

基于多模态超声指标和血清学指标的随机森林模型预测乙型病毒性肝炎相关性高危食管胃底静脉曲张的临床价值

冯斯奕, 黄玉洁, 刘广文, 程璟, 彭素好, 涂海斌

摘要 目的 基于多模态超声指标和血清学指标构建随机森林(RF)模型,探讨其预测乙型病毒性肝炎(以下简称乙肝)相关性高危食管胃底静脉曲张(EGV)的临床价值。方法 选取于我院接受胃镜检查的乙肝患者 317 例,根据 7:3 比例将其随机分为训练集 221 例(高危 EGV 病例 88 例,低危 EGV 病例 133 例)和验证集 96 例(高危 EGV 病例 39 例,低危 EGV 病例 57 例),比较训练集中高危 EGV 病例与低危 EGV 病例多模态超声指标和血清学指标的差异;通过 RF 算法递归特征消除(RFE)对变量的重要性进行排序,筛选出显著提升预测高危 EGV 效能的指标,构建 RF 模型,并使用 SHAP 对其进行可视化展示。绘制受试者工作特征(ROC)曲线分析 RF 模型及传统模型包括纤维化-4 指数(FIB-4)、血小板计数/脾脏长径比值(PSR)、肝脏硬度进展评分(LSPS)和终末期肝病(MELD)的诊断效能;绘制梯度提升曲线评估各模型在不同样本数量下的预测效能提升情况;预测分布图评估各模型对高危 EGV 病例与低危 EGV 病例的预测概率分布;利润曲线评估各模型在不同阈值下的潜在临床效益;累积获益曲线评估各模型在不同样本比例下识别高危 EGV 病例的能力。结果 与低危 EGV 病例比较,高危 EGV 病例年龄偏大,脾脏硬度(SSM)、脾脏长径及厚径、肌酐(Cr)、总胆红素(TBIL)、直接胆红素(DBIL)、间接胆红素(IBIL)、谷丙转氨酶(ALT)、血小板计数/脾脏长径均显著升高,肝血流量、门静脉流速、血小板计数(PLT)、白蛋白(ALB)均显著降低,且更容易形成侧支循环,差异均有统计学意义(均 $P < 0.05$)。根据 RFE 对上述差异有统计学意义的指标进行排序,排名前 10 的指标依次为 SSM、PLT、Cr、ALB、脾脏厚径、DBIL、门静脉流速、ALT、肝血流量和脾脏长径。当变量数目从 1 个逐渐增加至 7 个时,模型的曲线下面积(AUC)从 0.752 上升至 0.924;继续增加变量数目至 10 个时,AUC 仅小幅度增加,故选取排名前 7 的指标构建 RF 模型。RF、FIB-4、PSR、LSPS、MELD 模型在训练集中预测高危 EGV 的 AUC 分别为 0.97[95% 可信区间(CI):0.95~0.99]、0.47(95%CI:0.39~0.54)、0.61(95%CI:0.53~0.67)、0.45(95%CI:0.38~0.53)和 0.47(95%CI:0.39~0.55);在验证集中的 AUC 分别为 0.88(95%CI:0.81~0.94)、0.55(95%CI:0.43~0.66)、0.57(95%CI:0.45~0.68)、0.50(95%CI:0.39~0.63)和 0.51(95%CI:0.39~0.63);RF 模型在训练集、验证集中的 AUC 均显著高于其他模型,差异均有统计学意义(均 $P < 0.001$)。梯度提升曲线分析显示,RF 模型在训练集和验证集中的梯度提升值均高于其他模型。预测分布图分析显示,RF 模型在训练集和验证集中均能较好地地区分高危 EGV 病例与低危 EGV 病例;而 PSR、MELD、LSPS 和 FIB-4 模型的预测概率分布曲线重叠较多,表明这些模型难以有效区分二者。利润曲线分析显示,RF 模型在训练集和验证集中均能够在较低阈值(训练集 0.2,验证集 0.3)时保持较高的利润;PSR、MELD、LSPS 和 FIB-4 模型的利润曲线均低于 RF 模型,表明这些模型在识别高危 EGV 病例方面的效益较低。累积获益曲线分析显示,在训练集和验证集中,RF 模型的累积获益曲线明显优于传统模型。结论 RF 模型对乙肝相关性高危 EGV 的预测效能优于传统的 FIB-4、PSR、LSPS 和 MELD 模型,可作为其无创预测的有效工具,为临床决策提供参考。

关键词 超声检查;食管胃底静脉曲张;血清学指标;预测模型;随机森林

[中图法分类号]R445.1;R543.6

[文献标识码]A

Clinical value of random forest model based on multimodal ultrasonic indicators and serological markers in predicting high-risk esophagogastric varices related to hepatitis-B virus

基金项目:福建省自然科学基金项目(2022J011285、2023J011480)

作者单位:福建医科大学孟超肝胆医院超声医学科,福建 福州 350025

通讯作者:涂海斌,Email:thb861126@163.com

FENG Siyi, HUANG Yujie, LIU Guangwen, CHENG Jing, PENG Suyu, TU Haibin

Department of Ultrasound Medicine, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou 350025, China

ABSTRACT Objective To investigate the clinical value of a random forest (RF) model based on multimodal ultrasonic indicators and serological markers in predicting high-risk esophagogastric varices (EGV) related to hepatitis-B virus (HBV).

Methods A total of 317 HBV-infected patients who underwent gastroscopy at our hospital were enrolled. They were randomly divided into a training set of 221 cases (88 cases of high-risk EGV and 133 cases of low-risk EGV) and a validation set of 96 cases (39 cases of high-risk EGV and 57 cases of low-risk EGV) according to a ratio of 7:3. The differences in multimodal ultrasonic indicators and serological markers between high- and low-risk EGV patients in the training set were compared. The importance of variables were ranked by recursive feature elimination (RFE) algorithm of the RF to screen the indicators which significantly improve the prediction performance of high-risk EGV. A RF model was constructed and visualized by SHAP. Receiver operating characteristic (ROC) curve was drawn to analyze the diagnostic performance of the RF model and traditional models, including FIB-4, platelet count/splenic diameter ratio (PSR), liver stiffness progression score (LSPS), and end-stage liver disease (MELD). Gradient boosting curve was drawn to assess the prediction performance at different number samples. Prediction distribution map was drawn to assess the prediction probabilities for high-risk and non-high-risk EGV cases. Profit curve was drawn to evaluate the potential clinical benefits at different thresholds. Cumulative benefit curve was drawn to analyze the models' ability to identify high-risk EGV cases at different sample proportions.

