

可解释的机器学习预测模型用于子宫内膜癌患者的早期识别

佟丽艳¹, 刘巍², 安月盘², 马重¹

(1. 内蒙古医科大学研究生院, 内蒙古 呼和浩特 010110; 2. 内蒙古自治区妇幼保健院妇科, 内蒙古 呼和浩特 010020)

[摘要]目的 本研究旨在开发一种机器学习模型, 该模型利用患者通过无创伤检查所获得的临床数据建立, 用于对所有年龄段的妇女进行子宫内膜癌风险预测, 进而提高子宫内膜癌早期诊断的效率。方法 研究收集了2020年1月至2025年3月内蒙古自治区妇幼保健院344例子宫内膜癌(EC)患者及344例子宫内膜良性病变(包括子宫内膜息肉、子宫内膜增生不伴不典型增生)患者的临床数据, 从临床数据中识别出16个相关特征。采用10折交叉验证划分训练集和验证集, 分别构建极限梯度提升(XGBoost)、随机森林(RF)、Logistic回归(LR)和支持向量机(SVM)四种模型, 通过受试者工作特征曲线下面积(AUC)、灵敏度、特异度进行四种模型性能评估选出最优模型。此外, 选取其他医院142名患者组成的测试集对最优模型进行验证与敏感性分析, 最终对模型进行SHAP可解释性分析。结果 XGBoost模型在所有指标上表现最佳, 整体预测能力最强。在训练集中, 其AUC为0.96, 在测试集中, AUC达到了0.89, SHAP可解释性分析表明了年龄、血小板分布宽度(PDW)和甘油三酯-葡萄糖指数(TyG)作为预测因子的重要性。结论 本研究构建了一个子宫内膜癌早期诊断预测模型, 并识别出与子宫内膜癌发病风险相关的特征指标。该模型能有效辅助临床识别潜在高风险人群, 指导精准干预。

[关键词]子宫内膜癌; 机器学习; 模型; 早期诊断; 预测

Doi:10.3969/j.issn.1673-5293.2026.04.003

[中图分类号]R173

[文献标识码]A

[文章编号]1673-5293(2026)04-0020-08

Explainable machine learning prediction model for the early identification of patients with endometrial carcinoma

TONG Liyan¹, LIU Wei², AN Yuepan², MA Chong¹

(1. Graduate School, Inner Mongolia Medical University, Inner Mongolia Hohhot 010110, China; 2. Department of Gynecology, Inner Mongolia Maternal and Child Health Care Hospital, Inner Mongolia Hohhot 010020, China)

[Abstract] **Objective** This study aimed to develop a machine learning (ML) model based on clinical data obtained from noninvasive examinations to predict the risk of endometrial carcinoma (EC) in women of all ages, thereby improving the efficiency of early diagnosis of EC. **Methods** Clinical data were collected from 344 patients with EC and 344 patients with benign endometrial lesions (including endometrial polyps and endometrial hyperplasia without atypia) treated at Inner Mongolia Maternal and Child Health Care Hospital between January 2020 and March 2025. Sixteen relevant features were identified from the clinical dataset. The dataset was divided into training and validation sets using 10-fold cross-validation. Four models—eXtreme gradient boosting (XGBoost), random forest (RF), logistic regression (LR), and support vector machine (SVM)—were constructed. Model performance was evaluated using the area under the curve (AUC), sensitivity, and specificity to determine the optimal model. In addition, an external test set consisting of 142 patients from other hospitals was used to validate the optimal model and conduct sensitivity analysis. Finally, SHapley Additive exPlanations (SHAP) were applied to interpret the model. **Results** Among the four models, the XGBoost model demonstrated the best performance across all evaluation metrics and exhibited the strongest overall predictive ability. In the training set, the AUC was 0.96, while in the test set the AUC reached 0.89. SHAP analysis indicated that age, platelet distribution width (PDW), and the triglyceride-glucose index (TyG) were among the most important predictive factors. **Conclusion** This study developed a predictive model for the early diagnosis of EC and identified several feature indicators associated with the risk of EC. The model may effectively assist clinicians in identifying individuals at high risk and facilitate targeted clinical interventions.

