

基于经会阴盆底超声参数及超声影像组学的联合模型预测产后发生盆腔脏器脱垂的临床价值

黄丽华^{1,2}, 周煜皓², 文戈³

摘要 **目的** 基于会阴盆底超声(TPFU)参数及超声影像组学构建联合模型,探讨其预测产后发生盆腔脏器脱垂(POP)的临床价值。**方法** 选取南方医科大学南方医院增城院区行 TPFU 检查的产后女性 331 例,按照 7:3 的比例分为训练集 231 例(POP 患者 191 例,非 POP 患者 40 例)和测试集 100 例(POP 患者 75 例,非 POP 患者 25 例),收集训练集中 POP 患者和非 POP 患者临床资料、TPFU 参数,采用二元 Logistic 回归分析筛选预测产后发生 POP 的独立影响因素;同时基于训练集的超声图像,采用 3D-Slicer 软件勾画感兴趣区,使用 Pyradiomics 提取影像组学特征,经过标准化处理后,通过 *t* 检验、Spearman 秩相关系数法、最小冗余最大相关性法(mRMR)和最小绝对收缩和选择算子(LASSO)回归进行特征筛选,最终筛选出非零系数特征作为影像组学标签(Radsignature)。采用 5 种机器学习算法[Logistic 回归(LR)、支持向量机(SVM)、随机森林(RF)、极度随机树(ET)、轻量级梯度提升机(LightGBM)]分别基于独立影响因素、Radsignature 构建临床预测模型和超声影像组学模型;绘制受试者工作特征(ROC)曲线分析 5 种模型的诊断效能,筛选曲线下面积(AUC)高且泛化能力好的模型作为最优临床预测模型和最优超声影像组学模型进行后续研究;将筛选出的产后发生 POP 的独立影响因素及 Radsignature 纳入 Logistic 回归分析构建联合模型。绘制 ROC 曲线分析最优临床预测模型、最优超声影像组学模型和联合模型预测产后发生 POP 的诊断效能;Hosmer-Lemeshow 拟合优度检验和校准曲线评估各模型的校准度;临床决策曲线评估各模型的临床适用性。**结果** 二元 Logistic 回归分析显示,肛提肌裂孔横径为预测产后发生 POP 的独立影响因素($OR=1.139, P<0.05$)。在 5 种机器学习算法所构建的临床预测模型中,RF 模型为最优模型,其在训练集和测试集中预测产后发生 POP 的 AUC 分别为 0.845、0.820。基于 Radsignature 构建的影像组学模型中,RF 模型为最优模型,其在训练集和测试集中预测产后发生 POP 的 AUC 分别为 0.865、0.831。联合模型在训练集和测试集中预测产后发生 POP 的 AUC 分别为 0.897、0.856,均高于最优临床预测模型和最优超声影像组学模型,差异均有统计学意义(均 $P<0.05$)。Hosmer-Lemeshow 拟合优度检验显示,最优临床预测模型、最优超声影像组学模型和联合模型在训练集和测试集中的拟合度均较好($\chi^2=0.097, 0.060, 0.806, 0.099, 0.137, 0.107$, 均 $P>0.05$)。校准曲线分析显示,在训练集和测试集中最优临床预测模型和最优超声影像组学模型的预测曲线均较偏离标准曲线,联合模型的预测曲线均较贴近标准曲线,其校准度最高。临床决策曲线分析显示,在训练集中,最优临床预测模型在概率阈值为 67%~88% 时临床净获益最高,最优超声影像组学模型在概率阈值为 82%~85% 时临床净获益最高,联合模型在概率阈值为 34%~98% 时临床净获益最高;在测试集中,最优临床预测模型在概率阈值为 43%~54% 和 61%~88% 时临床净获益最高,最优超声影像组学模型在概率阈值为 54%~56% 和 69%~85% 时临床净获益最高,联合模型在概率阈值为 54%~92% 时临床净获益最高。**结论** 基于 TPFU 参数及超声影像组学的联合模型可用于预测产后发生 POP,具有一定的临床价值。

关键词 超声检查,经会阴;肛提肌裂孔;影像组学;盆腔脏器脱垂

[中图法分类号] R445.1

[文献标识码] A

Clinical value of a combined model based on transperineal pelvic floor ultrasound parameters and ultrasound radiomics for predicting the occurrence of postpartum pelvic organ prolapse

HUANG Lihua^{1,2}, ZHOU Yuhao², WEN Ge³

1.Department of Ultrasound Diagnosis, 3.Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China. 2.Department of Ultrasound Diagnosis, Nanfang Hospital Zengcheng Campus, Southern Medical University, Guangzhou 510515, China

