Journal of Hepatocellular Carcinoma

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

# Machine Learning Prognostic Model for Post-Radical Resection Hepatocellular Carcinoma in Hepatitis B Patients

Dalong Zhu[1,*], Alimu Tulahong[1,*], Abuduhaiwaier Abuduhelili[1,*], Chang Liu[1], Ayinuer Aierken[1], Yanze Lin[1], Tiemin Jiang[1], Renyong Lin[2], Yingmei Shao[1], Tuerganaili Aji[1]

[1]Department of Hepatobiliary and Echinococcosis Surgery, Digestive and Vascular Surgery Center, the First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang Uygur Autonomous Region, People's Republic of China; [2]State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Medical Research Institute, the First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang Uygur Autonomous Region, People's Republic of China

*These authors contributed equally to this work

Correspondence: Tuerganaili Aji, Department of Hepatobiliary and Echinococcosis Surgery, Digestive and Vascular Surgery Center, the First Affiliated Hospital of Xinjiang Medical University, 137 South Liyushan Road, Urumqi, Xinjiang Uygur Autonomous Region, 830054, People's Republic of China, Fax +8609914364556, Email tuergan78@sina.com

**Purpose:** Primary liver cancer, predominantly hepatocellular carcinoma (HCC), constitutes a substantial global health challenge, characterized by a poor prognosis, particularly in regions with high prevalence of hepatitis B virus (HBV) infection, such as China. This study sought to develop and validate a machine learning-based prognostic model to predict survival outcomes in patients with HBV-related HCC following radical resection, with the potential to inform personalized treatment strategies.

**Patients and Methods:** This study retrospectively analyzed clinical data from 146 patients at Xinjiang Medical University and 75 patients from The Cancer Genome Atlas (TCGA) database. A prognostic model was developed using a machine learning algorithm and evaluated for predictive performance using the concordance index (C-index), calibration curve, decision curve analysis (DCA), and receiver operating characteristic (ROC) curves.

**Results:** Key predictors for constructing the best model included body mass index (BMI), albumin (ALB) levels, surgical resection method (SRM), and the American Joint Committee on Cancer (AJCC) stage. The model achieved a C-index of 0.736 in the training set and performed well in both training and validation datasets. It accurately predicted 1-, 3-, and 5-year survival rates, with Area Under the Curve (AUC) values of 0.843, 0.797, and 0.758, respectively. Calibration curve analysis and Decision Curve Analysis (DCA) further validated the model's predictive capability, suggesting its potential use in clinical decision-making.

**Conclusion:** The study highlights the importance of BMI, ALB, SRM, and AJCC staging in predicting HBV-related HCC outcomes. The machine learning model aids clinicians in making better treatment decisions, potentially enhancing patient outcomes.

**Keywords:** hepatocellular carcinoma, hepatitis B virus, machine learning, prognostic model

## Introduction

Primary liver cancer is the sixth most common cancer and the third leading cause of cancer-related deaths worldwide, with hepatocellular carcinoma (HCC) accounting for 75% to 85% of cases.[1] In 2022, approximately 865,000 new cases of liver cancer and 758,000 death cases were reported globally, with China accounting for 42.51% of the global incidence and 41.75% of the global mortality, highlighting the significant health burden of liver cancer in China.[2,3] The main risk factors for liver cancer include chronic hepatitis B virus (HBV) infection, which is particularly prominent in China and Asia.[4] In recent years, significant advancements have been made in the clinical treatment of HCC, including liver transplantation, ablation therapy, and systemic antitumor treatments. Concurrently, in-depth multi-omics studies, such as genomics and radiomics, have greatly enhanced our understanding of the molecular mechanisms of HCC, promoting the

Received: 6 September 2024
Accepted: 8 February 2025
Published: 19 February 2025

modulation of the tumor immune microenvironment, the integration of imaging features with immune molecules, the fusion of traditional Chinese medicine with modern molecular biology, and the development of gene therapy.[5–7] These studies have provided new scientific evidence for HCC treatment and paved new avenues for future research directions and clinical applications. However, radical resection remains key for long-term survival in patients with liver cancer, with a 5-year survival rate of only 30% for HCC patients, emphasizing the importance of studying various factors affecting survival after curative resection of HCC and implementing personalized interventions and treatment strategies.[8,9] The development of postoperative predictive models is crucial for assisting clinicians in formulating more precise treatment plans and improving therapeutic outcomes.

Machine learning technology, by analyzing large-scale medical data (such as molecular markers, imaging data, and laboratory indicators), has shown tremendous potential in cancer diagnosis, prognosis assessment, and personalized treatment.[10–13] This study utilized multiple machine learning algorithms to develop and validate a postoperative prognostic model for patients with HBV-related HCC, aiming to provide scientific evidence for personalized treatment strategies and decision support to improve patient outcomes.

# Methods

## Inclusion of Subjects

In this study, patients diagnosed with HCC and treated at the First Affiliated Hospital of Xinjiang Medical University between January 2013 and December 2018, as well as those with HCC documented in The Cancer Genome Atlas (TCGA) database, were selected for analysis.[14] The cohort from the First Affiliated Hospital of Xinjiang Medical University was designated as the training set for the development of the prognostic nomogram. In contrast, the cohort from the TCGA database served as the validation set to evaluate the accuracy and generalizability of the model.

The inclusion criteria for the study were as follows: (1) Asian patients aged 18 years or older; (2) positive for hepatitis B surface antigen; (3) underwent radical resection; (4) clinical diagnosis of HCC prior to surgery with subsequent pathological confirmation of HCC post-surgery; (5) availability of comprehensive clinical data before, during, and after the surgical procedure.

The exclusion criteria included: (1) patients diagnosed with hepatitis A, hepatitis C, other hepatitis virus infections, or human immunodeficiency virus infection; (2) patients who had undergone chemotherapy, transcatheter arterial chemoembolization, radiation therapy, ablation treatment, systemic antitumor therapy, or similar interventions; (3) patients presenting with distant metastasis on preoperative imaging and classified as American Joint Committee on Cancer (AJCC) stage IV; (4) patients with a history of other malignancies; (5) patients with liver function classified as Child-Pugh class C; and (6) patients who died of non-tumor-related causes within one month post-surgery.

In this study, baseline data were systematically collected from the point of initial diagnosis of HCC. A comprehensive range of clinical parameters was gathered, including age, gender, HBV infection status, body mass index (BMI), family history of cancer, alpha-fetoprotein (AFP) levels, platelet count, total bilirubin levels, albumin (ALB) levels, creatinine levels, Child-Pugh liver function classification, AJCC stage, presence of vascular invasion (VI), surgical resection method (SRM), neoplasm histologic grade, albumin-bilirubin (ALBI) grade, and platelet-albumin-bilirubin (PALBI) grade. Furthermore, overall survival (OS) was defined as the interval from the date of surgery to either the date of follow-up or the date of patient death. The follow-up period ended on October 31, 2022.