[Key words] endometrial carcinoma; machine learning; model; early diagnosis; prediction

[收稿日期]2025-09-19

[基金项目]内蒙古自治区科技计划项目(2022YFSH0030); 内蒙古医学科学院公立医院科研联合基金项目(2024GLLH0207)

[作者简介]佟丽艳(2000—), 女, 硕士研究生, 主要从事妇科肿瘤相关研究。

[通讯作者]刘巍, 主任医师。

子宫内膜癌(endometrial cancer, EC)是妇科生殖系统三大恶性肿瘤之一,在中国,2022 年约有新发病例 77 700 例,死亡病例 13 500 例^[1]。由于经济水平提高,生活习惯的改变,代谢性疾病在全球范围内处于持续上升趋势,再加上许多发达国家老龄化趋势加剧,导致 EC 的发病率和死亡率均增加,并且呈现年轻化趋势,在欧美等发达地区,EC 已成为女性生殖系统最常见的恶性肿瘤,约占妇科恶性肿瘤新发病例的 50%^[2]。而 EC 的预后与临床分期有密切关系,早期诊断 EC,甚至通过常规检查预测 EC 发病风险,进而采取针对性预防或治疗措施,对改善 EC 预后,减少因 EC 造成的死亡或残疾将会有重大意义。在精准医疗时代,人们越来越关注通过无创手段识别高风险个体,以最大限度地提高早期临床干预的积极效果。目前,人工智能算法结合大数据挖掘技术将医院电子病历系统中的非结构化数据转换为有价值的决策信息已经成为可能。机器学习(machine learning, ML)在疾病预测方面表现出良好的性能,与传统的统计方法相比,ML 可以处理复杂的相互作用和非线性关系,致其在临床研究中的使用越来越多,可解释的人工智能方法的最新进展可以帮助解释 ML 模型,它便于临床医生和研究人员理解 ML 算法创建的决策过程和输出过程^[3]。本研究利用机器学习的优势开发了一个 EC 早期诊断预测模型,并识别出一些新的特征指标。

1 资料与方法

1.1 资料来源

本研究回顾性收集了 2020 年 1 月至 2025 年 3 月在内蒙古自治区妇幼保健院接受手术治疗,且病理诊断为 EC 或子宫内膜良性病变(包括子宫内膜息肉、子宫内膜增生不伴不典型增生)的 688 例患者的临床数据,另外还收集内蒙古自治区人民医院 2020 年 1 月至 2025 年 3 月的 142 例病理诊断为 EC 或子宫内膜良性病变患者临床数据作为测试集。

通过查阅文献及结合临床实际,本研究纳入以下特征变量进行预测模型构建:年龄、体质量指数(body mass index, BMI)、初潮年龄、是否绝经、是否有孕产史、是否有糖尿病史;经阴道超声检查是否有宫腔占位、内膜回声是否均匀、内膜或占位有无血流信号;糖类抗原 125(endometrial carcinoma 125,

CA125)、人附睾蛋白 4(human epididymis protein 4, HE4)、糖类抗原 153(endometrial carcinoma 153, CA153)、糖类抗原 199(endometrial carcinoma 199, CA199)、癌胚抗原(carcinoembryonic antigen, CEA)、血小板与淋巴细胞比值(platelet-to-lymphocyte ratio, PLR)、中性粒细胞与淋巴细胞比值(neutrophil-to-lymphocyte ratio, NLR)、血小板分布宽度(platelet distribution width, PDW)、红细胞分布宽度-变异系数(red cell distribution width-coefficient of variation, RDW-CV)、甘油三酯与高密度脂蛋白胆固醇比值(triglyceride to high-density lipoprotein cholesterol ratio, TG/HDL-C)、甘油三酯葡萄糖指数(triglyceride-glucose index, TyG)。本研究获得了内蒙古自治区妇幼保健院伦理委员会的批准(伦理审批号:[2025]伦审 K 第[007-2]号)。纳入标准:①术后病理确诊为 EC 的患者及术后病理证实为良性疾病的患者(手术治疗包括全子宫切除术或全子宫切除术加双侧附件切除术,可酌情辅以腹腔灌洗、盆腔淋巴结清扫和/或主动脉旁淋巴结清扫,或行诊断性刮宫及宫腔镜下电切术。);②患者术前未接受放疗、化疗、靶向治疗或激素治疗。排除标准:①排除患有其他系统恶性肿瘤、有凝血功能障碍病史、自身免疫性疾病、严重肝肾功能不全或临床病史不完整的患者;②病历资料不完善、缺失者;③血液样本采集前近一个月有影响炎症指标、肿瘤指标的药物服用的患者;④术前有全身或局部炎症,影响实验室检查结果者。