Results Compared with low-risk EGV cases, the high-risk EGV cases were older and exhibited significantly higher spleen stiffness (SSM), spleen length and thickness, serum creatinine (Cr), total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin (IBIL), alanine aminotransferase (ALT), and platelet count-to-spleen artery ratio, while hepatic blood flow, portal vein velocity, platelet count (PLT), and albumin (ALB) were significantly decreased, with a higher occurrence of collateral circulation, and the differences were statistically significant (all $P < 0.05$). Rank the above statistically significant indicators according to RFE. The top ten indicators were SSM, PLT, Cr, ALB, spleen thickness, DBIL, portal vein velocity, ALT, hepatic blood flow, and spleen length diameter. When the number of variables gradually increased from 1 to 7, the area under the curve (AUC) of the model rose from 0.752 to 0.924. When the number of variables was further increased to 10, the AUC was increased slightly. Therefore, the top 7 ranked indicators were selected to construct the RF model. The AUC of the RF, FIB-4, PSR, LSPS, and MELD models in predicting high-risk EGV in the training set were 0.97 [95% confidence interval (CI): 0.95~0.99], 0.47 (95%CI: 0.39~0.54), 0.61 (95%CI: 0.53~0.67), 0.45 (95%CI: 0.38~0.53), and 0.47 (95%CI: 0.39 - 0.55), respectively, the AUC in the validation set were 0.88 (95%CI: 0.81~0.94), 0.55 (95%CI: 0.43~0.66), 0.57 (95%CI: 0.45~0.68), 0.50 (95%CI: 0.39~0.63), and 0.51 (95%CI: 0.39~0.63), respectively. The AUC of the RF model in both the training set and the validation set were significantly higher than that of the other models, the differences were statistically significant (all $P < 0.001$). The gradient boosting curve showed that the gradient boosting values of the RF model in both the training set and the validation set were higher than that of other models. The prediction distribution map showed that the RF model can well distinguish high-risk from low-risk EGV cases in both the training set and the validation set. However, the prediction probability distribution curve of the PSR, MELD, LSPS, and FIB-4 models overlap significantly, indicating that these models had difficulty effectively differentiating the two. The profit curve showed that the RF model could maintain a high profit within a low threshold range (0.2 for the training set and 0.3 for the validation set) in both the training set and the validation set. The profit curve of the PSR, MELD, LSPS, and FIB-4 models were significantly lower than that of the RF model, indicating that these models were less effective in identifying high-risk EGV cases. The cumulative profit curve showed that the RF model has the highest profit in both the training set and the validation set remains at a high level compared with that in PSR, MELD, LSPS, and FIB-4 models.

Conclusion The RF model has better prediction performance for hepatitis B-related high-risk EGV than that of traditional FIB-4, PSR, LSPS, and MELD models. It can serve as an effective tool for non-invasive assessment, providing a reference for clinical prevention.

KEY WORDS Ultrasonography; Esophagogastric varices; Serological markers; Predictive model; Random forest

乙型病毒性肝炎(以下简称乙肝)相关性食管胃底静脉曲张(esophagogastric varices, EGV)作为慢性乙肝病程中危及生命的严重并发症^[1],其早期识别及干预对改善患者预后具有重要临床价值。胃镜检查是诊断 EGV 的金标准,但其侵入性和有创性导致临床应用受限^[2]。研究^[3]显示,基于超声图像或血清学标志物的单一模态预测模型在 EGV 风险评估中存在灵敏度和特异度不足的局限。在此背景下,机器学习算法,尤其是随机森林(random forest, RF)算法凭借其自适应特征选择能力和复杂非线性关系的解析优势^[4],致力于开发具有临床实用价值的无创筛查工具。本研究基于多模态超声指标和血清学指标构建 RF 模型,探讨其预测乙肝相关性高危 EGV 的临床价值,旨在实现高危 EGV 患者的精准分层,为优化临床决策、降低门静脉高压相关并发症风险及医疗成本提供依据,同时也为人工智能技术在肝病并发症预测领域的应用转化提供新的研究思路。

资料与方法

一、研究对象

选取 2019 年 1 月至 2024 年 3 月于我院接受胃镜检查的乙肝患者 317 例,男 232 例,女 85 例,年龄 28~73 岁,平均(53.5±12.3)岁。根据 7:3 比例将其随机分为训练集 221 例和验证集 96 例。纳入标准:①经血清学检测证实乙型肝炎病毒表面抗原(HBsAg)阳性 6 个月以上;②均接受胃镜、超声和实验室检查,临床资料完整;③年龄 18~75 周岁。排除标准:①合并遗传学、代谢性、药物性、自身免疫性肝病或肝脏恶性肿瘤;②行胃镜或介入等 EGV 治疗(如套扎、经颈静脉肝内门体静脉分流术、硬化治疗等);③脾脏厚径<2 cm 且多次测量脾脏硬度(SSM)不成功;④心、肾功能异常。本研究经我院医学伦理委员会批准(批准号:2022_028_01),所有患者均签署知情同意书。

二、仪器与方法

1. 多模态超声指标检测:使用西门子 Acuson Sequoia 彩色多普勒超声诊断仪,5C-1 腹部凸阵探头,频率 2~6 MHz。检查前患者空腹 8 h,首先行常规腹部超声获取门静脉内径(于距肝门约 1 cm 处测量)、门静脉流速(于距主干分叉处约 1 cm 处测量)、脾脏长径及厚径(当厚径>4 cm 时,判定为脾肿大)、脾静脉内径、脾静脉流速(于脾静脉胰后段脾侧测量)、肝血流量、脾血流量、脾血流量/肝血流量、侧支循环情况(探及胃左静脉、脐静脉、脾肾分流静脉中任一血管时,判定为

侧支循环形成);随后开启实时二维剪切波弹性成像(two-dimensional shear wave elasticity, 2D-SWE)模式,避开气体、大血管、胆道及肋骨等组织结构,设置取样框大小为 4 cm×3 cm,上缘置于肝包膜下 2 cm,深度不超过 6 cm,设定直径为 2.0 cm 感兴趣区,弹性测量量程为 0~40 kPa,嘱患者屏住呼吸 3~5 s(避免深吸气后屏气),测量肝脏硬度(LSM)。将探头置于脾脏最厚处,设置取样框大小为 3 cm×2 cm,上缘置于脾包膜下 1 cm,设定直径为 1.0 cm 感兴趣区,弹性测量量程为 0~50 kPa,测量 SSM。LSM 和 SSM 均于取样框内不同部位连续测量 5 次,取中位数作为最终结果。以上操作均由 10 年以上工作经验的超声医师完成。

2. 血清学指标检测:收集白细胞计数(WBC)、红细胞计数(RBC)、血小板计数(PLT)、谷丙转氨酶(ALT)、谷草转氨酶(AST)、碱性磷酸酶(ALP)、白蛋白(ALB)、总胆红素(TBIL)、直接胆红素(DBIL)、间接胆红素(IBIL)、凝血酶原时间(PT)、国际标准化比值(INR)、肌酐(Cr)。

3. 传统模型:包括肝纤维化(FIB)-4 模型[(年龄×AST)/(PLT×ALT)]^[5]、PSR 模型(PLT/脾脏长径)^[6]、肝脏硬度进展评分(LSPS)模型(LSM×脾脏长径/PLT)^[7]及终末期肝病(MELD)模型[MELD=3.8×Ln(TBIL)+11.2×Ln(INR)+9.6×Ln(Cr)+6.4(病因:胆汁性或酒精性赋值为 0,其他赋值为 1)]^[8]。

4. EGV 分级标准:依据《肝硬化门静脉高压食管胃静脉曲张出血的防治指南》^[9]将其分为:轻度,食管静脉为直线或者略有迂曲,无红色征;中度,食管静脉为直线或者略有迂曲,有红色征(即曲线静脉表面红斑、红色条纹和血疱)或者蛇形迂曲隆起;重度,曲张的静脉为蛇形迂曲隆起且有红色征或表现为串珠状、结节状或瘤状。其中,重度、中度 EGV 归为高危 EGV 病例;轻度 EGV、无静脉曲张归为低危 EGV 病例。本研究训练集中高危 EGV 病例 88 例,低危 EGV 病例 133 例;验证集中高危 EGV 病例 39 例,低危 EGV 病例 57 例。