The biological characteristics of tumors, such as tumor size, multifocality, the presence of liver cirrhosis as a comorbidity, and the existence of satellite nodules, are recognized prognostic factors for HCC and are associated with patients' progression-free survival and OS.[15–17] Although this study systematically collected the aforementioned baseline parameters, it was found during data integration that the TCGA database lacks standardized records for such tumor biological features. Notably, the AJCC staging system has integrated some key tumor biological information. To ensure the usability of cross-database data, this study ultimately decided to forgo the additional indicators and construct the prognostic model using only the common variables. Although this strategy may somewhat diminish the model's discriminative power, the model based on standardized indicators can still provide generalized references for treatment decision-making.

In this study, we utilized various combinations of seven machine learning algorithms—Cox Gradient Boosting (CoxBoost), Supervised Principal Components (SuperPC), Partial Least Squares Regression for Cox Models and Related Techniques (plsRcox), Stepwise Cox Regression (StepCox), Least Absolute Shrinkage and Selection Operator (Lasso), Survival Support Vector Machine (survivalSVM), and Ridge Regression (Ridge)—to construct prognostic models for patients with HBV-related HCC.[18] To assess the predictive performance of these models, we employed the concordance index (C-index) to analyze and identify the optimal model-incorporated variables as prognostic factors. Subsequently, we developed a novel nomogram to estimate the 1-, 3-, and 5-year survival rates of patients. To determine the accuracy and reliability of the nomogram, we utilized several statistical methods, including the C-index, calibration curves generated through 1000 bootstrap resampling, decision curve analysis (DCA), and receiver operating characteristic (ROC) curve analysis. In the description of fundamental characteristics, categorical variables are presented as frequencies (n) and percentages (%), and comparisons are made using the chi-squared test ($\chi^2$-test) or Fisher's exact test. For continuous variables that do not follow a normal distribution, the median (range) is provided. Where appropriate, the *t*-test or Mann–Whitney *U*-test was employed for analysis. All statistical analyses were conducted using R software version 4.4.1 and SPSS software version 26.0. A P-value of less than 0.05 was considered statistically significant.

# Results

## Subject Characteristics

After strict application of the inclusion and exclusion criteria, a total of 146 patients with HBV-related HCC who underwent partial hepatectomy at the First Affiliated Hospital of Xinjiang Medical University were eligible for this study. Simultaneously, 75 patients were selected from a cohort of 378 hCC patients in the TCGA database and included in the analysis. The study flow chart is presented in Figure 1, and the baseline clinical and pathological characteristics of the subjects are comprehensively detailed in Table 1. Additional unique features of the subjects in the training cohort are included in Supplementary Material Table S1. In the training cohort, the median follow-up duration for the subjects was 44.50 months, ranging from 26.00 to 72.00 months. The OS rates at 1-year, 3-year, and 5-year intervals were 85.62%, 69.01%, and 53.16%, respectively. Conversely, in the validation cohort, the median survival duration was 26.63 months, ranging from 15.69 to 52.71 months. The 1-year, 3-year, and 5-year OS rates for patients in the validation cohort were 94.45%, 88.59%, and 83.67%, respectively.

## Predictive Model Characteristics

The optimal model was identified through an analysis of the C-index and its average values across both the training and validation datasets of the machine learning model (Figure 2). Specifically, it was observed that the models integrating CoxBoost with plsRcox, Lasso with plsRcox, and StepCox with plsRcox showed consistent performance. The C-index values for these models were 0.736, 0.629, and 0.682 for the training set, validation set, and the average of the two sets, respectively. Among these models, BMI, ALB, SRM, and the AJCC stage have been identified as the four key variables for predicting the survival time of HCC patients. Furthermore, univariate and multivariate Cox regression analyses were conducted within the training cohort, yielding results that were unexpectedly consistent with those of the aforementioned model analysis (Supplementary Material Table S2). This consistency further substantiates the significance of these variables.

## Construction and Validation of Nomograms

Based on the identified factors, a nomogram was developed to predict 1-, 3-, and 5-year survival rates for patients with HBV-related HCC (Figure 3A). In the training cohort, the nomogram demonstrated a C-index of 0.736, reflecting a high level of predictive accuracy. The total score for each patient was calculated using the nomogram, and the cohort was subsequently stratified into high-risk and low-risk groups based on the median score. The survival duration of the low-risk cohort was significantly longer compared to that of the high-risk cohort across all groups (training cohort: $P < 0.001$, validation cohort: $P = 0.036$) (Figure 3B and C), thereby validating the efficacy of the nomogram. Calibration curve analysis indicated that within the training cohort (Figure 4A), the observed outcomes closely matched the survival rates

