

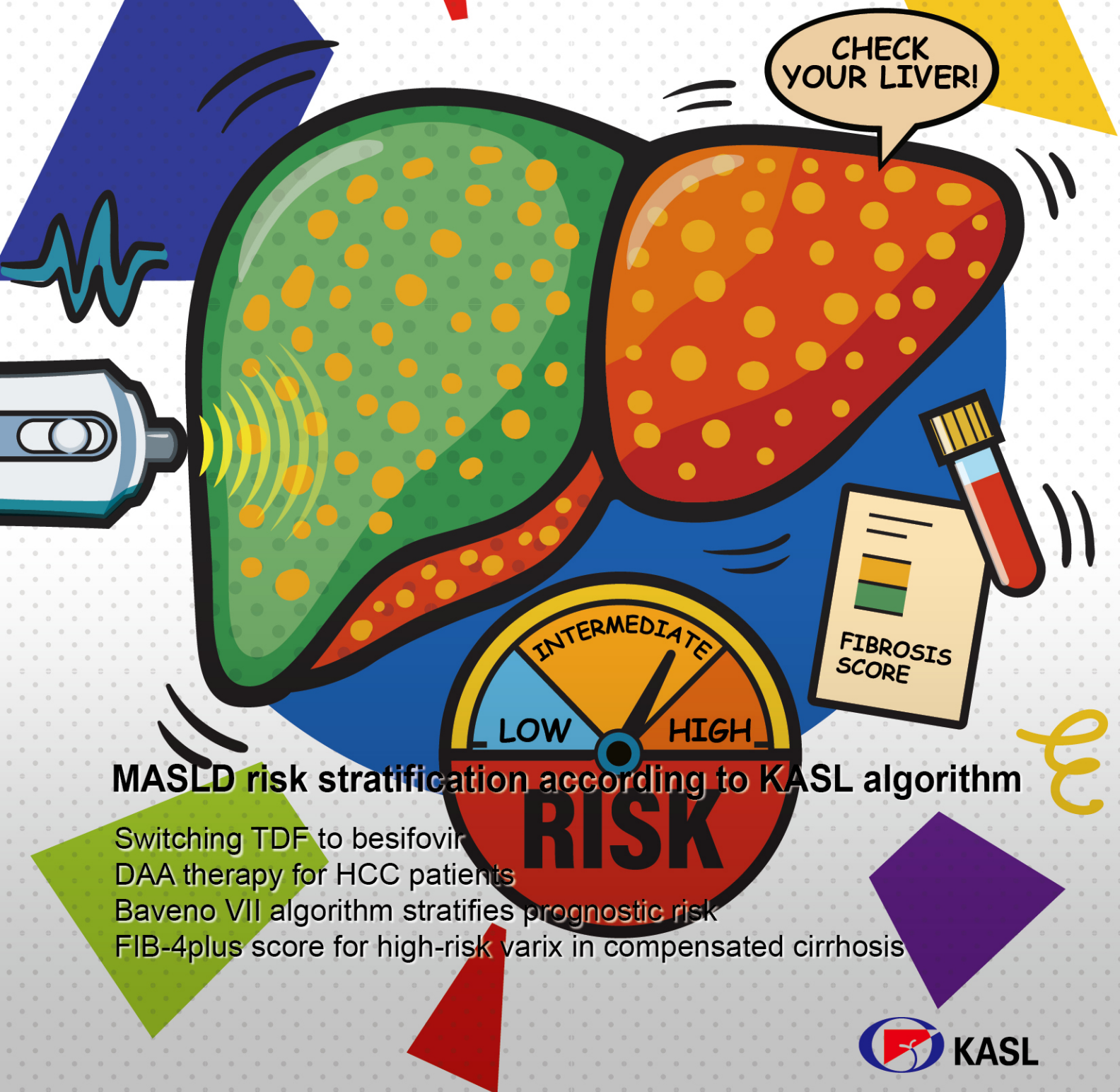
VOLUME 31 NUMBER 3 July 2025

pISSN 2287-2728
eISSN 2387-285X

CLINICAL and MOLECULAR HEPATOLOGY

The forum for latest knowledge of hepatobiliary diseases

CHECK
YOUR LIVER!



MASLD risk stratification according to KASL algorithm

Switching TDF to besifovir
DAA therapy for HCC patients

Baveno VII algorithm stratifies prognostic risk

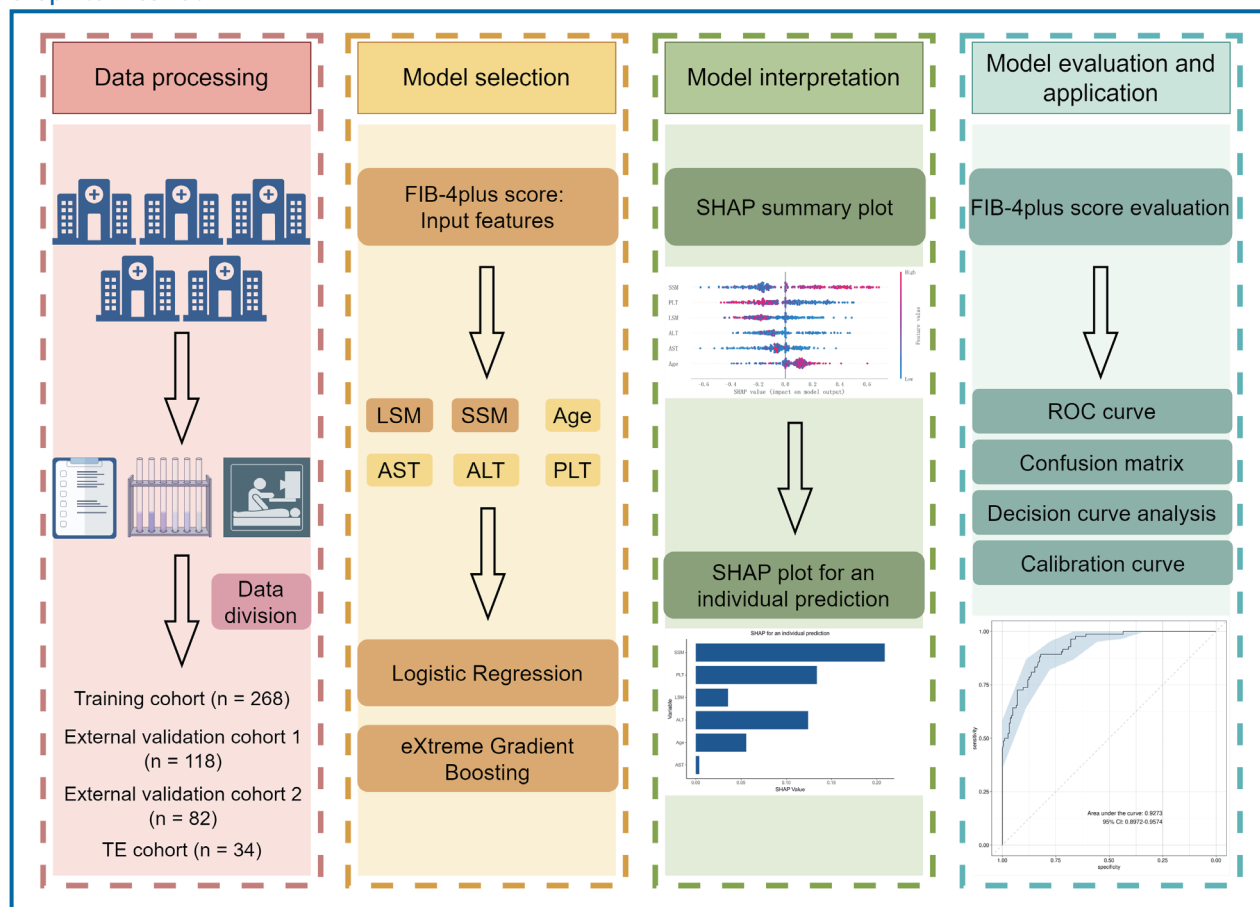
FIB-4plus score for high-risk varix in compensated cirrhosis

Fibrosis-4plus score: a novel machine learning-based tool for screening high-risk varices in compensated cirrhosis (CHESS2004): an international multicenter study

Bingtian Dong^{1,2,*}, Ruiling He^{3,4,*}, Shenghong Ju^{5,*}, Yuping Chen^{1,*}, Ivica Grgurevic^{6,*}, Jianzhong Ma^{7,*}, Ying Guo^{8,*}, Huizhen Fan^{9,*}, Qiang Yan^{10,*}, Chuan Liu¹, Huixiong Xu¹¹, Anita Madir¹², Kristian Podrug¹³, Jia Wang¹⁴, Linxue Qian¹⁴, Zhengzi Geng¹⁵, Shanghao Liu¹, Tao Ren¹⁶, Guo Zhang¹⁷, Kun Wang⁸, Meiqin Su¹⁸, Fei Chen¹⁹, Sumei Ma¹⁹, Liting Zhang²⁰, Zhaowei Tong²¹, Yonghe Zhou²², Xin Li²², Fanbin He¹⁵, Hui Huan¹⁶, Wenjuan Wang¹⁷, Yunxiao Liang¹⁷, Juan Tang²³, Fang Ai²⁴, Tingyu Wang²⁵, Liyun Zheng²⁶, Zhongwei Zhao²⁶, Jiansong Ji²⁶, Wei Liu²⁷, Jiaojiao Xu²⁷, Bo Liu²⁷, Xuemei Wang²⁸, Yao Zhang²⁸, Qiong Yan²⁹, Hui Liu³⁰, Xiaomei Chen³⁰, Shuhua Zhang³¹, Yihua Wang³¹, Yang Liu³¹, Li Yin³², Yanni Liu³², Yanqing Huang³², Li Bian³³, Ping An³³, Xin Zhang³³, Shaoting Zhang³⁴, Jinhua Shao³⁵, Xiangman Zhang³⁶, Wei Rao³⁷, Chaoxue Zhang², Christoph Frank Dietrich³⁸, Won Kim³⁹, and Xiaolong Qi¹

