· 流行病学研究 ·

BP 人工神经网络与 logistic 回归模型在石化企业员工蓄积性疲劳预测中比较：基于网络问卷调查

高鑫妮[1]，万保玉[2]，王雨潇[3]，王瑾[4]，杜鑫[5]，张巧耘[2]，尹立红[1]

1. 东南大学公共卫生学院 环境医学工程教育部重点实验室，南京 210009；
2. 江苏省疾病预防控制中心，南京 210009；
3. 深圳市职业病防治院，深圳 518020；
4. 中国疾病预防控制中心职业卫生与中毒控制所，北京 100050；
5. 中国石油化工股份有限公司金陵分公司，南京 210033
通信作者：尹立红，E-mail：lhyin@seu.edu.cn；张巧耘，E-mail：qycdc@163.com

【摘　要】目的　比较 BP 人工神经网络（BP-ANN）与 logistic 回归模型预测石化企业员工蓄积性疲劳风险的效果；评价石化企业员工蓄积性疲劳早期识别、干预的预测模型。**方法**　于 2021 年 7—10 月采用整群抽样方法在江苏省某石化企业抽取 4 066 名员工进行横断面研究。通过网络问卷方式调查员工的一般情况、工作情况、职业紧张状况、蓄积性疲劳患病情况。采用 SPSS 23.0 软件，通过非参数检验分析石化员工蓄积性疲劳的影响因素。建立 BP-ANN 和 logistic 回归两种预测模型并绘制 ROC 曲线，计算灵敏度、特异性、约登指数和 AUC 值。比较两者的预测效能。**结果**　共回收有效问卷 3 763 份。石化企业员工中蓄积性疲劳者占 63.2%（2 377/3 763）。综合两种模型分析结果，年龄、周均工作时间、轮班、有职业紧张和规律的体育锻炼是员工蓄积性疲劳的主要影响因素。两种模型的 AUC 值均>0.7，两种模型在灵敏度（BP 75.7%，logistic 75.6%）方面的表现相近。与 logistic 回归模型相比，BP-ANN 模型在特异性（BP 76.3%，logistic 69.3%）、约登指数（BP 51.9%，logistic 44.9%）和 AUC 值（BP 84.4%，logistic 79.9%）方面表现更佳。两种模型均可用于预测石化员工蓄积性疲劳发生率，但 BP-ANN 模型的总体效能优于传统 logistic 回归模型。**结论**　BP-ANN 模型和 logistic 回归模型可应用于石化企业员工蓄积性疲劳的预测；BP-ANN 模型能明显提高预测效能。

【关键词】蓄积性疲劳；石化企业员工；风险预测模型；BP 人工神经网络（BP-ANN）；logistic 回归

Comparison of back propagation artificial neural network and logistic regression models for predicting cumulative fatigue among petrochemical employees: a web-based questionnaire survey

GAO Xinni[1], WAN Baoyu[2], WANG Yuxiao[3], WANG Jin[4], DU Xin[5], ZHANG Qiaoyun[2], YIN Lihong[1]
(*1. Key Laboratory of Environmental Medicine Engineering of Ministry of Education*, *School of Public Health*, *Southeast University*, *Nanjing* 210009, *China*; *2. Jiangsu Provincial Center for Disease Control and Prevention*, *Nanjing* 210009, *China*; *3. Shenzhen Prevention and Treatment Center for Occupational Diseases*, *Shenzhen* 518020, *China*; *4. National Institute for Occupational Health and Poison Control*, *Chinese Center for Disease Control and Prevention*, *Beijing* 100050, *China*; *5. Sinopec Jinling Company*, *Nanjing* 210033, *China*)
Corresponding authors: YIN Lihong, E-mail: lhyin@seu.edu.cn; ZHANG Qiaoyun, E-mail: qycdc@163.com

【Abstract】 Objective To compare the performance of back propagation artificial neural network (BP-ANN) and logistic regression models in predicting the risk of cumulative fatigue among petrochemical employees and evaluate the prediction models for early identification and intervention of cumulative fatigue in these employees. **Methods** A total of 4 066 employees of a petrochemical enterprise in Jiangsu province were selected by cluster sampling for a cross-sectional study from July to October 2021. A web-based questionnaire was designed to investigate the general information, work conditions, occupational stress, and prevalence of cumulative fatigue among employees. SPSS 23.0 was used to analyze the influencing factors of cumulative fatigue among petrochemical employees through non-parametric tests. BP-ANN and logistic regression models were established and receiver operating characteristic (ROC) curves were plotted to compare the prediction performance of models based on sensitivity, specificity, Youden index, and area under the curve (AUC)

