

Machine Learning-Based Survival Analysis for Patients Receiving Lenvatinib for Unresectable Hepatocellular Carcinoma

Chien-Hung Lu^{1,*}, Ching-Wen Chang^{2,3,*}, San-Chi Chen⁴⁻⁶, Wei-Yu Kao^{1,2,7-9}, Wei-Yi Ting¹⁰, Chia-Hsun Lu¹¹, Kai-I Chuang¹², Cheng-Fu Ni¹², Yao-Yu Hsieh^{13,14}, Ming-Shun Wu^{7,8,15}, Chien-Wei Su^{4,16,17}, Chun-Chao Chang^{1,7,8}, Chih-Horng Wu¹⁸⁻²⁰

¹Division of Gastroenterology and Hepatology, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan; ²Graduate Institute of Metabolism and Obesity Sciences, Taipei Medical University, Taipei, Taiwan; ³Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD, USA; ⁴Division of Medical Oncology, Department of Oncology, Taipei Veterans General Hospital, Taipei, Taiwan; ⁵School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan; ⁶Institute of Clinical Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan; ⁷Division of Gastroenterology and Hepatology, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; ⁸TMU Research Center for Digestive Medicine, Taipei Medical University, Taipei, Taiwan; ⁹Taipei Cancer Center, Taipei Medical University, Taipei, Taiwan; ¹⁰Department of Medical Imaging, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan; ¹¹Department of Medical Imaging, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan; ¹²Department of Medical Imaging, Taipei Medical University Hospital, Taipei, Taiwan; ¹³Division of Hematology and Oncology, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan; ¹⁴Division of Hematology and Oncology, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; ¹⁵Division of Gastroenterology and Hepatology, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan; ¹⁶Division of Gastroenterology and Hepatology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan; ¹⁷Division of General Medicine, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan; ¹⁸Department of Medical Imaging and Radiology, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; ¹⁹Center of Minimal-Invasive Interventional Radiology, National Taiwan University Hospital, Taipei, Taiwan; ²⁰Hepatitis Research Center, National Taiwan University Hospital, Taipei, Taiwan

*These authors contributed equally to this work

Correspondence: Chih-Horng Wu, Department of Medical Imaging and Radiology, National Taiwan University Hospital and College of Medicine, No. 7 Chung-Shan South Road, Taipei City, 100, Taiwan, Tel +886 2 23123456 ext. 62570, Email chw1020@ntuh.gov.tw

Purpose: Lenvatinib is an effective treatment for patients with intermediate- to advanced-stage unresectable hepatocellular carcinoma (HCC). However, tumor response and survival outcomes vary widely. Traditional machine learning (ML) models have been developed to predict treatment response or survival status at discrete time points. However, an overall prediction of overall survival (OS) and progression-free survival (PFS) incorporating censored survival data is lacking. We aimed to conduct a comprehensive survival analysis of OS and PFS by using ML-based survival models.

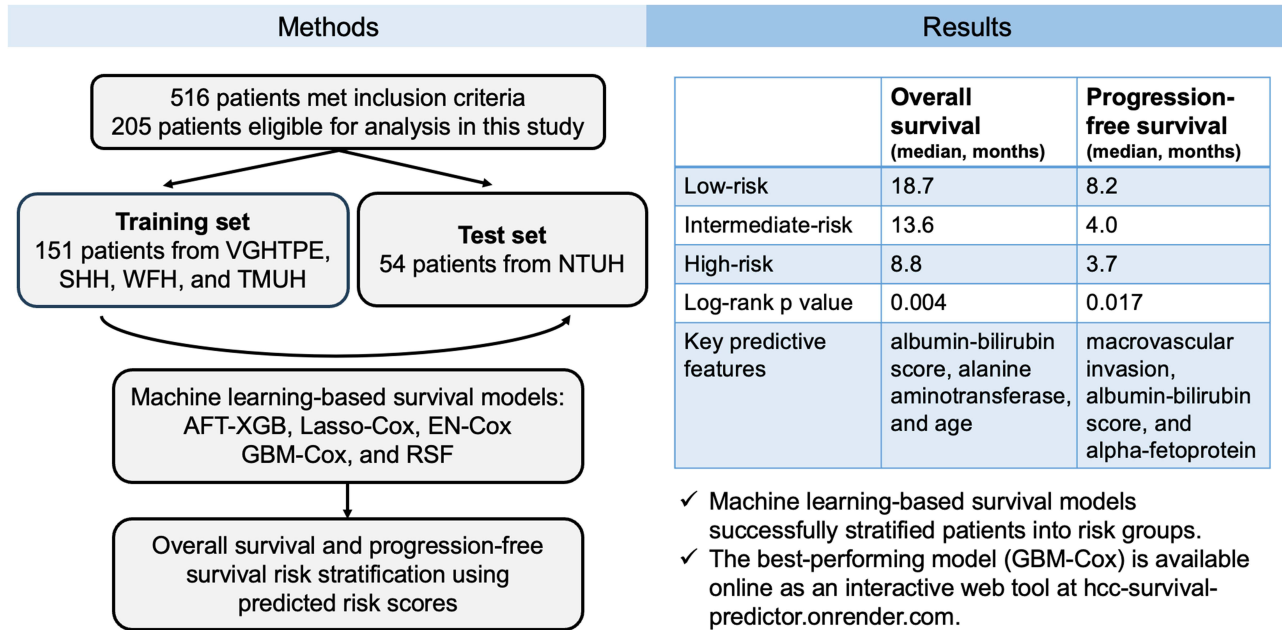
Patients and Methods: This multicenter, retrospective study included patients with unresectable HCC receiving lenvatinib across five healthcare centers. Demographic data, laboratory results, tumor characteristics, and survival outcomes were collected. Five ML-based survival models were developed and compared using Harrell's concordance index (C-index). The predicted risk scores were used to stratify patients into low-, intermediate-, and high-risk groups and validated in the test set.

