Journal of Hepatocellular Carcinoma

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Predictive Value of Hepatic Steatosis for Postoperative Recurrence in Hepatitis B-Related Hepatocellular Carcinoma: Development of a Machine Learning-Based Prognostic Model

Aoyun Hao, Changlei Li, Ruitao Sun, Bin Tan, Guanming Shao, Kun Li, Na Li, Weiyu Hu, Chao Qu, Jingyu Cao

Department of Hepatobiliary and Pancreatic Surgery, Affiliated Hospital of Qingdao University, Qingdao, Shandong, People's Republic of China

Correspondence: Jingyu Cao; Chao Qu, Department of Hepatobiliary and Pancreatic Surgery, Affiliated Hospital of Qingdao University, Qingdao, Shandong, People's Republic of China, Email cjy7027@163.com; 2011110479@bjmu.edu.cn

Background: With the increasing prevalence of obesity and type 2 diabetes, the number of patients with hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) coexisting with hepatic steatosis is steadily rising. However, the impact of hepatic steatosis on tumor recurrence following radical resection remains unclear.

Methods: We retrospectively analyzed a cohort of 733 HBV-infected patients diagnosed with HCC who underwent curative liver resection. Propensity score matching (PSM) was performed at a 1:2 ratio using 12 covariates to reduce selection bias and explore the association between preoperative hepatic steatosis and recurrence-free survival (RFS). Furthermore, we constructed a postoperative recurrence prediction model based on hepatic steatosis and other clinicopathological factors using 101 combinations of machine learning algorithms. The optimal model was identified through comprehensive evaluation and validation.

Results: After PSM, survival analysis revealed that patients without hepatic steatosis had significantly better RFS compared to those with steatosis. Multivariate Cox regression analysis confirmed that preoperative hepatic steatosis was an independent risk factor for recurrence following radical resection (P = 0.006, HR:1.564, 95% CI:1.137–2.150). A recurrence prediction model was developed using hepatic steatosis and additional clinicopathological features through machine learning. Among the 101 models tested, the Random Survival Forest (RSF) model exhibited the best predictive performance, achieving a C-index of 0.719 in the training cohort. The model demonstrated high predictive accuracy for 1-, 2-, and 3-year recurrence, with AUC of 0.782, 0.856, and 0.898, respectively. Compared to conventional staging systems such as BCLC and CNLC, our model achieved superior performance, and decision curve analysis (DCA) demonstrated favorable clinical utility.

Conclusion: Preoperative hepatic steatosis is an independent predictor of recurrence after radical resection in patients with HBV-related HCC. The RSF-based machine learning model incorporating hepatic steatosis and other clinicopathological factors effectively predicts postoperative recurrence risk and may facilitate personalized clinical decision-making in this patient population.

Keywords: hepatocellular carcinoma, recurrence, hepatic steatosis, machine learning

Introduction

Liver cancer is the sixth most common malignancy worldwide and the third leading cause of cancer-related mortality.[1] Hepatocellular carcinoma(HCC) accounts for approximately 90% of all liver cancer cases.[2] Radical resection remains the most effective therapeutic approach for HCC.[3,4] However, the postoperative recurrence rate remains high, with up to 70% of patients experiencing tumor recurrence following radical resection.[2] Given the considerable heterogeneity of HCC, it is essential to further explore potential risk factors associated with recurrence.

The development of HCC is typically associated with chronic liver diseases, such as hepatitis B virus (HBV) infection and nonalcoholic fatty liver disease (NAFLD). Although HBV infection does not directly promote hepatic steatosis,

Received: 26 May 2025
Accepted: 22 October 2025
Published: 1 November 2025

hepatic steatosis is increasingly observed in patients with chronic hepatitis B (CHB). In China, the prevalence of biopsy-confirmed hepatic steatosis among CHB patients has risen significantly—from 8.2% to 31.8% over the past decade.[5] This increase is partially attributed to the growing prevalence of obesity and type 2 diabetes, which contribute to lipid accumulation in hepatocytes.[6]

The relationship between chronic HBV infection and hepatic steatosis has been examined in several studies. However, a recent meta-analysis suggests that the impact of hepatic steatosis on HBV-related HCC (HBV-HCC) remains inconclusive.[7] In particular, the prognostic implications of hepatic steatosis for postoperative recurrence in HBV-HCC patients have not been clearly elucidated. While our NER (Neutrophil-to-Eosinophil Ratio)-based algorithm addressed hematological prognostic markers,[8] the current work establishes hepatic steatosis as a critical histopathological predictor for HBV-HCC recurrence after surgical resection. Furthermore, to provide a more individualized risk assessment, we developed a machine learning–based predictive model. To enhance the robustness of our findings and minimize the effects of confounding variables, we utilized propensity score matching (PSM) in our analysis.[9]

Patients and Methods

Patients

The cohort construction process is depicted in Figure 1. We retrospectively collected data from 1036 hepatitis B patients with HCC who underwent hepatectomy at the Qingdao University Affiliated Hospital from 2010 to 2023. Patients who test positive for the hepatitis B surface antigen (HBsAg) in pre-operative infectious disease screening are defined as having hepatitis B.[10] These patients were diagnosed with HCC through postoperative pathology, and underwent R0 resection. All patients had a Child-Pugh grade A liver function.

Continue screening patients according to the following exclusion criteria: (1) diagnosis of other concurrent malignancies; (2) receipt of preoperative interventions or other anti-tumor treatments; (3) postoperative liver transplantation; (4) presence of major vascular invasion confirmed by pathology; (5) coinfection with hepatitis C virus (HCV); and (6)

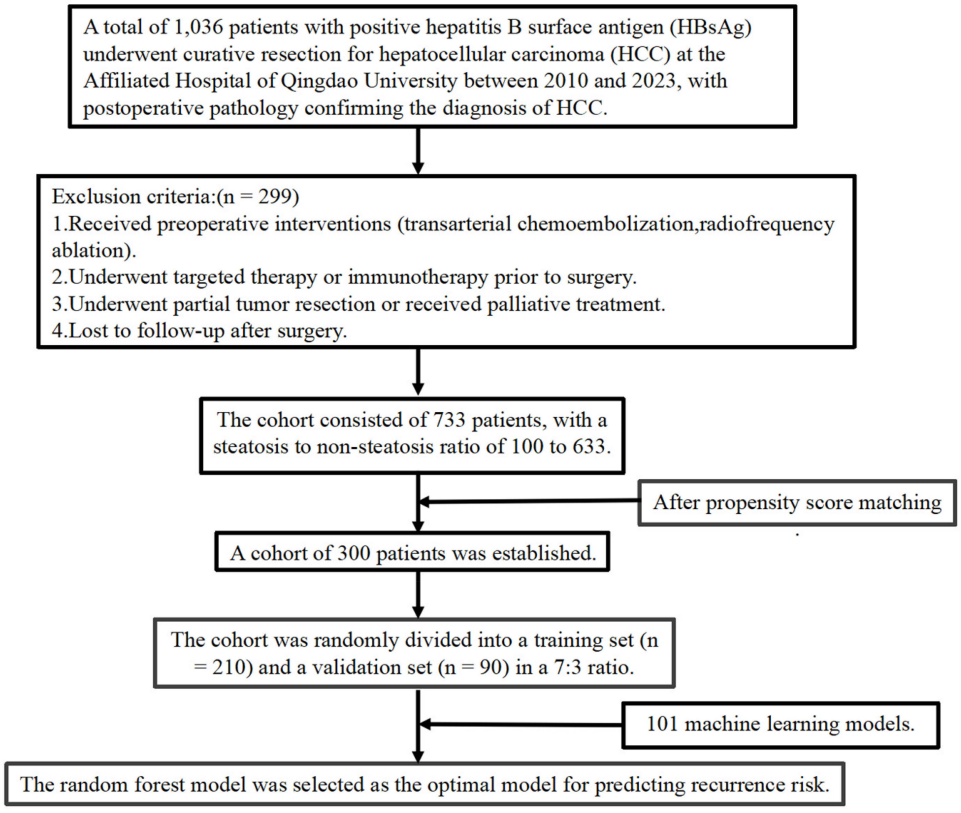

Figure 1 Flowchart of patient selection, matching, and modeling procedures.

incomplete baseline or follow-up data. In addition, patients with a history of alcohol abuse were excluded to avoid confounding effects on hepatic steatosis and liver fibrosis. Alcohol abuse was defined as a weekly alcohol intake exceeding 210 grams for men and 140 grams for women.[11]

In order to reduce the influence of selection bias and confounding factors, propensity score matching (PSM) was employed. The research protocol adhered to the ethical guidelines outlined in the Declaration of Helsinki. Prior to undergoing surgery, written consent was secured from each participant and their direct relatives, ensuring a comprehensive understanding of the process facilitating the utilization of tumor tissue and medical information for scientific investigation objectives.

