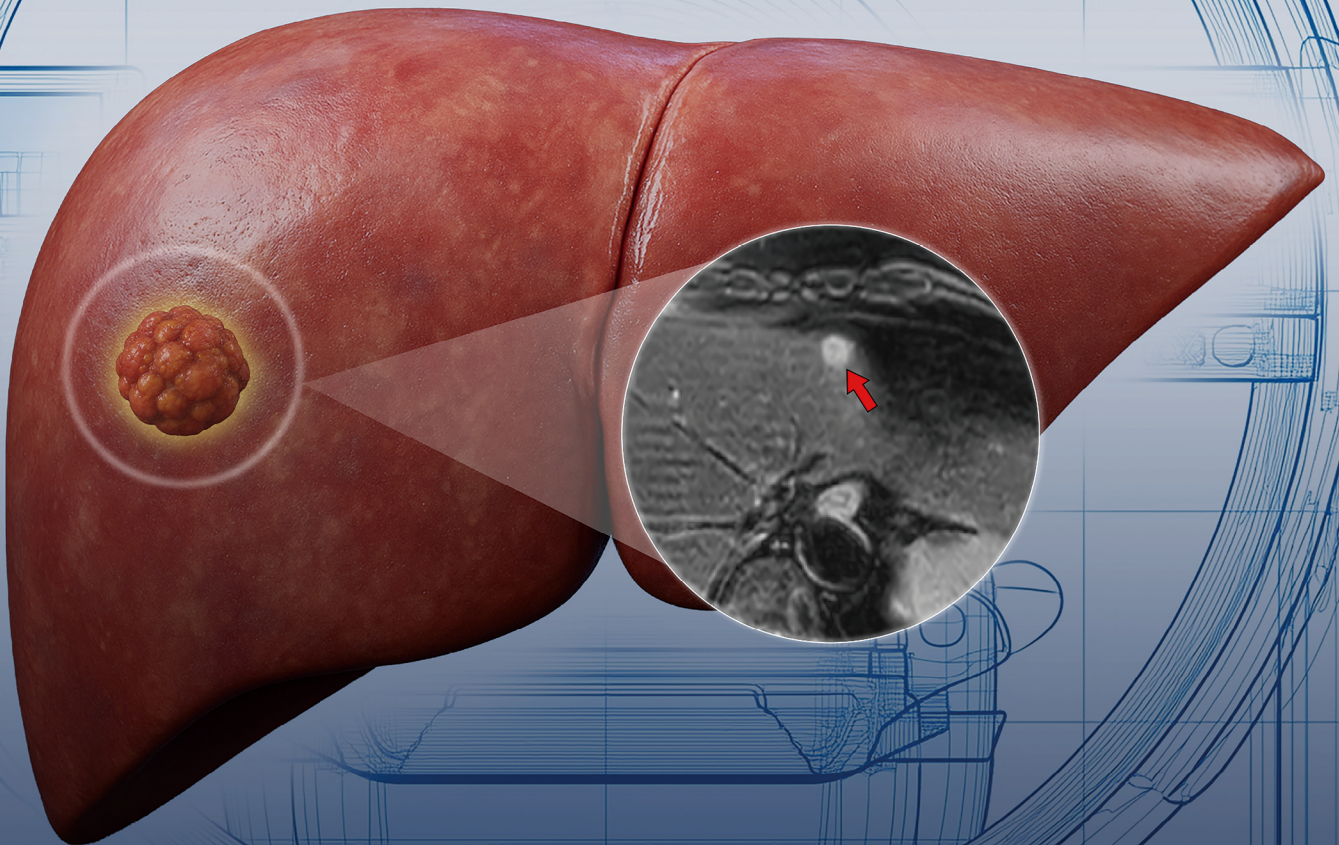


CLINICAL and MOLECULAR HEPATOLOGY

The forum for latest knowledge of hepatobiliary diseases



Non-contrast MRI for detection of recurrent HCC

Gene-BMI-wide interaction study and MASLD
Inflammatory imprint in chronic HCV patients
Risk factors of perinatal hepatitis B infection
Cholestasis and bezafibrate efficacy

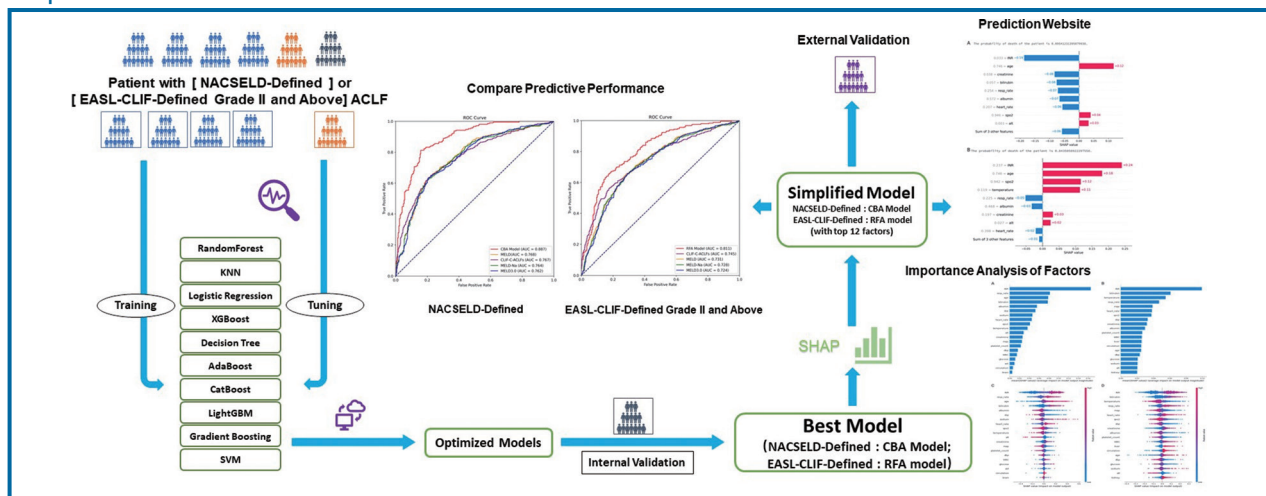
Original Article

Predictive machine learning model in intensive care unit patients with acute-on-chronic liver failure and two or more organ failures

Yee Hui Yeo^{1,*}, Mengyi Zhang^{2,*}, Martin S. McCoy^{3,*}, Jian Zu², Yingli He⁴, Yi Liu⁵, Juan Li⁴, Taotao Yan⁴, Yuan Wang⁵, Hirsh D. Trivedi^{1,6}, Ju Dong Yang^{1,6,7}, Vinay Sundaram^{1,6}, Xiaodan Sun², Zhujun Cao⁸, Chun-Ying Wu^{9,10,11}, Jonel Trebicka^{3,12,13}, and Fanpu Ji^{5,14,15,16,17}

¹Karsh Division of Gastroenterology and Hepatology, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ²School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China; ³Department of Internal Medicine B, University of Münster, Münster, Germany; ⁴Department of Infectious Diseases, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ⁵Department of Hepatology and Infectious Diseases, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ⁶Comprehensive Transplant Center, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ⁷Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ⁸Department of Infectious Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; ⁹Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei; ¹⁰Division of Translational Research, Taipei Veterans General Hospital, Taipei; ¹¹College of Public Health, China Medical University, Taichung, Taiwan; ¹²European Foundation for Study of Chronic Liver Failure, Barcelona, Spain; ¹³Department of Gastroenterology and Hepatology, Odense University Hospital, Odense, Denmark; ¹⁴Key Laboratory of Environment and Genes Related to Diseases (Xi'an Jiaotong University), Ministry of Education of China, Xi'an, China; ¹⁵Key Laboratory of Surgical Critical Care and Life Support (Xi'an Jiaotong University), Ministry of Education, Xi'an, China; ¹⁶Shaanxi Provincial Clinical Medical Research Center of Infectious Diseases, Xi'an, China; ¹⁷National & Local Joint Engineering Research Center of Biodiagnosis and Biotherapy, the Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Graphical Abstract



Study Highlights

- Machine learning model predicts 30-day mortality in ACLF patients with two or more organ failures.
- The CatBoost ACLF model (AUC 0.87) and Random Forest ACLF model (AUC 0.83) outperformed existing traditional scoring systems.
- An online prediction website was established to assist physicians in using machine learning prediction model to assess the mortality risk of ACLF patients.

Background/Aims: Prediction of short-term mortality in patients with acute-on-chronic liver failure (ACLF) admitted to the intensive care unit (ICU) may enhance effective management.

Methods: To develop, explain, and validate a predictive machine learning (ML) model for short-term mortality in patients with ACLF with two or more organ failures (OFs). Utilizing a large ICU cohort with detailed clinical information, we identified ACLF patients with two or more OFs according to the EASL-CLIF and NACSELD definitions. ML model was developed for each definition to predict 30-day mortality. The Shapley value was estimated to explain the models. Validation and calibration of these models were performed.

Results: Of 5,994 patients with cirrhosis admitted to ICU, 1,511 met NACSELD criteria, and 1,692 met EASL-CLIF grade II or higher criteria. The CatBoost ACLF (CBA) model had the greatest accuracy in the NACSELD cohort (area under curve [AUC] of 0.87), while the Random Forest ACLF (RFA) model performed best in the EASL-CLIF cohort (AUC of 0.83). Both models showed robust calibration. The models were explained by SHAP score analysis, yielding a rank list, and the top twelve predictors were selected. Both simplified models demonstrated similar performance (CBA model: AUC 0.89, RFA model: AUC 0.81) and significantly outperformed contemporary scoring systems, including CLIF-C ACLF and MELD 3.0. The models were validated in both internal and external cohorts. A simple-to-use online tool was created to predict mortality rates.

