

doi:10.3969/j.issn.1005-3697.2025.10.021

✧ 临床医学研究 ✧

基于血清维生素 D 水平和临床特征构建机器学习模型预测子宫肌瘤发生风险的临床研究

方春玲¹, 唐能欢², 胡莉¹, 胡辉权¹, 徐凡^{1,3}

(1. 川北医学院附属南充市中心医院妇科; 2. 川北医学院附属医院妇科, 四川 南充 637000; 3. 上海市第一妇婴保健院妇科, 上海 201204)

【摘要】目的: 开发和验证一种基于血清维生素 D 水平和其他临床特征评估子宫肌瘤发生风险的临床预测模型。**方法:** 回顾性分析 304 例子宫平滑肌瘤患者和 184 名正常人的相关临床资料, 根据收集资料的不同时间段, 分为训练集($n=192$)和验证集($n=296$)。通过血清维生素 D 水平和其他五个临床特征构建 5 种机器学习(ML)预测模型。曲线下面积(AUC)评价模型的效能, 并通过外部验证进行模型验证。**结果:** 五个机器学习模型都表现出较好的预测效能($AUC \geq 0.8$)。在模型的外部验证中, 相比于其他的机器学习算法, 支持向量机(SVM)($AUC=0.941$, $95\%CI: 0.927 \sim 0.969$, 敏感度为 0.968, 特异度为 0.801, Brier 评分为 0.148)展现出了更好的预测效能。**结论:** 血清维生素 D 水平与其他五种临床特征相结合展现出良好预测子宫肌瘤发生的潜力, 而支持向量机(SVM)模型能展现出更好的预测效能。

【关键词】 子宫肌瘤; 机器学习; 血清维生素 D 水平; 临床特征

【中图分类号】 R737.33

【文献标志码】 A

Clinical study on predicting the risk of uterine fibroids using machine learning models based on serum vitamin D levels and clinical characteristics

FANG Chun-ling¹, TANG Neng-huan², HU Li¹, HU Hui-quan¹, XU Fan^{1,3}

(1. Department of Gynecology, Affiliated Nanchong Central Hospital of North Sichuan Medical College; 2. Department of Gynecology, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, Sichuan; 3. Department of Gynecology, Shanghai First Maternity and Infant Hospital, Shanghai 201204, China)

【Abstract】Objective: To develop and validate a clinical prediction model for assessing the risk of uterine fibroid occurrence based on serum vitamin D levels and other clinical characteristics. **Methods:** A retrospective analysis was conducted on 304 patients with uterine leiomyomas and 184 healthy controls. They were divided into a training set ($n=192$) and a validation set ($n=296$) according to different time periods of data collection. Five machine learning (ML) predictive models were developed based on serum vitamin D levels and five other clinical features. The area under the curve (AUC) was used to evaluate the performance of the models, and external validation was performed to verify the models. **Results:** All five machine learning models demonstrated good predictive performance ($AUC \geq 0.8$). In external validation, the support vector machine (SVM) model exhibited superior predictive performance compared to other machine learning algorithms ($AUC=0.941$, $95\%CI: 0.927 \sim 0.969$, sensitivity=0.968, specificity=0.801, Brier score=0.148). **Conclusion:** The combination of serum vitamin D levels with five other clinical features shows promising potential for predicting the occurrence of uterine fibroids. Among the models tested, the support vector machine (SVM) model demonstrated the best predictive performance.

【Key words】 Uterine fibroids; Machine learning; Serum vitamin D levels; Clinical features

子宫肌瘤是妇科领域最常见的良性肿瘤, 其发病率因人群差异较大, 为 $5\% \sim 70\%$ ^[1-2]。患者常表现出性交疼痛、异常子宫出血、不孕症、盆腔疼痛、流产及早产等一系列临床症状^[3], 这些症状不仅降

低患者的生活质量, 也给医疗系统带来了巨大的经济压力^[4]。目前, 超声检查是诊断子宫肌瘤的主要手段, 但由于早期症状不明显, 早期诊断仍存在较大困难。近年来研究^[5]发现, 维生素 D 水平在评估子

基金项目: 川北医学院校级科研项目(CBY20-QA-Z06); 重庆市渝中区科技项目(20202723)