¹Liver Disease Center of Integrated Traditional Chinese and Western Medicine, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, Nurturing Center of Jiangsu Province for State Laboratory of AI Imaging & Interventional Radiology (Southeast University), Basic Medicine Research and Innovation Center of Ministry of Education, Zhongda Hospital, Southeast University, State Key Laboratory of Digital Medical Engineering, Nanjing; ²Department of Ultrasound, The First Affiliated Hospital of Anhui Medical University, Hefei; ³Department of Ultrasound, Donggang Branch the First Hospital of Lanzhou University; ⁴The First Clinical Medical College of Lanzhou University, Lanzhou; ⁵Nurturing Center of Jiangsu Province for State Laboratory of AI Imaging & Interventional Radiology, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, Nanjing, China; ⁶Department of Gastroenterology, Hepatology and Clinical Nutrition, University Hospital Dubrava, University of Zagreb School of Medicine and Faculty of Pharmacy and Biochemistry, Zagreb, Croatia; ⁷Department of General Surgery, The Third People's Hospital of Taiyuan; ⁸Department of Hepatology, The Third People's Hospital of Taiyuan, Taiyuan; ⁹Department of Gastroenterology, Yichun People's Hospital, Yichun; ¹⁰Division of HPB Section Department of General Surgery, Huzhou Central Hospital, The Affiliated Central Hospital of Huzhou University, Huzhou; ¹¹Department of Ultrasound, Zhongshan Hospital, Fudan University, Shanghai, China; ¹²Department of Gastroenterology, Hepatology and Clinical Nutrition, University Hospital Dubrava, Zagreb; ¹³Department of Gastroenterology and Hepatology, University Hospital Center Split, Split, Croatia; ¹⁴Department of Ultrasound, Beijing Friendship Hospital, Capital Medical University, Beijing; ¹⁵Department of Ultrasound, The People's Hospital of Linxia, Linxia; ¹⁶Department of Gastroenterology, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region, Sichuan; ¹⁷Department of Gastroenterology, Guangxi Hospital Division of the First Affiliated Hospital, Sun Yat-sen University; ¹⁸Department of Ultrasound, The Third People's Hospital of Taiyuan, Taiyuan; ¹⁹Department of Ultrasound, The First Hospital of Lanzhou University; ²⁰Department of Infectious Diseases, The First Hospital of Lanzhou University, Lanzhou; ²¹Huzhou Key Laboratory of Precision Medicine Research and Translation for Infectious Diseases, Huzhou Central Hospital, Huzhou; ²²Department of Ultrasonography, Tianjin Second People's Hospital, Tianjin; ²³Department of Infectious Disease, Zigong First People's Hospital; ²⁴Department of Ultrasound, Zigong First People's Hospital; ²⁵Department of Gastroscopy, Zigong First People's Hospital, Zigong; ²⁶Department of Radiology, Zhejiang Key Laboratory of Imaging and Interventional Medicine, Lishui Central Hospital, Li Shui; ²⁷Ultrasound Diagnosis Center, Shaanxi Provincial People's Hospital, Xi'an; ²⁸Department of Ultrasound, Ditan Hospital, Capital Medical University, Beijing; ²⁹Department of Gastroenterology, The Affiliated Hospital of Southwest Medical University; ³⁰Department of Ultrasound, The Affiliated Hospital of Southwest Medical University, Luzhou; ³¹Department of Ultrasound, North China University of Science and Technology Affiliated Hospital, Tangshan; ³²Department of Ultrasound, Jincheng People's Hospital, Jincheng; ³³Department of Hepatology, The Sixth People's Hospital of Shenyang, Shenyang; ³⁴Sensetime Research, Shanghai; ³⁵Wuxi Hisky Medical Technologies Co., Ltd., Wuxi; ³⁶Department of Medical and Marketing, Suzhou Hengrui Medical Devices Co., Ltd., Suzhou; ³⁷Shenzhen New Industries Biomedical Engineering Co., Ltd., Shenzhen, China; ³⁸Department General Internal Medicine (DAIM), Hospitals Hirslanden Bern Beau Site, Salem and Permanence, Bern, Switzerland; ³⁹Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul, Korea

Graphical Abstract



Study Highlights

- Fibrosis-4 score, liver stiffness measurement, and spleen stiffness measurement are widely used non-invasive tests; however, despite their association with different and potentially complementary aspects of liver characteristics, they only moderately predict esophageal varices.
- This study introduces the FIB-4plus score, which combines fibrosis-4, liver stiffness measurement, and spleen stiffness measurement using machine learning algorithms to better predict high-risk varices in compensated cirrhosis.
- Implementing the FIB-4plus score can lead to more accurate clinical identification of patients at increased risk for high-risk varices in compensated cirrhosis, hence improving patient outcomes by prioritizing those requiring further diagnostic procedures and interventions.

Background/Aims: A large percentage of patients undergoing esophagogastroduodenoscopy (EGD) screening do not have esophageal varices (EV) or have only small EV. We evaluated a large, international, multicenter cohort to develop a novel score, termed FIB-4plus, by combining the fibrosis-4 (FIB-4) score, liver stiffness measurement (LSM), and spleen stiffness measurement (SSM) to identify high-risk EV (HRV) in compensated cirrhosis.

Methods: This international cohort study involved patients with compensated cirrhosis from 17 Chinese hospitals and one Croatian institution (NCT04546360). Two-dimensional shear wave elastography-derived LSM and SSM values, and components of the FIB-4 score (i.e., age, aspartate aminotransferase, alanine aminotransferase, and platelet count [PLT]) were combined using machine learning algorithms (logistic regression [LR] and extreme gradient boosting [XGBoost]) to develop the LR-FIB-4plus and XGBoost-FIB-4plus models, respectively. Shapley Additive exPlanations method was used to interpret the model predictions.

Results: We analyzed data from 502 patients with compensated cirrhosis who underwent EGD screening. The XGBoost-FIB-4plus score demonstrated superior predictive performance for HRV, with an area under the receiver operating characteristic curve (AUROC) of 0.927 (95% confidence interval [CI] 0.897–0.957) in the training cohort (n=268), and 0.919 (95% CI 0.843–0.995) and 0.902 (95% CI 0.820–0.984) in the first (n=118) and second (n=82) external validation cohorts, respectively. Additionally, the XGBoost-FIB-4plus score exhibited high AUROC values for predicting EV across all cohorts. The FIB-4plus score outperformed the individual parameters (LSM, SSM, PLT, and FIB-4).

Conclusions: The FIB-4plus score effectively predicted EV and HRV in patients with compensated cirrhosis, providing clinicians with a valuable tool for optimizing patient management and outcomes. (*Clin Mol Hepatol* 2025;31:881-898)

Keywords: Esophageal and gastric varices; Machine learning; Diagnostic model; Elasticity imaging techniques

INTRODUCTION

Portal hypertension is a major complication of cirrhosis,

leading to significant morbidity and mortality due to its consequences, including ascites, esophageal varices (EV), variceal hemorrhage, and overt clinical decompensation.¹⁻⁴

Corresponding author : Xiaolong Qi

Liver Disease Center of Integrated Traditional Chinese and Western Medicine, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, Nurturing Center of Jiangsu Province for State Laboratory of AI Imaging & Interventional Radiology (Southeast University), Basic Medicine Research and Innovation Center of Ministry of Education, Zhongda Hospital, Southeast University, State Key Laboratory of Digital Medical Engineering, No.87, Dingjiaqiao, Nanjing, 210009, China
Tel: +86-18588602600, Fax: +86-025-83272121, E-mail: qixiaolong@vip.163.com
<https://orcid.org/0000-0002-3559-5855>

Won Kim

Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, 20 Boramae-ro 5-gil, Dongjak-gu, Seoul 07061, Korea
Tel: +82-2-870-2233, Fax: +82-2-831-2826, E-mail: drwon1@snu.ac.kr
<https://orcid.org/0000-0002-2926-1007>

*Bingtian Dong, Ruiling He, Shenghong Ju, Yuping Chen, Ivica Grgurevic, Jianzhong Ma, Ying Guo, Huizhen Fan, and Qiang Yan are joint first authors.

Received : Oct. 10, 2024 / **Revised :** Jan. 23, 2025 / **Accepted :** Feb. 3, 2025

Editor: Salvatore Piano, University of Padova, Italy

Abbreviations:

ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under the receiver operating characteristic curve; DCA, decision curve analysis; EGD, esophagogastroduodenoscopy; EV, esophageal varices; FIB-4, fibrosis-4; HBV, hepatitis B virus; HRV, high-risk varices; LR, logistic regression; LSM, liver stiffness measurement; NPV, negative predictive value; PLT, platelet; PPV, positive predictive value; SHAP, Shapley Additive exPlanations; SSM, spleen stiffness measurement; TE, transient elastography; XGBoost, eXtreme Gradient Boosting; 2D-SWE, two-dimensional shear wave elastography

Notably, EV occurs in approximately 50% of patients with cirrhosis.⁵ Variceal hemorrhage remains a life-threatening condition, with a six-week mortality rate of 15–25% following an acute EV hemorrhage and a one-year rebleeding risk of up to 60% without prophylaxis.^{6–8} Recently, the American Association for the Study of Liver Diseases practice guidelines recommended that patients with high-risk varices (HRV) undergo endoscopic surveillance and receive endoscopic variceal ligation to prevent variceal hemorrhage.² This highlights the importance of timely and accurate identification of HRV, which is crucial for optimizing clinical management and improving patient outcomes.

Esophagogastroduodenoscopy (EGD) is considered the gold standard for diagnosing EV.⁹ However, a large percentage of patients undergoing screening, particularly those with compensated cirrhosis, do not have EV or have only small EV without high-risk features, and therefore, do not require prophylactic therapy.¹⁰ These patients would not benefit from EGD but may suffer complications from the procedure. Therefore, developing a novel non-invasive tool for better detecting EV and predicting the risk of bleeding is urgently needed, ultimately reducing the burden of unnecessary EGDs.

Over the last decade, laboratory marker panels and elastography methods have received significant attention in routine clinical practice. The fibrosis-4 (FIB-4) score and liver stiffness measurement (LSM) are commonly used non-invasive tests for liver fibrosis in patients with chronic liver disease.^{11,12} As noted in the Baveno VI consensus guideline, utilizing LSM (<20 kPa) in combination with platelet (PLT) count (>150,000/μl) can identify patients with a very low likelihood of having HRV, making EGD screening unnecessary for these individuals.¹³ A recent study has demonstrated the value of FIB-4 in predicting the development of liver cirrhosis-related outcome events.¹⁴ Additionally, studies have confirmed that spleen stiffness measurement (SSM) correlates with portal hypertension and is considered as a promising non-invasive alternative to EGD.^{15,16} Indeed, the most recent Baveno VII consensus guideline has recommended that SSM be considered an additional tool for assessing portal hypertension and HRV.¹⁷

Despite these strong correlations, the ability of FIB-4, LSM, and SSM to predict EV is only moderate when considering each score separately.^{18,19} We proposed, for

the first time, the hypothesis that combining these three non-invasive markers can improve the overall performance, as FIB-4, LSM, and SSM are associated with different and potentially complementary aspects of liver characteristics.

