

# 基于贝叶斯网络的缺血性脑卒中筛查模型构建

张伯韬<sup>1,2,3</sup>, 仇率杰<sup>1,2,3</sup>, 孙爽爽<sup>1,2,3</sup>, 袁莹<sup>1,2,3</sup>, 胡锡峰<sup>1,2,3</sup>,  
贾晓峰<sup>4</sup>, 于媛媛<sup>2,5</sup>, 薛付忠<sup>1,2,3</sup>

(1.山东大学齐鲁医学院公共卫生学院生物统计学系,山东 济南 250012;

2.山东大学齐鲁医学院公共卫生学院健康医疗大数据研究院,山东 济南 250003;

3.山东大学齐鲁医院,山东 济南 250012;4.博兴县卫生健康保障中心网络信息办,山东 滨州 256500;

5.山东大学数据科学研究院,山东 济南 250100)

**摘要:**目的 依托大规模电子健康记录,结合贝叶斯网络不确定性推理的优势,构建缺血性脑卒中筛查模型。

**方法** 筛查模型开发队列来自于齐鲁全生命周期电子研究型数据库(Cheeloo Lifespan Electronic Health Research Data-library, Cheeloo LEAD),按照7:3比例分为训练集与测试集;外部验证队列来自国家健康医疗大数据研究院博兴合作中心数据库(博兴数据库)。采用单因素Logistic回归分析筛选与缺血性脑卒中发病显著相关的筛查因子,随后采用贝叶斯网络模型对筛查因子建模,利用禁忌搜索算法进行结构学习,利用贝叶斯估计算法进行参数学习,最终得到缺血性脑卒中筛查模型。从判别能力、校准能力两方面评价模型性能,并比较其与传统Logistic回归模型在缺血性脑卒中筛查中的表现。**结果** 开发队列共1 067 609例,31 019例患缺血性脑卒中;外部验证队列共386 773例,13 393例患缺血性脑卒中。经过单因素筛选得到67个筛查因子,最终构建的贝叶斯网络模型包括68个节点,440条有向边,其中缺血性脑卒中节点的父节点包括年龄、高血压病、缺血性心脏病、慢性下呼吸道疾病、其他脑血管病、发作性和阵发性疾患,累及认知、知觉、情绪状态和行为的症状和体征,训练集、测试集和外部验证队列的AUC分别为0.840(95%CI:0.838~0.843)、0.839(95%CI:0.836~0.843)和0.811(95%CI:0.808~0.814),模型的判别能力良好,并且校准能力仍旧表现较好。本研究构建的筛查模型在缺失数据下的表现仍优于传统的Logistic回归模型。**结论** 基于贝叶斯网络不确定性推理的优势,本研究成功构建了缺血性脑卒中筛查模型;模型具有较好的判别、校准能力,为早期缺血性脑卒中筛查提供了便捷、高效的方法。

**关键词:**电子健康记录;贝叶斯网络;Logistic回归;缺血性脑卒中;筛查模型

中图分类号:R743.3

文献标志码:A

## Development of the Bayesian network-based screening model for ischemic stroke

ZHANG Botao<sup>1,2,3</sup>, ZHANG Shuaijie<sup>1,2,3</sup>, SUN Shuangshuang<sup>1,2,3</sup>, YUAN Ying<sup>1,2,3</sup>, HU Xifeng<sup>1,2,3</sup>,  
JIA Xiaofeng<sup>4</sup>, YU Yuanyuan<sup>2,5</sup>, XUE Fuzhong<sup>1,2,3</sup>

(1. Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong, China; 2. Healthcare Big Data Research Institute, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250003, Shandong, China; 3. Qilu Hospital of Shandong University, Jinan 250012, Shandong, China; 4. Health and Wellness Assurance Center Network Information Office of Boxing County, Binzhou 256500, Shandong, China; 5. Data Science Institute, Shandong University, Jinan 250100, Shandong, China)

**Abstract: Objective** To develop a screening model for ischemic stroke by relying on large-scale electronic health

收稿日期:2024-07-12

基金项目:国家自然科学基金重点项目(82330108);国家自然科学基金面上项目(82173625);山东省重点研发计划项目(2021SFGC0504);中国博士后科学基金面上资助(2022M721921);山东省自然科学基金青年基金项目(ZR2023QH236)

通信作者:薛付忠。E-mail:xuefzh@sdu.edu.cn

于媛媛。E-mail:yu\_yy\_1993@163.com

records and combining the advantages of Bayesian network uncertainty inference. **Methods** The screening model derivation cohort was derived from the Cheeloo Lifespan Electronic Health Research Data-library (Cheeloo LEAD) and divided into training and testing sets in a 7:3 ratio. The external validation cohort was sourced from the Boxing Collaboration Center Database of the National Healthcare Big Data Research Institute (Boxing Database). The univariate Logistic regression analysis was used to screen for factors significantly associated with the ischemic stroke. These associated screening factors were used to develop the Bayesian network. The tabu search algorithm was employed for structure learning, while Bayesian estimation algorithm was used for parameter learning, ultimately leading to the development of the ischemic stroke screening model. The performance of the model was evaluated in terms of both discrimination and calibration abilities, and compared with the traditional Logistic regression model in screening for ischemic stroke. **Results** The derivation cohort included 1,067,609 individuals, among whom 31,019 suffered from ischemic stroke. The external validation cohort included 386,773 individuals, among whom 13,393 suffered from ischemic stroke. After the univariate screening, 67 screening factors were identified. The final Bayesian network model included 68 nodes and 440 directed edges. The parent nodes of the ischemic stroke node included age, hypertensive diseases, ischemic heart diseases, chronic lower respiratory diseases, other cerebrovascular diseases, episodic and paroxysmal disorders, and the symptoms and signs involved cognition, perception, emotional state and behavior. The AUC for the training set, testing set, and external validation cohort were 0.840 (95% CI: 0.838-0.843), 0.839 (95% CI: 0.836-0.843), and 0.811 (95% CI: 0.808-0.814), respectively, indicating good discrimination ability, and calibration ability also performed well. Our newly developed screening model continued to outperform the traditional Logistic regression screening model, even in the presence of missing data. **Conclusion** This study developed the ischemic stroke screening model with the advantage of Bayesian network uncertainty inference. The model has good discrimination and calibration abilities, providing a convenient and efficient method for early ischemic stroke screening.

**Key words:** Electronic health records; Bayesian network; Logistic regression; Ischemic stroke; Screening model

脑卒中以其高发病率、高患病率、高死亡率、高复发率和高致残率的特点,已成为世界上第二大常见疾病,分为出血性脑卒中和缺血性脑卒中<sup>[1]</sup>。2019年全球疾病负担卒中负担估计显示,卒中仍然是世界上第二大死亡原因,也是第三大残疾合并死亡原因,缺血性脑卒中占有所有新发卒中的62.4%<sup>[2]</sup>。全国高级卒中中心数据直报显示,我国2020年全年缺血性脑卒中占有所有卒中病种的72.72%<sup>[3]</sup>。因此,缺血性脑卒中高危个体的早发现、早诊断、早治疗是脑卒中防治的关键。

国外已经形成了稳定的脑卒中评分系统<sup>[4-8]</sup>,但主要适用于欧美人群。国内现有应用最为广泛的是缺血性心脑血管疾病10年风险预测模型<sup>[9]</sup>和China-PAR风险预测模型<sup>[10]</sup>,但二者均为长期预测,缺乏快速筛查缺血性脑卒中患者的能力。传统的筛查模型构建依托于横断面队列,通过问卷调查方式收集个体信息,耗时耗力并且样本量有限,限制了其外推能力。电子健康记录是以大规模长期随访方式非突发性收集的,包含患者的各方面医疗信息<sup>[11]</sup>,广泛应用到疾病风险预测中<sup>[12-13]</sup>,是改进传统筛查模型的契机。贝叶斯网络模型<sup>[14]</sup>凭借其不确定性推理的优势,已经在医学领域应用于风险评估和预测<sup>[15-16]</sup>,能够直观地展示网络结构以及变量之间