5. 临床资料收集:通过电子病历系统获取患者性别、年龄及 Childpugh 分级。

三、统计学处理

应用 Python 3.6 和 R 4.1 语言进行数据分析。计量资料先进行正态性和方差齐性检验,符合正态分布及方差齐者以 $\bar{x} \pm s$ 表示,采用两独立样本 t 检验;不符合正态分布或方差不齐者以 $M(Q_1, Q_3)$ 表示,采用 Mann-Whitney U 检验。计数资料以频数或率表示,采用 χ^2 检验或 Fisher 精确检验。采用 RF 算法递归特征

消除(RFE)对特征重要性进行排序,具体方法:通过构建多棵决策树评估每个特征的重要性,并计算每个特征在所有树中的平均重要性得分;然后根据得分对特征进行降序排序,筛选出可显著提升预测效能的关键特征;随后依据特征重要性排序,逐步纳入不同数量的特征构建预测模型,以确定最优特征组合。绘制受试者工作特征(ROC)曲线分析各模型预测乙肝相关高危 EGV 的诊断效能;曲线下面积(AUC)比较采用 DeLong 检验。绘制梯度提升曲线评估各模型在不同样本数量下的预测效能提升情况;预测分布图评估各模型对高危 EGV 病例与低危 EGV 病例的预测概率分布;利润曲线评估各模型在不同阈值下的潜在临床效益;累积获益曲线评估各模型在不同样本比例下识别高危 EGV 病例的能力。 $P<0.05$ 为差异有统计学意义。

结 果

一、训练集与验证集基线资料比较

训练集与验证集基线资料比较差异均无统计学意义,具有可比性。见表 1。

二、训练集中高危 EGV 病例与低危 EGV 病例基线资料比较

与低危 EGV 病例比较,高危 EGV 病例年龄偏大,SSM、脾脏长径及厚径、Cr、TBIL、DBIL、IBIL、ALT、PSR 均显著升高,肝血流量、门静脉流速、PLT、ALB 均显著降低,且更容易形成侧支循环,差异均有统计学意义(均 $P<0.05$);二者在性别、LSM、门静脉内径、门静脉流速、脾静脉内径、脾静脉流速、脾血流量/肝血流量、AST、ALP、RBC、WBC、PT、INR、FIB-4、LSPS、MELD、Childpugh 分级比较差异均无统计学意义。见表 2。不同程度 EGV 的 SSM、门静脉流速及肝血流量测量图见图 1~3。

三、RF 模型构建

根据 RFE 对上述差异有统计学意义的指标进行排序,排名前 10 的特征依次为 SSM、PLT、Cr、ALB、脾脏厚径、DBIL、门静脉流速、ALT、肝血流量、脾脏长径。见图 4。为秉承模型简约化,本研究拟采用尽量少的维度,达到最佳预测效果,依据重要性排序,分别纳入了不同数目的变量构建预测模型,并计算出对应的 AUC,结果显示当变量数目为 7 个时,AUC 达到 0.924,继续增加变量数目至 10 个时,AUC 仅小幅度增加,故选取排名前 7 的指标构建 RF 模型。见图 5。为解决机器学习预测过程难以理解的“黑箱”过程,本研究采用

表 1 训练集与验证集基线资料比较

指标	训练集(221)	验证集(96)	χ^2/t 值	<i>P</i> 值
高危 EGV(例)				
否	133	57	0.02	0.90
是	88	39		
年龄(岁)	53.6±12.5	53.4±12.0	0.27	0.79
性别(例)				
女	66	19	3.21	0.07
男	155	77		
LSM(kPa)	13.0±4.8	13.9±5.6	1.29	0.20
SSM(kPa)	14.9±4.8	15.7±5.0	1.93	0.06
门静脉内径(cm)	1.2±0.2	1.2±0.2	1.00	0.32
门静脉流速(cm/s)	27.6±7.0	27.1±6.0	0.16	0.87
脾脏长径(cm)	13.3±2.7	13.1±2.5	0.29	0.77
脾脏厚径(cm)	4.6±1.0	4.6±0.8	0.45	0.65
脾静脉内径(cm)	0.8±0.2	0.8±0.2	0.25	0.80
脾静脉流速(cm/s)	34.9±8.9	36.0±9.8	0.57	0.57
肝血流量(ml/s)	34.5±19.9	33.3±12.3	0.39	0.70
脾血流量(ml/s)	20.3±13.1	20.5±12.5	0.57	0.57
脾血流量/肝血流量	0.7±0.6	0.7±0.6	0.47	0.64
PLT(×10 ⁹ /L)	136.5±67.3	126.2±60.2	1.02	0.31
Cr(μmol/L)	76.1±30.0	74.4±17.7	0.74	0.46
AST(U/L)	40.7±44.7	32.1±17.6	0.77	0.44
ALT(U/L)	51.6±71.6	38.9±25.9	0.88	0.38
TBIL(μmol/L)	30.8±38.0	28.8±23.3	0.71	0.48
DBIL(μmol/L)	16.8±30.6	14.2±19.0	0.24	0.81
IBIL(μmol/L)	13.9±10.5	14.7±7.9	1.56	0.12
ALP(U/L)	110.5±60.2	104.1±42.7	0.04	0.97
RBC(×10 ¹² /L)	4.3±0.8	4.4±0.8	1.08	0.28
WBC(×10 ⁹ /L)	5.3±1.8	5.5±2.0	0.61	0.54
PT(s)	15.4±3.1	15.6±3.3	0.57	0.57
ALB(g/L)	40.0±9.2	39.3±8.0	0.57	0.57
INR	1.2±0.3	1.3±0.4	0.18	0.86
FIB-4	0.9±1.4	0.8±0.9	1.02	0.31
PSR	10.6±6.1	9.9±5.2	0.77	0.44
LSPS	1.6±1.3	1.9±1.5	1.48	0.14
MELD	30.2±2.0	30.3±1.6	0.94	0.35
Childpugh 分级(例)				
1 级	116	59	2.05	0.36
2 级	88	31		
3 级	17	6		
侧支循环(例)				
无	175	73	0.34	0.56
有	46	23		

SHAP 分析 RF 模型预测高危 EGV 的特征重要性,并对预测过程进行了可视化展示。见图 6。SHAP 值越高,表示该特征对高危 EGV 预测的贡献越大,本研究结果显示 SSM、PLT、Cr、ALB、脾脏厚径、DBIL 和门静脉流