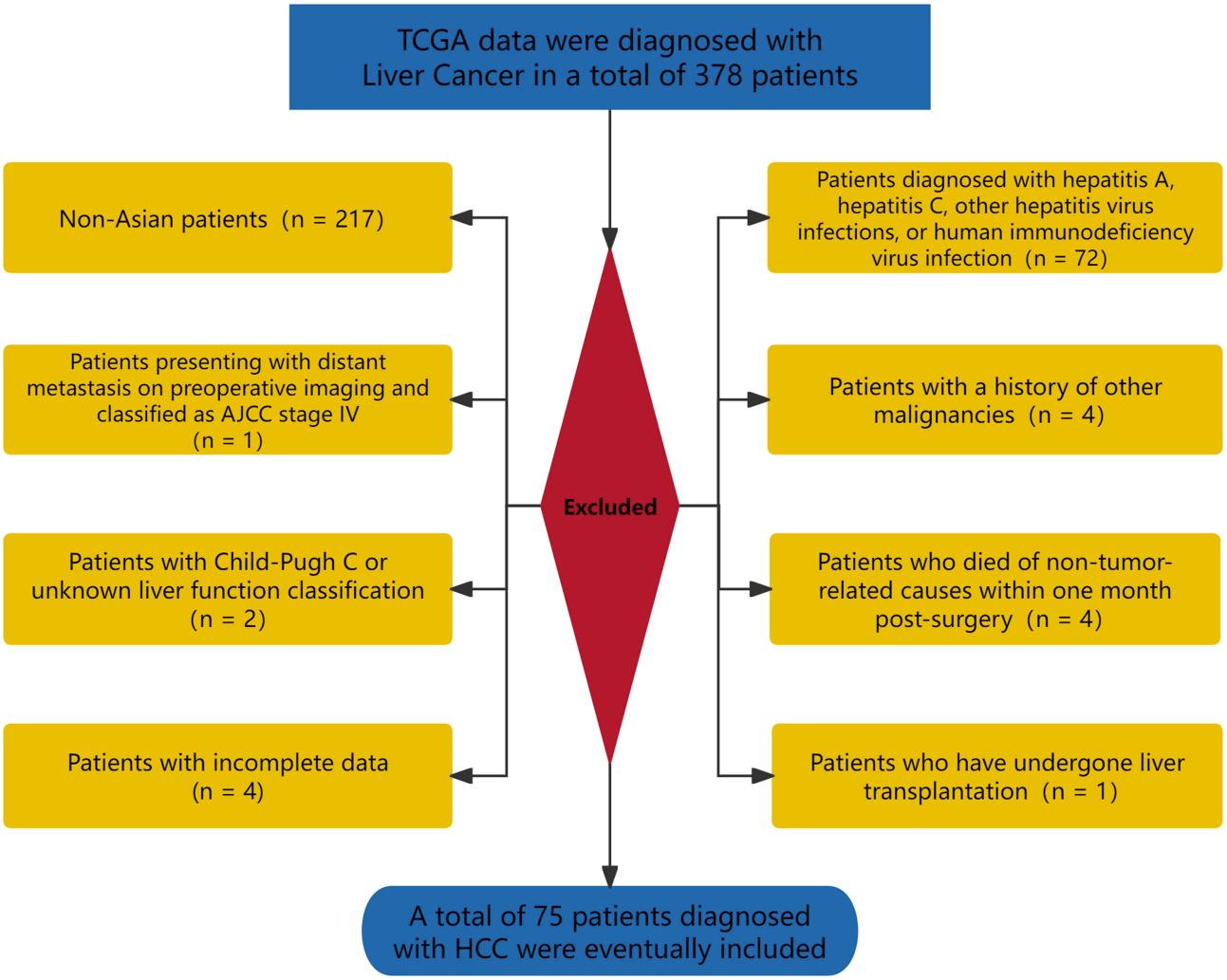


**Figure 1** The flow chart of patients in TCGA database was selected according to the exclusion criteria.

projected by the nomogram at 1, 3, and 5 years. DCA, as illustrated in Figure 4B, demonstrated that the net benefit of the nomogram exceeded that of both the "all treatment" and "no treatment" strategies, highlighting its potential utility in clinical decision-making. Additionally, the ROC curves, generated using the timeROC package in R software, confirmed the predictive capability of the nomogram. Specifically, the Area Under the Curve (AUC) values of the nomogram-related ROC curves for the training cohort were 0.843, 0.797, and 0.758 for 1-year, 3-year, and 5-year survival rates, respectively (Figure 4C), indicating strong performance in survival prediction. Subsequent analysis revealed that the AUC value of the nomogram for predicting 1-year survival was 0.843, surpassing the AUC values of individual clinical characteristics, including BMI, ALB, SRM, and the AJCC staging system, which were 0.619, 0.649, 0.734, and 0.719, respectively (Figure 4D).

## Discussion

The global incidence and mortality rates of HCC continue to rise,[1,19] with major risk factors including HBV and Hepatitis C Virus infections, excessive alcohol consumption, and exposure to aflatoxins. HBV infection is the leading cause of HCC worldwide, especially in China, East Asia, and Africa, where HBV-related HCC cases account for up to 80%.[4] Although HBV vaccination programs have significantly reduced the annual incidence and the number of new cases in China, there are still hundreds of thousands of new cases and deaths each year.[2] HCC patients can choose from a variety of treatment options, including liver transplantation, surgical resection, radiofrequency ablation, transarterial

**Table 1** Characteristics of the Cohorts

| Variable | Training Cohort 146 | Validation Cohort 75 | *P* |
|---|---|---|---|
| Gender | | | 0.79 |
| Female | 29 (19.86%) | 13 (17.33%) | |
| Male | 117 (80.14%) | 62 (82.67%) | |
| Age | | | 1.00 |
| <60 | 104 (71.23%) | 53 (70.67%) | |
| ≥60 | 42 (28.77%) | 22 (29.33%) | |
| BMI | 24.05 (21.90–26.10) | 24.13 (22.31–25.28) | 0.84 |
| AFP | | | 0.39 |
| ≤20 | 61 (41.78%) | 36 (48.00%) | |
| >20 - ≤400 | 48 (32.88%) | 18 (24.00%) | |
| >400 | 37 (25.34%) | 21 (28.00%) | |
| Total Bilirubin | | | 0.84 |
| Normal | 135 (92.47%) | 68 (90.67%) | |
| Higher_than_normal | 11 (7.53%) | 7 (9.33%) | |
| ALB | | | 0.05 |
| Normal | 88 (60.27%) | 56 (74.67%) | |
| Below_normal_values | 58 (39.73%) | 19 (25.33%) | |
| Platelet | | | 0.13 |
| Normal | 95 (65.07%) | 57 (76.00%) | |
| Below_normal_values | 51 (34.93%) | 18 (24.00%) | |
| Creatinine | | | 0.60 |
| Normal | 126 (86.30%) | 62 (82.67%) | |
| Below_normal_values | 20 (13.70%) | 13 (17.33%) | |
| Child-Pugh Classification | | | 0.88 |
| A | 136 (93.15%) | 71 (94.67%) | |
| B | 10 (6.85%) | 4 (5.33%) | |
| Vascular Invasion | | | 0.01 |
| None | 109 (74.66%) | 55 (73.33%) | |
| Micro | 19 (13.01%) | 18 (24.00%) | |
| Macro | 18 (12.33%) | 2 (2.67%) | |
| Surgical Resection Method | | | 0.00 |
| Segmentectomy_Single | 60 (41.10%) | 21 (28.00%) | |
| Segmentectomy_Multiple | 24 (16.44%) | 20 (26.67%) | |
| Lobectomy | 32 (21.92%) | 32 (42.67%) | |
| Extended_Lobectomy | 30 (20.55%) | 2 (2.67%) | |
| Neoplasm Histologic Grade | | | 0.00 |
| G1/2 | 98 (67.12%) | 28 (37.33%) | |
| G3/4 | 48 (32.88%) | 47 (62.67%) | |
| AJCC | | | 0.49 |
| Stage I | 105 (71.92%) | 54 (72.00%) | |
| Stage II | 25 (17.12%) | 16 (21.33%) | |
| Stage III | 16 (10.96%) | 5 (6.67%) | |
| ALBI | | | 0.74 |
| 1 | 88 (60.27%) | 48 (64.00%) | |
| 2 | 57 (39.04%) | 26 (34.67%) | |
| 3 | 1 (0.68%) | 1 (1.33%) | |
| PALBI | | | 0.66 |
| 1 | 79 (54.11%) | 43 (57.33%) | |
| 2 | 53 (36.30%) | 23 (30.67%) | |
| 3 | 14 (9.59%) | 9 (12.00%) | |