的概率依赖关系,并且相比于其他模型,能够在筛查因子缺失的情况下,利用现有变量进行推断。本研究依托大规模电子健康记录,基于贝叶斯网络模型构建缺血性脑卒中筛查模型,并从判别能力和校准能力两方面评价模型的性能。对开展廉价、高效率的缺血性脑卒中早期筛查,将被动医疗转变为主动健康管理具有重要意义。

## 1 资料与方法

### 1.1 资料来源

本研究开发队列来源于山东省齐鲁全生命周期电子研究型数据库(Cheeloo Lifespan Electronic Health Research Data-library, Cheeloo LEAD),收录了来自39个县(区)医疗服务机构的电子健康记录,涵盖2009年1月至2023年12月约500万样本的数据。外部验证队列基于国家健康医疗大数据研究院博兴合作中心数据库(博兴数据库),本研究包含了博兴县不同等级的医疗服务机构自2012年9月至2023年12月约70万样本的电子健康记录。Cheeloo LEAD数据库和博兴数据库包括个体基本信息、诊断记录、用药记录、手术记录、检验记录等变量,数据均通过链接医疗机构的电子健康记录系统

收集,并做标准化处理,确保数据的一致性。本研究使用个体年龄、性别以及诊断记录构建缺血性脑卒中筛查模型。

## 1.2 队列构建方法

开发队列回顾性地纳入了 Cheeloo LEAD 中 2015 年 1 月 1 日至 2017 年 12 月 31 日年龄  $\geq 18$  岁的个体。诊断记录中,疾病定义基于国际疾病分类第十版(ICD-10),缺血性脑卒中定义为 ICD-10 编码 I63,脑卒中定义为 ICD-10 编码 I60~I64。个体的基线时间:个体研究截止前的最后一次诊断时间,

其中,个体研究截止时间是诊断为缺血性脑卒中、死亡或 2017 年 12 月 31 日三者中最早的时间。排除标准:①在 2015 年 1 月 1 日之前死亡的个体;②基线时间之前患有脑卒中的个体;③2015 年 1 月 1 日至基线时间之间没有诊断记录的个体。最终开发队列包括 1 067 609 例,按 7:3 的比例随机分为训练集和测试集。外部验证队列中个体基线时间定义,队列纳入、排除标准及数据收集处理方式与开发队列类似。最终外部验证队列包括 386 773 例。上述两个队列的详细构建过程见图 1。

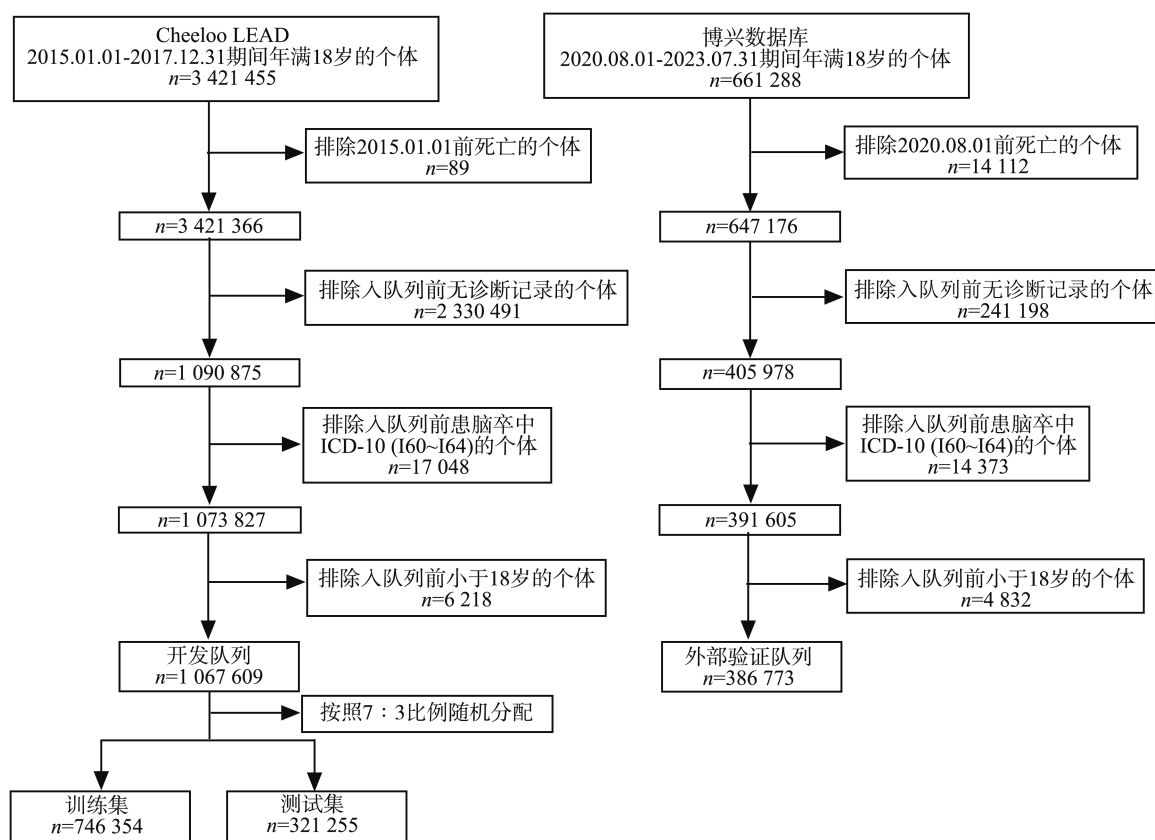


图 1 研究人群筛选过程流程图

Figure 1 Flow diagram of the screening process of the study population

## 1.3 统计学处理

### 1.3.1 数据预处理

年龄被处理为八分类变量: [18, 25)、[25, 35)、[35, 45)、[45, 55)、[55, 65)、[65, 75)、[75, 85), 以及  $\geq 85$  岁;性别被处理为二分类变量,男性为 1,女性为 0;诊断记录被处理为二分类变量,删除诊断时间异常的记录,凡是有明确诊断记录的均记为 1,否则记为 0。

### 1.3.2 队列基本情况描述

考虑到本研究的变量均为分类变量,采用频数和构成比 [ $n(\%)$ ] 描述开发队列与外部验证队列的基本情况,两组间的比较采用  $\chi^2$  检验。

### 1.3.3 单因素回归分析

采用单因素 Logistic 回归分析,在开发队列的训练集中筛选与缺血性脑卒中发病显著相关的筛查因子,并应用 Bonferroni 校正控制多重检验误差,并认为双侧  $P$  值  $< (0.05/\text{初始筛查因子总数})$  表示差异有统计学意义。

### 1.3.4 贝叶斯网络构建

贝叶斯网络模型是一种概率图模型,能够对一组有限随机变量的联合概率分布进行建模,通过有向无环图(directed acyclic graph, DAG)来描述变量之间的概率依赖关系。

在 DAG 中,节点代表随机变量集合  $U = \{X_1,$

$\dots, X_n\}$ , 而有向边 E 描述了变量之间的概率依赖关系。若有从  $X_i$  指向  $X_j$  的有向边, 那么  $X_i$  是父节点,  $X_j$  是子节点。每个节点都有一个条件概率分布表, 该表代表在给定父母节点状态下每个子节点的条件概率分布。贝叶斯网络可以表示为一组随机变量集合  $U = \{X_1, \dots, X_n\}$  的联合概率分布,  $\pi(X_i)$  代表的是节点  $X_i$  的所有父节点的集合。整个网络的联合概率分布可以通过所有节点基于其父节点状态的条件概率乘积来计算:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \pi(X_i))$$