作者简介: 方春玲(1997-), 女, 硕士, 住院医师。E-mail: 2568415032@qq.com

通讯作者: 徐凡。E-mail: xufanyidingxin@163.com

子宫肌瘤发生风险方面具有潜在价值。与此同时,随着人工智能技术的快速发展,机器学习算法在医学领域的应用日益广泛,其通过自主学习和优化经验的能力,为疾病预测和研究提供了新的技术路径^[6]。利用机器学习方法预测子宫肌瘤的发生风险,将有助于制定更精准的子宫肌瘤二级预防策略。基于此,本研究旨在通过分析血清维生素 D 水平及其他临床特征,构建机器学习预测模型,以评估子宫肌瘤的发生风险。

1 资料与方法

1.1 一般资料

选取 2021 年 10 月至 2022 年 6 月南充市中心医院收治的 133 例子宫肌瘤患者及同期 59 名健康体检志愿者作为训练集;同时纳入 2022 年 8 月至 2024 年 3 月南充市中心医院妇产科门诊确诊的 171 例子宫肌瘤患者及同期 125 名健康体检志愿者作为验证集。在本研究中,训练集包含 192 例,其中无子宫肌瘤(UF=0)59 例(30.7%),有子宫肌瘤(UF=1)133 例(69.3%);验证集包含 296 例,其中无子宫肌瘤(UF=0)125 例(42.2%),有子宫肌瘤(UF=1)171 例(57.8%)。所有受试者数据均用于机器学习模型的开发与验证。纳入标准:(1)无症状子宫肌瘤患者:①诊断为子宫肌瘤的绝经前妇女;②经阴道或腹部超声检查确诊为子宫肌瘤但无明显症状。(2)健康对照者:①无子宫肌瘤或其他与维生素 D 相关疾病的绝经前妇女;②超声学检查证实无子宫肌瘤患者。排除标准:(1)有症状的子宫肌瘤患者:①绝经妇女;②曾接受过治疗或既往诊断为子宫肌瘤的患者;③有症状的子宫肌瘤。(2)健康对照者:①绝经妇女;②先前诊断为子宫肌瘤或其他与维生素 D 相关疾病的患者;③数据缺失或并发其他肿瘤的患者。本研究通过南充市中心医院伦理委员会审批(2022076),所有参与者均知情同意。

1.2 检测方法

在本研究中,所有研究对象的外周血样本均采用标准化真空采血管进行采集。血清维生素 D 水平的检测使用罗氏 Elecsys 2010 全自动电化学发光分析仪,配套专用的 25-羟基维生素 D 检测试剂盒,所有检测步骤均严格遵循制造商提供的标准操作程序(SOP)进行,以确保检测结果的准确性和可靠性。

1.3 变量筛选及机器学习模型构建与验证

本研究通过 3 位具有丰富工作经验的妇产科医生和 1 名从事量表开发的统计学教授共同选择了 6 个变量,包括年龄、BMI、妊娠次数、产次、血清 LDH

水平以及血清维生素 D 水平。采用 R 软件构建机器学习预测模型,系统比较五种经典机器学习算法的性能和稳定性,包括逻辑回归(Logistic regression,LR)、支持向量机(support vector machine,SVM)、K 近邻(K-nearest neighbors,KNN)、朴素贝叶斯(naive bayes,NB)和人工神经网络(artificial neural network,ANN)。为全面评估模型的预测效能,我们计算了多项关键指标,包括受试者工作特征(ROC)曲线下面积(AUC)、敏感度、特异度及 Brier 评分(brier score,BS)。此外,为确保模型的可靠性和泛化能力,进一步采用外部验证法对模型的稳定性进行了严格评估。

1.4 统计学分析

本研究采用 SPSS 26.0 和 R studio 4.3.2 进行数据统计分析。分类变量用 $[n(\%)]$ 进行描述,组间比较用独立样本 χ^2 检验;符合正态分布的连续变量以 $(\bar{x}\pm s)$ 表示,组间比较用独立样本 t 检验。模型构建则采用 R studio 4.3.2 进行构建,并进行外部数据验证。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 一般资料