1.2 方法及工具

采用 SPSS 25.0 软件对基线资料进行统计学分析,计量资料以均数±标准差($\bar{x} \pm s$)表示,采用独立样本 t 检验,计数资料以例数(n)和率(%)表示,采用 χ^2 检验。采用 10 折交叉验证划分训练集和验证集,分别构建极限梯度提升(eXtreme gradient boosting, XGBoost)、随机森林(random forest, RF)、Logistic 回归(logistic regression, LR)、支持向量机(support vector machine, SVM)四种模型,通过受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)、灵敏度、特异度进行四种模型性能评估选出最优模型。此外,选取其他医院 142 例病理诊断为 EC 或子宫内膜

良性病变患者临床数据组成的测试集对最优模型进行验证与敏感性分析,最终对模型进行 SHAP 可解释性分析。

缺失值处理:去掉缺失占比大于 30%的变量,对其余变量进行多重填补,并利用填补的数据集和删除缺失值所在行的数据集分别建模进行敏感性分析。

参数调优:使用网格搜索法,在固定迭代次数、树深度、学习率、分裂所需的最小损失减少量(正则化参数)、特征采样比例、叶节点最小权重(正则化参数)以及样本采样比例这 7 个参数构成的 144 种组合中,寻找最优参数组合来构建模型。

工具:采用 R4.4.3 进行数据处理与建模,采用 mice 包进行缺失值填补,采用 caret 包构建预测模型,采用 pROC 包评估模型性能,采用 ggplot2、

patchwork 包绘图,采用 SHAPforxgboost 包进行 SHAP 分析。

2 结果

2.1 一般资料

本研究共收集了 830 例患者的相关数据资料,其中 EC 患者(病例组)455 例,子宫内膜良性病变(对照组)375 例。两组基线资料比较可以发现,除了两组的有无孕产史、有无宫腔占位、NLR 指标、RDW-CV 指标四个特征比较差异无统计学意义($P>0.05$),其余指标均有显著差异($P<0.05$),见表 1。且根据既往研究发现孕产史、宫腔占位、NLR 升高、RDW-CV 升高对 EC 的发生有影响,故本研究将所有特征变量均纳入预测模型。

表 1 两组基线资料比较[$\bar{x}\pm s, n(\%)$]

Table 1 Comparison of baseline characteristics between the two groups [$\bar{x}\pm s, n(\%)$]

特征	例数	病例组	对照组	t/χ^2	P
基本资料					
年龄(岁)	830	57.39±9.24	43.69±10.14	20.153	<0.001
BMI(kg/m ²)	827	26.84±3.74	24.20±3.54	10.367	<0.001
初潮年龄(岁)	830	13.35±0.80	14.05±0.78	-12.896	<0.001
是否绝经				211.952	<0.001
否	485	163(35.8)	322(85.9)		
是	345	292(64.2)	53(14.1)		
有无孕产史				0.660	0.520
无	41	25(5.5)	16(4.3)		
有	789	430(94.5)	359(95.7)		
有无糖尿病史				30.996	<0.001
无	755	391(85.9)	364(97.1)		
有	75	64(14.1)	11(2.9)		
超声指标					
有无宫腔占位				5.383	0.056
无	475	245(54.0)	230(61.3)		
有	354	209(46.0)	145(38.7)		
内膜回声是否均匀				48.468	<0.001
否	570	358(78.9)	212(56.5)		
是	259	96(21.1)	163(43.5)		

(转下表)

(续上表)

特征	例数	病例组	对照组	t/χ^2	P
内膜或占位有无血流信号				66.548	<0.001
无	556	250(55.1)	306(81.8)		
有	272	204(44.9)	68(18.2)		
血清学指标					
CA125(U/mL)	815	34.77±67.11	17.66±16.18	5.210	<0.001
HE4(pmol/L)	464	100.97±97.68	45.80±14.06	9.971	<0.001
CA153(U/mL)	255	12.77±9.22	10.04±4.54	2.902	0.004
CA199(U/mL)	392	27.67±61.18	10.27±6.77	4.195	<0.001
CEA(ng/mL)	320	2.00±2.27	1.20±0.58	4.793	<0.001
PLR	821	151.90±65.47	170.43±77.23	-3.637	<0.001
NLR	820	2.48±2.12	2.48±2.41	-0.020	0.984
PDW(fL)	814	13.31±2.78	11.65±1.93	10.007	<0.001
RDW-CV(%)	822	14.05±4.25	14.20±3.12	-0.559	0.576
TG/HDL-C	660	1.50±1.11	1.13±0.94	4.569	<0.001
TyG	658	8.84±0.57	8.41±0.50	10.312	<0.001