**Abbreviations**: BMI, body mass index; AFP, alpha-fetoprotein; ALB, albumin; AJCC, American Joint Committee on Cancer; ALBI, albumin-bilirubin; PALBI, platelet-albumin-bilirubin;


https://doi.org/10.2147/JHC.S495059

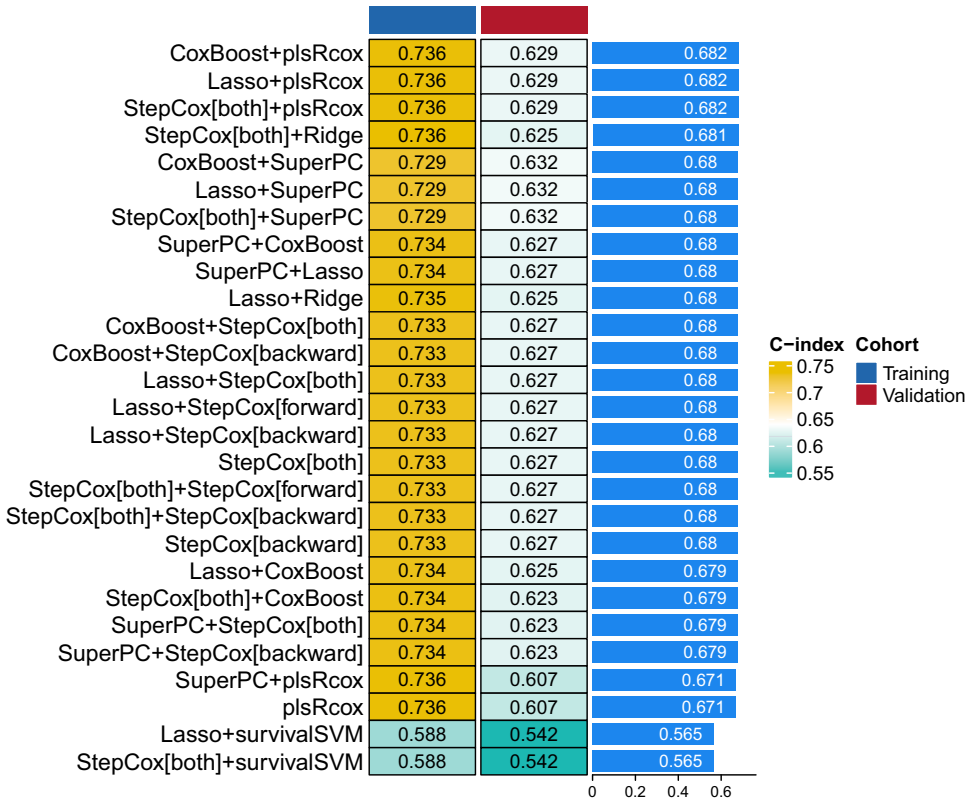


**Figure 2** Comparison of C-index among Prognostic models created by various algorithm combinations.

chemoembolization, and systemic therapy.[20] Surgical resection, as the main treatment for long-term patient survival, is limited by conditions and not all patients can undergo it.[21] The poor prognosis of HCC is mainly attributed to high recurrence rates and low cumulative survival rates, with late diagnosis limiting the feasibility of therapeutic interventions. However, advances in systemic therapies and neoadjuvant treatments have provided the possibility of surgical resection for some patients who were previously inoperable, increasing the likelihood of long-term survival.[22,23] The development of personalized postoperative treatment strategies is crucial for optimizing patient outcomes, requiring the assessment of various factors including tumor staging and liver function. Advances in detection technology have facilitated the application of genetic testing, immune microenvironment markers, tumor microbiota, and exosomes in the development of predictive models, offering new prospects for personalized treatment and prognosis assessment of HCC.[24–26] However, the availability of these complex detection technologies is limited in remote areas and primary healthcare facilities, restricting their application. Therefore, the development of predictive models relying on routine clinical laboratory indicators is particularly important.

The integration of machine learning technology in clinical oncology has improved the accuracy of prognosis prediction for cancer patients, provided information for the development of personalized treatment plans, and simplified clinical workflows.[10–13] It is anticipated that with continued technological advancements and clinical validation, the role of machine learning in cancer treatment will become increasingly significant. This study utilizes machine learning technology to develop a prognostic model as a tool to elucidate the prognosis of HBV-related HCC patients, enhancing the precision of clinical decision-making and potentially providing more personalized and effective treatment strategies for patients.

In this study, we identified BMI, ALB, SRM, and AJCC stage as significant prognostic risk factors for HBV-related HCC based on clinical practice. The incorporation of these variables enhances the predictive accuracy of the model and offers clinicians a more precise prognostic assessment tool. In recent years, the association between obesity, metabolic syndrome, and liver cancer has garnered increasing attention. BMI, as a critical indicator of obesity, has been extensively