Conclusions: We presented explainable, well-validated, and calibrated predictive models for ACLF patients with two or more OFs, which outperformed existing predictive scores. ([Clin Mol Hepatol 2025;31:1355-1371](#))

Keywords: Intensive care; Organ failure; Cirrhosis; NACSELD criteria; EASL-CLIF criteria

Corresponding author : Yee Hui Yeo

Karsh Division of Gastroenterology and Hepatology, Department of Medicine, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
Tel: +1-703-310-9292, Fax: +86-29-87678223, E-mail: yeehui.yeo@cshs.org
<https://orcid.org/0000-0002-2703-5954>

Jian Zu

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China
Tel: +86-29-82663149, Fax: +86-29-82668551, E-mail: jianzu@xjtu.edu.cn
<https://orcid.org/0000-0003-3633-0725>

Jonel Trebicka

Department of Internal Medicine B, University of Münster, Münster 48149, Germany
Tel: +49-251-83-59689, Fax: +49-251-83-48288, E-mail: Jonel.Trebicka@ukmuenster.de
<https://orcid.org/0000-0002-7028-3881>

Fanpu Ji

Department of Hepatology and Infectious Diseases, The Second Affiliated Hospital of Xian Jiaotong University, No.157 Xi Wu Road, Xi'an 710004, Shaanxi, China
Tel: +86-29-87678223, Fax: +86-29-87678223, E-mail: jifanpu1979@163.com
<https://orcid.org/0000-0002-1463-8035>

*These authors contributed equally to this work and are co-first authors.

Editor: Do Seon Song, The Catholic University of Korea, Korea

Received : May 28, 2025 / **Revised :** Aug. 11, 2025 / **Accepted :** Aug. 24, 2025

Abbreviations:

ACLF, acute on chronic liver failure; AdaBoost, Adaptive Boosting; ALT, alanine transaminase; APASL, Asian Pacific Association of Liver; AST, aspartate transaminase; AUC, area under curve; AUROC, area under the receiver-operating characteristic curve; BIDMC, Beth Israel Deaconess Medical Center; CatBoost, Categorical Boosting; CBA, CatBoost ACLF; CI, confidence interval; DBP, diastolic blood pressure; EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure; FIO₂, fraction of inspired oxygen; GBM, Gradient Boosting Machine; GCS, Glasgow Coma Scale; ICU, intensive care unit; IMC, intermediate care unit; INR, international normalized ratio; IQR, interquartile range; KNN, K-Nearest Neighbors; LightGBM, Light Gradient Boosting Machine; MAP, mean arterial pressure; MIMIC, Medical Information Mart for Intensive Care; ML, machine learning; NACSELD, North American Consortium for the Study of End-Stage Liver Disease; OFs, organ failures; PaO₂, partial pressure of arterial oxygen; RFA, Random Forest ACLF; SBP, systolic blood pressure; SHAP, SHapley Additive exPlanations; SpO₂, peripheral oxygen saturation; SVM, Support Vector Machine; UNOS, United Network for Organ Sharing; WBC, white blood cell count; XGBoost, eXtreme Gradient Boosting

INTRODUCTION

Acute-on-chronic liver failure (ACLF) represents a severe and prevalent complication among patients with cirrhosis globally.^{1,2} Especially in patients with ACLF and the development of two or more extrahepatic organ failures (OFs) following an acute insult,² intensive care unit (ICU)-admission is often required.³⁻⁶

In general, ACLF is associated with a high short-term mortality rate.^{1,2} Therefore, accurately determining disease severity and identifying key prognostic factors are crucial to optimizing care strategies. Especially, patients with ≥ 2 OFs, reflecting grade II ACLF according to the European Association for the Study of the Liver-Chronic Liver Failure (EASL-CLIF) definition and then North American Consortium for the Study of End-stage Liver Disease (NACSELD) definition of ACLF, may require huge medical resources. The conventional belief that an increased number of OFs directly correlates with higher mortality fails to account for the differential impact of specific OFs on patient prognosis.⁷ Clinical observations suggest that the mortality risk associated with ACLF is not uniformly influenced by all types of OF. However, there is a paucity of research exploring the hierarchical influence of different OFs on mortality.

The application of machine learning (ML) models in predicting outcomes for critically ill patients has seen significant growth.⁸⁻¹⁰ However, none have been developed for this specific ACLF patient population.¹¹ Moreover, the limited transparency of deep learning models hinders their clinical effectiveness and trustworthiness, as it constrains the interpretability of these models.¹² In this study, we aimed to utilize a large ICU cohort to develop explainable ML models capable of predicting short-term mortality in patients with ACLF ≥ 2 OFs and to elucidate the hierarchical relationships among predictive factors.

MATERIALS AND METHODS

Data source

This is an ICU-based retrospective cohort study. Data were obtained from the Medical Information Mart for Intensive Care- IV (MIMIC-IV) database (version 2.2), a publicly available database that documented de-identified electronic

health records from ICUs at the Beth Israel Deaconess Medical Center (BIDMC) between 2008–2019.¹³⁻¹⁵ Given the public availability and the de-identification of the data, approval from the local institutional board review was not obtained. The reporting of this study follows both STROBE¹⁶ and TRIPOD+AI statements (Supplementary Tables 1, 2).¹⁷

Definitions

The study explored two ACLF populations: those meeting NACSELD criteria and those with EASL-CLIF ACLF grade \geq II.^{18,19} Per the NACSELD definition, ACLF is characterized by the presence of two or more OFs in the kidney, brain, circulation, and respiration in patients with cirrhosis.²⁰ (for detailed information on definitions and patient selection criteria, see Supplementary Method and Supplementary Table 3).

Data description

A total of 24 variables were used, including demographic characteristics, drug use, vital signs, laboratory tests, and parameters of OFs (for detailed information on variables used, see Supplementary Method).

Study outcomes and design

The primary outcome was defined as the 30-day mortality rate. The MIMIC-IV database provides comprehensive mortality data through two reliable sources: the hospital information system and the Massachusetts State Registry of Vital Records and Statistics.²¹ This robust data linkage ensures complete mortality follow-up regardless of whether patients continued ICU care or were transferred to other wards. The secondary outcome involved interpreting the selected model to determine the predictors and their impact on 30-day mortality. We aimed to develop predictive models for patients with ACLF ≥ 2 OFs (NACSELD definition and grade two or higher ACLF, EASL-CLIF definition) (for detailed information on study design, see Supplementary Method).

Statistical analysis

All statistical analyses were performed using Python

(version 3.8.18) and R (version 4.3.2). Descriptive analysis was performed to summarize the baseline characteristics of patients with ACLF with at least two OFs.

To determine the required sample size for the prediction model, the `pmsampsize` package in R was used, following the criteria from Riley.²² The CLIF-C ACLF score served as the benchmark model, which was based on a cohort of 202 patients, 99 of whom (49%) died within 28 days. The model used three predictors and achieved an area under curve (AUC) of 0.8 for predicting 28-day mortality.⁴ The sample size calculation indicated that at least 385 samples, including 189 deaths, were required for reliable model development. The sample size in the study was sufficient to meet this requirement.

The ACLF cohort was divided into training (70%) and internal (30%) validation datasets through random sampling (Supplementary Fig. 1) (for detailed information on the predictive model development process, see Supplementary Method).

Model interpretation

Upon finalizing the predictive model, the SHapley Additive exPlanations (SHAP) were used for model interpretability. It quantified the contribution of each feature toward the overall prediction for a given observation.²³ A ranked list of variables according to the Shapley value was generated.

To enhance usability and practicality, we further refined the top-performing model for each definition by selecting the top twelve variables using embedding methods, packing methods, and SHAP calculations. Model performance was assessed in both training and internal validation cohorts. Additionally, AUC values of the final model were compared to four commonly used scores (MELD, MELD-Na, MELD 3.0, and CLIF-C ACLF) using the DeLong test.

External validation

In our external validation cohort, we validated our findings in a cohort of 498 patients with liver cirrhosis and admission to intermediate care unit (IMC)/ICU at a tertiary university clinic in Germany (for detailed information on external validation cohort, see Supplementary Method and Supplementary Fig. 2).

Website generation

We translated the findings into an interactive online calculator. This tool was designed to enable physicians and researchers to utilize or validate our model by providing a 30-day mortality rate and the direction and relative weight of the factors in predicting mortality. The latter feature indicates whether the predictors' impact is positive (increasing risk) or negative (decreasing risk).

RESULTS

Baseline characteristics of patients

Supplementary Fig. 3 delineates the patient selection process. From the MIMIC-IV database, which initially included 5,994 patients with cirrhosis, a total of 1,591 patients met the criteria of ACLF according to the NACSELD definition and 1,760 patients met the criteria of grade II and above ACLF according to the EASL-CLIF definition. Following the imputation of missing data for patients with five or fewer missing variables (Supplementary Table 4A, 4B), the final cohort for model development comprised 1,511 (NACSELD definition) and 1,692 (EASL-CLIF definition) patients with two or more OFs.

