

基于机器学习算法构建中重度围绝经期综合征预测模型



张 敏, 顾婷婷, 关 蔚, 刘想想, 施君瑶

上海市浦东新区妇幼保健院(中心)妇女群体保健科(上海 201206)

【摘要】目的 通过机器学习算法筛选围绝经期妇女发生围绝经期综合征(perimenopausal syndrome, PMS)的影响因素,构建围绝经期妇女中重度PMS发生风险的预测模型。**方法** 选取上海市浦东新区48个社区的围绝经期妇女作为研究对象,并根据Kupperman评分量表评价结果分为PMS正常或轻度组和中重度组。将数据随机分为训练集和测试集,使用Boruta算法和SHAP算法进行特征选择。使用逻辑回归(Logistic regression, LR)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)和梯度提升决策树(gradient boosting decision tree, GBDT)构建模型,并采用精确度、准确度、召回率、受试者工作特征曲线的曲线下面积(area under curve, AUC)、F1分数评估模型性能。**结果** 共纳入围绝经期妇女856例,其中PMS正常或轻度组557例、中重度组299例;训练集599例、验证集257例。使用Boruta算法以及SHAP算法纳入在职情况、运动、年龄、月经情况、就诊史、超重、居住地、健康教育史、户籍9个特征作为最终模型的预测变量。参数调优后,LR、RF、SVM、GBDT算法基于训练集的10折交叉验证的AUC分别为0.64、0.77、0.74、0.77。LR、RF、SVM、GBDT算法在测试集中的AUC分别为0.63、0.69、0.69、0.73;召回率分别为0.59、0.55、0.55、0.62。**结论** 构建的围绝经期妇女中重度PMS发生风险预测模型中,GBDT算法的预测表现最优,具有一定的预测价值,可为围绝经期妇女发生中重度PMS的早期识别和干预提供新思路和方法。

【关键词】 围绝经期综合征; 围绝经期妇女; 机器学习; 预测

Construction of a prediction model for moderate to severe perimenopausal syndrome based on machine learning algorithms

ZHANG Min, GU Tingting, GUAN Wei, LIU Xiangxiang, SHI Junyao

Department of Women's Health Care, Maternal & Child Health Care Institute of Pudong New Area, Shanghai 201206, China

Corresponding author: SHI Junyao, Email: shjuny622@126.com

【Abstract】Objective To identify risk factors for perimenopausal syndrome (PMS) among perimenopausal women using machine learning algorithms, and to construct a predictive model for the risk of developing moderate to severe PMS in perimenopausal women. **Methods** Perimenopausal women from 48 communities in Pudong New Area, Shanghai were selected as the study subjects. Based on the Kupperman index, participants were divided into

DOI: [10.12173/j.issn.1004-5511.202405116](https://doi.org/10.12173/j.issn.1004-5511.202405116)

基金项目: 上海市浦东新区卫生健康委员会面上项目(PW2020A-79); 上海市浦东新区第二轮医学学科建设项目(PWYgts2021-02)

通信作者: 施君瑶, 主任医师, Email: shjuny622@126.com

the normal or mild PMS group and the moderate to severe PMS group. The data was randomly divided into training set and testing set, and feature selection was performed using the Boruta algorithm and SHAP algorithm. Logistic regression (LR), random forest (RF), support vector machine (SVM), and gradient boosting decision tree (GBDT) were constructed, and model performances were evaluated using accuracy, precision, recall, area under curve(AUC) of the receiver operating characteristic curve, and F1-score. **Results** A total of 856 perimenopausal women were included in the study, of which 557 were in the normal or mild PMS group and 299 were in the moderate to severe PMS group; 599 were in the training set and 257 were in the testing set. 9 features (employment status, exercise, age, menstrual condition, medical history, obesity, residence area, history of health education, household register) were selected as predictors for the final model using the Boruta algorithm and SHAP analysis. After parameter tuning, the 10-fold cross-validation AUC of LR, RF, SVM, and GBDT models based on the training set were 0.64, 0.77, 0.74, and 0.77, respectively. The AUC of the LR, RF, SVM, and GBDT models based on the testing set were 0.63, 0.69, 0.69, and 0.73, respectively, with recall rates of 0.59, 0.55, 0.55, and 0.62. **Conclusion** Among the constructed predictive models for the risk of developing moderate to severe PMS in perimenopausal women, the GBDT model demonstrated the best predictive performance and has potential clinical value. This study provides a new approach for the early identification and intervention of moderate to severe PMS in perimenopausal women.