To test this hypothesis, we conducted a multicenter study incorporating the FIB-4 score (or its individual components as separate predictors), LSM, and SSM into a single model, termed FIB-4plus, to detect EV and HRV in patients with compensated cirrhosis whose EV severity was assessed using EGD. We used a novel machine learning method to develop and validate the FIB-4plus score for predicting EV and HRV in patients with compensated cirrhosis, aiming to provide a non-invasive and cost-effective alternative to EGD that enables the accurate identification of EV and HRV in this population.

MATERIALS AND METHODS

Study population and design

This international multicenter study included patients with compensated cirrhosis who underwent EGDs at 17 hospitals in China and one university hospital center in Croatia. Patients with compensated cirrhosis were prospectively recruited from 14 Chinese hospitals between October 2020 and December 2022 to form a training cohort for the development of the FIB-4plus score. It is retrospective analysis of prospectively collected data. Two additional independent cohorts were established for external validation: validation cohort 1, which comprised patients with compensated cirrhosis prospectively enrolled in two Chinese hospitals between March 2016 and December 2022, and validation cohort 2, which comprised patients with compensated cirrhosis retrospectively enrolled in a Croatian University Hospital Center between January 2016 and December 2022. These three cohorts included patients who underwent LSM and SSM using two-dimensional shear wave elastography (2D-SWE).

In addition, a cohort of eligible patients with compensated cirrhosis who underwent transient elastography (TE) to measure LSM and SSM was prospectively recruited from four Chinese hospitals between August 2021 and March 2022 to further validate the utility of our model in detecting EV and HRV in these patients.

This study was approved by the ethics committee of the principal investigator's hospital (approval number LDYYLL2020-246) and was registered with ClinicalTrials.gov (NCT04546360). All patients signed an informed consent form.

Inclusion and exclusion criteria

Adult participants (≥ 18 years) with compensated cirrhosis, confirmed by compatible clinical, liver histological biopsy, or radiologic findings, who underwent both EGD and 2D-SWE (training cohort, validation cohorts 1 and 2) or TE (TE cohort) for measuring LSM and SSM, were included. All patients were required to have time between 2D-SWE or TE measurements, laboratory tests, and EGD within 1 month. The exclusion criteria were as follows: (1) patients with previous or current decompensation (defined as cirrhotic patients with ascites, variceal hemorrhage, or hepatic encephalopathy, etc.); (2) individuals with noncirrhotic etiologies for portal hypertension; (3) unqualified acquisitions for LSM or SSM, as described in our previous studies;^{20,21} (4) patients who experienced cirrhosis regression following antiviral therapy (histology showing fibrosis stage $< F4$); (5) those who received non-selective beta-blockers or underwent endoscopic variceal ligation; and (6) those with concurrent hepatocellular carcinoma.

LSM and SSM measurements

In training and validation cohorts 1 and 2, both LSM and SSM were performed using the Aixplorer 2D-SWE system (SuperSonic Imagine, Aix-en-Provence, France) with an abdominal 3.5 MHz curved array probe (SuperSonic Imagine). In the TE cohort, LSM and SSM were performed using Fibroscan[®] (Echosens, Paris, France) equipped with an M or XL probe (Echosens) for LSM and a 100 Hz specific TE-probe (Echosens) for SSM. The patients fasted for at least 6 hours before scanning. Both 2D-SWE and TE examinations were conducted by experienced operators following standardized procedures; further details are provided in Supplementary Materials and Methods, Parts I and II.

Endoscopic evaluation of EV and HRV

Standard endoscopies were conducted by expert operators who were blinded to the clinical and laboratory data of the study population.²¹ Details are described in the Supplementary Materials and Methods, Part III. According to the Baveno VII consensus guidelines, HRV is defined as large varices (varix size, ≥ 5 mm), or red signs, or Child–Pugh C.¹⁷

Data collection

Clinical and laboratory data, including demographic characteristics, 2D-SWE or TE examination, routine blood tests, liver and renal biochemistry, coagulation, and etiology of cirrhosis, were recorded. Liver disease severity was evaluated by calculating the Child–Pugh score and the model for end-stage liver disease score.²² The FIB-4 score was calculated as follows: $FIB-4 = \text{age (years)} \times \text{aspartate aminotransferase (AST, U/L)} / (\text{PLT } [10^9/\text{L}] \times \text{alanine aminotransferase [ALT, U/L]}^{1/2})$.²³

Overview of machine learning models

Several approaches and strategies have been employed to construct a new model by combining LSM, SSM, and the FIB-4 score to develop an optimal FIB-4plus score. LSM, SSM, and FIB-4 values were combined as continuous variables to develop the model. In addition, we developed models that included the individual components of the FIB-4 score (i.e., age, AST, ALT, and PLT) as separate predictors rather than using the composite FIB-4 score as a single predictor. Additionally, univariable and multivariable logistic regression (LR) analyses were conducted to evaluate the significant associations between each variable of the FIB-4plus score (i.e., age, AST, ALT, PLT, LSM, and SSM) and the outcome, with all results showing statistical significance (all $P < 0.01$) (Supplementary Materials and Methods, Part IV).

In this study, two machine learning algorithms, LR and eXtreme Gradient Boosting (XGBoost), were used to develop the LR-FIB-4plus and XGBoost-FIB-4plus models, respectively. Here, we briefly describe the two algorithms. LR is a widely used machine learning algorithm, which is used to solve classification problems, particularly binary

classification problems. It predicts the probability that a single sample belongs to a certain class by utilizing a logistic function (usually the Sigmoid function) to map the output value of a linear regression between 0 and 1. The differences between statistical LR and machine learning LR are detailed in the Supplementary Materials and Methods, Part V. We developed an alternative model that integrated LSM, SSM, and the individual components of the FIB-4 score (i.e., age, AST, ALT, and PLT) as independent predictors, using statistical LR to predict EV and HRV. The performance metrics of this statistical LR model are detailed in the Supplementary Materials and Methods, Part VI. XGBoost, introduced by Chen and Guestrin²⁴ in 2016, is a prevalent tree-based machine learning ensemble algorithm. It integrates many tree models to form a strong classifier, resulting in better prediction results in terms of efficiency and accuracy.²⁵

Ten-fold cross-validation was used in the machine learning model-building process. We used the Shapley Additive exPlanations (SHAP) method to interpret the model predictions and determine the effects of features on the predictions. SHAP, a model-agnostic explanation technique, is useful for interpreting machine learning models at the cohort and patient levels. For each feature across all patients, the SHAP values were separately aggregated and averaged to measure the importance of the features for prediction. The importance plot of the SHAP features illustrates the overall importance of each feature. The larger the mean absolute SHAP value, the higher the importance of the feature for model prediction. The SHAP summary plot demonstrates the effect of each feature on the predictions. A positive SHAP value suggests that the outcome (i.e., HRV) is more likely to be due to the feature, whereas a negative SHAP value implies that the outcome is less likely to be due to the feature.

Outcomes

The primary outcome of this study was the development of a single score (i.e., FIB-4plus) that combines FIB-4, LSM, and SSM to predict HRV in patients with compensated cirrhosis using a machine learning algorithm. The secondary outcome was the prediction of EV, identified using this score.

Statistical analysis

Data are presented as continuous or categorical variables. We first used the Shapiro–Wilk test to determine whether the data followed a normal distribution. Data were given as mean±standard deviation for normally distributed data, while median and interquartile ranges were used for non-normally distributed data. Between-group comparisons were conducted using the *t*-test or the Mann–Whitney U-test, as appropriate. A two-tailed *P*-value <0.05 was considered significant. The IBM SPSS Statistics 26 software (IBM Co., Armonk, NY, USA) was used for this study.

The diagnostic performance of the FIB-4plus score was evaluated using several metrics, including accuracy, area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score, in both the training and validation cohorts. Among these metrics, the AUROC and F1-score served as primary indicators for model performance comparison. The DeLong test was used to assess the differences in the AUROCs.

We selected a rule-out cut-off that achieves sensitivity of ≥95% and a rule-in cut-off that achieved specificity of ≥95% for the outcome of HRV. For the outcome of EV, we selected cut-off values that achieve sensitivity or specificity >90%. A subgroup analysis was conducted to evaluate the model's performance in subgroups stratified by cirrhosis etiology. Additionally, LSM, SSM, PLT, and FIB-4 were used to predict EV and HRV in both the training and validation cohorts. The diagnostic performance of the rule-out and rule-in cut-offs for LSM and SSM in predicting EV and HRV was also assessed. Furthermore, we investigated the performance of the Baveno VI criteria¹³ and the Baveno VII algorithm¹⁷ for ruling out HRV using 2D-SWE. The performance of the FIB-4plus score was validated in a TE cohort. To further assess and validate the models' performance, decision curve analysis (DCA) and calibration curve were employed. The Python software (version 3.7, Python Software Foundation) was used in this study.

RESULTS

Patient characteristics

In total, 502 patients with compensated cirrhosis who underwent EGD in China and Croatia were included in the analysis (Fig. 1). The number of patients included at each center is detailed in the Supplementary Materials and Methods, Part VII. The baseline characteristics of the patients are summarized in Table 1. In the training cohort, the median age was 49.0 years. Among them, 67.2% were male, and 80.2% were diagnosed with hepatitis B virus

(HBV) infection. Additionally, 85.8% of the patients were undergoing treatment at the time of the elastography examination. Similarly, in validation cohort 1, the median age was 51.5 years, with 68.6% of the patients being male and 79.7% having an HBV infection. In validation cohort 2, the median age was 62.0 years, and 79.3% of the participants were male, with alcohol-related liver disease being the most common etiology (36.6%). In the TE cohort, both the median age (52.0 years) and the proportion of male participants (67.6%) were similar to those of the training cohort.

