Relationship between albumin-bilirubin score, Fibrosis-4 index and 28-day mortality in critically ill cirrhotic patients: a retrospective cohort analysis and machine learning-driven prediction models

Yantao Shu[1], Guangdong Wang[2], Bo Hui[1], Zhidong Wang[1], Yongqiang Xiong[1], Yanchen Li[3]*

[1]Department of Geriatric General Surgery, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China
[2]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China
[3]Department of Clinical Laboratory, Xi'an People's Hospital (Xi'an Fourth Hospital), Xincheng District, Xi'an, Shaanxi, China

***Corresponding author:**
Yanchen Li
Department of Clinical Laboratory
Xi'an People's Hospital (Xi'an Fourth Hospital)
Xincheng District
Xi'an, Shaanxi, China
E-mail: yanchenli16@163.com

Submitted: 18 May 2025; **Accepted:** 14 August 2025
Online publication: 15 October 2025

Arch Med Sci
DOI: https://doi.org/10.5114/aoms/209545

Abstract

Introduction: Critically ill patients with cirrhosis face substantially increased mortality. Reliable prognostic tools are essential for risk stratification and to guide clinical decision-making. Although prior studies have explored the predictive value of the albumin-bilirubin score (ALBI) and Fibrosis-4 (FIB-4) index in compensated cirrhosis and liver cancer, their role in predicting mortality among critically ill patients with cirrhosis has remained insufficiently investigated. This study was conducted to evaluate these associations in this high-risk population.

Material and methods: Data for 2,139 critically ill cirrhotic patients were drawn from Version 3.1 of the MIMIC-IV database. Associations were analyzed using restricted cubic splines and Cox proportional hazards regression. Survival analyses were performed using Kaplan-Meier curves. Feature importance of the ALBI score and FIB-4 index was assessed using the Boruta algorithm, and machine learning-driven predictive models were developed.

Results: Elevated ALBI scores were strongly correlated with increased 28-day all-cause mortality risk in patients with cirrhosis (HR = 1.69;95% CI: 1.48–1.93; $p < 0.05$). The FIB-4 index exhibited similar prognostic relevance. Boruta's feature selection indicated that both scores had high Z scores. Among predictive models, the random survival forest (RSF) approach performed best (AUC = 0.832).

Conclusions: The ALBI and FIB-4 scores are strongly associated with 28-day all-cause death rates in critically ill cirrhotic patients. Higher ALBI scores and FIB-4 index values are associated with greater mortality risks. Predictive models based on machine learning show strong performance. These findings suggest that the ALBI and FIB-4 scores may serve as potential predictors of adverse outcomes in critically ill cirrhotic patients.

Key words: cirrhosis, critical illness, ALBI score, FIB-4 index, mortality, machine learning.

Introduction

Characterized by progressive fibrosis and hepatic dysfunction, cirrhosis is recognized as the end-stage manifestation of progressive hepatic deterioration and predisposes individuals to life-threatening decompensation events, hepatocellular carcinoma (HCC), and significantly elevated mortality risk. In critically ill patients with cirrhosis, the convergence of impaired liver function, systemic inflammation, and multi-organ failure is associated with particularly high mortality rates [1, 2]. Reliable prognostic tools are therefore essential for risk stratification and to inform timely clinical decisions. Traditional tools such as the Child-Pugh score and the Model for End-Stage Liver Disease (MELD) have limitations. For example, the Child-Pugh score incorporates subjective parameters (e.g., ascites, hepatic encephalopathy), which reduces its objectivity. Its grading system is coarse (limited to grades A, B, and C), which restricts its capacity to capture incremental changes in liver function. For instance, the bilirubin levels of 60 µmol/l and 600 µmol/l are assigned the same weight. Moreover, the international normalized ratio (INR) used in both systems does not adequately reflect coagulopathy or hepatic dysfunction, and both international normalized ratio (INR) and creatinine are laboratory-dependent and influenced by non-hepatic factors [3, 4]. As novel liver function assessment tools, the albumin-bilirubin score (ALBI) score and Fibrosis-4 (FIB-4) index demonstrate many advantages. The ALBI score, calculated from two laboratory values (albumin and bilirubin), has been validated for assessment of hepatic function and as an important indicator in the prognostic assessment of HCC and chronic liver conditions, with the advantage of accurately reflecting subtle changes in liver dysfunction [5–7]. The FIB-4 index, which incorporates age, platelet count, alanine aminotransferase (ALT), and aspartate aminotransferase (AST), serves as a non-invasive marker for liver fibrosis, helps identify advanced fibrosis and cirrhosis, and has prognostic relevance for decompensation and long-term survival [8–10]. Recent studies have highlighted the synergistic potential of combining ALBI and FIB-4 to refine risk prediction in compensated cirrhosis, particularly for decompensatory events [11, 12]. However, the prognostic utility of these scores in critically ill cirrhotic patients – a population at extreme risk of short-term mortality – has not been thoroughly investigated. This investigation aims to address existing research gaps by evaluating the relationship between ALBI, FIB-4, and 28-day all-cause death rates in critically ill cirrhosis patients while also developing a machine learning-driven predictive model. Using retrospective cohort data, this study aims to evaluate the predictive value of ALBI and FIB-4 in acute settings and explore their integration with clinical and laboratory variables to enhance risk stratification. The findings could inform timely therapeutic decisions and thereby improve outcomes in this vulnerable population.

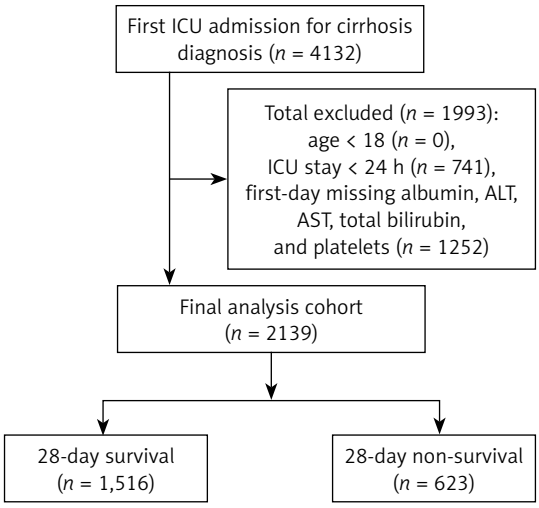

Figure 1. Selection of the study population from the MIMIC-IV database

Material and methods

Source of data

Version 3.1 of the MIMIC-IV database, an open-access ICU clinical resource, includes anonymized medical records from more than 90,000 critically ill patients requiring intensive care. The database is widely used for research [13]. One of the authors, YCL (Certificate number: 14171994), complied with the database's access requirements and was responsible for extracting the relevant data for this study. The patient inclusion process for this study involved identifying individuals receiving care in ICUs for the diagnosis of cirrhosis. A total of 4,132 patients were initially considered. 1,993 patients were excluded based on the predefined exclusion criteria: pediatric population (aged < 18 years; n = 0), critical care duration under 24 h (n = 741), and incomplete essential parameters during the first 24-hour critical care monitoring – specifically albumin, ALT, AST, total bilirubin measurements, and platelet counts (n = 1,252). The final cohort for analysis comprised 2,139 patients. The cohort was stratified according to 28-day clinical outcomes, resulting in two distinct categories: patients who survived beyond this period (n = 1,516) and those who experienced mortality within this timeframe (n = 623) (Figure 1).

Data extraction

Data for this study were extracted from the MIMIC-IV database using PostgreSQL software.