The distribution of demographic and clinical characteristics of these patients with two or more OFs is presented in Supplementary Table 5A and 5B, stratified by severity levels. In the NACSELD criteria-defined ACLF cohort, the median age was 59 years (interquartile range [IQR]: 52–67 years), and 63.3% of the cohort was male. The prevalence of OFs was as follows: respiratory failure (83.5%), circulatory failure (73.5%), brain failure (50.3%), and kidney failure (22.7%). Among those with grade II and above ACLF (EASL-CLIF definition), the median age was 59 years (IQR: 52–68 years), and 63.5% were male. circulatory failure (59.0%) exhibited the highest prevalence, followed by lung failure (56.7%), kidney failure (50.6%), and coagulation failure (33.9%). Liver failure (23.2%) was the least common OF. Notably, in both definitions, patients with more advanced ACLF exhibited significantly higher levels of aspartate transaminase (AST), alanine transaminase (ALT), total bilirubin, international normalized ratio (INR), and creatinine, along with lower albumin levels. Table 1 and Table 2

Table 1. Distribution of demographic and clinical characteristics in patients with acute-on-chronic liver failure (ACLF) according to the NACSELD ACLF definition, categorized by survival status at day 30

Variable	Overall (n=1,511)	Survival (n=788)	Death (n=723)	P-value
Age, years	59.0 (52.0, 67.0)	58.0 (51.0, 66.0)	60.0 (52.0, 69.0)	0.002
Male	956 (63.3)	499 (63.3)	457 (63.2)	0.963
Etiology				
ALD	826 (54.7)	429 (54.4)	397 (54.9)	0.855
HBV	32 (2.1)	13 (1.6)	19 (2.6)	0.187
HCV	125 (8.3)	63 (8.0)	62 (8.6)	0.683
MASLD	89 (5.9)	49 (6.2)	40 (5.5)	0.572
Immune-related	85 (5.6)	55 (7.0)	30 (4.1)	0.017
Genetic-related	12 (0.8)	2 (0.3)	10 (1.4)	0.014
Cryptogenic cirrhosis	473 (31.3)	241 (30.6)	232 (32.1)	0.529
Others	18 (1.2)	9 (1.1)	9 (1.2)	0.854
Complication				
CKD	229 (15.2)	109 (34.4)	120 (16.6)	0.134
COPD	183 (12.1)	96 (12.2)	87 (12.0)	0.929
IHD	155 (10.3)	92 (11.7)	63 (8.7)	0.058
Hypertension	530 (35.1)	291 (36.9)	239 (33.1)	0.115
Antibiotic	1,196 (79.2)	613 (77.8)	583 (80.6)	0.174
Bleeding	606 (40.1)	290 (36.8)	316 (43.7)	0.006
Organ failure				
Brain	760 (50.3)	396 (50.3)	364 (50.3)	0.972
Circulation	1,110 (73.5)	526 (66.8)	584 (80.8)	<0.001
Kidney	343 (22.7)	144 (18.3)	199 (27.5)	<0.001
Respiratory	1,262 (83.5)	684 (86.8)	578 (79.9)	<0.001
Mechanical ventilation				
Respiratory failure	1,016 (67.2)	525 (66.6)	491 (67.9)	0.595
Brain failure	143 (9.5)	95 (12.1)	48 (6.6)	<0.001
Other	103 (6.8)	64 (8.1)	39 (5.4)	0.036
Score				
MELD	22.6 (14.2, 31.2)	16.3 (11.3, 25.1)	28.8 (20.8, 35.3)	<0.001
MELD-Na	23.6 (14.5, 32.3)	17.4 (11.5, 26.3)	29.7 (21.5, 35.9)	<0.001
Vital signs				
Heart rate (bpm)	89.3 (77.3, 103.0)	85.4 (75.4, 99.2)	94.0 (80.9, 106.0)	<0.001
SBP (mmHg)	107.3 (99.7, 117.0)	110.7 (102.7, 121.3)	103.7 (95.7, 112.6)	<0.001
DBP (mmHg)	58.0 (51.7, 64.5)	59.7 (54.1, 66.5)	55.6 (49.4, 61.7)	<0.001
MAP (mmHg)	72.1 (66.3, 78.6)	74.3 (69.1, 81.2)	69.3 (63.3, 75.9)	<0.001
Respiratory rate (bpm)	19.3 (16.6, 22.6)	18.2 (16.0, 20.8)	20.8 (17.8, 24.0)	<0.001
Temperature (°C)	36.8 (36.5, 37.1)	36.9 (36.6, 37.2)	36.7 (36.4, 37.0)	<0.001
SpO ₂ (%)	97.5 (95.8, 98.9)	98.1 (96.6, 99.2)	96.7 (95.0, 98.3)	<0.001
Laboratory testing				
Glucose (mg/dL)	135.5 (110.3, 172.0)	139.0 (116.0, 177.5)	131.0 (103.1, 166.3)	<0.001
AST (U/L)	84.0 (45.0, 191.5)	76.0 (43.4, 188.2)	90.0 (48.0, 195.2)	0.900
ALT (U/L)	41.0 (24.0, 102.2)	42.0 (24.0, 119.9)	40.3 (24.0, 88.7)	0.018

Table 1. Continued

Variable	Overall (n=1,511)	Survival (n=788)	Death (n=723)	P-value
Bilirubin (mg/dL)	3.8 (1.6, 9.5)	2.5 (1.3, 5.4)	5.9 (2.4, 15.2)	<0.001
INR	1.8 (1.4, 2.3)	1.5 (1.3, 1.9)	2.2 (1.7, 2.7)	<0.001
WBC (K/ μ L)	10.8 (7.0, 15.9)	9.9 (6.4, 14.4)	12.3 (8.1, 17.9)	<0.001
Platelet count (K/ μ L)	98.7 (64.0, 150.5)	106.0 (72.6, 157.5)	88.0 (57.6, 139.1)	<0.001
Creatinine (mg/dL)	1.5 (0.9, 2.8)	1.2 (0.8, 2.2)	2.0 (1.2, 3.3)	<0.001
Sodium (mEq/L)	138.0 (134.0, 141.7)	138.8 (135.5, 141.5)	137.0 (133.0, 142.0)	0.005
Albumin (g/dL)	3.0 (2.6, 3.4)	3.1 (2.8, 3.4)	2.9 (2.5, 3.4)	<0.001

Values are presented as median (interquartile range) or number (%).

NACSELD, North American Consortium for the Study of End-Stage Liver Disease; ALD, alcoholic liver disease; HBV, hepatitis B virus; HCV, hepatitis C virus; MASLD, metabolic dysfunction-associated steatotic liver disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; SpO₂, peripheral oxygen saturation; ALT, alanine transaminase; AST, aspartate transaminase; DBP, diastolic blood pressure; INR, international normalized ratio; MAP, mean arterial pressure; SBP, systolic blood pressure; WBC, white blood cell count.

show the demographic distribution, clinical characteristics and etiology (Supplementary Table 6) of ACLF patients with two or more OFs, grouped according to their 30-day survival status. For both NACSELD and EASL-CLIF definitions, the group that survived had a notably lower prevalence of kidney and circulatory failures compared to those who did not survive. Vital signs and the majority of laboratory measurements exhibited significant differences between the survival and mortality groups. Based on EASL-CLIF criteria, when partial pressure of arterial oxygen (PaO₂)/fraction of inspired oxygen (FiO₂) \leq 200 or peripheral oxygen saturation (SpO₂)/FiO₂ \leq 214, we could consider the patient to have respiratory failure. We counted the actual rates of respiratory failure and brain failure in patients receiving mechanical ventilation in the NACSELD and EASL-CLIF cohorts (Tables 1, 2 and Supplementary Table 5).

The mortality probability on day 30 in the NACSELD ACLF cohort was 47.8% (Fig. 1A), while it was 49.2% in the EASL-CLIF Grade II and above ACLF cohort (Fig. 1B). The EASL-CLIF cohort was able to further differentiate mortality risks among subgroups, showing distinct mortality rates for patients with 2, 3, and 4–6 OFs. Specifically, mortality rates were 37.8%, 55.2%, and 72.1% for patients with 2, 3, and 4–6 OFs, respectively. In comparison, the stratification of the NACSELD cohort was less discriminative (Fig. 1C, 1D).

Model development

The predictive performance of the top 10 combinations of predictive models determined using the training dataset

(left panel) and the results of internal validation (right panel) are demonstrated in Supplementary Table 7A (NACSELD definition) and 7B (EASL-CLIF definition). In the NACSELD ACLF cohort, the CatBoost ACLF (CBA) model exhibited the highest accuracy for predicting 30-day mortality in the overall cohort in both training and internal validation datasets, achieving an area under the receiver-operating characteristic curve (AUROC) of 0.87, sensitivity of 0.86, and specificity of 0.72 (Supplementary Table 7A). In the EASL-CLIF ACLF cohort, the Random Forest ACLF (RFA) model performed best, achieving an AUROC of 0.83, sensitivity of 0.82, and specificity of 0.67 (Supplementary Table 7B). Figure 2A and 2B display the ROC curves and relevant AUC values for the validation model in the NACSELD and EASL-CLIF ACLF cohorts, respectively.

The comparison between the CBA and RFA models' predicted result and actual mortality exhibited consistency (Fig. 2C [NACSELD ACLF cohort] and 2D [EASL-CLIF ACLF cohort]). The categorization of patients based on the mortality probability yielded three distinct mortality probability curves (Fig. 2E [NACSELD ACLF cohort] and 2F [EASL-CLIF ACLF cohort]). The risk stratifications in both cohorts were more effective than conventional stratification, as exhibited in Figure 1C and 1D.