开放获取：CC BY-NC-ND 4.0　DOI：10.11847/zgggws1145418
基金项目：中国疾控中心职业卫生所职业健康风险评估与国家职业卫生标准制定（131031109000160004）；中国石油化工股份有限公司金陵分公司横向课题（31700000-19-ZC0607-0013）
第一作者：高鑫妮（1998—），硕士在读，研究方向：职业健康促进与职业紧张。
收稿日期：2025-04-25　修回日期：2025-08-05　录用日期：2025-09-03
利益冲突：不存在　伦理审查：已获取　出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

value. **Results** A total of 3 763 valid questionnaires were collected, which showed that 63.2% (2 377/3 763) of petrochemical employees experienced cumulative fatigue. The comprehensive analysis of the two models indicated that age, weekly working hours, shift work, occupational stress, and regular exercise were the key factors influencing employees' cumulative fatigue. Both models had AUC values > 0.7. Sensitivity was comparable between the BP-ANN and logistic models (BP 75.7%, logistic 75.6%). The BP-ANN model achieved higher specificity (BP 76.3%, logistic 69.3%), Youden index (BP 51.9%, logistic 44.9%), and AUC value (BP 84.4%, logistic 79.9%) than the logistic model. Both models can be used to predict the incidence of cumulative fatigue among petrochemical employees, while the overall prediction performance of the BP-ANN model was superior to that of the logistic model. **Conclusions** Both BP-ANN and logistic regression models can be adopted to predict cumulative fatigue among petrochemical employees, with the BP-ANN model showing better prediction performance.

【Keywords】 cumulative fatigue; petrochemical employees; risk prediction models; back propagation artificial neural network (BP-ANN); logistic regression

近年来，职业人群的过劳问题越来越受到社会关注。随着中国经济的快速发展和社会的剧烈变化，劳动者因过劳而导致的极端现象“过劳死”常有发生，并且因过劳导致身心健康受损的现象也较为普遍[1]。蓄积性疲劳是指由于长时间的体力消耗、长期的情绪压力或睡眠不足，导致个人身体机能下降、健康状况恶化、情绪紊乱或工作效率降低的一种状态[2]。蓄积性疲劳是“过劳”的早期表现，如能及时实施针对性干预，可减缓对身心健康的损害，降低由此导致的工作失误和事故，并有效遏制“过劳死”的发生[3]。导致蓄积性疲劳的危险因素较为复杂，通过构建预测模型，可精准评估个体在特定情境下蓄积性疲劳风险。人工神经网络（artificial neural network, ANN）可以克服线性模型的局限性[4]，并分析多个输入输出变量之间的强耦合与高度非线性相关性。因此，在许多领域得到应用。在非线性人工神经网络模型中，特别是 BP 人工神经网络（back propagation artificial neural network, BP-ANN），BP-ANN 模型具有自学习性、结构简单、仿真能力强等特点，可以提高预测精度[5]。石化企业员工工作任务繁重，除接触职业性危害因素外，还存在轮班、长工时、高职业紧张等各种健康危害因素，长期承受着较高的生理和心理负荷[6]。有研究显示，江苏省某石化企业员工整体的蓄积性疲劳程度处在较高的水平[7]。为全面预防员工的蓄积性疲劳，有必要探讨危险因素的影响，并开发蓄积性疲劳的早期预测模型。目前，关于 BP-ANN 在蓄积性疲劳中的具体应用研究较为有限。本研究采用 BP-ANN 模型构建石化企业员工蓄积性疲劳预测模型，并与传统 logistic 回归模型比较，以期为该人群通过适宜的蓄积性疲劳预测模型应用、精准预测员工的蓄积性疲劳发生风险。同时可根据预测结果提出针对性的干预措施，为企业管理者提供科学的决策依据。

1 对象与方法

1.1 *研究对象* 采用整群抽样法，2021 年 7—10 月通过企业版问卷星对江苏省某石油化工企业的全体员工进行问卷调查。研究对象纳入遵循以下标准：(1)员工需在本岗位工作至少 6 个月；(2)年龄需满≥18 岁的在岗员工；(3)无精神病史；(4)自愿参与调查，并在了解研究内容后签署知情同意书。本次共收集 4 066 份问卷，有效问卷 3 763 份，有效回收率 92.5%。本研究已通过中国疾病预防控制中心职业卫生与中毒控制所医学伦理委员会审批（批号：NIOHP202108）。