Results: 205 patients were included for training and validation. Among the five ML models, the GBM-Cox model achieved the highest C-indices for both OS (0.617) and PFS (0.645) prediction. The predicted risk scores stratified the patients into low-, intermediate-, and high-risk groups for OS (median, 18.7 vs 13.6 vs 8.8 months; $p = 0.004$) and PFS (median, 8.2 vs 4.0 vs 3.7; $p = 0.017$). The most influential prognostic factors included albumin-bilirubin (ALBI) score, alanine aminotransferase, and age for OS, and macrovascular invasion, ALBI score, and alpha-fetoprotein for PFS.

Conclusion: ML-based survival models successfully stratified patients into low-, intermediate-, and high-risk groups for OS and PFS. Key features included ALBI score and alanine aminotransferase for OS, and macrovascular invasion and ALBI score for PFS. These models have the potential to guide clinicians' treatment decisions and provide prognostic evaluations. Future prospective studies with larger cohorts, as well as integration of imaging biomarkers are warranted to optimize these predictive models.

Keywords: survival models, cox proportional hazards model, gradient boosting machine, random survival forest, overall survival, progression-free survival

Graphical Abstract



Introduction

Hepatocellular carcinoma (HCC), the predominant form of liver cancer, is a major global health concern and a leading cause of cancer-related mortality worldwide.¹ The past decade has seen advancements in treatment, particularly systemic therapies, which have brought the treatment of HCC into a new era. In 2018, lenvatinib emerged as an effective first-line treatment for unresectable HCC based on the REFLECT trial, which demonstrated its noninferiority over sorafenib in terms of overall survival (OS).² Its effectiveness has also been confirmed in real-world studies.^{3,4} A further analysis of the REFLECT trial reported an objective response rate (ORR) of 18.8% according to the RECIST (Response Evaluation Criteria in Solid Tumors) 1.1 criteria and 40.6% according to the modified RECIST criteria.⁴

Although the advent of immunotherapy has replaced lenvatinib as the first-line treatment for unresectable HCC since 2020,⁵ lenvatinib still serves as an effective treatment for patients who are intolerant or resistant to immunotherapies. Additionally, in patients with metabolic dysfunction-associated steatotic liver diseases, lenvatinib reportedly has comparable efficacy to immunotherapy.⁶ However, the treatment response rate is modest, and survival outcomes vary across patients, which may be related to patient demographics, tumor features, metabolic phenotypes, and immunological characteristics.^{2,4,7,8} It is highly important to identify which patient population is likely to benefit from treatment and to predict patient prognosis, which is essential in guiding treatment plans.

Several studies have explored machine learning (ML) approaches to predict outcomes for HCC patients treated with lenvatinib. Most of them adopt traditional ML algorithms, such as logistic regression, decision trees, or support vector machine, designed for binary or multiclass prediction of treatment response or survival status at specific time points.^{9–12} However, treatment responders do not necessarily guarantee better OS,^{13,14} and an overall evaluation of survival outcomes is usually more important for clinicians than prediction at discrete time points. Furthermore, traditional ML

algorithms require training data with known event status over a defined time period, reflecting the inherent limitations of these methods.¹⁵ Without incorporating censored survival data, their results potentially limited the reliability of survival predictions. Survival models manage censored survival data, among which the Cox proportional hazards (CoxPH) model is the most commonly used. It can be integrated with coefficient regularization methods such as Lasso (least absolute shrinkage and selection operator) or elastic net regularization and serve as effective ML models.^{16,17} Recently, novel ML-based survival models have been developed that adapt ML algorithms, such as random forest and gradient boosting, including the random survival forest (RSF),¹⁸ gradient boosting machine CoxPH (GBM-Cox) model,¹⁹ and accelerated failure time XGBoost (AFT-XGB) model.²⁰ In this study, we aimed to conduct a comprehensive survival analysis of OS and progression-free survival (PFS) for patients with unresectable HCC receiving lenvatinib by using various ML-based survival models and to share our best-performing model publicly as an interactive web-based tool.