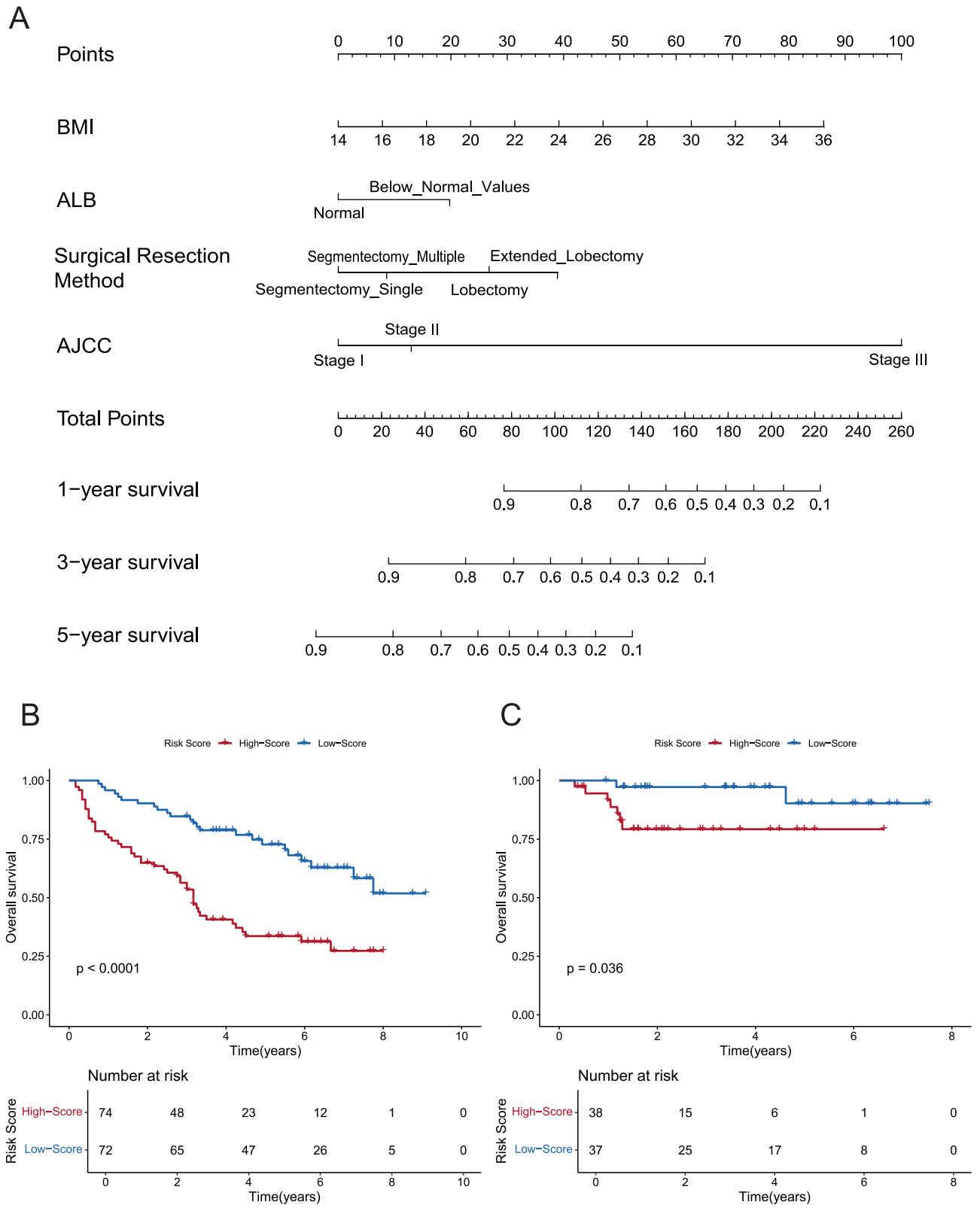


**Figure 3** Nomogram model for predicting 1-, 3- and 5-year OS in patients with HCC (**A**). The nomogram is used by summing all the points determined by each variable on the scale. The total score projected on the bottom scale represents the probability of 1-, 3- and 5-year survival. The Kaplan-Meier curve of OS of patients with HCC in the training cohort (**B**) and the verification cohort (**C**).


https://doi.org/10.2147/JHC.S495059

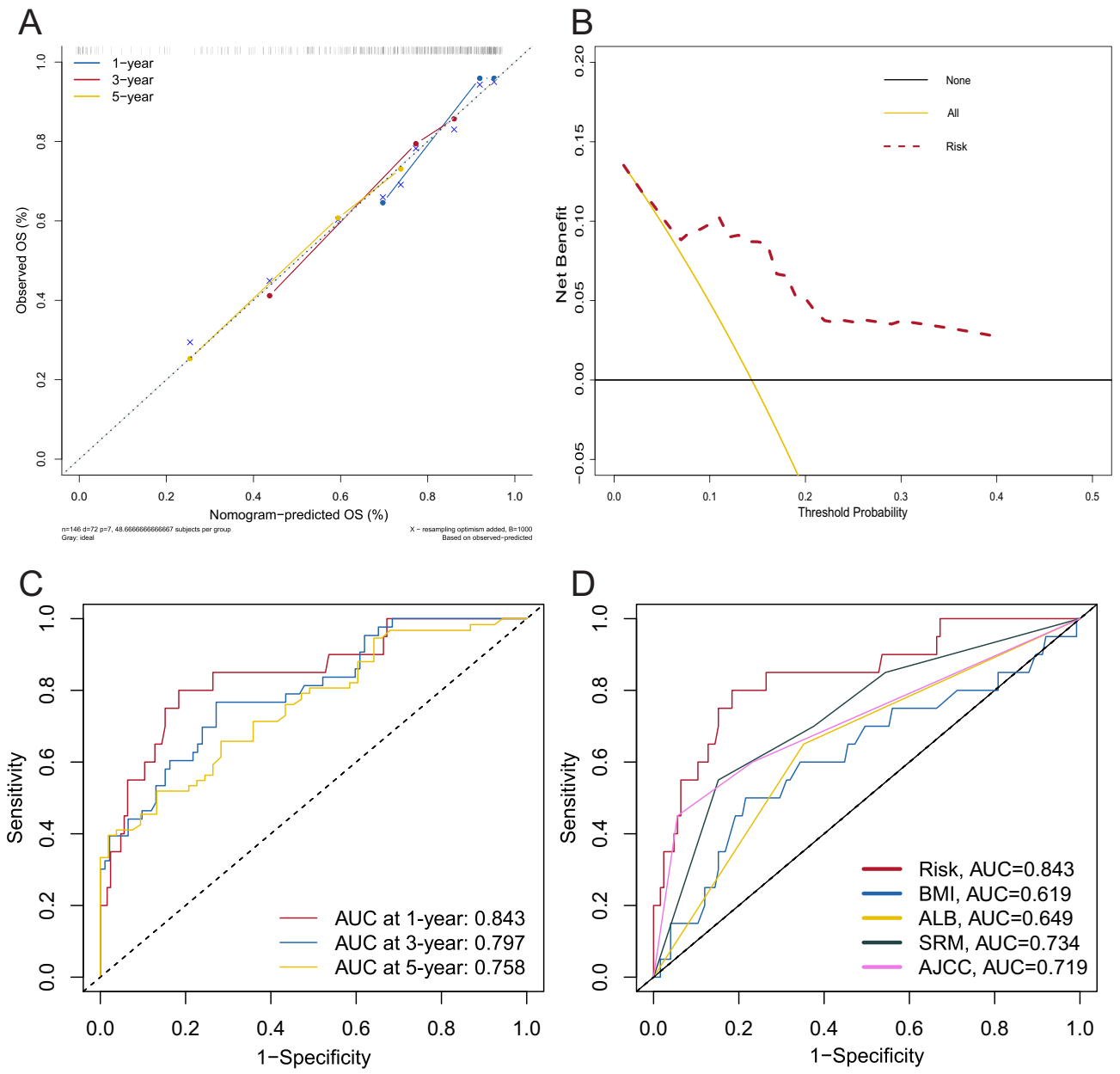


**Figure 4** The calibration curves of 1-, 3- and 5-year OS were predicted in the training cohort (**A**). DCA of the nomogram indicated the net clinical benefits (**B**). AUC values of the nomogram-related time-dependent ROC curves for the training cohort for 1-, 3-, and 5-year survival rates (**C**). AUC values of the nomogram and other clinical characteristics related time-dependent ROC curves for the training cohort survival rates (**D**).

examined in numerous studies. Obesity and overweight have been demonstrated to increase the risk of HCC.[27,28] The substantial correlation between adult weight gain and the risk of HCC may be attributable to heightened insulin resistance, inflammation, and oxidative stress.[27] In individuals with cirrhosis, obesity may further exacerbate the risk of HCC by disrupting protein homeostasis and amino acid metabolism.[29] Other studies have demonstrated a significant association between BMI and the risk of HCC in men, indicating a 30% increase in HCC risk per 5 kg/m² increase in BMI. Conversely, this association is not statistically significant in women, potentially due to the higher accumulation of visceral fat in men compared to women.[30,31] Regarding prognosis, patients with HCC and a high body mass index (BMI $\geq$ 25.0 kg/m²) exhibit a lower survival rate compared to those with a normal BMI (18.5–24.9 kg/m²). Post liver resection, individuals with elevated BMI are at an increased risk of complications, including bleeding and infection, as well as adverse events such as perioperative blood transfusion and extended postoperative hospital stays. These complications

are frequently associated with comorbid conditions such as diabetes, fatty liver disease, and non-alcoholic steatohepatitis.[32] Simultaneously, a low BMI (defined as BMI < 18.5 kg/m²) poses similar challenges for patients, although due to different underlying factors. A low BMI is often correlated with sarcopenia and reduced albumin levels, which can impair the regenerative capacity of the remaining liver and diminish the body's immune function, thereby weakening the patient's ability to manage physiological stress.[33,34] However, other studies have reported no significant differences between high BMI and normal BMI with respect to postoperative complication rates and postoperative mortality rates.[35] This lack of significant difference may be attributable to the inclusion of patients with low BMI within the "normal BMI" category, thereby mitigating any disparities when compared to patients with high BMI. In the context of liver transplantation, numerous studies have investigated the relationship between BMI and transplantation outcomes. Currently, there is no definitive evidence to suggest that BMI significantly influences the recurrence rate or long-term survival of patients following transplantation for HCC.[36] We are considering whether the BMI of the donor may affect the quality of the liver and thus the recurrence rate and survival rate of the recipient after transplantation, but there is currently a lack of relevant research to draw a definitive conclusion.