【Keywords】 Perimenopausal syndrome; Perimenopausal women; Machine learning; Prediction

围绝经期综合征 (perimenopausal syndrome, PMS) 是指女性在围绝经期由于卵巢功能减退和激素水平波动而引发的一系列身心症状, 包括潮热、盗汗以及失眠等^[1-2]。研究表明, 约有 70%~80% 的围绝经期妇女会有不同程度的 PMS 症状, 其中约 20% 的妇女症状较为严重, 影响其日常生活和工作^[3]。针对 PMS 的干预措施主要包括药物治疗、生活方式干预和心理治疗等^[4-6]。然而, 药物治疗可能带来副作用以及增加并发症的风险^[7]; 生活方式干预和心理治疗的长期依从性也较差。

将健康管理的重点从疾病治疗转移到疾病预防和健康促进, 有效预测 PMS 发生风险进而采取措施预防 PMS 的发生是应对 PMS 的有效策略之一^[8]。人工智能, 尤其是机器学习技术的发展, 为解决上述问题提供了新的可能。机器学习通过从大数据中发现潜在的模式和规律的方式对事件进行预测^[9]。例如, 在围绝经妇女人群中, Wang 等^[10]基于健康检查数据开发了能够早期识别代谢综合征高风险人群的机器预测模型; Ali 等^[11]构建了能够预测围绝经期妇女抑郁症的机器学习模

型。然而, 目前尚缺乏使用机器学习模型预测围绝经期妇女 PMS 发生风险的研究。因此, 本研究利用机器学习技术构建围绝经期妇女 PMS 发生风险的预测模型, 以期为围绝经期妇女 PMS 的预防和健康提供管理提供参考。

1 资料与方法

1.1 研究对象

本研究以围绝经期妇女为研究对象。2022 年 7 月至 2023 年 6 月, 对上海市浦东新区辖区内 48 个社区的围绝经期妇女进行调查。纳入标准: ①社区常住妇女; ②年龄 40~60 周岁; ③在 10 个月之内发生两次相邻月经周期长度的变化大于 7 d, 绝经定义为月经停止超过 12 个月^[12]。排除标准: ①有严重器质性疾病; ②有精神疾病或心理障碍; ③临床结局无法判断者; ④正在服用可能影响 PMS 症状的药物者 (如激素替代治疗等); ⑤不能理解调查内容或无法有效沟通者。本研究经上海市浦东新区妇幼保健中心伦理委员会审批 [批号: (2020) 伦理第 (005) 号], 并获得研究对象知情同意书。

1.2 资料收集

在参考相关文献的基础上自行设计问卷内容, 包含: ①一般人口学资料, 包括年龄、居住地(城镇或农村)、户籍、职业(工人、公司职员、国家公务员、科技人员、医务人员、教师、私营业主以及家庭主妇)、教育程度、经济状况、在职情况、是否照看儿童、是否超重、是否不规律运动(规律运动定义为每周至少进行三次中等强度运动, 每次至少 30 min)、是否抽烟(每周吸烟至少一次, 每次至少一根香烟)、是否饮酒(每周饮酒至少一次, 每次饮酒至少一杯)等; ②月经婚育史, 包括是否绝经、婚姻状况; ③围绝经期相关因素, 包括健康教育史(接受关于围绝经期健康的教育或培训)、对更年期态度、就诊史(是否有过针对围绝经期相关症状的就诊记录)等; ④围绝经期症状和 PMS 严重程度。

参考《更年期妇女健康管理专家共识(基层版)》^[13], 使用改良版 Kupperman 评分量表评估 PMS 的严重程度^[14]。该量表评估 13 项症状, 包括潮热出汗、感觉异常、失眠、易激动、抑郁、头晕、疲乏无力、骨关节肌肉痛、头痛、心悸、皮肤蚁走感、性交痛和泌尿系统症状。每个条目评分以症状程度乘以症状指数, 其中潮热出汗的症状指数为 4, 感觉异常、失眠、易激动、性交痛、泌尿系症状的症状指数为 2, 其余症状指数为 1^[15]。总分范围为 0~63 分, > 30 分为重度、16~30 分为中度、6~15 分为轻度以及 < 6 分为正常^[14]。根据评估结果, 将研究对象分为 PMS 正常或轻度组和 PMS 中重度组。

1.3 质量控制

调查场所为上海市浦东新区辖区内 48 个社区的社区卫生服务中心。调查员均为社区卫生服务中心在职医生, 且经过统一培训, 培训内容包括语言规范性和标准、调查的目的和意义、调查问卷的条目解释及填写方式等。通过对在社区卫生服务中心门诊就诊或参加专题讲座的研究对象随机分发调查问卷和知情同意书。问卷以匿名方式, 由调查对象独立完成, 调查员当场答疑、回收、核查。缺失内容超过 20% 的问卷予以剔除。共发放 893 份问卷, 最终回收有效问卷 856 份, 有效回收率为 95.86%。

1.4 统计学分析

采用 Python 3.10.9 软件对数据进行分析。分

类变量使用频数和百分比($n, %$)表示, 组间比较采用 χ^2 检验或 Fisher 精确检验, 多分类变量采用“独热编码”的方式转换为二分类变量, 以更好地适用于机器学习算法。在拟合模型前, 使用众数对缺失率不足 20% 的数据进行缺失值插补, 使用标准化法对数据进行缩放, 使用 SMOTE 算法处理不平衡数据^[16], 并按照 7:3 的比例把数据随机分为训练集和测试集, 使用 Boruta 算法和 SHAP 算法进行特征选择^[17-18]。基于训练集, 使用逻辑回归(Logistic regression, LR)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)和梯度提升决策树(gradient boosting decision tree, GBDT) 4 种算法构建围绝经期妇女发生中重度 PMS 的风险预测模型。使用 Optuna 算法进行超参数选择, 然后基于训练集进行 10 折交叉验证评估预测模型性能, 同时在测试集上验证预测模型的性能。采用受试者工作特征(receiver operating characteristic, ROC)曲线的曲线下面积(area under curve, AUC)、精确度、准确度、召回率、F1 分数等指标评价模型性能。 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 一般情况