2.2 缺失值处理

HE4 四个变量,对其余变量进行多重填补。
去掉缺失占比>30%的 CA153、CEA、CA199、

表 2 数据缺失情况
Table 2 Summary of missing data

变量	训练-验证集 缺失比例(%)	测试集 缺失比例(%)	变量	训练-验证集 缺失比例(%)	测试集 缺失比例(%)
年龄	0.000	0.000	PDW	1.890	2.113
BMI	0.000	2.113	RDW-CV	0.727	2.113
初潮年龄	0.000	0.000	TG/HDL-C	20.349	21.127
CA125	2.180	0.000	TyG	20.640	21.127
HE4	40.988	59.155	是否有宫腔占位	0.000	0.704
CA153	64.826	90.845	内膜回声是否均匀	0.000	0.704
CA199	51.453	59.155	内膜或占位有无血流信号	0.145	0.704
CEA	61.483	61.268	是否绝经	0.000	0.000
PLR	0.872	2.113	孕产史	0.000	0.000
NLR	1.017	2.113	糖尿病史	0.000	0.000

2.3 模型的构建及评估

结果:XGBoost 在所有指标上表现最佳,AUC 为 0.9563,灵敏度为 0.8866,特异度为 0.8780,其整体预测能力最强。Logistic 回归和 SVM 性能稍逊,AUC 分别为 0.9371 和 0.9320,灵敏度和特异度均

低于集成模型(XGBoost 和 RF)。在本数据集上,集成模型显著优于传统模型(Logistic 回归和 SVM),表明复杂非线性关系可能主导预测任务。XGBoost 的优异表现使其成为本研究的首选。

表 3 四种模型性能均值比较

Table 3 Comparison of the mean performance of the four models

模型	AUC	灵敏度	特异度
LR	0.9371	0.8579	0.8725
RF	0.9489	0.8897	0.8548
XGBoost	0.9563	0.8866	0.8780
SVM	0.9320	0.8657	0.8518

2.4 参数调优

表 4 调参前后模型性能比较

Table 4 Comparison of model performance before and after hyperparameter tuning

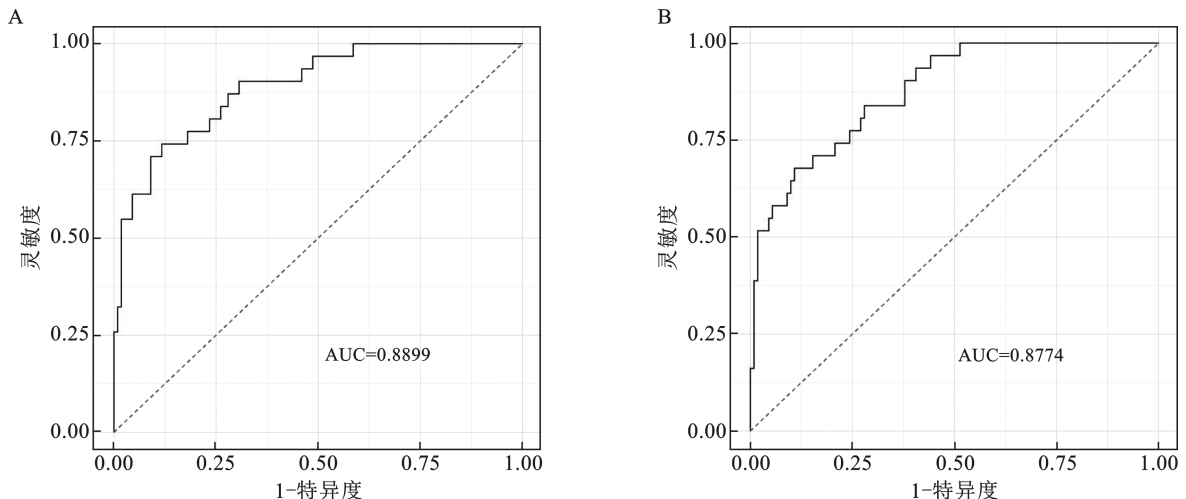
模型	AUC	灵敏度	特异度
XGBoost-Tuned	0.9425	0.8838	0.8553
XGBoost-Untuned	0.9563	0.8866	0.8780

根据上一步的结论选择 XGBoost 作为基础模型进行参数调优,通过网格搜索对 XGBoost 模型评

估的超参数调优后发现,与未调优模型相比,性能提升非常有限。表明默认参数已接近该模型的最优性能,鉴于调优的收益甚微,在进一步的研究中采用未调优的 XGBoost 模型。