本研究共纳入 488 例研究对象,其中训练集 192 例,验证集 296 例。训练集和验证集两组研究对象在年龄(40.8 ± 8.7 岁 *vs.* 40.2 ± 8.9 岁)、BMI (23.7 ± 3.0 kg/m² *vs.* 23.5 ± 3.3 kg/m²)、妊娠次数(2.9 ± 1.5 次 *vs.* 2.78 ± 1.9 次)、产次(1.4 ± 0.8 次 *vs.* 1.4 ± 0.8 次)、血清维生素 D 水平(14.2 ± 5.4 nmol/L)及乳酸脱氢酶水平(170.4 ± 29.5 U/L *vs.* 167.2 ± 29.2 U/L)方面均未表现出显著差异($P>0.05$)。在子宫肌瘤患病情况方面,训练组的患病率(69.3%)高于验证组(57.8%),差异具有统计学意义($P=0.010$)。见表 1。

表 1 训练集与验证集两组研究对象的临床特征 $[\bar{x}\pm s,n(\%)]$

临床特点	训练集($n=192$)	验证集($n=296$)	t/χ^2 值	P 值
年龄(岁)	40.8± 8.7	40.2± 8.9	0.746	0.456
BMI(kg/m ²)	23.7± 3.0	23.5± 3.3	0.449	0.653
妊娠次数(次)	2.9± 1.5	2.78± 1.9	1.051	0.294
产次(次)	1.4± 0.8	1.4± 0.8	1.074	0.283
血清维生素 D 水平(nmol/L)	13.2± 5.4	14.2± 5.4	1.862	0.063
乳酸脱氢酶(U/L)	170.4±29.5	167.2±29.2	1.195	0.232
子宫肌瘤(UF)			6.558	0.010
无(UF=0)	59(30.7)	125(42.2)		
有(UF=1)	133(69.3)	171(57.8)		

2.2 基于维生素 D 水平和其他五个临床特征构建和验证机器学习预测模型

本研究显示,支持向量机(SVM)、K 近邻(KNN)和人工神经网络(ANN)三种预测模型展现出卓越的预测性能,其受试者工作特征曲线下面积(AUC)均达到 0.90 以上,优于其他模型。相比之下,逻辑回归(LR)和朴素贝叶斯(NB)模型的预测效能相对较低,其 AUC 为 0.894 和 0.869。见表 2。

表 2 基于维生素 D 水平和训练集上其他五个临床特征的五种机器学习预测模型

模型	组别	AUC 值	95%CI	敏感度	特异度	Brier 评分
LR	训练集	0.894	0.850~0.938	0.898	0.729	0.122
SVM	训练集	0.938	0.902~0.972	0.830	0.902	0.093
KNN	训练集	0.928	0.895~0.961	0.915	0.789	0.098
NB	训练集	0.869	0.817~0.921	0.847	0.751	0.135
ANN	训练集	0.901	0.856~0.946	0.796	0.872	0.107

LR:逻辑回归;SVM:支持向量机;KNN:K-近邻;NB:朴素贝叶斯;ANN:人工神经网络。

2.3 外部验证评估模型预测性能

本研究整合血清维生素 D 水平及其他五个关键临床特征,系统构建了基于逻辑回归(LR)、支持向量机(SVM)、K 近邻(KNN)、朴素贝叶斯(NB)和人工神经网络(ANN)五种机器学习算法的预测模型。通过严格的外部验证评估各模型的预测性能,结果显示,SVM 模型展现出优异的预测效能,其外部验证 AUC 为 0.941,同时敏感度为 0.968,特异度为 0.801,BRIER 评分为 0.148。见表 3 及图 1。

2.4 各预测变量重要性分析

基于支持向量机(SVM)模型的各变量重要性排序结果,其算法采用 SVM 排列重要性算法进行分析。条形图显示,在所有变量中,年龄、维生素 D 水平以及产次是影响子宫肌瘤发生的主要因素,其中年龄的重要性最高,其次是维生素 D 水平和产次。见图 2。

表 3 外部验证模型性能

模型	组别	AUC 值	95%CI	敏感度	特异度	BRIER 评分
LR	验证集	0.862	0.821~0.903	0.624	0.883	0.158
SVM	验证集	0.941	0.927~0.969	0.968	0.801	0.148
KNN	验证集	0.856	0.814~0.898	0.648	0.894	0.152
NB	验证集	0.869	0.817~0.921	0.304	0.953	0.148
ANN	验证集	0.846	0.803~0.890	0.904	0.701	0.212