Clinicopathological Findings

Baseline clinical and laboratory data were collected, including demographic information (age, sex, date of surgery), liver function tests (total protein, albumin, total bilirubin, ALT, AST, γ-GT), complete blood count (hemoglobin, red blood cell count, platelet count, neutrophil count, eosinophil count, basophil count), and tumor markers (alpha-fetoprotein, AFP). HBV serological markers (HBsAg, HBeAg, anti-HBe, anti-HBc) were also recorded. Tumor-related characteristics were assessed, including maximum tumor diameter, tumor number, microvascular invasion, capsule invasion, presence of satellite nodules, and histological differentiation. Hepatic steatosis was defined histologically as the presence of lipid droplets in more than 5% of hepatocytes on hematoxylin-eosin–stained liver tissue sections.[11] Patients with intratumoral steatosis identified on pathological examination but without hepatic steatosis in the surrounding liver tissue were excluded from the cohort.

Follow-Up

Postoperative follow-up was conducted through outpatient visits. Patients were followed monthly during the first 3 months, every 3 months for the first 2 years, and every 3–6 months thereafter. Tumor recurrence was monitored via imaging modalities. If elevated AFP levels were detected, further imaging such as contrast-enhanced CT or PET-CT was performed. In the event of recurrence, patients received additional treatment, including reoperation, microwave ablation, or transcatheter arterial chemoembolization (TACE). Prognostic data were obtained through clinical visits or telephone interviews. The final follow-up date was September 2023. Recurrence-free survival (RFS) was defined as the time from the date of surgery to the date of first documented recurrence (either local or distant metastasis) or liver cancer–related death.

Propensity Score Matching (PSM)

To reduce potential selection bias, PSM was performed using a 1:2 nearest-neighbor matching algorithm with a caliper width of 0.1 (Supplemental Figure 1). We employed the cutoff package in R to convert continuous variables into binary variables prior to propensity score matching. This package determines optimal cut-off values based on the maximization of statistical discrimination, such as log-rank statistics or the area under the curve (AUC). Although some baseline characteristics have internationally recognized clinical thresholds, we chose to derive data-driven cut-offs based on statistical criteria to identify thresholds more suitable for our specific cohort. This approach enhances the sensitivity and specificity of subsequent analyses. Standard thresholds may not adequately capture the characteristics of the study population, whereas the data-adaptive method ensures better balance between groups and strengthens the robustness of the propensity score model. The following baseline variables were included in the model for matching: sex (male vs female), age (≥68 years), tumor diameter (≥3.6 cm), total protein (≥73.4 g/L), total bilirubin (≥14.89 μmol/L), prothrombin time (≥12.10s), international normalized ratio (INR ≥ 0.98), thrombin time (≥1.36s), D-dimer (≥60 ng/mL), antithrombin III (≥66.70 mg/L), fibrinogen (≥2.78 g/L), white blood cell count (≥5.12×10^9/L), bile acid (≥0.10 μmol/L), alkaline phosphatase (≥96 U/L), ALT (≥18.6 U/L), AST (≥25.9 U/L), triglycerides (≥1.20 mmol/L), total cholesterol (≥3.42 mmol/L), HDL-C (≥1.06 mmol/L), LDL-C (≥2.77 mmol/L), AFP (≥209.8 ng/mL), preoperative blood glucose (≥5.63 mmol/L), and body mass index (BMI ≥ 22 kg/m²).

Model Construction, Evaluation, and Explanation

Ten machine learning techniques and a total of 101 algorithmic permutations were employed to develop predictive models, encompassing Random Survival Forest (RSF), Elastic Net (Enet), among others. Lasso, Ridge, Stepwise Cox regression, CoxBoost, Cox partial least squares regression (plsRcox), Supervised Principal Components (SuperPC), Generalized Boosted Regression Model (GBM), and other statistical techniques, and Survival Support Vector Machine (survival-SVM). For every algorithm, the Harrell's concordance index (C-index) was computed on both the training and testing datasets. The predictive model that exhibited the highest average C-index across all datasets was deemed the most optimal. Discrimination modeling was further assessed through the calculation of the area under the receiver operating characteristic curve (AUC). To assess clinical applicability, decision curve analysis (DCA) and calibration plots were employed. Extending our established machine learning framework, this study employed SHapley Additive exPlanations (SHAP) values to quantify feature contributions and determine variable importance rankings in model predictions.[8]

Statistical Analysis

The normality of continuous variables was assessed using the Shapiro–Wilk test. For variables with normal distributions, intergroup comparisons were performed using independent samples *t*-tests, while non-normally distributed variables were compared using the Mann–Whitney *U*-test. Variables with a missing data rate exceeding 20% were excluded from the analysis, whereas those with missing data below 20% were imputed using multiple imputation techniques. All covariates were dichotomized, and categorical variables were compared using the chi-square (χ^2) test. Hazard ratios (HRs) and 95% confidence intervals (CIs) for recurrence risk were estimated using univariate and multivariate Cox proportional hazards regression models. Kaplan– Meier survival curves were constructed to estimate recurrence-free survival, and the Log rank test was used to compare survival distributions between groups. Variables with a p-value < 0.05 were considered statistically significant. All statistical analyses and model construction were conducted using R software (version 4.3.2).

Results

Baseline Clinical Characteristics and Propensity Score Matching

Baseline clinical characteristics of the 733 patients are summarized in Table 1, among whom 100 (13.6%) were diagnosed with hepatic steatosis involving more than 5% of hepatocytes. The cohort included 133 females and 600 males, with a median age of 54 years (range: 30–87 years). Before propensity score matching (PSM), significant differences were observed between the steatosis and non-steatosis groups in terms of microvascular invasion (MVI), age, neutrophil count, eosinophil count, alpha-fetoprotein (AFP), tumor size, preoperative blood glucose, triglycerides, and cholesterol levels. To reduce the impact of confounding variables, we performed PSM based on the following

Table 1 The Baseline Characteristics of the Patients Before PSM

Variable		Non-Steatosis (n=633)	Steatosis (n=100)	*P* Value
Sex	Male	520	80	0.604
	Female	113	20	
Age (year)		58.01±9.73	61.21±8.74	**0.002**
BMI (kg/m^2)		24.50±3.95	24.67±3.40	0.677
Diabetes	Yes	554	83	0.213
	No	79	17	
Cirrhosis	Yes	249	41	0.752
	No	384	59	
Number	Simple	605	89	**0.009**
	Multiple	29	11	

(*Continued*)

Table 1 (Continued).

Variable		Non-Steatosis (n=633)	Steatosis (n=100)	*P* Value
Size (cm)		4.63±3.17	3.58±2.47	**0.002**
MVI	Yes	310	34	**0.005**
	No	323	66	
AFP (ng/mL)		3087.82±24197.52	6683.84±59814.77	0.290
Blood glucose (mmol/L)		5.89±10.86	6.15±2.11	0.811
TP (g/L)		69.73±6.09	69.43±6.54	0.650
ALB (g/L)		42.47±5.84	41.81±4.07	0.276
TBIL (μmol/L)		19.26±8.88	17.95±7.52	0.163
ALT (U/L)		49.12±81.48	44.92±69.02	0.626
AST (U/L)		41.11±56.23	35.55±48.53	0.351
GGT (U/L)		59.58±84.70	67.57±115.04	0.408
WBC (10^9/L)		14.84±169.42	5.46±1.65	0.580
Prothrombin time (s)		11.09±1.36	11.40±1.37	**0.033**
INR		4.78±0.98	18.63±0.10	**<0.001**
D-dimer (ng/mL)		305.79±675.34	349.38±339.57	0.528
AntithrombinIII (mg/L)		82.58±20.51	83.76±17.99	0.586
Fibrinogen (g/L)		2.76±3.23	3.44±7.92	0.131
BA (μmol/L)		0.13±0.18	0.16±0.26	0.174
ALK (U/L)		81.15±38.06	83.79±53.95	0.548
HDL (mmol/L)		1.34±0.35	1.26±0.31	**0.032**
LDL (mmol/L)		2.58±0.76	2.67±0.62	0.252
TG (mmol/L)		0.96±0.65	1.22±0.71	**<0.001**
TC (mmol/L)		4.45±1.07	4.52±0.94	0.510

Note: Bold values indicate statistically significant differences ($p < 0.05$).

covariates: age, sex, tumor size, tumor number, MVI, AFP, preoperative blood glucose, neutrophils, eosinophils, basophils, triglycerides, total cholesterol, HDL-C, LDL-C, body mass index (BMI), cirrhosis, and type 2 diabetes mellitus. This process resulted in a well-balanced matched cohort of 100 patient pairs. The standardized mean differences (SMDs) for all variables were below 0.1, except for BMI, which had a marginally higher SMD of 0.11, indicating excellent comparability between groups. The matched characteristics are shown in Table 2. In the matched cohort, the median recurrence-free survival (RFS) for patients with hepatic steatosis was 18.7 months [95% CI: 21.8–33.5], compared to 24.1 months [95% CI: 29.0–36.2] in patients without steatosis.