2.5 模型验证与敏感性分析

为了评估本研究最佳模型在不同缺失值处理方式下的稳健性,本研究设计并实施了敏感性分析。通过比较模型在填补数据与删除缺失行数据上训练的两个模型在测试集上的预测效果,可以判断缺失值处理策略是否对模型性能造成显著影响,从而增强研究结论的可靠性与说服力。此外,敏感性分析也为选择最合适的缺失值处理方法提供了依据。在外部验证集上,最佳模型(XGBoost)在不同缺失值处理策略下的预测性能表现一致良好。结果如图 1,填补数据上训练的模型的 AUC 为 0.8899。删除缺失行数据上训练的模型的 AUC 为 0.8774。



注:A 为填补数据训练的模型在测试集上的预测效果,B 为删除缺失行数据训练的模型在测试集上的预测效果

图 1 缺失值处理策略对模型性能的影响

Fig. 1 Impact of missing-value handling strategies on model performance

从上述结果中可以得出,不同缺失值处理方法对模型的预测能力影响较小。模型对缺失值处理策略具有较好的稳健性。填补数据可作为首选策略,从而最大化利用样本,同时保持模型性能。敏感性分析结果表明,不同缺失值处理策略下模型性能差异极小,模型具有良好的稳健性。

2.6 SHAP 分析与可视化

如图 2,变量的重要性从高到低前三位依次是:年龄(平均绝对 SHAP 值 1.557)、PDW(平均绝对 SHAP 值 0.841)、TyG 指数(平均绝对 SHAP 值

0.690)。重要性最低的是:有糖尿病病史(平均绝对 SHAP 值 0.035)。随着年龄、PDW、TyG、CA125 等的增大,EC 发生的风险也增大。

年龄与患病风险大体上呈正相关关系。随着标准化年龄的升高,模型预测的正向贡献增强,提示年龄是影响 EC 患病风险的重要因素,这与临床实际结果一致(见图 3)。

TyG 与患病风险呈正相关关系。低 TyG 标准化值可能与较低的患病风险相关,而高 TyG 标准化值(尤其是超过 0)显著增加患病概率(见图 4)。

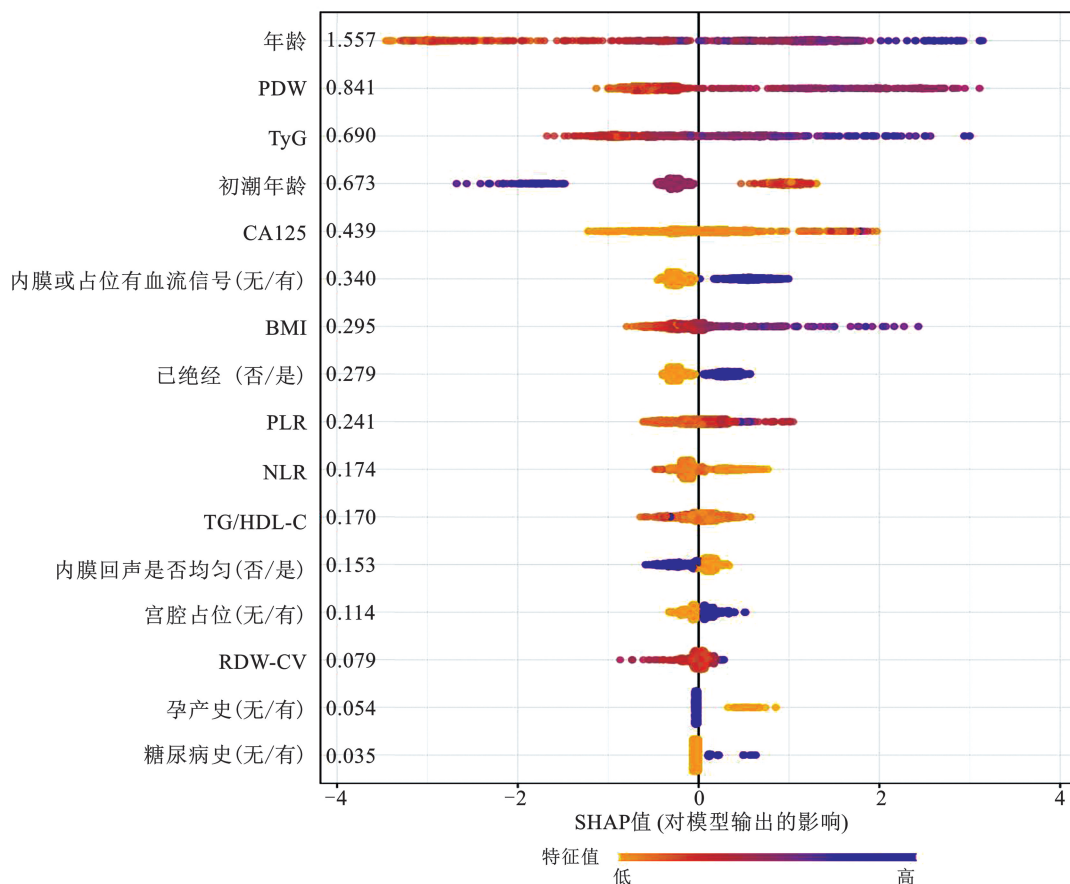
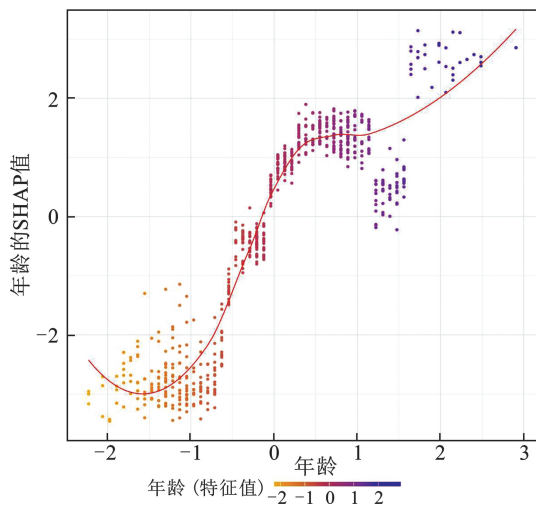