ABSTRACT Objective To construct a combined model based on transperineal pelvic floor ultrasound (TPFU) parameters and ultrasound radiomics, and to explore its clinical value for predicting the occurrence of postpartum pelvic organ prolapse (POP). **Methods** A total of 331 postpartum women who underwent TPFU examination in Zengcheng Campus of Southern Hospital of Southern Medical University were selected. They were divided into a training set of 231 cases (191 POP patients and 40 non-POP patients) and a testing set of 100 cases (75 POP patients and 25 non-POP patients) at a ratio of 7:3. The clinical data and TPFU parameters of POP and non-POP patients in the training set were collected. Binary Logistic regression analysis was used to screen independent influencing factors for predicting the occurrence of postpartum POP. At the same time, based on the ultrasound images of the training set, 3D-Slicer software was used to delineate the region of interest, and Pyradiomics was used to extract radiomics features. After standardization, *t*-test, Spearman rank correlation coefficient method, mRMR, and LASSO were used for feature screening. Finally, the non-zero coefficient features were selected as radiomics labels (Radsignature). 5 machine learning algorithms (LR, SVM, RF, ET, and LightGBM) were used to construct clinical prediction models and ultrasound radiomics models based on independent influencing factors and Radsignature, respectively. Receiver operating characteristic (ROC) curve was drawn to analyze the diagnostic efficacy of 5 models, and select the model with high area under the curve (AUC) and good generalization ability as the optimal clinical prediction model and ultrasound radiomics model for subsequent research. The selected independent influencing factors and Radsignature were incorporated into Logistic regression analysis to construct a combined model. ROC curve was drawn to analyze the diagnostic efficacy of the optimal clinical prediction model, optimal ultrasound radiomics model, and combined model for predicting the occurrence of postpartum POP. Calibration curve and Hosmer-Lemeshow goodness-of-fit test were used to evaluate the calibration of each model. Clinical decision curve was used to evaluate the clinical applicability of each model. **Results** Binary Logistic regression analysis showed that the transverse diameter of the anal sphincter hiatus (TD) was an independent influencing factor for predicting the occurrence of postpartum POP ($OR=1.139, P<0.05$). Among the clinical prediction models constructed by 5 machine learning algorithms, the RF model was the optimal model, with the AUC of 0.845 and 0.820 for predicting the occurrence of postpartum POP in the training and testing sets, respectively. The RF model was the optimal model among the ultrasound radiomics models constructed by Radsignature, with the AUC of 0.865 and 0.831 for predicting the occurrence of postpartum POP in the training and testing sets, respectively. The AUC of combined model for predicting the occurrence of postpartum POP in the training and testing sets were 0.897 and 0.856, respectively, which were higher than those of the optimal clinical prediction model and ultrasound radiomics model, and the differences were statistically significant (both $P<0.05$). The Hosmer-Lemeshow goodness-of-fit test showed that the optimal clinical prediction model, optimal ultrasound radiomics model, and combined model all had good goodness-of-fit in both the training and testing sets ($\chi^2=0.097, 0.060, 0.806, 0.099, 0.137, 0.107$, all $P>0.05$). The calibration curve showed that the prediction curve of the optimal clinical prediction model and ultrasound radiomics model in both the training and testing sets deviate from the standard curve, while the prediction curve of the combined model was close to the standard curve, indicating the best calibration degree. The clinical decision curve analysis showed that in the training set, the optimal clinical prediction model had the highest clinical net benefit at probability thresholds of 67%~88%, the optimal ultrasound radiomics model had the highest clinical net benefit at probability thresholds of 82%~85%, and the combined model had the highest clinical net benefit at probability thresholds of 34%~98%. In the testing set, the optimal clinical prediction model had the highest clinical net benefit at probability thresholds of 43%~54% and 61%~88%, the optimal ultrasound radiomics model had the highest clinical net benefit at probability thresholds of 54%~56% and 69%~85%, and the combined model had the highest clinical net benefit at probability thresholds of 54%~92%. **Conclusion** The combined model based on TPFU parameters and ultrasound radiomics can be used to predict the occurrence of postpartum POP, which has a certain clinical value.

KEY WORDS Ultrasonography; transperineal; Levator hiatus; Radiomics; Pelvic organ prolapse

盆底功能障碍性疾病 (pelvic floor dysfunction, PFD) 是由盆底支持结构损伤或功能异常引发的一系列临床综合征, 主要表现为盆腔器官脱垂 (pelvic organ prolapse, POP)、尿失禁、排便障碍及性功能障碍

等^[1]。流行病学调查^[2]显示, 目前全球范围内 POP 患病率均处于较高水平 (25.0%~41.8%)。该病不仅导致盆底解剖结构异常和泌尿生殖系统功能障碍, 还会引发焦虑等心理问题, 显著降低患者生活质量^[3]。经

会阴盆底超声(transperineal pelvic floor ultrasound, TPFU)因其独特的动态评估优势,已成为重要的影像学检查手段。该技术可多平面实时显示静息、缩肛及最大Valsalva动作下盆底器官的三维空间位置变化,并能精准量化肛提肌裂孔(LH)等关键解剖结构的形态学参数^[4]。影像组学技术的发展为超声诊断赋予了新的维度,即通过高通量提取图像特征信息,构建可量化的生物标志物体系,为疾病预测和疗效评估提供了新思路^[5]。本研究基于TPFU图像提取和筛选的影像组学特征,并结合临床资料及TPFU参数构建联合模型,探讨其在预测产后发生POP的临床价值。

资料与方法

一、研究对象

选取2024年1~8月于南方医科大学南方医院增城院区行TPFU检查的产后女性331例,年龄19~42岁,平均 (30.45 ± 4.04) 岁。按照7:3比例将其随机分为训练集231例(POP患者191例,非POP患者40例)和测试集100例(POP患者75例,非POP患者25例)。纳入标准:①均于产后半年内完成TPFU检查;②POP符合盆腔器官脱垂定量(pelvic organ prolapse quantitation, POP-Q)分期法的诊断标准^[6],妊娠前资料完整且排除POP;③单胎足月妊娠;④近半年内未进行盆底康复相关治疗。排除标准:①新生儿为巨大儿(即出生体质量 >4 kg);②妊娠合并糖尿病、高血压及子痫等并发症;③无法通过沟通独立完成TPFU检查或有严重精神障碍者。本研究经医院医学伦理委员会批准(批准号:NFEC-2024-616),所有研究对象均签署知情同意书。

二、仪器与方法

(一)图像采集

使用GE Voluson E10彩色多普勒超声诊断仪, RM6C容积探头,频率2~6 MHz。检查前受检者排空膀胱和直肠,取膀胱截石位,将探头置于会阴部,启动4D扫描模式,嘱其行最大Valsalva动作并持续时间 ≥ 6 s,获取最大Valsalva动作下LH图像,测量其面积(LHA)、周长(C)、横径(TD)及前后径(APD),计算前后径与横径比值(APD/TD),并存取图像备用。以上操作均由2名具有4年以上TPFU工作经验的住院医师和主治医师共同完成,所有超声参数均重复测量3次取平均值。