构建贝叶斯网络模型包含结构学习和参数学习<sup>[17]</sup>两部分。使用基于搜索-评分<sup>[18]</sup>结构学习方法中的禁忌搜索算法<sup>[19]</sup>来学习网络结构。通过 200 次 Bootstrap<sup>[20]</sup> 采样学习, 得到高置信度连接作为网络的先验结构白名单, 并且设置年龄、性别节点不可被其他节点影响的约束黑名单, 从而获得高质量的鲁棒网络结构。采用贝叶斯估计算法进行参数学习, 计算网络中每个节点的条件概率表。

贝叶斯网络的概率推理是给出一组证据变量来计算查询变量的后验概率分布<sup>[21]</sup>。本研究推理过程采用基于采样算法的似然加权近似推理算法, 能够有效处理大型网络, 更好地反映数据的分布<sup>[22-23]</sup>, 推断个体是否患缺血性脑卒中, 实现筛查的目的。

### 1.3.5 模型评价

本研究使用 ROC 曲线下面积 (area under the curve, AUC)、校准曲线的截距和斜率来评估基于贝叶斯网络的缺血性脑卒中筛查模型的判别能力和校准能力, 并与传统的 Logistic 回归比较, 验证贝叶斯网络筛查模型在缺失数据背景下的应用潜力。将测试集与外部验证队列中除年龄、性别外的其他变量按照 50% 的比例随机缺失。比较直接使用缺失数据集的贝叶斯网络筛查模型和基于随机森林插补的 Logistic 回归筛查模型的表现, 采用 Delong 检验比较两个模型的 AUC 值。

上述方法均使用 R 软件 (4.3.0) 进行统计分析和数据可视化处理, 其中贝叶斯网络模型结构学习、参数学习及概率推理均通过 bnlearn 包实现, 测试集和外部验证队列的随机缺失与插补通过 missForest 包实现。

## 2 结果

### 2.1 队列的一般情况

开发队列包括 1 067 609 例, 其中 31 019 例患缺血性脑卒中。外部验证队列包括 386 773 例, 其中 13 393 例患缺血性脑卒中。训练集中共纳入初始筛查因子 170 个, 本文仅展示后续单因素分析筛选得到的与缺血性脑卒中发病显著相关的筛查因子。见表 1。

表 1 开发队列与外部验证队列基线特征/ $n$ (%)

Table 1 Baseline characteristics of the derivation cohort and external validation cohort / $n$ (%)

变量	开发队列( $n=1\ 067\ 609$ )		$P$	外部验证队列( $n=386\ 773$ )		$P$
	未患缺血性脑卒中 ( $n=1\ 036\ 590$ )	患缺血性脑卒中 ( $n=31\ 019$ )		未患缺血性脑卒中 ( $n=373\ 380$ )	患缺血性脑卒中 ( $n=13\ 393$ )	
年龄/岁			<0.001			<0.001
[18,25)	55 169(5.3)	134(0.4)		36 557(9.8)	26(0.2)	
[25,35)	301 009(29.0)	1 089(3.5)		60 112(16.1)	123(0.9)	
[35,45)	181 444(17.5)	1 259(4.1)		61 611(16.5)	398(3.0)	
[45,55)	193 227(18.6)	4 100(13.2)		86 504(23.2)	1 928(14.4)	
[55,65)	147 778(14.3)	8 012(25.8)		57 301(15.3)	2 889(21.6)	
[65,75)	100 689(9.7)	9 139(29.5)		48 865(13.1)	5 065(37.8)	
[75,85)	47 203(4.6)	6 092(19.6)		17 030(4.6)	2 462(18.4)	
$\geq 85$	10 071(1.0)	1 194(3.8)		5 400(1.4)	502(3.7)	
性别			<0.001			0.096
男	383 814(37.0)	15 004(48.4)		190 902(51.1)	6 946(51.9)	
女	652 776(63.0)	16 015(51.6)		182 478(48.9)	6 447(48.1)	
主要影响中枢神经系统的全身萎缩	204(0.0)	23(0.1)	<0.001	61(0.0)	7(0.1)	0.006
锥体外系和运动疾患	2 435(0.2)	334(1.1)	<0.001	520(0.1)	72(0.5)	<0.001
发作性和阵发性疾患	48 557(4.7)	4 533(14.6)	<0.001	28 062(7.5)	2 127(15.9)	<0.001
神经、神经根和神经丛疾患	8 612(0.8)	485(1.6)	<0.001	3 687(1.0)	217(1.6)	<0.001

