

OPEN ACCESS

编辑

朱波，

北京协和医院（CAMS），

中国

审查者：：

郭向阳，北京

大学，中国

左志毅，

美国弗吉尼亚大学

*通信：

洪昭

mazui_zhaohong@pkuph.edu.cn

专业部分：这篇文

章被提交给骨科手术、

外科前沿》杂志的一

个栏目

收到：2020年11月28日

背景：接受髌部骨折修复手术的老年患者由于年龄增长、合并症和身体虚弱，出现谵妄的风险增加。但目前识别住院病人中谵妄高风险的方法准确度一般，而且需要额外的问卷调查。人工智能使得建立基于电子健康数据的机器学习模型来预测意外谵妄风险成为可能。

方法：我们对2018年6月1日至2019年5月31日期间在脊柱或全身麻醉下接受髌部骨折骨科修复的老年患者（≥65岁）进行了一项回顾性病例对照研究。审查了麻醉记录和病历，收集人口统计学、手术、麻醉特征和虚弱指数，以探索术后谵妄的潜在风险因素。在住院期间，由受过训练的护士使用混乱评估法（CAM）每12小时评估一次谵妄。构建了四个机器学习风险模型来预测术后谵妄的发生率：随机森林、eXtreme Gradient Boosting（XGBoosting）、支持向量机（SVM）和多层感知（MLP）。采用K-fold交叉验证法来完成内部验证和性能评估。

结果：纳入约245名患者，术后谵妄影响了12.2%（30/245）的患者。多重逻辑回归显示，痴呆/中风史[OR 3.063, 95% CI（1.231, 7.624）]、输血[OR 2.631, 95% CI（1.055、

6.559）]，以及准备时间[OR 1.476, 95% CI（1.170, 1.862）]与术后谵妄有关，实现了接收器操作曲线（AUC）下面积为0.779, 95% CI（0.703, 0.856）。

机器学习模型预测术后谵妄发生的准确性为83.67%至87.75%。机器学习方法检测出16个导致谵妄发生的风险因素。准备时间、虚弱指数、手术期间使用血管压缩机、痴呆/中风史、手术时间和麻醉是导致谵妄的六个最重要的风险因素。



使用电子 图表数据 的机器学 习算法预 测老年髌 部骨折手 术后的谵 妄：一个 回顾性的 病例对照 研究

结论：电子病历衍生的机器学习模型可以生成医院特定的谵妄预测模型，并计算出风险因素对谵妄发生的贡献。需要进一步的研究来评估电子病历衍生的机器学习模型对接受髋部骨折修复手术的老年患者谵妄风险的检测意义和适用性。

关键词：谵妄，髋部骨折，老年人，机器学习，手术

简介

随着预期寿命的延长，预计在未来30年内，全世界每年将发生220万至450万例髋部骨折（1）。手术修复始终是必要的，以缓解疼痛和促进早期动员。围手术期的护理需要关注提供足够的镇痛和预防谵妄（2，3）。谵妄是一种波动的意识状态，在接受髋部骨折修复手术的患者中很常见，其发病率为36.5%（4）。它还与住院时间延长、术后1年死亡率增加、功能恢复不佳、费用高和入住养老院有关。（5）.预防策略通常是非药物性的，如保持病人的方向和环境保持互动，这些都是资源和人员密集型的（6）。如果可以准确预测谵妄风险，就可以针对高危患者进行预防，以改善患者的预后并节省资源。

人们开发了一些预测模型，以帮助识别髋部骨折后出现谵妄的高风险患者，但大多数模型往往依赖于医护人员进行的问卷调查（如迷你精神状态检查）、非例行的临床数据（护理人员主观疾病严重程度评估），或额外的计算（如急性生理和慢性健康评估得分），这些都很费时费力，增加了临床的难度。到目前为止，这些模型的预测准确性是中等的，达到了0.69-0.81的接受者操作曲线下的面积（AUCs）（7）。因此，寻找更方便、更适用的工具，以更高的准确度预测临床谵妄是很重要的。

机器学习是一种数据分析技术，它使用计算方法直接从数据中“学习”信息，而不依赖预先确定的方程式作为模型。适当的机器学习模型由于能够从多个模块中学习，因此能够实现局部的特定预测适应数据和数据噪音的稳健性，并能适应个别医院的小规模样本。此外，机器学习有可能分析出不同并发症的潜在机制（8）。黄博士建立了一个基于电子健康记录的机器学习模型来估计新住院病人的谵妄风险，实现了0.848至0.855的准确率（9）。

在这项单中心的回顾性病例对照研究中，我们开发并验证了四个机器学习模型，以预测老年髋部骨折患者的术后

准确率在83.67%至87.75%之间的图表自动预测谵妄的风险。

方法

我们对在脊柱或全身麻醉下接受低能量髋部骨折骨科修复的老年患者（≥65岁）进行了一项回顾性病例对照研究。我们审查了麻醉记录和病历，以收集人口统计学、手术、麻醉特征、虚弱指数以及术后谵妄的发生情况。我们分析了这些变量与谵妄发生率之间的关系，以揭示潜在的风险因素，并用多变量回归和机器学习模型计算预测的准确性。该研究得到了北京大学人民医院机构审查委员会的批准，由于是回顾性设计，所以放弃了知情同意。本文符合STROBE的病例对照研究指南。