(二)图像分割、特征提取及特征筛选

由2名具有4年以上TPFU检查经验的超声医师使用3D-Slicer软件(5.0.2)手动勾画LH感兴趣区(ROI),如边界不清晰时,由另一名具有7年以上TPFU检查经验的副主任医师共同商议确定最终ROI。见图1。采用图像去噪、灰度标准化对超声图像进行预处理。超声影像组学的所有特征均按照生物标志物标准化倡议(IBSI)采用基于Python开发的软件包Pyradiomics(v3.0.1)进行提取。提取特征后,使用Z-score归一化对提取的所有特征进行标准化,以确保分析的一致性;然后对上述归一化的特征进行 t 检验,保留 $P < 0.05$ 的特征;随后采用Spearman秩相关分析,对于具有高线性相关性($r > 0.9$)的特征仅保留一个,而后采用最小冗余最大相关性法(mRMR)、最小绝对值收缩和选择算子(LASSO)回归筛选非零系数特征作为影像组学标签(Radsignature),用于构建超声影像组学模型。

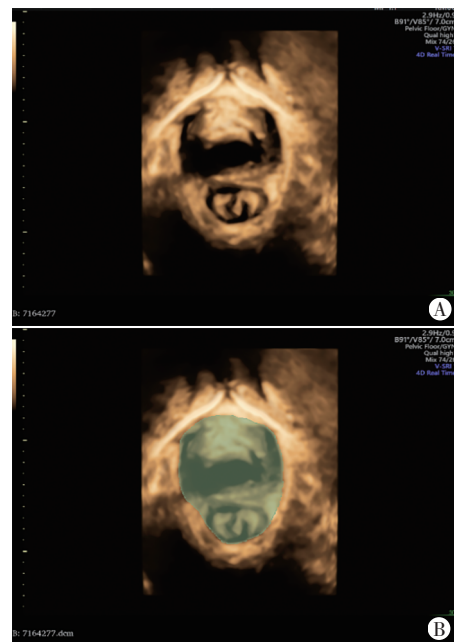


图1 最大Valsalva动作下LH图像(A)及对应的ROI勾画图(B)

(三)各模型构建

1. 临床预测模型:将训练集的临床资料和TPFU参数纳入二元Logistic回归分析,筛选预测产后发生POP的独立影响因素;随后采用Logistic回归(LR)、支持向量机(SVM)、随机森林(RF)、极度随机树(ET)、轻量级梯度提升机(LightGBM)5种机器学习算法构建临床预测模型,绘制受试者工作特征(ROC)曲线分析各模型预测产后发生POP的诊断效能,选取曲线下面积(AUC)高且泛化能力好(即AUC在训练集和测试集中差异小)的模型作为最优临床预测模型。

2. 超声影像组学模型:将筛选出的Radsignature采

用 LR、SVM、RF、ET、LightGBM 5 种机器学习算法构建超声影像组学模型,绘制 ROC 曲线分析各模型预测产后发生 POP 的诊断效能,选取 AUC 高且泛化能力好的模型作为最优超声影像组学模型。

3. 联合模型:将筛选出的产后发生 POP 的独立影响因素及 Radsignature 纳入 Logistic 回归分析,构建联合模型。

三、统计学处理

应用 Python 3.7.12、Statsmodels 0.13.2 语言进行数据分析。计量资料以 $\bar{x} \pm s$ 表示,组间比较采用 t 检验;计数资料以例或率表示,组间比较采用 χ^2 检验。绘制 ROC 曲线分析最优临床预测模型、最优超声影像组学模型和联合模型预测产后发生 POP 的诊断效能; Hosmer-Lemeshow 拟合优度检验和校准曲线评估各模型的校准度;临床决策曲线评估各模型的临床适用性;AUC 比较采用 DeLong 检验。 $P < 0.05$ 为差异有统计学意义。

结 果

一、训练集与验证集基线特征比较

训练集与验证集基线特征比较差异均无统计学意义。见表 1。

二、临床预测模型的构建

1. 单因素 Logistic 回归分析显示,年龄、分娩方式、LHA、C、TD 和 APD 均为预测产后发生 POP 的影响因

素(均 $P < 0.05$);多因素 Logistic 回归分析显示,TD 为预测产后发生 POP 的独立影响因素($OR = 1.139, P < 0.05$)。见表 2。

2. 临床预测模型的诊断效能比较:采用 5 种机器学习算法(LR、SVM、RF、ET、LightGBM)进行了临床预测模型的构建及效能评估,结果显示 RF 模型在训练集和测试集中预测产后 POP 发生的 AUC 均大于 0.80,其中 RF 模型在训练集与测试集中 AUC 变化最小,其泛化能力好,为最优临床预测模型。见表 3。

三、超声影像组学模型的构建

1. 特征筛选结果:共提取了 1561 个超声影像组学特征,包括一阶特征 306 个、形状特征 14 个和 GLCM、GLSZM、NGTDM、GLRLM、GLDM 等一系列纹理特征共 1241 个。通过 t 检验、Spearman 秩相关系数分析、mRMR 及 LASSO 回归筛选,最终筛选出 12 个非零系数的超声影像组学特征。见图 2。

2. 超声影像组学模型的诊断效能比较:采用 5 种机器学习算法进行了超声影像组学模型的构建及效能评估,结果显示各模型在训练集和测试集中预测产后发生 POP 的 AUC 均大于 0.75,其中 SVM 模型和 LightGBM 模型在训练集中预测产后发生 POP 的 AUC 分别为 0.925、0.930,但其在测试集中的 AUC 均明显下降,分别为 0.767、0.810;RF 模型在训练集与测试集中 AUC 变化最小,其泛化能力好,为最优超声影像组学模型。见表 4。

