 的有效性。

参考文献

- [1] Arabi Y, Haddad S, Goraj R, et al. Assessment of performance of four mortality prediction systems in a Saudi Arabian intensive care unit. *Crit Care*, 2002, 6: 166-174.
- [2] Chiavone PA, Sens YA. Evaluation of APACHE II system among intensive care patients at a teaching hospital. *Sao Paulo Med J*, 2003, 121:53-57.
- [3] Chan CH, Chan EY, Ng DK, et al. Application of artificial neural networks to establish a predictive mortality risk model in children admitted to a paediatric intensive care unit. *Singapore Med J*, 2006, 47:928-934.
- [4] Le Gall JR, Neumann A, Hemery F, et al. Mortality prediction using SAPS II; an update for French intensive care units. *Crit Care*, 2005, 9:645-652.
- [5] Buchman TG, Kubos KL, Seidler AJ, et al. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Crit Care Med*, 1994, 22:750-762.

- [6] Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res*, 1993, 26;220-229.
- [7] Penny W, Frost D. Neural networks in clinical medicine. *Med Decis Making*, 1996, 16;386-398.
- [8] Clermont G, Angus DC, DiRusso SM, et al. Predicting hospital mortality for patients in the intensive care unit; a comparison of artificial neural networks with logistic regression models. *Crit Care Med*, 2001, 29; 291-296.
- [9] Frize M, Ennett CM, Stevenson M, et al. Clinical decision support systems for intensive care units; using artificial neural networks. *Med Eng Phys*, 2001, 23; 217-225.
- [10] Nimgaonkar A, Karnad DR, Sudarshan S, et al. Prediction of mortality in an Indian intensive care unit. Comparison between APACHE II and artificial neural networks. *Intensive Care Med*, 2004, 30;248-253.
- [11] Hanson CW 3rd, Marshall BE. Artificial intelligence applications in the intensive care unit. *Crit Care Med*, 2001, 29;427-435.
- [12] Wong LS, Young JD. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks. *Anaesthesia*, 1999, 54; 1048-1054.
- [13] Knaus WA, Draper EA, Wagner DP, et al. APACHE II; a severity of disease classification system. *Crit Care Med*, 1985, 13;818-829.
- [14] Tom M. *Decision Tree Learning*. Machine Learning. New York; McGraw-Hill, 1997. 55-60.
- [15] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*, 2002, 35; 352-359.
- [16] 周国鹏, 宋以信, 赵志杰. APACHE 评价系统对 75 岁以上老年重症病人预后评价价值. *中华内科杂志*, 2005, 44;251-253.
- [17] Lertakyamanee J, Somprakit P, Vorakitpokaton P, et al. APACHE II in a postoperative intensive care unit in Thailand. *J Med Assoc Thai*, 1997, 80; 169-177.
- [18] Moreno R, Morais P. Outcome prediction in intensive care; results of a prospective, multicenter, Portuguese study. *Intensive Care Med*, 1997, 23; 177-186.
- [19] Knaus WA, Wagner DP, Draper EA, et al. The APACHE II prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 1991, 100;1619-1636.

(上接第 404 页)

参考文献

- [1] Planel E, Richter KE, Nolan CE, et al. Anesthesia leads to tau hyperphosphorylation through inhibition of phosphatase activity by hypothermia. *J Neurosci*, 2007, 27; 3090-3097.
- [2] Panza F, D'Introno A, Colacicco AM, et al. Lipid metabolism in cognitive decline and dementia. *Brain Res Rev*, 2006, 51; 275-292.
- [3] 李光伟, 潘孝仁, Lillioja S. 检测人群胰岛素敏感性的一项新指数. *中华内科杂志*, 1993, 32; 656-660.
- [4] Obisesan TO. Hypertension and cognitive function. *Clin Geriatr Med*, 2009, 25; 259-288.
- [5] Cole GM, Frautschy SA. The role of insulin and neurotrophic factor signaling in brain aging and Alzheimer's disease. *Exp Gerontol*, 2007, 42; 10-21.
- [6] Hämäläinen A, Grau-Olivares M, Tervo S, et al. Apolipoprotein E epsilon 4 allele is associated with increased atrophy in progressive mild cognitive impairment; a voxel-based morphometric study. *Neurodegener Dis*, 2008, 5; 186-189.
- [7] Craft S, Watson GS. Insulin and neurodegenerative disease; shared and specific mechanisms. *Lancet Neurol*, 2004, 3; 169-178.
- [8] Zhao WQ, Alokun DL. Role of insulin and insulin receptors in learning and memory. *Mol Cell Endocrinol*, 2001, 177; 125-134.
- [9] Hoyern S. Brain glucose and energy metabolism abnormalities in sporadic Alzheimer's disease. Causes and consequences; an update. *Exp Gerontol*, 2000, 35; 1363-1372.
- [10] Fernandes ML, Saad MJ, Velloso LA. Effects of age on elements of insulin signaling pathway in central nervous system of rats. *Endocrine*, 2001, 16; 227-234.