The decision curve analysis shows the net benefits of the best model for ACLF patients with two or more OFs under different probability thresholds (Supplementary Fig. 4). In both NACSELD and EASL-CLIF cohorts, the predictive models exhibited a slow downward trend, and the net benefit remained above zero at most thresholds, demonstrat-

Table 2. Distribution of demographic and clinical characteristics in patients with acute-on-chronic liver failure (ACLF) according to the EASL-CLIF definition grade II and above, categorized by survival status at day 30

Variable	Overall (n=1,692)	Survival (n=860)	Death (n=832)	P-value
Age, years	59.0 (52.0, 68.0)	59.0 (51.0, 67.0)	60.0 (52.0, 69.0)	0.003
Male	1,074 (63.5)	550 (64.0)	524 (63.0)	0.678
Etiology				
ALD	920 (54.4)	456 (53.0)	464 (55.8)	0.257
HBV	38 (2.2)	17 (2.0)	21 (2.5)	0.448
HCV	140 (8.3)	70 (8.1)	70 (8.4)	0.838
MASLD	103 (6.1)	57 (6.6)	46 (5.5)	0.344
Immune-related	94 (5.6)	57 (6.6)	37 (4.4)	0.050
Genetic-related	14 (0.8)	4 (0.5)	10 (1.2)	0.095
Cryptogenic cirrhosis	532 (31.4)	269 (31.3)	263 (31.6)	0.883
Others	23 (1.4)	13 (1.5)	10 (1.2)	0.583
Complication				
CKD	300 (17.7)	154 (17.9)	146 (17.5)	0.847
COPD	198 (11.7)	106 (12.3)	92 (11.1)	0.417
IHD	179 (10.6)	106 (12.3)	73 (8.8)	0.017
Hypertension	591 (34.9)	316 (36.7)	275 (33.1)	0.111
Antibiotic	1,223 (72.3)	636 (74.0)	587 (70.6)	0.118
Bleeding	652 (38.5)	298 (34.7)	354 (42.5)	<0.001
Organ failure				
Liver	392 (23.2)	126 (14.7)	266 (32.0)	<0.001
Kidney	856 (50.6)	368 (42.8)	488 (58.7)	<0.001
Coagulation	573 (33.9)	202 (23.5)	371 (44.6)	<0.001
Brain	754 (44.6)	383 (44.5)	371 (44.6)	0.981
Circulation	999 (59.0)	482 (56.0)	517 (62.1)	0.011
Lung	960 (56.7)	534 (62.1)	426 (51.2)	<0.001
Mechanical ventilation				
Respiratory failure	946 (55.9)	526 (61.2)	420 (50.5)	<0.001
Brain failure	8 (0.5)	4 (0.5)	4 (0.5)	0.963
Other	6 (0.4)	4 (0.5)	2 (0.2)	0.437
Score				
MELD	24.4 (16.1, 31.5)	19.6 (12.8, 27.2)	28.8 (21.8, 34.7)	<0.001
MELD-Na	25.8 (16.7, 32.8)	20.8 (13.0, 28.5)	30.0 (22.8, 35.5)	<0.001
Vital signs				
Heart rate (bpm)	89.8 (77.4, 103.3)	86.3 (75.5, 98.7)	94.4 (80.8, 107.2)	<0.001
SBP (mmHg)	107.1 (99.0, 117.4)	110.1 (102.0, 121.4)	104.4 (95.8, 113.7)	<0.001
DBP (mmHg)	58.2 (52.0, 64.6)	59.6 (54.0, 66.3)	56.5 (49.7, 63.0)	<0.001
MAP (mmHg)	72.0 (66.2, 78.7)	74.2 (68.2, 81.0)	69.8 (63.8, 76.3)	<0.001
Respiratory rate (bpm)	19.2 (16.6, 22.4)	18.2 (16.0, 20.8)	20.4 (17.4, 23.8)	<0.001
Temperature (°C)	36.7 (36.5, 37.0)	36.8 (36.6, 37.1)	36.6 (36.4, 36.9)	<0.001
SpO ₂ (%)	97.3 (95.6, 98.8)	97.8 (96.3, 99.1)	96.7 (94.8, 98.3)	<0.001
Laboratory testing				
Glucose (mg/dL)	134.3 (109.4, 170.9)	139.0 (115.1, 179.0)	130.0 (103.0, 161.5)	<0.001

Table 2. Continued

Variable	Overall (n=1,692)	Survival (n=860)	Death (n=832)	P-value
AST (U/L)	85.0 (44.0, 202.2)	79.8 (41.3, 210.1)	87.0 (48.0, 192.1)	0.758
ALT (U/L)	41.0 (23.0, 107.0)	41.4 (23.0, 134.1)	41.0 (24.0, 89.3)	0.002
Bilirubin (mg/dL)	3.7 (1.7, 10.3)	2.6 (1.3, 6.2)	5.6 (2.4, 15.2)	<0.001
INR	1.9 (1.5, 2.5)	1.6 (1.3, 2.0)	2.2 (1.7, 2.7)	<0.001
WBC (K/ μ L)	10.9 (7.1, 15.9)	10.0 (6.5, 14.8)	12.0 (7.9, 17.6)	<0.001
Platelet count (K/ μ L)	100.0 (64.0, 152.0)	108.1 (73.6, 158.3)	88.6 (57.7, 142.8)	<0.001
Creatinine (mg/dL)	1.8 (1.0, 2.9)	1.4 (0.8, 2.4)	2.0 (1.2, 3.2)	<0.001
Sodium (mEq/L)	137.7 (133.5, 141.3)	138.3 (134.5, 141.2)	136.7 (132.0, 141.5)	0.002
Albumin (g/dL)	3.0 (2.6, 3.4)	3.0 (2.7, 3.3)	3.0 (2.5, 3.4)	0.008

Values are presented as median (interquartile range) or number (%).

EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure; ALD, alcoholic liver disease; HBV, hepatitis B virus; HCV, hepatitis C virus; MASLD, metabolic dysfunction-associated steatotic liver disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; SpO₂, peripheral oxygen saturation; ALT, alanine transaminase; AST, aspartate transaminase; DBP, diastolic blood pressure; INR, international normalized ratio; MAP, mean arterial pressure; SBP, systolic blood pressure; WBC, white blood cell count.

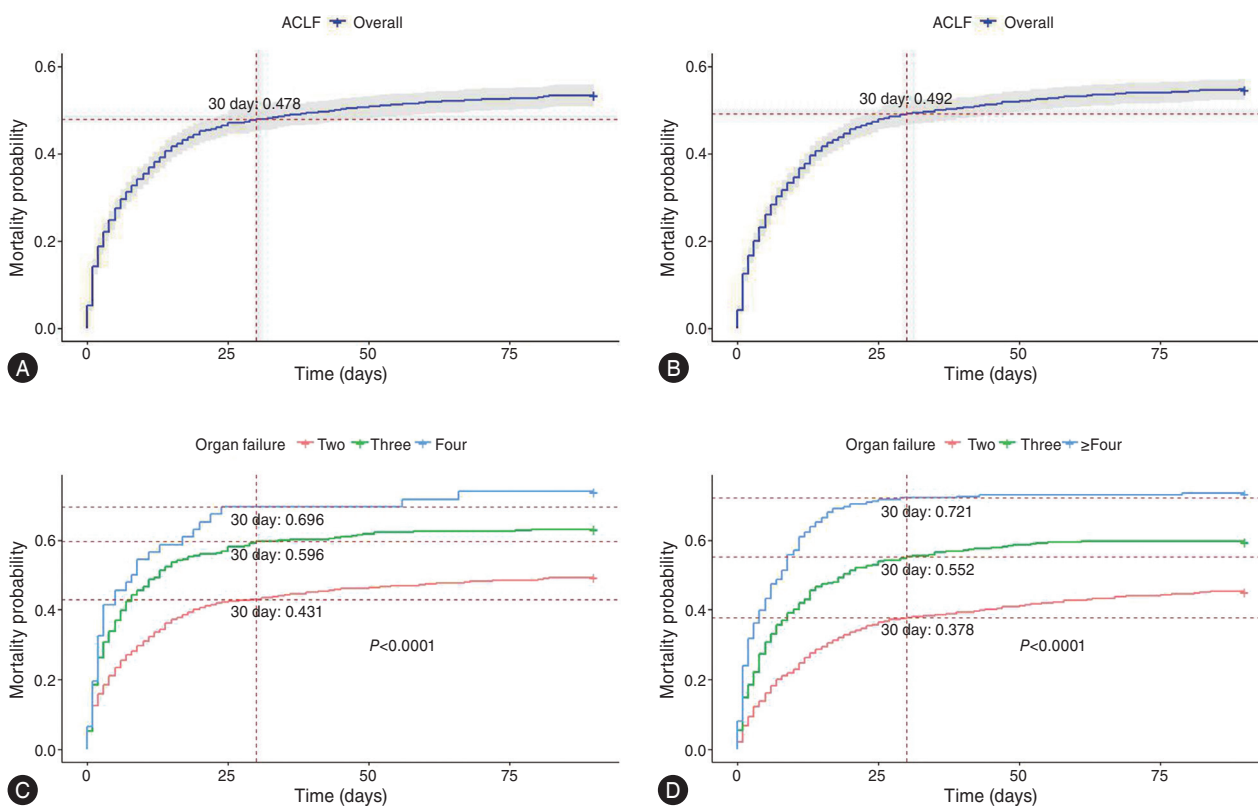


Figure 1. Mortality probability of patients with acute-on-chronic liver failure at day 30. (A) Cohort according to the NACSELD ACLF definition. The dotted line denotes the mortality probability at day 30. (B) Cohort according to the EASL-CLIF definition grade II and above ACLF. The dotted line denotes the mortality probability at day 30. (C) The NACSELD ACLF definition subgroup analysis by the number of organ failures. (D) The EASL-CLIF definition grade II and above ACLF subgroup analysis by the number of organ failures. The red line denotes patients with ACLF with two organ failures, while the green and blue represent three and four organ failures, respectively. P-value was computed using log-rank test to determine the difference in mortality probability across the severity subgroups. NACSELD, North American Consortium for the Study of End-Stage Liver Disease; ACLF, acute-on-chronic liver failure; EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure.