纳入 856 名围绝经期妇女, 其中 PMS 正常或轻度组 557 人, PMS 中重度组 299 人。年龄、学历、是否公司职员、在职情况、是否照看儿童、是否超重、月经情况、是否不规律运动、是否饮酒以及就诊史在两组间具有显著差异($P < 0.05$)。

按照 7:3 随机将原始数据集分为训练集和测试集, 其中训练集 599 例(PMS 正常或轻度 393 例, PMS 中重度 206 例), 验证集 257 例(PMS 正常或轻度 164 例, PMS 中重度 93 例)。PMS 发生情况及各特征在两个数据集间无显著差异, 具体见表 1。

2.2 机器学习特征选择

基于 RF 算法, 分别使用 Boruta 算法和 SHAP 算法计算排名前 10 的主要特征, 然后求两种算法的交集。由图 1 可知, 两种算法计算获得的前 10 个重要特征中均包含在职情况、运动、年龄、月经情况、就诊史、超重、居住地、健康教育史、户籍共 9 个特征。因此, 将该 9 个特征作为机器学习算法的输入特征。

表1 不同组别和数据集中围绝经期妇女基线特征情况 (n, %)

Table 1. Baseline characteristics of perimenopausal women in different groups and datasets (n, %)

特征	PMS正常或轻度 (n=557)	PMS中重度 (n=299)	χ^2 值	P值	训练集 (n=599)	验证集 (n=257)	χ^2 值	P值
年龄 (岁)			25.71	<0.01			<0.01	1.00
<50	253 (45.42)	82 (27.42)			234 (39.07)	101 (39.30)		
≥50	304 (54.58)	217 (72.58)			365 (60.93)	156 (60.70)		
城镇居住地	263 (47.22)	149 (49.83)	0.43	0.51	297 (49.58)	115 (44.75)	1.50	0.22
上海户籍	441 (79.17)	246 (82.27)	0.99	0.32	480 (80.13)	207 (80.54)	<0.01	0.96
学历			6.29	0.01			0.16	0.68
大专及以上	223 (40.04)	93 (31.10)			218 (36.39)	98 (38.13)		
高中及以下	334 (59.96)	206 (68.90)			381 (63.61)	159 (61.87)		
职业 [#]								
工人	91 (16.34)	45 (15.05)	0.15	0.69	96 (16.03)	40 (15.56)	<0.01	0.95
公司职员	107 (19.21)	38 (12.71)	5.39	0.02	108 (18.03)	37 (14.40)	1.44	0.23
国家公务员	16 (2.87)	6 (2.01)	0.29	0.59	14 (2.34)	8 (3.11)	0.18	0.67
科技人员	7 (1.26)	1 (0.33)		0.27 [*]	8 (1.34)	0 (0.00)		0.11 [*]
医务人员	75 (13.46)	27 (9.03)	3.24	0.07	68 (11.35)	34 (13.23)	0.44	0.51
教师	24 (4.31)	17 (5.69)	0.54	0.46	28 (4.67)	13 (5.06)	<0.01	0.95
私营业主	10 (1.80)	8 (2.68)	0.37	0.54	13 (2.17)	5 (1.95)	<0.01	1.00
家庭主妇	72 (12.93)	49 (16.39)	1.65	0.20	88 (14.69)	33 (12.84)	0.37	0.54
在职情况			25.47	<0.01			0.51	0.47
在职	365 (65.53)	142 (47.49)			360 (60.10)	147 (57.20)		
离退休/下岗	192 (34.47)	157 (52.51)			239 (39.90)	110 (42.80)		
月收入情况 (元)			1.78	0.18			0.17	0.68
≤2000	97 (17.41)	64 (21.40)			110 (18.36)	51 (19.84)		
>2000	460 (82.59)	235 (78.60)			489 (81.64)	206 (80.16)		
婚姻情况			1.00	0.32			0.08	0.78
无配偶	30 (5.39)	22 (7.36)			35 (5.84)	17 (6.61)		
有配偶	527 (94.61)	277 (92.64)			564 (94.16)	240 (93.39)		
儿童照看	132 (23.70)	94 (31.44)	5.61	0.02	150 (25.04)	76 (29.57)	1.67	0.20
超重	136 (24.42)	94 (31.44)	4.53	0.03	159 (26.54)	71 (27.63)	0.06	0.81
月经情况			17.41	<0.01			<0.01	0.98
未绝经	312 (56.01)	122 (40.80)			303 (50.58)	131 (50.97)		
已绝经	245 (43.99)	177 (59.20)			296 (49.42)	126 (49.03)		
不规律运动	343 (61.58)	209 (69.90)	5.52	0.02	390 (65.11)	162 (63.04)	0.25	0.61
抽烟	1 (0.18)	3 (1.00)		0.13 [*]	2 (0.33)	2 (0.78)		0.59 [*]
饮酒	2 (0.36)	6 (2.01)	4.06	0.04	6 (1.00)	2 (0.78)	<0.01	1.00
健康教育史	391 (70.20)	217 (72.58)	0.43	0.51	432 (72.12)	176 (68.48)	0.99	0.32
对更年期态度			3.37	0.07			0.83	0.36
积极	474 (85.10)	239 (79.93)			504 (84.14)	209 (81.32)		
不积极	83 (14.90)	60 (20.07)			95 (15.86)	48 (18.68)		
就诊史	382 (68.58)	179 (59.87)	6.16	0.01	397 (66.28)	164 (63.81)	0.38	0.54