图 2 SHAP 蜂群图

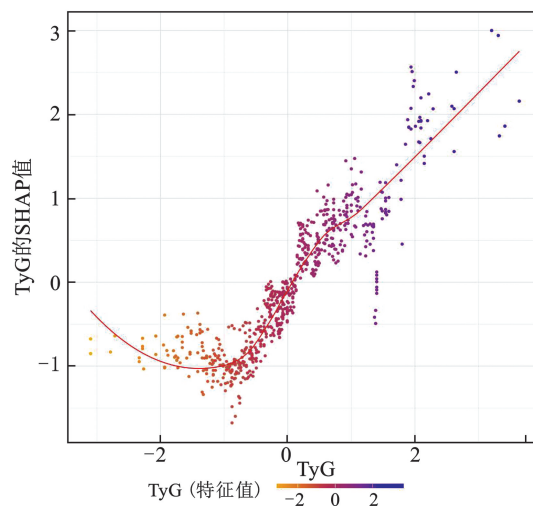
Fig. 2 SHAP beeswarm plot



注：年龄数据已进行标准化处理。

图 3 年龄 SHAP 依赖图

Fig. 3 SHAP dependence plot for age



注：TyG 数据已进行标准化处理。

图 4 TyG SHAP 依赖图

Fig. 4 SHAP dependence plot for TyG

3 讨论

在全球范围内,EC 仍然是发病率和致死率最高的三大妇科恶性肿瘤之一,每年新增确诊病例约为 420 000 例,凸显了 EC 早期诊断以改善临床预后的

迫切需要^[4]。宫腔镜下子宫内膜活检及诊断性刮宫术目前仍是诊断 EC 的金标准,其优势在于精准定位病变的同时取样,明显降低了 EC 的漏诊率。然而,内膜活检或刮宫术是一种侵入性诊断手段,除具有疼痛、出血和子宫穿孔等潜在风险^[5]外,患者对创

伤性诊断的抗拒也在一定程度上限制了其早期应用,因此使得一部分早期患者病情延误。因此,人们越来越关注开发非创伤性诊断方法,实现 EC 的预测或早期诊断,并实现大规模临床应用。哈特等^[6]基于个人临床基本信息使用随机森林(RF)来预测绝经后 EC 的发生风险。由于种族、地区和生活习惯的地理差异,该模型并不完全适用于中国妇女。Pergialiotis 等^[7]开发的模型对有绝经后出血症状的 EC 患者预测具有较高的准确性,但本研究模型在无限定人群(包含绝经前女性)中实现了更高的综合性能(AUC=0.96),且通过外部验证(AUC=0.89)证实了模型的稳健性,该模型突破了传统筛查对绝经状态或特定症状的依赖,为 EC 的早期识别提供了普适性工具。