表2 训练集中高危EGV病例与低危EGV病例基线资料比较

指标	低危EGV病例 (133)	高危EGV病例 (88)	χ^2/t 值	P值
年龄(岁)	52.3±13.4	55.9±10.0	2.41	0.02
性别(例)				
女	44	22	2.85	0.25
男	89	66		
LSM(kPa)	13.2±5.7	13.6±4.6	1.14	0.26
SSM(kPa)	13.7±4.4	17.2±4.5	5.67	<0.01
门静脉内径(cm)	1.2±0.2	1.2±0.2	1.04	0.30
门静脉流速(cm/s)	28.0±5.9	15.8±6.0	2.78	0.01
脾脏长径(cm)	10.6±2.8	13.1±2.1	2.42	0.03
脾脏厚径(cm)	3.6±0.9	4.9±0.9	3.21	<0.01
脾静脉内径(cm)	0.9±0.2	0.8±0.2	1.62	0.12
脾静脉流速(cm/s)	36.4±9.0	34.3±9.3	1.94	0.05
肝血流量(ml/s)	35.2±16.7	20.4±10.2	2.33	0.02
脾血流量(ml/s)	22.4±14.1	18.0±10.1	2.22	0.03
脾血流量/肝血流量	0.7±0.5	0.7±0.6	0.60	0.55
PLT(×10 ⁹ /L)	148.0±75.1	92.1±49.3	2.31	0.02
Cr(μmol/L)	75.7±35.1	98.7±12.6	3.12	<0.01
AST(U/L)	47.4±63.9	46.0±57.6	1.42	0.16
ALT(U/L)	29.0±42.0	41.1±13.9	2.81	0.02
TBIL(μmol/L)	27.9±38.7	33.7±28.9	3.89	<0.01
DBIL(μmol/L)	10.5±3.12	18.6±2.43	4.02	<0.01
IBIL(μmol/L)	13.4±9.7	14.8±7.4	2.28	0.02
ALP(U/L)	104.8±53.7	112.2±57.4	0.79	0.43
RBC(×10 ¹² /L)	4.3±0.9	4.2±0.7	0.55	0.60
WBC(×10 ⁹ /L)	5.5±2.1	5.2±1.7	1.42	0.16
PT(s)	15.4±3.1	15.8±3.2	1.49	0.14
ALB(g/L)	40.7±8.9	25.6±8.7	6.25	<0.01
INR	1.2±0.3	1.3±0.4	1.36	0.18
FIB-4	1.0±1.6	0.7±0.6	0.01	1.00
PSR	9.7±6.4	10.8±4.3	3.18	<0.01
LSPS	2.0±1.6	1.4±0.7	1.43	0.16
MELD	30.4±2.2	30.0±1.4	0.50	0.62
Childpugh 分级(例)				
1 级	77	39	5.11	0.08
2 级	49	39		
3 级	7	10		
侧支循环(例)				
无	123	52	6.25	<0.01
有	10	36		

速对 RF 模型预测高危 EGV 病例的贡献较大。

四、各模型预测效能验证

1. ROC 曲线分析: FIB-4、PSR、LSPS 和 MELD 模型在训练集中预测高危 EGV 的 AUC 分别为 0.47[(95% 可信区间(CI): 0.39~0.54)]、0.61(95%CI: 0.53~0.67)、0.45(95%CI: 0.38~0.53) 和 0.47(95%CI: 0.39~0.55); 在验

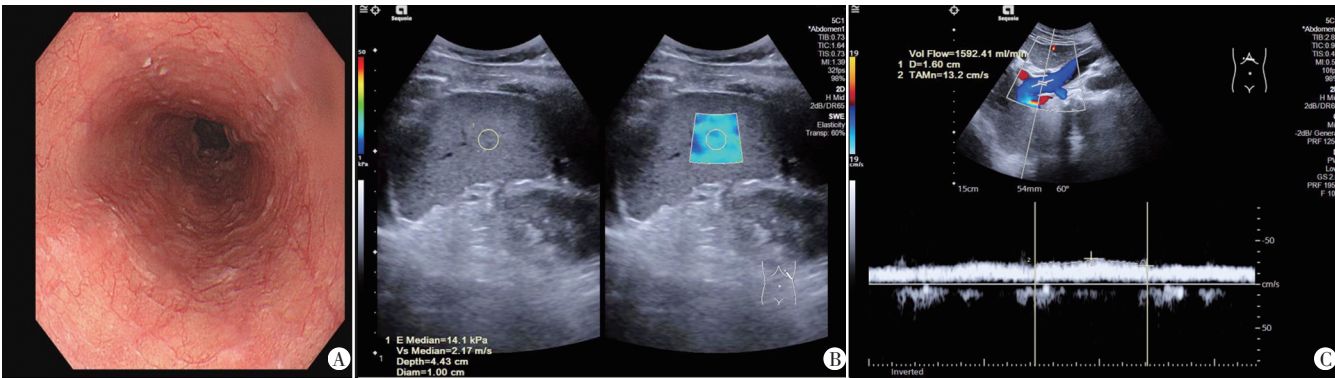
证集中的 AUC 分别为 0.55(95%CI: 0.43~0.66)、0.57(95%CI: 0.45~0.68)、0.50(95%CI: 0.39~0.63) 和 0.51(95%CI: 0.39~0.63), RF 模型在训练集、验证集中的 AUC 分别为 0.97(95%CI: 0.95~0.99)、0.88(95%CI: 0.81~0.94), 均显著高于其他模型, 差异均有统计学意义(均 $P<0.001$)。见图 7。

2. 梯度提升曲线分析: RF 模型的梯度提升曲线始终位于最高点, 表明 RF 模型在训练集和验证集中均表现出最佳的预测效能。在训练集中, 随着样本量的增加, RF 模型的提升值逐渐减小, 但总体上保持在较高水平; 相比之下, MELD、PSR、LSPS 和 FIB-4 模型的提升曲线均低于 RF 模型, 且在样本量较少时波动较大, 表明这些模型在预测高危 EGV 病例方面的效能相对较差。在验证集中, MELD、PSR、LSPS 和 FIB-4 模型的提升曲线在样本量较少时也表现出较大的波动, 且整体提升值低于 RF 模型。见图 8。

3. 预测分布图分析: RF 模型在训练集中能够较好地地区分阳性病例(高危 EGV 病例)与阴性病例(低危 EGV 病例), 阳性病例的预测概率集中在较高区间, 阴性病例的预测概率集中在较低区间, 表明 RF 模型具有较强的区分能力; 而 PSR、MELD、LSPS 和 FIB-4 模型的预测概率分布曲线重叠较多, 表明这些模型难以有效区分阳性病例与阴性病例。在验证集中, RF 模型的区分能力略有下降, 但仍明显优于其他模型, 表明 RF 模型具有较好的泛化能力。见图 9。

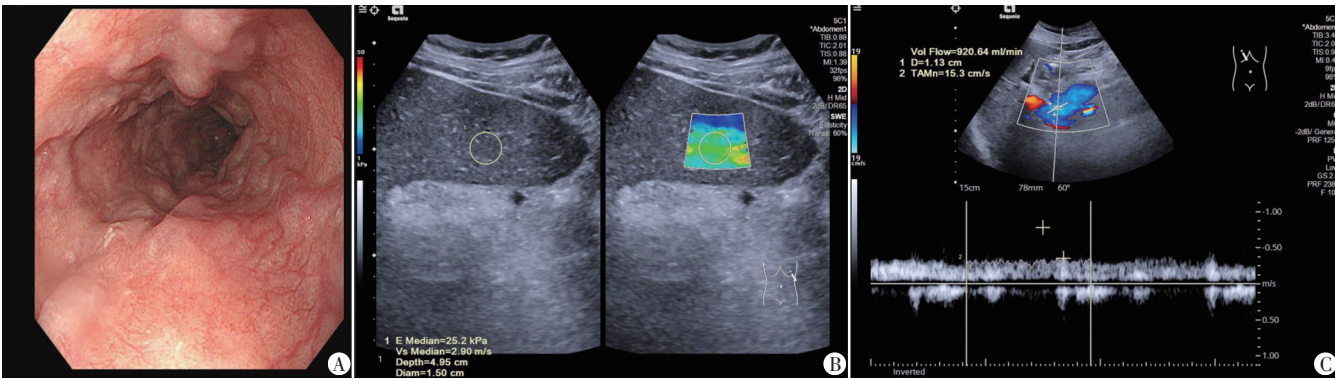
4. 利润曲线分析: RF 模型在训练集和验证集中均表现出色, 能够在较低阈值范围内保持较高的利润。训练集中, 当预测概率阈值为 0.2 时, RF 模型能够识别出 90% 的高危 EGV 病例, 同时误判率较低, 从而获得较高的利润; 验证集中, 当预测概率阈值为 0.3 时, RF 模型也能达到较高的利润水平, 可识别出 85% 的高危 EGV 病例。RF 模型的利润曲线始终位于最高点, 表明 RF 模型在训练集和验证集中的利润最大, 尤其在低阈值范围内(0.0~0.8)利润保持在高水平, 随着概率阈值增加, 利润逐渐下降。PSR、MELD、LSPS 和 FIB-4 模型的利润曲线均低于 RF 模型, 表明这些模型在识别高危 EGV 病例方面的效益较低。见图 10。

