



DOI: 10.11817/j.issn.1672-7347.2020.190655

http://xbyxb.csu.edu.cn/xbwk/fileup/PDF/2020101215.pdf

随机森林算法在产后抑郁风险预测中的应用

肖美丽¹, 晏春丽², 付冰³, 杨淑平⁴, 朱姝娟³, 杨东琪^{1,5}, 雷倍美^{1,6}, 黄瑞瑞^{1,7}, 雷俊^{1,3}

(1. 中南大学湘雅护理学院, 长沙 410013; 2. 中南大学湘雅三医院肿瘤科, 长沙 410013;
3. 中南大学湘雅三医院妇产科, 长沙 410013; 4. 中南大学数学与统计学院, 长沙 410083;
5. 河南省人民医院妇科, 郑州 450000; 6. 中南大学湘雅医院耳鼻喉科, 长沙 410008;
7. 湖南省医药学院护理学院, 湖南 怀化 418000)

[摘要] **目的:** 探讨随机森林算法在产后抑郁影响因素的筛选和风险预测中的应用效果。**方法:** 选取2017年6月至2018年6月在湖南省长沙市某三甲医院接受产前检查并在该医院分娩, 符合纳入和排除标准的孕早期妇女为研究对象。入组时, 使用自编的调查问卷、中文版爱丁堡产后抑郁量表(Edinburgh Postnatal Depression Scale, EPDS)调查研究对象的人口经济学、心理社会学、生物学和产科及其他特征; 产后4~6周内, 采用中文版EPDS进行抑郁评分和自编的产后资料问卷收集分娩和产后资料。采用R软件在训练数据集上建立产后抑郁风险预测的随机森林模型, 在验证数据集上采用预测准确率、灵敏度、特异度、阳性预测值、阴性预测值和曲线下面积(area under curve, AUC)评价模型的预测效果。**结果:** 共调查406例研究对象, 其中150例的EPDS得分 ≥ 9 , 产后抑郁的发生率为36.9%。通过随机森林算法对训练集建立的模型在验证集上验证, 得出预测准确度为80.10%, 灵敏度为61.40%, 特异度为89.10%, 阳性预测值为73.00%, 阴性预测值为82.80%, AUC值0.833。采用随机森林算法通过变量重要性评分对产后抑郁影响因素的重要程度进行排序, 得出排名前10位的重要预测变量为产前抑郁、产后经济担忧程度、产后工作担忧程度、孕早期血清游离三碘甲腺原氨酸、孕晚期高密度脂蛋白、向婴幼儿发脾气、孕早期血清总胆固醇、孕早期三酰甘油、孕晚期血细胞比容和三酰甘油。**结论:** 随机森林算法在产后抑郁的风险预测中具有较大优势, 通过综合评价机制能从复杂的多因素中识别出产后抑郁的重要影响因素, 并进行定量分析。这对识别产后抑郁关键因素, 进行及时、有效干预具有重要意义。

[关键词] 随机森林; 产后抑郁; 影响因素; 风险预测

Risk prediction for postpartum depression based on random forest

XIAO Meili¹, YAN Chunli², FU Bing³, YANG Shuping⁴, ZHU Shujuan³,
YANG Dongqi^{1,5}, LEI Beimei^{1,6}, HUANG Ruirui^{1,7}, LEI Jun^{1,3}

收稿日期(Date of reception): 2019-09-23

第一作者(First author): 肖美丽, Email: 786001878@qq.com, ORCID: 0000-0001-9756-5400

通信作者(Corresponding author): 雷俊, Email: lei junbao@126.com, ORCID: 0000-0002-0430-4672

基金项目(Foundation item): 国家自然科学基金(81874267); 湖南省重点研发计划项目(2018SK2068); 中南大学湘雅三医院“新湘雅”人才工程项目(20170305)。This work was supported by the National Natural Science Foundation (81874267), Hunan Provincial Major Research Development Program (2018SK2068), and “New Talent” Program of the Third Xiangya Hospital of Central South University (20170305), China.