1.2 *方法*

1.2.1 基本情况调查 采用由中国疾病预防控制中心职业卫生与中毒控制所编制的基本情况调查表，收集研究对象的一般人口学和个人健康行为等数据，包括年龄、性别、教育程度、婚姻状况以及体育锻炼等信息。

1.2.2 岗位特征情况调查 采用项目组自行设计的问卷调查员工岗位特征信息，包括职务、岗位工龄、周均工作时间、是否上夜班及是否轮班等。

1.2.3 职业紧张状况调查 采用职业紧张核心量表（Core Occupational Stress Scale, COSS）评估受访者的职业紧张状况[8]。该量表包括 4 个维度的 17 个条目，采用 5 级赋分法，赋值范围为 1（完全不同意）～5 分（非常同意）。反向计分条目（社会支持和自主性）转换（6–实际得分）后计算总分，总分为 17～85 分.以职业紧张总分≥50 判定为有职业紧张；总分越高，表明职业紧张程度越严重。本研究中该量表的 Cronbach′s α 系数为 0.865。

1.2.4 蓄积性疲劳程度测量 采用“劳动者疲劳蓄积度自己诊断调查表”评估研究对象的蓄积性疲劳情况[9]。该量表由两部分组成.第一部分评估疲劳自感症状，包括 13 个条目，每个条目按 0～3 分制计分，将各条目的合计总分分为以下 4 个

等级：Ⅰ级（0～4 分）、Ⅱ级（5～10 分）、Ⅲ级（11～20 分）和Ⅳ级（21 分或以上）；第二部分评估工作状况，包括 7 个条目，评分方法同第一部分。合计总分分为 4 个等级：0 分为 A 级，1～2 分为 B 级，3～5 分为 C 级，6 分或以上为 D 级。根据“工作负担分数表[9]”测算“工作负担度分数”，得分≥2 分表示存在蓄积性疲劳，分数越高员工的蓄积性疲劳程度越高。本研究中该量表的 Cronbach's α 系数为 0.937。

1.3 质量控制　(1)调查前期：与被调查单位沟通以获得支持和配合；根据预调查结果完善调查问卷及调查方案。培训调查员，统一调查要求；借助企业版问卷星平台，建立统一的在线数据采集平台，设置逻辑项和漏题提醒。(2)调查中期：员工集中调查，介绍调查目的，明确注意事项，及时解答，给予足够的填表时间，不催促，减少环境因素带来的影响。(3)调查后期：问卷经系统初步审核后，由专业人员 2 次审核。

1.4 数据整理　导出数据到 Excel。检查数据一致性与完整性，剔除重复数据、逻辑不符、明显异常值等无效数据，对数据进行清洗。

1.5 统计分析　使用 SPSS 23.0 软件进行统计分析。

1.5.1 描述性分析　采用百分比形式描述计数资料。对于计量资料，先进行 *K-S* 正态性检验，数据不符合正态分布（P<0.05），以 [M（P_{25}～P_{75}）] 的形式表示；采用 Mann-Whitney U 检验或 Kruskal-Wallis H 检验单因素分析蓄积性疲劳的影响因素，检验统计量为 Z 值或 H 值。检验水准 α=0.05。

1.5.2 logistic 回归预测模型构建　基于单因素分析结果，筛选出差异有统计学意义（P<0.05）的影响因素作为输入变量（自变量），以是否出现蓄积性疲劳作为输出变量（因变量），采用向前逐步回归法构建 logistic 回归预测模型。模型公式为 logit P=$\beta_0+\beta_1X_1+\ldots+\beta_nX_n$，在该模型中，$P$ 为预测概率，其范围在 0～1，X 是自变量；预测概率 P 与 logit P 呈正相关，logit P 值越大，表明该石化企业员工发生蓄积性疲劳的风险越高。

1.5.3 BP-ANN 预测模型构建　采用多层感知器（MLP）构建神经网络模型。模型参数如下，样本拆分，按照 7:3 的比例分为训练集与测试集；输入层中的神经元为单因素分析中差异有统计学意义的影响因素；Li Z M 等[10] 的研究表明，一个包含单层隐藏层、足够数量的神经元、并使用非多项式激活函数的神经网络，能够逼近任意可测函数，达到任何所需的准确度。因此，本研究选择单隐藏层的设计策略。隐藏层中的神经元通过 MLP 中的自动估计算法确定，算法在给定神经元数量范围内逐步增设隐藏层神经元，选定测试误差最低的神经元数量。根据 SPSS 官方文档，本研究神经元数量范围设为 1～50 个，适合中等规模分类任务；输出层的神经元为预测变量。Hussein B M 等[11] 的研究表明，在训练机器学习模型时采用早停法可有效防止过拟合。本研究在神经网络训练过程中使用了停止准则，隐藏层每增加 1 个神经元使用留出的测试集计算误差，当发现测试集的误差在一个连续步骤中未减少，则停止迭代。模型训练采用反向传播算法。Sato N 等[12] 的研究显示，标度共轭梯度算法比现有的自适应方法更快地最小化训练损失。因此，在训练优化方面，本研究使用标度共轭梯度算法，加速收敛并减少训练误差。