Table 2 Absolute Standardized Mean Differences (SMD) After Propensity Score Matching for Baseline Covariates

Variable		Non-Steatosis (n=215)	Steatosis (n=85)	SMD
Sex	Male	166	67	<0.1
	Female	46	18	
Age (years)	≥ 68	40	20	<0.1
	< 68	175	65	
BMI (kg/m^2)	≥ 22	194	75	0.11
	< 22	21	10	
Diabetes	Yes	27	14	<0.1
	No	188	71	

(*Continued*)

https://doi.org/10.2147/JHC.S536629

Table 2 (Continued).

Variable		Non-Steatosis (n=215)	Steatosis (n=85)	SMD
Cirrhosis	Yes	128	50	<0.1
	No	87	35	
Size (cm)	≥ 3.6	105	33	<0.1
	< 3.6	110	52	
Number	Multiple	14	7	<0.1
	Single	201	78	
MVI	Yes	82	33	<0.1
	No	133	52	
AFP (ng/mL)	≥ 209.8	66	20	<0.1
	< 209.8	149	65	
Blood glucose (mmol/L)	≥ 5.63	69	30	<0.1
	< 5.63	146	55	
TP (g/L)	≥ 73.40	59	22	<0.1
	< 73.40	156	63	
ALB (g/L)	≥ 44.72	52	22	<0.1
	< 44.72	163	63	
TBIL (μmol/L)	≥ 14.89	129	51	<0.1
	< 14.89	86	34	
ALT (U/L)	≥ 18.60	84	32	<0.1
	< 18.60	131	53	
AST (U/L)	≥ 25.90	126	48	<0.1
	< 25.90	89	37	
GGT (U/L)	≥ 44.00	90	34	<0.1
	< 44.00	125	51	
WBC (10^9/L)	≥ 5.12	116	45	<0.1
	< 5.12	99	40	
Prothrombin Time (s)	≥ 12.10	123	48	<0.1
	< 12.10	92	37	
INR	≥ 0.98	109	44	<0.1
	< 0.98	106	41	
D-dimer (ng/mL)	≥ 60	4	1	<0.1
	< 60	211	84	
Antithrombin III (%)	≥ 66.70	45	19	<0.1
	< 66.70	170	66	
Fibrinogen (g/L)	≥ 2.78	71	29	<0.1
	< 2.78	144	56	
BA (μmol/L)	≥ 0.10	97	38	<0.1
	< 0.10	118	47	
ALK (U/L)	≥ 96	46	16	<0.1
	< 96	169	69	
HDL (mmol/L)	≥ 1.06	157	62	<0.1
	< 1.06	58	23	
LDL (mmol/L)	≥ 2.77	87	35	<0.1
	< 2.77	128	50	
TG (mmol/L)	≥ 1.20	75	32	<0.1
	< 1.20	140	53	
TC (mmol/L)	≥ 3.42	195	78	<0.1
	< 3.42	20	7	

Clinicopathological Factors Associated with Prognosis in HBV-HCC

As of the final follow-up in September 2023, a total of 437 patients experienced tumor recurrence, including 70 (16.0%) with hepatic steatosis and 367 (84.0%) without. Univariate Cox regression identified several variables significantly associated with recurrence, including hepatic steatosis, microvascular invasion (MVI), tumor size and number, albumin, alpha-fetoprotein (AFP), white blood cell (WBC) count, high-density lipoprotein cholesterol (HDL-C), total bilirubin (TBIL), and fibrinogen (Table 3).

In the multivariate Cox regression model, hepatic steatosis emerged as a significant independent risk factor (HR:1.564, 95% CI: 1.137–2.150, $\boldsymbol{P} < 0.006$), alongside MVI (HR:1.382, 95% CI: 1.011–1.890, $\boldsymbol{P} = 0.043$), tumor number (HR:1.518, 95% CI: 1.063–2.168, $\boldsymbol{P} = 0.022$), tumor size (HR:1.880, 95% CI: 1.389–2.545, $\boldsymbol{P} < 0.001$), AFP (HR:1.457, 95% CI: 1.046–2.031, $\boldsymbol{P} = 0.026$), and WBC count (HR:1.466, 95% CI: 1.075–2.000, $\boldsymbol{P} = 0.016$).

On the other hand, HDL-C (HR:0.638, 95% CI: 0.459–0.887, $\boldsymbol{P} = 0.008$) and albumin (HR:0.687, 95% CI: 0.473–0.998, $\boldsymbol{P} = 0.049$) were identified as protective factors against recurrence.

Table 3 Univariate and Multivariate Cox Regression Analyses of Risk Factors Associated with Recurrence-Free Survival (PFS) After Surgery in the Total Cohort

Variables	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Sex	1.237(0.855–1.790)	0.259		
Age	1.096(0.770–1.559)	0.610		
Bmi	1.134(0.668–1.925)	0.641		
Diabetes	0.734(0.468–1.149)	0.176		
Cirrhosis	1.055(0.779–1.429)	0.728		
Number	1.511(1.077–2.120)	**0.017**	1.518(1.063–2.168)	**0.022**
Size	1.749(1.303–2.349)	**<0.001**	1.880(1.389–2.545)	**<0.001**
Mvi	1.685(1.251–2.270)	**0.001**	1.382(1.011–1.890)	**0.043**
Steatosis	1.432(1.048–1.956)	**0.024**	1.564(1.137–2.150)	**0.006**
AFP	1.591(1.160–2.180)	**0.004**	1.457(1.046–2.031)	**0.026**
Blood glucose	1.048(0.768–1.431)	0.766		
TP	0.854(0.610–1.195)	0.357		
ALB	0.692(0.482–0.995)	**0.047**	0.687(0.473–0.998)	**0.049**
TBIL	0.682(0.507–0.918)	**0.012**	0.696(0.514–0.943)	**0.019**
ALT	0.973(0.720–1.314)	0.858		
AST	1.153(0.855–1.556)	0.350		
GGT	1.385(1.031–1.860)	**0.030**	1.168(0.851–1.603)	**0.337**
WBC	1.401(1.037–1.892)	**0.028**	1.466(1.075–2.000)	**0.016**
Prothrombin time	1.329(0.986–1.791)	0.061		
INR	1.125(0.838–1.509)	0.433		
D-dimer	1.955(0.485–7.882)	0.346		
Antithrombin III	0.738(0.503–1.081)	0.119		
Fibrinogen	1.488(1.099–2.016)	**0.010**	1.306(0.951–1.795)	0.099
BA	1.213(0.904–1.626)	0.198		
ALK	1.239(0.871–1.761)	0.233		
HDL	0.703(0.511–0.967)	**0.030**	0.638(0.459–0.887)	**0.008**
LDL	1.166(0.869–1.566)	0.306		
TG	1.030(0.759–1.397)	0.851		
TC	0.848(0.521–1.381)	0.507		

Note: Bold values indicate statistically significant differences ($p < 0.05$).

Recurrence-Free Survival (RFS) Analysis

Median follow-up of 58.8 months (95% CI, 53.0 to 68.2 months) calculated by the reverse Kaplan-Meier method. Based on the results of multivariate Cox regression, Kaplan–Meier (KM) survival curves were plotted for hepatic steatosis status under different clinical subgroups. Variables that remained statistically significant in the multivariate model included MVI, tumor size, tumor number, albumin, AFP, WBC count, HDL-C, and total bilirubin. In the overall cohort, patients with hepatic steatosis showed worse Recurrence-free survival (RFS) than those without (Figure 2A). In patients with negative MVI, tumor diameter < 3.6 cm, solitary tumor, albumin > 44.72 g/L, AFP < 209.8 ng/mL, WBC count < 5.12×10^9/L, HDL-C > 1.06 mmol/L, and total bilirubin < 14.89 μmol/L, the presence of hepatic steatosis was consistently associated with shorter RFS (Figure 2B–E).

