

基于机器学习的产妇阴道分娩后创伤后应激障碍预测模型的构建与评价

李夏楠¹,刘玲²,刘冬梅²,王美²

(1.安徽省临泉县人民医院 妇产科,安徽 临泉 236400;2.安徽省临泉县人民医院 护理部)

【摘要】 目的 基于机器学习分析产妇自然阴道分娩后创伤后应激障碍(posttraumatic stress disorder, PTSD)的危险因素并构建预测模型,探讨预测模型的预测价值。方法 采用便利抽样法选取2023年12月至2024年8月安徽省某院的产妇400例为研究对象,根据其是否发生 PTSD 分为未发生 PTSD 组($n=346$)和发生 PTSD 组($n=54$)。基于随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、K-最近邻(K-nearest neighbor, KNN)、逻辑回归(Logistic regression, LR) 4个机器学习模型构建产妇分娩后 PTSD 预测模型。采用准确度、精确度、接受者操作特性曲线(receiver operating characteristic curve, ROC)曲线下面积(area under the curve, AUC)等评价模型的稳健性及临床实用性。另纳入同期安徽省另一所医院的200例产妇进行外部验证。结果 产妇产后 PTSD 的发生率为13.50%。年龄、喂养方式、计划怀孕、母婴皮肤接触、急产、妊娠糖尿病等均是产后是否发生 PTSD 的独立影响因素(均 $P<0.05$)。RF模型对产妇 PTSD 的发生率的预测效果最好($AUC=0.882$)、综合预测效能最佳。在外部验证中,RF、SVM、KNN和LR模型的AUC值分别为0.842、0.734、0.725、0.767。结论 基于RF算法构建的预测模型具有较优质的预测性能,可以更加简单、快速、有效地识别产后 PTSD。

【关键词】 产后;创伤后应激障碍;机器学习;预测模型

doi:10.3969/j.issn.2097-1826.2025.10.008

【中图分类号】 R473.71 【文献标识码】 A 【文章编号】 2097-1826(2025)10-0031-05

A Prediction Model for Post-traumatic Stress Disorder After Vaginal Delivery in Parturients Based on Machine Learning: Construction and Evaluation

LI Xianan¹, LIU Ling², LIU Dongmei², WANG Mei² (1. Obstetrics and Gynecology Department, Linquan County People's Hospital, Linquan 236400, Anhui Province, China; 2. Nursing Department, Linquan County People's Hospital)
Corresponding author: LIU Dongmei, Tel: 0558-6403676

【Abstract】 Objective To analyze the risk factors of posttraumatic stress disorder (PTSD) after natural vaginal delivery in parturients and develop a predictive model based on machine learning, and to explore its predictive value. **Methods** A total of 400 parturients from a hospital in Anhui Province from December 2023 to August 2024 were selected by the convenience sampling method, and were divided into the group without PTSD ($n=346$) and the group with PTSD ($n=54$) according to the occurrence of PTSD. Based on 4 machine learning models, namely random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and Logistic regression (LR), a post-delivery PTSD prediction model was developed. The robustness and clinical practicability of the model were evaluated by using accuracy, precision, area under the curve (AUC) of the receiver operating characteristic curve (ROC), etc. Another 200 parturients from another hospital during the same period were included for external validation. **Results** The incidence of PTSD in parturients was 13.50%. Age, feeding method, planned pregnancy, skin-to-skin contact between mother and baby, acute labor, gestational diabetes, etc. were all independent influencing factors for the occurrence of PTSD after delivery (all $P<0.05$). The RF model has the best predictive effect on the incidence of PTSD in parturients ($AUC=0.882$) and the best comprehensive predictive efficacy. In external validation, the AUC values of the RF, SVM, KNN and LR models were 0.842, 0.734, 0.725 and 0.767, respectively. **Conclusions** The prediction model developed based on the RF algorithm has superior prediction performance and can identify postpartum PTSD more simply, quickly and effectively.