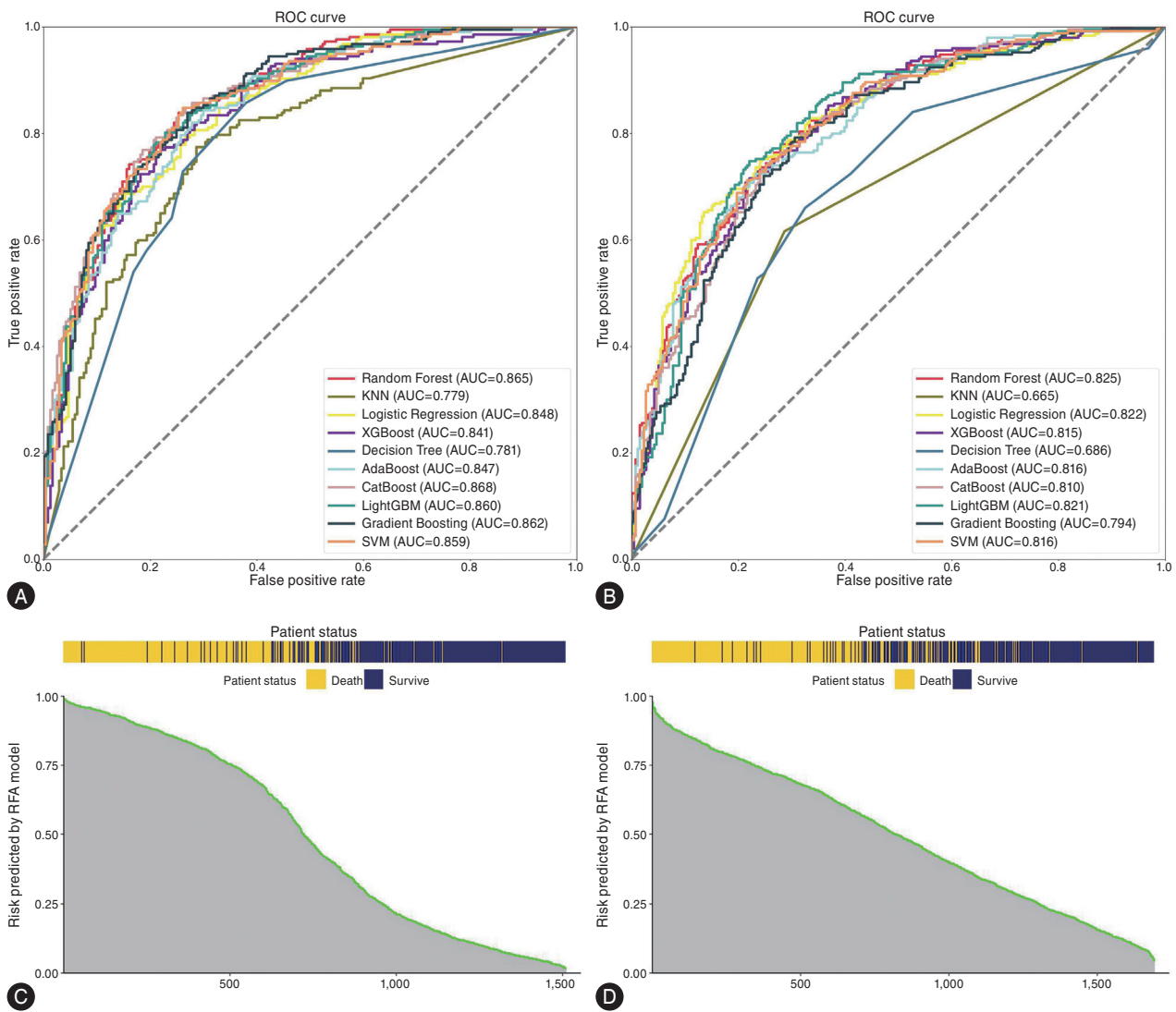


Figure 2. Predictive performance for 30-day mortality of acute-on-chronic liver failure (ACLF). (A) Areas under the receiver operating characteristics (AUROC) curves of the 10 top-performing machine learning models with the cohort according to the NACSELD ACLF definition. (B) AUROC curves of the 10 top-performing machine learning models with the cohort according to the EASL-CLIF definition grade II and above ACLF. The AUC was calculated using the trapezoidal rule, which approximates the area under the ROC curve by summing the areas of trapezoids formed under the curve. (C) Comparison of CBA model predicted mortality rates with actual survival status. (D) Comparison of RFA model predicted mortality rates with actual survival status. The green line represents the predicted probability of death, and the patient status bar represents the patient's 30-day true survival status. The comparison highlights the accuracy of the models in predicting mortality outcomes. (E) Mortality probability of patients with ACLF at day 30 categorized by the prediction of the CBA model. (F) Mortality probability of patients with ACLF at day 30 categorized by the prediction of the RFA model. The red line represents high risk ACLF patients, while green and blue represent low risk and medium risk patients, respectively. The *P*-value is calculated using logarithmic rank test to determine the differences in mortality probability among subgroups of different severity levels. NACSELD, North American Consortium for the Study of End-Stage Liver Disease; EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure; AUC, area under curve; ROC, receiver operating characteristics; CBA, CatBoost ACLF; RFA, Random Forest ACLF; KNN, K-Nearest Neighbors; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; CatBoost, Categorical Boosting; LightGBM, Light Gradient Boosting Machine; SVM, Support Vector Machine.

ing strong clinical utility across most thresholds.

To explain the best-performing models, the variables were screened and ranked according to predictor importance analysis as indicated by multiple methods

(Supplementary Table 8). Of note, the rank lists of SHAP values of both ACLF cohorts showed that INR is the top-ranked predictor (Supplementary Fig. 5).

To enhance the usability of the CBA model and RFA mod-

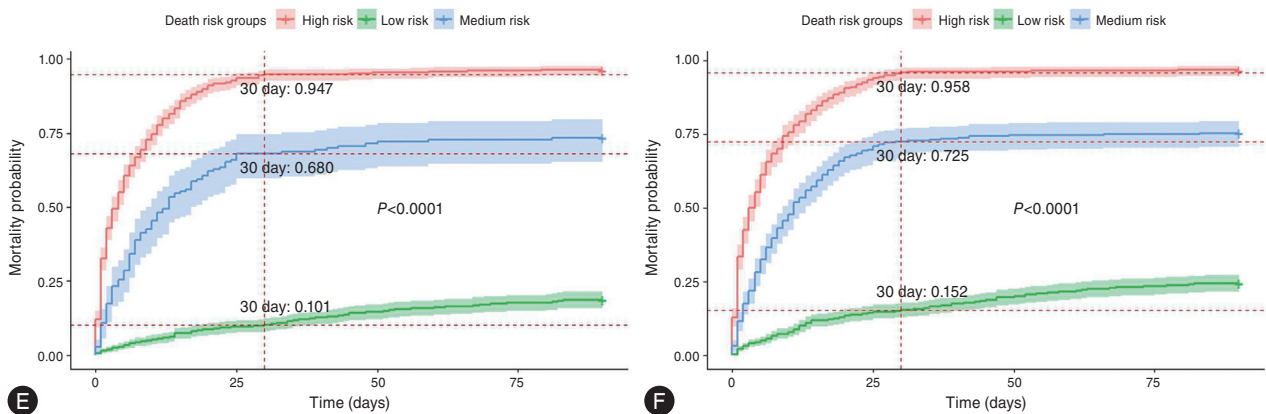


Figure 2. Continued.

Table 3. The predictive performance of the best model for NACSELD ACLF and EASL-CLIF definition grade II and above ACLF cohorts using all or the 12 top-performing predictors

	Metrics	12 factors
CBA model (NACSELD definition)	Specificity	0.814
	Sensitivity	0.813
	Precision	0.813
	F1 score	0.813
	AUC	0.887
RFA model (EASL-CLIF definition grade II and above)	Specificity	0.779
	Sensitivity	0.696
	Precision	0.753
	F1 score	0.723
	AUC	0.811

NACSELD, North American Consortium for the Study of End-Stage Liver Disease; ACLF, acute-on-chronic liver failure; EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure; AUC, area under the curve; CBA, CatBoost ACLF; RFA, Random Forest ACLF.

el, we focused on the top 12 factors identified in both the overall model and the severity subgroup models (Table 3). Using the models with top-performing variables only did not compromise the overall predictive performance. The AUC of the CBA model to predict the 30-day mortality in patients with NACSELD-define ACLF was 0.89, with both sensitivity and specificity of 0.81 (12 factors: INR, respiratory rate, age, total bilirubin, albumin, systolic blood pressure [SBP], sodium, heart rate, SpO₂, body temperature, ALT, creatinine). The AUC of the RFA model in the EASL-CLIF ACLF cohort was 0.81, with a sensitivity of 0.78 and a specificity of 0.70 (12 factors: INR, total bilirubin, body temperature, respiratory rate, mean arterial pressure [MAP],

heart rate, SpO₂, SBP, creatinine, platelet count, white blood cell count [WBC], albumin).

Model comparison and validation

Furthermore, both CBA and RFA models demonstrated significantly higher AUC values compared to conventional predictive models (MELD, MELD-Na, MELD 3.0, and CLIF-C ACLF scores) for predicting 30-day mortality in the respective cohorts (Fig. 3A, 3B). Figure 3C and 3D show the calibration curves of CBA and RFA models compared to conventional prediction models. The calibration curves of the CBA and RFA models were closest to the reference line, with the lowest Brier scores, which significantly outperformed conventional predictive models. The *P*-values of the DeLong test that compared either the CBA or RFA models to the conventional predictive models were all below 0.05 (Supplementary Table 9).

Using the external validation cohort (Supplementary Table 10), the use of these 12 top-performing variables showed a consistent predictive accuracy, with an AUC of 0.79 for the CBA model and 0.73 for the RFA model.

To further validate the effectiveness of the model, we validated it in three subgroups of ACLF patients: patients with alcohol-associated liver disease, patients treated with antibiotics, and patients with bleeding. The results showed that our model performed well in different subgroups, with certain stability and wide applicability (Supplementary Table 11).