LR:逻辑回归;SVM:支持向量机;KNN:K-近邻;NB:朴素贝叶斯;ANN:人工神经网络。

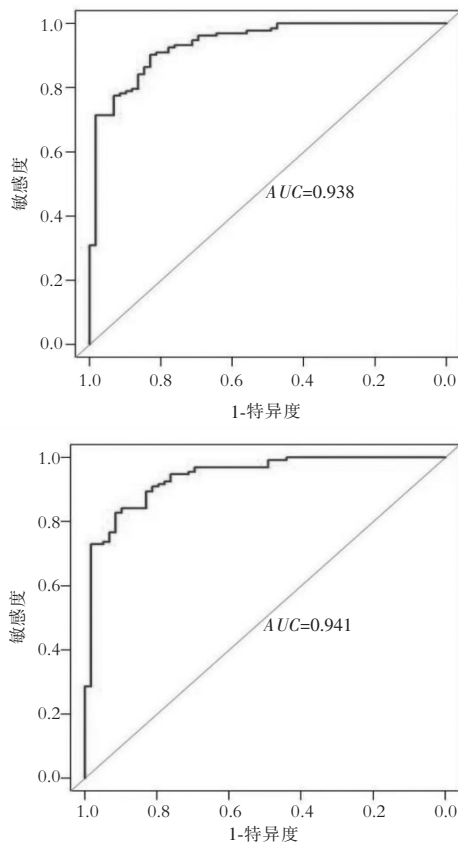


图 1 SVM(支持向量机)训练集、验证集 ROC 曲线

A. 训练集的 ROC 曲线;B. 验证集的 ROC 曲线。

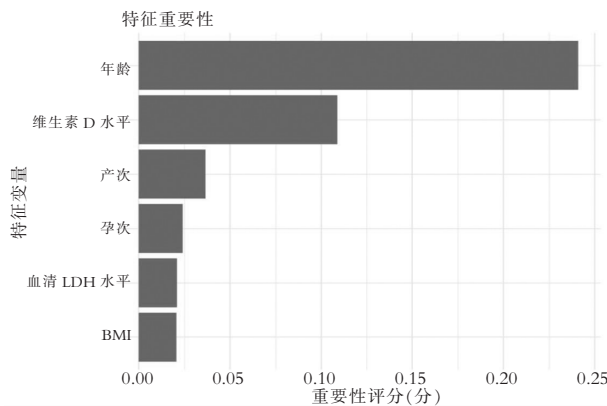


图 2 纳入各变量的重要性评分

3 讨论

子宫肌瘤是女性生殖系统中最常见的良性肿瘤。临床流行病学研究表明^[2],70%~80%的患者无明显临床症状,仅 20%~30%的患者会表现出不同程度的临床征象。对于出现症状的患者群体,手术治疗仍然是目前临床实践中的主要干预手段。然而,这种治疗模式不仅会显著增加医疗系统的经济负担,其远期疗效和预后转归仍缺乏大规模、多中心的前瞻性研究数据支持^[7]。因此,构建科学的风险评估体系,探索新型的疾病管理策略,以实现对子宫肌瘤发生风险的早期预测和有效干预,已成为当前

妇科领域的重要研究方向^[8]。

近年来,人工智能技术的蓬勃发展为医学研究带来了革命性变革,其中机器学习技术在医疗大数据分析 and 疾病预测模型构建方面取得了显著突破^[9-11]。特别是在肿瘤领域,多种机器学习算法[如支持向量机(SVM)、K 近邻(KNN)和逻辑回归(LR)等]已成功应用于乳腺癌、宫颈癌和卵巢癌的风险分层、早期诊断和预后评估,展现出优异的预测效能^[12-14]。本研究首次创新性地构建并应用了基于集成学习的机器学习模型,用于子宫肌瘤发生风险的预测分析。通过系统比较多种模型的预测性能,结果表明,维生素 D 水平是预测子宫肌瘤的重要特征之一,其中 SVM 模型在预测子宫肌瘤风险时表现出较高的准确性。

SVM 模型优于其他模型的可能机制解释如下:支持向量机(SVM)在分类和回归任务中表现出色,主要得益于其最大化间隔的机制,提升了泛化能力并降低了过拟合风险。SVM 通过核函数处理非线性关系,尤其适合中小型数据集,因为决策边界仅由支持向量决定。此外,SVM 在高维数据处理中表现优异,并通过正则化参数灵活调节模型复杂度。SVM 的可解释性和稀疏解特性使其在需要解释的场景中更具实用性。尽管在大数据集上训练较慢且需调参,SVM 在文本分类、图像识别等复杂任务中仍具有显著优势^[15]。