(1. Xiangya Nursing School, Central South University, Changsha 410013; 2. Department of Oncology, Third Xiangya Hospital, Central South University, Changsha 410013; 3. Department of Obstetrics and Gynaecology, Third Xiangya Hospital, Central South University, Changsha 410013; 4. School of Mathematics and Statistics, Central South University, Changsha 410013; 5. Department of Gynaecology, Henan Provincial People's Hospital, Zhengzhou 450000; 6. Department of Otolaryngology, Xiangya Hospital, Central South University; 7. School of Nursing, Hunan University of Medicine, Huaihua Hunan 418000, China)

ABSTRACT

Objective: To explore the application of random forest algorithm in screening the risk factors and predictive values for postpartum depression.

Methods: We recruited the participants from a tertiary hospital between June 2017 and June 2018 in Changsha City, and followed up from pregnancy up to 4–6 weeks postpartum. Demographic economics, psychosocial, biological, obstetric, and other factors were assessed at first trimesters with self-designed obstetric information questionnaire and the Chinese version of Edinburgh Postnatal Depression Scale (EPDS). During 4–6 weeks after delivery, the Chinese version of EPDS was used to score depression and self-designed questionnaire to collect data of delivery and postpartum. The data of subjects were randomly divided into the training data set and the verification data set according to the ratio of 3 : 1. The training data set was used to establish the random forest model of postpartum depression, and the verification data set was used to verify the predictive effects via the accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and AUC index.

Results: A total of 406 participants were in final analysis. Among them, 150 of whom had EPDS score ≥ 9 , and the incidence of postpartum depression was 36.9%. The predictive effects of random forest model in the verification data set were at accuracy of 80.10%, sensitivity of 61.40%, specificity of 89.10%, positive predictive value of 73.00%, negative predictive value of 82.80%, and AUC index of 0.833. The top 10 predictive influential factors that screening by the variable importance measure in random forest model was antenatal depression, economic worries after delivery, work worries after delivery, free triiodothyronine in first trimesters, high-density lipoprotein in third trimester, venting temper to infants, total serum cholesterol and serum triglyceride in first trimester, hematocrit and serum triglyceride in third trimester.

Conclusion: Random forest has a great advantage in risk prediction for postpartum depression. Through comprehensive evaluation mechanism, it can identify the important influential factors for postpartum depression from complex multi-factors and conduct quantitative analysis, which is of great significance to identify the key factors for postpartum depression and carry out timely and effective intervention.

KEY WORDS

random forest; postpartum depression; influencing factors; risk prediction

产后抑郁(postpartum depression, PPD)是由生物、心理、社会等多种因素综合作用所致的非精神障碍性综合征。世界范围内的流行病学调查^[1]显示:

PPD的发病率为0.1%~26.3%,且呈上升趋势,已成为影响母婴健康的全球公共卫生问题。PPD发病机制复杂且尚不明确。现有研究多集中于PPD影响因素

的验证和确定,如通过量表对PPD危险因素进行评估,采用logistic回归通过OR判断某一因素对PPD的风险性^[2-3]。对于多因素综合作用下个体患PPD的风险评估研究较少,且多采用传统的统计学方法综合生物、心理、社会等多维度探讨PPD的个体患病风险,易出现过拟合、预测精度和效度低等问题^[4]。因此,如何综合生物、心理、社会等多维度因素,构建PPD风险预测模型,筛选出PPD的关键影响因素,并对个体患PPD的风险进行有效判别是PPD风险预测研究的关键。近年来,随机森林算法在疾病风险预测、预警和预后中广泛应用,其能对医学数据中的混杂数据、缺失值或离群值及较高维度的数据进行有效地处理,不易出现过拟合,预测精度也较准确^[5-6]。已有研究^[6-7]将随机森林算法用于精神分裂症、抑郁症等的基因组学研究和个体患病的分类判别中,结果显示随机森林算法在筛选重要代谢物和有效判别个体患病中具有较大的优势。PPD与精神分裂症、抑郁症等均为多因素综合作用所致,但目前将随机森林算法用于PPD的风险预测,对患病率定量分析和寻找关键影响因素的研究较少。因此,本研究通过收集人口经济学、生物学、心理社会学、产科及其他特征等多维度因素,建立PPD风险预测的随机森林模型,探讨随机森林算法在PPD影响因素的筛选和风险预测研究中的应用效果。