表 1 训练集与测试集基线特征比较

分类	年龄 (岁)	孕次(例)		产次(例)				分娩方式(例)		LHA (cm ²)	C (cm)	TD (cm)	APD (cm)	APD/TD
		<3 次	≥3 次	1 次	2 次	3 次	4 次	剖宫产	阴道分娩					
训练集(231)	30.52±4.07	162	69	118	90	21	2	56	175	22.16±6.33	17.41±2.51	4.57±0.59	6.12±1.09	1.34±0.19
测试集(100)	30.27±3.97	74	26	58	33	9	0	26	74	20.52±5.41	16.92±2.26	4.50±0.64	5.96±1.02	1.33±0.21
t/χ^2 值	0.525	0.511		2.162				0.041		1.970	1.671	0.992	1.215	0.967
P 值	0.600	0.475		0.539				0.840		0.049	0.096	0.322	0.225	0.779

LHA: 肛提肌裂孔面积;C: 肛提肌裂孔周长;TD: 肛提肌裂孔横径;APD: 肛提肌裂孔前后径;APD/TD: 肛提肌裂孔前后径与横径比值

表 2 二元 Logistic 回归分析结果

变量	单因素 Logistic 回归分析			多因素 Logistic 回归分析		
	OR 值	OR 值的 95% 可信区间	P 值	OR 值	OR 值的 95% 可信区间	P 值
年龄	1.013	1.004~1.022	<0.05	1.008	1.000~1.016	0.087
孕次	1.009	0.980~1.040	0.607			
产次	1.061	1.007~1.119	0.064			
分娩方式	1.213	1.117~1.317	<0.05	1.081	0.999~1.170	0.103
LHA	1.031	1.025~1.037	<0.05	1.006	0.978~1.035	0.721
C	1.077	1.063~1.092	<0.05	1.039	0.960~1.125	0.425
TD	1.336	1.267~1.409	<0.05	1.139	1.052~1.234	<0.05
APD	1.161	1.125~1.197	<0.05	0.985	0.918~1.058	0.730
APD/TD	1.198	0.993~1.445	0.114			

表 3 临床预测模型中各种机器学习算法在训练集和测试集中预测产后 POP 发生的诊断效能

机器学习算法	AUC	AUC 的 95% 可信区间	P 值	准确率 (%)	灵敏度	特异度	阳性预测值	阴性预测值
LR								
训练集	0.864	0.810~0.918	<0.001	64.1	0.576	0.950	0.982	0.319
测试集	0.808	0.714~0.902	<0.001	59.0	0.480	0.920	0.947	0.371
SVM								
训练集	0.891	0.848~0.934	<0.001	81.8	0.801	0.900	0.975	0.486
测试集	0.817	0.726~0.908	<0.001	75.0	0.787	0.640	0.868	0.500
RF								
训练集	0.845	0.793~0.896	<0.001	81.0	0.832	0.700	0.930	0.467
测试集	0.820	0.737~0.903	<0.001	79.0	0.813	0.720	0.897	0.562
ET								
训练集	0.851	0.791~0.911	<0.001	68.8	0.654	0.850	0.954	0.340
测试集	0.815	0.725~0.905	<0.001	67.0	0.613	0.840	0.920	0.420
LightGBM								
训练集	0.889	0.846~0.932	<0.001	78.8	0.764	0.900	0.973	0.444
测试集	0.820	0.732~0.907	<0.001	71.0	0.680	0.800	0.911	0.455

AUC:曲线下面积

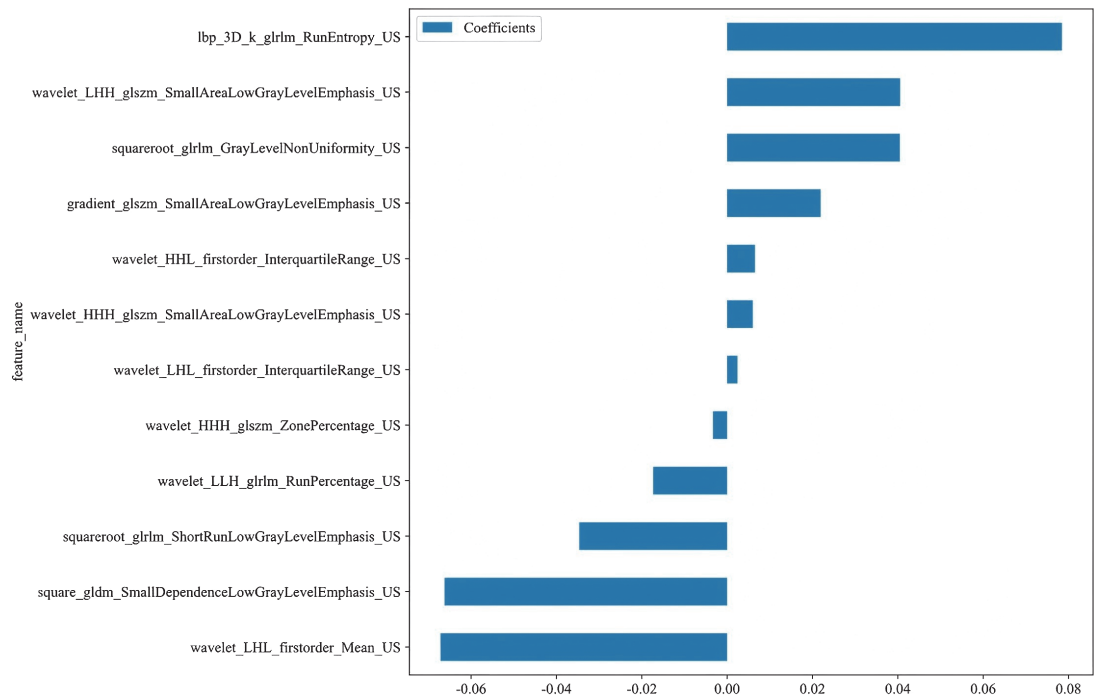


图 2 LASSO 回归最终筛选出的超声影像组学特征

四、联合模型的构建

将产后是否发生 POP 作为因变量(发生 POP=1, 未发生 POP=0),筛选的独立影响因素 TD 及 12 个影像组学特征 Radsignature 作为自变量,纳入 Logistic 回归分析构建联合模型。绘制的列线图见图 3。