维生素 D 与子宫肌瘤之间的关联机制已得到广泛研究和证实。Wise 等^[16]的研究及一项 Meta 分析^[17]表明,维生素 D 缺乏可能参与子宫肌瘤的发病机制,且血清维生素 D 水平降低与子宫肌瘤发生风险呈相关性。前期研究^[18]也证实了血清维生素 D 水平与无症状子宫肌瘤患者之间的相关性;本研究在此基础上,结合六种临床特征构建机器学习模型,进一步验证了血清维生素 D 水平与无症状子宫肌瘤之间的关联,为维生素 D 在子宫肌瘤发病机制中的作用提供了更深入的证据支持。

在分子遗传学层面,研究提示,维生素 D 与平滑肌瘤的关联可能受到特定基因多态性(如 rs12800438 和 rs6058017)的调控^[19]。近年来的机制研究^[16,20]进一步阐明,维生素 D 可通过抑制 TGF- β 、Wnt/ β -连环蛋白和 mTOR 等关键信号通路,调控子宫平滑肌瘤的纤维化进程,从而影响肿瘤的生长。这些发现已在动物模型和人类模型研究^[21-22]中得到验证,为维生素 D 在子宫肌瘤发生发展中的生物学作用提供了有力证据。基于上述研究证据,维生素 D 水平不仅可作为评估子宫肌瘤发生风险的重要生物学指标,还可能为子宫肌瘤的预

防和治疗提供新的干预靶点。

李红^[23]的研究系统分析了子宫肌瘤的流行病学特征,发现其发病高危因素主要包括年龄大于 40 岁、初潮年龄早、未生育史以及肥胖等。其中,未生育(即妊娠次数为 0)被证实是子宫肌瘤发生的独立危险因素。其次,年龄与子宫肌瘤发病风险的相关性也在多项研究^[24-25]中得到证实。作为典型的激素依赖性肿瘤,子宫肌瘤的生长与雌激素水平呈现相关性^[26]。临床流行病学数据显示,子宫肌瘤的发病率在育龄期女性中升高,其中 30~50 岁年龄段为高发期。这种年龄分布特征与女性生育期雌激素水平的变化趋势相一致。随着女性年龄增长,子宫肌瘤体积可能逐渐增大,这一过程可能与长期暴露于较高水平的雌激素环境有关。值得注意的是,绝经后女性由于雌激素水平显著下降,子宫肌瘤的生长通常趋于停滞,甚至可能出现体积缩小,这一现象进一步证实了雌激素在子宫肌瘤发生发展中的关键调控作用。

综上,本研究首次将血清维生素 D 水平与临床特征相结合,为子宫肌瘤的风险预测提供新的视角和方向。通过对比多种机器学习算法,验证了支持向量机(SVM)在预测子宫肌瘤风险中的优越性,并结合可解释性分析增强模型的可信度。这一研究为子宫肌瘤的早期筛查、风险分层和个性化预防提供了重要的技术手段,有望帮助临床医生更早识别高风险人群并制定精准的预防策略,从而降低疾病进展风险和医疗成本。

参考文献

[1] Al-Hendy A, Myers ER, Stewart E. Uterine fibroids: burden and unmet medical need[J]. *Seminars in Reproductive Medicine*, 2017, 35(6): 473-480.

[2] Giuliani E, As-Sanie S, Marsh EE. Epidemiology and management of uterine fibroids[J]. *International Journal of Gynaecology and Obstetrics*, 2020, 149(1): 3-9.

[3] Vannuccini S, Petraglia F, Carmona F, et al. The modern management of uterine fibroids-related abnormal uterine bleeding[J]. *Fertility and Sterility*, 2024, 122(1): 20-30.

[4] Pynnä K, Vuorela P, Lodenius L, et al. Cost-effectiveness of hysterectomy for benign gynecological conditions: a systematic review[J]. *Acta Obstetrica et Gynecologica Scandinavica*, 2014, 93(3): 225-232.

[5] Islam MS, Akhtar MM, Segars JH. Vitamin D deficiency and uterine fibroids: an opportunity for treatment or prevention? [J]. *Fertility and Sterility*, 2021, 115(5): 1175-1176.

[6] Armand TPT, Nfor KA, Kim JI, et al. Applications of artificial intelligence, machine learning, and deep learning in nutrition: a systematic review[J]. *Nutrients*, 2024, 16(7): 1073.