1.5.4 模型预测性能比较　绘制两种模型的受试者工作特征曲线（receiver operating characteristic curve, ROC），并计算曲线下面积（area under the curve, AUC）值以及最大诊断分界值对应的灵敏度、特异性和约登指数，比较两种模型的预测效果，以评价模型优劣。

2 结　果

2.1 一般情况及蓄积性疲劳检出率（表 1）　纳入研究的 3 763 名员工中，男性 3 046 人（占 80.9%），女性为 717 人（占 19.1%）；年龄 [M（P_{25}～P_{75}）] 为 48.0（42.0～53.0）岁；工龄 [M（P_{25}～P_{75}）] 为 11.0（4.2～23.3）年；受教育程度主要为大专或高职及以下（占 67.3%）；以一般工作人员为主（占 78.7%）；大多数员工每周工作时间少于 40 h（占 62.4%）；轮班及夜班人员分别占 54.2% 及 58.0%；大多数员工有规律的体育锻炼习惯（占 66.9%）。职业紧张检出 1 293 人（占 34.4%），COSS 总分 [M（P_{25}～P_{75}）] 为 45.0（40.0～52.0）分。蓄积性疲劳检出 2 377 人（占 63.2%），分数 [M（P_{25}～P_{75}）] 为 3.0（0.0～5.0）分。除婚姻状况组别（Z/H=6.290）在蓄积性疲劳分布中差异无统计学意义（P=0.053）外，年龄（Z/H=35.890）、性别（Z/H= −10.403）、受教育程度（Z/H=100.390）、平均月收入（Z/H=41.490）、职务（Z/H=67.860）、规律的体育锻炼（Z/H= −11.210）、工龄（Z/H=47.830）、周均工作时间（Z/H=48.430）、夜班（Z/H= −24.360）、职业紧张（Z/H= −30.050）和轮班组别（Z/H= −24.850）的蓄积性疲劳分布比较，差异均有统计学意义（P<0.001）。

表 1 不同特征研究对象蓄积性疲劳得分比较 [$M(P_{25}\sim P_{75})$]

Table 1 Comparison of scores of cumulative fatigue among participants with different characteristics [$M\ (P_{25}\sim P_{75})$]

变量	人数	构成比/%	蓄积性疲劳			
			M	$P_{25}\sim P_{75}$	Z/H值	P值
年龄组/岁					35.890	<0.001
≤30	443	11.8	2.0	0.0～5.0		
31～<41	452	12.0	3.0	1.0～5.0		
41～<51	1 469	39.0	3.0	0.0～5.0		
≥51	1 399	37.2	2.0	0.0～4.0		
性别					−10.403	<0.001
男性	3 046	80.9	3.0	0.0～5.0		
女性	717	19.1	1.0	0.0～4.0		
学历					100.390	<0.001
初中及以下	161	4.3	4.0	0.0～5.0		
高中或中专	1 202	31.9	4.0	1.0～6.0		
大专或高职	1 171	31.1	3.0	0.0～5.0		
大学本科	992	26.4	2.0	0.0～4.0		
研究生及以上	237	6.3	2.0	0.0～4.0		
婚姻状况					6.290	0.053
未婚	374	9.9	2.5	0.0～4.0		
已婚	3 128	83.2	3.0	0.0～5.0		
离婚或丧偶	261	6.9	3.0	1.0～5.0		
平均月收入/元					41.490	<0.001
<3 000	59	1.6	3.0	1.0～5.0		
3 000～<5 000	635	16.9	2.0	0.0～4.0		
5 000～<7 000	891	23.7	3.0	0.0～5.0		
7 000～<9 000	896	23.8	3.0	0.0～6.0		
9 000～<11 000	729	19.4	3.0	1.0～5.0		
≥11 000	553	14.6	2.0	0.0～4.0		
职务					67.860	<0.001
一般工作人员	2 962	78.7	3.0	0.0～5.0		
班组长	553	14.7	4.0	1.0～6.0		
中层及以上干部	248	6.6	1.0	0.0～3.0		
规律的体育锻炼					−11.210	<0.001
否	1 245	33.1	4.0	1.0～6.0		
是	2 518	66.9	2.0	0.0～4.0		
岗位工龄/年					47.830	<0.001
≤5	1 033	27.5	2.0	0.0～4.0		
6～<11	850	22.5	3.0	1.0～6.0		
11～<21	777	20.7	3.0	0.0～5.0		
≥21	1 103	29.3	3.0	0.0～5.0		
每周工时/h					48.430	<0.001
≤40	2 349	62.4	2.0	0.0～4.0		
41～<51	686	18.2	3.0	1.0～5.0		
51～<61	527	14.1	3.0	1.0～5.0		
≥61	201	5.3	4.0	2.0～6.0		
夜班					−24.360	<0.001
否	1 582	42.0	1.0	0.0～3.0		
是	2 181	58.0	4.0	2.0～6.0		
轮班					−24.850	<0.001
否	1 722	45.8	1.0	0.0～3.0		
是	2 041	54.2	4.0	2.0～6.0		
职业紧张					−30.050	<0.001
否	2 470	65.6	1.5	0.0～4.0		
是	1 293	34.4	5.0	3.0～7.0		