研究协议

我们搜索了麻醉信息管理系统（AIMS）的北京大学人民医院老年患者（≥65岁）在2018年6月1日至2019年5月31日期间接受髋部骨折修复手术。诊断包括股骨颈骨折、股骨间骨折或股骨下骨折。从患者的医疗记录中提取了人口统计学、手术和麻醉学特征，并记录了术后谵妄的发生率和住院时间。

围手术期管理

早期医疗优化和早期手术是最初的目标的护理。在整个住院期间，同时使用老年疼痛协议和谵妄检测程序。所有的病人都是来自于骨科病房。

急诊科。泰乐士®

（口服羟考酮5毫克和

谵妄。这些模型使用了从医院获得的临床变量

对乙酰氨基酚325毫克)被用作术前镇痛剂; A

手术前在手术室进行单次髂筋膜室阻滞(0.3%罗哌卡因30毫升),然后使用舒芬太尼进行病人控制的镇痛,作为术后镇痛方案。患者离开手术室后,立即提供助听器和眼镜。但由于手脚短小,没有提供电脑游戏或认知康复等互动活动。

到达手术室后,我们为病人建立了标准的监测器,包括心电图、脉搏血氧仪和

动脉管。脊柱麻醉或全身麻醉是由主治麻醉师决定的。脊柱麻醉是通过L3-L4椎间盘完成的，单剂量的0.5%罗哌卡因2.2-3毫升。用依托咪酯、顺阿曲库铵和舒芬太尼诱导全身麻醉，并插入气管导管；手术期间双光谱指数（BIS）的目标是40至60。所有接受全身麻醉的患者都在手术室拔管。术中低血压（ $<90/60$ mmHg）用肾上腺素或苯肾上腺素治疗。启动红细胞输注的标准是血红蛋白浓度 <9 g-dl⁻¹。那些血压不稳定或有氧气依赖的病人在被转移到重症监护室（ICU）后，被送往医院。手术，进行严密监测。

变量的定义

外科手术包括半关节置换术、全髋关节置换术、全髋关节置换术。

关节置换术，或空心铆钉内固定治疗股骨颈骨折，股骨近端钉反转术（PFNA）治疗股骨间和股骨下骨折。记录了合并症，如痴呆症、中风史、高血压、糖尿病、冠心病、心房颤动和慢性肺部疾病。人口统计学、手术和麻醉数据都是从AIMS中自动提取的。我们使用了以前验证过的由病历得出的虚弱指数，包括年龄 >70 ，术前体重指数 <18.5 ，血红蛋白 $<35\%$ ，白蛋白 <3.4 g-dl⁻¹，肌酐 >176.8 umol-l⁻¹（10）。此后，虚弱状况的数字可以自动添加，以实现每个病人的虚弱指数。

成果的定义

主要的结果是，在治疗期间谵妄的发生率。

谵妄被定义为急性的、短暂的、波动的、通常可逆转的注意力、认知或注意力水平的干扰。谵妄是由训练有素的护士在住院期间每12小时使用混乱评估法（CAM）进行评估，直至出院。

候选预测器

共有33个变量被列入候选预测因素、

包括人口统计学特征、并发疾病、手术数据、麻醉数据、血液制品的消耗和准备时间。由于痴呆和中风史是已知的谵妄风险因素（6），而这两种情况在研究中很少见，所以这两个变量被合并起来，以代表基线时脑功能受损。准备时间被定义为髌部骨折诊断和手术之间的日历天数。大多数患者缺失的变量，如血气分析，没有被输入逻辑回归或机器学习模型中。

队列建设和内部验证

数据集被分为衍生队列（训练集）。

根据计算机生成的随机数字列表，以4:1的比例对验证队列（测试集）进行验证。采用K-fold交叉验证法来完成内部验证。

验证和性能评估。K-折交叉验证方法将训练数据集分为k个折，使用(k-1)个折学习分类器，并通过在剩余折中测试分类器来计算误差值。误差的k折交叉验证估计是每个折中的误差的平均值。所提模型的超参数通过k-fold验证方法进行优化，可以有效避免在小数据集情况下的过拟合和欠拟合(11)。

预测模型

构建了四个风险模型来预测发病率术后谵妄的分类方法：随机森林、eXtreme Gradient Boosting (XGBoost)、支持向量机(SVM)和多层感知器(MLP)。随机森林是一种集合学习方法，通过构建众多的决策树进行分类操作。XGBoost是一种梯度提升决策树的实现。SVM是有监督的学习模型，有相关的学习算法，分析用于分类和回归分析的数据。MLP是一种人工神经网络，从一组输入中生成一组输出。机器学习方法也被用来分析每个变量对谵妄发生的贡献、和相关系数，一个介于-1和+1之间的数字，被计算出来，代表变量与其他变量之间的关联。