[7] Marsh EE, Wegienka G, Williams DR. Uterine fibroids[J]. *Ja-*

- ma, 2024, 331(17):1492.
- [8] American college of obstetricians and gynecologists' committee on practice bulletins - gynecology. Management of symptomatic uterine leiomyomas; ACOG practice bulletin, number 228 [J]. *Obstetrics & Gynecology*, 2021, 137(6):e100—e115.
- [9] Eckhardt CM, Madjarova SJ, Williams RJ, *et al.* Unsupervised machine learning methods and emerging applications in healthcare[J]. *Knee Surgery, Sports Traumatology, Arthroscopy*, 2023, 31(2):376—381.
- [10] Motwani A, Shukla PK, Pawar M. Ubiquitous and smart healthcare monitoring frameworks based on machine learning: a comprehensive review [J]. *Artificial Intelligence in Medicine*, 2022, 134: 102431.
- [11] Silva GFS, Fagundes TP, Teixeira BC, *et al.* Machine learning for hypertension prediction: a systematic review[J]. *Current Hypertension Reports*, 2022, 24(11):523—533.
- [12] Clift AK, Dodwell D, Lord S, *et al.* Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study[J]. *BMJ (Clinical Research Ed)*, 2023, 381:e073800.
- [13] Kan Y, Dong D, Zhang Y, *et al.* Radiomic signature as a predictive factor for lymph node metastasis in early-stage cervical cancer[J]. *Journal of Magnetic Resonance Imaging*, 2019, 49(1):304—310.
- [14] Boehm KM, Aherne EA, Ellenson L, *et al.* Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer[J]. *Nature Cancer*, 2022, 3(6):723—733.
- [15] Zhang J, Lai Z, Kong H, *et al.* Learning the optimal discriminant SVM with feature extraction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(4):2897—2911.
- [16] Wise LA, Ruiz-Narváez EA, Haddad SA, *et al.* Polymorphisms in vitamin D - related genes and risk of uterine leiomyomata [J]. *Fertility and Sterility*, 2014, 102(2):503—510. e1.
- [17] Mohammadi R, Tabrizi R, Hessami K, *et al.* Correlation of low serum vitamin-D with uterine leiomyoma: a systematic review and meta-analysis[J]. *Reproductive Biology and Endocrinology*, 2020, 18(1):85.
- [18] Xu F, Li F, Li L, *et al.* Vitamin D as a risk factor for the presence of asymptomatic uterine fibroids in premenopausal Han Chinese women [J]. *Fertility and Sterility*, 2021, 115(5):1288—1293.
- [19] Islam MS, Ciavattini A, Petraglia F, *et al.* Extracellular matrix in uterine leiomyoma pathogenesis: a potential target for future therapeutics[J]. *Human Reproduction Update*, 2018, 24(1):59—85.
- [20] Corachán A, Ferrero H, Aguilar A, *et al.* Inhibition of tumor cell proliferation in human uterine leiomyomas by vitamin D via Wnt/ β -catenin pathway[J]. *Fertility and Sterility*, 2019, 111(2):397—407.
- [21] Ciavattini A, Delli Carpini G, Serri M, *et al.* Hypovitaminosis D and “small burden” uterine fibroids: Opportunity for a vitamin D supplementation[J]. *Medicine*, 2016, 95(52):e5698.
- [22] Elhusseini H, Elkafas H, Abdelaziz M, *et al.* Diet-induced vitamin D deficiency triggers inflammation and DNA damage profile in murine myometrium [J]. *International Journal of Women's Health*, 2018, 10:503—514.
- [23] 李红. 子宫肌瘤快速生长相关因素的临床回顾性研究[D]. 重庆:重庆医科大学, 2024.
- [24] 娄净, 李春明, 周坚红. 子宫肌瘤病因与预防[J]. *实用妇产科杂志*, 2024, 40(2):81—84.
- [25] 沈杨, 许茜, 徐洁, 等. 子宫肌瘤危险因素流行病学调查研究[J]. *实用妇产科杂志*, 2013, 29(3):189—193.
- [26] 纪晓菁, 王巍. 非绝经期女性子宫肌瘤与代谢综合征相关指标相关性分析[J]. *华南预防医学*, 2020, 46(5):551—553.

(收稿日期: 2025—02—03)

修回日期: 2025—03—10)