1 资料与方法

1.1 一般资料

选取2017年6月至2018年6月在长沙市某三甲医院进行定期产前检查,计划在该医院分娩且产后42 d在该医院进行产后随访的孕早期妇女,初次入组时间为孕8~12周,分别在孕22~24周、孕36~38周、产后4~6周进行3次随访。纳入标准:1)孕早期女性;2)单活胎妊娠者;3)妊娠前甲状腺功能、血清脂质无异常者;4)能正确理解问卷内容并独立填写问卷者;5)自愿参与本研究者。排除标准:1)妊娠前有抑郁疾病史或精神疾病史者;2)既往患有严重躯体疾病或脑外伤者;3)孕期患有严重产科合并症和/或并发症者;4)孕期出现不良孕产史,如流产、胎儿畸形、死胎等者;5)在随访过程中,因身体不适合参与本研究或中途退出者;6)资料缺失比例高达20%者。

1.2 方法

1.2.1 研究工具

调查问卷:1)人口经济学问卷,包括年龄、受教育水平、居住地、职业、家庭人均月收入、有无医

疗保险。2)心理社会学问卷,包括性格特征、夫妻关系、婆媳关系、中文版爱丁堡PPD量表(Edinburgh Postnatal Depression Scale, EPDS)评分、新生儿健康满意度、产后工作担忧程度、产后经济担忧程度。3)产科及其他特征问卷,主要包括妊娠期和产褥期资料。妊娠期资料包括孕周、预产期、孕产史、流产史、妊娠合并症;产褥期资料包括分娩孕周、分娩方式、新生儿资料(性别、出生体重、Apgar得分等)、产后出血量、子宫复旧、母乳量、喂养方式、母婴同室、母婴血型不合、睡眠情况、夜晚起床次数等。4)生物学因素问卷。从医院病历系统中提取孕早期、孕中期和孕晚期指标,如血细胞比容(hematocrit, Hct)、血糖(blood sugar, BS)、血清三酰甘油(triglyceride, TG)、高密度脂蛋白(high-density lipoprotein, HDL)、低密度脂蛋白(low-density lipoprotein, LDL)、血清总胆固醇(total cholesterol, TC)、甲状腺过氧化物酶抗体(thyroid peroxidase antibody, TPO-Ab)、游离甲状腺素(free thyroxine, FT₄)、游离三碘甲状腺原氨酸(free triiodothyronine, FT₃)、促甲状腺素(thyroid stimulating hormone, TSH)。

其中EPDS采用香港中文大学Lee等^[8]编译和修订的中文版EPDS,该量表涉及心境、乐趣、自责、焦虑、恐惧、失眠、应付能力、悲伤、哭泣和自伤10个条目,评分为4个等级(0分表示从未,1分表示偶尔,2分表示经常,3分表示总是),总分范围0~30,评分越高,表示抑郁症状越严重。该量表在中国大陆的Cronbach' α 系数为0.79,重测信度为0.85^[9]。由于研究目的和研究人群不同,EPDS的界值有所不同。本研究采用国内推荐界值^[10]: <9分为无抑郁, ≥ 9 分为抑郁, ≥ 13 分为重度抑郁。其余问卷为自制。

1.2.2 随机森林模型的建立

随机森林是一种机器学习方法,通过多个决策树对数据进行综合分类,关联性检验、预测和解释^[11]。本研究中结局变量是指所研究对象是否发生PPD,是决策进行分类的最终目的。而解释变量为一系列可能引起PPD发生的危险因素,如血脂的变化,激素的改变等,用于对结局变量进行分类。建立随机森林模型的步骤^[4]如下:1)通过Bootstrap自助法从原始数据中有放回地抽取 n 个训练样本建立 n 棵树(ntree)。2)在生成树的过程中,在每个树节点处从所有变量中随机抽取 m 个变量(mtry),在 m 个变量中选择最具有分类能力的变量对数据进行分类。3)在Bootstrap中未被抽到的数据为测试样本,用来评价每棵树的性能。 n 棵树组成一个随机森林,用于数据

的综合判别和分类,如图1。

1.2.3 变量重要性评分

随机森林算法在应用过程中得到的主要结果为变量的重要性评分,其目的是对解释变量在结局发展中的重要性进行评价^[5]。对测试样本中的变量进行重新排列,对重新排列前后出现的误差率进行再次

计算,比较两个误差率之间是否存在差值,差值即为每棵树的重要评分标准。在随机森林模型的建立过程中,结局变量的分类主要是依据解释变量的分类程度,变量的重要性评分越高表示该变量的应用能力越强。