Albumin serves as a critical biomarker for evaluating the nutritional and immune status of patients. Traditionally, the Child-Pugh score system has been widely used as a primary indicator for predicting postoperative complications in liver disease. However, recent advancements in the prognostic assessment of liver cancer have led to the development of various novel scoring systems, including the ALBI, PALBI, and age-male-ALBI-platelets score.[37] These emerging systems prominently incorporate albumin as a core component, highlighting its pivotal role in assessing the risk associated with liver cancer. Research indicates that ALB exhibits direct growth-inhibitory properties across various malignancies, notably HCC and breast cancer. Concurrently, ALB levels are closely associated with survival outcomes in liver cancer patients, with decreased ALB levels typically signifying a poor prognosis.[38,39] In numerous clinical cohort studies of HCC, low ALB levels have been inversely correlated with invasive characteristics of liver cancer, such as increased tumor diameter, higher incidence of portal vein thrombosis, multiple tumor foci, and elevated AFP levels, all of which contribute to a worse prognosis for HCC patients.[40] Other studies have demonstrated that the supplementation of albumin in the culture medium of HCC cell lines can attenuate cell proliferation and reduce AFP production. Conversely, diminished levels of albumin may facilitate the migration and invasion of HCC cells by upregulating the expression of urokinase-type plasminogen activator receptor and matrix metalloproteinases.[41] Although the precise mechanisms by which albumin influences HCC progression remain to be elucidated, these findings emphasize the significance of albumin in the prognosis and treatment of HCC.

Our findings indicate that surgical resection and the AJCC staging are significant prognostic factors for HBV-related HCC. Specifically, the study revealed that an increased extent of surgical resection and a higher AJCC stage correlate with a decreased overall survival rate in patients. This observation may be attributable to the propensity for larger tumors to exhibit vascular invasion and multiple lesions, as well as an elevated risk of distant metastasis.[42,43] Furthermore, insufficient residual liver function post-resection remains a critical concern. Resection of larger tumors leads to a diminished residual liver volume, thereby elevating perioperative morbidity and reducing overall survival rates.[44] Despite variability in the tumor size thresholds across different studies, the conclusions remain largely consistent. Several authors have undertaken comprehensive investigations into solitary HCC in patients with cirrhosis.[45] For instance, Lang et al demonstrated that for tumors with diameters ranging from 3 to 5 cm, extensive hepatectomy was associated with improved disease-free survival and overall survival.[46] However, Dahiya et al did not identify a significant difference between the two groups; it did, however, reveal that patients in the small resection group exhibited more severe liver disease ($P = 0.005$).[47] We hypothesize that the increased severity of liver disease in the small resection group may have mitigated the differences in final outcomes between the two groups. Additionally, the tumor's location is a critical determinant in selecting the type of surgical resection and significantly influences the patient's prognosis. Numerous studies have demonstrated that tumors situated on the left side of the liver are associated with lower OS and recurrence-free survival rates compared to those on the right side.[48] Tang et al proposed that recurrence following liver resection for HCC predominantly occurs intrahepatically. They suggested that left-sided HCCs result in a larger residual liver volume post-resection and may have narrower resection margins, thereby elevating the risk of recurrence.[46] Despite the

limitations of the traditional AJCC TNM staging system in prognosticating cancer patients, it is widely recognized as a key component of the new prognostic model and remains integral to the assessment of HCC prognosis.

Although several nomograms have been developed to predict the prognosis of HCC, models specifically tailored for HBV-related HCC remain scarce.[49,50] For instance, in the study conducted by Su et al, the C-index of the training set achieved a value of 0.743, while the AUC values for predicting 1-year and 3-year survival rates were 0.784 and 0.779, respectively.[49] Our study demonstrated a comparable C-index to that of Su et al, but exhibited superior AUC values, suggesting that our model possesses enhanced predictive performance. It is important to note that the study conducted by Su et al did not specifically include patients with HCC who had undergone radical resection. Conversely, the C-index of the nomogram developed by Mo et al for patients with HCC post-radical resection was 0.680, a value that is notably lower than that of our model.[50] Emphasis should be placed on the fact that this study is primarily based on a specific patient population associated with HBV-related HCC. Therefore, the applicability of the model in patients with HCC not triggered by HBV infection or in other populations still needs further validation. In China, the main cause of HCC is HBV infection, while in other regions, HCC is more often caused by HCV infection or alcoholic liver disease, which may lead to significant differences in tumor characteristics among patients. Thus, the applicability of this model may be stronger among Chinese patients, but its application in other regions still requires further confirmation through external validation. Furthermore, there may be differences in the predictive effectiveness of the model for patients of different genders due to the varying roles of BMI in the pathogenesis of HCC. Studies have shown that BMI exhibits a more significant correlation in male HCC patients, while its impact is less pronounced in female patients.[30,31] This phenomenon indicates that future research needs to pay more attention to the potential impact of gender differences on the predictive performance of the model and further optimize the model to adapt to the characteristics of different patient groups.

Although our study provides valuable insights, it is important to recognize its inherent limitations. First, as a retrospective study, the possibility of selection bias cannot be completely ruled out, which may affect the generalizability and applicability of the findings. Second, while we analyzed independent datasets from our center and the TCGA database, both have limited sample sizes, and differences in baseline information may lead to decreased performance of the model on the validation set. To mitigate these shortcomings, we employed cross-validation; however, the statistical power and extrapolative capacity of the model remain constrained. Lastly, it is important to note that the data sets utilized in our study were derived from earlier years. Given the recent widespread adoption of advanced surgical techniques, such as laparoscopic and robotic liver resection, in the treatment of HCC, the applicability of our prediction model in contemporary clinical practice may be limited. To address these limitations, we hope that future research endeavors collect data from a broader range of centers, incorporating larger sample sizes and more comprehensive indicators. The acquisition of such data will facilitate the further validation and optimization of our model, thereby enhancing its predictive accuracy and clinical relevance. We posit that through the continuous refinement and updating of the model, we can offer clinicians more precise and personalized decision support for treatment, especially in the context of managing HBV-related HCC.