续表

变量	开发队列 ( $n=1\ 067\ 609$ )		$P$	外部验证队列 ( $n=386\ 773$ )		$P$
	未患缺血性脑卒中 ( $n=1\ 036\ 590$ )	患缺血性脑卒中 ( $n=31\ 019$ )		未患缺血性脑卒中 ( $n=373\ 380$ )	患缺血性脑卒中 ( $n=13\ 393$ )	
多神经病和周围神经系统的其他疾患	2 908(0.3)	183(0.6)	<0.001	690(0.2)	64(0.5)	<0.001
神经系统的其他紊乱	6 280(0.6)	548(1.8)	<0.001	1 247(0.3)	82(0.6)	<0.001
晶状体疾病	21 674(2.1)	1 579(5.1)	<0.001	4 906(1.3)	528(3.9)	<0.001
视神经和视路疾患	1 279(0.1)	92(0.3)	<0.001	463(0.1)	49(0.4)	<0.001
外耳疾病	6 966(0.7)	263(0.8)	<0.001	3 015(0.8)	85(0.6)	0.031
非化脓性中耳炎	8 398(0.8)	311(1.0)	<0.001	8 253(2.2)	230(1.7)	<0.001
内耳疾病	6 588(0.6)	439(1.4)	<0.001	7 809(2.1)	726(5.4)	<0.001
高血压病	150 772(14.5)	12 673(40.9)	<0.001	65 896(17.6)	6 545(48.9)	<0.001
其他脑血管病	32 395(3.1)	5 747(18.5)	<0.001	12 326(3.3)	2 131(15.9)	<0.001
动脉、小动脉和毛细血管疾病	15 115(1.5)	1 162(3.7)	<0.001	2 989(0.8)	427(3.2)	<0.001
静脉、淋巴管和淋巴结疾病,不可归类在他处者	32 037(3.1)	1 317(4.2)	<0.001	11 952(3.2)	442(3.3)	0.538
循环系统其他和未特指的疾患	684(0.1)	70(0.2)	<0.001	178(0.0)	17(0.1)	<0.001
急性上呼吸道感染	12 8601(12.4)	4 860(15.7)	<0.001	153 356(41.1)	5 938(44.3)	<0.001
其他上呼吸道疾病	39 313(3.8)	1 349(4.3)	<0.001	65 448(17.5)	1 795(13.4)	<0.001
慢性下呼吸道疾病	72 012(6.9)	4 254(13.7)	<0.001	27 372(7.3)	1 291(9.6)	<0.001
缺血性心脏病	109 135(10.5)	9 537(30.7)	<0.001	26 720(7.2)	3 221(24.0)	<0.001
其他胸膜疾病	19 802(1.9)	671(2.2)	0.001	2 569(0.7)	149(1.1)	<0.001
口腔、唾液腺和颌部疾病	51 529(5.0)	1 975(6.4)	<0.001	24 778(6.6)	806(6.0)	0.005
食管、胃和十二指肠疾病	122 699(11.8)	5 553(17.9)	<0.001	72 956(19.5)	3 374(25.2)	<0.001
以皮肤和黏膜损伤为特征的病毒感染	12 488(1.2)	536(1.7)	<0.001	5 298(1.4)	209(1.6)	0.186
疝	7 070(0.7)	307(1.0)	<0.001	1 826(0.5)	112(0.8)	<0.001
非感染性小肠炎和结肠炎	17 953(1.7)	921(3.0)	<0.001	4 510(1.2)	174(1.3)	0.363
肝脏疾病	50 761(4.9)	2 092(6.7)	<0.001	11 183(3.0)	589(4.4)	<0.001
胆囊、胆道和胰腺疾病	30 271(2.9)	1 564(5.0)	<0.001	6 721(1.8)	396(3.0)	<0.001
皮肤炎和湿疹	32 436(3.1)	1 048(3.4)	0.014	15 931(4.3)	444(3.3)	<0.001
荨麻疹和红斑	7 528(0.7)	264(0.9)	0.012	2 928(0.8)	67(0.5)	<0.001
皮肤附件疾病	9 751(0.9)	278(0.9)	0.441	5 461(1.5)	76(0.6)	<0.001
炎性多关节病	22 168(2.1)	1 422(4.6)	<0.001	23 189(6.2)	1 044(7.8)	<0.001
关节病	22 424(2.2)	1 267(4.1)	<0.001	10 386(2.8)	535(4.0)	<0.001
脊椎病	41 774(4.0)	2 089(6.7)	<0.001	13 969(3.7)	722(5.4)	<0.001
其他软组织疾病	22 447(2.2)	885(2.9)	<0.001	13 158(3.5)	489(3.7)	0.447
肺源性心脏病和肺循环疾病	6 703(0.6)	459(1.5)	<0.001	1 704(0.5)	185(1.4)	<0.001
肾小球疾病	5 266(0.5)	363(1.2)	<0.001	591(0.2)	33(0.2)	0.017
肾小管-间质疾病	7 500(0.7)	228(0.7)	0.840	2 335(0.6)	79(0.6)	0.648
肾功能衰竭	8 255(0.8)	681(2.2)	<0.001	1 391(0.4)	149(1.1)	<0.001
肾和输尿管的其他疾患	16 124(1.6)	857(2.8)	<0.001	3 801(1.0)	272(2.0)	<0.001
泌尿系统的其他疾病	23 561(2.3)	1 392(4.5)	<0.001	11 140(3.0)	474(3.5)	<0.001
肠道传染病	11 979(1.2)	526(1.7)	<0.001	20 654(5.5)	443(3.3)	<0.001
累及循环和呼吸系统的症状和体征	64 061(6.2)	3 262(10.5)	<0.001	31 388(8.4)	1 208(9.0)	0.013
累及皮肤和皮下组织的症状和体征	12 455(1.2)	533(1.7)	<0.001	4 891(1.3)	255(1.9)	<0.001
累及神经和肌肉骨骼系统的症状和体征	1 287(0.1)	108(0.3)	<0.001	289(0.1)	28(0.2)	<0.001
累及认知、知觉、情绪状态和行为的症状和体征	32 473(3.1)	3 690(11.9)	<0.001	17 081(4.6)	1 377(10.3)	<0.001

续表

变量	开发队列( $n=1\ 067\ 609$ )		$P$	外部验证队列( $n=386\ 773$ )		$P$
	未患缺血性脑卒中 ( $n=1\ 036\ 590$ )	患缺血性脑卒中 ( $n=31\ 019$ )		未患缺血性脑卒中 ( $n=373\ 380$ )	患缺血性脑卒中 ( $n=13\ 393$ )	
一般症状和体征	192 925(18.6)	6 732(21.7)	<0.001	24 532(6.6)	1 031(7.7)	<0.001
消化器官恶性肿瘤	14 518(1.4)	581(1.9)	<0.001			
不明确、继发和未特指部位的恶性肿瘤	20 255(2.0)	727(2.3)	<0.001			
淋巴、造血和相关组织的恶性肿瘤	6 798(0.7)	138(0.4)	<0.001			
动态未定或动态未知的肿瘤	12 651(1.2)	413(1.3)	0.084	1 334(0.4)	64(0.5)	0.027
涉及免疫机制的某些疾患	6 067(0.6)	227(0.7)	0.001	1 299(0.3)	48(0.4)	0.898
其他急性下呼吸道感染	37 163(3.6)	1 795(5.8)	<0.001	12 442(3.3)	400(3.0)	0.030
其他类型的心脏病	42 952(4.1)	3 300(10.6)	<0.001	13 429(3.6)	1 619(12.1)	<0.001
良性肿瘤	26 776(2.6)	674(2.2)	<0.001	5 336(1.4)	202(1.5)	0.471
呼吸系统的其他疾病	28 238(2.7)	1 489(4.8)	<0.001	10 320(2.8)	446(3.3)	<0.001
肠的其他疾病	38 718(3.7)	1 390(4.5)	<0.001	15 604(4.2)	703(5.2)	<0.001
甲状腺疾病	39 238(3.8)	1 589(5.1)	<0.001	9 110(2.4)	411(3.1)	<0.001
糖尿病	82 546(8.0)	6 393(20.6)	<0.001	24 802(6.6)	2 335(17.4)	<0.001
营养不良	1 141(0.1)	37(0.1)	0.693	552(0.1)	26(0.2)	0.212
其他营养缺乏	4 450(0.4)	103(0.3)	0.011	2 119(0.6)	61(0.5)	0.100
代谢紊乱	49 325(4.8)	2 961(9.5)	<0.001	22 918(6.1)	1 837(13.7)	<0.001
器质性(包括症状性)精神障碍	3 005(0.3)	224(0.7)	<0.001	366(0.1)	21(0.2)	0.048
心境(情感)障碍	7 582(0.7)	413(1.3)	<0.001	805(0.2)	62(0.5)	<0.001
神经症,与压力有关的和躯体形式的疾病	50 209(4.8)	2 647(8.5)	<0.001	11 453(3.1)	734(5.5)	<0.001

注:本研究中的变量均为分类变量,“年龄”变量为八分类变量,用于比较各年龄组之间的差异;“性别”变量为二分类变量,用于比较“男”与“女”的差异;其他变量为二分类变量,分别用于比较“有此疾病”与“无此疾病”的差异。

## 2.2 单因素分析结果

后得到 67 个与缺血性脑卒中发病显著相关的筛查

在训练集中,经单因素 Logistic 回归分析筛选

因子。见表 2。

表 2 缺血性脑卒中的单因素 Logistic 回归分析  
Table 2 Univariate Logistic regression analysis of ischemic stroke

变量	$Beta$	$SE$	$OR(95\%CI)$	$Wald$	$P$
年龄/岁					
[18, 25)			1		
[25, 35)	0.335	0.106	1.398(1.135~1.723)	3.149	0.002
[35, 45)	0.982	0.106	2.670(2.170~3.284)	9.292	<0.001
[45, 55)	2.109	0.102	8.243(6.751~10.064)	20.707	<0.001
[55, 65)	3.043	0.101	20.974(17.206~25.568)	30.117	<0.001
[65, 75)	3.546	0.101	34.690(28.461~42.282)	35.123	<0.001
[75, 85)	3.896	0.101	49.194(40.324~60.016)	38.404	<0.001
$\geq 85$	3.801	0.107	44.737(36.297~55.138)	35.634	<0.001
性别					
女			1		
男	0.474	0.014	1.606(1.564~1.651)	34.313	<0.001
其他脑血管病	1.158	0.020	3.185(3.064~3.310)	58.594	<0.001
主要影响中枢神经系统的全身萎缩	1.151	0.254	3.162(1.923~5.198)	4.538	<0.001
累及认知、知觉、情绪状态和行为的症状和体征	1.038	0.023	2.823(2.698~2.953)	45.014	<0.001
循环系统其他和未特指的疾患	0.809	0.154	2.246(1.660~3.041)	5.238	<0.001
发作性和阵发性疾患	0.790	0.021	2.204(2.116~2.296)	38.092	<0.001
锥体外系和运动疾患	0.764	0.073	2.146(1.859~2.478)	10.409	<0.001