Based on EGD, 84 (31.3%), 22 (18.6%), 19 (23.2%), and

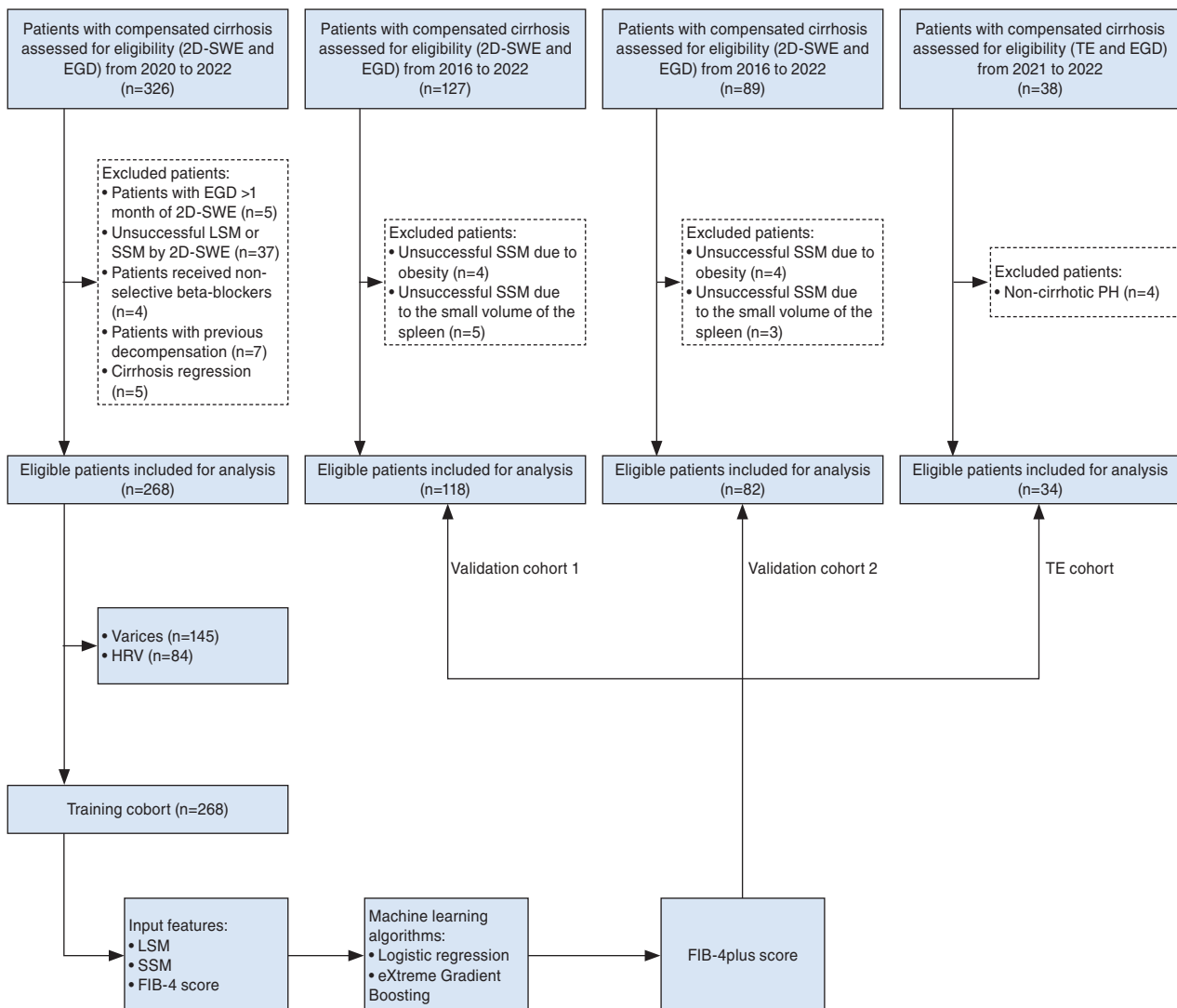


Figure 1. Study design and flow chart of the enrolled patients. EGD, esophagogastroduodenoscopy; FIB-4, fibrosis-4; HRV, high-risk varices; LSM, liver stiffness measurement; PH, portal hypertension; SSM, spleen stiffness measurement; TE, transient elastography; 2D-SWE, two-dimensional shear wave elastography.

Table 1. Patient characteristics for the cohort of 502 patients with compensated cirrhosis

Characteristics	Training cohort (n=268)	Validation cohort 1 (n=118)	Validation cohort 2 (n=82)	TE cohort (n=34)
Male	180/268 (67.2)	81/118 (68.6)	65/82 (79.3)	23/34 (67.6)
Age (yr)	49.0 (42.0–55.0)	51.5 (44.8–59.0)	62.0 (57.0–68.0)	52.0 (47.0–60.3)
BMI (kg/m ²), n=401	24.1 (22.0–26.1)	24.5 (22.5–26.3)	27.4 (24.3–32.2)	23.5 (21.2–25.7)
Laboratory data				
ALT (U/L)	38.0 (26.0–68.0)	25.0 (19.0–39.0)	33.0 (25.0–67.0)	24.0 (19.0–51.0)
AST (U/L)	40.0 (28.7–71.2)	27.0 (21.0–37.0)	46.0 (29.0–67.0)	28.3 (23.0–50.8)
PLT (10 ⁹ /L)	108.0 (72.0–154.0)	120.0 (81.0–168.0)	141.0 (106.0–193.0)	98.0 (71.0–173.0)
INR, n=500	1.1 (1.0–1.3)	1.2 (1.1–1.3)	1.1 (1.0–1.3)	1.1 (1.1–1.2)
Albumin (g/L), n=499	39.8 (35.1–44.4)	42.0 (38.0–45.0)	40.0 (36.0–46.0)	42.0 (38.0–45.0)
Total bilirubin (μmol/L)	22.0 (16.6–31.6)	17.6 (12.7–22.5)	19.0 (13.8–26.3)	16.5 (12.3–26.4)
Serum creatinine (mg/dL)	0.7 (0.6–0.8)	0.7 (0.6–0.8)	0.9 (0.7–1.0)	0.7 (0.6–0.9)
Etiology of cirrhosis				
HBV	215/268 (80.2)	94/118 (79.7)	6/82 (7.3)	29/34 (85.3)
HCV	18/268 (6.7)	15/118 (12.7)	11/82 (13.4)	1/34 (2.9)
MASLD	2/268 (0.7)	1/118 (0.8)	15/82 (18.3)	0 (0)
Alcohol	7/268 (2.6)	1/118 (0.8)	30/82 (36.6)	3/34 (8.8)
AILD*	11/268 (4.1)	2/118 (1.7)	9/82 (11.0)	0 (0)
Other	15/268 (5.6)	5/118 (4.2)	11/82 (13.4)	1/34 (2.9)
2D-SWE or TE examination†				
LSM (kPa)	14.1 (10.8–19.3)	11.8 (9.7–15.2)	22.3 (15.1–29.8)	9.7 (7.2–18.1)
SSM (kPa)	32.7 (24.9–40.0)	28.1 (23.9–32.8)	31.7 (25.6–39.3)	32.8 (23.6–46.2)
Child–Pugh score, n=420	5.0 (5.0–6.0)	5.0 (5.0–5.0)	-	5.0 (5.0–6.0)
Child–Pugh class (A/B), n=468	223 (83.2)/45 (16.8)	115 (97.5)/3 (2.5)	74 (90.2)/8 (9.8)	-
MELD score, n=497	9.0 (7.6–11.1)	10.1 (8.5–14.0)	9.4 (7.5–11.0)	9.4 (6.2–11.6)
FIB-4	3.1 (2.1–5.8)	2.3 (1.6–3.9)	3.0 (2.2–5.2)	2.9 (1.7–5.7)
EV	145/268 (54.1)	73/118 (61.9)	33/82 (40.2)	21/34 (61.8)
HRV	84/268 (31.3)	22/118 (18.6)	19/82 (23.2)	9/34 (26.5)

Values presented as number (%), median (interquartile range).

AILD, autoimmune liver disease; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; EV, esophageal varices; FIB-4, fibrosis-4; HBV, hepatitis B virus; HCV, hepatitis C virus; HRV, high-risk varices; INR, international normalized ratio; LSM, liver stiffness measurement; MASLD, metabolic dysfunction-associated steatotic liver disease; MELD, Model for End-stage Liver Disease; PLT, platelet; SSM, spleen stiffness measurement; TE, transient elastography; 2D-SWE, two-dimensional shear wave elastography.

*AILD includes autoimmune hepatitis, primary biliary cholangitis, and primary sclerosing cholangitis. †LSM and SSM were performed using 2D-SWE in the training cohort and the validation cohort 1 and 2. In the TE cohort, LSM and SSM were performed using TE.

9 (26.5%) patients had HRV in the training, the validation cohort 1, the validation cohort 2, and the TE cohorts, respectively. A comparison of the population characteristics of patients with HRV and without HRV is presented in Supplementary Table 1. The comparison of characteristics between patients who underwent 2D-SWE and those who underwent TE is provided in Supplementary Table 2.

Performance of the FIB-4plus score in the prediction of EV and HRV

We evaluated the diagnostic performance of the FIB-4plus score in predicting EV and HRV, as summarized in Table 2 and illustrated in Figure 2. Across all three cohorts, the XGBoost-FIB-4plus score demonstrated superior performance in predicting HRV, achieving an AUROC of 0.927 (95% confidence interval [CI] 0.897–0.957) in the training cohort, and 0.919 (95% CI 0.843–0.995) and 0.902 (95% CI 0.820–0.984) in the first and second external validation cohorts, respectively (Fig. 2). The XGBoost-FIB-4plus score achieved an accuracy, sensitivity, specificity, PPV, NPV, and F1-score of 0.825, 0.726, 0.924, 0.813, 0.880, and 0.767, respectively, in the training cohort (Table 2).

For detecting EV, the XGBoost-FIB-4plus score presented an AUROC of 0.900 (95% CI 0.864–0.935). The XGBoost-FIB-4plus score achieved an accuracy, sensitivity, specificity, PPV, NPV, and F1-score of 0.798, 0.848, 0.748, 0.799, 0.807, and 0.823, respectively, in the training cohort (Table 2, Supplementary Fig. 1). In the validation cohorts, the XGBoost-FIB-4plus score consistently achieved AUROCs exceeding 0.80 for predicting EV.