5. 累积获益曲线分析: 训练集中, RF 模型的累积获益曲线明显优于传统模型, 抽取前 20% 的样本即可识别出约 60% 的高危 EGV 病例, 抽取前 40% 的样本可识别出约 85% 的高危 EGV 病例, 体现出极高的识别率; PSR 模型表现次之, 抽取前 40% 的样本时仅能识别出约 45% 的高危 EGV 病例; LSPS、MELD 和 FIB-4



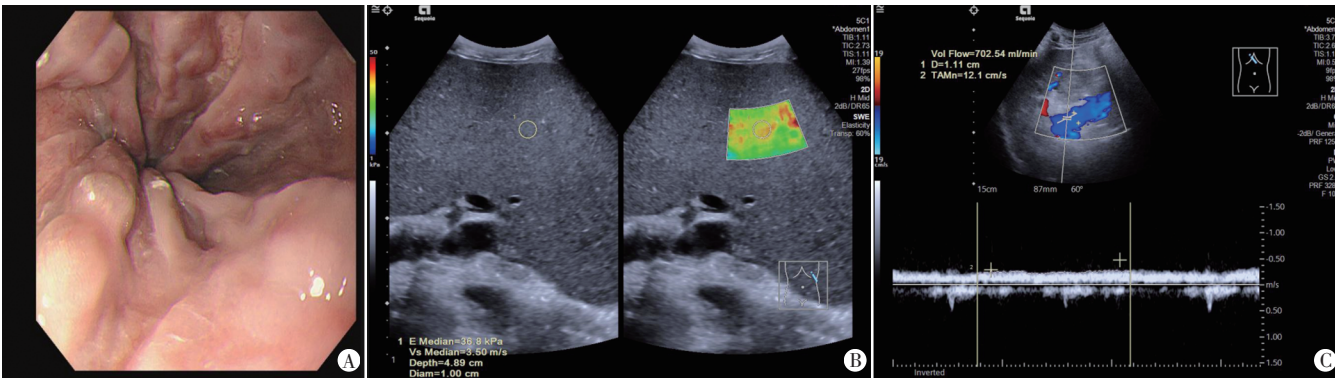
A: 胃镜显示轻度 EGV; B: SSM 为 14.1 kPa; C: 门静脉流速 13.2 cm/s, 肝血流量 1592.41 ml/min

图 1 低危 EGV 病例(男, 42 岁)胃镜及超声图像



A: 胃镜显示中度 EGV; B: SSM 为 25.2 kPa; C: 门静脉流速 15.3 cm/s, 肝血流量 920.64 ml/min

图 2 高危 EGV 病例(男, 58 岁)胃镜及超声图像



A: 胃镜显示重度 EGV; B: SSM 为 36.8 kPa; C: 门静脉流速 12.1 cm/s, 肝血流量 702.54 ml/min

图 3 高危 EGV 病例(男, 65 岁)胃镜及超声图像

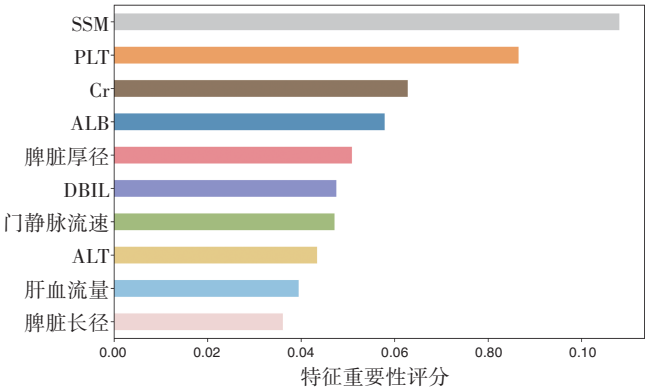


图 4 根据 RFE 排名前 10 的特征

模型的累积获益更低,抽取前 40% 样本时识别率分别约为 28%、28% 和 26%。验证集中,RF 模型同样保持了优异表现,虽然累积获益率较训练集略有下降,但仍明显优于其他模型;RF 模型在抽取前 20% 和 40% 样本时可分别识别出约 45%、80% 的高危 EGV 病例。其他传统模型在验证集中表现与训练集相似,其中 PSR 模型表现相对较好,但在抽取前 40% 样本时识别率约 42%;而 LSPS、MELD 和 FIB-4 模型的识别率更低,抽取前 40% 样本时识别率均低于 40%。见图 11。

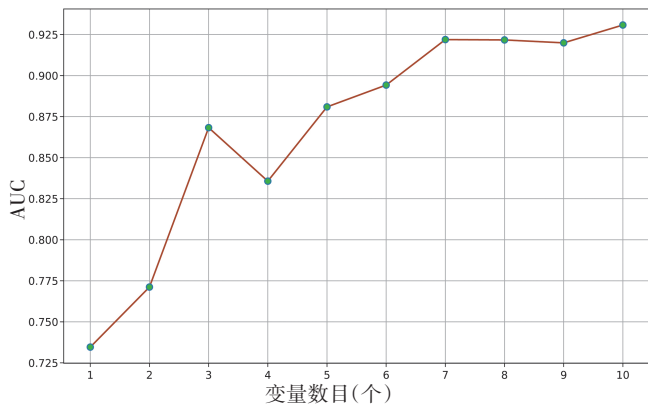
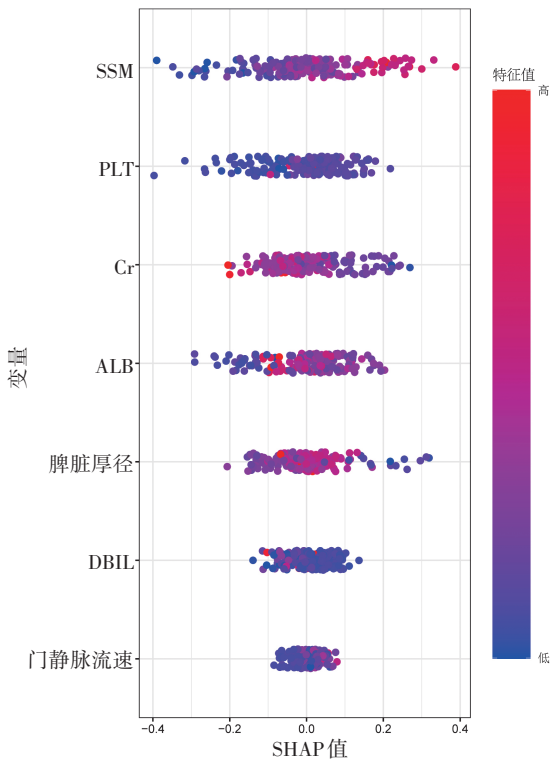
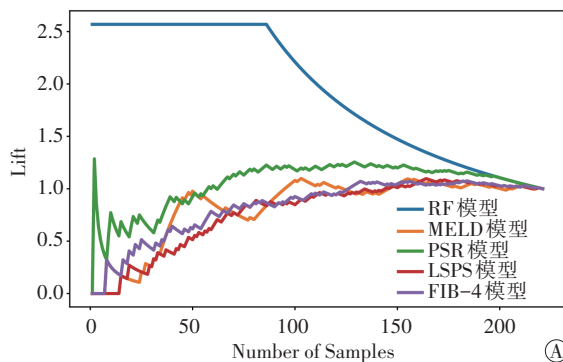