五、各模型的效能及验证

1.ROC 曲线分析:最优临床预测模型、最优超声影像组学模型和联合模型在训练集中预测产后发生 POP 的 AUC 分别为 0.845(95% 可信区间:0.793~0.896)、

0.865(95% 可信区间:0.812~0.918)和 0.897(95% 可信区间:0.854~0.940),在测试集中预测产后发生 POP 的 AUC 分别为 0.820(95% 可信区间:0.727~0.903)、0.831(95% 可信区间:0.750~0.913)和 0.856(95% 可信区间:0.782~0.931);且联合模型的 AUC 高于最优临床预测模型、最优超声影像组,差异均有统计学意义(均 $P<0.05$);最优临床预测模型与最优超声影像组学 AUC 比较差异无统计学意义。见表 5 和图 4。

2.校准度分析:Hosmer-Lemeshow 拟合优度检验

表 4 超声影像组学模型中各种机器学习算法在训练集和测试集中预测产后发生 POP 的诊断效能

机器学习算法	AUC	AUC 的 95% 可信区间	P 值	准确率 (%)	灵敏度	特异度	阳性预测值	阴性预测值
LR								
训练集	0.882	0.832~0.932	<0.001	71.0	0.660	0.950	0.984	0.369
测试集	0.801	0.712~0.889	<0.001	74.0	0.720	0.800	0.915	0.488
SVM								
训练集	0.925	0.885~0.964	<0.001	87.9	0.932	0.625	0.922	0.658
测试集	0.767	0.673~0.861	<0.001	76.0	0.893	0.360	0.807	0.529
RF								
训练集	0.865	0.812~0.918	<0.001	73.2	0.712	0.825	0.951	0.375
测试集	0.831	0.750~0.913	<0.001	76.0	0.747	0.800	0.918	0.513
ET								
训练集	0.905	0.855~0.956	<0.001	81.8	0.801	0.900	0.975	0.486
测试集	0.805	0.712~0.898	<0.001	77.0	0.800	0.680	0.882	0.531
LightGBM								
训练集	0.930	0.895~0.964	<0.001	85.3	0.832	0.950	0.988	0.543
测试集	0.810	0.725~0.895	<0.001	76.0	0.800	0.640	0.870	0.516

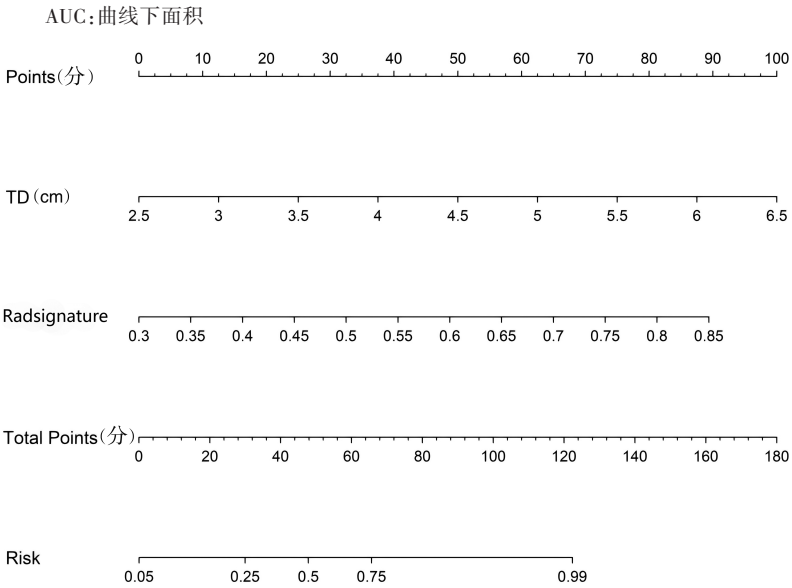


图 3 联合模型的列线图

显示,最优临床预测模型、超声影像组学模型和联合模型在训练集和测试集中的拟合度均较好($\chi^2=0.097$ 、 0.060 、 0.806 、 0.099 、 0.137 、 0.107 ,均 $P>0.05$)。校准曲线分析显示,在训练集和测试集中最优临床预测模型和超声影像组学的预测曲线均较偏离标准曲线,联合模型的预测曲线均较贴合标准曲线,其预测产后发生 POP 的校准度更高。见图 5。

3. 临床适用性:临床决策曲线分析显示,在训练集中,最优临床预测模型在概率阈值为 67%~88% 时临床净获益最高,最优超声影像组学模型在概率阈值为 82%~85% 时临床净获益最高,联合模型在概率阈值为 34%~98% 时临床净获益最高;在测试集中,最优临床预测模型在概率阈值为 43%~54% 和 61%~88% 时临床净获益最高,最优超声影像组学模型在概率阈值为 54%~56% 和 69%~85% 时临床净获益最高,联合模型在概率阈值为 54%~92%