The data exclusively incorporated critical care parameters documented during the initial 24-hour monitoring period. Demographic variables included age, sex, race (white and other), and vital signs (systolic blood pressure, heart rate, oxygen saturation). Physiological scoring systems used were the Logistic Organ Dysfunction Score, Sequential Organ Failure Assessment, and Charlson Comorbidity Index. Comorbid conditions evaluated included acute kidney injury, diabetes, and hypertension. Biomarker analyses quantified white and red blood cell counts (WBC and RBC), platelet count, electrolyte levels (potassium [K] and sodium [Na]), glucose, albumin, total bilirubin, ALT, AST, and INR. Therapeutic interventions included continuous renal replacement therapy, administration of epinephrine, neuromuscular blockade, and mechanical ventilation. Event-related data, including the duration of hospital and intensive care unit (ICU) stays, as well as hospital and ICU mortality rates, were also collected. Supplementary Table SI demonstrates the absence of data. To minimize potential bias, Features (missing rate > 20%) were excluded. Missing values (< 0%) were addressed using multiple imputation with a random forest approach, implemented in the "mice" package (R software).

Description and clinical findings

The study used clinical and laboratory parameters, including the FIB-4 score and ALBI index. The ALBI score and FIB-4 index are derived from the following formulae: FIB-4 = (age × AST)/(platelet count × √ALT); ALBI score = ($\log_{10}$ bilirubin [µmol/l] × 0.66) + (albumin [g/l] × -0.0852). AST and ALT were quantified in international units per liter (IU/l), platelet count was expressed as $\times 10^9$ cells/l, albumin was recorded in grams per liter (g/dl), and total bilirubin values were assessed in micromoles per liter (µmol/l) [14, 15]. The principal endpoint was the four-week all-cause mortality rates, a common endpoint for evaluating short-term prognosis in critically ill patients.

Statistical analysis

Associative analysis

Data distribution patterns were assessed through Shapiro-Wilk normality evaluation. Normally distributed metrics underwent parametric analysis with results expressed as arithmetic mean ± standard deviation. Non-normal distributions were examined via non-parametric tests, presented as median values and interquartile ranges (IQR). For inter-group comparisons: Continuous variables with normal distribution between prognostic cohorts were analyzed using the independent *t*-test, while non-normally distributed variables were assessed using the Mann-Whitney *U* test. Qualitative parameters were examined through the χ^2 test. Prognostic trajectories across FIB-4 score and ALBI index quartile-stratified cohorts were calculated through Kaplan-Meier analysis. Survival curves for each FIB-4 and ALBI quartile were plotted to assess the differences in survival across these groups. Intergroup survival differences were assessed through the log-rank test. Multivariable Cox regression modeling with the proportional hazards assumption determined the independent risk factors of clinical outcomes, which assessed the association between clinical variables including ALBI and FIB-4, and 28-day mortality. Three models were developed: Model 1 (unadjusted); Model 2, adjusted for sex, age, and race; Model 3, further adjusted for Charlson Comorbidity Index, RBC, WBC, INR, and continuous renal replacement therapy (CRRT). The ALBI score was stratified into quartiles, with the first quartile serving as the reference group. Relationships were quantified using hazard ratio (HR) values with 95% confidence intervals across quartile strata. Similarly, the FIB-4 index was analyzed both as a continuous variable (per 5-unit increase) and as quartiles, using the first quartile as the reference. Restricted cubic spline (RCS) regression with three knots was applied to model the non-linear relationships between continuous variables, such as FIB-4 and ALBI, and 28-day mortality risk. Interaction terms in Cox regression were used to evaluate whether the associations of FIB-4 and ALBI with 28-day mortality were modified by covariates including age, sex, race, diabetes, hypertension, and CRRT status.

Development and evaluation of machine learning models

Machine learning models were developed and evaluated as follows. First, variables used to calculate FIB-4 and ALBI were excluded due to significant collinearity with these two indices. The remaining variables were then input into the Boruta algorithm to assess the relative importance of FIB-4, ALBI, and other variables. The important variables identified by the Boruta algorithm (green variables) were used to construct machine learning models. The complete dataset was randomly partitioned, with 70% designated for model development and the remaining 30% reserved for verification purposes. The training cohort was used for model construction and hyperparameter optimization via grid search, while the validation cohort was employed for model performance evaluation. The outcome of interest was 28-day mortality. Multiple machine learning models were trained, including the Cox Proportional Hazards Model (CoxPH), Multilayer Perceptron (MLP), Conditional

Table I. Baseline characteristics

Characteristic	Overall *N* = 2,139	28 d survival *N* = 1,516	28 d non-survival *N* = 623	*P*-value
Demographics				
Age [years]	59.64 (51.97, 66.96)	59.15 (51.13, 65.76)	61.31 (53.52, 70.38)	< 0.001
Sex, *n* (%)				0.335
Female	753.00 (35.20%)	524.00 (34.56%)	229.00 (36.76%)	
Male	1,386.00 (64.80%)	992.00 (65.44%)	394.00 (63.24%)	
Race, *n* (%)				0.019
White	1,482.00 (69.28%)	1,073.00 (70.78%)	409.00 (65.65%)	
Other	657.00 (30.72%)	443.00 (29.22%)	214.00 (34.35%)	
Vital signs				
Heart rate [bpm]	88.20 (77.71, 99.58)	86.81 (76.39, 97.80)	92.23 (80.90, 103.60)	< 0.001
SBP [mm Hg]	110.00 (101.85, 122.04)	112.44 (103.19, 124.94)	106.39 (99.24, 114.44)	< 0.001
SpO_2 (%)	97.08 (95.50, 98.60)	97.30 (95.78, 98.73)	96.56 (95.00, 98.21)	< 0.001
Clinical score				
SOFA	4.00 (1.00, 6.00)	3.00 (1.00, 6.00)	5.00 (1.00, 8.00)	< 0.001
LODS	7.00 (4.00, 9.00)	6.00 (4.00, 8.00)	8.00 (6.00, 10.00)	< 0.001
Charlson Comorbidity Index	6.00 (4.00, 8.00)	5.00 (4.00, 7.00)	6.00 (4.00, 8.00)	< 0.001
Comorbidities, *n* (%)				
AKI	1,870.00 (87.42%)	1,273.00 (83.97%)	597.00 (95.83%)	< 0.001
Diabetes	638.00 (29.83%)	465.00 (30.67%)	173.00 (27.77%)	0.182
Hypertension	1,030.00 (48.15%)	736.00 (48.55%)	294.00 (47.19%)	0.568
Laboratory tests				
RBC [10^9/l]	2.98 (2.57, 3.43)	3.06 (2.65, 3.47)	2.75 (2.41, 3.30)	< 0.001
WBC [10^9/l]	10.00 (6.63, 14.75)	9.31 (6.32, 13.33)	12.20 (7.90, 18.00)	< 0.001
Platelet [10^9/l]	96.00 (64.00, 148.00)	96.58 (67.00, 147.50)	94.50 (59.00, 148.50)	0.079
Sodium [mmol/l]	137.00 (133.00, 140.33)	137.50 (134.00, 140.50)	135.50 (131.00, 139.75)	< 0.001
Potassium [mmol/l]	4.17 (3.80, 4.65)	4.13 (3.80, 4.55)	4.26 (3.80, 4.90)	< 0.001
Glucose [mg/dl]	131.50 (106.00, 173.67)	133.50 (106.88, 182.00)	127.00 (103.00, 157.33)	< 0.001
Anion gap	14.80 (12.00, 18.20)	14.00 (11.50, 17.00)	17.33 (14.00, 21.50)	< 0.001
BUN [mg/dl]	28.40 (16.75, 49.00)	25.00 (16.00, 42.00)	40.00 (22.50, 66.00)	< 0.001
Albumin [mg/dl]	2.96 (2.53, 3.40)	3.00 (2.60, 3.40)	2.90 (2.50, 3.40)	0.051
ALT [U/l]	37.25 (21.00, 96.00)	36.50 (21.00, 106.08)	40.25 (22.00, 84.00)	0.706
AST [U/l]	79.00 (42.33, 207.33)	72.25 (41.00, 204.08)	91.00 (47.50, 212.00)	0.007
Bilirubin total [mg/dl]	3.15 (1.35, 7.90)	2.52 (1.20, 5.73)	5.87 (2.10, 17.25)	< 0.001
INR	1.70 (1.43, 2.17)	1.60 (1.39, 1.95)	2.10 (1.67, 2.60)	< 0.001
ALBI score	–1.35 (–1.81, –0.95)	–1.41 (–1.89, –1.03)	–1.15 (–1.62, –0.73)	< 0.001
FIB-4 index	8.40 (4.47, 16.82)	7.91 (4.08, 15.78)	9.60 (5.40, 19.17)	< 0.001
Treatment, *n* (%)				
Epinephrine	76.00 (3.55%)	24.00 (1.58%)	52.00 (8.35%)	< 0.001
Neuromuscular blockade	84.00 (3.93%)	31.00 (2.04%)	53.00 (8.51%)	< 0.001
MV	1,743.00 (81.49%)	1,210.00 (79.82%)	533.00 (85.55%)	0.002
CRRT	365.00 (17.06%)	190.00 (12.53%)	175.00 (28.09%)	< 0.001
Event				
LOS hospital [day]	11.85 (6.39, 21.31)	13.26 (7.10, 25.81)	8.89 (4.45, 14.65)	< 0.001
Hospital mortality *n* (%)	563.00 (26.32%)	55.00 (3.63%)	508.00 (81.54%)	< 0.001
LOS ICU [day]	3.14 (1.88, 6.25)	2.98 (1.84, 5.78)	3.85 (2.01, 7.21)	< 0.001
ICU mortality *n* (%)	371.00 (17.34%)	18.00 (1.19%)	353.00 (56.66%)	< 0.001