2.2　员工蓄积性疲劳 logistic 模型分析(表 2)　根据石化企业员工蓄积性疲劳的多因素 logistic 回归模型分析结果，建立如下风险预测模型：logit P=−0.870+0.627×X_1+0.407×X_2+0.324×X_3+0.240×X_4+0.724×X_5+1.009×X_6+1.284×X_7−0.553×X_8+1.767×X_9；其中 X_1=年龄 31～<41 岁；X_2=年龄 41～<51 岁；X_3=性别男；X_4=工时 41～<51 h；X_5=工时 51～<61 h；X_6=工时 61 h 以上；X_7=轮班(是)；X_8=规律的体育锻炼(是)；X_9=职业紧张(是)。Hosmer-Lemeshow 拟合优度检验结果 P>0.05，表明模型拟合优度较好。

表 2　蓄积性疲劳的多因素 logistic 回归模型分析
Table 2　Multivariate logistic regression model analysis of cumulative fatigue

变量	β	$S_{\bar{x}}$	Wald χ^2值	P值	OR值	95%CI
年龄组/岁						
≤30					1.000	
31～<41	0.627	0.161	15.191	0.000	1.871	1.365～2.564
41～<51	0.407	0.127	10.222	0.001	1.502	1.171～1.928
性别						
女性					1.000	
男性	0.324	0.105	9.561	0.002	1.382	1.126～1.697
每周工时/h						
≤40					1.000	
41～<51	0.240	0.107	5.067	0.024	1.271	1.032～1.567
51～<61	0.724	0.125	33.455	0.000	2.063	1.614～2.637
≥61	1.099	0.209	27.576	0.000	3.002	1.992～4.525
规律的体育锻炼						
否					1.000	
是	−0.553	0.086	41.561	0.000	0.575	0.486～0.681
轮班						
否					1.000	
是	1.284	0.083	238.766	0.000	3.612	3.069～4.251
职业紧张						
否					1.000	
是	1.767	0.101	307.887	0.000	5.852	4.804～7.128
常量	−0.870	0.153	32.465	0.000		

注：引入年龄、性别、受教育程度、平均月收入、职务、规律的体育锻炼、工龄、周工作时间、夜班、职业紧张、轮班变量进行调整，表中仅列出差异有统计学意义的结果。

2.3　蓄积性疲劳 BP-ANN 模型分析　将单因素分析结果中员工蓄积性疲劳差异有统计学意义的变量作为预测变量构成模型输入层的 11 个输入神经元。根据 SPSS 软件中的 BP-ANN 系统，为该隐藏层自动匹配 8 个神经元。输出层由两个神经元组成，是否发生蓄积性疲劳(未发生=0，发生=1)。将 3 763 个样本，采用 7:3 的分割比例随机抽取 2 601 个样本纳入训练集，其余 1 162 个样本分配到测试集，得到一个输入层数为 1、输入层神经元个数为 11，隐藏层数为 1、隐藏层神经元个数为 8，输出层数为 1、输出层神经元个数为 2 的神经网络模型。模型自变量标准化重要性排序依次为职业紧张、周均工作时间、轮班、年龄、平均月收入、规律的体育锻炼、学历、工龄、职务、夜班和性别。