Prediction website

To promote the accessibility and practical application of

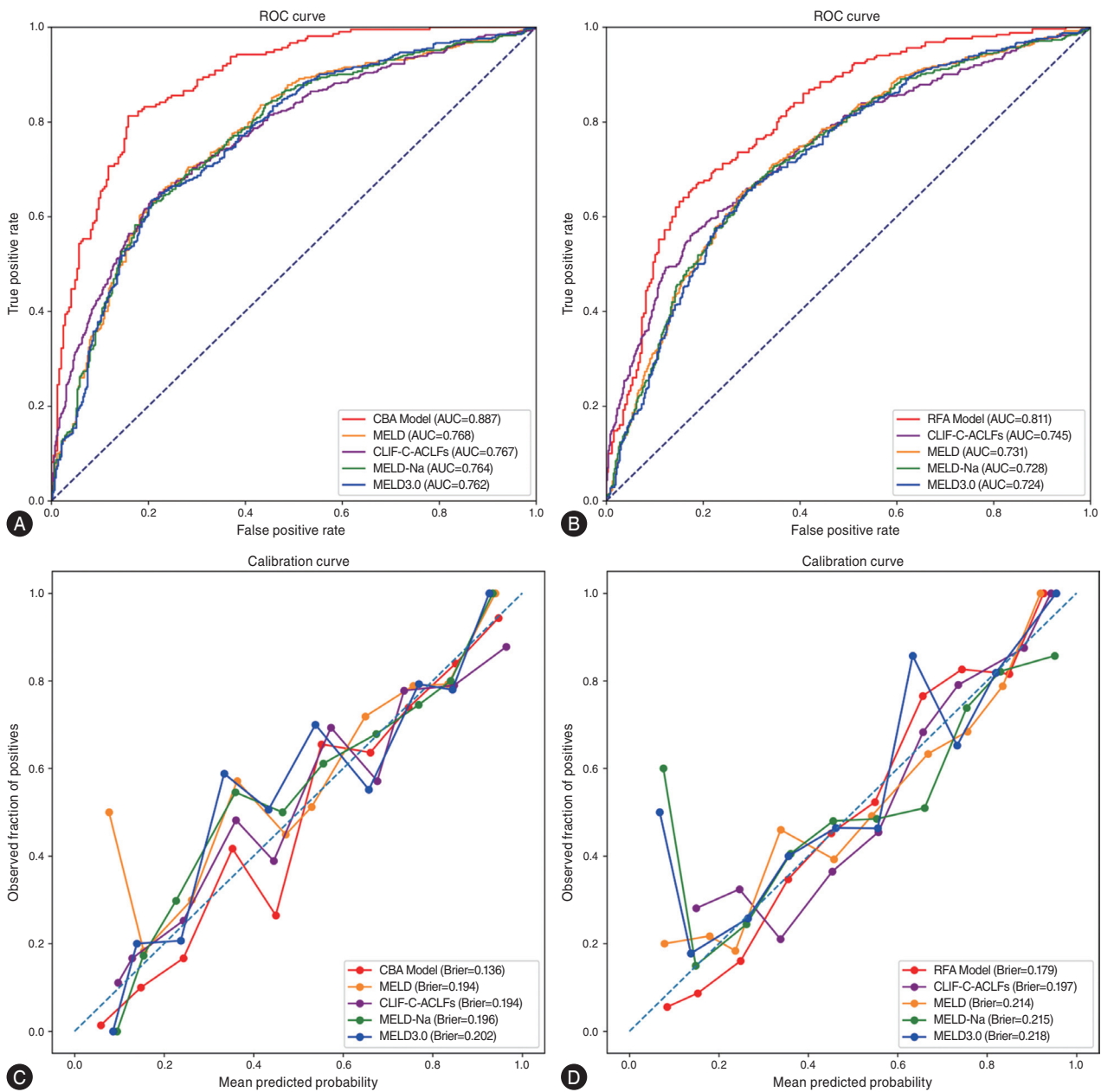
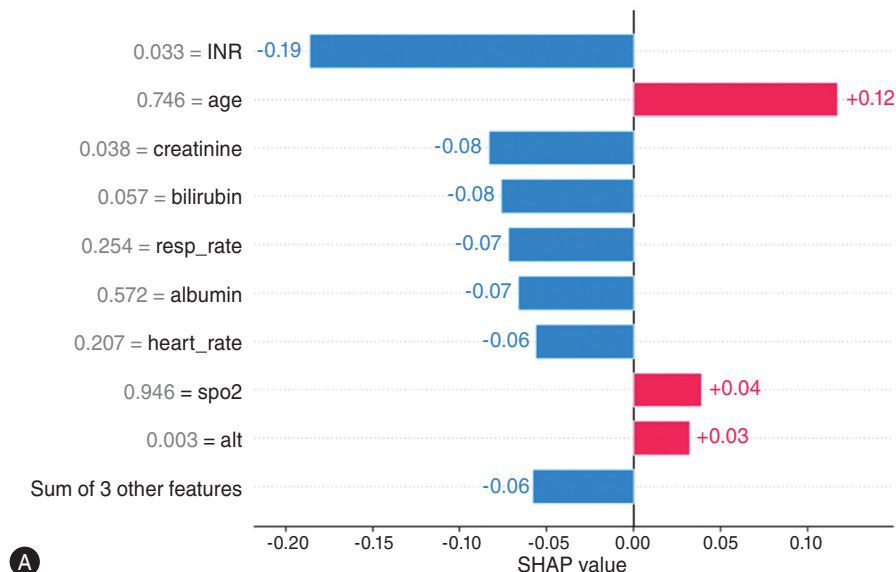


Figure 3. Comparison of predictive performance between machine learning models and conventional predictive scores for acute-on-chronic liver failure (ACLF). (A) Cohort according to the NACSELD ACLF definition. (B) Cohort according to the EASL-CLIF definition grade II and above ACLF. The ROC curve illustrates the predictive performance of various machine learning models and conventional scoring systems for predicting 30-day mortality in patients with ACLF. The AUC was calculated using the trapezoidal rule, which approximates the area under the ROC curve by summing the areas of trapezoids formed under the curve. Calibration curves of CBA model, RFA model, and conventional predictive models (MELD, MELD-Na, MELD 3.0, and CLIF-C ACLF scores) for predicting 30-day mortality. (C) Cohort according to the NACSELD ACLF definition. (D) Cohort according to the EASL-CLIF definition grade II and above ACLF. NACSELD, North American Consortium for the Study of End-Stage Liver Disease; EASL-CLIF, European Association for the Study of the Liver-Chronic Liver Failure; AUC, area under curve; ROC, receiver operating characteristics; CBA, CatBoost ACLF; RFA, Random Forest ACLF.

these models, we developed a dedicated website. The platform is designed to assist physicians and researchers worldwide in utilizing our predictive models for ACLF pa-

tients with two or more OFs to estimate the mortality risk. The website (<https://aclf-death-prediction-nacseld-and-easl-clif.streamlit.app/>) allowed both individual patient data

The probability of death of the patient is 0.09541231395079938.



The probability of death of the patient is 0.8435058922297558.

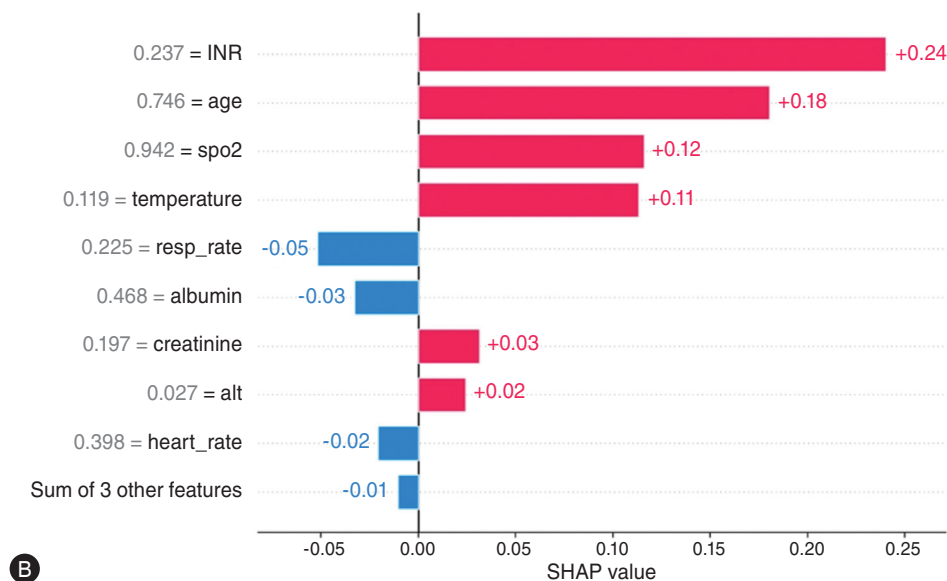


Figure 4. Examples of website output including survival rate at day 30 and the direction and magnitude of contributions towards survival. (A) A case of a patient with acute-on-chronic liver failure who survived within 30 days. Patient information: INR (1.28), age (74 year), bilirubin (2.95 mg/dL), respiratory rate (16.67 bpm), albumin (3.50 g/dL), sodium (136.67 mEq/L), heart rate (67.71 bpm), systolic blood pressure (99.22 mmHg), SpO₂ (95.00%), alanine transaminase (20.50 IU/L), temperature (37.04°C), creatinine (0.87 mg/dL). (B) A case of a patient with acute-on-chronic liver failure who died within 30 days. Patient information: INR (3.63), age (74 year), bilirubin (3.40 mg/dL), respiratory rate (15.85 bpm), albumin (3.00 g/dL), sodium (139.67mEq/L), heart rate (88.34 bpm), systolic blood pressure (100.34 mmHg), SpO₂ (94.67%), alanine transaminase (94.00 IU/L), temperature (35.59°C), creatinine (3.43 mg/dL). This figure illustrates the SHAP values for two patients, showing the probability of death and the contribution of various features to the model's prediction. The blue and left indicate a positive contribution to survival, and the red and right indicate a negative contribution. INR, international normalized ratio; SpO₂, peripheral oxygen saturation; SHAP, SHapley Additive exPlanations.

input and uploading of a table of patient data. Figure 4 shows case examples of the model's predictions for pa-

tients, illustrating cases where patients survived or succumbed by day 30, respectively. Notably, the model dy-

namically adjusts the ranking and direction of variables (positive or negative predictive factors) based on the specific input data provided for each patient. Figure 4A demonstrates a predicted 30-day mortality probability of 9.5% for the patient, indicating low short-term risk, which aligns with the observed survival outcome. In contrast, Figure 4B shows mortality probability of 84.4% for another patient. The high-risk prediction was primarily driven by elevated INR and advanced age, and the prediction was consistent with the patient's mortality outcome observed at 30 days.