SBP – systolic blood pressure, SpO_2 – oxygen saturation, LODS – Logistic Organ Dysfunction Score, SOFA – Sequential Organ Failure Assessment, AKI – acute kidney injury, RBC – red blood cells, WBC – white blood cells, BUN – blood urea nitrogen, INR – international normalized ratio, AST – aspartate aminotransferase, ALT – alanine aminotransferase, FIB-4 – Fibrosis-4, ALBI – albumin-bilirubin score, CRRT – continuous renal replacement therapy, MV – mechanical ventilation, ICU – intensive care unit, LOS – length of stay.

Inference Trees (CTree), Extreme Gradient Boosting (XGBoost), Elastic Net (Enet), Generalized Boosted Regression Model (GBM), and Random Survival Forest (RSF). The predictive accuracy of the models was assessed using three analytical methodologies: discriminatory capacity analysis via receiver operating characteristic (ROC) curves, clinical utility evaluation through decision curve analysis (DCA), and predictive accuracy plots for model calibration. Additionally, Shapley Additive Explanations (SHAP) were employed to deconstruct the predictive mechanisms of the top-performing model, quantifying feature contributions through additive game-theoretic principles. All statistical analyses were performed using R statistical environment (version 4.4.2), and statistical significance was set at a two-sided *p*-value of less than 0.05.

Results

Baseline characteristics

Table I shows that non-survivors had a higher mean age compared to survivors (61.31 vs. 59.15 years, $p < 0.001$) and a greater proportion of individuals identified as "other" races (34.35% vs. 29.22%, $p = 0.019$). They also exhibited higher heart rates (92.23 vs. 86.81, $p < 0.001$), lower systolic pressure (106.39 vs. 112.44, $p < 0.001$), and lower SpO_2 levels (96.56 vs. 97.30, $p < 0.001$). Clinical scores were worse among non-survivors, including higher SOFA (Sequential Organ Failure Assessment) (5 vs. 3, $p < 0.001$), LODS (Logistic Organ Dysfunction Score) (8 vs. 6, $p < 0.001$), Charlson Comorbidity Index (6 vs. 5, $p < 0.001$). Non-survivors had a higher incidence of AKI (95.83% vs. 83.97%, $p < 0.001$). In terms of laboratory findings, non-survivors had higher WBC counts (12.20 vs. 9.31, $p < 0.001$), lower RBC counts (2.75 vs. 3.06, $p < 0.001$), lower sodium levels (135.50 vs. 137.50, $p < 0.001$), and higher total bilirubin (5.87 vs. 2.52, $p < 0.001$). INR levels were also elevated among non-survivors (2.10 vs. 1.60, $p < 0.001$), along with higher ALBI scores (–1.15 vs. –1.41, $p < 0.001$) and FIB-4 index values (9.60 vs. 7.91, $p < 0.001$. Regarding treatments, non-survivors were more likely to receive epinephrine (8.35% vs. 1.58%, $p < 0.001$) and CRRT (28.09% vs. 12.53%, $p < 0.001$). Finally, non-survivors had shorter hospital stays (8.89 vs. 13.26 days, $p < 0.001$) and longer ICU stays (3.85 vs. 2.98 days, $p < 0.001$), along with higher hospital (81.54% vs. 3.63%, $p < 0.001$) and ICU mortality rates (56.66% vs. 1.19%, $p < 0.001$).

Kaplan-Meier survival curve

Kaplan-Meier analysis demonstrates significantly different survival across both FIB-4 and ALBI quartiles (log-rank $p < 0.001$) (Figure 2). Quartile 1 of both FIB-4 and ALBI showed the highest survival probability, while quartile 4 showed the lowest.

Cox proportional hazards regression analysis

Supplementary Table SII summarized the univariate Cox regression analysis conducted in critically ill patients with cirrhosis, incorporating variables that were statistically significant ($p < 0.05$), as well as clinically relevant factors identified through expert opinion and prior knowledge.

Table II presents the results of the multivariate Cox regression analysis for 28-day mortality in Model 3, which adjusts for sex, age, race, Charlson Comorbidity Index, RBC, WBC, INR, and CRRT. In this model, the ALBI score was a significant predictor of mortality, with a hazard ratio of 1.69 (1.48–1.93, $p < 0.001$) for each unit increase. For ALBI score quartiles, the highest quartile (quartile 4) had the greatest risk of mortality (HR = 2.24 [1.77–2.83], $p < 0.001$), and a significant trend across the quartiles was observed (p for trend < 0.001). The FIB-4 index per 5-unit increase was also a significant indicator of mortality (HR = 1.02 [1.01–1.03], $p < 0.001$). For FIB-4 index quartiles, the highest quartile (quartile 4) had the greatest mortality risk (HR = 1.60 [1.25–2.05], $p < 0.001$), with a significant trend across quartiles (p for trend < 0.001). The variables included in Model 3 were determined through the Boruta algorithm, univariate Cox regression, and expert clinical recommendations.

Non-linear relationship analysis using RCS for 28-day mortality

The RCS analysis reveals a statistically significant overall relationship between ALBI score and 28-day mortality (p-value < 0.001, Figure 3). The p-value for non-linearity was 0.053, suggesting a potential non-linear trend, but it did not reach statistical significance. For FIB-4 index, the overall p-value was 0.001, indicating a significant association with mortality, and the p-value for non-linearity was 0.020, demonstrating a significant non-linear relationship.