续表

变量	Beta	SE	OR(95%CI)	Wald	P
神经系统的其他紊乱	0.690	0.056	1.994(1.788~2.223)	12.429	<0.001
肾小球疾病	0.688	0.067	1.989(1.743~2.270)	10.204	<0.001
累及神经和肌肉骨骼系统的症状和体征	0.618	0.123	1.855(1.457~2.362)	5.015	<0.001
神经症,与压力有关的和躯体形式的疾病	0.597	0.026	1.817(1.727~1.910)	23.245	<0.001
其他营养缺乏	0.570	0.123	1.769(1.389~2.253)	4.621	<0.001
高血压病	0.564	0.015	1.758(1.707~1.812)	37.179	<0.001
内耳疾病	0.555	0.061	1.741(1.546~1.962)	9.138	<0.001
心境(情感)障碍	0.520	0.062	1.682(1.488~1.900)	8.339	<0.001
糖尿病	0.518	0.018	1.679(1.622~1.739)	29.081	<0.001
视神经和视路疾患	0.502	0.128	1.652(1.285~2.124)	3.917	<0.001
多神经病和周围神经系统的其他疾患	0.499	0.094	1.646(1.370~1.978)	5.324	<0.001
外耳疾病	0.490	0.077	1.632(1.404~1.897)	6.393	<0.001
荨麻疹和红斑	0.442	0.077	1.555(1.338~1.808)	5.744	<0.001
皮肤附件疾病	0.418	0.074	1.519(1.313~1.758)	5.615	<0.001
非感染性小肠炎和结肠炎	0.402	0.042	1.494(1.376~1.622)	9.594	<0.001
炎性多关节病	0.392	0.034	1.481(1.384~1.584)	11.396	<0.001
肾功能衰竭	0.380	0.050	1.463(1.326~1.614)	7.569	<0.001
缺血性心脏病	0.359	0.016	1.432(1.387~1.479)	21.856	<0.001
器质性(包括症状性)精神障碍	0.359	0.088	1.432(1.205~1.701)	4.087	<0.001
累及皮肤和皮下组织的症状和体征	0.340	0.054	1.404(1.263~1.562)	6.259	<0.001
非化脓性中耳炎	0.308	0.070	1.360(1.185~1.561)	4.372	<0.001
以皮肤和黏膜损伤为特征的病毒感染	0.295	0.054	1.343(1.207~1.493)	5.445	<0.001
神经、神经根和神经丛疾患	0.288	0.059	1.334(1.188~1.498)	4.871	<0.001
泌尿系统的其他疾病	0.265	0.035	1.303(1.217~1.395)	7.614	<0.001
急性上呼吸道感染	0.260	0.019	1.297(1.249~1.348)	13.383	<0.001
皮肤炎和湿疹	0.240	0.039	1.272(1.178~1.373)	6.131	<0.001
口腔、唾液腺和颌部疾病	0.235	0.029	1.265(1.195~1.338)	8.147	<0.001
肠道传染病	0.226	0.055	1.254(1.126~1.396)	4.115	<0.001
代谢紊乱	0.222	0.024	1.249(1.190~1.310)	9.062	<0.001
累及循环和呼吸系统的症状和体征	0.211	0.023	1.234(1.179~1.292)	9.043	<0.001
脊椎病	0.209	0.028	1.232(1.165~1.302)	7.353	<0.001
一般症状和体征	0.189	0.017	1.208(1.168~1.249)	10.982	<0.001
其他上呼吸道疾病	0.188	0.035	1.207(1.127~1.293)	5.391	<0.001
其他急性下呼吸道感染	0.182	0.031	1.200(1.130~1.274)	5.937	<0.001
动脉、小动脉和毛细血管疾病	0.178	0.038	1.195(1.108~1.288)	4.638	<0.001
其他软组织疾病	0.170	0.043	1.185(1.090~1.288)	3.990	<0.001
关节病	0.168	0.035	1.183(1.104~1.268)	4.759	<0.001
静脉、淋巴管和淋巴结疾病,不可归类在他处者	0.153	0.035	1.165(1.088~1.248)	4.387	<0.001
甲状腺疾病	0.143	0.032	1.154(1.084~1.229)	4.459	<0.001
其他类型的心脏病	0.137	0.024	1.147(1.094~1.202)	5.700	<0.001
慢性下呼吸道疾病	0.096	0.021	1.101(1.056~1.148)	4.559	<0.001
食管、胃和十二指肠疾病	0.082	0.018	1.086(1.047~1.126)	4.463	<0.001
胆囊、胆道和胰腺疾病	-0.124	0.033	0.883(0.829~0.942)	-3.806	<0.001
晶状体疾病	-0.133	0.033	0.875(0.821~0.933)	-4.054	<0.001
肝脏疾病	-0.160	0.028	0.852(0.806~0.901)	-5.610	<0.001
肠的其他疾病	-0.184	0.034	0.832(0.778~0.889)	-5.413	<0.001
呼吸系统的其他疾病	-0.251	0.034	0.778(0.728~0.831)	-7.478	<0.001
肾和输尿管的其他疾患	-0.261	0.044	0.770(0.707~0.839)	-5.987	<0.001
肾小管-间质疾病	-0.347	0.083	0.707(0.601~0.832)	-4.183	<0.001
动态未定或动态未知的肿瘤	-0.349	0.061	0.705(0.626~0.795)	-5.736	<0.001
良性肿瘤	-0.350	0.049	0.705(0.640~0.775)	-7.190	<0.001

续表

变量	Beta	SE	OR(95%CI)	Wald	P
肺源性心脏病和肺循环疾病	-0.374	0.060	0.688(0.612~0.774)	-6.215	<0.001
涉及免疫机制的某些疾患	-0.390	0.082	0.677(0.577~0.795)	-4.764	<0.001
其他胸膜疾病	-0.408	0.048	0.665(0.605~0.731)	-8.472	<0.001
疝	-0.448	0.071	0.639(0.556~0.734)	-6.325	<0.001
淋巴、造血和相关组织的恶性肿瘤	-0.463	0.109	0.629(0.508~0.779)	-4.248	<0.001
不明确、继发和未特指部位的恶性肿瘤	-0.564	0.047	0.569(0.519~0.623)	-12.103	<0.001
消化器官恶性肿瘤	-0.595	0.052	0.552(0.498~0.611)	-11.352	<0.001
营养不良	-1.095	0.236	0.334(0.211~0.531)	-4.638	<0.001

注:本研究中的变量均为分类变量,“年龄”变量为八分类变量,以“[18, 25)”作为参照;“性别”变量为二分类变量,以“女”作为参照;其他变量为二分类变量,以“无此疾病”作为参照。

### 2.3 贝叶斯网络模型

最终构建的贝叶斯网络模型包括 68 个节点, 440 条有向边,其中缺血性脑卒中节点的父节点包括以下 7 个:年龄,高血压病,缺血性心脏病,慢性下呼吸道疾病,其他脑血管病,发作性和阵发性疾患,

累及认知、知觉、情绪状态和行为的症状和体征。由于网络中的节点数目较多,每一个节点的条件概率表篇幅过大,文中仅展示缺血性脑卒中节点条件概率表的一部分。见图 2、表 3。