注：[#]职业为多分类变量，经独热编码处理；^{*}组间比较采用Fisher精确检验。

2.3 模型超参数调优

基于训练集，采用10折交叉验证分别在默认超参数和使用Optuna算法调优情况下，将筛选

的9个特征作为输入特征，以AUC为评价指标评价模型性能。结果显示，LR、RF、SVM和GBDT在超参数调优前的AUC分别为0.64、0.75、0.71

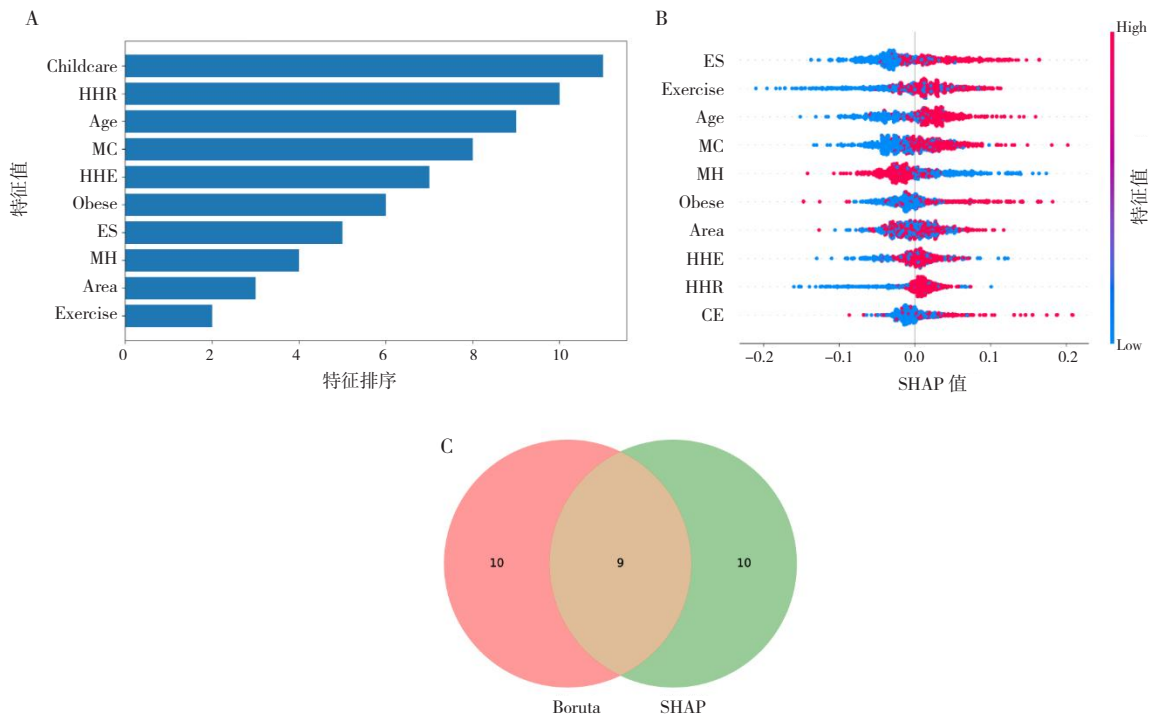


图1 Boruta算法与SHAP算法特征选择

Figure 1. Feature selection based on Boruta algorithm and SHAP algorithm

注：A. 基于Boruta算法的特征重要性排名前10的特征（条形图越短，排名越靠前）；B. 基于SHAP算法的特征重要性排名前10的特征；C. Boruta算法与SHAP算法获得的前10个重要特征的交集；ES. 在职情况；Exercise. 运动；Age. 年龄；MC. 月经情况；MH. 就诊史；Obese. 超重；Area. 居住地；HHE. 健康教育史；HHR. 户籍；Childcare. 照看儿童；CE. 公司职员。

和0.76，在参数调优后的AUC分别为0.64、0.77、0.74和0.77。参数调优后，除LR外，其他模型的AUC均有所提升，见图2。

2.4 模型验证

经模型调优确认最终的超参数后，基于测试集对模型性能进行评估。LR、RF、SVM和GBDT在

测试集中的AUC分别为0.63、0.69、0.69和0.73（图3-A）；4种模型的召回率分别为0.59、0.55、0.55以及0.62（图3-B）。各模型基于准确度、精确度以及F1分数的表现，具体见表2。其中GBDT模型的综合表现最好，其基于测试集的准确度、精确度以及F1分数分别为0.67、0.56以及0.59。