Interaction analysis

The interaction analysis for 28-day mortality demonstrates that both ALBI score and FIB-4 index are significant predictors across all subgroups, with overall p-values for both < 0.001 (Figure 4). For ALBI, the p-values for interaction were non-significant in all subgroups (age: $p = 0.989$, sex: $p = 0.464$, race: $p = 0.204$, diabetes: $p = 0.059$, hypertension: $p = 0.944$, CRRT: $p = 0.858$), suggesting no significant effect modification. For FIB-4, the p-values for interaction were also non-significant

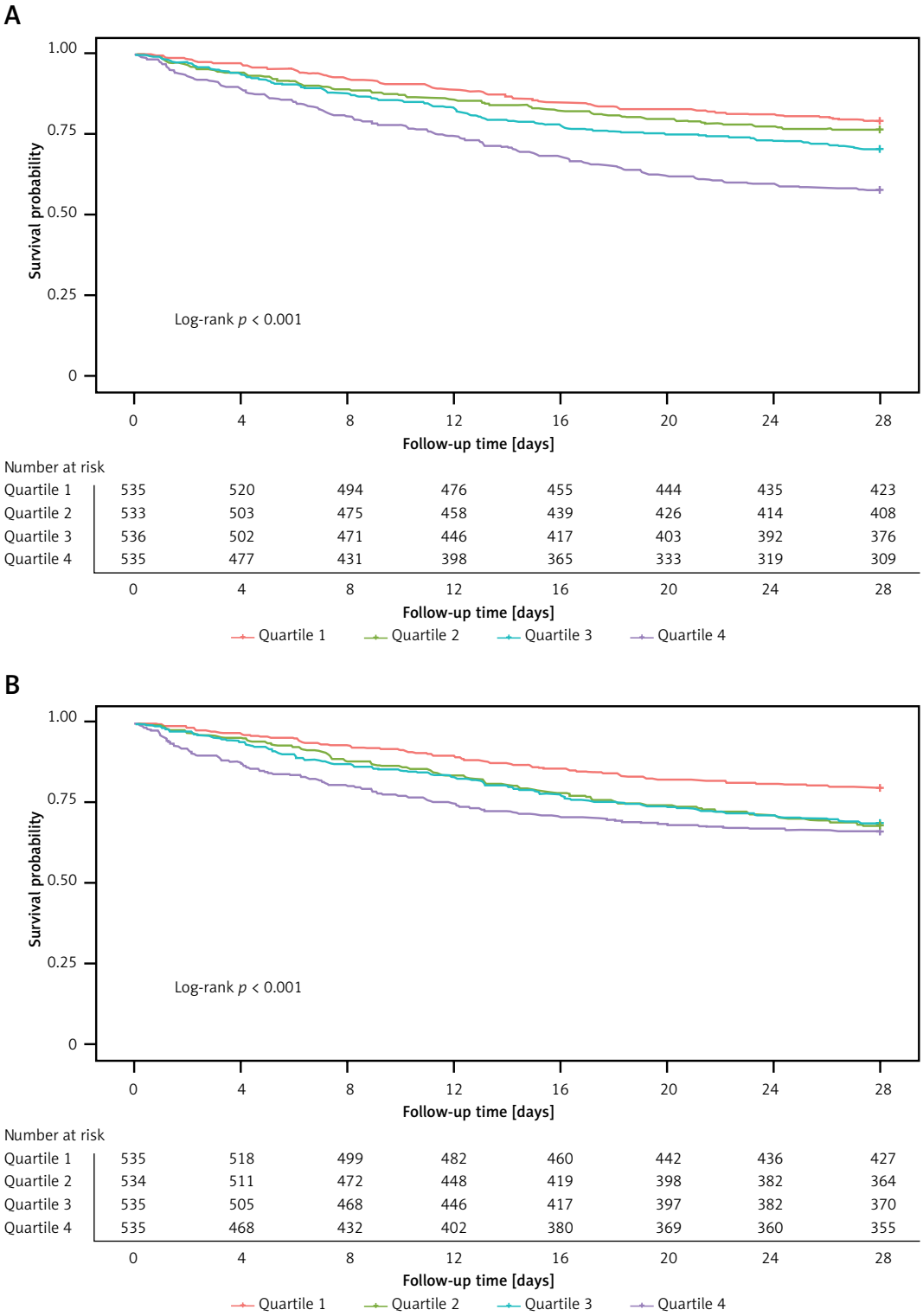

Figure 2. 28-day KM survival curve. KM curves showing survival rates at 28 days for each quartile. **A** – ALBI: quartile 1 (-3.78 – -1.81), quartile 2 (-1.81– -1.35), quartile 3 (-1.35– -0.95), quartile 4 (-0.95–0.44). ALBI: albumin-bilirubin score. **B** – FIB-4: quartile 1 (0.17–4.47), quartile 2 (4.47–8.40), quartile 3 (8.40–16.82), quartile 4 (16.21–623.21)

(sex: $p = 0.425$, race: $p = 0.539$, diabetes: $p = 0.877$, hypertension: $p = 0.335$, CRRT: $p = 0.143$), indicating no significant interaction except for age ($p = 0.046$).

Boruta algorithm

Table II. Multivariate Cox regression

Variables	Model 1		Model 2		Model 3	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
ALBI	1.78 (1.56–2.03)	< 0.001	1.95 (1.71–2.23)	< 0.001	1.69 (1.48–1.93)	< 0.001
ALBI quantile						
1	1.00 (Reference)		1.00 (Reference)		1.00 (Reference)	
2	1.16 (0.90–1.49)	0.263	1.28 (0.99–1.65)	0.062	1.25 (0.97–1.62)	0.087
3	1.52 (1.19–1.93)	< 0.001	1.70 (1.33–2.17)	< 0.001	1.46 (1.14–1.87)	0.002
4	2.37 (1.89–2.97)	< 0.001	2.74 (2.17–3.46)	< 0.001	2.24 (1.77–2.83)	< 0.001
P for trend		< 0.001		< 0.001		< 0.001
FIB-4 per 5 units	1.02 (1.02–1.03)	< 0.001	1.02 (1.01–1.03)	< 0.001	1.02 (1.01–1.03)	< 0.001
FIB-4 Quantile						
1	1.00 (Reference)		1.00 (Reference)		1.00 (Reference)	
2	1.67 (1.32–2.13)	< 0.001	1.61 (1.27–2.05)	< 0.001	1.46 (1.14–1.87)	0.002
3	1.64 (1.28–2.08)	< 0.001	1.58 (1.24–2.02)	< 0.001	1.42 (1.11–1.81)	0.006
4	1.92 (1.51–2.44)	< 0.001	1.79 (1.41–2.28)	< 0.001	1.60 (1.25–2.05)	< 0.001
P for trend		< 0.001		< 0.001		< 0.001

Model 1: Crude. Model 2: Adjust: age, sex, race. Model 3: Adjust: age, sex, race, Charlson Comorbidity Index, RBC, WBC, INR, CRRT. HR – hazard ratio, CI – confidence interval. FIB-4 – Fibrosis-4, ALBI – albumin-bilirubin score, RBC – red blood cells, WBC – white blood cells, BUN – blood urea nitrogen, INR – international normalized ratio, CRRT – continuous renal replacement therapy.

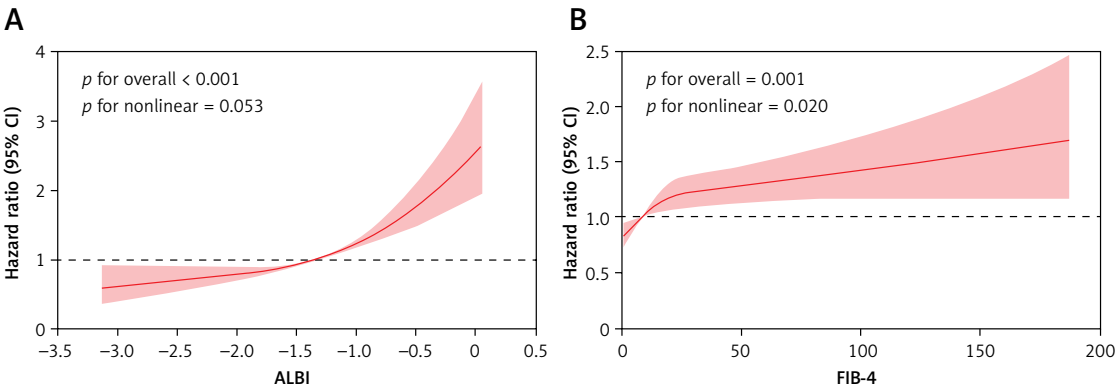

Figure 3. A – RCS analysis of ALBI with 28-day all-cause mortality. **B** – RCS analysis of FIB-4 with 28-day all-cause mortality. Curves represent estimated adjusted hazard ratios, and shaded ribbons represent 95% confidence intervals. The horizontal dashed line represents a hazard ratio of 1.0

HR – hazard ratio, CI – confidence interval.