的事件。相关系数是通过层间相关性传播(LRP)方法计算的，该方法将相关性从输出层递归地传播到输入层(12, 13)。具有相关系数的风险因素>0.5被认为是发生谵妄的重要预测因素。

所有的机器学习模型都是在Keras框架下用Python构建的。模型报告符合《生物医学研究中机器学习预测模型的开发和报告指南》(14)。

统计分析

连续变量以平均值±SD或中位数表示，其中四分位数范围(IQR)。分类变量表示为百分比。所有的P值都是双尾的， $P<0.05$ 被认为是有统计学意义的。统计分析使用SPSS 22.0统计软件包(SPSS Inc., Chicago, IL, U.S.A)进行。对连续变量采用独立t检验或Mann-Whitney U检验，对分类变量采用chi-square检验或Fisher's exact检验来评估组间差异，如适用。采用多变量逻辑回归法来确定术后谵妄的风险因素。

结果

在本研究中，在AIMS中使用“65岁及以上”“接受”“腕部骨折修复”的连续患者有245人，其中最年长为99岁(图1A)

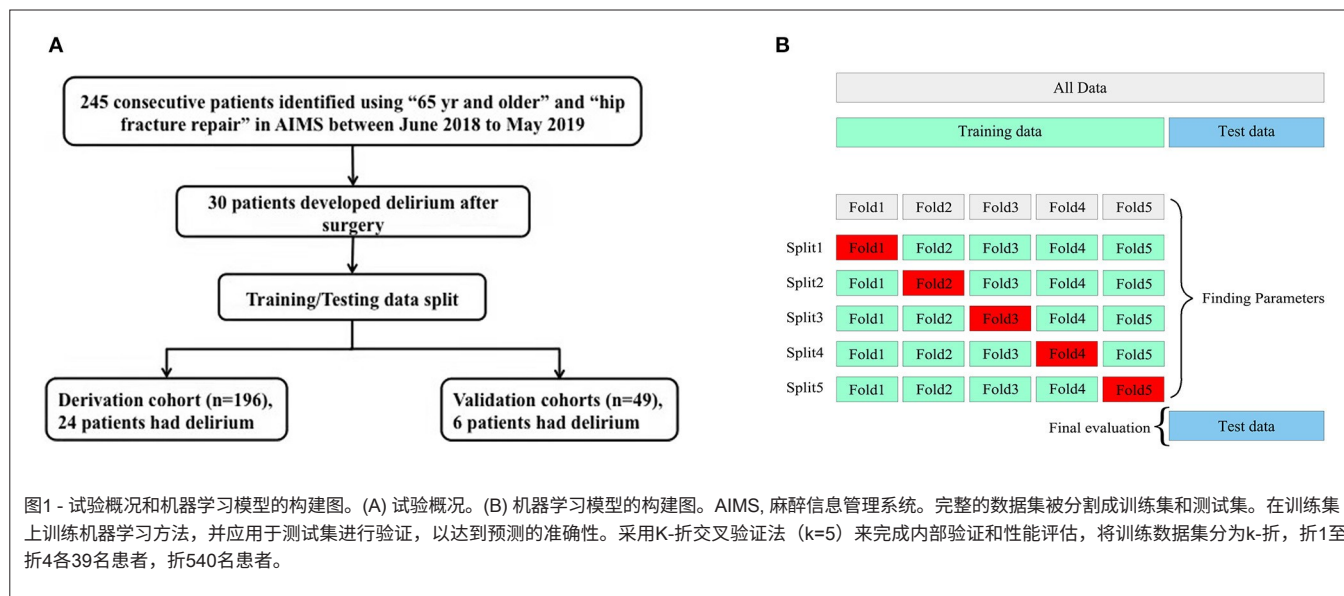


图1 - 试验概况和机器学习模型的构建图。(A) 试验概况。(B) 机器学习模型的构建图。AIMS, 麻醉信息管理系统。完整的数据集被分割成训练集和测试集。在训练集上训练机器学习方法, 并应用于测试集进行验证, 以达到预测的准确性。采用K-折交叉验证法 (k=5) 来完成内部验证和性能评估, 将训练数据集分为k-折, 折1至折4各39名患者, 折540名患者。

表1)。患者在诊断为腕部骨折后被送入骨科病房。

在245名患者中, 住院期间术后谵妄的发生率为12.2% (30/245)。所有的谵妄病例都是在术后48小时内观察到的。术后出现谵妄的患者, 其术后住院时间明显长于无谵妄的患者[13 (6) 天vs.8 (7) 天, $P = 0.013$].