Machine Learning Model Development and Performance Evaluation

Patients with HBV-HCC were randomly divided into a training set (n = 210) and a validation set (n = 90) in a 3:1 ratio. Complete clinical and pathological information for both sets is provided in the Supplemental Table 1. In the training set, variables identified as independent prognostic factors through multivariate Cox analysis were used for model construction. Seven features—including hepatic steatosis, MVI, HDL-C, LDL-C, triglycerides, albumin, and tumor number—were selected to build the model. A total of 101 combinations across 10 machine learning algorithms were tested (Figure 3A). The Random Survival Forest (RSF) model showed the highest concordance index, and when the number of trees reached 1000, the error rate plateaued (Figure 3B). The final RSF model achieved a C-index of 0.719 in the training set and 0.634 in the validation set. The feature importance ranking from the random forest model revealed that hepatic steatosis was among the most influential predictors contributing to the model's performance (Figure 3C).

In the training set, the area under the curve (AUC) for 1-year, 2-year, and 3-year predictions were 0.782, 0.856, and 0.898, respectively. In the validation set, the AUCs were 0.660, 0.772, and 0.775, respectively (Figure 3D and E). Calibration curves demonstrated good consistency between predicted and observed outcomes (Figure 4A and B). Decision curve analysis (DCA) in the full cohort showed that the developed model offered higher net clinical benefit compared to widely used staging systems, including the Barcelona Clinic Liver Cancer (BCLC) and China Liver Cancer (CNLC) staging systems, for predicting 1-, 3-, and 5-year recurrence risk (Figure 4C–H). Overall, the machine learning model developed in this study outperformed other conventional models in predictive accuracy.

To elucidate the contribution of individual clinical features to the risk stratification model, we applied Shapley Additive Explanations (SHAP) analysis to the Random Survival Forest (RSF) model. As shown in Figure 5A, Tumor Size was identified as the most influential variable, exhibiting the highest mean SHAP value (13.72), followed by microvascular invasion (MVI, 8.37), total bilirubin (TBIL, 7.79), and Hepatic Steatosis (4.71). These variables exerted a considerable impact on the predicted relapse risk. The SHAP summary plot in Figure 5B provides a more granular view of feature influence at the individual patient level. High values of Tumor Size, MVI, and TBIL were predominantly associated with increased relapse risk, as indicated by the rightward shift of yellow dots. Conversely, low levels of albumin and high tumor number showed a relatively weaker effect on the model output. These findings confirm that tumor burden and hepatic function-related indices are key determinants in relapse prediction among HBV-related HCC patients.

Risk Stratification and Prognostic Classification

Following model construction, a recurrence risk score was calculated for each patient. Based on the optimal threshold (54.19), patients in the training set were classified into high- and low-risk groups. A total of 104 patients with scores below 54.19 were defined as low risk, while those with scores above the threshold were classified as high risk. Kaplan–Meier survival analysis was performed to visualize differences in RFS between the two risk groups. The results confirmed that the model exhibited good performance in distinguishing between high- and low-risk patients for recurrence after radical resection of HBV-related HCC (Figure 5C and D).

To further illustrate the clinical and pathological heterogeneity of patients with and without hepatic steatosis, we selected two representative cases from our cohort. These cases emphasize the potential role of hepatic steatosis in promoting early recurrence following radical resection for HBV-related HCC (Figure 6A–D).

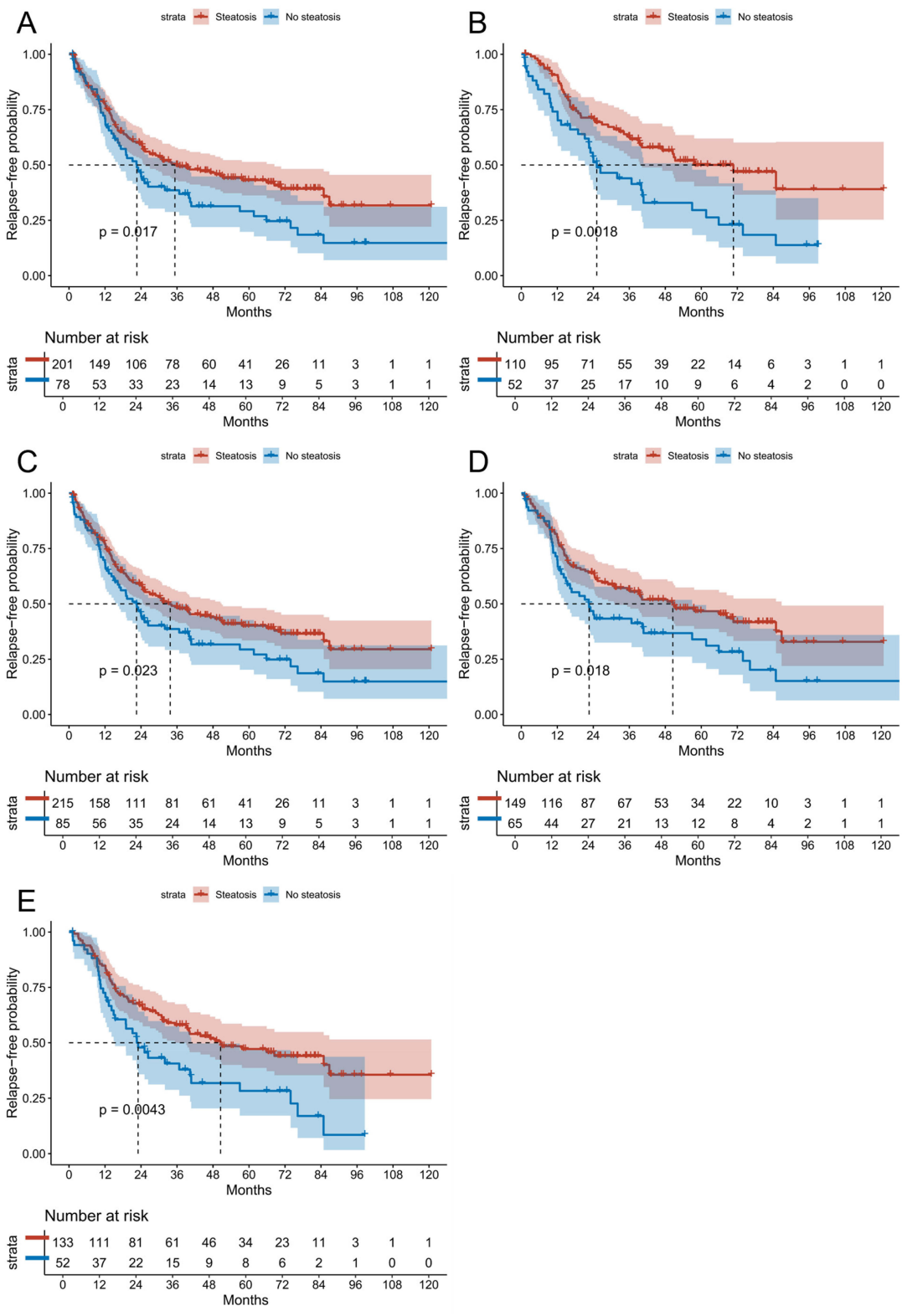

Figure 2 Kaplan–Meier curves for recurrence-free survival (RFS) in patients with and without hepatic steatosis. (**A**) RFS curves for the overall cohort stratified by hepatic steatosis status. (**B**–**E**) Subgroup analyses of RFS based on hepatic steatosis status in patients with (**B**) tumor number; (**C**) tumor diameter; (**D**) alpha-fetoprotein level; (**E**) microvascular invasion.