The Boruta algorithm identifies several important variables for predicting 28-day mortality (Figure 5). The green boxes in the plot represent the confirmed important variables, which include LODS, anion gap, INR, RBC, glucose, WBC, BUN, ALBI score, epinephrine, neuromuscular blockade, SOFA, serum sodium, SpO_2, CRRT, serum potassium, SBP, and FIB-4 index.

Development and evaluation of machine learning model

The machine learning models were developed using the important variables identified by the Boruta algorithm and evaluated on the testing set for 28-day mortality (t = 28). The trained models include CoxPH, CTree, GBM, Enet, MLP, RSF, and XGBoost. The performance of these models, measured by AUC, was as follows: CoxPH (0.822, 0.782–0.862), CTree (0.765, 0.717–0.813), GBM (0.825, 0.784–0.865), Enet (0.824, 0.785–0.863), MLP (0.698, 0.647–0.749), RSF (0.832, 0.794–0.869), and XGBoost (0.830, 0.790–0.870) (Figure 6). The RSF and XGBoost models achieved the highest AUROC values (0.832 and 0.830, respectively), indicating the best performance in predicting 28-day mortality. In terms of decision curve analysis (Supplementary Figure S1), RSF shows the best net benefit across a wide range of threshold probabilities, outperforming the other models. The calibration curve for RSF also demonstrated better calibration than the other models. Overall, RSF was the best performing model, in terms of both discrimination and calibration.

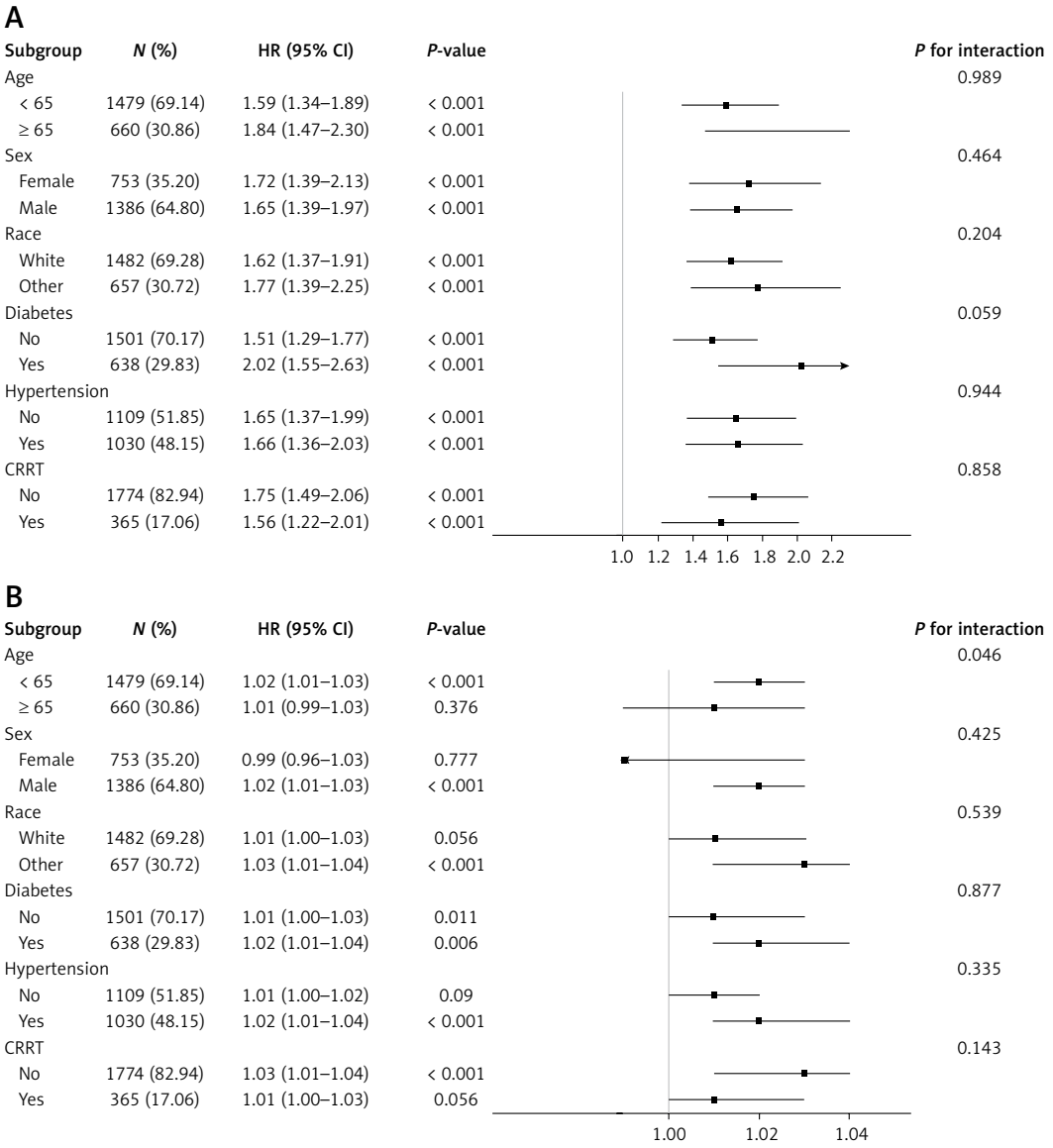

Figure 4. Subgroup forest plot for 28-day all-cause mortality. **A** – Interaction analysis of ALBI score across all subgroups. **B** – Interaction analysis of FIB-4 index across all subgroups. Adjusted for age, sex, race, diabetes, hypertension, and CRRT

HR – hazard ratio, CI – confidence interval.

SHAP explanation for the best performing model

SHAP values were used to interpret the RSF model for 28-day mortality (Figure 7). The ALBI score was the fifth most important variable, where higher ALBI score values correlated with a greater risk of 28-day mortality. The FIB-4 index was an important variable, contributing significantly to the model's predictions. Other important variables included the anion gap and LODS, which played key roles in determining mortality risk.

Discussion

The results of this study indicated a significant association between ALBI, FIB-4, and 28-day mortality in critically ill patients with cirrhosis. Patients in the highest quartiles of ALBI score had the greatest risk of unfavorable clinical outcomes. This was also observed with the FIB-4 index. After adjusting for various confounders, these findings remained consistent. The interaction analysis for 28-day mortality also showed that ALBI and FIB-4 were significant predictors across all subgroups.