图5 不同数目的变量所对应的AUC

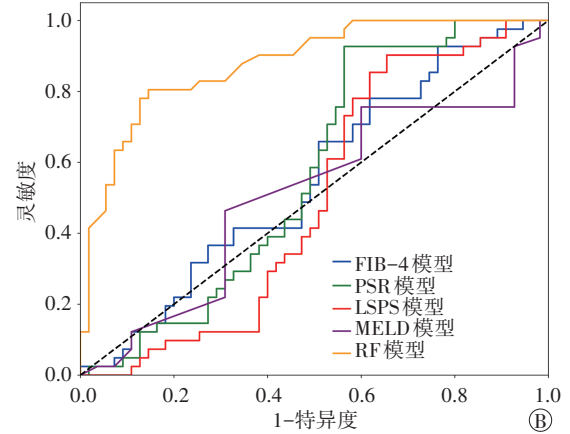
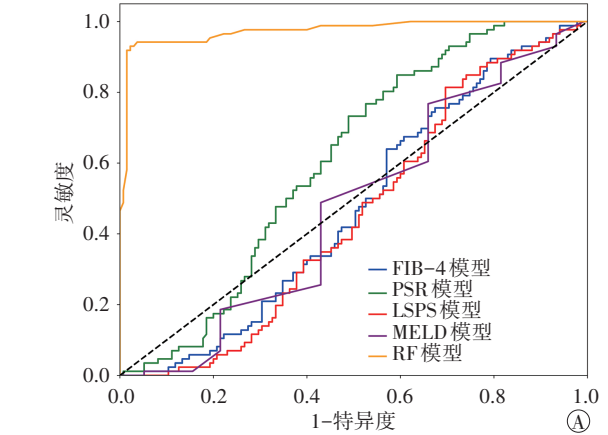


图中每一行代表一个特征,每个点代表一个样本。点的颜色代表特征值的大小,红色表示高值预示发生高危EGV,蓝色表示低值预示不容易发生高危EGV

图6 基于SHAP值的RF模型预测高危EGV的特征



A: 训练集; B: 验证集



A: 训练集; B: 验证集

图7 各模型在训练集和验证集中预测高危EGV的ROC曲线图

讨论

既往研究^[10]显示,肝硬化群体中EGV患者30 d死亡率较无EGV患者高3.01倍,并发症发生率高1.28倍(均 $P<0.001$),故构建精准动态风险评估体系已成为改善患者预后的关键。当前临床诊疗指南^[11-12]虽推荐胃镜检查作为EGV评估金标准,但由于侵入性操作导致的患者依从性下降及医疗资源区域性差异,高危人群的持续监测与预防性干预仍面临实施障碍。本研究基于多模态超声指标和血清学指标构建RF模型,

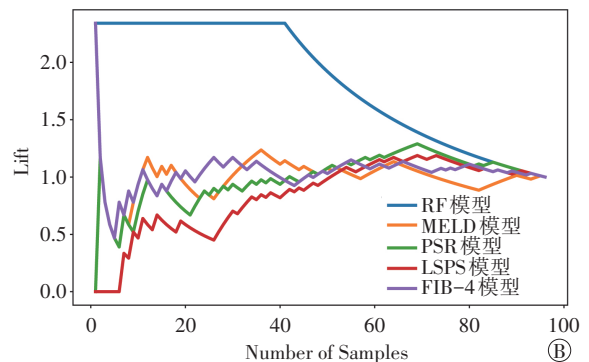


图8 各模型在训练集和验证集中预测高危EGV的梯度提升曲线图

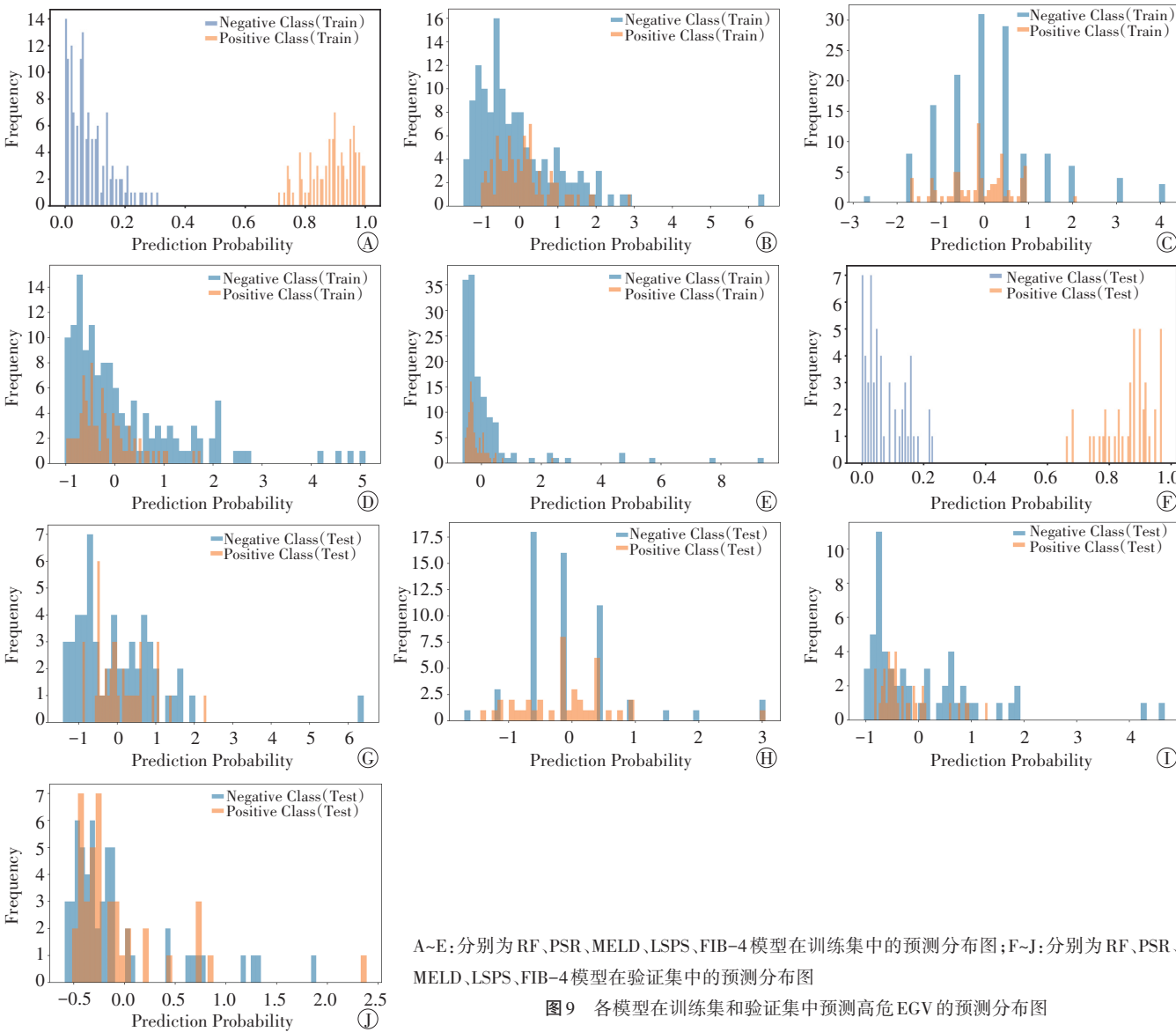
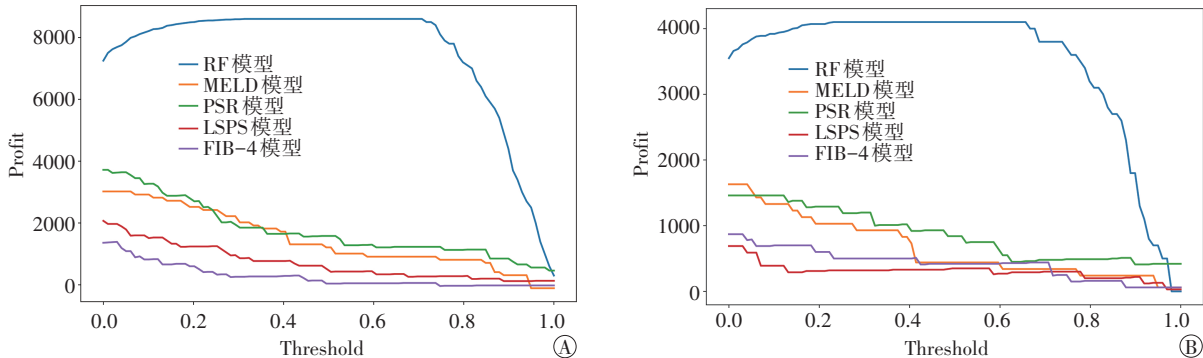