肺炎影响了37名 (15.1%) 患者, 两名 (0.8%) 患者出现深静脉血栓, 28名 (11.4%) 患者术后被送入ICU。没有患者在住院期间死亡 (表1)。

谵妄发展的多变量Logistic回归

12个变量之间存在显著差异。

谵妄组和非谵妄组, 分别是痴呆/中风史、麻醉类型、麻醉时间、手术时间、输血 (围手术期输红细胞/新鲜冷冻血浆量和输血病人数)、术中输液量和手术前准备时间 ($P < 0.1$)。有六个因素进入多变量逻辑回归, 包括痴呆/中风史、麻醉类型、麻醉时间、术中输液量、输血和术前准备时间。多重逻辑回归显示, 痴呆/中风史[OR 3.063, 95% CI (1.231, 7.624)], 输血[OR 2.631, 95% CI (1.055, 6.559)], 和准备时间[OR 1.476, 95% CI (1.170, 1.862)]与术后谵妄有关 (表2, 3), 达到AUC下面积0.779, 95% CI (0.703, 0.856) (图2)。

机器学习模型训练和验证

构建了四个机器学习模型, 具有类似

预测术后谵妄的准确性, 范围从

如混淆矩阵中所示 (表3), 83.67%至87.75%。K-折 ($k = 5$) 交叉验证被部署来完成内部验证和性能评估 (图1B)。在这项研究中、

k-fold交叉验证法将训练数据集分为五折, 使用四折学习分类器, 并通过在剩余的折中测试分类器来计算误差值。5折交叉验证法对误差的估计是在每个折中所犯错误的平均值。所提模型的超参数通过5折验证方法进行了优化, 可以有效避免在小数据集情况下的过拟合和欠拟合。真阳性率和假阳性率也被描绘出来, 位于逻辑回归的ROC曲线左侧和上方, 表明机器学习模型的性能优于逻辑回归 (图2)。

用LRP方法计算了代表每个变量对术后谵妄发生率贡献的相关系数。准备时间、虚弱指数、手术中使用血管抑制剂、痴呆/中风史、手术时间和麻醉是谵妄的六个最重要预测因素 (相关系数 >0.5), 其相关系数见图3。

讨论

在这项回顾性病例对照研究中, 利用电子病历数据生成了四种预测谵妄的机器学习模型, 预测谵妄的准确率为83.67%至87.75%。使用机器学习方法还检测了16个风险因素对谵妄发展的贡献。准备时间、虚弱指数、使用

表1 |围手术期数据的单变量和多变量分析。

	谵妄组 (n = 30)	无谵妄组(n = 215)	统计学价值	P值	或	95%CI	
年龄(岁)	79.3 (7.6)	80.0 (7.1)	-0.613	0.540			
男性[n (%)]	9 (30%)	56 (26.0%)	0.211	0.646			
高度 (厘米)	159.8 (7.9)	160.0 (8.1)	-0.159	0.874			
重量 (公斤)	57.7 (11.1)	57.9 (11.3)	0.032	0.975			
BMI	22.5 (3.8)	22.6 (4.3)	0.10	0.919			
辽宁省			3.202	0.362			
1	0 (0%)	6 (2.8%)					
2	21 (70%)	129 (60%)					
3	8 (26.7%)	78 (36.3%)					
4	1 (3.3%)	2 (0.9%)					
高血压	19 (63.3%)	131 (60.9%)	0.064	0.844			
糖尿病	7 (23.3%)	53 (24.7%)	0.025	1.0			
冠状动脉疾病	5 (16.7%)	62 (28.8%)	1.963	0.194			
痴呆症: 中风史	12 (40%)	40 (18.6%)	7.208	0.015*	3.063	1.231	7.624
虚弱指数	2.1 (1.0)	1.9 (1.0)	-0.810	0.418			
血细胞比容	0.32 (0.06)	0.33 (0.06)	-1.235	0.226			
白蛋白 (g·dl) ⁻¹	34.8 (4.1)	35.9 (3.6)	1.455	0.147			
肌酐 (umol·l) ⁻¹	84 (39)	63 (31)	-0.794	0.427			
诊断			5.478	0.242			
股骨颈骨折	6 (20%)	75 (34.9%)					
肱骨间骨折	19 (63.3%)	116 (53.9%)					
肱骨下骨折	5 (16.7%)	24 (11.2%)					
手术过程			2.646	0.266			
关节置换术/半关节置换术	4 (13.3%)	48 (22.3%)					
空心铆钉	2 (6.7%)	27 (12.6%)					
PFNA	24 (80%)	140 (65.1%)					
全身麻醉[n (%)]	8 (26.7%)	25 (11.6%)	5.109	0.04*	2.788	0.966	8.047
血气分析							
pH值	7.44 (0.04)	7.45 (0.03)	0.961	0.338			
氧饱和度	70 (33.2)	77.4 (14.3)	-0.176	0.861			
手术时间 (分钟)	130 (100)	75 (40)	-2.390	0.017*			
麻醉时间 (分钟)	228 (93)	163 (45)	-2.413	0.016*	1.009	0.998	1.019
注入量 (100毫升)	12 (5.63)	9 (5)	-1.924	0.054	0.988	0.914	1.067
失血量 (毫升)	200 (250)	100 (135)	-1.076	0.282			
术中红细胞输注 (毫升)	0 (150)	0 (0)	-2.546	0.011*			
术中新鲜冷冻 血浆输注 (毫升)	0 (400)	0 (0)	-3.576	<0.001*			
患者在术中接受 血制品[n (%)]	10 (33.3%)	31 (14.4%)	6.759	0.017*	2.631	1.055	6.559
术后红细胞输注 (毫升)	200 (400)	0 (1000)	-3.813	<0.001*			
术后新鲜冷冻 血浆输注 (毫升)	0 (400)	0 (0)	-4.012	<0.001*			
患者在术后接受 血制品[n (%)]	16 (53.3%)	42 (19.5%)	16.6	<0.001*			
准备时间 (天)	3 (2)	2 (3)	-2.822	0.04*	1.476	1.170	1.862
术后停留时间 (日)	13 (6)	8 (7)	-2.477	0.013*			
血管抑制剂的剂量	0 (1)	1 (2)	0.354	0.722			
入住ICU[n (%)]	5 (16.6%)	19 (8.8%)	0.177	0.189			
肺炎[n (%)]	5 (15.1%)	32 (15.0%)	0.0001	0.993			
深静脉血栓/PE[n (%)]	2 (6.0%)	0 (0%)	14.451	0.017*			