表 5 最优临床预测模型、最优超声影像组学模型及联合模型在训练集和验证集中预测产后发生 POP 的诊断效能

方法	AUC	AUC 的 95% 可信区间	P 值	准确率 (%)	灵敏度	特异度	阳性预测值	阴性预测值
最优临床预测模型								
训练集	0.845	0.793~0.896	<0.001	81.0	0.832	0.700	0.930	0.467
测试集	0.820	0.737~0.902	<0.001	79.0	0.813	0.720	0.897	0.562
最优超声影像组学模型								
训练集	0.865	0.811~0.917	<0.001	73.2	0.712	0.825	0.951	0.375
测试集	0.831	0.750~0.912	<0.001	76.0	0.747	0.800	0.918	0.513
联合模型								
训练集	0.897	0.854~0.940	<0.001	84.8	0.901	0.600	0.915	0.562
测试集	0.856	0.781~0.930	<0.001	80.0	0.853	0.640	0.877	0.593

AUC:曲线下面积

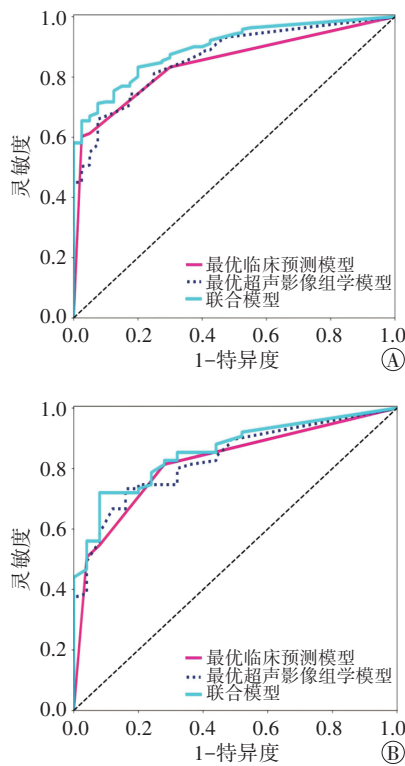


图 4 各模型在训练集(A)和测试集(B)中预测产后发生POP的ROC曲线图

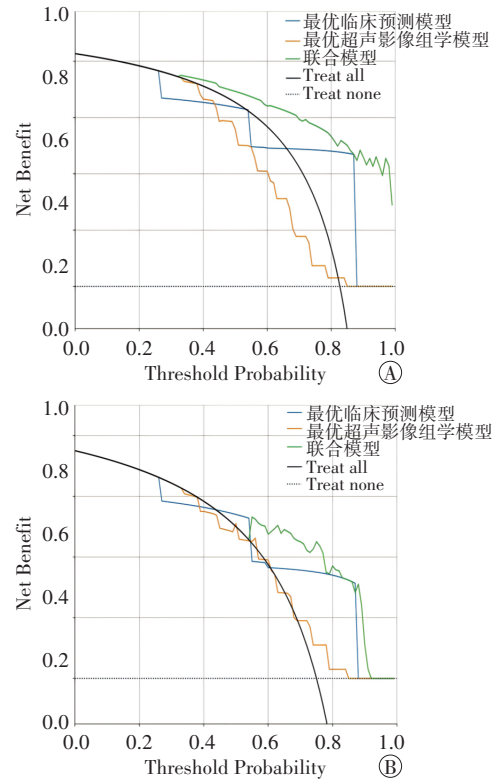


图 6 各模型在训练集(A)和测试集(B)中预测产后发生POP的临床决策曲线图

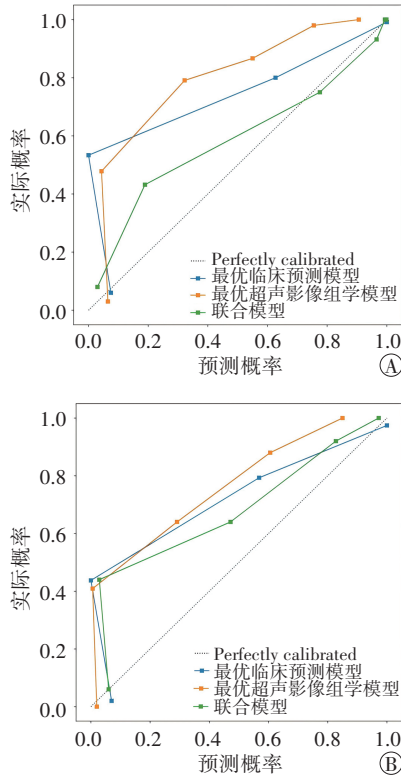


图 5 各模型在训练集(A)和测试集(B)中预测产后发生POP的校准曲线图

时临床净获益最高;其中以联合模型的临床适用性最好。见图6。

讨 论

研究^[7]显示,我国POP的患病率接近20%,其中症状性POP在成年女性中的占比约为9.6%。但约19.6%的产后患者对于脱垂症状羞于启齿而错失产后42 d至半年内的盆底康复“黄金时期”^[8]。近年来随着盆底康复知识的推广,POP逐渐受到关注。目前临床推荐将POP-Q分期法作为诊断POP的“金标准”^[6],其具有细致、客观、可靠等优点,但操作较繁琐复杂、学习周期长、难以掌握,且未涵盖阴道轴向问题,具有一定的临床局限。因此寻找更简单、有效的方法预测产后发生POP成为当前研究热点。TPFU具有操作简便且能动态成像的优点,可以实时观察LH的形态变化和扩张程度。《盆底超声检查中国专家共识(2022版)》^[9]指出,LH的扩张程度与POP紧密相关,尤其是在Valsalva动作下更为明显,LH扩张越大,POP的程度越严重。因此,本研究拟基于临床资料、TPFU参数及超声影像组学特征构建联合模型,探讨其预测产后发生POP的临床价值。

本研究单因素Logistic回归分析显示,年龄、分娩方式、LHA、C、TD和APD均为预测产后发生POP的影响因素(均 $P<0.05$);多因素Logistic回归分析显示,TD为预测产后发生POP的独立影响因素($OR=1.139$,