3.1 模型构建与性能评估

本研究预测模型的构建使用了 16 个特征变量,包括人口统计学特征、超声指标和实验室指标,这些变量通常包括在大多数医院的常规体检及门诊检查中,使得该新模型更加实用。该模型包含了大多数统计学模型中基本的风险因素,例如年龄,绝经状态和 BMI 等。除此以外,还包括实验室指标及超声指标,如 PDW、TyG、CA125、内膜或占位有无血流信号等,本研究首次将 TyG 指数纳入预测模型。研究结果表明,XGBoost 模型在预测性能上显著优于其他模型(AUC=0.96),且通过 SHAP 可解释性分析揭示了年龄、血小板分布宽度(PDW)和甘油三酯-葡萄糖指数(TyG)作为关键预测因子的重要性,SHAP 依赖图显示年龄增长与 EC 风险呈正相关,符合 EC 高发于围绝经期及老年女性的流行病学特征。研究证实了 TyG 指数对 EC 的预测价值,这与 EC 发病的代谢机制^[8](如肥胖、糖尿病)高度吻合。还发现了 PDW 的重要预测价值,提示血小板介导的炎症微环境可能参与 EC 发生,符合肿瘤的炎症特征。

3.2 关键预测因子解读

本研究首次将 TyG 指数纳入预测模型,同时发现 TyG 指数对模型预测的贡献度较高且 TyG 指数与 EC 患病风险呈正相关。TyG 指数是近年来引起广泛关注的一个代谢性指标,流行病学和临床研究表明,代谢综合征是 EC 进展中的重要致病条件,其

中胰岛素抵抗在致癌过程中发挥着关键作用^[9]。肥胖会导致胰岛素抵抗,这反过来又会导致性激素代谢的变化、慢性炎症和循环胰岛素水平升高。在胰岛素抵抗的患者中,胰腺产生更多胰岛素以弥补组织对胰岛素反应的下降,从而导致高胰岛素血症,过量的胰岛素会参与肿瘤进展^[10]。TyG 目前已被广泛认可并用于预测胰岛素抵抗的发生。与胰岛素抵抗指数不同,其计算时不需要胰岛素值,通过血清甘油三酯和血糖水平计算得出,计算公式为: $TyG = \ln[TG(mg/dL) \times FPG(mg/dL) / 2]$ ^[11-12],因此 TyG 在患者体检时具有普遍适用性。一项研究结果表明^[13],无论是否调整炎症标志物,TyG 指数与 EC 之间均存在很强的相关性,TyG 指数可能是 EC 进展的潜在标志物,且表明 TyG 指数对 EC 有独立影响。另一项分析 NHANES 数据库的研究表明,TyG 指数可能成为美国女性生殖系统癌症的一个潜在的生物标志物^[14]。上述两项研究结果均与本研究结果相符。

在本研究中,另一个特征指标 PDW 也展示出了较高的预测性能,既往的研究多关注于 PDW 在肿瘤预后方面的影响^[15],较少有研究指出其在疾病预测中的作用。肿瘤的十大特征中,血常规指标的变化与肿瘤促进性炎症和免疫逃逸的发生密切相关,其中中性粒细胞、淋巴细胞及血小板的动态变化与肿瘤微环境中炎症与免疫系统的交互作用较为突出^[16-17]。血小板数量和功能异常已被证实是许多癌症的副肿瘤综合征的一部分,癌细胞诱导血小板活化和聚集,进而导致肿瘤进展、侵袭、血管生成、转移和免疫逃避^[18]。血小板作为肿瘤转移的“帮凶”,在肿瘤患者体内会有所升高。血小板通过包裹循环中的肿瘤细胞,保护其免受血流剪切力破坏及免疫清除,血小板释放的血小板第四因子(platlet factor 4, PF4)等介质促进肿瘤细胞与血管内皮黏附,加速 EC 远处转移^[17]。PDW 升高提示血小板活化异常性,可能与 EC 相关的血管异常或凝血紊乱有关,Ural 等^[19]的研究中 EC 组 PDW 显著升高($P < 0.001$),与本研究结果类似,再次证实了 PDW 可能与 EC 的发生密切相关。

3.3 研究局限性与转化前景

虽然本研究通过简单易得的非创伤性指标构建

了一个具有良好预测性能及普遍适用性的 EC 早期预测模型,但仍存在一些局限。第一,研究所纳入数据来自单一临床机构,不可避免的存在样本代表性不足、数据同质化问题。第二,本研究是回顾性研究,收集数据中产生了缺失及不平衡的数据,例如 CA153、CEA、CA199、HE4 因缺失率高被删除,其中 HE4 已被其他研究^[20]证实对 EC 的诊断有重要意义,希望未来能通过改善研究方案、增加样本量来解决这个问题。第三,该模型中糖尿病病史对预测结果的贡献较低,与临床实际不一致,考虑与测试工具的差异有关。