数值为平均值 (SD) 或数量 (比例) 或中位数 (IQR)。PFNA, 股骨近端钉子反转; 虚弱指数, 以下虚弱情况的总和, 年龄>70岁, 术前体重指数<18.5, 血细胞比容<35%, 白蛋白<3.4 g·dl⁻¹, 肌酐>176.8 umol·l⁻¹。每种虚弱状况将得到1分。DVT, 深静脉血栓; PE, 肺栓塞。准备时间定义为髋部骨折诊断和手术之间的天数。统计学价值为归一化连续变量的t值, 非参数变量的M-W, 分类变量的检验为χ²。* P < 0.05。

表2 | 术后谵妄的多变量分析 准备时间被定义为髌部骨折诊断和手术之间的日历天数。

	β	标准误差	谷仓	P值	或	95%CI	
痴呆症: 中风史	1.120	0.465	5.792	0.016*	3.063	1.231	7.624
全身麻醉	1.025	0.541	3.595	0.058	2.788	0.966	8.047
麻醉时间 (分钟)	0.009	0.005	2.658	0.103	1.009	0.998	1.019
术中输液 (毫升)	0.000	0.000	0.176	0.675	1.000	0.999	1.001
接受输血的患者	0.967	0.466	4.307	0.038*	2.631	1.055	6.559
准备时间 (天)	0.389	0.119	10.576	0.001*	1.476	1.170	1.862

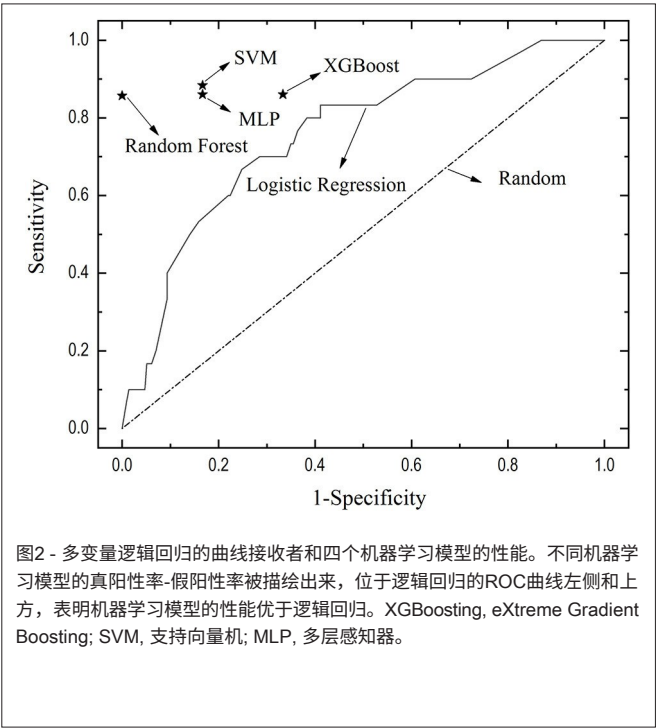
*P < 0.05.

表3 | 机器学习模型的混淆矩阵。

模型	准确度	假阴性	真正的正面	假阳性	真正的负面
随机森林	85.71%	16.28%	83.72%	0.00%	100.00%
XGBoost	83.67%	13.95%	86.05%	33.34%	66.66%
证券公司	87.75%	11.63%	88.37%	16.67%	83.33%
MLP	85.71%	13.95%	86.05%	16.67%	83.33%

准确度, (真真+真假) /ALL。

XGBoosting, eXtreme Gradient Boosting; SVM, 支持向量机; MLP, 多层感知。



手术中使用血管压缩机、痴呆/中风史、手术时间和麻醉是老年髌部骨折手术后谵妄的六个最重要预测因素。

医院可能有个性化和局部的特征, 可能无法实施临床指南或试验的重大发现。机器学习模型可以为整合风险估计方面的个性化和局部特征提供机会, 这反过来又为

指导临床实践。机器学习模型的高准确率可能是由于它们比传统的回归技术使用了更多的数学运算来更好地定义风险因素和结果之间的复杂关系 (9)。这四种机器学习模型在对小规模样本进行准确预测方面表现出色, 这使得它在评估医院特定的住院和死亡风险方面更加实用 (8)。