【Key words】 postpartum; posttraumatic stress disorder; machine learning; predictive model

[Mil Nurs, 2025, 42(10): 31-34, 39]

【收稿日期】 2024-10-12 【修回日期】 2025-07-02
【基金项目】 安徽省妇幼健康科研课题(2020FY14)
【作者简介】 李夏楠, 本科, 副主任护师, 电话: 0558-6402679
【通信作者】 刘冬梅, 电话: 0558-6403676

产后创伤后应激障碍(posttraumatic stress disorder, PTSD)是发生于产后以分娩为应激源的压力性障碍, 症状持续时间超过1个月, 症状包括过度警觉、

回避与分娩相关的刺激、与分娩相关的负面认知、情绪和功能损害^[1]。有研究^[2]显示,产妇产后 PTSD 发病率较高,它不仅会降低自身的身心健康水平,而且会影响配偶和家庭的和谐,甚至会影响婴幼儿的认知、情绪以及智力发展^[3]。现有研究^[4]表明,分娩方式与 PTSD 有关,而临床医务人员易低估自然阴道分娩与 PTSD 风险的联系。现阶段国内对产后 PTSD 的研究主要集中在危险因素横断面调查,并根据影响因素构建了各种评估模型。但各模型存在差异,进而影响模型评估效能,因此选择最佳模型对精准判断产后 PTSD 发生具有重要意义。机器学习已被广泛应用于疾病诊断、并发症、预后等评估模型中^[5],使得构建更为精准的自然分娩产后 PTSD 发生模型成为可能。本研究分别基于随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、K-最近邻(K-nearest neighbor, KNN)、逻辑回归(Logistic regression, LR)4 个机器学习模型构建了产妇产后 PTSD 发生的风险模型,为临床筛选最佳模型、实施临床干预提供参考依据。

1 对象与方法

1.1 研究对象 以便利抽样法选取 2023 年 12 月至 2024 年 8 月安徽省临泉县某院住院分娩产妇 400 例为研究对象。纳入标准:(1)年龄 ≥ 18 岁;(2)自然阴道分娩;(3)有阅读和理解能力;(4)均知情同意,签署知情同意书。排除标准:(1)认知障碍;(2)合并严重心脑血管等重要器官疾病;(3)产后非正常分娩结局者。(4)既往已有其他 PTSD 等精神障碍;(5)病历资料不完整。根据影响因素分析中样本量估计原则,样本量是自变量的 5~10 倍^[6]。本研究涉及自变量 17 个,选取 10 倍进行计算,考虑 20% 的失访率,所需的样本量为 204 例。本研究最终确定内部实验组样本量为 400 例,外部验证组样本量为 200 例。模型组样本量源于安徽省临泉县某院,外部验证组样本量源于同地区另一所医院。本研究经医院伦理委员会批准(2023005)。

1.2 方法

1.2.1 研究工具

1.2.1.1 一般情况调查表 自行设计,包括年龄、产次、胎次、文化程度、计划怀孕(夫妻双方意愿一致并备孕)、双胞胎、择期、母婴皮肤接触(产后 30 min 内首次接触且接触时长 ≥ 60 min)、喂养方式[(母乳(纯母乳+混合)/人工)、会阴撕裂(未撕裂、撕裂)、产中存在高血压、产时发热、会阴或盆腔疼痛、急产、阴道出血过多(胎儿娩出后 24 h 阴道出血量 ≥ 500 mL)、妊娠糖尿病等,所有信息由医院电子病历系统收集。

1.2.1.2 围产期创伤后应激障碍问卷(perinatal

post-traumatic stress disorder questionnaire, PPQ)