综上所述,该模型可以利用非创伤性检查结果更早的预测 EC 发病风险,帮助临床医生选择合适的临床策略来预防或早期诊断 EC,以改善预后。将来可以从实验设计、提升临床转化等方面进一步改进研究,为子宫内膜的预测及早期诊断提供更多研究支撑。

[参考文献]

- [1]Zheng R S,Chen R,Han B F,et al. Cancer incidence and mortality in China, 2022 [J]. *Zhonghua Zhong Liu Za Zhi*,2024,46(3):221-231.
- [2]Henley S J,Ward E M,Scott S,et al. Annual report to the nation on the status of cancer,part I:national cancer statistics[J]. *Cancer*,2020,126(10):2225-2249.
- [3]Saeed W,Omlin C W P. Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities[J]. *ArXiv*,2021,abs/2111.06420.
- [4]Crosbie E J,Kitson S J,McAlpine J N,et al. Endometrial cancer[J]. *Lancet*,2022,399(10333):1412-1428.
- [5]O'Flynn H,Ryan N A J,Narine N,et al. Diagnostic accuracy of cytology for the detection of endometrial cancer in urine and vaginal samples[J]. *Nat Commun*,2021,12(1):952.
- [6]Hart G R,Yan V,Huang G S,et al. Population-based screening for endometrial cancer:human vs. machine intelligence[J]. *Front Artif Intell*,2020,3:539879.
- [7]Pergialiotis V,Pouliakis A,Parthenis C,et al. The utility of artificial neural networks and classification and regression trees for the prediction of endometrial cancer in postmenopausal women[J]. *Public Health*,2018,164:1-6.
- [8]李婵玉,邓洁. 肥胖与妇科肿瘤关系的研究进展[J]. *中国妇幼健康研究*,2020,31(10):1441-1444.
- [9]Chiefari E,Mirabelli M,La Vignera S,et al. Insulin resistance and cancer:in search for a causal link[J]. *Int J Mol Sci*,2021,22(20):11137.
- [10]Iyengar N M,Hudis C A,Dannenberg A J. Obesity and cancer:local and systemic mechanisms[J]. *Annu Rev Med*,2015,66:297-309.
- [11]Tao L C,Xu J N,Wang T T,et al. Triglyceride-glucose index as a marker in cardiovascular diseases:landscape and limitations[J]. *Cardiovasc Diabetol*,2022,21(1):68.
- [12]Xiang Q,Xu H,Zhan J,et al. Association between the triglyceride-glucose index and vitamin D status in type 2 diabetes mellitus[J]. *Nutrients*,2023,15(3):639.
- [13]Shi H,Guo F,Zheng K,et al. Triglyceride-glucose index (TyG index) and endometrial carcinoma risk:a retrospective cohort study[J]. *Int J Gynaecol Obstet*,2024,164(1):298-304.
- [14]Shi H,Zhou L,Yang S,et al. The relationship between triglyceride and glycose (TyG) index and the risk of gynaecologic and breast cancers[J]. *Clin Nutr ESPEN*,2022,51:345-352.
- [15]Lin H,Zhong W,Zhong L,et al. The inflammatory markers combined with CA125 may predict postoperative survival in endometrial cancer[J]. *J Obstet Gynaecol*,2024,44(1):2373937.
- [16]de Visser K E,Joyce J A. The evolving tumor microenvironment:from cancer initiation to metastatic outgrowth[J]. *Cancer Cell*,2023,41(3):374-403.
- [17]Sun Y,Li T,Ding L,et al. Platelet-mediated circulating tumor cell evasion from natural killer cell killing through immune checkpoint CD155-TIGIT[J]. *Hepatology*,2025,81(3):791-807.
- [18]Allensworth S K,Langstraat C L,Martin J R,et al. Evaluating the prognostic significance of preoperative thrombocytosis in epithelial ovarian cancer[J]. *Gynecol Oncol*,2013,130(3):499-504.
- [19]Ural Ü M, chitoğlu i,Tekin Y B,et al. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios in patients with endometrial hyperplasia and endometrial cancer[J]. *J Obstet Gynaecol Res*,2015,41(3):445-448.
- [20]Liu J,Han L,Sun Q,et al. Meta-analysis of the diagnostic accuracy of HE4 for endometrial carcinoma[J]. *Eur J Obstet Gynecol Reprod Biol*,2020,252:404-411.

[专业责任编辑:安瑞芳]

[中文编辑:刘昕田;英文编辑:陈子怡]