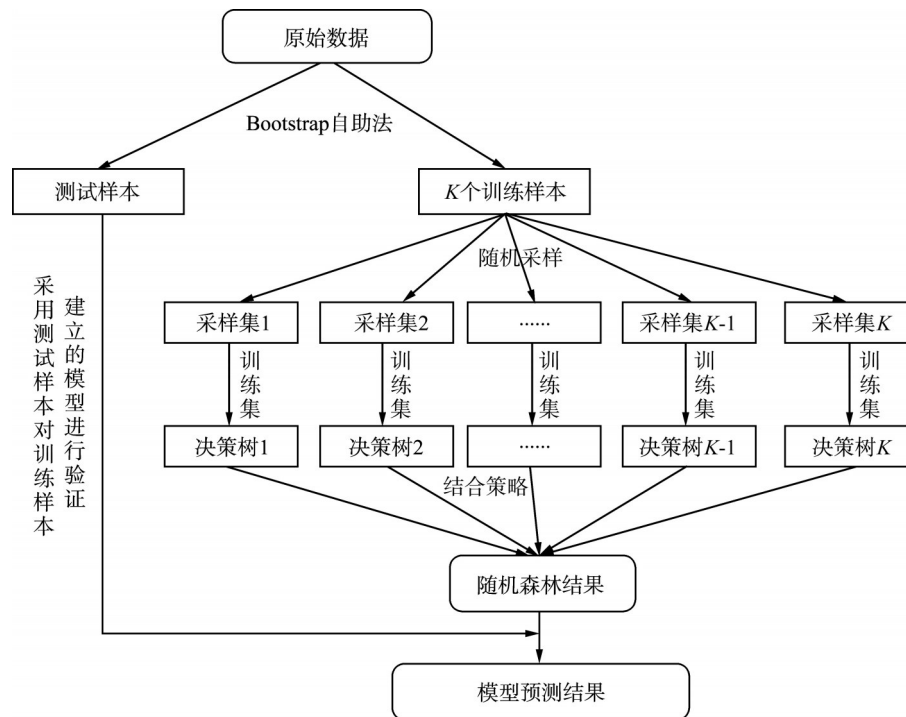


图1 PPD风险预测的随机森林模型建立过程

Figure 1 Development process of random forest model for postpartum depression risk prediction

1.2.4 数据收集过程

采用问卷调查法,在孕早期(入组时)使用孕早期女性调查问卷收集研究对象的一般资料、妊娠资料和心理社会学资料;在征求研究对象同意后,在孕早期(8~12周)、孕中期(22~24周)和孕晚期(32~38周)抽取其静脉血进行常规检查时,采集全血标本检测甲状腺相关激素、血清脂质水平;在产后4~6周内,使用产妇心理调查问卷收集研究对象的EPDS得分、分娩资料、产科及其他资料。本研究入组孕早期女性560例,中途退出/失联20例,流产3例,未采集孕早期血液标本24例;孕中期第1次随访时,中途退出/失联15例,死胎3例,未采集孕中期血液标本23例;孕晚期第2次随访时,中途退出/失联19例,未采集孕晚期血液标本18例;产后4~6周第3次随访时,中途退出/失联18例,产后问卷填写不完整11例。最终纳入资料完整,符合纳入和排除标准的孕

妇406例,其中PPD组150例,对照组256例,回收有效率为72.50%。

1.3 统计学处理

采用SPSS 18.0统计软件包对数据进行单因素分析,符合正态分布的计量资料采用均数±标准差($\bar{x} \pm s$)表示,组间比较采用 t 检验;计数资料采用频数或百分比表示,组间比较采用 χ^2 检验。等级资料采用中位数和四分位间距表示,组间比较采用秩和检验。采用R软件中的Package random forest (Version 4.6)软件包进行多因素分析, $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 基本特征

入组406例研究对象,孕周为6~12周,年龄为

20~38(28.61±3.11)岁, 其中独生子女101例。文化程度: 大学及以上242例, 高中/中专109例, 初中45例, 小学及以下10例。居住地: 农村26例, 城镇42例, 城市338例。有医疗保险340例。月收入: ≤4 000元101例, 4 001~6 000元209例, >6 000元96例。以EPDS评分<9为临界值, 将406例研究对象分为PPD组和对照组, 两组一般资料差异无统计学意义($P>0.05$), 具有可比性。