2.4　两种模型预测效果比较(表 3，图 1、2)　ROC 曲线分析显示，logistic 回归模型 AUC 值为 79.9%，灵敏度为 75.6%，特异性为 69.3%，约登指数为 44.9%；BP-ANN 模型的 AUC 值为 84.4%，灵敏度为 75.7%，特异性为 76.3%，约登指数为 51.9%。AUC 值取值范围为 0.5～1.0，AUC 值越高，模型整体性能越好；灵敏度、特异性、约登指数越高，模型对正样本、负样本、正负样本整体预测的能力越好。两模型 AUC 值均>0.7，表明两种模型在预测员工蓄积性疲劳方面均展现出较好的效果，且 BP-ANN 的 AUC 值更高。同时，BP-ANN 模型灵敏度、特异性、约登指数均高于 logistic 回归模型。表明 BP-ANN 模型预测性能更佳。

表 3 logistic 回归模型与 BP-ANN 模型预测性能比较
Table 3 Comparison of prediction performance of logistic regression and BP-ANN models

模型	灵敏度/%	特异性/%	约登指数/%	AUC/%
logistic回归模型	75.6	69.3	44.9	79.9
BP-ANN模型	75.7	76.3	51.9	84.4

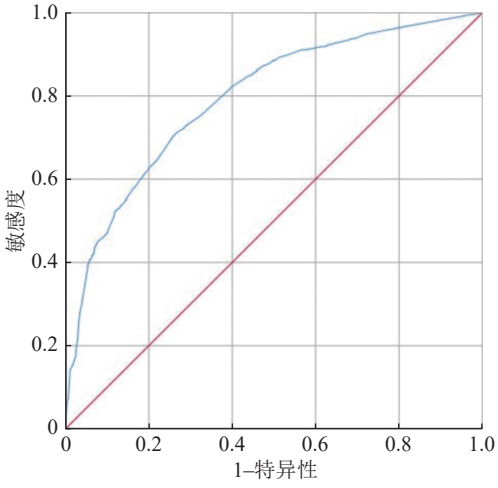


图 1 logistic 回归模型预测蓄积性疲劳的 ROC 曲线
Fig. 1 ROC curve of the logistic regression model in predicting cumulative fatigue

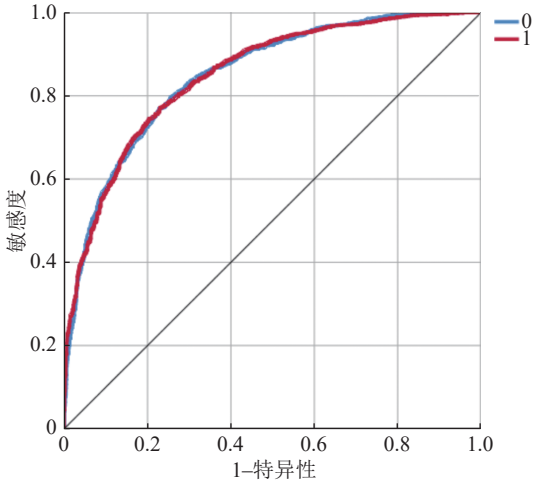


图 2 BP-ANN 模型预测蓄积性疲劳的 ROC 曲线
Fig. 2 ROC curve of the BP-ANN model in predicting cumulative fatigue

3 讨　论

本研究首次采用 BP-ANN 模型预测石化企业员工的蓄积性疲劳发生风险，并比较 BP-ANN 模型与 logistic 模型的预测性能。结果显示，BP-ANN 模型预测性能优于 logistic 模型，可用于员工蓄积性疲劳的早期筛查，避免发展为过劳等更严重的症状，帮助企业制定更有效的排班计划和干预策略，提升员工健康和生产效率并降低安全事故发生率。蓄积性疲劳的影响因素众多，作用方式复杂，不同因素存在交互作用[7]。只有在用户明确定义的情况下应用传统 logistic 回归模型预测对复杂关系和交互作用等进行建模[13]。且多变量间复杂的相互关系可导致分析误差增加[14]，存在较大局限性。人工神经网络能够识别和对潜在的复杂共线性关系建模，并能隐式检测交互作用。BP-ANN 模型适用于多变量间复杂的非线性分析，比传统模型解释力更强，而且不易受因素间干扰的影响。该模型具有方法学上的优势，使其在预测疾病方面更加有效[15–16]。关于蓄积性疲劳的研究比较有限，主要聚焦于蓄积性疲劳的影响因素研究，Lu Y S 等[17] 对广东省医疗卫生机构职工的蓄积性疲劳影响因素进行了研究，但并未建立蓄积性疲劳的早期预测模型。相关文献显示，在处理复杂数据时，BP-ANN 预测性能往往优于 logistic 回归模型，如周欣彤等[18] 对医院感染病例的预测研究，Yao L 等[19] 对糖尿病视网膜病变的预测研究，刘佩芳等[20] 对口腔医生肌肉骨骼疾患及张鑫宇[[21] 对护士睡眠障碍风险预测研究中，BP-ANN 均展现了更好的预测效能，与本研究结果一致。