https://doi.org/10.2147/JHC.S536629

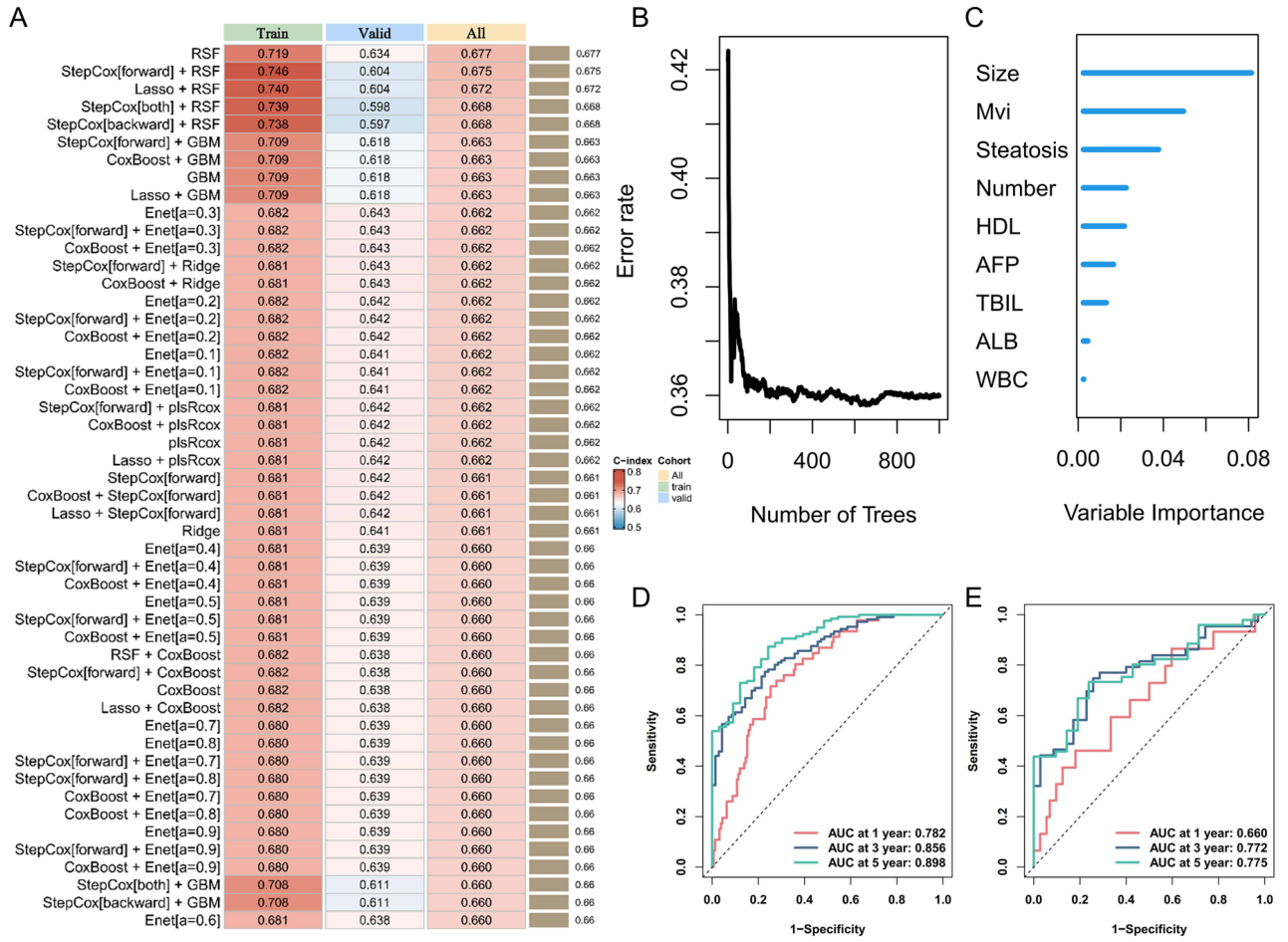

Figure 3 Machine learning model selection and performance evaluation. (**A**) Comparison of 101 machine learning model combinations based on C-index in the training, validation, and overall cohorts. The Random Survival Forest (RSF) model achieved the highest overall performance. (**B**) Error rate curve of the RSF model as a function of the number of trees, showing stabilization after approximately 300 trees. (**C**) Variable importance plot of the RSF model, highlighting the top contributing features. Time-dependent ROC and corresponding AUC for relapse-free survival prediction at 1, 3, and 5 years in the training cohort (**D**) and validation cohort (**E**).

Discussion

Previous studies have suggested that hepatic steatosis may contribute to the development of HCC in patients with hepatitis B. Xu et al investigated the impact of hepatic steatosis on postoperative outcomes in hepatocellular carcinoma (HCC) patients undergoing laparoscopic hepatectomy but found no significant association.[12] Our findings on the significant negative impact of MAFLD on long-term survival appear to contrast with previous reports by Yoon et al and Lin et al, which concluded that fatty liver disease had no significant prognostic effect in patients with HBV-related HCC after resection.[13,14] Several methodological distinctions may explain this discrepancy. The study by Yoon et al,[13] which also employed PSM, included a notably smaller cohort compared to our present study. The larger sample size in our analysis provides greater statistical power to detect a significant difference, potentially uncovering a effect that was underpowered in earlier, smaller studies. Furthermore, the study by Lin et al,[14] focused specifically on the spectrum of hepatic steatosis and steatohepatitis by excluding patients without liver fat accumulation. While this design offers valuable insights into the disease severity, it precludes a direct comparison of outcomes between patients with versus without hepatic steatosis—which was the central question of our investigation. Our study design, which included all comers and then stratified them based on hepatic steatosis diagnosis, may more accurately reflect the real-world prognostic implication of concurrently having hepatic steatosis in the entire HBV-HCC population undergoing surgery. Therefore, the enhanced power of our larger cohort and the differences in patient selection criteria likely contribute to the novel identification of hepatic steatosis as an independent risk factor for worse survival in this setting. In our study, we

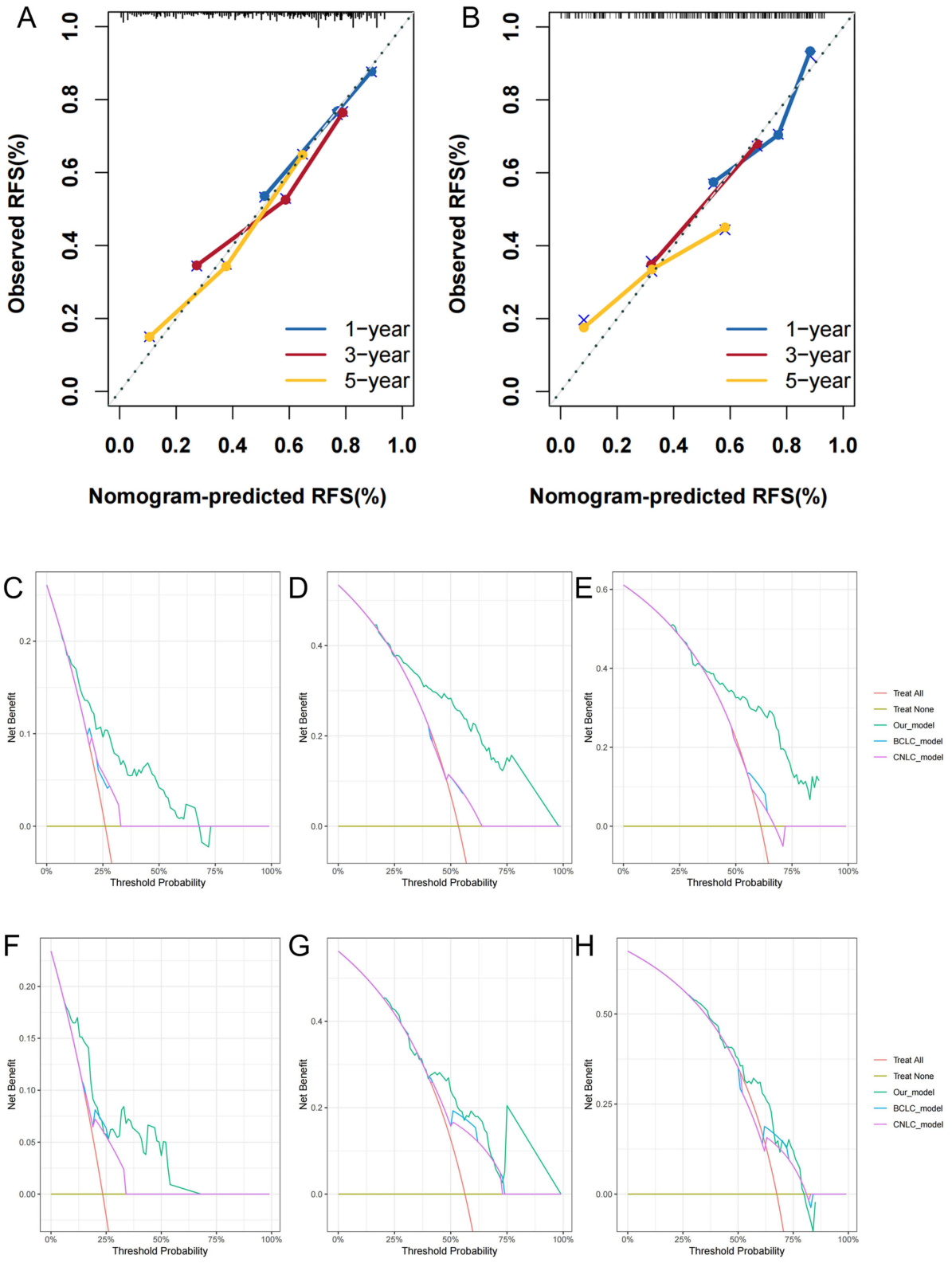

Figure 4 Calibration curves and decision curve analysis (DCA) of the predictive model. (**A** and **B**): Calibration curves at 1, 3, and 5 years in the training cohort (**A**) and validation cohort (**B**). (**C**–**H**): DCA curves at 1, 3, and 5 years in the training cohort (**C**–**E**) and validation cohort (**F**–**H**), respectively.