$P<0.05$)。其机制可能为妊娠对盆底肌的压迫,以及分娩时胎头通过产道对肛提肌产生明显横向拉伸,导致盆底肌肉纤维的过度牵拉甚至撕裂,引起 TD 扩大、盆底支撑力削弱,促使产后 POP 的发生^[10]。而其他影响因素可能因共线性关系被排除(如 LHA、C)或通过 TD 间接发挥作用如年龄及分娩方式。与既往研究^[11-13]结论一致。

机器学习是人工智能的一个重要分支,其通过计算机系统的学习和自动化推理从数据中获取知识和经验,并用于分类、回归、聚类和预测等^[14]。从学习方式上,机器学习可分为监督学习、无监督学习、半监督学习及强化学习 4 种。常见的监督学习算法有 LR、SVM、RF、ET、LightGBM 及 K 近邻(KNN)等^[15]。LR 常用于二分类问题,其通过结合线性函数和逻辑函数将特征映射到概率值上实现分类,具有简单、高效和可解释性强等优点,因此在风险评估和疾病诊断中得到广泛应用,但其对数据规模和质量要求较高,对多分类问题处理不灵活,容易出现过拟合,因此具有一定的局限性^[16]。SVM 则适用于分类和回归任务,泛化能力强,在处理高维数据和非线性问题上具有显著优势,但是对大规模数据集训练效率低,对缺失数据敏感,需要进行预处理^[17]。RF 结合了多个决策树的预测结果,适用于分类和回归任务,准确性高,抗过拟合能力强,减少了模型过拟合的风险,并且对异常值和缺失值的容忍度较高,能够有效处理数据中的噪声问题,但其可解释性较差。ET 主要处理离散型特征,其通过随机选择特征子集和使用随机阈值进行节点分裂,减少了寻找最优分割点的计算量,因此训练速度快,但是也存在可能过拟合的缺点。LightGBM 的核心原理是构建一个弱模型的集合,每一步均试图减小损失函数的梯度,具有高效、准确、灵活、低内存消耗的优点,但其对噪声敏感并具有参数调优需求^[18]。不同的机器学习算法在临床疾病诊断与预后评估中发挥着不同的效能,为了筛选出预测产后发生 POP 的最优临床预测模型,本研究将二元 Logistic 回归分析筛选的独立影响因素 TD 作为输入变量,采用 5 种机器学习算法进行了临床预测模型的构建及诊断效能评估,结果显示 SVM 模型和 LightGBM 模型在训练集中预测产后发生 POP 的 AUC 分别为 0.891、0.889,但在测试集中的 AUC 均明显下降,分别为 0.817、0.820。RF 模型在训练集与测试集之间保持了更一致的诊断效能,预测产后发生 POP 的 AUC 分别为 0.845、0.820,其 AUC 高且泛化能力好,为最优临床预测模型。分析原因为本研

究产后 POP 的样本量小且相对不均衡,而 RF 模型为 Bootstrap 抽样和随机特征选择,Bootstrap 抽样可增加数据的多样性,降低过拟合风险,随机特征选择可使模型关注不同特征的组合,避免被无关特征干扰,提升泛化能力,尤其适用于小样本的数据,因此 RF 模型在本研究预测产后发生 POP 中具有良好的临床价值,与时欣然等^[19]研究结论一致。

超声影像组学可通过深度学习从图像中提取高维特征,捕捉传统参数难以描述的细微形态学变化,从而弥补传统盆底超声的不足^[5]。本研究共提取 1561 个影像组学特征,最终筛选出 12 个非零系数特征,其中包括 9 个高阶特征和 3 个一阶特征。高阶特征是通过相关的滤波器处理及变换后得到的,可以捕捉超声图像中更复杂的纹理及局部信息,对于预测产后发生 POP 具有一定的诊断价值。本研究中预测产后发生 POP 最重要的 3 个超声影像组学特征分别为 lbp_3D_k_glrIm_RunEntropy、square_gldm_SmallDependenceLowGrayLevelEmphasis 和 wavelet_LHL_firstorder_Mean,均为高阶特征,其中 Mean 反映了超声图像中所有体素的灰度值平均值,可用于评估 LH 病变区域与周围组织之间的灰度平均值差异;RunEntropy 是运行熵,用于量化超声图像中纹理的复杂性和有序性,该值越高,表明超声图像中有序纹理的比例越低,无序纹理的比例越高,从而反映 LH 病变内部更为复杂的纹理结构;而 SmallDependenceLowGrayLevelEmphasis 是小依赖低灰度强调,用于描述超声图像中低灰度值和小依赖的联合分布,其数值越大,表明超声图像中低灰度区域的依赖更小,图像的同质化纹理越少,纹理越不均匀,其能够捕捉和量化 LH 病变局部的纹理特性,有助于识别和分析异常区域、更好地预测 POP 的发生^[20-22]。本研究采用 5 种机器学习算法对 LASSO 回归最终筛选出的超声影像组学特征进行了超声影像组学模型的构建及效能评估,结果显示 SVM 模型和 LightGBM 模型在训练集中预测产后发生 POP 的 AUC 分别为 0.925、0.930,但在测试集的 AUC 均明显下降,分别为 0.767、0.810。RF 模型在训练集与测试集中一致性高,AUC 分别为 0.865、0.831,其 AUC 高且泛化能力好,为最优超声影像组学模型。分析原因为即使通过特征筛选,预测产后 POP 发生的超声影像组学特征中仍可能包含一定的噪声,如伪影或超声设备导致的特征波动等,而 RF 模型在高维医学数据中具有对缺失值和噪声不敏感的优势^[23],能够避免模型被噪声误导,抗过拟合