2.2 PPD发生率

产后4~6周随访时, 406例研究对象EPDS得分为0~22(7.52±3.225)。其中150例EPDS得分≥9, PPD的发生率为36.9%, 22例(5.42%)存在严重的抑郁(EPDS得分≥13)。

2.3 调查数据的预处理

对406例研究对象的一般人口学资料、心理社会学资料、生物学资料和产科及其他特征资料进行单因素分析, 将 $P<0.2$ 的变量用于随机森林模型的构建, 最终纳入49个变量。1)人口经济学因素: 职业。2)心理社会学因素: 性格特征、产前抑郁、婆媳关系、夫妻关系、新生儿健康满意度、产后工作担忧程度、产后经济担忧程度、向婴幼儿发脾气、婴幼儿教育能力担忧程度。3)产科及其他特征: 怀孕次数、新生儿性别、母婴同室、母乳量、喂养方式、子宫复旧、产褥期感染、睡眠情况和夜晚起床次数。4)生物学因素: 孕早期、孕中期和孕晚期的Hct, BS, TG, HDL, LDL, TC, ATPO, FT₄, FT₃, TSH。

2.4 随机森林模型的建立和评价

在随机森林模型的建立过程中, 涉及的参数主

要包括n_{tree}和m_{try}。在纳入的406例研究对象中, 采用Bootstrap自助法抽取66.50%(270例)数据建立随机森林训练模型, 设置不同参数下的随机森林模型并在验证集上验证, 发现当n_{tree}=1 000, m_{try}=15时, 随机森林模型的预测效果最佳, 预测准确率为80.10%、灵敏度为61.40%、特异度为89.10%、阳性预测值为73.00%、阴性预测值为82.80%、曲线下面积(AUC)为0.833, 具体分类结果见表1。训练集和验证集AUC分别为0.776(图2A)和0.833(图2B)。

表1 PPD风险预测的随机森林模型在验证数据集中的分类结果

Table 1 Classification of random forest model for postpartum depression risk prediction in validation data set

随机森林模型预测PPD	实际/例		合计/例
	无	有	
非PPD	82	17	99
PPD	10	27	37
合计/例	92	44	136

2.5 随机森林模型的变量重要性评分

以随机森林模型预测精度的平均下降量对PPD随机森林模型的总体变量重要性评分进行排序, 排在前10位的重要预测变量为: 产前抑郁、产后经济担忧程度、产后工作担忧程度、FT₃孕早期、HDL孕晚期、向婴幼儿发脾气、TCHOL孕早期、TG孕早期、Hct孕晚期、TG孕晚期(表2)。

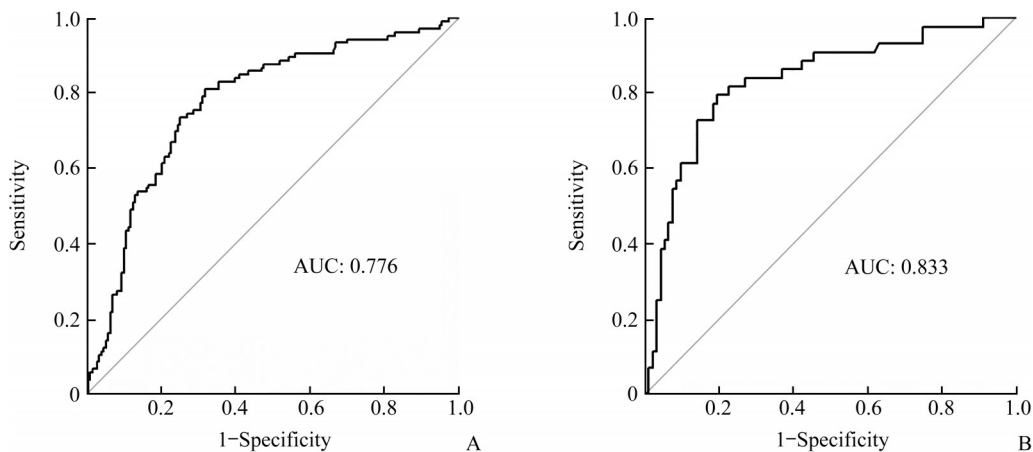


图2 PPD风险预测的随机森林模型训练集和验证集上ROC曲线

Figure 2 ROC curve of random forest model for postpartum depression risk prediction in training data set and validation data set

A: Area under ROC curve in training data set; B: Area under ROC curve in validation data set.