本研究创建的机器学习模型通过病历参数的 "快照" 提供了谵妄发生的预测概率, 其准确、客观且易于获取, 因此很容易被纳入医院的工作流程 (15)。那么, 资源密集型的谵妄预防措施应适用于髌部骨折的高风险老年患者 (3)。

在本研究中, 多变量逻辑回归在术后谵妄预测中得到的AUC为0.779, 这与以前的研究相似 (16)。现有的预测模型, 如Vochteloo模型, 使用9个项目的模型来实现加分公式, 以5分作为分界点 (17)。擅离职守 (年龄 >79岁, 不能向后拼写 世界, 迷失方向的地方, 以及较高的护士评分的疾病严重程度) 模型是另一个预测新入院病人谵妄的有力工具 (18)。这两个模型都需要额外的问卷调查, 对谵妄的预测准确率适中, 从65%到75%不等 (16)。

多变量逻辑回归发现, 痴呆/中风史、食用血液制品和准备时间较长是老年髌部骨折患者术后谵妄的三个风险因素。对于传统的多变量逻辑回归, 建议每个输入变量至少有5-

10个结果事件 (19)。痴呆/中风史可被视为住院患者谵妄的易感风险因素。

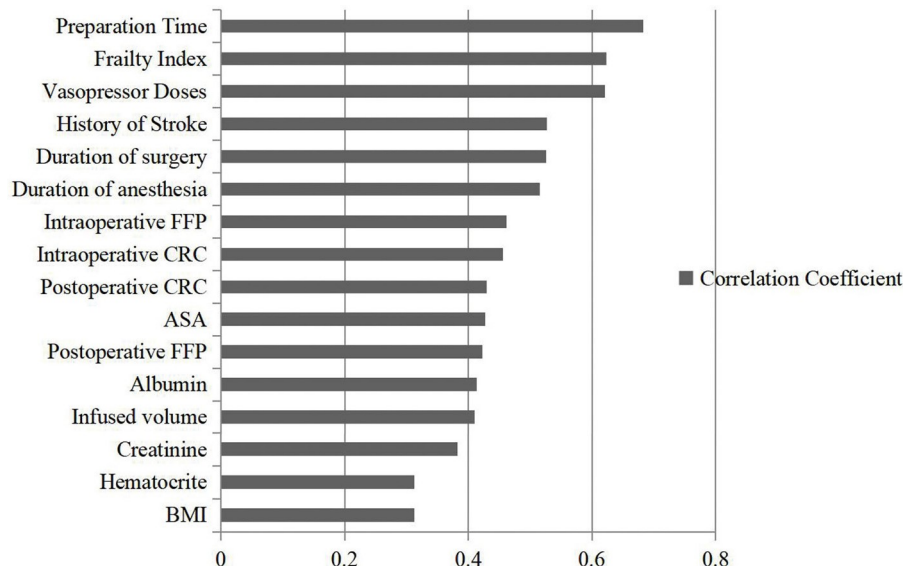


图3 | 通过机器学习方法检测的不同变量的相关系数。相关系数是一个介于-1和+1之间的数字，计算后表示变量和事件的线性依赖关系。系数最大的预测变量被认为是最重要的预测变量；而相关系数次大的预测变量是次重要的变量，以此类推。

患者(6)。然而，另外两个因素，即血液制品的消耗和较长的准备时间是诱发因素，这意味着它们可以被调整和改善。关于输血的研究得出了有争议的结果，因为贫血和同源输血对术后谵妄都有不利的影响。贫血与脑灌注不足有关，而同源性输血是谵妄患者的保护因素。

测量的最低血红蛋白水平小于或等于 $9.7\text{g}\cdot\text{dl}^{-1}$ (20)。而根据Leuzinger等人的研究，老年患者与未输血的患者相比，接受红细胞输注的髋部骨折患者的住院时间更长，感染率更高 (21)。因此，在这种情况下，避免贫血和建立全院输血标准是至关重要的。根据一些国际共识，髋部骨折修复手术应该每周7天都可以进行 (22)，但在这个中心，周末出现在急诊科的病人有时直到周一才接受手术。

相反，通过机器学习方法计算了16个变量与谵妄发生的相关系数，显示了机器学习的可行性，可以更好地定义多个风险因素和结果之间的复杂关系。除了通过多元逻辑回归确定的三个风险因素外，虚弱指数和术中的血管抑制剂剂量也与谵妄的发生有关。虚弱是接受外科手术的老年患者发病率和死亡率的一个重要指标 (23, 24)。术中的血管收缩剂剂量可能被视为低血压发作的反映，这与谵妄的出现有关 (25)。