Both the LR-FIB-4plus and XGBoost-FIB-4plus models showed improved diagnostic performance compared with LSM, SSM, PLT, or FIB-4 alone. Their diagnostic performance for detecting EV and HRV across all cohorts is shown in Supplementary Table 3 and Supplementary Figure 2. For HRV, the AUROCs for LSM, SSM, PLT, and FIB-4 were 0.610 (95% CI 0.539–0.680), 0.792 (95% CI 0.735–0.849), 0.753 (95% CI 0.694–0.813), and 0.686 (95% CI 0.619–0.752) in the training cohort, 0.775 (95% CI 0.662–0.889), 0.881 (95% CI 0.777–0.985), 0.764 (95% CI 0.655–0.874), and 0.773 (95% CI 0.676–0.869) in the validation cohort 1, and 0.766 (95% CI 0.643–0.889), 0.798 (95% CI 0.686–0.910), 0.614 (95% CI 0.462–0.766), and

0.584 (95% CI 0.432–0.736) in the validation cohort 2, respectively. Details of LSM and SSM performance in predicting EV and HRV are presented in Supplementary Table 4. Overall, the SSM demonstrated higher accuracy and AUROC compared with the LSM, PLT, and FIB-4. Furthermore, we investigated the performance of the Baveno VI criteria and the Baveno VII algorithm for ruling out HRV using 2D-SWE (Supplementary Table 5). However, the results were not satisfactory.

We also included the composite FIB-4 score as a single predictor for model development. Supplementary Table 6 shows the diagnostic performance of the FIB-4plus score, which combines the LSM, SSM, and the FIB-4 score as continuous variables for predicting HRV. The XGBoost-FIB-4plus score outperformed the LR-FIB-4plus score, achieving AUROCs of 0.869 (95% CI 0.822–0.916), 0.905 (95% CI 0.830–0.980), and 0.842 (95% CI 0.745–0.938) in the training cohort, validation cohort 1, and validation cohort 2, respectively (Supplementary Fig. 3).

Finally, the XGBoost-FIB-4plus score, which combines LSM, SSM, and the individual components of the FIB-4 score (i.e., age, AST, ALT, and PLT) as separate predictors, demonstrated the highest discrimination ability (Table 2; Fig. 2). An open web page for HRV detection is freely accessible at <https://FIB-4plus.shinyapps.io/dynnomapp/>.

Performance of the XGBoost-FIB-4plus score for ruling out and ruling in EV and HRV

By applying a rule-out cut-off value of 0.420, the XGBoost-FIB-4plus score achieved sensitivity and NPV greater than 0.95 in both the training cohort and validation cohort 1, outperforming the Baveno VI criteria and the Baveno VII algorithm. Similarly, using a rule-in cut-off value of 0.640, the XGBoost-FIB-4plus score demonstrated specificity and PPV greater than 0.95. These results highlight the superior performance of the XGBoost-FIB-4plus score in both ruling out and ruling in HRV, supporting its potential utility in clinical practice. Detailed results are presented in Supplementary Table 7.

Model interpretation

Figure 3 and Supplementary Figure 4 show the SHAP feature importance and SHAP summary plots for the LR-

Table 2. Diagnostic performance of the machine learning models for predicting EV and HRV

EV	Accuracy	AUROC (95% CI)	SEN	SPE	PPV	NPV	F1-score
Training cohort							
LR-FIB-4plus	0.718	0.805 (0.753–0.857)	0.786	0.650	0.726	0.721	0.755
XGBoost-FIB-4plus	0.798	0.900 (0.864–0.935)	0.848	0.748	0.799	0.807	0.823
Validation cohort 1							
LR-FIB-4plus	0.710	0.776 (0.699–0.862)	0.932	0.489	0.747	0.815	0.829
XGBoost-FIB-4plus	0.768	0.816 (0.736–0.897)	0.959	0.578	0.787	0.897	0.864
Validation cohort 2							
LR-FIB-4plus	0.787	0.846 (0.754–0.938)	0.697	0.878	0.793	0.811	0.742
XGBoost-FIB-4plus	0.731	0.831 (0.744–0.918)	0.667	0.796	0.688	0.780	0.677
HRV							
Training cohort							
LR-FIB-4plus	0.700	0.836 (0.787–0.885)	0.536	0.864	0.643	0.803	0.584
XGBoost-FIB-4plus	0.825	0.927 (0.897–0.957)	0.726	0.924	0.813	0.880	0.767
Validation cohort 1							
LR-FIB-4plus	0.797	0.898 (0.801–0.994)	0.636	0.958	0.778	0.920	0.700
XGBoost-FIB-4plus	0.867	0.919 (0.843–0.995)	0.818	0.917	0.692	0.957	0.750
Validation cohort 2							
LR-FIB-4plus	0.618	0.835 (0.736–0.933)	0.316	0.921	0.546	0.817	0.400
XGBoost-FIB-4plus	0.808	0.902 (0.820–0.984)	0.632	0.984	0.846	0.923	0.750

The FIB-4plus score was developed by combining LSM, SSM, and the individual components of the fibrosis-4 (FIB-4) score (i.e., aspartate aminotransferase, alanine aminotransferase, age, and platelet count).
In the training cohort and the validation cohort 1 and 2, LSM and SSM were performed using two-dimensional shear wave elastography.
AUROC, area under the receiver-operating characteristic curve; CI, confidence interval; EV, esophageal varices; HRV, high-risk varices; LR, logistic regression; LSM, liver stiffness measurement; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity; SSM, spleen stiffness measurement; XGBoost, eXtreme Gradient Boosting.

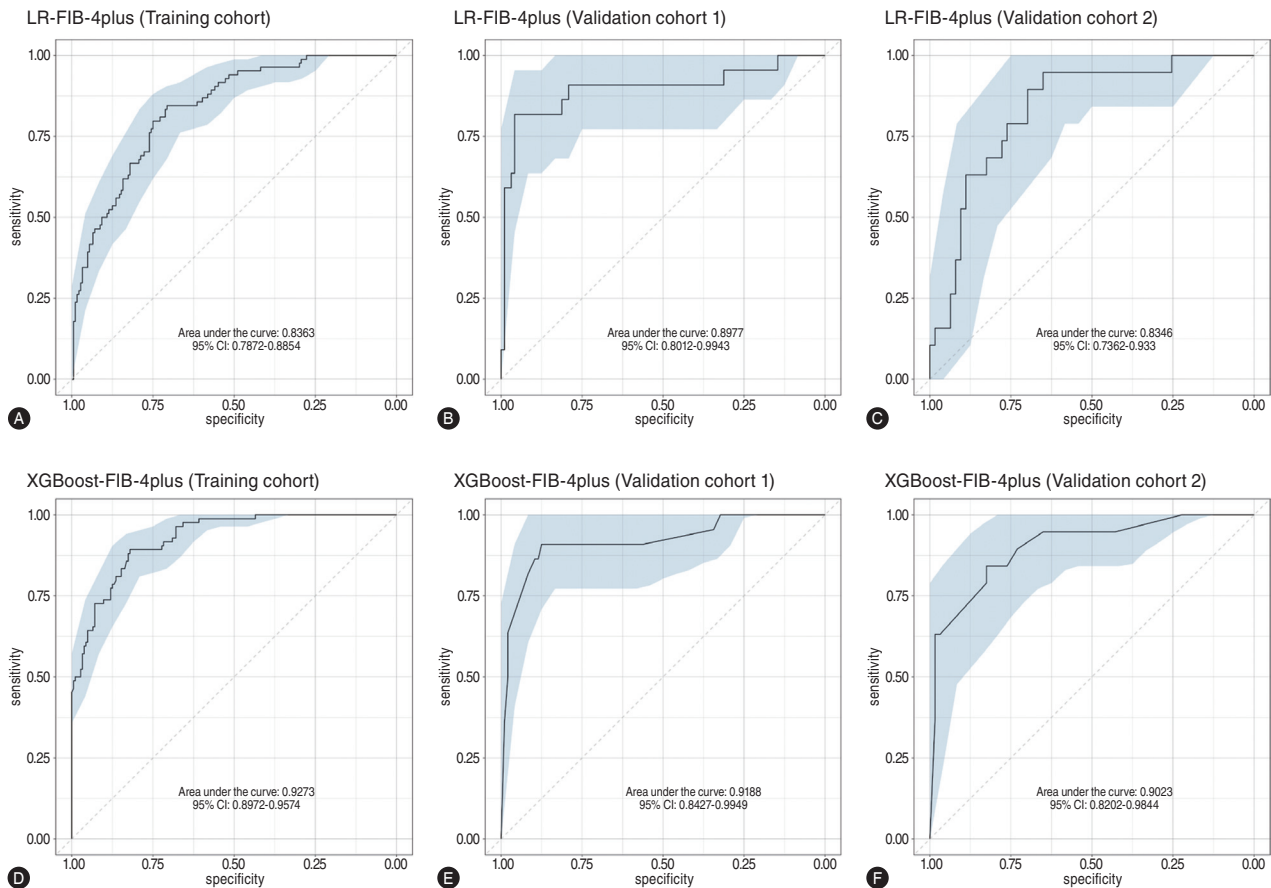


Figure 2. Diagnostic performance of the machine learning models for HRV. (A) The LR-FIB-4plus model in the training cohort, (B) The LR-FIB-4plus model in the validation cohort 1, (C) The LR-FIB-4plus model in the validation cohort 2, (D) The XGBoost-FIB-4plus model in the training cohort, (E) The XGBoost-FIB-4plus model in the validation cohort 1, (F) The XGBoost-FIB-4plus model in the validation cohort 2. The FIB-4plus score was developed by combining LSM, SSM, and the individual components of the fibrosis-4 (FIB-4) score (i.e., aspartate aminotransferase, alanine aminotransferase, age, and platelet count). CI, confidence interval; FIB-4, fibrosis-4 score; HRV, high-risk esophageal varices; LR, logistic regression; LSM, liver stiffness measurement; SSM, spleen stiffness measurement; XGBoost, eXtreme Gradient Boosting.