表2 PPD风险预测的随机森林模型变量的重要性评分

Table 2 Importance scores of input variable in random forest model for postpartum depression risk prediction

变量	预测精度的平均下降量	排序	变量	预测精度的平均下降量	排序
产前抑郁	37.392	1	LDL _{孕晚期}	0.663	26
产后经济担忧程度	17.272	2	FT _{3孕中期}	0.603	27
产后工作担忧程度	10.903	3	夫妻关系	0.555	28
FT _{3孕早期}	6.688	4	BS _{孕中期}	0.208	29
HDL _{孕晚期}	6.238	5	FT _{4孕中期}	0.147	30
向婴幼儿发脾气	4.613	6	ATPO-Ab _{孕中期}	0.102	31
TCHOL _{孕早期}	3.768	7	喂养方式	-0.137	32
TG _{孕早期}	3.076	8	Hct _{孕中期}	-0.402	33
Hct _{孕晚期}	3.066	9	LDL _{孕早期}	-0.421	34
TG _{孕晚期}	2.948	10	FT _{4孕中期}	-0.566	35
夜晚起床次数	2.773	11	FT _{3孕晚期}	-0.639	36
BS _{孕晚期}	2.679	12	HDL _{孕早期}	-0.810	37
怀孕次数	2.512	13	新生儿健康满意度	-0.986	38
LDL _{孕中期}	2.392	14	母婴同室	-1.160	39
睡眠情况	2.367	15	TSH _{孕中期}	-1.186	40
婴幼儿教育能力担忧程度	1.957	16	HDL _{孕中期}	-1.193	41
FT _{4孕晚期}	1.786	17	婆媳关系	-1.208	42
TCHOL _{孕晚期}	1.720	18	BS _{孕早期}	-1.243	43
TSH _{孕晚期}	1.478	19	新生儿性别	-1.597	44
TG _{孕中期}	1.238	20	性格特征	-1.948	45
TSH _{孕早期}	1.176	21	ATPO-Ab _{孕晚期}	-2.526	46
母乳量	1.110	22	职业	-2.553	47
ATPO-Ab _{孕早期}	1.077	23	产褥期感染	-2.684	48
TCHOL _{孕中期}	0.995	24	Hct _{孕早期}	-3.942	49
子宫复旧	0.866	25			

3 讨论

由于评估工具、评估时间、评估标准、样本量及各地文化背景等差异, 研究报道的不同国家或地区PPD的发生率也存在差异。一项在亚洲文化背景下, 基于17个亚洲国家、64项研究的PPD的系统综述^[12]显示: 亚洲地区PPD的发生率为3.5%~63.3%。2013年对中国PPD的发生率进行系统分析^[13]发现: PPD的发生率为1.1%~52.1%, 总体发生率约为14.7%。本研究结果显示: 在产后4~6周采用EPDS评估研究对象PPD的发生率为36.9%(EPDS≥9分)。这与上述研究结果较为一致。但本研究与国内其他相关研究比较, PPD的发生率稍高, 如山东一项PPD的纵向研究^[14]显示产后1周PPD(EPDS得分≥10)的发生率约为23.1%; 李华丽等^[15]对西安市1994名产妇的调查发现PPD(EPDS得分≥13)的发生率为33.7%。分析原因可能是, 本研究对PPD的评估时间在4~6

周, 为抑郁的高发时间段, 这在一定程度上增加了PPD的阳性筛查率。此外, 随着我国对围产期孕、产妇产心理保健的重视, 社交媒体对PPD的症状和危害的科普, 以及社会大众对孕产妇抑郁的认知提高, 在筛查过程中, 产妇会主动地表达自己的内心感受, 咨询PPD的相关知识, 提高了PPD的筛查阳性率。因此, 在我国生育政策调整下, 应关注产妇心理健康, 对PPD筛查阳性的孕妇, 提供相应的心理咨询服务, 指导和帮助其正确地面对PPD情绪, 从而降低PPD对母婴的危害。

本研究采用随机森林算法, 综合人口经济学、心理社会学、生物学和产科及其他因素等多个维度对PPD进行风险预测并进行验证, 发现随机森林算法作为一种非线性建模工具, 在处理不同类型变量(计量资料、计数资料和等级资料)时具有很好的容忍性^[5]。在建模的过程中, 随机森林算法能综合考虑各类PPD的影响因素, 并给出PPD影响因素的重要性