## Conclusions

Our machine learning-based models, which incorporate BMI, ALB level, surgical resection status, and the AJCC staging nomogram, demonstrated strong performance in predicting survival outcomes following radical resection of HBV-related HCC. These models provide precise prognostic evaluations for this patient cohort, thereby assisting clinicians in making informed clinical decisions.

## Abbreviations

HCC, hepatocellular carcinoma; HBV, hepatitis B virus; TCGA, The Cancer Genome Atlas; AJCC, American Joint Committee on Cancer; BMI, body mass index; AFP, alpha-fetoprotein; ALB, albumin; VI, vascular invasion; SRM, surgical resection method; ALBI, albumin-bilirubin; PALBI, platelet-albumin-bilirubin; OS, overall survival; CoxBoost, Cox Gradient Boosting; SuperPC, Supervised Principal Components; plsRcox, Partial Least Squares Regression for Cox Models and Related Techniques; StepCox, Stepwise Cox Regression; Lasso, Least Absolute Shrinkage and Selection

Operator; survivalSVM, Survival Support Vector Machine; Ridge, Ridge Regression; C-index, concordance index; DCA, decision curve analysis; ROC, receiver operating characteristic; AUC, Area Under the Curve.

## Data Sharing Statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Ethics Approval and Consent to Participate

This retrospective study received approval from the Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University (approval number: K202104-14) and complied with the standards of the Declaration of Helsinki. Informed consent was obtained from each patient or their relatives.

## 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 in 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

Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01D17); State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia (SKL-HIDCA-2024-2); Xinjiang Uygur Autonomous Region Graduate Innovation Program (XJ2024G153).

## Disclosure

The authors declare that they have no competing interests.

## 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 a Cancer J Clinicians*. 2024;74(3):229–263. doi:10.3322/caac.21834
2. Xie D, Shi J, Zhou J, Fan J, Gao Q. Clinical practice guidelines and real-life practice in hepatocellular carcinoma: a Chinese perspective. *Clin Mol Hepatol*. 2023;29(2):206–216. doi:10.3350/cmh.2022.0402
3. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. *J Natl Cancer Center*. 2024;4(1):47–53. doi:10.1016/j.jncc.2024.01.006
4. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health*. 2020;8(2):e180–e190. doi:10.1016/S2214-109X(19)30488-7
5. Guo BJ, Ruan Y, Wang YJ, et al. Jiedu Recipe, a compound Chinese herbal medicine, inhibits cancer stemness in hepatocellular carcinoma via Wnt/beta-catenin pathway under hypoxia. *J Integr Med*. 2023;21(5):474–486. doi:10.1016/j.joim.2023.06.008
6. Wu J, Liu W, Qiu X, et al. A Noninvasive Approach to Evaluate Tumor Immune Microenvironment and Predict Outcomes in Hepatocellular Carcinoma. *Phenomics*. 2023;3(6):549–564. doi:10.1007/s43657-023-00136-8
7. Ran G, Feng X, Xie Y, et al. The use of miR122 and its target sequence in adeno-associated virus-mediated trichosanthin gene therapy. *J Integr Med*. 2021;19(6):515–525. doi:10.1016/j.joim.2021.09.004
8. YaJing H, ZhiMing Z, WeiMeng H, et al. Epidemiological characteristics and prognosis of hepatocellular carcinoma: a single-center observational real-world cohort study of 1302 cases. *J Clin Hepatol*. 2019;5(35):1002–1007 doi:10.3969/j.issn.1001-5256.2019.05.014
9. Forner A, Reig, M, Bruix, J. Hepatocellular carcinoma. *Lancet*. 2018;10127(391):1301–1314 doi:10.1016/S0140-6736(18)30010-2
10. Nam D, Chapiro J, Paradis V, Seraphin TP, Kather JN. Artificial intelligence in liver diseases: improving diagnostics, prognostics and response prediction. *JHEP Rep*. 2022;4(4):100443. doi:10.1016/j.jhepr.2022.100443
11. Zeng Q, Klein C, Caruso S, et al. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *J Hepatol*. 2022;77(1):116–127. doi:10.1016/j.jhep.2022.01.018
12. Ramírez-Mejía MM, Méndez-Sánchez N. From prediction to prevention: machine learning revolutionizes hepatocellular carcinoma recurrence monitoring. *World J Gastroenterol*. 2024;30(7):631–635. doi:10.3748/wjg.v30.i7.631
13. Zhou SN, Jv DW, Meng XF, et al. Feasibility of machine learning-based modeling and prediction using multiple centers data to assess intrahepatic cholangiocarcinoma outcomes. *Ann Med*. 2023;55(1):215–223. doi:10.1080/07853890.2022.2160008
14. The Cancer Genome Atlas Program. Center for Cancer Genomics. The National Cancer Institute. 2024. https://portal.gdc.cancer.gov. Accessed 12, Jun 2024.