FIB-4plus and XGBoost-FIB-4plus models. This explainable method offers two types of insights: (1) a global explanation, which highlights feature-level contributions to the overall model predictions, and (2) a local explanation, which provides case-specific insights into individual predictions. Features were ranked based on the mean absolute SHAP values, reflecting their relative importance in influencing predictions. A higher SHAP value for a feature indicates a stronger association with the likelihood of EV or HRV. To predict HRV using the XGBoost-FIB-4plus model, the feature importance was ranked as follows: SSM, PLT, LSM, ALT, AST, and age. SSM and PLT consistently emerged as the two most critical predictors across both the LR-FIB-4plus and XGBoost-FIB-4plus models (Fig. 3). This systematic feature ranking reinforces the dominant role of SSM and PLT

in prediction models and highlights the interpretability of these models in clinical decision making.

Performance of the FIB-4plus score for predicting EV and HRV in patients with HBV infection

Subgroup analysis was conducted to evaluate the diagnostic performance of the XGBoost-FIB-4plus score specifically in patients with HBV infection. Given the limited number of HBV-infected patients in the validation cohort 2, the analysis focused on the training cohort and validation cohort 1. The XGBoost-FIB-4plus score achieved AUROCs of 0.923 (95% CI 0.887–0.960) and 0.905 (95% CI 0.798–1.000) in the training cohort and validation cohort 1, respectively

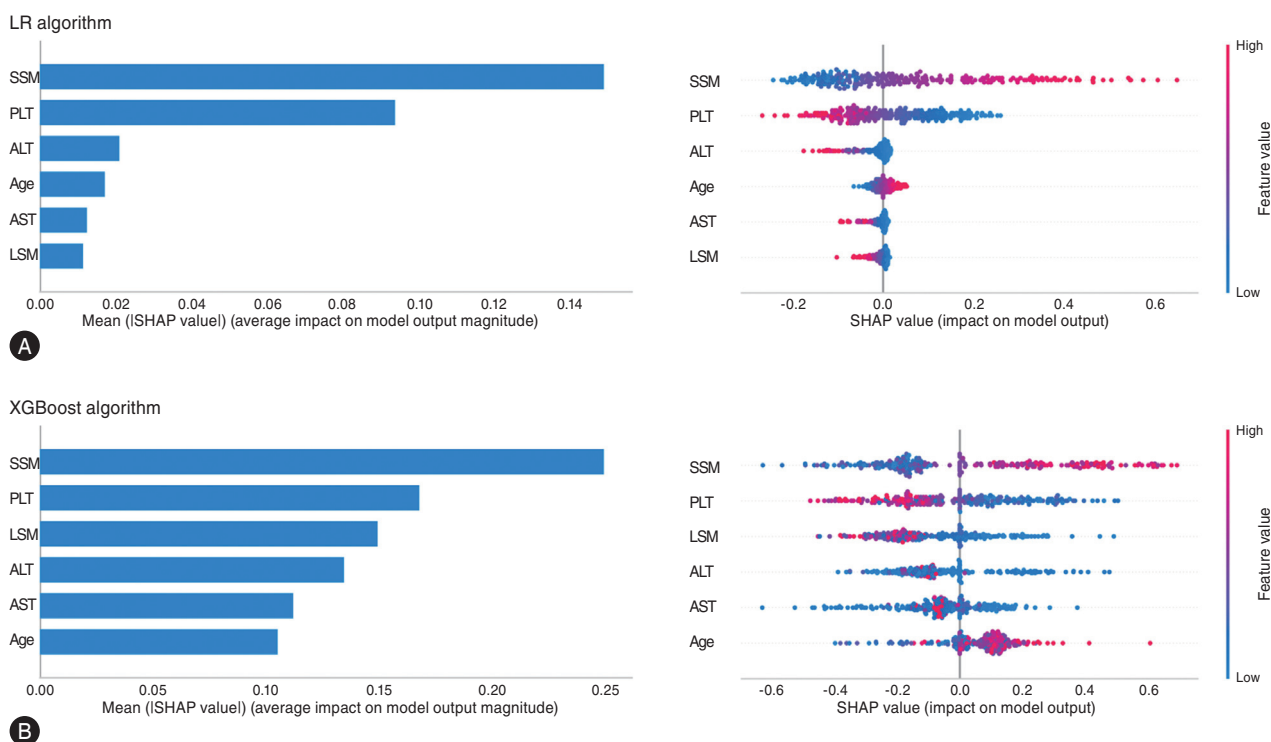


Figure 3. Shapley Additive exPlanations (SHAP) plot: the effects of clinical features for predicting high-risk varices in the (A) logistic regression (LR) and (B) eXtreme Gradient Boosting (XGBoost). ALT, alanine aminotransferase; AST, aspartate aminotransferase; LSM, liver stiffness measurement; PLT, platelet; SSM, spleen stiffness measurement.

(Supplementary Fig. 5; Supplementary Table 8).

Validation of TE-based models in the prospective external validation cohort

The performance of the FIB-4plus score was further assessed in the TE cohort, which utilized TE-derived LSM and SSM values along with the individual components of the FIB-4 score as separate predictors. The XGBoost-FIB-4plus score demonstrated consistently high diagnostic accuracy, achieving an AUROC of 0.904 (95% CI 0.799–1.000) for predicting HRV, whereas the LR-FIB-4plus score achieved a slightly lower AUROC of 0.831 (95% CI 0.686–0.977) (Supplementary Fig. 6). In the detection of EV, the XGBoost-FIB-4plus score achieved an AUROC exceeding 0.9 (Supplementary Table 9).

Performance of the XGBoost model combining LSM, SSM, and PLT for predicting EV and HRV

We also developed an alternative model that combined

LSM, SSM, and PLT using the XGBoost algorithm to predict EV and HRV. The ROC curves of this model are shown in Supplementary Figure 7, and the importance of these three clinical features, as assessed using the SHAP method, is displayed in Supplementary Figure 8. To predict the HRV, the XGBoost model achieved AUROCs of 0.855 (95% CI 0.808–0.901), 0.891 (95% CI 0.812–0.971), and 0.828 (95% CI 0.726–0.930) for the training, validation cohort 1, and validation cohort 2, respectively (Supplementary Fig. 7). For EV detection, the AUROCs ranged from 0.780 to 0.831. Therefore, the performance of the XGBoost model combining the LSM, SSM, and PLT was lower than that of the FIB-4plus model that we developed earlier.

Evaluation of the FIB-4plus score performance for EV and HRV diagnosis by DCA and calibration curve

The DCA for the XGBoost-FIB-4plus score consistently showed a net benefit in the training cohort and all three validation cohorts across various threshold probabilities. The

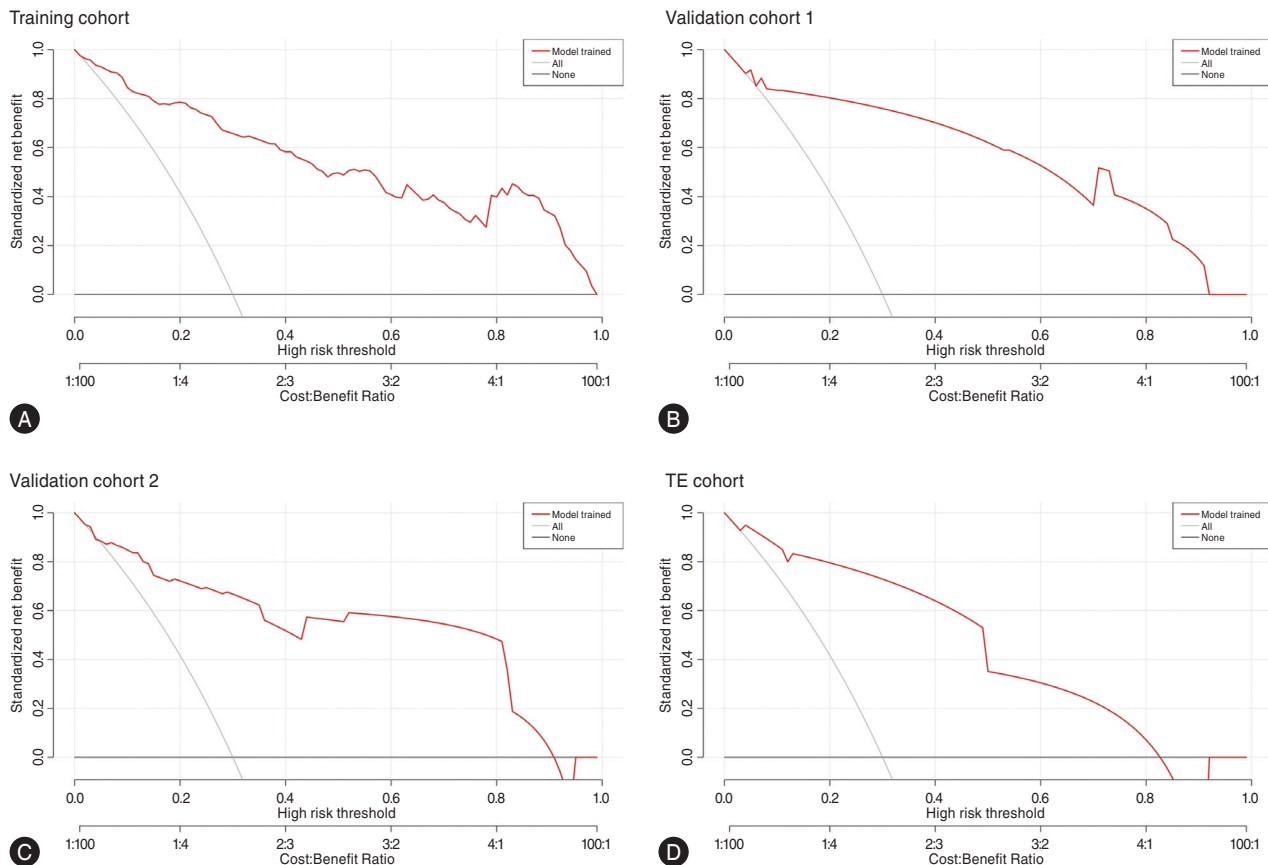


Figure 4. Net benefits of the XGBoost-FIB-4plus model by decision curve analysis for predicting HRV. (A) Training cohort, (B) Validation cohort 1, (C) Validation cohort 2, (D) TE cohort. HRV, high-risk varices; TE, transient elastography; XGBoost, eXtreme Gradient Boosting.

model outperformed the ‘treat none’ strategy, highlighting its practical utility in decision making (Fig. 4, Supplementary Fig. 9). Furthermore, calibration curves demonstrated good diagnostic performance of the XGBoost-FIB-4plus model across the three validation cohorts (Supplementary Figs. 9, 10).