The Boruta algorithm, an all-relevant feature selection method, uses shadow features to identify predictive features through iterative permutation-based importance testing [16]. The results demonstrate that both ALBI and FIB-4 prominently feature in the green area, with high Z scores in feature selection, particularly the ALBI score. This

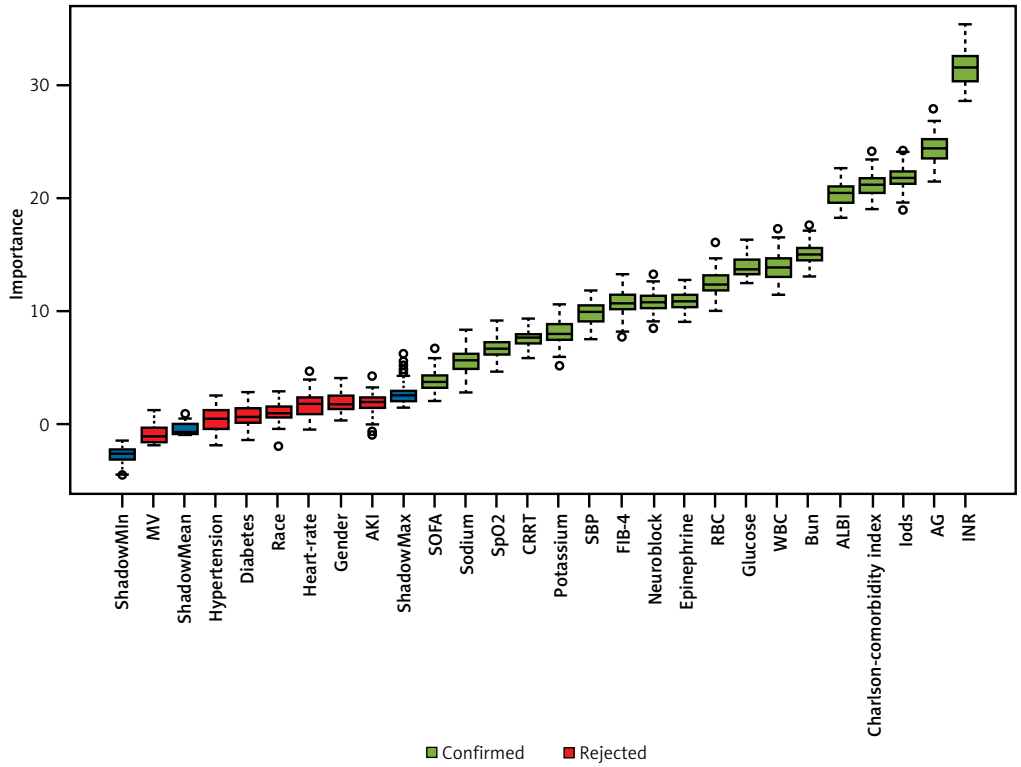

Figure 5. Feature selection based on the Boruta algorithm. The horizontal axis is the name of each variable, and the vertical axis is the *Z* value of each variable. The box plot shows the *Z* value of each variable during model calculation. The green boxes represent important variables, and the red boxes represent unimportant variables

reveals that ALBI and FIB-4 are significantly associated with short-term adverse clinical outcomes. The results also suggest that ALBI and FIB-4 may be significant predictors of 28-day all-cause mortality in cirrhotic patients. However, this does not imply that they are decisive factors; other variables, such as LODS, anion gap, INR, RBC, glucose, WBC, and BUN, are also important features.

With the transformative innovations of artificial intelligence, machine learning algorithms are widely adopted for predicting patient prognosis and treatment outcomes in medical research [17]. We developed and trained seven machine learning models. The results indicate that all these predictive models performed well, especially the RSF and XGBoost models. SHAP values were used to interpret the RSF model. It can be reasonably inferred that ALBI and FIB-4 may be robust predictive indicators.

Previous studies have shown that the conventional Child-Pugh score and the MELD score both have limitations [3, 4, 18, 19]. The Child-Pugh score incorporates subjective indicators, such as ascites and hepatic encephalopathy, and uses broad categories (Classes A/B/C), which limits granularity and sensitivity to small changes in disease severity [7]. The MELD score incorporates INR, creatinine, and other indicators, showing a strong dependence on renal function [20]. The INR in both systems cannot adequately reflect coagulopathy or liver function, while both INR and creatinine measurements are laboratory-dependent and influenced by non-liver factors [3, 4].

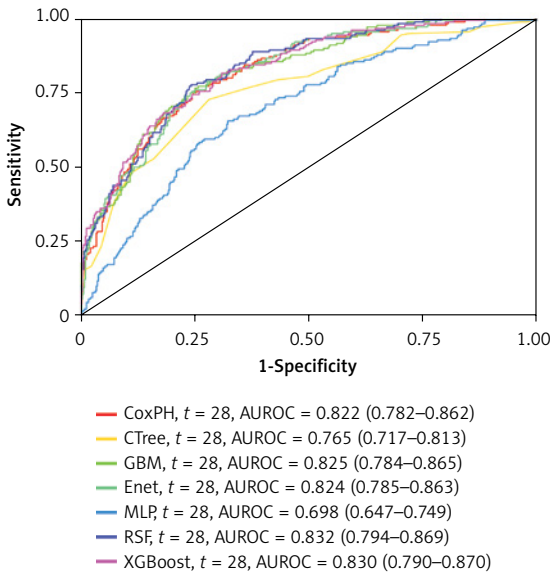

Figure 6. ROC curves of the machine learning algorithms

CoxPH – Cox Proportional Hazards, CTree – Conditional Inference Trees, GBM – Generalized Boosted Regression Model, Enet – Elastic Net, MLP – Multilayer Perceptron, RSF – Random Survival Forest, XGBoost – Extreme Gradient Boosting, ROC – receiver operating characteristic, AUC – area under the curve.