评分,具有较高的预测准确率(80.10%)。这说明随机森林算法在筛选和鉴别PPD阳性患者中具有较大的优势。然而就变量的解释程度而言,随机森林算法建立模型的过程是不断进行重复抽样、随机选择特征和决策分类的,其建立模型的过程相当于一个“黑匣子”,无法看清内部结构,对于每一个变量及变量间的解释较为困难。因此,未来可考虑将传统的统计学方法和随机森林等多种算法进行融合,探索PPD的预测效能和提高变量的解释程度。此外,由于本研究中收集的大部分数据均来自孕产妇常规检查的指标,而随机森林模型在PPD风险预测中显示了较好的优越性,这也提示随机森林算法建立的PPD风险预测模型可识别出常规产检指标中的高危因素,有可能做到在不增加孕产妇经济负担和时间负担的情况下,通过常规产检指标识别出PPD高危人群,做到早发现、早识别和早干预。

随机森林在不断生长和分类判别的过程中,还可以通过分值法对各特征变量的重要性进行评估,并依据分值大小对各特征变量的重要性进行降序排列,从而判断各特征变量对结局变量的影响和筛选出相对重要的变量^[16]。本研究采用平均预测精度下降量对PPD风险预测的随机森林模型中的影响因素的重要程度进行排序,发现心理学因素和生物学因素对PPD的影响较大,其中心理学因素对PPD影响较大的因素为产前抑郁、产后工作担忧、产后经济担忧,生物学因素中对PPD影响较大的因素主要集中在孕早期和孕晚期的血清脂质和血清甲状腺相关激素。目前关于PPD的研究也多集中在心理学和生物学两方面^[17],在心理学中探讨最多的为产前抑郁、既往精神病史对PPD的影响,且一致认为产前抑郁和既往精神史是PPD最强的预测因子之一^[17-18]。在生物学因素中,孕早期和孕晚期血清脂质和血清甲状腺相关激素对PPD的影响较大,这提示孕早期和孕晚期血清脂质和血清甲状腺相关激素的变化可能是PPD发生的生物标志物。

本研究通过随机森林算法从人口经济学、心理学、社会学、生物学、产科及其他特征因素等多个维度构建PPD风险预测模型并进行验证,分析了可能与PPD相关的因素。随机森林通过综合评价机制(变量重要性评分)从复杂的多维度因素中识别出重要的影响因素,并定量估计各因素对PPD的影响,有助于及时判别个体患PPD的特征、危险人群和发现可能的发生机制,在PPD的早期筛查、早期识别和早期干预中具有较大的指导意义。但本研究还存在一定的局限性,如生物学因素主要为孕早期、孕中期和孕晚期的数据,未纳入产褥期的生物学数据;对于

心理学因素的评估,仅在孕早期采用量表筛查了孕妇的心理学中的产前抑郁,部分因素未用量表测量,且未考虑丈夫和新生儿等其他因素对产妇抑郁的影响。在未来的研究中,将全面、综合考虑各类影响,扩大PPD的时间测量点,以寻找PPD的高危时间点。

利益冲突声明: 作者声称无任何利益冲突。

参考文献

- [1] Norhayati MN, Hazlina NHN, Asrenee AR, et al. Magnitude and risk factors for postpartum symptoms: a literature review [J]. *J Affect Disord*, 2015, 175: 32-34.
- [2] 蔡飞亚, 况利, 王我, 等. 基于社会心理因素的产后抑郁症模型的建立和评价[J]. *第二军医大学学报*, 2017, 38(4): 476-481.
CAI Feiya, KUANG Li, WANG Wo, et al. Prediction model for postpartum depression based on social psychological factors: establishment and evaluation[J]. *Academic Journal of Second Military Medical University*, 2017, 38(4): 476-481.
- [3] 尹航. 孕中、孕晚期单胺类神经递质水平与产后抑郁的关系研究[D]. 长沙: 中南大学, 2018.
YIN Hang. The levels of monoamine neurotransmitters in the second and third trimester of pregnancy and postpartum depression[D]. Changsha: Central South University, 2018.
- [4] 韩玉, 施海龙, 曲波, 等. 随机森林方法在医学中的应用[J]. *中国预防医学杂志*, 2014, 15(1): 79-81.
HAN Yu, SHI Hailong, QU Bo, et al. Application of random forest method in medicine[J]. *China Preventive Medicine*, 2014, 15(1): 79-81.
- [5] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. *统计与信息论坛*, 2011, 26(3): 32-38.
FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests[J]. *Statistics Information Forum*, 2011, 26(3): 32-38.
- [6] Huang L, Jin Y, Gao Y, et al. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest[J]. *Neurobiol Aging*, 2016, 46: 180-191.
- [7] 于建民. 老年人抑郁症患病风险评估方法研究[D]. 北京: 北京理工大学, 2016.
YU Jianmin. The research on risk assessment method of elderly depression[D]. Beijing: Beijing Institute of Technology, 2016.
- [8] Lee DTS, Yip SK, Chiu HFK, et al. Detecting postnatal depression in Chinese women[J]. *Br J Psychiatry*, 1998, 172(5): 433-437.
- [9] Wang Y, Guo X, Lau Y, et al. Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale[J]. *Inter J Nursg Stud*, 2009, 46(6): 813-823.
- [10] 张慧琳, 李乐之. 三种国外产后抑郁量表应用的比较分析[J].