15. Toyoda H, Lai PBS, O'Beirne J, et al. Long-term impact of liver function on curative therapy for hepatocellular carcinoma: application of the ALBI grade. *Br J Cancer*. 2016;114(7):744–750. doi:10.1038/bjc.2016.33
16. Takayama T, Yamazaki S, Matsuyama Y, et al. Prognostic grade for resecting hepatocellular carcinoma: multicentre retrospective study. *Br J Surg*. 2021;108(4):412–418. doi:10.1093/bjs/znaa109
17. Liu F, Yan T, Cui D, Jiang J. Identification and validation of a prognostic model based on four genes related to satellite nodules in hepatocellular carcinoma. *Sci Rep*. 2024;14(1):15633. doi:10.1038/s41598-024-66610-z
18. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev mol Cell Biol*. 2022;23(1):40–55. doi:10.1038/s41580-021-00407-0
19. Alawyia B, Constantinou C. Hepatocellular Carcinoma: a Narrative Review on Current Knowledge and Future Prospects. *Curr Treat Options Oncol*. 2023;24(7):711–724. doi:10.1007/s11864-023-01098-9
20. Brown ZJ, Tsilimigras DI, Ruff SM, et al. Management of Hepatocellular Carcinoma: a Review. *JAMA Surg*. 2023;158(4):410–420 doi:10.1001/jamasurg.2022.7989
21. Vibert MSKM E, Schwartz M, Olthoff KM. Advances in resection and transplantation for hepatocellular carcinoma. *J Hepatol*. 2020;2(72):262–276. doi:10.1016/j.jhep.2019.11.017
22. Chick RC, Ruff SM, Pawlik TM. Neoadjuvant systemic therapy for hepatocellular carcinoma. *Front Immunol*. 2024;15:1355812. doi:10.3389/fimmu.2024.1355812
23. Zhao H, Cai J. Chinese expert consensus on neoadjuvant and conversion therapies for hepatocellular carcinoma. *World J Gastroenterol*. 2021;27(47):8069–8080. doi:10.3748/wjg.v27.i47.8069
24. Yuan W, Xu Y, Wu Z, et al. Cellular senescence-related genes: predicting prognosis in hepatocellular carcinoma. *BMC Cancer*. 2023;23(1):1001. doi:10.1186/s12885-023-11288-1
25. Sun L, Ke X, Guan A, et al. Intratumoural microbiome can predict the prognosis of hepatocellular carcinoma after surgery. *Clin Transl Med*. 2023;13(7):e1331. doi:10.1002/ctm2.1331
26. Kocheise L, Schoenlein M, Behrends B, et al. EpCAM-positive circulating tumor cells and serum AFP levels predict outcome after curative resection of hepatocellular carcinoma. *Sci Rep*. 2023;13(1):20827. doi:10.1038/s41598-023-47580-0
27. Simon TG, Kim MN, Luo X, et al. Adiposity, Adulthood Weight Change, and Risk of Incident Hepatocellular Carcinoma. *Cancer Prevent Res*. 2021;14(10):945–954. doi:10.1158/1940-6207.CAPR-20-0549
28. Jun BG, Kim M, Shin HS, Yi JJ, Yi SW. Impact of overweight and obesity on the risk of hepatocellular carcinoma: a prospective cohort study in 14.3 million Koreans. *Br J Cancer*. 2022;127(1):109–115. doi:10.1038/s41416-022-01771-0
29. Anand AC. Nutrition and Muscle in Cirrhosis. *J Clin Exp Hepatol*. 2017;7(4):340–357. doi:10.1016/j.jceh.2017.11.001
30. Saeed U, Nordsletten M, Myklebust TÅ, et al. Cancer risk and survival according to body mass index in hepatobiliary malignancies: a nationwide registry-based cohort study. *HPB*. 2023;25(11):1382–1392. doi:10.1016/j.hpb.2023.07.882
31. Setiawan VW, Lim U, Lipworth L, et al. Sex and Ethnic Differences in the Association of Obesity With Risk of Hepatocellular Carcinoma. *Clin Gastroenterol Hepatol*. 2016;14(2):309–316. doi:10.1016/j.cgh.2015.09.015
32. Yu JJ, Liang L, Lu L, et al. Association between body mass index and postoperative morbidity after liver resection of hepatocellular carcinoma: a multicenter study of 1324 patients. *HPB*. 2020;22(2):289–297. doi:10.1016/j.hpb.2019.06.021
33. Yang J, Chen K, Zheng C, et al. Impact of sarcopenia on outcomes of patients undergoing liver resection for hepatocellular carcinoma. *J Cachexia, Sarcopenia Muscle*. 2022;13(5):2383–2392. doi:10.1002/jcsm.13040
34. Jin W, Jiang S, Chen A, Chen Y. Effect of Preoperative Malnutrition Based on Albumin and BMI on Hepatocellular Carcinoma Surgery and Prediction of Risk Factors of Complications. *J Gastrointest Cancer*. 2024;55(2):511–518. doi:10.1007/s12029-023-01008-0
35. Liu X, Xu J. Body Mass Index and Waistline are Predictors of Survival for Hepatocellular Carcinoma After Hepatectomy. *Med Sci Monit*. 2015;21:2203–2209. doi:10.12659/MSM.894202
36. El-Domiaty N, Saliba F, Karam V, et al. Impact of body mass index on hepatocellular carcinoma recurrence after liver transplantation through long-term follow-up. *Hepatobiliary Surg Nutr*. 2021;10(5):598–609. doi:10.21037/hbsn.2020.04.01
37. 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
38. Carr BI, Guerra V. Serum Inflammation Parameters and Survival in Hepatocellular Carcinoma Patients: importance of Albumin and Gamma-Glutamyltranspeptidase. *Oncology*. 2023;101(5):313–320. doi:10.1159/000527650
39. Bağırsakçı E, Şahin E, Atabey N, Erdal E, Guerra V, Carr BI. Role of Albumin in Growth Inhibition in Hepatocellular Carcinoma. *Oncology*. 2017;93(2):136–142. doi:10.1159/000471807
40. Carr BI, Guerra V. Serum Albumin Levels in Relation to Tumor Parameters in Hepatocellular Carcinoma Patients. *Int J Biol Markers*. 2017;32(4):391–396. doi:10.5301/ijbm.5000300
41. Fu X, Yang Y, Zhang D. Molecular mechanism of albumin in suppressing invasion and metastasis of hepatocellular carcinoma. *Liver Int*. 2022;42(3):696–709. doi:10.1111/liv.15115
42. Shehta A, Elsabbagh AM, Medhat M, et al. Impact of tumor size on the outcomes of hepatic resection for hepatocellular carcinoma: a retrospective study. *BMC Surg*. 2024;24(1):7. doi:10.1186/s12893-023-02296-w
43. Dai CY, Lin CY, Tsai PC, et al. Impact of tumor size on the prognosis of hepatocellular carcinoma in patients who underwent liver resection. *J Chin Med Assoc*. 2018;81(2):155–163. doi:10.1016/j.jcma.2017.06.018
44. Cha SW, Sohn JH, Kim SH, et al. Interaction between the tumor microenvironment and resection margin in different gross types of hepatocellular carcinoma. *J Gastroenterol Hepatol*. 2020;35(4):648–653. doi:10.1111/jgh.14848
45. Yang A, Xiao W, Chen D, et al. The power of tumor sizes in predicting the survival of solitary hepatocellular carcinoma patients. *Cancer Med*. 2018;7(12):6040–6050. doi:10.1002/cam4.1873
46. Lang BH, Poon RT, Fan ST, Wong J. Perioperative and long-term outcome of major hepatic resection for small solitary hepatocellular carcinoma in patients with cirrhosis. *Arch Surg*. 2003;138(11):1207–1213. doi:10.1001/archsurg.138.11.1207
47. Dahiya D, Wu TJ, Lee CF, Chan KM, Lee WC, Chen MF. Minor versus major hepatic resection for small hepatocellular carcinoma (HCC) in cirrhotic patients: a 20-year experience. *Surgery*. 2010;147(5):676–685. doi:10.1016/j.surg.2009.10.043

48. Tang SC, Lin KY, Huang TF, et al. Association of primary tumor location with long-term oncological prognosis following hepatectomy for hepatocellular carcinoma: a multicenter propensity score matching analysis. *Eur J Surg Oncol*. 2023;49(7):1234–1241. doi:10.1016/j.ejso.2023.02.001
49. Su K, Shen Q, Tong J, et al. Construction and validation of a nomogram for HBV-related hepatocellular carcinoma: a large, multicenter study. *Ann Hepatol*. 2023;28(4):101109. doi:10.1016/j.aohep.2023.101109
50. Mo Q, Liu Y, Zhou Z, et al. Prognostic Value of Aspartate Transaminase/Alanine Transaminase Ratio in Patients With Hepatitis B Virus-Related Hepatocellular Carcinoma Undergoing Hepatectomy. *Front Oncol*. 2022;12:876900. doi:10.3389/fonc.2022.876900