DISCUSSION

Similar to previous studies,^{10,26} our current study highlights that a significant proportion (45.8%) of patients who underwent EGD did not have EV. Notably, the presence of EV on EGD serves as one of the surrogate markers of clinically significant portal hypertension.² This finding underscores that, for well-compensated cirrhotic patients undergoing screening, EGD may lead to unnecessary costs and potential complications without any

added benefits.^{10,21} Consequently, there is a growing interest in developing an accurate and non-invasive tool for the detection of EV in compensated cirrhosis, given the invasive and costly nature of the gold standard method.

We first validated the suboptimal performance of the Baveno VI criteria¹³ and the Baveno VII algorithm¹⁷ (Baveno VI criteria combined with SSM ≤ 40 kPa) in ruling out HRV using 2D-SWE. Similarly, our recent study showed that the Baveno VII algorithm was less effective for ruling out clinically significant portal hypertension, with an NPV $<90\%$.²⁰ To address this, we developed and validated a novel single model, termed FIB-4plus, designed to detect EV and HRV in patients with compensated cirrhosis. The FIB-4plus score is not only highly accurate but is also entirely based on routinely collected data, including elastography examination and laboratory test parameters.

Using a large international multicenter cohort of patients with compensated cirrhosis, we combined LSM, SSM, and

the FIB-4 score into the FIB-4plus score for the first time. The FIB-4plus score demonstrated excellent diagnostic performance in detecting HRV in both the training and validation cohorts. It consistently performed well in the subgroup analysis of patients with HBV infection. Importantly, the FIB-4plus score showed superior predictive performance for EV and HRV in compensated cirrhosis compared with LSM, SSM, PLT, or FIB-4 alone. Furthermore, the XGBoost-FIB-4plus score, trained using the XGBoost algorithm, achieved an AUROC of 0.927 in the training cohort and 0.919 and 0.902 in the first and second external validation cohorts, respectively. The accuracy values ranged from 0.80 to 0.90, with sensitivity and NPV exceeding 0.95 for ruling out HRV, suggesting that the FIB-4plus score is an accurate and non-invasive tool for HRV in patients with compensated cirrhosis. Further analysis using DCA and calibration curves confirmed the practical applicability of the FIB-4plus score in predicting HRV. Thus, our findings indicate that the FIB-4plus score has the potential to predict EV and HRV in patients with compensated cirrhosis, thereby offering clinicians a valuable tool to guide clinical management and improve patient outcomes.

In this study, we chose to combine LSM, SSM, and FIB-4 values because each parameter measures a different yet complementary aspect of advanced liver disease.²⁷ In recent years, numerous studies have demonstrated that LSM is a promising non-invasive tool that assesses the degree of liver fibrosis, which is crucial in diagnosing conditions such as cirrhosis and portal hypertension.^{28,29} The correlation between LSM and the extent of liver fibrosis makes it valuable in predicting the progression and potential complications of liver diseases.^{30,31} At the same time, performing SSM alongside LSM during the same examination is simple and fast. The SSM can reflect changes related to portal hypertension and its severity, irrespective of the etiology.²⁸ High SSM values are associated with a higher risk of variceal bleeding and other portal hypertension-related complications. A recent meta-analysis has shown that, compared with LSM, the diagnostic performance of SSM was significantly better for detecting the presence of EV.³² The FIB-4 score, based on blood laboratory parameters, is one of the most widely adopted serum markers, which can be increased in patients with advanced liver disease.³¹ However, there is

only little data on the combination of two or more non-invasive tests to predict portal hypertension and/or its related complications.^{15,21,26,27}

Interestingly, a recent study by Vutien et al.²⁷ introduced the FIB-5 score, a system that combines the FIB-4 score and TE-derived LSM. This research demonstrated that the FIB-5 score offers superior predictive accuracy for assessing the risk of complications (ascites, hepatic encephalopathy, or variceal bleeding) related to portal hypertension in individuals with compensated liver disease (AUROC=0.868), outperforming the predictive capabilities of the FIB-4 score (AUROC=0.672) and LSM (AUROC=0.688) when used independently.

With the advancement of machine learning, clinicians can now transform large volumes of data into practical models, significantly enhancing their ability to diagnose diseases more effectively.^{25,33} A machine learning model was previously described for the non-invasive detection of EV and HRV in patients with cirrhosis.¹⁰ This approach utilized a machine learning algorithm with routine laboratory parameters, including international normalized ratio, AST, PLT, urea nitrogen, hemoglobin, and the presence of ascites. In the validation cohort, the AUROC ranged from 0.75 to 0.82. This is an example of a non-invasive assessment of HRV using a machine learning algorithm, but its predictive capability was not comparable to that we observed across all three datasets in our study (AUROC ranged from 0.902 to 0.927). Our previous study built a diagnostic model that combined TE-derived LSM, PLT, and total bilirubin based on the light gradient boosting machine algorithm,²¹ which achieved an AUROC of 0.74 to 0.86 in detecting HRV, and spared more screening EGDs in all cohorts as compared with the Baveno VI criteria.

In this study, we further designed and combined the LSM, SSM, and the FIB-4 score into a single, more accurate score (FIB-4plus), offering clinicians a valuable tool for predicting HRV in patients with compensated cirrhosis. Both LR-FIB-4plus and XGBoost-FIB-4plus scores demonstrated higher specificity and overall classification accuracy for predicting HRV than LSM, SSM, PLT, or FIB-4 alone. One particularly noteworthy finding was that SSM emerged as the most important clinical feature in the prediction, representing a more useful parameter for HRV detection. As our results showed, SSM demonstrated superior accuracy and AUROC in predicting

HRV in patients with compensated cirrhosis compared with LSM, PLT, and FIB-4. Researchers have confirmed that SSM is significantly associated with portal hypertension and may be a superior marker for predicting clinical outcomes compared with LSM in patients with cirrhosis.^{15,16,28,32,34,35} In contrast, LSM exhibited a poor correlation with the prediction of clinically significant portal hypertension when used for detecting hepatic venous pressure gradient >12 mmHg, likely due to the increasing influence of extra-hepatic factors on the progression of portal hypertension.³⁶ In fact, both the European Association for the Study of the Liver guidelines³⁷ and the Baveno VII consensus¹⁷ emphasized that SSM plays an important role in the diagnosis of portal hypertension and HRV in patients with compensated advanced chronic liver disease.

Subgroup analyses can provide a more in-depth understanding of the diagnostic performance of the FIB-4plus score in specific patient populations. This study assessed the performance of the FIB-4plus score in a subgroup of patients with HBV infections. Notably, for the XGBoost-FIB-4plus score, an AUROC of over 0.9 was achieved in both the training and validation cohorts. In the training cohort, the FIB-4plus score demonstrated an NPV of 0.865 for ruling out HRV, whereas in the validation cohort, the NPV was 0.946. In addition to the two external validations based on 2D-SWE measurements, we prospectively recruited a cohort of eligible patients with compensated cirrhosis who underwent TE examinations at four Chinese hospitals. We further validated our model for the detection of EV and HRV in these patients. The FIB-4plus score, developed by including the TE-derived LSM and SSM, maintained a good performance in predicting HRV. In the TE cohort, the XGBoost-FIB-4plus score achieved an AUROC of over 0.9. The consistent results across different analyses proved the generalizability and stability of the FIB-4plus score, regardless of the elastographic modality. However, further validation in other ethnic populations is required to determine its broader applicability.

The strengths of this study include its large sample size of patients from two countries, with 502 patients from different institutions included for analysis. This study enrolled a large number of well-characterized, compensated patients. Another important strength is that

we used robust modeling techniques and different strategies to develop the optimal LR and XGBoost models. Specifically, the XGBoost-FIB-4plus score, which included the individual components of the FIB-4 score (i.e., age, AST, ALT, and PLT) as separate predictors rather than the composite FIB-4 score as a single predictor, demonstrated the highest discrimination. Ultimately, we developed a new and readily available scoring system, termed FIB-4plus, based on routinely collected data, including elastography examinations and laboratory test parameters.

There are also limitations to this study that should be acknowledged. First, the FIB-4plus score was developed using data collected prospectively from Chinese patients and externally validated in the Chinese and Croatian populations. Nevertheless, validation studies involving diverse and representative ethnic groups are necessary to ensure generalizability. Second, the etiology of cirrhosis in our study was unevenly distributed, with virus-infected cirrhosis accounting for the majority of cases. Therefore, we performed a subgroup analysis of patients with HBV infection. Future prospective studies on patients with cirrhosis with other etiologies are warranted to validate our model. Additionally, we did not further investigate the potential factors influencing LSM based on different etiologies. Third, we prospectively recruited a cohort of patients who underwent TE examinations at four Chinese hospitals and validated our model for these patients. However, our cohort did not undergo simultaneous 2D-SWE or TE examinations. Thus, we were unable to compare the performance of the FIB-4plus score (which utilizes LSM and SSM from 2D-SWE) and the Baveno criteria (which uses LSM and SSM from TE) in avoiding endoscopic screening. Nevertheless, we did evaluate the performance of the Baveno VI criteria¹³ and the Baveno VII algorithm¹⁷ for ruling out HRV using 2D-SWE. Further studies are required to validate the FIB-4plus score for avoiding endoscopic screening in comparison with the Baveno criteria. Fourth, we did not explore the impact of antiviral therapy and viral suppression status on the efficacy of this score. Further research is needed in the future to investigate the effect of antiviral treatment and viral suppression status on the outcomes studied. Finally, we did not evaluate the prognostic performance of the FIB-4plus score in predicting long-term clinical outcomes, such as decompensation (e.g., ascites, HRV bleeding, and

hepatic encephalopathy) or mortality. Nevertheless, we planned to validate the prognostic relevance of the FIB-4plus score in our prospective observational cohort of patients with compensated cirrhosis in terms of predicting complications related to portal hypertension.

In conclusion, we combined the LSM, SSM, and the FIB-4 score into a single scoring system (FIB-4plus) that could predict HRV in patients with compensated cirrhosis. This novel scoring system is an effective tool for clinicians and holds great promise for guiding clinical management and improving patient outcomes.

Authors' contribution

Study concept and design: Bingtian Dong, Yuping Chen, Ruiling He, Shenghong Ju, and Xiaolong Qi.