本研究采用杨涵^[7]于 2019 年汉化的中文版本,包含再体验(3 个条目)、高警觉(5 个条目)、逃避行为(6 个条目)3 个维度。量表采用 Likert 5 级评分,从“从不发生”至“经常发生”依次计 0~4 分。总分 0~56 分, ≥ 19 分为存在创伤后应激障碍。该量表的 Cronbach's α 系数为 0.84。本研究以 PPQ ≥ 19 分且持续症状 ≥ 1 个月判断为 PTSD 阳性。

1.2.2 资料收集及质量控制 为保障研究资料的一致性及有效性,2 名调查员均经过统一培训,熟悉医院电子病历系统的使用及调查表的填写方法,通过医院电子病历系统获取资料;由经过培训的调查员在产妇出院 1 个月后采用电话随访的形式评估产妇是否发生 PTSD。收集的资料经双人核对后录入,以确保数据的准确性。对于缺失变量则使用该变量的均值或相似样本均值来补充。调查表收集完成后,由专人保管并抽取 15% 内容进行复核。本研究通过电话随访发放问卷 623 份,有效回收 600 份(其中 200 份为外部验证组),问卷有效回收率为 96.31%。

1.2.3 统计学处理 采用 SPSS 26.0 和 R 4.2.0 软件进行分析,正态分布的计量资料以 $\bar{x} \pm s$ 表示,采用 t 检验;计数资料以例数和百分比表示,采用 χ^2 检验或秩和检验。采用 Lasso 回归特征变量筛选,将筛选得到的特征变量纳入 Logistic 回归进行多因素分析,以 $P < 0.05$ 或 $P < 0.01$ 为差异有统计学意义。借助 R 语言 rms、glmnet、randomForest、ggplot2 等程序分别构建 RF、SVM、KNN、LR 等 4 个机器学习预测模型。采用准确度、精确度、召回率、F1 值、接受者操作特性曲线(receiver operating characteristic curve, ROC)曲线下面积(area under the curve, AUC)等来评价模型的稳健性及临床实用性。

2 结果

2.1 内部实验组研究对象的一般资料 400 例产妇产年龄 20~46 岁,平均(27.86 \pm 4.11)岁;初产妇 180 例、经产妇 220 例;G1P1 168 例、G2P2 93 例、G3P3 8 例、G2P1 49 例、其他 82 例;小学及以下 9 例、初中 66 例、高中 79 例、专科 135 例、本科及以上 111 例;计划怀孕 366 例;双胞胎 21 例;择期分娩 330 例;母婴皮肤接触 371 例;母乳喂养 370 例;会阴撕裂 272 例;产中存在高血压 28 例;产时发热 22 例;会阴疼痛 379 例、盆腔疼痛 21 例;急产 20 例;阴道出血过多 21 例;妊娠糖尿病 64 例。

2.2 产妇发生 PTSD 的单因素分析 根据产妇是否发生 PTSD 分为未发生 PTSD 组($n=346$)和发生 PTSD 组($n=54$)。两组产妇在年龄、文化程度、喂养方式以及是否计划怀孕、双胞胎、择期、母婴皮

肤接触、急产、妊娠糖尿病等方面的差异均有统计学意义(均 $P < 0.05$)。本文仅列出差异有统计学意义的项目,见表 1。

2.3 产妇发生 PTSD 影响因素 Lasso 回归分析

Lasso 回归在一般线性回归的基础上加入正则项,可对不蕴含有用信息的特征进行减少权重操作,在保证最佳拟合误差的同时,使“参数”尽可能简单,从而达到减少过拟合目的。因此,为避免过度拟合,将单因素分析中有意义的变量纳入 Lasso 回归进一步筛选, λ 为惩罚系数,随 λ 改变模型变量不断减少,以 10 倍交叉验证误差为 $\lambda + SE$ 时为最优质值,结果共筛选出年龄、计划怀孕、母婴皮肤接触、喂养方式、是否急产、有无妊娠糖尿病等 6 个影响因素。

表 1 发生 PTSD 产妇的单因素分析(N=400)