图9 各模型在训练集和验证集中预测高危EGV的预测分布图



A: 训练集; B: 验证集

图10 各模型在训练集和验证集中预测高危EGV的利润曲线图

探讨其预测乙肝相关性高危EGV的临床价值。

本研究通过比较训练集中高危EGV病例与低危EGV病例基线资料发现,高危EGV病例年龄偏大,SSM、脾脏长径及厚径、Cr、TBIL、DBIL、IBIL均显著升

高,肝血流量、门静脉流速、PLT、ALB均显著降低,且更容易形成侧支循环,与低危EGV病例比较差异均有统计学意义(均 $P<0.05$);与既往研究^[13-15]结果一致,表明年龄、SSM、脾脏长径及厚径、Cr、TBIL、DBIL、

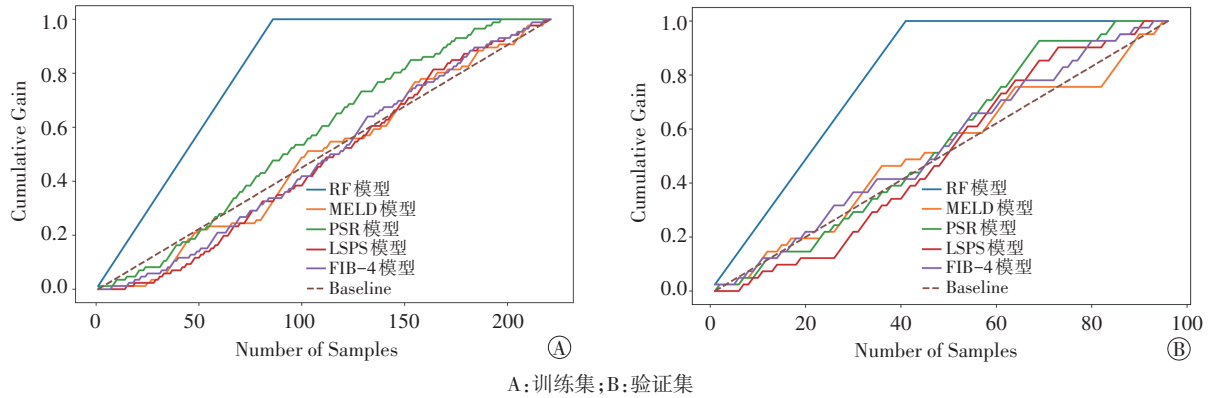


图 11 各模型在训练集和验证集中预测高危 EGV 的累积获益曲线图

IBIL、肝血流量、门静脉流速、PLT、ALB、侧支循环形成均与高危 EGV 有一定关系。然而传统的单一超声指标或血清学指标在预测高危 EGV 方面存在以下局限性：①尽管单一超声弹性成像指标（如 LSM 或 SSM）在 EGV 风险分层中具有一定应用价值，如 SSM<40 kPa 对排除高危 EGV 具有较高的阴性预测值，但其阳性检出率较低——仅能识别 21% 的高危病例^[16]；且此类指标的诊断效能易受脂肪肝变性干扰^[17-18]。脾脏形态学参数如脾脏长径、脾脏面积作为门静脉高压的间接评估指标亦存在显著局限性，其诊断特异度易受非门静脉高压因素干扰（如感染性疾病、骨髓增殖性肿瘤等）；且在门静脉高压早期阶段（如肝硬化代偿期），脾脏形态学改变常滞后于血流动力学变化^[19-20]。因此通过其预测门静脉高压并由此预测高危 EGV 的价值有限。②传统血清学指标如 PLT、Cr、ALB、胆红素等虽可通过不同维度反映肝功能损害程度及门静脉高压状态，并间接提示 EGV 风险等级^[21]，但这些指标分别表征肝脏合成功能（ALB）、门静脉血流动力学（PLT）、肾脏灌注（Cr）及胆汁代谢（胆红素），而 EGV 的形成涉及血管活性物质失衡、肝窦内皮细胞损伤及侧支循环血管重塑等复杂机制。在临床实践中，当患者合并自发性细菌性腹膜炎、肝肾综合征或多器官功能障碍时，单一指标的截断值（如 PLT<100×10⁹/L）往往因多重混杂因素干扰而丧失鉴别诊断能力。研究^[21-22]显示，基于血清学指标构建的模型预测高危 EGV 的 AUC 普遍低于 0.75，灵敏度及特异度分别为 62%~78%、65%~83%，难以满足精准分层的临床需求。

为了解决上述问题，本研究根据 RFE 进行排序，筛选出排名前 10 的特征依次为 SSM、PLT、Cr、ALB、脾脏厚径、DBIL、门静脉流速、ALT、肝血流量、脾脏长径。为秉承模型简约化，本研究拟采用尽量少的维度，达到最佳预测效果，依据重要性排序，分别纳入了不同

数目的变量构建预测模型，并计算出对应的 AUC，当变量数目为 7 个时，AUC 达到 0.924，随后虽然变量数目增加，但 AUC 仅小幅度增加，因此本研究纳入排名前 7 的变量作为预测模型的组成部分。为解决机器学习预测过程难以理解的“黑箱”过程，本研究采用 SHAP 对预测过程进行了可视化展示，结果显示 SSM、PLT、Cr、ALB、脾脏厚径、DBIL 和门静脉流速对 RF 模型预测高危 EGV 病例的贡献较大。