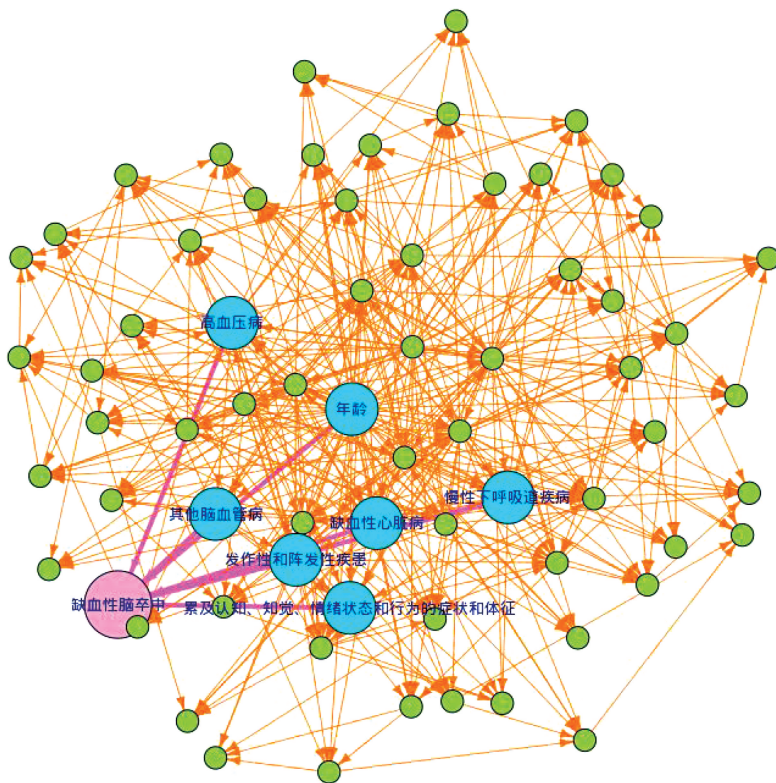


图 2 缺血性脑卒中筛查模型的贝叶斯网络结构

Figure 2 Bayesian network structure of the ischemic stroke screening model

表 3 缺血性脑卒中节点的条件概率表(部分)

Table 3 Conditional probability table for the ischemic stroke node (partial)

缺血性脑卒中	年龄	发作性和阵发性疾患	高血压病	其他脑血管病	慢性下呼吸道疾病	缺血性心脏病	累及认知、知觉、情绪状态和行为的症状和体征	概率
0	7	1	1	1	0	1	0	0.808
1	2	0	0	1	0	0	0	0.068
0	5	0	0	0	0	1	1	0.837
1	1	1	0	0	1	0	1	<0.001
0	4	0	1	1	1	0	1	0.962

续表

缺血性脑卒中	年龄	发作性和阵发性疾患	高血压病	其他脑血管病	慢性下呼吸道疾病	缺血性心脏病	累及认知、知觉、情绪状态和行为的症状和体征	概率
0	4	0	0	1	1	0	0	0.943
1	3	0	1	1	1	0	0	0.058
0	0	0	1	0	0	0	1	>0.999
0	3	0	1	1	0	0	0	0.900
0	5	1	0	0	0	1	1	0.876
0	1	1	1	1	0	0	0	>0.999
1	2	0	0	1	0	0	1	0.029
0	0	1	1	0	1	0	0	>0.999
1	5	0	0	1	1	1	0	0.135
0	3	0	1	0	0	0	0	0.968
1	0	1	1	0	1	1	1	0.500
0	4	0	1	0	1	0	0	0.943
0	1	0	1	1	0	1	1	0.667
0	4	0	0	1	0	1	1	0.978
0	4	1	0	1	0	0	0	0.892
1	2	0	0	1	1	0	1	0.001
0	7	0	0	0	0	0	0	0.920
0	1	1	1	1	1	1	1	0.727
1	2	0	0	0	1	1	1	0.100
0	2	0	1	0	1	0	0	0.967
0	7	0	1	1	1	1	1	0.583
1	3	0	1	1	0	0	1	0.075
0	2	0	1	1	0	0	0	0.922
0	4	1	1	0	0	0	1	0.914
1	4	1	0	0	1	0	0	0.054
0	7	0	1	1	1	0	0	0.800
1	4	1	0	0	0	1	1	0.089

## 2.4 模型评价

训练集、测试集和外部验证队列的 AUC 分别为 0.840 (95% CI: 0.838 ~ 0.843)、0.839 (95% CI: 0.836 ~ 0.843) 和 0.811 (95% CI: 0.808 ~ 0.814), 筛查模型的判别能力较好; 筛查模型的校准能力较好。在测试集与外部验证队列中变量随机缺失比例为 50% 时, 贝叶斯网络筛查模型的 AUC 均高于经过插补处理的传统 Logistic 回归筛查模型的 AUC, 并且

Delong 检验证实二者 AUC 值均有差异; 校准曲线的斜率和截距表明贝叶斯网络筛查模型的预测概率与实际发生概率一致性良好, 能够准确估计缺血性脑卒中的风险, 而 Logistic 回归筛查模型的预测概率高于实际发生概率, 高估缺血性脑卒中的风险, 证明贝叶斯网络模型在缺失数据背景下拥有巨大应用潜力。见图 3、4、表 4。

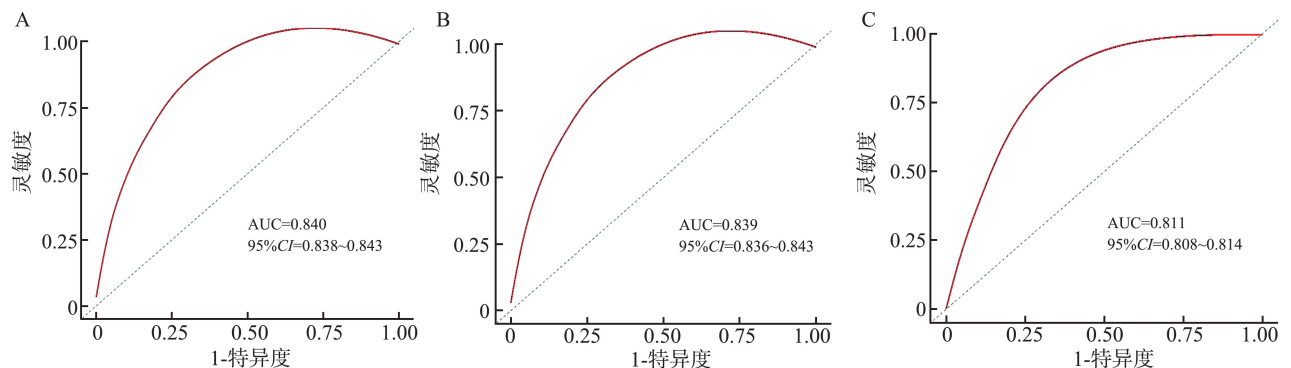


图 3 基于贝叶斯网络的缺血性脑卒中筛查模型的 ROC 曲线

A: 训练集; B: 测试集; C: 外部验证。

Figure 3 ROC curve of the Bayesian network-based ischemic stroke screening model

A: Training set; B: Testing set; C: External validation.