项目	未发生 PTSD(n=346)		发生 PTSD (n=54)		t 或 z 或 χ^2	P
	PTSD(n=346)	(n=54)	t 或 z 或 χ^2	P		
年龄(岁, $\bar{x} \pm s$)	27.58 ± 4.01	29.61 ± 4.32	3.419	0.003		
文化程度[n(%)]			14.852	0.005		
小学及以下	9(2.60)	0(0.00)				
初中	65(18.79)	1(1.85)				
高中	62(17.92)	17(31.48)				
专科	113(32.66)	22(40.74)				
本科及以上	97(28.03)	14(25.93)				
计划怀孕[n(%)]			35.836	<0.001		
否	18(5.20)	16(29.63)				
是	328(94.80)	38(70.37)				
双胞胎[n(%)]			4.311	0.038		
否	331(95.66)	48(88.89)				
是	15(4.34)	6(11.11)				
择期[n(%)]			10.840	<0.001		
否	52(15.03)	18(33.33)				
是	294(84.97)	36(66.67)				
母婴皮肤接触[n(%)]			46.498	<0.001		
否	13(3.76)	16(29.63)				
是	333(96.24)	38(70.37)				
喂养方式[n(%)]			10.925	<0.001		
母乳喂养	326(94.22)	44(81.48)				
人工喂养	20(5.78)	10(18.52)				
急产[n(%)]			12.660	<0.001		
否	334(96.53)	46(85.19)				
是	12(3.47)	8(14.81)				
妊娠糖尿病[n(%)]			20.556	<0.001		
否	302(87.28)	34(62.96)				
是	44(12.72)	20(37.04)				

2.4 产妇发生 PTSD 的多因素 Logistic 回归分析

以是否发生 PTSD 为因变量,以 Lasso 回归有统计学意义的因素为自变量。变量赋值分类资料发生 PTSD、计划怀孕、母婴皮肤接触、喂养方式、急产、妊娠糖尿病均为二分类变量,赋值无=0、有=1,计量资料年龄原值录入,进行 Logistic 回归分析,变量筛选采用逐步法(模型纳入水准为 0.05,剔除水准为 0.10)。结果表明,似然比卡方检验提示回归模型有统计学意义($P < 0.05$),H-L 拟合优度检验提示模型具有较好的校准度($\chi^2 = 4.787, P = 0.781$),年龄、

计划怀孕、皮肤接触、喂养方式、急产、妊娠糖尿病均是产后是否发生 PTSD 的独立影响因素,见表 2。

2.5 ROC 曲线分析 4 种预测模型对产妇发生 PTSD 的预测效能 同时将年龄、计划怀孕、母婴皮肤接触、喂养方式、急产、妊娠糖尿病 6 个自变量进入个机器学习模型中,分析其在各模型的预测效果。在 RF、SVM、KNN 和 LR 4 个模型中 ROC 曲线分析结果显示,RF 模型对产妇 PTSD 的发生率的预测效果最好($AUC = 0.882$),以 RF 模型在预测产妇 PTSD 的发生率的能力方面更优。

表 2 产妇发生 PTSD 的多因素 Logistic 回归分析(n=400)

项目	β	Sb	Wald χ^2	P	OR	95%CI
常量	-2.962	1.464	4.097	0.043	0.052	-
年龄	0.095	0.041	5.335	0.021	1.100	1.014~1.192
计划怀孕	-2.254	0.459	24.152	<0.001	0.105	0.042~0.258
母婴皮肤接触	-2.155	0.484	19.819	<0.001	0.116	0.045~0.299
喂养方式	1.532	0.516	8.808	0.003	4.627	1.682~12.726
急产	1.451	0.518	7.852	0.005	4.266	1.547~11.770
妊娠糖尿病	1.325	0.396	11.192	0.001	3.763	1.731~8.179