https://doi.org/10.2147/JHC.S536629

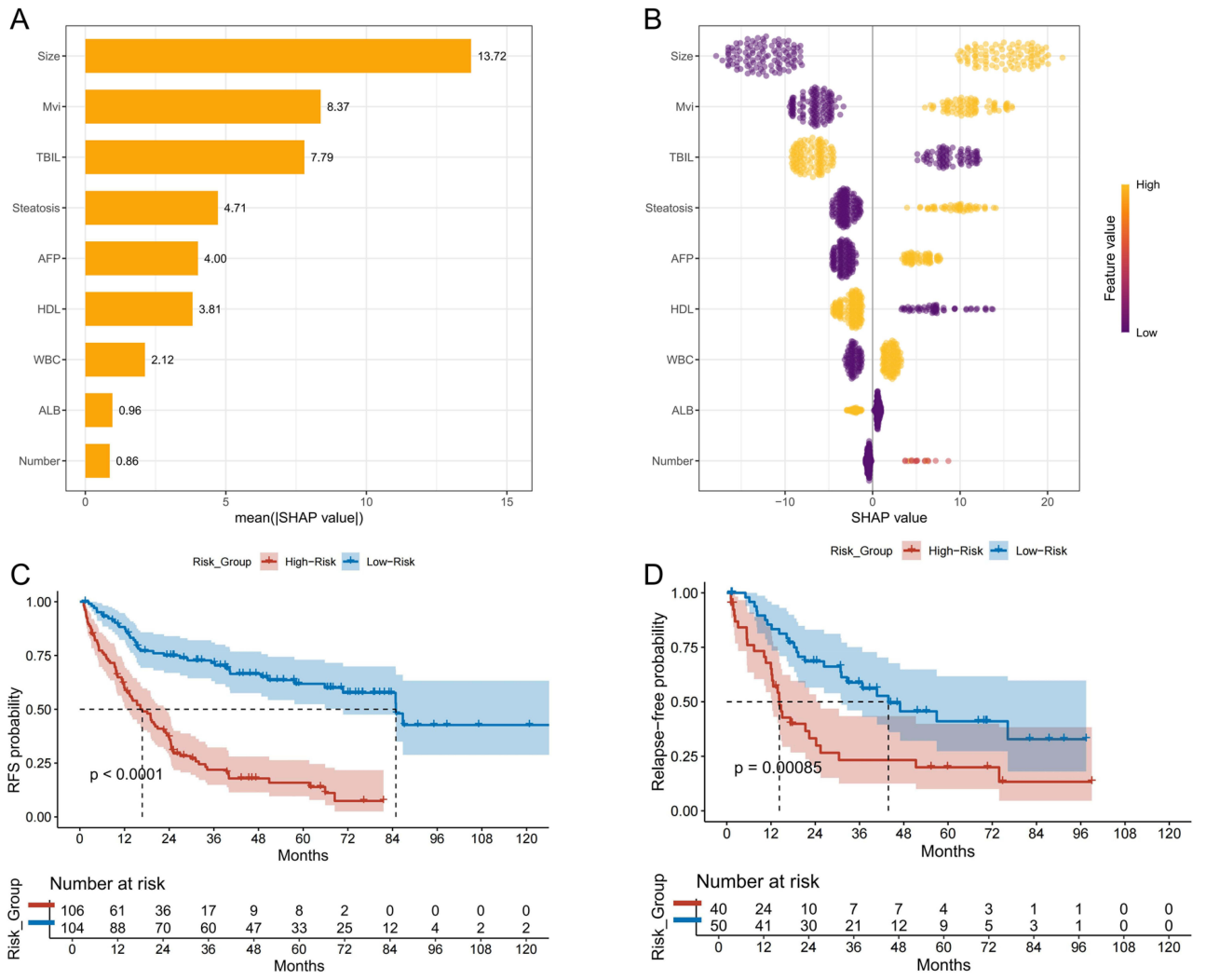

Figure 5 SHAP analysis and risk stratification based on the RSF model. (**A**) Mean SHAP values of the top nine features contributing to the Random Survival Forest (RSF) model, ranked by importance. (**B**) SHAP summary plot showing the distribution of SHAP values for each feature. Each dot represents an individual case, colored by the feature value (purple: low; yellow: high), illustrating the impact of each feature on model output. (**C** and **D**) Kaplan–Meier curves for recurrence-free survival (RFS) stratified by RSF-derived risk groups in the training cohort (**C**) and validation cohort (**D**).

demonstrated that hepatic steatosis significantly promotes tumor recurrence following radical resection in patients with HBV-related HCC. Kaplan–Meier survival analysis revealed that patients without hepatic steatosis had significantly better recurrence-free survival (RFS) compared to those with steatosis. Furthermore, multivariate Cox regression analysis identified hepatic steatosis as an independent risk factor for tumor recurrence. To strengthen the validity of our findings, we applied propensity score matching (PSM) to balance baseline characteristics, thereby simulating a randomized controlled trial and minimizing potential confounding bias. Additionally, we employed a comprehensive ensemble of 101 machine learning models to identify the most effective predictive approach. The Random Survival Forest (RSF) model demonstrated the highest predictive performance. Using SHAP value analysis, hepatic steatosis emerged as the fourth most important variable influencing the model's predictions. Decision curve analysis (DCA) further confirmed that our model offers greater clinical benefit than traditional staging systems, such as the Barcelona Clinic Liver Cancer (BCLC) and China Liver Cancer (CNLC) systems, in predicting recurrence risk.

Although numerous clinical and pathological factors influencing HCC prognosis have been reported, accurately identifying and evaluating these predictors remains challenging. Hepatic steatosis is commonly associated with hyperglycemia, insulin resistance, and oxidative stress.[15] Epidemiologically, it is more prevalent among individuals with