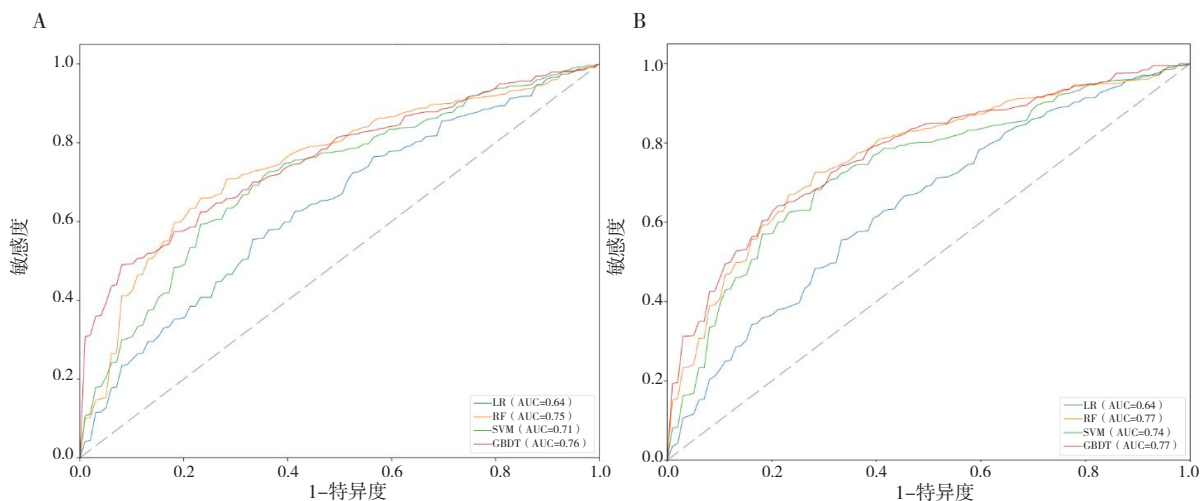


图2 各模型10折交叉验证的ROC曲线

Figure 2. ROC curves of ten-fold cross validation for each model

注：A. 默认超参数下各模型的10折交叉验证的ROC曲线；B. 调参后的各模型10折交叉验证ROC曲线；LR. 逻辑回归；RF. 随机森林；SVM. 支持向量机；GBDT. 梯度提升决策树。

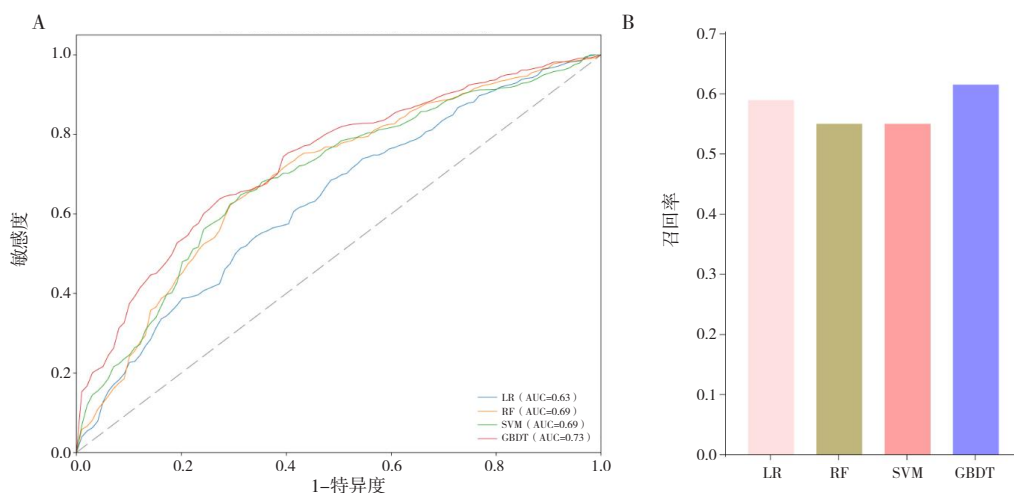


图3 基于独立测试集的模型性能评估

Figure 3. Model performance evaluation based on independent test sets

注：A. 基于测试集的各模型性能验证ROC曲线；B. 基于测试集的各模型的召回率；LR. 逻辑回归；RF. 随机森林；SVM. 支持向量机；GBDT. 梯度提升决策树。

表2 各模型评价指标结果

Table 2. Results of evaluation indicators for each model

模型	准确度	精确度	召回率	AUC	F1 分数
LR	0.61	0.47	0.59	0.63	0.52
RF	0.63	0.49	0.55	0.69	0.52
SVM	0.61	0.46	0.55	0.69	0.50
GBDT	0.67	0.56	0.62	0.73	0.59

注：LR. 逻辑回归；RF. 随机森林；SVM. 支持向量机；GBDT. 梯度提升决策树。

3 讨论

本研究利用机器学习算法，基于上海市浦东新区围绝经期妇女的流行病学调查数据，构建了中重度 PMS 发生风险的预测模型。研究结果表明，经过超参数调优后，相较于其他 3 种模型，GBDT 算法的预测表现最优，AUC 为 0.73，召回率和 F1 分数分别为 0.62 和 0.59，准确度为 0.67，具有较好的综合性能，模型对阳性结果的识别度有显著提升。

以往关于 PMS 的研究主要集中在其发病机制、症状评估及治疗措施等方面^[19-22]，缺乏对 PMS 发生风险预测的研究。本研究基于机器学习模型提供了一种新的预测中重度 PMS 的方法，为围绝经期妇女的健康管理提供了新思路。此外，通过预测模型可以在社区保健中心或妇幼保健院进行大规模筛查，及时发现高风险个体并提供相应的预防和治疗措施。这种模式在公共卫生管理中尤为重要，可以减少医疗成本，提高整体健康