2.6 4 种机器学习预测模型性能的比较

RF、SVM、KNN、LR 等 4 种算法建立的 PTSD 预测模型中,RF 模型的综合预测效能最佳,精准度、召回率、精确度、F1 得分均最高。RF 模型的整体预测能力优于其他 3 种模型,见表 3。

表 3 4 种机器学习模型对产妇发生 PTSD 的预测模型性能比较

模型	准确度	召回率	精确度	F1 得分	AUC	95%CI
RF	0.973	0.971	0.971	0.968	0.882	0.819~0.945
SVM	0.915	0.915	0.909	0.905	0.807	0.718~0.896
KNN	0.861	0.861	0.862	0.835	0.821	0.773~0.909
LR	0.918	0.918	0.913	0.908	0.857	0.778~0.936

2.7 4 种机器学习预测模型的外部验证比较 基于 RF 诊断模型的 AUC 值最高(0.842),SVM、KNN 和 LR 模型的 AUC 值分别为 0.734、0.725、0.767。

2.8 RF 模型中的特征重要性比较 在 RF 模型中,排名前 10 位的特征分别为年龄、母婴皮肤接触、文化程度、计划怀孕、胎次、妊娠糖尿病、会阴撕裂、择期、喂养方式、急产,见图 1。

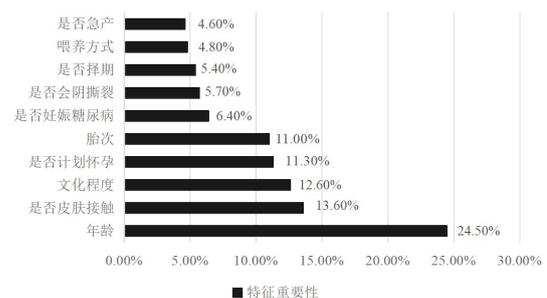


图 1 RF 模型中的特征重要性比较

3 讨论

3.1 产妇产道分娩后发生 PTSD 的现状 本研究中产妇产道分娩后 PTSD 的发生率为 13.5%，与以往研究^[3-4]结果基本一致。现阶段，国内对产后 PTSD 危险因素的横断面调查结论多同质化，而产后 PTSD 的预测模型研究仅限于列线图预测、模型结构方程模型等^[8]，无法对比预测模型的优越性，机器学习在产后 PTSD 预后评估方面的运用成为目前研究的重点。本研究显示，年龄、计划怀孕、母婴皮肤接触、喂养方式、急产、妊娠糖尿病均是产后是否发生 PTSD 的独立影响因素(均 $P < 0.05$)。其原因分析可能为低龄/高龄产妇身心适应较难、非计划怀孕心理准备不足、母婴皮肤接触缺失影响角色认同、母乳喂养困难引发自我否定、急产带来应激失控感、妊娠糖尿病叠加身心压力等。建议加强孕产期心理评估与教育，尤其对年龄偏大、计划外怀孕、妊娠糖尿病等高危产妇，提升其应对能力和控制感；规范母婴皮肤接触流程，构建母乳喂养支持体系以促进情感联结与激素调节；应用有效的预测模型进行产后早期心理筛查，为筛查阳性及高风险产妇提供个体化的心理干预，如认知行为疗法、眼动脱敏再处理法等，并构建产科-精神心理科协作的支持体系。

3.2 产妇产道分娩后发生 PTSD 的危险因素分析
3.2.1 年龄 本研究显示，发生 PTSD 组的产妇平均年龄为(29.61±4.32)岁，高于未发生 PTSD 组；年龄是产后 PTSD 发生的独立影响因素，且在 RF 模型中排名第 1 位。但 Suttora 等^[9]研究结果显示，年龄与围产期 PTSD 无显著关系，分析其原因可能为在中国“二胎”“三胎”政策的相继开放的背景下，产妇平均年龄有所增加，加之高龄产妇更易受到家庭、工作等的影响，产后可能更易发生 PTSD。