DISCUSSION

In this study, we developed interpretable predictive models to predict 30-day mortality for patients with ACLF with two or more OFs, defined by NACSELD and EASL-CLIF criteria, that outperformed existing scoring systems including MELD, MELD 3.0, MELD-Na, and CLIF-C ACLF scores. We also presented hierarchical insights into the clinical and demographic factors influencing the survival of patients with ACLF, categorizing by severity of ACLF. The establishment of a user-friendly, fully automated online tool for mortality prediction without the need for data processing underscores the practical applicability of our findings, facilitating personalized patient care. The diagnostic performance of these models was validated using an internal validation cohort and an independent cohort with the same diagnosis criteria and eligibility, including both NACSELD and EASL-CLIF definitions.

Utilizing a large ICU cohort with detailed clinical information allowed us to establish a robust predictive model. Previous studies using the United Network for Organ Sharing (UNOS) or National Inpatient Database^{24,25} often lack the granularity of clinical data. Although there have been high quality studies to facilitate policy-making,^{26,27} the data was found to be discordant with manual chart review,²⁸ which could hinder their use in developing predictive models.

The clinical characteristics of our ACLF cohort, defined by both NACSELD and EASL-CLIF criteria, exhibit notable differences in previous studies. A prior study comparing the utility of the NACSELD and CLIF-C ACLF definitions showed that the NACSELD definition was superior in predicting overall 28-day mortality.²⁹ The CLIF-C ACLF criteria define respiratory failure based on parameters such as

SpO₂, PaO₂, and FiO₂, while the NACSELD criteria emphasize the use of mechanical ventilation only. This reliance on intubation status may misclassify patients who are mechanically ventilated for airway protection due to hepatic encephalopathy, rather than true respiratory failure. Recent research, including a study by Schulz et al.,³⁰ has indicated that respiratory failure and the necessity for mechanical ventilation are independent risk factors for increased short-term mortality. Recent consensus statements on OF definitions in cirrhosis have called for refinement of existing criteria. Future studies that adopt more physiologically nuanced definitions are needed.³¹

The NACSELD-ACLF score, developed through multi-variable regression modeling using a multi-centered, prospective cohort of ACLF patients,⁷ was not externally validated in the original study. In a subsequent study involving 276 consecutive cirrhotic patients with ACLF, the predictive performance of the NACSELD-ACLF score (AUC=0.67, 95% confidence interval [CI]: 0.61–0.74) was found to be significantly lower than that of the CLIF-C ACLF score (AUC=0.79, 95% CI: 0.73–0.85) and the CLIF-C-lactate score (AUC=0.80, 95% CI: 0.75–0.86).³² Notably, the AUC of the CLIF-C ACLF score in our study was 0.77, aligning with the findings of this subsequent validating article. Additionally, our study, based on baseline data collected at the time of ACLF diagnosis, aligns with the methodology of the VOCAL-Penn model (derived from 74,790 patients meeting the EASL-CLIF criteria to predict 28- and 90-day mortality, achieving AUC values ranging from 0.79 to 0.82).³³ Notably, our model achieved equivalent predictive performance using a significantly smaller cohort of 1,692 ACLF patients meeting the EASL-CLIF criteria, while better distinguishing mortality risk levels.

This study has demonstrated that both the CBA and RFA models significantly outperform the CLIF-C ACLF score in predicting mortality in ACLF patients, with the RFA model performing particularly well in the EASL-CLIF cohort. Notably, the performance of the CLIF-C ACLF score varies across prior literature. A systematic review of the predictive performance of the CLIF-C ACLF score reported a wide variation in AUC values.³⁴ This variability could be attributed to differences in the definition of ACLF, inclusion/exclusion criteria of studies (we focused on ICU patients), baseline severity of ACLF among patients (we focused on grade II and above), and variations in patient management across

different countries. To provide an example with high comparability, a landmark study by Engelmann et al.,⁴ which included an ICU cohort of 202 patients with EASL-CLIF criteria-defined ACLF between 2005 and 2012, reported an AUC of 0.80 for 28-day mortality using the CLIF-C ACLF score. Another study focusing on 249 patients who met the EASL-CLIF definition of ACLF and were admitted to the ICU demonstrated an AUC of 0.83 in predicting 28-day mortality.³⁵ Our study focused on ICU patients with two or more OFs, which is known for capturing a more severe form of ACLF. Consequently, our study presents a well-validated and calibrated predictive score for these ACLF patients with high short-term mortality.

An unresolved issue in ACLF research has been the hierarchical relationship among OFs. Our study leveraged SHAP score analysis to elucidate the importance of each feature in the predictive model. We provided a hierarchical ranking of OFs, and the top-ranked variables were in line with the CLIF-C ACLF criteria, where age and WBC are the only two non-OF-related variables included.^{6,19} Additionally, our study brings new insights into protein synthesis in ACLF, which is consistent with other established predictive models.³⁶ Markers of protein synthesis, such as INR and albumin, were highly ranked in all our models. INR was also one of the diagnostic criteria according to both the EASL-CLIF and the Asian Pacific Association of Liver (APASL) definitions,^{19,37} and INR was also used as an additional diagnostic indicator for HBV-ACLF.^{38,39} Importantly, the significance of both INR and albumin escalated markedly when comparing ACLF patients with two OFs to those with ≥ 3 OFs, which signifies the progressive impact of impaired protein synthesis on mortality risk as ACLF severity increases. This highlights the critical role of liver function markers in prognostic models.¹⁹ Consistently, in the EASL-CLIF ACLF cohort, the RFA model's feature importance analysis showed that coagulation failure (represented by INR) and liver failure (represented by total bilirubin) were among the top-ranked predictors. This suggests that the inclusion of these OFs, which are not part of the NACSELD criteria, provides an added layer of granularity in assessing mortality risk and advocates for a modification of the NACSELD criteria.

In terms of methodology, we used the K-Nearest Neighbors (KNN) algorithm to fill in missing data. Studies have shown that KNN interpolation can effectively solve the

missing data in medical problems.⁴⁰ We also employed a cross-validation method in selecting the optimal resampling method and parameters. Previous studies have shown that cross-validation can optimize the predictive performance of ML and prevent model overfitting.^{41,42}

While our study offered significant insights, several limitations warrant acknowledgment. Firstly, ICU patients tend to present with more severe conditions at baseline, this ICU setting likely captures the majority of the patients with ACLF defined by the NACSELD criteria with at least two OFs. The 30-day mortality rate of our cohort was also similar to prior literature. Furthermore, given the close monitoring and minimal missing data, data deriving from the ICU often provided a more accurate and timely reflection of patient conditions. Secondly, the absence of data on West Haven grading represents a limitation. We used a Glasgow Coma Scale (GCS) ≤ 13 as a proxy, yielding a similar AUC in the external validation cohort, but future studies incorporating West Haven grading to define brain failure are necessary. Thirdly, the presence of infection can significantly impact the short-term mortality rate of patients with ACLF.^{6,7} The use of antibiotics as a marker of infection in this study could cause limitations, as antibiotics does not necessarily imply microbiologically confirmed infection. However, the use of culture positivity may introduce a more significant misclassification bias due to high false-negative rates influenced by prior empirical treatment, suboptimal sampling, or atypical pathogens. Furthermore, variables such as vital signs and laboratory tests can fluctuate dynamically, especially in critical diseases like ACLF. Despite our careful definition of baseline clinical characteristics, there is a potential risk of misclassification. Additionally, some variables may be influenced by treatment; therefore, our study relied solely on baseline data. Future research should consider the impact of therapeutic interventions on these measurements. Given the limited number of patients undergoing liver transplantation in the database, those who received transplants were excluded from the study. Therefore, the impact of liver transplantation on patient mortality was not considered in the analysis of the study. Lastly, the evolution of certain variables may be more crucial than one-time baseline data in prognosticating patients with ACLF. This is a common challenge across disease models, highlighting the need for future models that allow for real-time prediction and adaptation to the dynamic nature of patient conditions.

In conclusion, we developed and validated explainable ML models, including both CBA and RFA, to predict mortality and demonstrated a hierarchical relationship among OFs. Looking forward, this study lays the groundwork for more personalized and dynamic approaches to ACLF management, ultimately improving patient outcomes.

Authors' contribution

Study conception: J.T., Y.H.Y., F.J. Study design and data analysis: Y.H.Y., M.Z., F.J., J.Z., J.T. Drafting of the manuscript: M.Z., Y.H.Y. Data interpretation and critical review of the manuscript: Y.H.Y., M.Z., M.S.M., J.Z., Y.H., Y.L., J.L., T.Y., Y.W., H.D.T., J.D.Y., V.S., X.S., Z.C., C.Y.W., J.T., F.J.

Acknowledgements

We would like to acknowledge Dr. Vinay Sundaram, whose wisdom and invaluable insights were instrumental in initiating this study just weeks before his passing. His passion and thoughtful contributions continue to resonate within our team, and we honor his memory as we carry forward the work inspired by his guidance.

The study was supported by National Natural Science Foundation of China (82170626, 12471470, 12071366, 82473291), Shaanxi Province "Three Qin Scholars" Innovation Team Project (2023001) and the Fundamental Research Funds for the Central Universities (xtr062023003) and the Health and Wellness Scientific Research and Innovation Project of Shaanxi Province (2025TD-09).