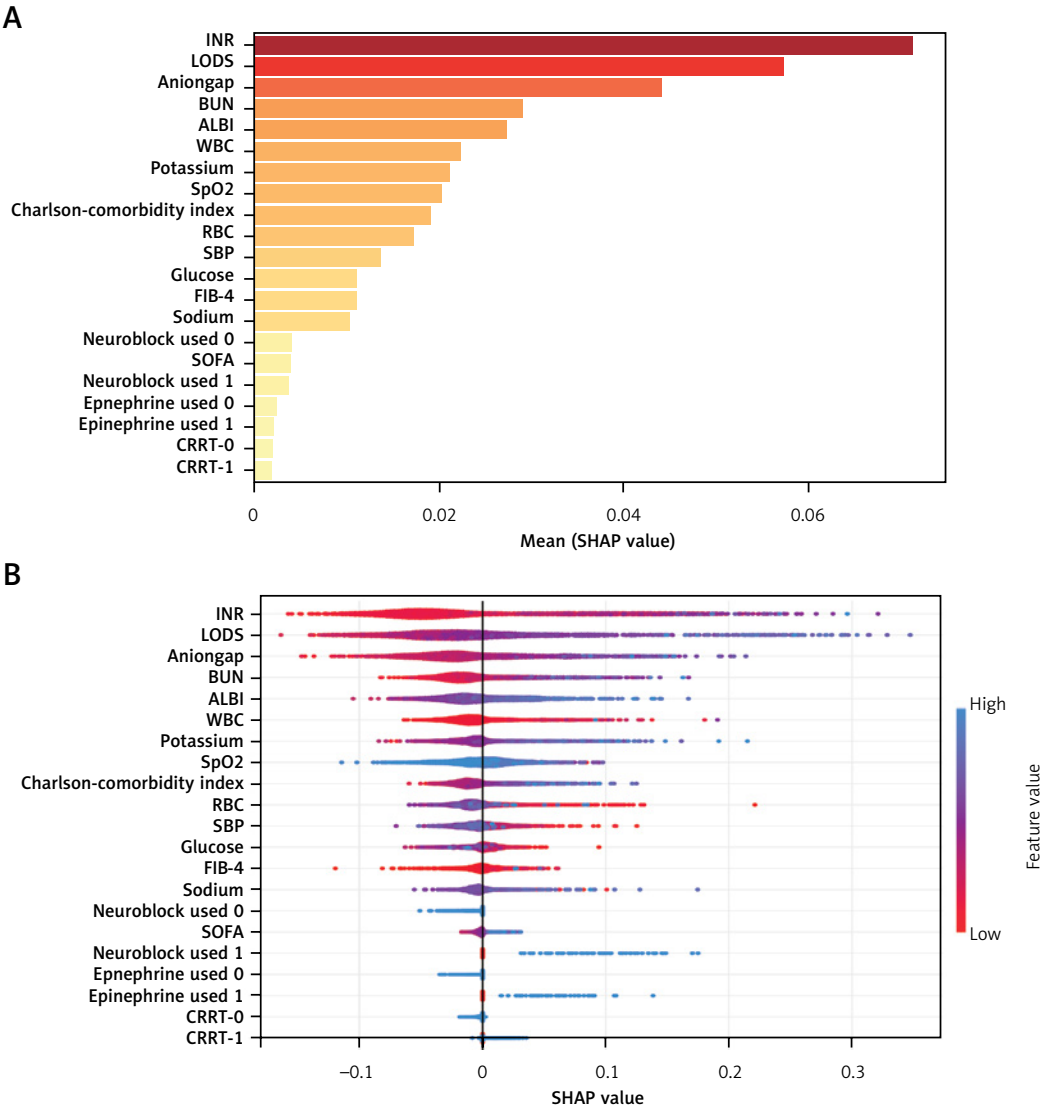

Figure 7. RSF model explanation by the SHAP method. **A** – SHAP summary bar plot. This plot evaluates the contribution of each feature to the model using mean SHAP values, displayed in descending order. **B** – SHAP summary dot plot. The risk of 28-day mortality increases with the SHAP values of the features. Each dot represents a patient's SHAP value for a given feature, with red indicating higher feature values and blue indicating lower values. Dots are stacked vertically to show density

The ALBI score, calculated based on two laboratory values (albumin and bilirubin), is simple to determine and avoids the influence of subjective indicators. The ALBI score's high-resolution quantification enables precise monitoring of subclinical hepatic function alterations in compensated cirrhosis cohorts [7, 21]. It has been widely used to assess the degree of liver function impairment, especially in patients with end-stage liver disease and HCC. It can also indirectly predict mortality risk by reflecting the deterioration of liver function [7]. Prior research demonstrated that the ALBI score was related to the prognosis of HCC and identified different prognostic subgroups across diverse HCC stages [21, 22]. The prognostic value of the ALBI score for patients who received HCC treatments, such as hepatic resection, transarterial therapies, locoregional ablative therapies, and systemic therapies, has also been demonstrated [7, 23–26]. The ALBI score had been further validated for prognostication of outcomes in patients with benign hepatic diseases. Previous research established that the ALBI score could predict longer-term mortality in patients with cirrhosis [7, 27]. Several investigations further supported its clinical equivalence to the MELD in predicting short-time mortality risk in patients with decompensated cirrhosis [28]. Moreover, many reports have demonstrated a significant association between the ALBI score and non-hepatological diseases, such as heart failure, acute kidney injury, acute pancreatitis, sepsis, and aortic dissection [7, 29, 30].

The FIB-4 index, incorporating age, platelet count, ALT, and AST, can non-invasively assess the progression level of liver fibrosis and indirectly reflect liver reserve function [31, 32]. The FIB-4 index has been widely adopted for evaluating the extent of liver fibrosis and predicting cirrhosis and HCC [33, 34]. A longitudinal study in Germany among 248,224 outpatients showed that the FIB-4 index was associated with the incidence of HCC [35, 36]. A retrospective study demonstrated that the FIB-4 index predicted cirrhosis and HCC among individuals with chronic HCV infection [37]. Prior studies have shown that the FIB-4 index predicts both long-term and short-term mortality in decompensated cirrhotic patients [38]. Similar to the ALBI score, many studies have demonstrated a significant association of the FIB-4 index with survival outcomes in many non-hepatological conditions, including cardiovascular events, ischemic stroke, and sepsis [39, 40]. Miele *et al.* [40] demonstrated the prognostic capacity of the FIB-4 index across four distinct SARS-CoV-2 pandemic waves for mortality prediction. Guan *et al.* [41] established the association of the FIB-4 index with mortality among patients with diabetes.

Since the indicators included in ALBI and FIB-4 do not overlap, combining both is speculated to have higher predictive value. Recent studies have confirmed this hypothesis [42]. Yibo Tian *et al.* [11] found that the ALBI-FIB4 score accurately predicts posthepatectomy liver failure (PHLF) in patients with HCC. Liao *et al.* [12] reported that combined preoperative ALBI and FIB-4 significantly predicted postoperative HCC recurrence following curative hepatectomy. Previous studies have not explored the correlation between ALBI, FIB-4, and the prognosis of patients with cirrhosis in the ICU. This study fills this gap in the literature. The results showed that ALBI and FIB-4 strongly correlated with the 28-day all-cause death risk in patients with cirrhosis in the ICU. Moreover, the machine learning model constructed by combining ALBI, FIB-4, and other important indicators also demonstrates good predictive value.

However, this study has several limitations. Firstly, this was a single-center retrospective cohort study. Secondly, ALBI and FIB-4 can only reflect liver function reserve; neither of them includes a detailed evaluation of physical performance and nutritional status. ICU liver cirrhosis patients often have liver function damage and other multiple organ dysfunctions, such as hepatorenal syndrome, coagulation disorders, gastrointestinal bleeding, and hepatic encephalopathy, all of which can lead to the patient's death. ALBI and FIB-4 represent only part of the clinical picture. Future studies that incorporate a broader range of variables, such as key biomarkers, genetic markers, and other organ dysfunctions, may enhance the model's predictive accuracy and clinical applicability. Thirdly, this study did not stratify patients by underlying cause (e.g., viral hepatitis, alcoholic cirrhosis, non-alcoholic fatty liver disease), which may influence the prognostic significance of ALBI and FIB-4. Meanwhile, although the predictive models performed well, the absence of external validation (such as other databases or prospective cohorts) increased the risk of overfitting. Multicenter studies are needed to confirm the model's generalizability and robustness. Fourth, we did not compare the predictive value of ALBI and FIB-4 with the traditional Child-Pugh score and MELD score, which could assess incremental predictive value. Furthermore, with the rapid development of artificial intelligence, we should explore more advanced machine learning algorithms to optimize the model.

In conclusion, this study demonstrated that both ALBI and FIB-4 exhibited significant associations with 28-day all-cause mortality among critically ill cirrhotic patients. Our findings revealed a strong positive correlation between elevated ALBI scores and increased risks of adverse clinical outcomes. Similarly, higher FIB-4 index values demonstrated a parallel association with unfavorable prognosis. Future multicenter, prospective design, and external validation cohorts are still needed to validate these results.

Availability of data and material

Open-access datasets were analyzed in this study. The original datasets are retrievable at https://mimic.mit.edu/.

Funding

No external funding.

Ethical approval

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

References

1. López-Sánchez GN, Dóminguez-Pérez M, Uribe M, et al. The fibrogenic process and the unleashing of acute-on-chronic liver failure. Clin Mol Hepatol 2020; 26: 7-15.
2. Xiong J, Zhang M, Guo X, et al. Acute kidney injury in critically ill cirrhotic patients with spontaneous bacterial peritonitis: a comparison of KDIGO and ICA criteria. Arch Med Sci 2019; 16: 569-76.
3. Kumada T, Toyoda H, Tada T, et al. Changes in background liver function in patients with hepatocellular carcinoma over 30 years: comparison of Child-Pugh classification and albumin bilirubin grade. Liver Cancer 2020; 9: 518-28.