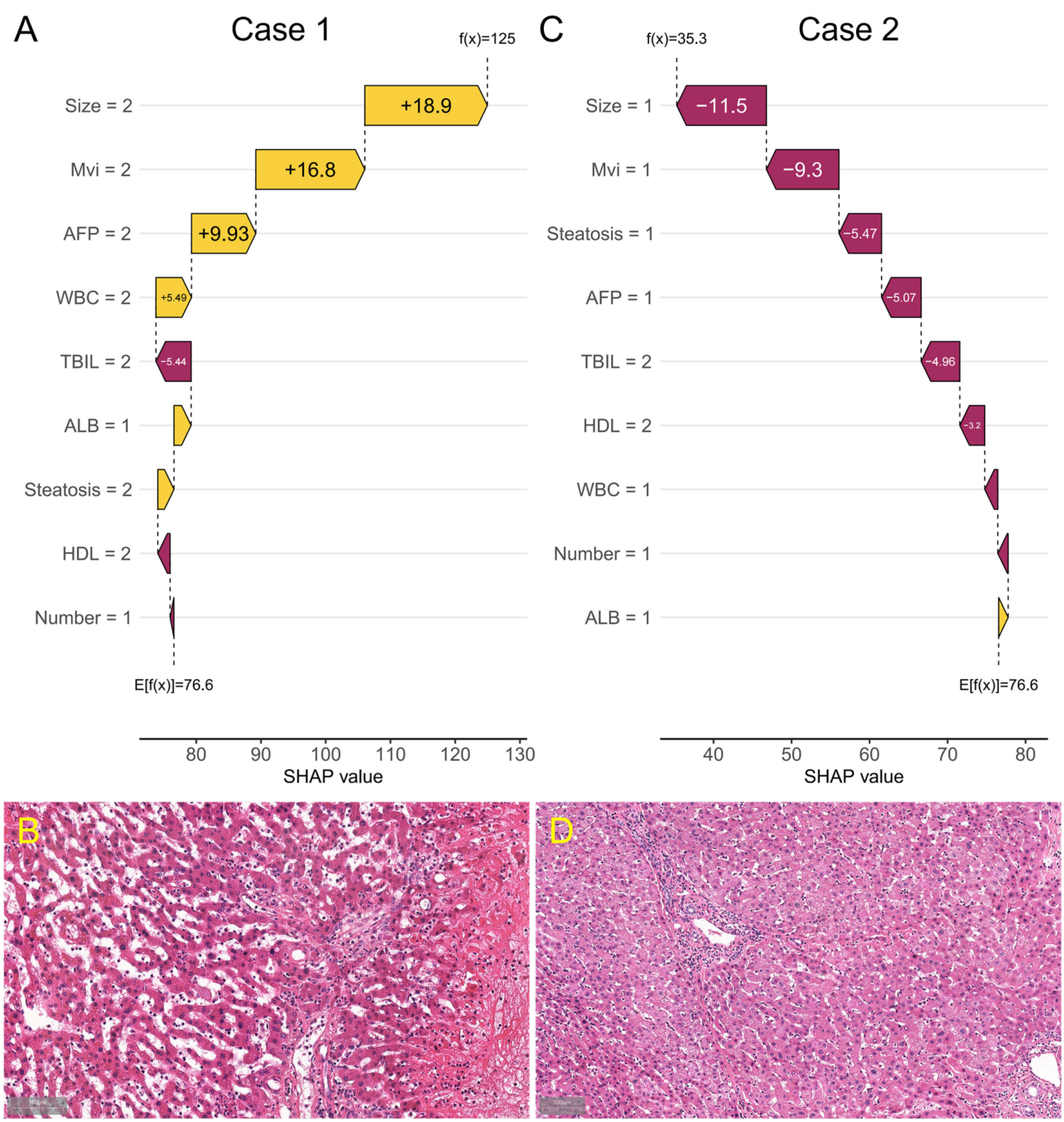

Figure 6 Clinical and Pathological Heterogeneity and Prognosis of Patients with and without Hepatic Steatosis. Case 1: A 53-year-old male patient with HBV-related HCC, presenting with a 4.8 cm tumor exhibiting vascular invasion and with elevated serum AFP (15.20 ng/mL). This patient was classified into the high-risk group (**A**). Histopathological examination revealed a moderately differentiated tumor surrounded by steatotic liver tissue, with prominent macrovesicular steatosis (**B**). Despite undergoing radical resection, the patient developed intrahepatic recurrence 19 months post-surgery. Case 2: A 60-year-old male patient with HBV-related HCC, presenting with a smaller 3 cm tumor, free of microvascular invasion and with elevated serum AFP (5.05 ng/mL). This patient was classified into the low-risk group (**C**). Histopathological examination showed a poorly differentiated tumor without evidence of hepatic steatosis (**D**). Despite receiving similar surgical treatment, the patient remained recurrence-free during an 82-month follow-up period.

obesity and type 2 diabetes.[5] As a central organ in lipid homeostasis, the liver is intricately involved in metabolic dysfunction, which is increasingly recognized as a key factor in hepatocarcinogenesis.[16] Several lipid-synthesizing enzymes are significantly upregulated in HCC. Moreover, hepatocellular stress—such as endoplasmic reticulum (ER) stress and mitochondrial dysfunction—can exacerbate insulin resistance and promote hepatic fat accumulation.[17] In

https://doi.org/10.2147/JHC.S536629

patients with nonalcoholic fatty liver disease (NAFLD), macrophage activation leads to the release of inflammatory cytokines, including tumor necrosis factor (TNF) and interleukin-6 (IL-6), thereby facilitating the progression from simple steatosis to steatohepatitis and ultimately to HCC.[18] Hepatic steatosis may also reduce CD4+ T cell levels, impair immune surveillance, and contribute to immune evasion.[19] Recent work by Hiroki et al highlighted the association between hepatic steatosis and immune exhaustion within the tumor immune microenvironment (TIME) in liver cancer.[20]

It remains uncertain whether metabolic dysregulation is a cause or consequence of HCC. However, several large-scale studies have shown that hepatitis B surface antigen (HBsAg)-positive patients often exhibit dysregulated lipid metabolism.[7,21] This represents a shift in our understanding of the metabolic consequences of hepatitis B virus (HBV) infection, which, similar to hepatitis C virus, has been implicated in the development of hepatic steatosis.[22] HBV may disrupt lipid homeostasis in hepatocytes, resulting in the persistent accumulation of triglycerides, free fatty acids, and cholesterol.[6] One hypothesis suggests that intracellular lipid droplets may alter the localization of HBV antigens and inhibit viral replication.[23] The HBx protein has been shown to upregulate fatty acid-binding proteins, further promoting lipid accumulation. Interestingly, HBsAg-positive individuals often exhibit lower serum cholesterol and triglyceride levels and a reduced risk of hyperlipidemia.[24–26] This paradox complicates clinical identification of hepatic steatosis in HBV-infected individuals based solely on the presence or absence of metabolic syndrome. The interaction between hepatic steatosis and HBV-related hepatitis is thus multifaceted, and their relationship may be synergistic or antagonistic. Further mechanistic studies are needed to fully elucidate this complex interplay.

Previous studies have identified tumor diameter, tumor number, alpha-fetoprotein (AFP) levels, and microvascular invasion (MVI) as significant predictors of recurrence.[27] Patients with smaller tumors, solitary lesions, negative MVI, and low AFP levels generally exhibit better outcomes.[28] To eliminate the influence of these known prognostic factors, we performed subgroup analyses. Notably, even among patients with tumors <3.6 cm, solitary lesions, no MVI, and AFP <209.8 ng/mL, hepatic steatosis remained associated with poorer prognosis. This suggests that hepatic steatosis may influence prognosis independently of tumor burden. Whether lifestyle modifications—such as weight loss or improved insulin sensitivity—can mitigate this effect remains to be investigated. The aMAP score, composed of albumin and bilirubin levels, has been used to predict the risk of HCC development in chronic hepatitis patients.[29] In our study, albumin was identified as a protective factor, while bilirubin was associated with increased recurrence risk. Studies have indicated that high-density lipoprotein (HDL) may exert anti-tumor effects through multiple mechanisms, including its abilities to mitigate inflammation, regulate lipid metabolism, and modulate immune responses. All of these mechanisms are closely associated with cancer development and progression.[30] In insulin-resistant states, HDL-C levels decline, and intrahepatic triglyceride accumulation accelerates HDL-C degradation.[31,32] Given HDL-C's essential role in cholesterol efflux, therapeutic strategies aimed at enhancing HDL-C production and function may represent a promising direction in the management of hepatic steatosis in HBV-infected patients.

Machine learning has become increasingly valuable in clinical research, aiding in disease diagnosis, risk stratification, and treatment decision-making.[33,34] In terms of modeling, machine learning empowers systems to learn from errors, analyze data, identify patterns, and make informed decisions with minimal human intervention.[35] In our study, we tested 101 machine learning models to select the optimal predictor of postoperative recurrence. Importantly, the resulting model provides interpretable insights into the relative importance of risk and protective factors, facilitating potential clinical policy.

Nevertheless, several limitations should be acknowledged. The number of patients with hepatic steatosis in our cohort was relatively small, which limited more precise stratification. Larger sample sizes and multi-center validation are needed to confirm these findings. Moreover, to ensure diagnostic accuracy, we did not rely on imaging or the fatty liver index (FLI) to identify hepatic steatosis.[36] While liver biopsy remains the gold standard, it is invasive and impractical for routine use. Noninvasive imaging techniques such as ultrasound and magnetic resonance imaging-proton density fat fraction (MRI-PDFF) have shown promise but require further refinement. We hope that future advancements in diagnostic technology will enable more accurate identification of hepatic steatosis, thus facilitating precision medicine.[37]

Conclusion

This study shows that preoperative hepatic steatosis is an independent risk factor for postoperative recurrence in HBV-related HCC. The Random Survival Forest model, incorporating hepatic steatosis and clinical-pathological variables, offers a reliable tool for predicting recurrence risk and may assist clinicians in optimizing follow-up and personalizing treatment strategies.

Abbreviations

RSF, Random Survival Forest; Enet, Elastic Net; plsRcox, Partial Least Squares Regression for Cox; SuperPC, Supervised Principal Components; GBM, Gradient Boosting Machine; Survival-SVM, Survival Support Vector Machine; AUC, area under the curve; MVI, microvascular invasion; AFP, alpha-fetoprotein; WBC, white blood cell count; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TBIL, total bilirubin; TP, total protein; ALB, albumin; ALK, alkaline phosphatase; INR, international normalized ratio; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides; TC, total cholesterol.

Data Sharing Statement

The data that support the findings of this study are available from the Affiliated Hospital of Qingdao University, but restrictions apply to the availability of these data, which were used under licence for the current study and so are not publicly available. The data are, however, available from the authors upon reasonable request (E-mail: cjy7027@163.com).

Ethics Approval and Consent to Participate

The protocol of this study was approved by the ethical review board of Qingdao University. (No. QYFY WZLL 30019).

Consent for Publication

Written informed consent for publication was obtained from all participants.

Acknowledgments

Thank all the staff authors for their contributions to this study. The authors extend their sincere gratitude to Bixiao Cao for her invaluable intellectual contribution to the interpretation, particularly in elucidating the reasons for discrepancies between our findings and existing literature.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis, and interpretation, or all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by the project supported by the Taishan Scholar Foundation of Shandong Province, the Shandong Provincial Natural Science Foundation (ZR2024QH265), the Shandong Provincial Medical and Health Science and Technology Development Plan (202425020784), the Qingdao Natural Science Foundation (24-4-4-zrjj-103-jch).