- 中华护理杂志, 2007, 42(2): 186-188.
- ZHANG Huilin, LI Lezhi. Comparative analysis of 3 postpartum depression scales abroad[J]. Chinese Journal of Nursing, 2007, 42(2): 186-188.
- [11] 姚登举. 面向医学数据的随机森林特征选择及分类方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2016.
- YAO Dengju. Research on feature selection and classification method based on random forest for medical datasets[D]. Harbin: Harbin Engineering University, 2016.
- [12] Klainin P, Arthur DG. Postpartum depression in Asian cultures: A literature review[J]. Int J Nurs Stud, 2009, 46(10): 1355-1373.
- [13] 钱耀荣, 晏晓颖. 中国产后抑郁发生率的系统分析[J]. 中国实用护理杂志, 2013, 29(12): 1-3.
- QIAN Yaorong, YAN Xiaoying. Prevalence of postpartum depression in China: a systematic analysis[J]. Chinese Journal of Practical Nursing, 2013, 29(12): 1-3.
- [14] 龙周婷. 产后抑郁预测因素的纵向研究[D]. 济南: 山东大学, 2014.
- LONG Zhouting. Predictors of postpartum depression: A longitudinal research[D]. Jinan: Shandong University, 2014.
- [15] 李华丽, 李维玲, 刘冬红. 西安某医院1994名产妇产后抑郁发生的相关因素分析[J]. 中国妇幼健康研究, 2017, 28(11): 1374-1376.
- LI Huali, LI Weiling, LIU Donghong. Related factors of postpartum depression in 1994 cases in a hospital in Xi'an[J]. Chinese Journal of Woman and Child Health Research, 2017, 28(11): 1374-1376.
- [16] 武晓岩, 李康. 随机森林方法在基因表达数据分析中的应用及研究进展[J]. 中国卫生统计, 2009, 26(4): 437-440.
- WU Xiaoyan, LI Kang. Application and research progress of random forest method in gene expression data analysis[J]. Chinese Journal of Health Statistics, 2009, 26(4): 437-440.
- [17] Yim IS, Tanner SL, Guardino CM, et al. Biological and psychosocial predictors of postpartum depression: systematic review and call for integration[J]. Annu Rev Clin Psychol, 2015, 11: 99-137.
- [18] Pampaka D, Papatheodorou SI, Alseaidan M, et al. Postnatal depressive symptoms in women with and without antenatal depressive symptoms: results from a prospective cohort study [J]. Arch Womens Ment Health, 2019, 22(1): 93-103.

(本文编辑 彭敏宁)

本文引用: 肖美丽, 晏春丽, 付冰, 杨淑平, 朱姝娟, 杨东琪, 雷倍美, 黄瑞瑞, 雷俊. 随机森林算法在产后抑郁风险预测中的应用[J]. 中南大学学报(医学版), 2020, 45(10): 1215-1222. DOI: 10.11817/j.issn.1672-7347.2020.190655

Cite this article as: XIAO Meili, YAN Chunli, FU Bing, YANG Shuping, ZHU Shujuan, YANG Dongqi, LEI Beimei, HUANG Ruirui, LEI Jun. Risk prediction for postpartum depression based on random forest[J]. Journal of Central South University. Medical Science, 2020, 45(10): 1215-1222. DOI: 10.11817/j.issn.1672-7347.2020.190655