水平。现有研究的诊断工具只能判断当前，否患有中重度 PMS，无法预测未来患病风险，而机器学习模型可以根据既往数据预测未来患病可能性，对高危人群提前进行健康干预，从而实现“治未病”的目标^[23]。在 PMS 普查中，可以先使用诊断工具筛查当前患者，再用机器学习模型预测高危人群，或先用模型筛查高危对象，再用诊断工具确认当前病情，从而提高筛查效率。

本研究综合了两种特征重要性算法，对纳入机器学习的特征进行了筛选。最终，在职情况、运动、年龄、月经情况、就诊史、超重、居住地、健康教育史以及户籍等 9 个特征被纳入模型。上述大部分特征在以往的研究中已被证实与 PMS 发生风险相关。例如，年龄是影响 PMS 发生的重要因素，随着年龄增长，PMS 的发生率逐渐升高；超重与内分泌紊乱、炎症反应等密切相关，也是 PMS 的危险因素^[12, 24]，在本研究中，无论是基于单因素分析还是基于 SHAP 算法的结果，都支持上述结论。此外，刘红雪等^[25]的研究表明，绝经

后 PMS 的患病率显著高于正常月经和绝经过渡期的患病率, 并且 PMS 的患病率在受教育情况、职业现状方面的差异具有统计学意义, 本研究的结果也具有一致性; 然而, 该研究指出郊区女性患病率显著高于城区, 而本研究结果显示, 居住地不同(城镇/郊区)的研究对象, 中重度 PMS 的患病率无显著差异, 这种结果的差异或可归因于地区间发展的不平衡。另外, 已有系列研究表明运动对于围绝经期妇女的健康具有积极意义。比如, 方梦辰等^[26]的研究表明中低强度的有氧运动可对围绝经期和绝经后妇女心率变异性产生积极影响; 也有研究表明增加体育运动可降低或缓解 PMS 的发生^[27-28], 该结论也和本研究具有一致性。除上述报道较多的 PMS 影响因素外, 本研究还纳入了文献较少出现的“就诊史”以及“户籍”两个因素。综上, 本研究纳入的中重度 PMS 相关影响因素, 基本与文献报道一致, 进一步提升了模型的可解释性, 保证了模型的可靠性。

本研究仍存在一定局限性。一是样本量限制了模型的学习能力; 二是特征选择存在不足, 未能充分挖掘影响中重度 PMS 发生的所有潜在因素; 三是模型选择和参数调优有待优化; 四是研究对象的招募主要集中在社区卫生服务中心, 研究对象的代表性可能不足。以上因素导致本研究模型的 AUC、召回率等指标虽然在可接受范围内, 但仍有较大提升空间。未来仍需扩大样本量和样本来源, 探索更高效的算法以改善模型性能以及扩大模型的适用场景。

综上所述, 本研究构建了基于机器学习算法的中重度 PMS 发生风险的预测模型, 其中 GBDT 算法的预测表现最优, 具有一定的预测价值, 可为围绝经期妇女中重度 PMS 发生风险的早期识别和早期干预提供新的思路和方法。

参考文献

- 1 郭丽, 刘春, 巩爱玲. 更年期综合征的研究进展 [J]. 中国当代医药, 2019, 26(33): 21-23. [Guo L, Liu C, Gong AL. Research progress of climacteric syndrome[J]. China Modern Medicine, 2019, 26(33): 21-23.] DOI: [10.3969/j.issn.1674-4721.2019.33.007](https://doi.org/10.3969/j.issn.1674-4721.2019.33.007).
- 2 Zeng M, Li M, Li M, et al. 2-Phenylacetamide isolated from the seeds of lepidium apetalum and its estrogen-like effects in vitro and in vivo[J]. *Molecules*, 2018, 23(9): 2293. DOI: [10.3390/molecules23092293](https://doi.org/10.3390/molecules23092293).
- 3 Xu HW, Du W, He L, et al. Effectiveness and safety of warm needle acupuncture on insomnia in climacteric women: protocol for a systematic review and Meta-analysis[J]. *Medicine (Baltimore)*, 2019, 98(20): e15637. DOI: [10.1097/MD.00000000000015637](https://doi.org/10.1097/MD.00000000000015637).
- 4 Xue B, Chen X, Wang X, et al. Application of multivariate statistical analysis and network pharmacology to explore the mechanism of Danggui Liuhuang Tang in treating perimenopausal syndrome[J]. *J Ethnopharmacol*, 2022, 284: 114543. DOI: [10.1016/j.jep.2021.114543](https://doi.org/10.1016/j.jep.2021.114543).
- 5 Kargozar R, Azizi H, Salari R. A review of effective herbal medicines in controlling menopausal symptoms[J]. *Electron Physician*, 2017, 9(11): 5826-5833. DOI: [10.19082/5826](https://doi.org/10.19082/5826).
- 6 冯华. 健康教育对围绝经期综合征妇女的生活质量和心理状态的影响 [J]. 世界最新医学信息文摘(连续型电子期刊), 2020, 20(41): 248, 250. [Feng H. The impact of health education on the quality of life and psychological status of women with perimenopausal syndrome[J]. *World Latest Medicine Information (Electronic version)*, 2020, 20(41): 248, 250.] DOI: [10.13429/j.cnki.cjcr.2015.06.048](https://doi.org/10.13429/j.cnki.cjcr.2015.06.048).
- 7 Qin XY, Niu ZC, Han XL, et al. Anti-perimenopausal osteoporosis effects of Erzhi formula via regulation of bone resorption through osteoclast differentiation: a network pharmacology-integrated experimental study[J]. *J Ethnopharmacol*, 2021, 270: 113815. DOI: [10.1016/j.jep.2021.113815](https://doi.org/10.1016/j.jep.2021.113815).
- 8 李文先, 曹秀菁. 中国妇幼保健的进展与成效 [J]. 中华疾病控制杂志, 2022, 26(9): 993-995, 1107. [Li WX, Cao XJ. Progress and effects of women and children's health in China[J]. *Chinese Journal of Disease Control & Prevention*, 2022, 26(9): 993-995, 1107.] DOI: [10.16462/j.cnki.zhjbkz.2022.09.001](https://doi.org/10.16462/j.cnki.zhjbkz.2022.09.001).
- 9 Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects[J]. *Science*, 2015, 349(6245): 255-260. DOI: [10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415).
- 10 Wang XX, Wang ZJ, Chen SC, et al. Risk prediction model of metabolic syndrome in perimenopausal women based on machine learning[J]. *Int J Med Inform*, 2024, 188: 105480. DOI: [10.1016/j.ijmedinf.2024.105480](https://doi.org/10.1016/j.ijmedinf.2024.105480).
- 11 Ali MM, Algashamy HAA, Alzidi E, et al. Development and performance analysis of machine learning methods for predicting depression among menopausal women[J].