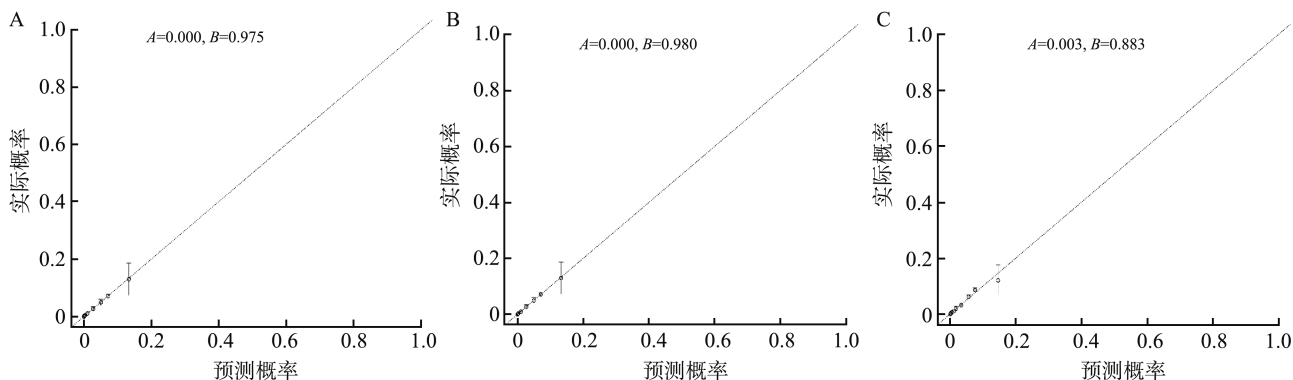


图4 基于贝叶斯网络的缺血性脑卒中筛查模型的校准曲线

A: 训练集; B: 测试集; C: 外部验证。

Figure 4 Calibration curve of the Bayesian network-based ischemic stroke screening model

A: Training set; B: Testing set; C: External validation.

表4 数据集缺失比例50%背景下贝叶斯网络筛查模型与 Logistic 回归筛查模型的模型对比

Table 4 The model comparison between the Bayesian network screening model and the Logistic regression screening model in the context of a 50% missing data rate

数据集	判别能力		校准能力	
	AUC (95% CI)	DeLong 检验 ( $P$ )	斜率 (95% CI)	截距 (95% CI)
缺失比例50%的测试集(基于贝叶斯网络的缺血性脑卒中筛查模型)	0.830(0.827~0.834)	<0.001	1.025(1.013~1.037)	0.000(-0.001~0.000)
缺失比例50%的测试集(基于 Logistic 回归的缺血性脑卒中筛查模型)	0.809(0.805~0.812)		0.327(0.244~0.410)	0.007(-0.002~0.017)
缺失比例50%的外部验证队列(基于贝叶斯网络的缺血性脑卒中筛查模型)	0.809(0.806~0.812)	<0.001	1.000(0.932~1.068)	0.001(-0.003~0.004)
缺失比例50%的外部验证队列(基于 Logistic 回归的缺血性脑卒中筛查模型)	0.797(0.794~0.800)		0.336(0.234~0.437)	0.011(-0.002~0.024)

### 3 讨论

在目前有效特异性治疗手段相对缺乏的情况下,早期预防是减少缺血性脑卒中疾病负担的最佳途径,国际上公认使用传统的缺血性脑卒中风险评估量表(如改良的弗明汉卒中量表、汇集队列方程等)以及一些利用 Logistic 回归等模型构建的短期筛查模型,来识别缺血性脑卒中发生的高危人群。然而,对大规模人群进行风险评估时,上述传统工具通常依赖于大量问卷量表来获取筛查因子,耗费大量的人力、物力资源,因而很难将其推广到公共卫生服务实践中。

考虑到电子健康记录具备样本量大、易于访问获取、个体医疗信息丰富等优点,本研究依托 Cheeloo LEAD 的优势,使用 BN 模型构建了缺血性脑卒中筛查模型,外部验证队列中模型的 AUC 为 0.811(95% CI:0.808~0.814),表明该筛查模型具备强大的风险预警功能,能够准确筛查出缺血性脑卒中的高危个体。在缺失数据的情况下,该筛查模型

的表现优于经过插补处理的传统 Logistic 回归筛查模型,证明其在应对缺失数据时依然具有较大的潜力。医生通过调取患者的电子健康记录能够迅速、高效地识别出高危患者,无需依赖传统的填写量表问卷等,并且模型具有很强的适应性,随着电子健康记录的实时更新,模型的条件概率表也能够动态调整,使得筛查过程自动化、灵活化、精准化。

贝叶斯网络能够灵活处理并可视化变量之间的复杂条件依赖关系,帮助临床医生更好地理解 and 解释各个因素对缺血性脑卒中风险的影响路径,使其在解释和应用模型时更具信心。医生对筛查出的高危人群进行合理的诊断,随后根据患者的具体情况实施个性化干预,重点干预与其疾病风险密切相关的因素。网络结构中缺血性脑卒中节点有如下 7 个父节点:年龄、高血压病、缺血性心脏病、慢性下呼吸道疾病、其他脑血管病,发作性和阵发性疾患,累及认知、知觉、情绪状态和行为的症状和体征。年龄是脑卒中最重要的且不可改变的危险因素<sup>[24]</sup>,在 45 岁后每 10 年发病率增加一倍,70% 以上的脑卒中发生于 65 岁以上人群<sup>[25]</sup>。高血压是脑卒中的主要可

改变危险因素,INTERSTROKE 研究揭示高血压病史显著增加脑卒中风险 ( $OR = 2.64, 99\% CI: 2.26 \sim 3.08$ )<sup>[26]</sup>。缺血性心脏病主要包括心肌梗死,是脑卒中发生的重要危险因素。在心肌梗死发作后的 30 d 内,心肌梗死组与匹配的普通人群相比,调整后的脑卒中率比(stroke rate ratio, SRR)为 31.9 ( $95\% CI: 28.4 \sim 35.8$ ),表明心肌梗死组发生缺血性脑卒中的风险是匹配的普通人群的 31.9 倍<sup>[27]</sup>。慢性下呼吸道疾病主要包括慢性阻塞性肺疾病和哮喘,患者的呼吸道问题会加重脑卒中的风险<sup>[28-29]</sup>。其他脑血管疾病主要包括无症状严重颈动脉狭窄和颅内动脉粥样硬化性狭窄,是缺血性脑卒中的重要危险因素,斑块阻塞血管并且容易破溃脱落引起缺血性脑卒中<sup>[30-31]</sup>。发作性和阵发性疾患包括短暂性脑缺血发作、偏头痛和睡眠障碍,也会增加脑卒中风险,受短暂性脑缺血发作影响的个体患缺血性脑卒中的短期风险很高,偏头痛通过多种复杂的机制影响缺血性脑卒中的发生,睡眠障碍与卒中发生存在正向因果关系 ( $OR = 1.264, 95\% CI: 1.122 \sim 1.423$ )<sup>[32-34]</sup>。累及认知、知觉、情绪状态和行为的症状和体征主要指的是缺血性脑卒中发生的前兆诊断,包括头晕、眩晕、意识障碍、精神错乱、肢体麻木等。上述节点均是缺血性脑卒中的危险因素,侧面证实了本研究所构建的筛查模型的合理性与准确性,并且这些危险因素也为临床医生提供了健康干预的重点。

综上所述,本研究基于 Cheeloo LEAD 构建了缺血性脑卒中贝叶斯网络筛查模型。该模型在内部和外部验证中表现良好,网络结构和条件概率表直观地展示了筛查因子之间的概率依赖关系,医生能够使用该模型实现早期筛查、早期诊断、早期干预等一系列流程,有效地降低缺血性脑卒中的发病率和死亡率。

## 参考文献:

[1] Li XM, Bian D, Yu JH, et al. Using machine learning models to improve stroke risk level classification methods of China national stroke screening[J]. BMC Med Inform Decis Mak, 2019, 19(1): 261. doi: 10.1186/s12911-019-0998-2.

[2] GBD Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019[J]. Lancet Neurol, 2021, 20(10): 795-820.