4. Dai B, Guissi NEI, Sulyok LF, et al. Advantages of using indocyanine green in liver transplantation: a narrative review. Ann Transl Med 2022; 10: 110.
5. da Fonseca LG, de Melo MAZ, da Silveira THM, et al. Prognostic role of albumin-bilirubin (ALBI) score and Child-Pugh classification in patients with advanced hepatocellular carcinoma under systemic treatment. Ecancermedicalscience 2024; 18: 1748.
6. Taylor GA, Fagenson AM, Kuo LE, et al. Predicting operative outcomes in patients with liver disease: albumin-bilirubin score vs model for end-stage liver disease-sodium score. J Am Coll Surg 2020; 232: 470-80.
7. Toyoda H, Johnson PJ. The ALBI score: from liver function in patients with HCC to a general measure of liver function. JHEP Rep 2022; 4: 100557.
8. Mózes FE, Lee JA, Selvaraj EA, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. Gut 2022; 71: 1006-19.
9. McPherson S, Hardy T, Dufour JF, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol 2017; 112: 740-51.
10. Rasmussen DN, Thiele M, Johansen S, et al. Prognostic performance of 7 biomarkers compared to liver biopsy in early alcohol-related liver disease. J Hepatol 2021; 75: 1017-25.
11. Tian YB, Niu H, Xu F, et al. ALBI score combined with FIB-4 index to predict post-hepatectomy liver failure in patients with hepatocellular carcinoma. Sci Rep 2024; 14: 8034.
12. Liao R, Li DW, Du CY, et al. Combined preoperative ALBI and FIB-4 is associated with recurrence of hepatocellular carcinoma after curative hepatectomy. J Gastrointest Surg 2018; 22: 1679-87.
13. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-V, a freely accessible electronic health record dataset. Sci Data 2023; 10: 1.
14. Petroff D, Berg T, Wiegand J. Transitioning FIB-4 score: from fibrosis screening tool to key biomarker for clinical endpoints. J Hepatol 2024; 81: e228-9.
15. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol 2015; 33: 550.
16. Tsagris M, Tsamardinos I. Feature selection with the R package MXM. F1000Res 2018; 7: 1505.
17. Hsu WH, Ko AT, Weng CS, et al. Explainable machine learning model for predicting skeletal muscle loss during surgery and adjuvant chemotherapy in ovarian cancer. J Cachexia Sarcopenia Muscle 2023; 14: 2044-53.
18. Allaire M, Goumard C, Lim C, et al. New frontiers in liver resection for hepatocellular carcinoma. JHEP Rep 2020; 2: 100134.
19. Legaz I, Bolarin JM, Campillo JA, et al. Pretransplant ascites and encephalopathy and their influence on survival and liver graft rejection in alcoholic cirrhosis disease. Arch Med Sci 2019; 17: 682-93.
20. Zhou XD, Chen QF, Sun DQ, et al. Remodeling the model for end-stage liver disease for predicting mortality risk in critically ill patients with cirrhosis and acute kidney injury. Hepatol Commun 2017; 1: 748-56.
21. Kudo M, Finn RS, Cheng AL, et al. Albumin-bilirubin grade analyses of atezolizumab plus bevacizumab versus sorafenib in patients with unresectable hepatocellular carcinoma: a post hoc analysis of the phase III IMbrave150 study. Liver Cancer 2023; 12: 479-93.
22. Fagenson AM, Gleeson EM, Pitt HA, et al. Albumin-bilirubin score vs model for end-stage liver disease in predicting post-hepatectomy outcomes. J Am Coll Surg 2020; 230: 637-45.
23. Lee YH, Koh YS, Hur YH, et al. Effectiveness of the albumin-bilirubin score as a prognostic factor for early recurrence after curative hepatic resection for hepatocellular carcinoma. Ann Hepatobiliary Pancreat Surg 2018; 22: 335-43.
24. Tai K, Kuramitsu K, Kido M, et al. Impact of albumin-bilirubin score on short- and long-term survival after living-donor liver transplantation: a retrospective study. Transplant Proc 2020; 52: 910-9.
25. Nam JY, Choe AR, Sinn DH, et al. A differential risk assessment and decision model for transarterial chemoembolization in hepatocellular carcinoma based on hepatic function. BMC Cancer 2020; 20: 504.
26. Lee CH, You GR, Jo HG, et al. Albumin-bilirubin grade as a valuable predictor of recurrence and prognosis in patients with hepatocellular carcinoma following radiofrequency ablation. Cancers (Basel) 2024; 16: 4167.
27. Fragaki M, Sifaki-Pistolla D, Orfanoudaki E, et al. Comparative evaluation of ALBI, MELD, and Child-Pugh scores in prognosis of cirrhosis: is ALBI the new alternative? Ann Gastroenterol 2019; 32: 626-32.
28. Oikonomou T, Goulis L, Doumtsis P, et al. ALBI and PALBI grades are associated with the outcome of patients with stable decompensated cirrhosis. Ann Hepatol 2019; 18: 126-36.
29. Wang J, Wang K, Feng G, et al. Association between the albumin-bilirubin (ALBI) score and all-cause mortality risk in intensive care unit patients with heart failure. Glob Heart 2024; 19: 97.
30. Gou E, Yang Q, Chen J, et al. Association between albumin-bilirubin score and in-hospital mortality in patients with sepsis: evidence from two large databases. Heliyon 2024; 10: e34697.
31. Anstee QM, Castera L, Loomba R. Impact of non-invasive biomarkers on hepatology practice: past, present and future. J Hepatol 2022; 76: 1362-78.
32. Wang Y, Yang X, Wang S. Progress and prospects of elastography techniques in the evaluation of fibrosis in chronic liver disease. Arch Med Sci 2024; 20: 1784-92.
33. Kjaergaard M, Lindvig KP, Thorhauge KH, et al. Using the ELF test, FIB-4 and NAFLD fibrosis score to screen the population for liver disease. J Hepatol 2023; 79: 277-86.
34. Parikh ND, Mehta M, Tapper EB. FIB-4 and APRI for cirrhosis detection in a privately insured national cohort. JHEP Rep 2023; 6: 100925.
35. Loosen SH, Kostev K, Demir M, et al. An elevated FIB-4 score is associated with an increased incidence of liver cancer: a longitudinal analysis among 248,224 outpatients in Germany. Eur J Cancer 2022; 168: 41-50.
36. Li X, Xu H, Gao P. Fibrosis index based on 4 factors (FIB-4) predicts liver cirrhosis and hepatocellular carcinoma in chronic hepatitis C virus (HCV) patients. Med Sci Monit 2019; 25: 7243-50.
37. Lee J, Vali Y, Boursier J, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: a systematic review. Liver Int 2021; 41: 261-70.
38. Ma X, Zhu Y, Yeo YH, et al. The impact of an increased fibrosis-4 index and the severity of hepatic steatosis on mortality in individuals living with diabetes. Hepatol Int 2024; 18: 952-63.
39. Albayrak T, Yuksel B. Prognostic value of fibrosis 4 (FIB-4) index in sepsis patients. J Pers Med 2024; 14: 531.

40. Miele L, Dajko M, Savino MC, et al. Fib-4 score is able to predict intra-hospital mortality in 4 different SARS-COV2 waves. Intern Emerg Med 2023; 18: 1415-27.
41. Guan L, Li L, Zou Y, et al. Association between FIB-4, all-cause mortality, cardiovascular mortality, and cardiovascular disease risk among diabetic individuals: NHANES 1999-2008. Front Cardiovasc Med 2023; 10: 1172178.
42. Akabane M, Kawashima J, Altaf A, et al. Dynamic ALBI score and FIB-4 index trends to predict complications after resection of hepatocellular carcinoma: a K-means clustering approach. Eur J Surg Oncol 2025; 51: 109723.