通过深度访谈发现，石化企业员工对蓄积性疲劳的筛查和预防意识较低。因此，建立预测模型，对早期识别和预防其是否患有蓄积性疲劳具有重要意义。本研究利用石化企业员工的人口学特征、工作特征、生活习惯等数据，训练并验证了蓄积性疲劳的 BP-ANN 预测模型。BP-ANN 的 AUC 值为 84.4%，高于 logistic 回归模型的 79.9%。与 logistic 回归模型相比，BP-ANN 具有较好的预测性能。本研究结果显示，63.2%（2 377/3 763）的石化企业员工发生了蓄积性疲劳，高于其他行业如初级卫生保健系统、医疗卫生系统员工的发生率[17,22]。可能是石化行业职业危险因素较多、易发生急性事故所致，加之职业紧张程度较高[23]，均易增加员工疲劳程度。本研究两种预测模型均显示职业紧张、周均工作时间、轮班、年龄、规律的体育锻炼为重要影响因素。其中高职业紧张、周均工作时间在 41 h 以上、参与轮班及年龄 31～50 岁是蓄积性疲劳的危险因素，与相关文献结果一致[24]。而规律的体育锻炼则是员工蓄积性疲劳的保护因素，与万保玉等[7] 的研究一致。

本研究结果提示，应该从个人、企业两个层面着手，预防石化企业员工的蓄积性疲劳。首先，从员工个人角度出发，提高职业健康和安全意识，定期参加体育锻炼，提升抗压能力。不规律的轮班和夜班工作，易扰乱睡眠节律，而充足的睡眠可有效缓解疲劳；其次，企业、组织层面可以通过以下途径缓解员工疲劳：(1)根据员工的年龄、工作量、工作性质和特点，科学评估考核频率；(2)合理分配工作任务，有效地减轻工作带来

的压力，使员工在良好的情绪下高效工作；(3)关注员工的职业紧张状况，针对高职业紧张分布的岗位员工及时提供相应的心理援助，针对已经出现职业紧张的员工考虑实行岗位适配性调控。在蓄积性疲劳的初期阶段，个人、组织和企业应努力缓解员工的蓄积性疲劳，以防止症状加重。

本研究仍存在一定的局限性：(1)本研究仅限于石化企业员工的蓄积性疲劳预测，推广到其他行业有待验证。(2)研究为横断面研究，需要未来的前瞻性研究以证实模型在企业中应用时对员工蓄积性疲劳的改善效果。(3)本研究纳入的输入变量有限，未来有必要探索与蓄积性疲劳更多相关的指标以提高模型的预测性能。

综上，BP-ANN 模型在预测石化企业员工蓄积性疲劳方面展现出了较高的拟合度和风险识别能力，其预测效能较好，且相较于 logistic 回归模型具有一定的优越性。可根据早期预测结果评估石化企业员工蓄积性疲劳发生风险，提示可能具有较好的应用前景。此外，还可以在数据挖掘方面通过机器学习结合其他模型实现更高级预测模型的创建，为职业人群蓄积性疲劳预防提供更科学合理的指导。

数据可用性声明　本研究的原始数据因涉及参与者隐私保护及企业保密信息无法公开共享，但经脱敏处理后的数据可在合理请求下向通信作者申请获取。申请需通过伦理委员会审批，并签署数据使用协议。

致谢　感谢中国疾病预防控制中心职业卫生所项目组专家的指导及参与本研究的企业人员的支持

参考文献

[1] 朱逸飞, 郝艳华, 张桥, 等. 医生过劳生成机理理论模型构建 [J]. 中国医院, 2024, 28(6): 7-11.

[2] Lu Y S, Li Z, Chen Q S, et al. Association of working hours and cumulative fatigue among Chinese primary health care professionals[J]. Frontiers in Public Health, 2023, 11: 1193942.

[3] Huang H W, Xia X, Zhao W X, et al. Overwork, job embeddedness and turnover intention among Chinese knowledge workers[J]. Asia Pacific Journal of Human Resources, 2021, 59(3): 442-459.