Supervision of the study: Xiaolong Qi.

Acquisition of data and technical support: Ruiling He, Shenghong Ju, Ivica Grgurevic, Jianzhong Ma, Ying Guo, Huizhen Fan, Qiang Yan, Chuan Liu, Huixiong Xu, Anita Madir, Kristian Podrug, Jia Wang, Linxue Qian, Zhengzi Geng, Shanghao Liu, Tao Ren, Guo Zhang, Kun Wang, Meiqin Su, Fei Chen, Sumei Ma, Liting Zhang, Zhaowei Tong, Yonghe Zhou, Xin Li, Fanbin He, Hui Huan, Wenjuan Wang, Yunxiao Liang, Juan Tang, Fang Ai, Tingyu Wang, Liyun Zheng, Zhongwei Zhao, Jiansong Ji, Wei Liu, Jiaojiao Xu, Bo Liu, Xuemei Wang, Yao Zhang, Qiong Yan, Hui Liu, Xiaomei Chen, Shuhua Zhang, Yihua Wang, Yang Liu, Li Yin, Yanni Liu, Yanqing Huang, Li Bian, Ping An, Xin Zhang, Shaoting Zhang, Jinhua Shao, Xiangman Zhang, Wei Rao, Chaoxue Zhang, Christoph Frank Dietrich, and Won Kim.

Interpretation of data: Bingtian Dong, Yuping Chen, and Ruiling He.

Drafting of the manuscript: Bingtian Dong and Yuping Chen.

All authors approved this version for submission.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (92359304, 82402413); The Key Research and Development Program of Jiangsu Province (BE2023767a); the Fundamental Research Fund of Southeast University (3290002303A2); Changjiang Scholars Talent Cultivation Project of Zhongda Hospital of Southeast University (2023YJXYRCPY03); Research

Personnel Cultivation Programme of Zhongda Hospital Southeast University (CZXM-GSP-RC125, CZXM-GSP-RC119); China Postdoctoral Science Foundation (2024M750461); Natural Science Foundation of Jiangsu Province (BK20241681); Health Research Program of Anhui (AHWJ2023A30169).

Conflicts of Interest

The authors have no conflicts to disclose.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Clinical and Molecular Hepatology website (<http://www.e-cmh.org>).

REFERENCES

1. Qi X, Berzigotti A, Cardenas A, Sarin SK. Emerging non-invasive approaches for diagnosis and monitoring of portal hypertension. *Lancet Gastroenterol Hepatol* 2018;3:708-719.
2. Kaplan DE, Ripoll C, Thiele M, Fortune BE, Simonetto DA, Garcia-Tsao G, et al. AASLD Practice Guidance on risk stratification and management of portal hypertension and varices in cirrhosis. *Hepatology* 2024;79:1180-1211.
3. Liu C, You H, Zeng QL, Wong YJ, Wang B, Grgurevic I, et al. Carvedilol to prevent hepatic decompensation of cirrhosis in patients with clinically significant portal hypertension stratified by new non-invasive model (CHESS2306). *Clin Mol Hepatol* 2025;31:105-118.
4. Boike JR, Thornburg BG, Asrani SK, Fallon MB, Fortune BE, Izzy MJ, et al. North American practice-based recommendations for transjugular intrahepatic portosystemic shunts in portal hypertension. *Clin Gastroenterol Hepatol* 2022;20:1636-1662.e1636.
5. Liu Y, Tan HY, Zhang XG, Zhen YH, Gao F, Lu XF. Prediction of high-risk esophageal varices in patients with chronic liver disease with point and 2D shear wave elastography: a systematic review and meta-analysis. *Eur Radiol* 2022;32:4616-4627.
6. Reverter E, Tandon P, Augustin S, Turon F, Casu S, Bastiampillai R, et al. A MELD-based model to determine risk of mortality among patients with acute variceal bleeding. *Gastroenterology* 2014;146:412-419.e413.
7. Tapper EB, Friderici J, Borman ZA, Alexander J, Bonder A,

- Nuruzzaman N, et al. A multicenter evaluation of adherence to 4 major elements of the Baveno guidelines and outcomes for patients with acute variceal hemorrhage. *J Clin Gastroenterol* 2018;52:172-177.
8. Wang X, Hu B, Li Y, Lin W, Feng Z, Gao Y, et al. Nationwide survey analysis of esophagogastric varices in portal hypertension based on endoscopic management in China. *Port Hypertens Cirrhosis* 2024;03:129-138.
9. Pan J, Li Z, Liao Z, Liu T, Wang N. Capsule endoscopy in portal hypertension and varices. *Port Hypertens Cirrhosis* 2024;03:171-172.
10. Dong TS, Kalani A, Aby ES, Le L, Luu K, Hauer M, et al. Machine learning-based development and validation of a scoring system for screening high-risk esophageal varices. *Clin Gastroenterol Hepatol* 2019;17:1894-1901.e1891.
11. Barr RG, Wilson SR, Rubens D, Garcia-Tsao G, Ferraioli G. Update to the society of radiologists in ultrasound liver elastography consensus statement. *Radiology* 2020;296:263-274.
12. Loomba R, Adams LA. Advances in non-invasive assessment of hepatic fibrosis. *Gut* 2020;69:1343-1352.
13. de Franchis R. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. *J Hepatol* 2015;63:743-752.
14. Åberg F, Asteljoki J, Männistö V, Luukkonen PK. Combined use of the CLivD score and FIB-4 for prediction of liver-related outcomes in the population. *Hepatology* 2024;80:163-172.
15. Dajti E, Ravaioli F, Zyklus R, Rautou PE, Elkrif L, Grgurevic I, et al. Accuracy of spleen stiffness measurement for the diagnosis of clinically significant portal hypertension in patients with compensated advanced chronic liver disease: a systematic review and individual patient data meta-analysis. *Lancet Gastroenterol Hepatol* 2023;8:816-828.
16. Yoo JJ, Kim SG. The rise of non-invasive tools in the diagnosis of portal hypertension: Validation of the Baveno VII consensus. *Clin Mol Hepatol* 2023;29:102-104.
17. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol* 2022;76:959-974.
18. Zhang X, Chen C, Yan C, Song T. Accuracy of 2D and point shear wave elastography-based measurements for diagnosis of esophageal varices: a systematic review and meta-analysis. *Diagn Interv Radiol* 2022;28:138-148.
19. Singh S, Eaton JE, Murad MH, Tanaka H, Iijima H, Talwalkar JA. Accuracy of spleen stiffness measurement in detection of esophageal varices in patients with chronic liver disease: systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2014;12:935-945.e934.
20. He R, Liu C, Grgurevic I, Guo Y, Xu H, Liu J, et al. Validation of Baveno VII criteria for clinically significant portal hypertension by two-dimensional shear wave elastography. *Hepatol Int* 2024;18:1020-1028.
21. Huang Y, Li J, Zheng T, Ji D, Wong YJ, You H, et al. Development and validation of a machine learning-based model for varices screening in compensated cirrhosis (CHESS2001): an international multicenter study. *Gastrointest Endosc* 2023;97:435-444.e432.
22. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* 2006;44:217-231.
23. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;43:1317-1325.
24. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA: Association for Computing Machinery, 2016:785-794.
25. Rui F, Yeo YH, Xu L, Zheng Q, Xu X, Ni W, et al. Development of a machine learning-based model to predict hepatic inflammation in chronic hepatitis B patients with concurrent hepatic steatosis: a cohort study. *EClinicalMedicine* 2024;68:102419.
26. Colecchia A, Ravaioli F, Marasco G, Colli A, Dajti E, Di Biase AR, et al. A combined model based on spleen stiffness measurement and Baveno VI criteria to rule out high-risk varices in advanced chronic liver disease. *J Hepatol* 2018;69:308-317.
27. Vutien P, Berry K, Feng Z, VoPham T, He Q, Green PK, et al. Combining FIB-4 and liver stiffness into the FIB-5, a single model that accurately predicts complications of portal hypertension. *Am J Gastroenterol* 2022;117:1999-2008.
28. Reiberger T. The value of liver and spleen stiffness for evaluation of portal hypertension in compensated cirrhosis. *Hepatol Commun* 2022;6:950-964.
29. Pons M, Augustin S, Scheiner B, Guillaume M, Rosselli M, Rodrigues SG, et al. Noninvasive diagnosis of portal hypertension in patients with compensated advanced chronic liver disease. *Am J Gastroenterol* 2021;116:723-732.
30. Semmler G, Yang Z, Fritz L, Köck F, Hofer BS, Balcar L, et al. Dynamics in liver stiffness measurements predict outcomes in advanced chronic liver disease. *Gastroenterology* 2023;165:1041-1052.

31. Anstee QM, Castera L, Loomba R. Impact of non-invasive biomarkers on hepatology practice: Past, present and future. *J Hepatol* 2022;76:1362-1378.
32. Manatsathit W, Samant H, Kapur S, Ingviya T, Esmadi M, Wijarnpreecha K, et al. Accuracy of liver stiffness, spleen stiffness, and LS-spleen diameter to platelet ratio score in detection of esophageal varices: Systemic review and meta-analysis. *J Gastroenterol Hepatol* 2018;33:1696-1706.
33. Reiniš J, Petrenko O, Simbrunner B, Hofer BS, Schepis F, Scoppettuolo M, et al. Assessment of portal hypertension severity using machine learning models in patients with compensated cirrhosis. *J Hepatol* 2023;78:390-400.
34. Vanderschueren E, Armandi A, Kwanten W, Cassiman D, Francque S, Schattenberg JM, et al. Spleen stiffness-based algorithms are superior to Baveno VI criteria to rule out varices needing treatment in patients with advanced chronic liver disease. *Am J Gastroenterol* 2024;119:1515-1524.
35. Takuma Y, Morimoto Y, Takabatake H, Toshikuni N, Tomokuni J, Sahara A, et al. Measurement of spleen stiffness with acoustic radiation force impulse imaging predicts mortality and hepatic decompensation in patients with liver cirrhosis. *Clin Gastroenterol Hepatol* 2017;15:1782-1790.e1784.
36. Vizzutti F, Arena U, Romanelli RG, Rega L, Foschi M, Colagrande S, et al. Liver stiffness measurement predicts severe portal hypertension in patients with HCV-related cirrhosis. *Hepatology* 2007;45:1290-1297.
37. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol* 2021;75:659-689.