机器学习模型除了准确率高之外，还有另一个优点，使用电子病历导出的参数和确认不同风险因素的相关系数，也就是说，机器学习模型的准确率可以随着时间的推移而提高，眼科医生和视网膜专家之间糖尿病视网膜病变分级的可比性提高就是一个例子 (26, 27)。随着人工智能的广泛使用和对谵妄的更多了解，更复杂和准确的预测正在构建中，更多的特征可以加入到模型中，如虚弱或神经炎症的生物标志物来研究谵妄的潜在机制。

这项研究有一些局限性。首先，所有的变量都是从电子病历和AIMS中获得的，一些数据可能会被遗漏，特别是微生物学、放射学或与谵妄的潜在机制有关的生物标志物。我们鼓励更多的研究来调查谵妄发生的基本机制。第二，一些变量，如入院前的行走能力或财务状况没有记录在病历中。但从病历中得出的虚弱指数可以作为患者功能能力的一个指标。

总之，电子病历衍生的机器学习模型（随机森林、SVM、XGBoost和MLP）可以生成医院特定的谵妄预测模型，并计算出风险因素对谵妄发生的贡献。需要进一步的研究来评估电子病历衍生的机器学习模型对于检测老年髋部骨折修复手术患者谵妄风险的意义和适用性。

数据可用性声明

支持本文结论的原始数据将由作者提供，没有不适当的保留。

道德声明

涉及人类参与者的研究由北京大学人民医院机构审查委员会审查和批准。根据国家法律和机构的要求，本研究不需要书面知情同意书参与。

作者的贡献

HZ帮助设计研究，进行研究，并准备稿件。JY和YP帮助进行了机器模型的构建。YF帮助设计了研究。所有作者都参与了文章的撰写，并批准了提交的版本。

参考文献

- Veronese N, Maggi S. 髌部骨折的流行病学和社会成本。 *Injury*. (2018) 49:1458-60. doi: 10.1016/j.injury.2018.04.015
- Berian JR, Zhou L, Russell MM, Hornor MA, Cohen ME, Finlayson E, et al. Postoperative delirium as a target for surgical quality improvement. *Ann Surg*. (2018) 268:93-9. doi: 10.1097/SLA.0000000000002436
- Hughes CG, Boncyk CS, Culley DJ, Fleisher LA, Leung JM, McDonagh DL, et al. American Society for enhanced recovery and perioperative quality initiative joint consensus statement on postoperative delirium prevention. *Anesth Analg*. (2020) 130:1572-90. doi: 10.1213/ANE.0000000000004641
- Siebert FE, Neufeld KJ, Gottschalk A, Bigelow GE, Oh ES, Rosenberg PB, et al. 老年患者接受髌部骨折修复时镇静深度对术后谵妄的影响：STRIDE 随机临床试验。 *JAMA Surg*. (2018) 153:987-95. doi: 10.1001/jamasurg.2018.2602
- Inouye SK, Westendorp RG, Saczynski JS. 老年人的谵妄。 *柳叶刀*。 (2014) 383:911-22. doi: 10.1016/S0140-6736 (13) 60688-1
- Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borzodina A, et al. 欧洲麻醉学会基于证据和共识的术后谵妄指南。 *Eur J Anaesthesiol*. (2017) 34:192-214. DOI: 10.1097/EJA.0000000000000594
- Zhang X, Tong DK, Ji F, Duan XZ, Liu PZ, Qin S, et al. Predictive nomogram for postoperative delirium in elderly patients with a hip fracture. *Injury*. (2019) 50:392-7. doi: 10.1016/j.injury.2018.10.034
- Mathis MR, Kheterpal S, Najarian K. Artificial intelligence for anesthesia: what the practicing clinician needs to know: more than black magic for the art of the dark. *Anesthesiology*. (2018) 129:619-22. DOI: 10.1097/ALN.0000000000002384
- Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. 基于电子健康记录的机器学习模型的开发和验证，以估计没有已知认知障碍的新住院病人的谵妄风险。 *JAMA Netw Open*. (2018) 1:e181018. doi: 10.1001/jamanetworkopen.2018.1018
- Amrock LG, Neuman MD, Lin HM, Deiner S. 常规术前数据可以预测老年人的不良后果吗？开发并验证了一个简单的风险模型，其中包括从图表中得出的虚弱评分。 *J Am Coll Surg*. (2014) 219:684-94. doi:

资金来源

本研究得到了北京大学人民医院科研发展基金项目（RGY 2019-17）的支持，北京，中国。

鸣谢

我们感谢来自中国北京大学健康科学中心应用语言学系的张健女士，她提供了医学写作的编辑。我们感谢来自中国北京大学临床研究中心的严晓燕博士提供的统计分析帮助。

补充材料

这篇文章的补充材料可以在网上找到：
<https://www.frontiersin.org/articles/10.3389/fsurg.2021.634629/full#supplementary-material>