能力好。

本研究将筛选的独立影响因素 TD 及 Radsignature 联合构建预测模型,结果显示联合模型在训练集和测试集中预测产后发生 POP 的 AUC、准确率、灵敏度分别为 0.897、84.8%、0.901 和 0.856、80.0%、0.853,均高于最优临床预测模型和最优超声影像组学模型,其中联合模型的 AUC 高于最优临床预测模型、最优超声影像组,差异均有统计学意义(均 $P < 0.05$)。校准曲线分析显示联合模型的校准度较高。临床决策曲线分析显示,训练集和测试集中,联合模型在概率阈值范围为 34%~98% 和 54%~92% 时的临床净获益最高。表明联合模型在预测产后发生 POP 中具有良好的临床应用价值,既弥补了最优临床预测模型仅依赖 TD 而难以观察盆底结构微损伤的局限,同时也弥补了最优影像组学模型仅依赖超声影像组学特征而忽视 TD 的不足。

综上所述,基于盆底超声参数及超声影像组学的联合模型可用于预测产后发生 POP,具有一定的临床价值。但本研究为回顾性、单中心研究,今后可联合多中心进行前瞻性研究进一步验证。

参考文献

- [1] 侯晓,李霞,孙群,等.女性盆底功能障碍性疾病诊治流程及物理康复技术临床应用——定义、流行病学、发病机制及物理康复技术概要[J].生殖医学杂志,2024,33(3):277-282.
- [2] Hadizadeh-Talasaz Z, Khadivzadeh T, Mohajeri T, et al. Worldwide prevalence of pelvic organ prolapse: a systematic review and Meta-analysis[J]. Iran J Public Health, 2024, 53(3): 524-538.
- [3] Li J, Zhao X, Li J, et al. Pelvic organ prolapse after delivery: effects on sexual function, quality of life, and psychological health[J]. J Sex Med, 2023, 20(12): 1384-1390.
- [4] Bahrami S, Khatri G, Sheridan AD, et al. Pelvic floor ultrasound: when, why, and how? [J]. Abdom Radiol (NY), 2019, 44(4): 1395-1413.
- [5] McCague C, Ramlee S, Reinius M, et al. Introduction to radiomics for a clinical audience[J]. Clin Radiol, 2023, 78(2): 83-98.
- [6] Madhu C, Swift S, Moloney-Geany S, et al. How to use the pelvic organ prolapse quantification (POP-Q) system? [J]. Neurourol Urodyn, 2018, 37(S6): S39-S43.
- [7] Chen J, Yu J, Morse A, et al. Effectiveness of self-cut vs mesh-kit titanium-coated polypropylene mesh for transvaginal treatment of severe pelvic organ prolapse: a multicenter randomized noninferiority clinical trial[J]. JAMA Netw Open, 2022, 5(9): e2231869.
- [8] 何丽,张海燕,王薇,等.重庆市产后 42 天妇女盆腔器官脱垂的危险因素及康复效果分析[J].中国妇产科临床杂志,2020,21(1): 71-72.
- [9] 中华医学会超声医学分会妇产超声学组.盆底超声检查中国专家共识(2022 版)[J].中华超声影像学杂志,2022,31(3): 185-191.
- [10] DeLancey JO. What's new in the functional anatomy of pelvic organ prolapse? [J]. Curr Opin Obstet Gynecol, 2016, 28(5): 420-429.
- [11] Youssef A, Del MA, Nedu B, et al. Feasibility and reproducibility of new technique for measurement of transverse diameter of levator ani muscle hiatus using two-dimensional transperineal ultrasound in nulliparous women with term pregnancy [J]. Ultrasound Obstet Gynecol, 2024, 63(6): 824-832.
- [12] 吴凤姬,陈梅.经会阴三维超声测量肛提肌裂孔面积及形态诊断产后子宫脱垂的价值[J].现代医用影像学,2024,33(6): 1099-1102.
- [13] O'Leary BD, Keane DP. Effect of the length of the second stage of labor on pelvic floor dysfunction [J]. Am J Obstet Gynecol MFM, 2023, 5(2): 100795.
- [14] Greener JG, Kandathil SM, Moffat L, et al. A guide to machine learning for biologists [J]. Nat Rev Mol Cell Biol, 2022, 23(1): 40-55.
- [15] Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer [J]. Behav Ther, 2020, 51(5): 675-687.
- [16] Wang QQ, Yu SC, Qi X, et al. Overview of Logistic regression model analysis and application [J]. Zhonghua Yu Fang Yi Xue Za Zhi, 2019, 53(9): 955-960.
- [17] Valkenburg D, Rousseau AJ, Geubbelmans M, et al. Support vector machines [J]. Am J Orthod Dentofacial Orthop, 2023, 164(5): 754-757.
- [18] Hu X, Yin S, Zhang X, et al. Blood pressure stratification using photoplethysmography and light gradient boosting machine [J]. Front Physiol, 2023, 14: 1072273.
- [19] 时欣然,庞震,乔婷,等.基于机器学习的女性压力性尿失禁发病风险预测模型建立及效能评价[J].现代泌尿外科杂志,2025,30(3): 196-206.
- [20] Chun SH, Suh YJ, Han K, et al. Differentiation of left atrial appendage thrombus from circulatory stasis using cardiac CT radiomics in patients with valvular heart disease [J]. Eur Radiol, 2021, 31(2): 1130-1139.
- [21] Bicci E, Calamandrei L, Di Finizio A, et al. Predicting response to exclusive combined radio-chemotherapy in naso-oropharyngeal cancer: the role of texture analysis [J]. Diagnostics (Basel), 2024, 14(10): 1036-1052.
- [22] Huang ZN, Zhang HX, Sun YQ, et al. Multi-cohort study in gastric cancer to develop CT-based radiomic models to predict pathological response to neoadjuvant immunotherapy [J]. J Transl Med, 2025, 23(1): 362-378.
- [23] Gao D, Feng W, Qiao Y, et al. Development and validation of a random forest model to predict functional outcome in patients with intracerebral hemorrhage [J]. Neurol Sci, 2023, 44(10): 3615-3627.

(收稿日期:2024-12-04)