[4] Nejatdarabi S, Mohebbi M. Predicting the rehydration process of mushroom powder by multiple linear regression (MLR) and artificial neural network (ANN) in different rehydration medium[J]. Journal of Food Measurement and Characterization, 2023, 17(2): 1962-1973.

[5] 马晓梅, 隋美丽, 段广才, 等. 手足口病重症化危险因素 BP 神经网络模型预测分析 [J]. 中国公共卫生, 2014, 30(6): 758-761.

[6] 杜鑫, 高茜茜, 李胜男, 等. 某石化企业员工职业紧张状况及影响因素分析 [J]. 中国工业医学杂志, 2022, 35(4): 349-353.

[7] 万保玉, 苏禹, 高茜茜, 等. 生活满意度和轮班及其交互作用对石化企业员工蓄积性疲劳的影响 [J]. 环境与职业医学, 2023, 40(9): 1039-1045.

[8] 王瑾, 张巧耘, 陈惠清, 等. 中国职业人群职业紧张测量核心量表编制 [J]. 中华预防医学杂志, 2020, 54(11): 1184-1189.

[9] 黄河, 耿东, 丑纪岳. 疲劳蓄积度自测与过劳预防 [J]. 中国人力资源开发, 2009(8): 35-37,74.

[10] Li Z M, Li Y N. A comparative study on the prediction of the BP artificial neural network model and the ARIMA model in the incidence of AIDS[J]. BMC Medical Informatics and Decision Making, 2020, 20(1): 143.

[11] Hussein B M, Shareef S M. An empirical study on the correlation between early stopping patience and epochs in deep learning[J]. ITM Web of Conferences, 2024, 64: 01003.

[12] Sato N, Izumi K, Iiduka H. Scaled conjugate gradient method for nonconvex optimization in deep neural networks[J]. Journal of Machine Learning Research, 2024, 25(395): 1-37.

[13] Sutradhar R, Barbera L. Comparing an artificial neural network to logistic regression for predicting ED visit risk among patients with cancer: a population-based cohort study[J]. Journal of Pain and Symptom Management, 2020, 60(1): 1-9.

[14] 李禄伟, 黄倩, 施佳成, 等. 基于三种统计学方法构建的超重及肥胖人群高血压发病预测模型的分析比较 [J]. 现代预防医学, 2021, 48(11): 2061-2066.

[15] Wei W, Yang X. Comparison of diagnosis accuracy between a backpropagation artificial neural network model and linear regression in digestive disease patients: an empirical research[J]. Computational and Mathematical Methods in Medicine, 2021, 2021(1): 6662779.

[16] Rezaeian A, Rezaeian M, Khatami S F, et al. Prediction of mortality of premature neonates using neural network and logistic regression[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(3): 1269-1277.

[17] Lu Y S, Li Z, Fan Y T, et al. The mediating role of cumulative fatigue on the association between occupational stress and depressive symptoms: a cross-sectional study among 1 327 Chinese primary healthcare professionals[J]. International Journal of Environmental Research and Public Health, 2022, 19(23): 15477.

[18] 周欣彤, 于晓松. 神经网络模型建立及在医院感染病例预警中应用 [J]. 中国公共卫生, 2019, 35(4): 445-450.

[19] Yao L, Zhong Y, Wu J, et al. Multivariable logistic regression and back propagation artificial neural network to predict diabetic retinopathy[J]. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 2019, 12: 1943-1951.

[20] 刘佩芳, 沈波, 许旭艳, 等. 两种模型用于口腔医生工作相关肌肉骨骼疾患预测的比较 [J]. 中国工业医学杂志, 2022, 35(4): 301-306.

[21] 张鑫宇. 基于 logistic 回归和人工神经网络的安宁疗护试点病房护士睡眠障碍风险预测 [D]. 锦州: 锦州医科大学, 2023.

[22] 李智, 卢瑀石, 范玉婷, 等. 心理资本在卫生系统职工职业紧张与蓄积性疲劳关系中的中介作用 [J]. 现代预防医学, 2023, 50(1): 53-59.

[23] 苏禹, 高茜茜, 刘晓曼, 等. 南京市某石化企业员工职业紧张状况及其对早期健康效应影响 [J]. 中国公共卫生, 2023, 39(7): 870-876.

[24] 王瑾, 李霜, 刘晓曼, 等. 互联网企业员工疲劳及影响因素研究 [J]. 中华劳动卫生职业病杂志, 2018, 36(12): 904-907.