3.2.2 母婴皮肤接触 陈颖等^[10]研究表明，早期母婴皮肤接触[OR=0.20, 95%CI(0.16, 0.25)]是产后 PTSD 的保护因素，这与本研究的研究结果基本一致。本研究结果还显示，RF 模型中的特征重要性占比中，是否母婴皮肤接触排第 2 位，这表明与胎儿皮肤接触对产妇情绪的重要性，围产期分离等分娩中产妇的感知觉以及情绪是影响产后 PTSD 的重要因素。

3.2.3 计划怀孕 本研究显示，计划怀孕是产后 PTSD 发生的独立影响因素，且在 RF 模型特征重要性排名第 4 位。分析原因可能为非计划怀孕的产妇可能缺乏充足的心理准备和社会支持，更易产生消极情绪，而产妇情绪、低水平的心理一致感以及社会支持也是影响产后 PTSD 的重要因素^[11]。

3.2.4 喂养方式 孟亦心等^[12]研究表明，母乳喂养

是剖宫产产妇发生 PTSD 的保护因素。这与本研究的研究结果基本一致。RF 模型特征重要性该因素排第 9 位，这表明无论自然分娩还是剖宫产产妇，母乳喂养更易降低产妇 PTSD 发生率。

3.2.5 急产 有研究^[12]表明，急诊手术是产妇发生 PTSD 的危险因素。本研究显示，急产是产后 PTSD 发生的独立影响因素之一，且在 RF 模型特征重要性中排第 10 位。分析原因可能是：一方面急产的产妇因急性应激可能产生负性情绪及对胎儿的担心；另一方面，Grekin 等^[13]研究表明，产程失控、器械辅助分娩是产后 PTSD 的重要危险因素。

3.2.6 妊娠糖尿病 本研究显示，妊娠糖尿病是产后 PTSD 发生的独立影响因素，妊娠糖尿病产妇产后 PTSD 风险增加 3.763 倍，且在 RF 模型特征重要性中排第 6 位，这与李晓燕等^[8]研究结果显示结果相似。妊娠糖尿病加重孕妇精神担忧和应激反应，影响机体正常代谢，增加不良妊娠和产后 PTSD 发生风险。

3.2.7 其他 本研究发现，胎次在产后 PTSD 的危险因素中虽然不是独立影响因素，但是在 RF 模型特征重要性中排第 5 位。吴琴等^[14]研究发现，分娩次数是 PTSD 发生的影响因素，还有研究^[15]表明，妊娠间隔时间 < 2 年是产后 PTSD 的危险因素，说明再次分娩有可能增加产后 PTSD 发生的危险性；同时，养育 2 个或多个孩子，产妇体力和精力上的消耗会加重 PTSD 患病发生率。

3.3 模型的准确性与实用性分析 本研究发现，将构建的 RF、SVM、KNN 和 LR 模型在测试集中进行验证，各模型的 AUC 值分别为 0.882、0.807、0.821、0.857，在外部验证中上述模型的 AUC 值分别为 0.842、0.734、0.725、0.767。从模型算法和运用原理来看，RF 模型可自动分析因子间的交互作用和非线性作用，可通过有目的放飞对数据进行利用，对数据要求低^[16]。临床医护人员可以采用 RF 模型，更加简单、快速、有效地识别产后 PTSD。

4 小结

本研究分别基于 RF、SVM、KNN 和 LR 构建了产妇产道分娩后创伤后应激障碍发生风险模型，模型的构建规避了单一指标评估效能有限的弊端，有助于医护人员识别高危患者并早期实施干预，降低产妇产道分娩后 PTSD 的发生。但本研究中母乳喂养包括纯母乳与混合喂养，未对母乳喂养方式进行亚组分析，有待于开展前瞻性队列研究进行深入探讨。另外，由于本研究为单中心研究，且受样本量的限制具有一定局限性，未来还需要开展多中心大样本研究以验证本研究构建模型的可扩展性和有效性。

(下转第 39 页)