[3] 《中国脑卒中防治报告 2021》编写组.《中国脑卒中防治报告 2021》概要[J]. 中国脑血管病杂志, 2023, 20

(11): 783-793.

Report on Stroke Prevention and Treatment in China Writing Group. Brief report on stroke prevention and treatment in China, 2021 [J]. Chinese Journal of Cerebrovascular Diseases, 2023, 20(11): 783-793.

[4] Wolf PA, D'Agostino RB, Belanger AJ, et al. Probability of stroke: a risk profile from the Framingham Study[J]. Stroke, 1991, 22(3): 312-318.

[5] D'Agostino RB, Wolf PA, Belanger AJ, et al. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study[J]. Stroke, 1994, 25(1): 40-43.

[6] Andrus B, Lacaille D. 2013 ACC/AHA guideline on the assessment of cardiovascular risk[J]. J Am Coll Cardiol, 2014, 63(25): 2886. doi:10.1016/j.jacc.2014.02.606.

[7] Parmar P, Krishnamurthi R, Ikram MA, et al. The Stroke Riskometer(TM) App: validation of a data collection tool and stroke risk predictor[J]. Int J Stroke, 2015, 10(2): 231-244.

[8] Vartiainen E, Laatikainen T, Peltonen M, et al. Predicting coronary heart disease and stroke: the FINRISK calculator[J]. Glob Heart, 2016, 11(2): 213-216.

[9] 国家“十五”攻关“冠心病、脑卒中综合危险度评估及干预方案的研究”课题组. 国人缺血性心血管病发病危险的评估方法及简易评估工具的开发研究[J]. 中华心血管病杂志, 2003, 31(12): 893-901.

The Collaborative Research Group of the National 10th Five Year Plan Project: a Study on Evaluation and Intervention of the Coronary Heart Disease and Stroke Integrated Risk. A study on evaluation of the risk of ischemic cardiovascular diseases in Chinese and the development of simplified tools for the evaluation[J]. Chinese Journal of Cardiology, 2003, 31(12): 893-901.

[10] Yang XL, Li JX, Hu DS, et al. Predicting the 10-year risks of atherosclerotic cardiovascular disease in Chinese population: the china-PAR project (prediction for ASCVD risk in China) [J]. Circulation, 2016, 134(19): 1430-1440.

[11] Carroll RJ, Eyster AE, Denny JC. Intelligent use and clinical benefits of electronic health records in rheumatoid arthritis[J]. Expert Rev Clin Immunol, 2015, 11(3): 329-337.

[12] Liao WQ, Coupland CAC, Burchardt J, et al. Predicting the future risk of lung cancer: development, and internal and external validation of the CanPredict (lung) model in 19.67 million people and evaluation of model performance against seven other risk prediction models[J]. Lancet Respir Med, 2023, 11(8): 685-697.

[13] Han YT, Zhu X, Hu YZ, et al. Electronic health record-based absolute risk prediction model for esophageal cancer in the Chinese population: model development and ex-

- ternal validation[J]. *JMIR Public Health Surveill*, 2023, 9; e43725. doi:10.2196/43725.
- [14] Kyburg HE, Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference[J]. *J Philos*, 1991, 88(8): 434. doi:10.2307/2026705.
- [15] Yuan BY, Wang CB, Fan ZX, et al. A Bayesian network-based approach for identifying risk factors and predicting ischemic stroke in infective endocarditis patients [J]. *Front Cardiovasc Med*, 2023, 10: 1294229. doi:10.3389/fcvm.2023.1294229.
- [16] Fan ZX, Wang CB, Fang LB, et al. Risk factors and a Bayesian network model to predict ischemic stroke in patients with dilated cardiomyopathy[J]. *Front Neurosci*, 2022, 16: 1043922. doi:10.3389/fnins.2022.1043922.
- [17] Larrañaga P, Karshenas H, Bielza C, et al. A review on evolutionary algorithms in Bayesian network learning and inference tasks[J]. *Inf Sci*, 2013, 233: 109-125. doi:10.1016/j.ins.2012.12.051.
- [18] Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data[J]. *Mach Learn*, 1995, 20(3): 197-243.
- [19] Glover F. Artificial intelligence, heuristic frameworks and tabu search[J]. *Manag Decis Econ*, 1990, 11(5): 365-375.
- [20] Friedman N, Goldszmidt M, Wyner A. Data analysis with Bayesian networks: a bootstrap approach[EB/OL]. (2013-01-23) [2024-07-01]. 2013, arXiv: 1301.6695. <http://arxiv.org/abs/1301.6695>.
- [21] Peng Y, Zhang SY, Pan R. Bayesian network reasoning with uncertain evidences [J]. *Int J Unc Fuzz Knowl Based Syst*, 2010, 18(5): 539-564.
- [22] Fung R, Chang KC. Weighing and integrating evidence for stochastic simulation in Bayesian networks [J]. *Machine Intelligence and Pattern Recognition*, 1990, 10: 209-219. doi:10.1016/B978-0-444-88738-2.50023-3.
- [23] Shachter RD, Peot MA. Simulation approaches to general probabilistic inference on belief networks[J]. *Machine Intelligence and Pattern Recognition*, 1990, 10: 221-231. doi:http://dx.doi.org/.
- [24] 鞠柠蔚, 蒋红, 宋润琛, 等. 脑卒中患病影响因素的研究现状[J]. *中外医学研究*, 2024, 22(14): 165-171.
- JU Ningwei, JIANG Hong, SONG Runchen, et al. Research status of influencing factors of stroke [J]. 2024, 22(14): 165-171.
- [25] Kelly-Hayes M. Influence of age and health behaviors on stroke risk: lessons from longitudinal studies[J]. *J Am Geriatr Soc*, 2010, 58(Suppl 2): S325-S328.
- [26] O'Donnell MJ, Xavier D, Liu LS, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study[J]. *Lancet*, 2010, 376(9735): 112-123.
- [27] Sundbøll J, Horváth-Puhó E, Schmidt M, et al. Long-term risk of stroke in myocardial infarction survivors thirty-year population-based cohort study[J]. *Stroke*, 2016, 47(7): 1727-1733.
- [28] 李丹. 老年慢性阻塞性肺疾病合并卒中中的临床分析[J]. *中国卒中杂志*, 2023, 18(10): 1216.
- [29] Wen LY, Ni H, Li KS, et al. Asthma and risk of stroke: a systematic review and meta-analysis [J]. *J Stroke Cerebrovasc Dis*, 2016, 25(3): 497-503.
- [30] Chang RW, Tucker LY, Rothenberg KA, et al. Incidence of ischemic stroke in patients with asymptomatic severe carotid stenosis without surgical intervention[J]. *JAMA*, 2022, 327(20): 1974-1982.
- [31] Holmstedt CA, Turan TN, Chimowitz MI. Atherosclerotic intracranial arterial stenosis: risk factors, diagnosis, and treatment[J]. *Lancet Neurol*, 2013, 12(11): 1106-1114.
- [32] Johnston SC, Gress DR, Browner WS, et al. Short-term prognosis after emergency department diagnosis of TIA [J]. *JAMA*, 2000, 284(22): 2901-2906.
- [33] 潘飞豹, 周冀英, 谭戈. 偏头痛与脑卒中: 复杂的关联[J]. *重庆医学*, 2013, 42(22): 2665-2667.
- [34] 张亦舒, 李晓宁, 苗晨欣. 睡眠障碍与卒中风险的因果关系: 孟德尔随机化研究[J]. *现代医学*, 2023, 51(11): 1559-1565.
- ZHANG Yishu, LI Xiaoning, MIAO Chenxin. Sleep disorder and stroke risk: a Mendelian randomization study [J]. *Modern Medical Journal*, 2023, 51(11): 1559-1565.

(编辑: 相峰)