- Healthcare Analytics, 2023, 3: 100202. DOI: [10.1016/j.health.2023.100202](https://doi.org/10.1016/j.health.2023.100202).
- 12 刘成伟, 徐黎贤, 朱其舟, 等. 南昌地区围绝经期女性健康状况与需求调查[J]. 江西医药, 2020, 55(7): 804–806. [Liu CW, Xu LX, Zhu QZ, et al. Health status and demand of perimenopausal women in Nanchang[J]. Jiangxi Medical Journal, 2020, 55(7): 804–806.] DOI: [10.3969/j.issn.1006-2238.2020.07.003](https://doi.org/10.3969/j.issn.1006-2238.2020.07.003).
- 13 中国医师协会全科医师分会, 北京妇产学会社区与基层分会. 更年期妇女健康管理专家共识(基层版)[J]. 中国全科医学, 2021, 24(11): 1317–1324. [Chinese Medical Doctor Association (CMDA)'s General Practitioners Sub-association, The Primary Care Branch of Beijing Institute of Obstetrics & Gynecology. Consensus on health management of climacteric women in primary medical institutions edition[J]. Chinese General Practice, 2021, 24(11): 1317–1324.] DOI: [10.12114/j.issn.1007-9572.2021.00.402](https://doi.org/10.12114/j.issn.1007-9572.2021.00.402).
- 14 Tao MF, Shao HF, Li CB, et al. Correlation between the modified Kupperman index and the menopause rating scale in Chinese women[J]. Patient Preference and Adherence, 2013; 223–229. DOI: [10.2147/PPA.S42852](https://doi.org/10.2147/PPA.S42852).
- 15 国家药品监督管理局. 关于印发中药、天然药物治疗冠心病心绞痛和女性更年期综合征临床研究技术指导原则的通知[EB/OL]. (2011-07-08) [2024-08-01]. https://www.nmpa.gov.cn/directory/web/nmpa/xxgk/fgwj/gzwy/gzwyjyp/20110708113101528_13.html.
- 16 Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 341–378. DOI: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- 17 Kursa MB, Jankowski A, Rudnicki WR. Boruta—a system for feature selection[J]. Fundam Informatiae, 2010, 101: 271–285. DOI: [10.3233/FI-2010-288](https://doi.org/10.3233/FI-2010-288).
- 18 Fryer D, Strümke I, Nguyen H. Shapley values for feature selection: the good, the bad, and the axioms[J]. IEEE Access, 2021, 9: 144352–144360. DOI: [10.1109/ACCESS.2021.3119110](https://doi.org/10.1109/ACCESS.2021.3119110).
- 19 张倩, 刘筱茂, 张梅奎. 围绝经期失眠的发生机制与治疗进展[J]. 解放军医学院学报, 2023, 44(9): 1056–1059. [Zhang Q, Liu XM, Zhang MK. Pathogenesis and treatment progress of perimenopausal insomnia[J]. Academic Journal of Chinese PLA Medical School, 2023, 44(9): 1056–1059.] DOI: [10.12435/j.issn.2095-5227.2023.049](https://doi.org/10.12435/j.issn.2095-5227.2023.049).
- 20 胡波, 汤洪, 张新平. 围绝经期女性抑郁情绪与围绝经期症状的相关性研究[J]. 中国妇幼保健, 2020, 35(21): 4042–4045. [Hu B, Tang H, Zhang XP. Study on the correlation between depression and menopausal symptoms in menopausal women[J]. Maternal and Child Health Care of China, 2020, 35(21): 4042–4045.] DOI: [10.19829/j.zgfybj.issn.1001-4411.2020.21.045](https://doi.org/10.19829/j.zgfybj.issn.1001-4411.2020.21.045).
- 21 王洪雪, 王杰琼, 高杰. 中药治疗围绝经期综合征的机制研究[J]. 世界科学技术: 中医药现代化, 2020, 22(6): 1822–1827. [Wang HX, Wang JQ, Gao J. Study on the mechanism of Chinese medicine in treating perimenopausal syndrome[J]. Modernization of Traditional Chinese Medicine and Materia Medica—World Science and Technology, 2020, 22(6): 1822–1827.] DOI: [10.11842/wst.20191010005](https://doi.org/10.11842/wst.20191010005).
- 22 Modzelewski S, Oracz A, Żukow X, et al. Premenstrual syndrome: new insights into etiology and review of treatment methods[J]. Front Psychiatry, 2024, 15: 1363875. DOI: [10.3389/fpsy.2024.1363875](https://doi.org/10.3389/fpsy.2024.1363875).
- 23 叶萍萍. 运用“治未病”思想防治妇科病症经验举隅[J]. 临床合理用药杂志, 2022, 15(8): 171–173, 177. [Ye PP. Examples of applying the concept of "treating before illness" to prevent and treat gynecological conditions[J]. Chinese Journal of Clinical Rational Drug Use, 2022, 15(8): 171–173, 177.] DOI: [10.15887/j.cnki.13-1389/r.2022.08.053](https://doi.org/10.15887/j.cnki.13-1389/r.2022.08.053).
- 24 张竞, 阿地拉·阿里木, 郭艳英, 等. 乌鲁木齐市天山区维吾尔族女性围绝经期和绝经后代谢综合征现状及危险因素分析[J]. 中华全科医学, 2023, 21(1): 73–76. [Zhang J, Adila A, Guo YY, et al. Prevalence and risk factors of metabolic syndrome among pre-menopausal and post-menopausal women of Uygur in Tianshan district of Urumqi[J]. Chinese Journal of General Practice, 2023, 21(1): 73–76.] DOI: [10.16766/j.cnki.issn.1674-4152.002815](https://doi.org/10.16766/j.cnki.issn.1674-4152.002815).
- 25 刘红雪, 郑建华, 郁琦. 哈尔滨地区围绝经期女性健康现状及保健需求分析[J]. 现代生物医学进展, 2012, 12(21): 4065–4068, 4078. [Liu HX, Zheng JH, Yu Q. Analysis of healthy conditions and healthcare service demand of perimenopausal women in Harbin[J]. Progress in Modern Biomedicine, 2012, 12(21): 4065–4068, 4078.] DOI: [10.13241/j.cnki.pmb.2012.21.026](https://doi.org/10.13241/j.cnki.pmb.2012.21.026).
- 26 方梦辰, 张培珍. 运动对围绝经期和绝经后妇女心率变