10.1016/j.jamcollsurg.2014.04.018

- Rodriguez JD, Perez A, Lozano JA. 预测误差估计中K-折交叉验证的敏感性分析。 *IEEE Trans Pattern Anal Mach Intell*. (2010) 32:569-75. doi: 10.1109/TPAMI.2009.187
- Bach S, Binder A, Montavon G, Klauschen F, Muller KR, Samek W. On Pixel-Wise Explanations for nonlinear classifier

- 通过层级相关性传播的方式进行决策。 *PLoS ONE*.(2015) 10:e0130140. doi: 10.1371/journal.pone.0130140
13. Ding YZ, Liu Y, Luan HB, Sun MS.可视化和理解神经机器翻译。In: *计算语言学协会第55届年会论文集 (Acl 2017)*。Vol. 1.Vancouver, BC: Association for Computational Linguistics (2017). p. 1150-9 (2017). doi: 10.18653/v1/P17-1106
 14. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. 生物医学研究中开发和报告机器学习预测模型的指南: 多学科观点。 *J Med Internet Res*. (2016) 18:e323. doi: 10.2196/jmir.5870
 15. McIsaac DI, MacDonald DB, Aucoin SD.围手术期临床医生的虚弱: 一个叙述性的回顾。 *Anesth Analg*. (2020) 130:1450-60. doi: 10.1213/ANE.0000000000004602
 16. Whitlock EL, Braehler MR, Kaplan JA, Finlayson E, Rogers SE, Douglas V, et al. 基于电子病历的围手术期谵妄风险分层工具的衍生、验证、持续性能和临床影响。 *Anesth Analg*. (2020) 131:1901-10. doi: 10.1213/ANE.0000000000005085
 17. Vochteloo AJ, Moerman S, van der Burg BL, de Boo M, de Vries MR, Niesten DD, et al. 髌部骨折患者的谵妄风险筛查和氟哌啶醇预防计划是识别高风险患者的有用工具, 但并不能降低谵妄的发生率。 *BMC Geriatr*. (2011) 11:39. doi: 10.1186/1471-2318-11-39
 18. Brown EG, Josephson SA, Anderson N, Reid M, Lee M, Douglas VC.预测住院病人的谵妄: 临床实践中的AWOL谵妄风险分层评分。 *Geriatr Nurs*.(2017) 38:567-72. doi: 10.1016/j.gerinurse.2017.04.006
 19. Vittinghoff E, McCulloch CE.放宽Logistic和Cox回归中每个变量10个事件的规则。 *Am J Epidemiol*.(2007) 165:710-8. doi: 10.1093/aje/kwk052
 20. van der Zanden V, Beishuizen SJ, Scholtens RM, de Jonghe A, de Rooij SE, van Munster BC.输血对谵妄发病率的影响。 *J Am Med Dir Assoc*. (2016) 17:748-53. doi: 10.1016/j.jamda.2016.04.008
 21. Leuzinger E, Poblete B, Konrad CJ, Hansen D. 目前老年髌部骨折患者的输血做法与现行指南仍有差异, 以及对结果的影响: 一项回顾性观察研究。 *Eur J Anaesthesiol*. (2018) 35:972-9. doi: 10.1097/EJA.0000000000000883
 22. White SM, Altermatt F, Barry J, Ben-David B, Coburn M, Coluzzi F, et al. International Fragility Fracture Network Delphi consensus statement on anaesthesia principles for patients with hip fracture. *Anaesthesia*.(2018) 73:863-74. doi: 10.1111/anae.14225

23. Robinson TN, Walston JD, Brummel NE, Deiner S, Brown CH, Kennedy M, et al. 外科医生的虚弱：国家老龄化研究所专家虚弱会议的回顾。 *J Am Coll Surg.* (2015) 221:1083-92. doi: 10.1016/j.jamcollsurg.2015.08.428
24. McIsaac DI, Bryson GL, van Walraven C. 虚弱与重大选择性非心脏手术后1年死亡率的关系：一项基于人口的队列研究。 *JAMA Surg.* (2016) 151:538-45. doi: 10.1001/jamasurg.2015.5085
25. White SM MI, Griffiths R, Johansen A, Wakeman R, Boulton C, Plant F. 对英国前瞻性麻醉冲刺实践审计（ASAP-2）中11,085例腕部骨折手术后的结果进行二次分析。 *麻醉。* (2016) 71:506-11. doi: 10.1111/anae.13415
26. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. 开发和验证用于检测视网膜眼底照片中糖尿病视网膜病变的深度学习算法。 *JAMA.* (2016) 316:2402-10. doi: 10.1001/jama.2016.17216
27. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader variability and importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology.* (2018) 125:1264-72. doi: 10.1016/j.phtha.2018.01.034

利益冲突：作者声明，该研究是在没有任何商业或财务关系的情况下进行的，可被视为潜在的利益冲突。

Copyright © 2021 Zhao, You, Peng and Feng. 这是一篇根据知识共享署名许可 (CC BY) 条款发布的开放性文章。允许在其他论坛上使用、传播或复制，但必须注明原作者和版权所有，并按照公认的学术惯例引用本杂志上的原始出版物。任何不符合这些条款的使用、传播或复制都是不允许的。