During the preparation of this work the authors used GPT-4 in order to assist with editing and grammar refinement. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflicts of Interest

Jonel Trebicka: Speaker and/or consulting fees: Versantis, Gore, Boehringer-Ingelheim, Falk, Grifols, Genfit and CSL Behring. Fanpu Ji: Speaker: Gilead Sciences, MSD and Ascletis. Consulting/advisory board: Gilead, MSD. All other authors do not have conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Clinical and Mo-

lecular Hepatology website (<http://www.e-cmh.org>).

REFERENCES

1. Moreau R, Jalan R, Gines P, Pavesi M, Angeli P, Cordoba J, et al; CANONIC Study Investigators of the EASL-CLIF Consortium. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology* 2013;144:1426-1437, 1437.e1-9.
2. Arroyo V, Moreau R, Jalan R. Acute-on-chronic liver failure. *N Engl J Med* 2020;382:2137-2145.
3. Dong V, Karvellas CJ. Acute-on-chronic liver failure: objective admission and support criteria in the intensive care unit. *JHEP Rep* 2019;1:44-52.
4. Engelmann C, Thomsen KL, Zakeri N, Sheikh M, Agarwal B, Jalan R, et al. Validation of CLIF-C ACLF score to define a threshold for futility of intensive care support for patients with acute-on-chronic liver failure. *Crit Care* 2018;22:254.
5. Karvellas CJ, Garcia-Lopez E, Fernandez J, Saliba F, Sy E, Jalan R, et al; Chronic Liver Failure Consortium and European Foundation for the Study of Chronic Liver Failure. Dynamic prognostication in critically ill cirrhotic patients with multiorgan failure in ICUs in Europe and North America: a multicenter analysis. *Crit Care Med* 2018;46:1783-1791.
6. Jalan R, Saliba F, Pavesi M, Amoros A, Moreau R, Ginès P, et al; CANONIC study investigators of the EASL-CLIF Consortium. Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. *J Hepatol* 2014;61:1038-1047.
7. O'Leary JG, Reddy KR, Garcia-Tsao G, Biggins SW, Wong F, Fallon MB, et al. NACSELD acute-on-chronic liver failure (NACSELD-ACLF) score predicts 30-day survival in hospitalized patients with cirrhosis. *Hepatology* 2018;67:2367-2374.
8. Zheng MH, Shi KQ, Lin XF, Xiao DD, Chen LL, Liu WY, et al. A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network. *J Viral Hepat* 2013;20:248-255.
9. Shi KQ, Zhou YY, Yan HD, Li H, Wu FL, Xie YY, et al. Classification and regression tree analysis of acute-on-chronic hepatitis B liver failure: seeing the forest for the trees. *J Viral Hepat* 2017;24:132-140.
10. Hou Y, Zhang Q, Gao F, Mao D, Li J, Gong Z, et al. Artificial neural network-based models used for predicting 28- and 90-day mortality of patients with hepatitis B-associated acute-on-

- chronic liver failure. *BMC Gastroenterol* 2020;20:75.
11. Gary PJ, Lal A, Simonetto DA, Gajic O, Gallo de Moraes A. Acute on chronic liver failure: prognostic models and artificial intelligence applications. *Hepatol Commun* 2023;7:e0095.
 12. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep* 2019;49:15-21.
 13. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 2.2). *PhysioNet* web site, <<https://doi.org/10.13026/6mm1-ek67>>. Accessed 1 Sep 2023.
 14. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 2023;10:1. Erratum in: *Sci Data* 2023;10:31. Erratum in: *Sci Data* 2023;10:219.
 15. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101:E215-E220.
 16. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth* 2019;13(Suppl 1):S31-S34.
 17. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. Erratum in: *BMJ* 2024;385:q902.
 18. Bajaj JS, O'Leary JG, Reddy KR, Wong F, Biggins SW, Patton H, et al; North American Consortium For The Study Of End-Stage Liver Disease (NACSELD). Survival in infection-related acute-on-chronic liver failure is defined by extrahepatic organ failures. *Hepatology* 2014;60:250-256.
 19. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on acute-on-chronic liver failure. *J Hepatol* 2023;79:461-491. Erratum in: *J Hepatol* 2024;81:370.
 20. Gambino C, Piano S, Angeli P. Acute-on-chronic liver failure in cirrhosis. *J Clin Med* 2021;10:4406.
 21. Registry of Vital Records and Statistics (RVRS). Vital records administration and statistics. Massachusetts government web site, <<https://www.mass.gov/orgs/registry-of-vital-records-and-statistics>>. Accessed 20 Jul 2025.
 22. Riley RD. Correction to: Minimum sample size for developing a multivariable prediction model: part II-binary and time-to-event outcomes by Riley RD, Snell KI, Ensor J, et al. *Stat Med* 2019;38:5672. Erratum for: *Stat Med* 2019;38:1276-1296.
 23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, eds. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Red Hook: Curran Associates, Inc., 2017:4765-4774.
 24. Abdallah MA, Kuo YF, Asrani S, Wong RJ, Ahmed A, Kwo P, et al. Validating a novel score based on interaction between ACLF grade and MELD score to predict waitlist mortality. *J Hepatol* 2021;74:1355-1361.
 25. Rosenblatt R, Shen N, Tafesh Z, Cohen-Mekelburg S, Crawford CV, Kumar S, et al. The North American Consortium for the Study of End-Stage Liver Disease-acute-on-chronic liver failure score accurately predicts survival: an external validation using a national cohort. *Liver Transpl* 2020;26:187-195.
 26. Sundaram V, Jalan R, Wu T, Volk ML, Asrani SK, Klein AS, et al. Factors associated with survival of patients with severe acute-on-chronic liver failure before and after liver transplantation. *Gastroenterology* 2019;156:1381-1391.e3.
 27. Sundaram V, Lindenmeyer CC, Shetty K, Rahimi RS, Al-Attar A, Flocco G, et al; Multi-Organ Dysfunction and Evaluation for Liver Transplantation (MODEL) Consortium. Patients with acute-on-chronic liver failure have greater healthcare resource utilization after liver transplantation. *Clin Gastroenterol Hepatol* 2023;21:704-712.e3.
 28. Lee BP, Cullaro G, Vosooghi A, Yao F, Panchal S, Goldberg DS, et al. Discordance in categorization of acute-on-chronic liver failure in the United Network for Organ Sharing database. *J Hepatol* 2022;76:1122-1126.
 29. Cao Z, Liu Y, Cai M, Xu Y, Xiang X, Zhao G, et al. The use of NACSELD and EASL-CLIF classification systems of ACLF in the prediction of prognosis in hospitalized patients with cirrhosis. *Am J Gastroenterol* 2020;115:2026-2035.
 30. Schulz MS, Mengers J, Gu W, Drolz A, Ferstl PG, Amoros A, et al. Pulmonary impairment independently determines mortality in critically ill patients with acute-on-chronic liver failure. *Liver Int* 2023;43:180-193.
 31. Bajaj JS, Shawcross DL, Choudhury A, Karvellas CJ, O'Leary JG, Trebicka J, et al. Defining organ failures in patients with cirrhosis: consensus statements. *Gastroenterology* 2025:S0016-5085(25)00844-3.
 32. Lin SH, Chen WT, Tsai MH, Kuo WL, Wang SF, Liu Y, et al. Comparing three profoundly influential prognostic scores in cirrhotic patients with acute-on-chronic-liver failure admitted to the ICU: prediction of one-month mortality-a retrospective cohort study. *Diagnostics (Basel)* 2023;13:3160.
 33. Mahmud N, Hubbard RA, Kaplan DE, Taddei TH, Goldberg DS. Risk prediction scores for acute on chronic liver failure de-

- velopment and mortality. *Liver Int* 2020;40:1159-1167.
34. Rashed E, Soldera J. CLIF-SOFA and CLIF-C scores for the prognostication of acute-on-chronic liver failure and acute decompensation of cirrhosis: a systematic review. *World J Hepatol* 2022;14:2025-2043.
35. Chen BH, Tseng HJ, Chen WT, Chen PC, Ho YP, Huang CH, et al. Comparing eight prognostic scores in predicting mortality of patients with acute-on-chronic liver failure who were admitted to an ICU: a single-center experience. *J Clin Med* 2020;9:1540.
36. Wang Y, Dong F, Sun S, Wang X, Zheng X, Huang Y, et al. Increased INR values predict accelerating deterioration and high short-term mortality among patients hospitalized with cirrhosis or advanced fibrosis. *Front Med (Lausanne)* 2021;8:762291.
37. Sarin SK, Choudhury A, Sharma MK, Maiwall R, Al Mahtab M, Rahman S, et al; APASL ACLF Research Consortium (AARC) for APASL ACLF working Party. Acute-on-chronic liver failure: consensus recommendations of the Asian Pacific association for the study of the liver (APASL): an update. *Hepatol Int* 2019;13:353-390. Erratum in: *Hepatol Int* 2019;13:826-828.
38. Li J, Liang X, You S, Feng T, Zhou X, Zhu B, et al; Chinese Group on the Study of Severe Hepatitis B (COSSH). Development and validation of a new prognostic score for hepatitis B virus-related acute-on-chronic liver failure. *J Hepatol* 2021;75:1104-1115.
39. Wu T, Li J, Shao L, Xin J, Jiang L, Zhou Q, et al; Chinese Group on the Study of Severe Hepatitis B (COSSH). Development of diagnostic criteria and a prognostic score for hepatitis B virus-related acute-on-chronic liver failure. *Gut* 2018;67:2181-2191.
40. Karamti H, Alharthi R, Anizi AA, Alhebshi RM, Eshmawi AA, Alsubai S, et al. Improving prediction of cervical cancer using KNN imputed SMOTE features and multi-model ensemble learning approach. *Cancers (Basel)* 2023;15:4412.
41. Park JY, Hsu TC, Hu JR, Chen CY, Hsu WT, Lee M, et al. Predicting sepsis mortality in a population-based national database: machine learning approach. *J Med Internet Res* 2022;24:e29982.
42. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. *Algorithms* 2023;16:88.