- 异性的调节[J]. 中南大学学报(医学版), 2024, 49(4): 516-525. [Fang MC, Zhang PZ. Regulation of exercise on heart rate variability in perimenopausal and postmenopausal women[J]. Journal of Central South University (Medical Science), 2024, 49(4): 516-525.] DOI: [10.11817/j.issn.1672-7347.2024.230399](https://doi.org/10.11817/j.issn.1672-7347.2024.230399).
- 27 江雪芳, 田丰莲, 黄晓晖, 等. 围绝经期妇女运动对预防围绝经期综合征的影响[J]. 中国临床医学, 2004, 11(6): 1091-1092. [Jiang XF, Tian FL, Huang XH, et al. The study of the relationship between athletics and perimenopausal syndrome in Guangdong[J]. Chinese Journal of Clinical Medicine, 2004, 11(6): 1091-1092.] DOI: [10.3969/j.issn.1008-6358.2004.06.064](https://doi.org/10.3969/j.issn.1008-6358.2004.06.064).
- 28 李青, 王艳娜, 杨娜, 等. 运动对女性围绝经期综合征影响的系统评价[J]. 中国老年学杂志, 2017, 37(19): 4875-4878. [Li Q, Wang YN, Yang N, et al. A systematic review of the effects of exercise on menopausal symptoms in women[J]. Chinese Journal of Gerontology, 2017, 37(19): 4875-4878.] DOI: [10.3969/j.issn.1005-9202.2017.19.085](https://doi.org/10.3969/j.issn.1005-9202.2017.19.085).
- 收稿日期: 2024 年 05 月 28 日 修回日期: 2024 年 07 月 01 日
本文编辑: 李绪辉 曹越

引用本文: 张敏, 顾婷婷, 关蔚, 等. 基于机器学习算法构建中重度围绝经期综合征预测模型[J]. 医学新知, 2024, 34(8): 871-879. DOI: [10.12173/j.issn.1004-5511.202405116](https://doi.org/10.12173/j.issn.1004-5511.202405116)
Zhang M, Gu TT, Guan W, et al. Construction of a prediction model for moderate to severe perimenopausal syndrome based on machine learning algorithms[J]. Yixue Xinzhi Zazhi, 2024, 34(8): 871-879. DOI: [10.12173/j.issn.1004-5511.202405116](https://doi.org/10.12173/j.issn.1004-5511.202405116)