Disclosure

The authors declare no conflict of interest, financial or otherwise.

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–263. doi:10.3322/caac.21834
2. Galle PR, Forner A, Llovet JM, et al. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182–236. doi:10.1016/j.jhep.2018.03.019
3. Nevola R, Ruocco R, Criscuolo L, et al. Predictors of early and late hepatocellular carcinoma recurrence. *World J Gastroenterol*. 2023;29 (8):1243–1260. doi:10.3748/wjg.v29.i8.1243
4. Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol*. 2022;76 (3):681–693. doi:10.1016/j.jhep.2021.11.018
5. Shi YW, Yang RX, Fan JG. Chronic hepatitis B infection with concomitant hepatic steatosis: current evidence and opinion. *World J Gastroenterol*. 2021;27(26):3971–3983. doi:10.3748/wjg.v27.i26.3971
6. Cheng Y, Shao S, Wang Z, et al. From lipotoxicity to pan-lipotoxicity. *Cell Discov*. 2025;11(1):27. doi:10.1038/s41421-025-00787-z

7. Mao X, Cheung KS, Peng C, et al. Steatosis, HBV-related HCC, cirrhosis, and HBsAg seroclearance: a systematic review and meta-analysis. *Hepatology*. 2023;77(5):1735–1745. doi:10.1002/hep.32792
8. Shao G, Ma Y, Qu C, et al. Machine learning model based on the neutrophil-to-eosinophil ratio predicts the recurrence of hepatocellular carcinoma after surgery. *J Hepatocell Carcinoma*. 2024;11:679–691. doi:10.2147/jhc.S455612
9. Wan F. Propensity Score Matching: should we use it in designing observational studies? *BMC Med Res Methodol*. 2025;25(1):25. doi:10.1186/s12874-025-02481-w
10. Wu X, Quan D, Li W, et al. Clinical results of an HBV-specific T-cell receptor-T-cell therapy (SCG101) in patients with HBV-related hepatocellular carcinoma treated in an investigator-initiated, interventional trial. *Gut*. 2025:gutjnl–2025–335456. doi:10.1136/gutjnl-2025-335456
11. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American association for the study of liver diseases. *Hepatology*. 2018;67(1):328–357. doi:10.1002/hep.29367
12. Xu H, Liu Y, Wei Y. Impact of metabolic dysfunction-associated fatty liver disease on the outcomes following laparoscopic hepatectomy for hepatocellular carcinoma. *Surg Endosc*. 2024;38(11):6456–6463. doi:10.1007/s00464-024-11239-2
13. Yoon JS, Lee HY, Chung SW, et al. Prognostic impact of concurrent nonalcoholic fatty liver disease in patients with chronic hepatitis B-related hepatocellular carcinoma. *J Gastroenterol Hepatol*. 2020;35(11):1960–1968. doi:10.1111/jgh.15026
14. Lin YP, Lin SH, Wang CC, et al. Impact of MAFLD on HBV-related stage 0/A hepatocellular carcinoma after curative resection. *J Pers Med*. 2021;11(8):684. doi:10.3390/jpm11080684
15. Israelsen M, Francque S, Tsochatzis EA, Krag A. Steatotic liver disease. *Lancet*. 2024;404(10464):1761–1778. doi:10.1016/s0140-6736(24)01811-7
16. Du D, Liu C, Qin M, et al. Metabolic dysregulation and emerging therapeutical targets for hepatocellular carcinoma. *Acta Pharm Sin B*. 2022;12(2):558–580. doi:10.1016/j.apsb.2021.09.019
17. Wang J, He W, Tsai PJ, et al. Mutual interaction between endoplasmic reticulum and mitochondria in nonalcoholic fatty liver disease. *Lipids Health Dis*. 2020;19(1):72. doi:10.1186/s12944-020-01210-0
18. Huby T, Gautier EL. Immune cell-mediated features of non-alcoholic steatohepatitis. *Nat Rev Immunol*. 2022;22(7):429–443. doi:10.1038/s41577-021-00639-3
19. Ma C, Kesarwala AH, Eggert T, et al. NAFLD causes selective CD4(+) T lymphocyte loss and promotes hepatocarcinogenesis. *Nature*. 2016;531(7593):253–257. doi:10.1038/nature16969
20. Murai H, Kodama T, Maesaka K, et al. Multiomics identifies the link between intratumor steatosis and the exhausted tumor immune microenvironment in hepatocellular carcinoma. *Hepatology*. 2023;77(1):77–91. doi:10.1002/hep.32573
21. Huang SC, Liu CJ. Chronic hepatitis B with concurrent metabolic dysfunction-associated fatty liver disease: challenges and perspectives. *Clin Mol Hepatol*. 2023;29(2):320–331. doi:10.3350/cmh.2022.0422
22. Wang MM, Wang GS, Shen F, Chen GY, Pan Q, Fan JG. Hepatic steatosis is highly prevalent in hepatitis B patients and negatively associated with virological factors. *Dig Dis Sci*. 2014;59(10):2571–2579. doi:10.1007/s10620-014-3180-9
23. Wu YL, Peng XE, Zhu YB, Yan XL, Chen WN, Lin X. Hepatitis B virus X protein induces hepatic steatosis by enhancing the expression of liver fatty acid binding protein. *J Virol*. 2016;90(4):1729–1740. doi:10.1128/jvi.02604-15
24. Liu PT, Hwang AC, Chen JD. Combined effects of hepatitis B virus infection and elevated alanine aminotransferase levels on dyslipidemia. *Metabolism*. 2013;62(2):220–225. doi:10.1016/j.metabol.2012.07.022
25. Li H, Xu QY, Xie Y, Luo JJ, Cao HX, Pan Q. Effects of chronic HBV infection on lipid metabolism in non-alcoholic fatty liver disease: a lipidomic analysis. *Ann Hepatol*. 2021;24:100316. doi:10.1016/j.aohep.2021.100316
26. Chen JY, Wang JH, Lin CY, et al. Lower prevalence of hypercholesterolemia and hyperglyceridemia found in subjects with seropositivity for both hepatitis B and C strains independently. *J Gastroenterol Hepatol*. 2010;25(11):1763–1768. doi:10.1111/j.1440-1746.2010.06300.x
27. Liang L, Li C, Wang MD, et al. Development and validation of a novel online calculator for estimating survival benefit of adjuvant transcatheter arterial chemoembolization in patients undergoing surgery for hepatocellular carcinoma. *J Hematol Oncol*. 2021;14(1):165. doi:10.1186/s13045-021-01180-5
28. Yang Z, Fu Y, Wang Q, et al. Dynamic changes of serum α-fetoprotein predict the prognosis of bevacizumab plus immunotherapy in hepatocellular carcinoma. *Int J Surg*. 2025;111(1):751–760. doi:10.1097/js9.0000000000001860
29. Fan R, Papatheodoridis G, Sun J, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. *J Hepatol*. 2020;73(6):1368–1378. doi:10.1016/j.jhep.2020.07.025
30. Onwuka JU, Okekunle AP, Olutola OM, Akpa OM, Feng R. Lipid profile and risk of ovarian tumours: a meta-analysis. *BMC Cancer*. 2020;20(1):200. doi:10.1186/s12885-020-6679-9
31. Xu L, Yang Q, Zhou J. Mechanisms of abnormal lipid metabolism in the pathogenesis of disease. *Int J Mol Sci*. 2024;25(15):8465. doi:10.3390/ijms25158465
32. Nagarajan SR, Cross E, Sanna F, Hodson L. Dysregulation of hepatic metabolism with obesity: factors influencing glucose and lipid metabolism. *Proc Nutr Soc*. 2022;81(1):1–11. doi:10.1017/s0029665121003761
33. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*. 2023;186(8):1772–1791. doi:10.1016/j.cell.2023.01.035
34. Kumar Y, Gupta S, Singla R, Hu YC. A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Arch Comput Methods Eng*. 2022;29(4):2043–2070. doi:10.1007/s11831-021-09648-w
35. Al-Tashi Q, Saad MB, Muneer A, et al. Machine learning models for the identification of prognostic and predictive cancer biomarkers: a systematic review. *Int J Mol Sci*. 2023;24(9):7781. doi:10.3390/ijms24097781
36. Kaneva AM, Bojko ER. Fatty liver index (FLI): more than a marker of hepatic steatosis. *J Physiol Biochem*. 2024;80(1):11–26. doi:10.1007/s13105-023-00991-z
37. Starekova J, Hernando D, Pickhardt PJ, Reeder SB. Quantification of liver fat content with CT and MRI: state of the art. *Radiology*. 2021;301(2):250–262. doi:10.1148/radiol.2021204288