在整合多模态数据方面，机器学习在医学领域已取得重要进展^[23]，特别是 RF 算法，能高效整合多源信息，克服了单一指标的局限性，具备处理高维数据、非线性关系及特征间交互作用的能力^[20]。本研究结果显示，RF 模型在训练集和验证集中均展现出良好的预测效能（AUC=0.97、0.88，均 $P<0.05$ ）。为了评估 RF 模型的相对优势，本研究将其与临床常用的 FIB-4、PSR、LSPS 和 MELD 模型进行了比较，结果显示，无论是在训练集还是验证集，本研究构建的 RF 模型的 AUC 均显著高于 FIB-4、PSR、LSPS 和 MELD 模型，差异均有统计学意义（均 $P<0.001$ ）；且 RF 模型在 ROC 曲线、梯度提升值和预测分布方面均表现出显著优势。特别在利润曲线分析中，当预测概率阈值为 0.2 时，RF 模型在训练集中能够识别出 90% 的高危 EGV 病例，同时保持较低的误判率；当预测概率阈值为 0.3 时，RF 模型在验证集中可以识别出 85% 的高危 EGV 病例。表明 RF 算法能够更好地捕捉不同指标之间的复杂交互关系，提高预测准确性和稳定性。另外，本研究结果显示 RF 模型在训练集的 AUC 高于验证集（0.97 vs. 0.88），提示模型可能存在一定程度的过拟合现象。相反，FIB-4、LSPS、MELD 等传统模型在验证集中的 AUC 略高于其在训练集中的数据，分析原因为：①这些相对简单的传统模型可能在训练过程中未能充分捕捉训练数据的复杂性，甚至可能受到训练集中某些特定噪声或异

常值分布的影响;而在数据分布可能略有不同的独立验证集上,这些简单模型反而表现出与其固有预测能力更相符的水平;②虽然已进行随机划分训练集与验证集,但二者之间可能存在的样本特征分布差异也可能导致这一现象。尽管存在这种训练集与验证集 AUC 的差异,但在关键的、更能反映模型真实应用价值的验证集上,本研究构建的 RF 模型的 AUC 仍然显著高于传统模型(均 $P < 0.001$),进一步证实了 RF 模型整合多维信息的优越性及普适性。

本研究的局限性:①不同超声设备测量的 LSM 及 SSM 未达到统一标准,今后需进一步研究以完善诊断截断值的统一;②血清学指标可能受非肝病因素影响,需进一步优化和完善预测模型;③为回顾性研究,且样本量有限,今后需通过前瞻性、多中心、大样本研究获取更多多样化的数据进行训练和外部验证。

综上所述,RF 模型对乙肝相关性高危 EGV 的预测效能优于传统的 FIB-4、PSR、LSPS 和 MELD 模型,可作为其无创评估的有效工具,为临床决策提供参考。

参考文献

- [1] Pallio S, Melita G, Shahini E, et al. Diagnosis and management of esophagogastric varices[J]. *Diagnostics (Basel)*, 2023, 13(6):65-74.
- [2] Mathur N, Patel H, Mangalia R, et al. Correlation of non-invasive parameters with upper gastrointestinal endoscopy findings for presence and grading of gastroesophageal varices in liver cirrhosis patients[J]. *J Family Med Prim Care*, 2024, 13(4):1347-1353.
- [3] Madir A, Barisic Jaman M, Milosevic M, et al. Improved applicability and diagnostic accuracy of the novel spleen-dedicated transient elastography device for high-risk esophageal varices[J]. *Diagnostics (Basel)*, 2024, 14(7):65-73.
- [4] Zhao L, Tie J, Wang G, et al. Efficacy of TIPS plus extrahepatic collateral embolisation in real-world data: a validation study[J]. *BMJ Open Gastroenterol*, 2024, 11(1):87-102.
- [5] Mignot V, Chirica C, Tron L, et al. Early screening for chronic liver disease: impact of a FIB-4 first integrated care pathway to identify patients with significant fibrosis[J]. *Sci Rep*, 2024, 14(1):20720-20731.
- [6] Du YC, Jiang D, Wu J. Predicting the severity of esophageal varices in patients with hepatic cirrhosis using non-invasive markers[J]. *Risk Manag Healthc Policy*, 2023, 16(3):1555-1566.
- [7] Wang X, Han H, Yang J, et al. Liver stiffness-spleen diameter to platelet ratio score (LSPS model) predicts variceal rebleeding for cirrhotic patients[J]. *Eur J Gastroenterol Hepatol*, 2023, 35(4):488-496.
- [8] Matoba D, Noda T, Kobayashi S, et al. Analysis of short-term and long-term outcomes of living donor liver transplantation for patients with a high model for end-stage liver disease score[J]. *Transplant Proc*, 2023, 55(4):893-897.
- [9] 中华医学会肝病学分会, 中华医学会消化病学分会, 中华医学会内镜学分会. 肝硬化门静脉高压食管胃静脉曲张出血的防治指南[J]. *实用肝脏病杂志*, 2016, 19(5):641-656.
- [10] Kadry Z, Schaefer EW, Shah RA, et al. Portal hypertension: an underestimated entity?[J]. *Ann Surg*, 2016, 263(5):986-991.
- [11] Bhattarai S, Dewan KR, Shrestha G, et al. Non-invasive predictors of gastro-oesophageal varices[J]. *JNMA J Nepal Med Assoc*, 2017, 56(207):298-303.
- [12] Alam R, Karim A, Rukunuzzaman M, et al. Non-endoscopic predictors of esophageal varices in children with chronic liver disease and their utility in resource-constrained countries[J]. *Indian J Gastroenterol*, 2019, 38(4):310-316.
- [13] Kotwal V, Mbachi C, Wang Y, et al. A novel score to predict esophageal varices in patients with compensated advanced chronic liver disease[J]. *Dig Dis Sci*, 2021, 66(6):2084-2091.
- [14] Li Y, Li L, Weng HL, et al. Computed tomography vs liver stiffness measurement and magnetic resonance imaging in evaluating esophageal varices in cirrhotic patients: a systematic review and Meta-analysis[J]. *World J Gastroenterol*, 2020, 26(18):2247-2267.
- [15] Zhao H, Zhang X, Huang B, et al. Application of machine learning methods for predicting esophageal variceal bleeding in patients with cirrhosis[J]. *Eur Radiol*, 2025, 35(3):1440-1450.
- [16] Xuan J, Shi Z. Shear wave elastography measured liver stiffness-spleen size-to-platelet ratio for the prediction of high-risk oesophageal varices: a Meta-analysis[J]. *Eur J Gastroenterol Hepatol*, 2023, 35(7):753-760.
- [17] Rigamonti C, Cittone MG, Manfredi GF, et al. Spleen stiffness measurement predicts decompensation and rules out high-risk oesophageal varices in primary biliary cholangitis[J]. *JHEP Rep*, 2023, 6(1):100952.
- [18] Kumada T, Toyoda H, Yasuda S, et al. Liver stiffness measurements by 2D shear-wave elastography: effect of steatosis on fibrosis evaluation[J]. *Am J Roentgenol*, 2022, 219(4):604-612.
- [19] Indiran V, Vinod Singh N, Ramachandra Prasad T, et al. Does coronal oblique length of spleen on CT reflect splenic index? [J]. *Abdom Radiol (NY)*, 2017, 42(5):1444-1448.
- [20] Ahmad AK, Atzori S, Maurice J, et al. Non-invasive splenic parameters of portal hypertension: assessment and utility[J]. *World J Hepatol*, 2020, 12(11):1055-1066.
- [21] Garcia-Tsao G, Abraldes JG, Rich NE, et al. AGA clinical practice update on the use of vasoactive drugs and intravenous albumin in cirrhosis: expert review[J]. *Gastroenterology*, 2024, 166(1):202-210.
- [22] Khadka D, Prajapati S, Sudhamshu KC, et al. Significance of non-invasive markers as predictor of esophageal varices in liver cirrhosis[J]. *JNMA J Nepal Med Assoc*, 2017, 56(208):412-416.
- [23] Murillo Pineda MI, Siu Xiao T, Sanabria Herrera EJ, et al. The prediction and treatment of bleeding esophageal varices in the artificial intelligence era: a review[J]. *Cureus*, 2024, 16(3):e55786-e55797.

(收稿日期:2024-08-08)