Materials and Methods

Study Population and Dataset

This retrospective, multicenter study included patients ≥ 18 years old with intermediate- to advanced-stage HCC treated with lenvatinib at Taipei Veterans General Hospital (VGHTPE), National Taiwan University Hospital (NTUH), Shuang Ho Hospital (SHH), Taipei Municipal Wan Fang Hospital (WFH), and Taipei Medical University Hospital (TMUH) between December 2019 and April 2022. Patients were included in the study if they met the following criteria: (1) a diagnosis of HCC through pathological or radiological assessment with typical imaging patterns; (2) availability of CT or MRI imaging within one month prior to initiating lenvatinib; and (3) treatment with lenvatinib as a first-, second-, or third-line therapy. The exclusion criteria were as follows: (1) concurrent local therapy or immunotherapy during lenvatinib treatment; (2) lenvatinib treatment duration of less than two months; (3) Child–Pugh score of class C; (4) absence of intrahepatic tumors; and (5) prior systemic treatment exceeding three lines. The demographic and clinical variables were collected. Radiological assessments were conducted every 2 to 3 months via contrast-enhanced CT or MRI scans. OS was defined as the time from the start of lenvatinib treatment to either death or the most recent follow-up. PFS was defined as the time from the start of lenvatinib treatment to the occurrence of disease progression, death, or the most recent follow-up. The final follow-up date was December 31, 2022.

This study complied with the principles of the Declaration of Helsinki, institutional guidelines, the Medical Care Act, and the Personal Data Protection Act of Taiwan. The institutional review boards of the participating hospitals approved the study (VGHTPE: 2023–09–012BC; NTUH: 202405146RINE; and TMU-JIRB: N202308023).

Study Variables

The variables that were used as input features included age, sex, hepatitis etiology, Barcelona Clinic Liver Cancer (BCLC) stage, alpha-fetoprotein (AFP), creatinine, albumin–bilirubin (ALBI) score, alanine aminotransferase (ALT), platelet count, tumor radiologic burden on the basis of the Up-to-7 criteria,²¹ presence of main portal vein thrombosis, macrovascular invasion, metastasis, and prior systemic treatment. Body weight, body height, and Child–Pugh class were not included as input variables because their percentage of missing values was $\geq 10\%$. For those input variables with missing values of $< 10\%$, imputation was performed using the mode for categorical variables and the average for continuous variables. Continuous variables were standardized during data preprocessing. Treatment response was evaluated based on the RECIST 1.1 criteria. The ORR was calculated as the rate of complete response or partial response.

Sample Size Justification

Given the expected limited sample size, an a priori sample size justification was conducted to evaluate statistical power and the risk of overfitting. The Riley method was applied using the `pmsamplesize` package in R (version 4.4.2).²² As no previous studies have utilized ML-based survival models in patients receiving lenvatinib for unresectable HCC, a previously published CoxPH model for OS prediction was used as a reference. The study reported an AUROC of 0.80 with seven predictors among 351 enrolled patients.²³ The analysis indicated that a minimum sample size of 490 would be

required when the shrinkage coefficient is set to 0.9, and 216 when set to 0.8, with the latter associated with an increased risk of overfitting. To mitigate such risks, regularized (Lasso and elastic net) and tree-based ML algorithms were employed to reduce the influence of weakly predictive variables.^{16–18,24}

Survival Analysis via Machine Learning

The entire dataset was collected under the same enrollment protocol to ensure consistent baseline characteristics across centers and divided into a training set and a test set based on the volume of the healthcare centers involved. Patient data from VGHTPE, SHH, WFH, and TMUH (74% of the dataset) were used for model training and internal validation, whereas those from NTUH comprised the test set (26% of the dataset) and served as external validation. The dataset collection and splitting process is illustrated in Figure 1.

We developed and validated five ML survival models, including the RSF, Lasso-regularized CoxPH (Lasso-Cox), elastic net-regularized CoxPH (EN-Cox), GBM-Cox, and AFT-XGB models. A conventional CoxPH model was developed for comparison. RSF is an ensemble learning and nonparametric method that extends from the random forest algorithm to manage right-censored survival data.^{18,25} It is constructed via an ensemble of survival trees and estimation of the cumulative hazard function. The split of survival trees is judged by log-rank statistics. After training the survival trees, RSF aggregates the results via ensemble methods. Lasso regularization, or the L1 penalty, effectively performs feature selection by penalizing features with lower predictive ability, leading to some coefficients being shrunk toward zero.^{17,26} Elastic net regularization, which combines L1 and L2 (Ridge) penalties, effectively handles situations with highly correlated features and balances between feature selection and coefficient shrinkage.¹⁶ The GBM-Cox model combines GBM, an ensemble learning tree-based method, with the CoxPH model by using the negative log partial likelihood as its cost function.²⁷ The AFT model, in contrast to the proportional hazards model, which estimates covariates' multiplicative effects on the hazard function, directly models the time to event by assessing how covariates accelerate or decelerate survival time.²⁸ XGBoost, a widely used and efficient tree-based ML algorithm, has been adapted to integrate with the AFT model to build an effective survival model (AFT-XGB).²⁰ The training process involved 5-fold

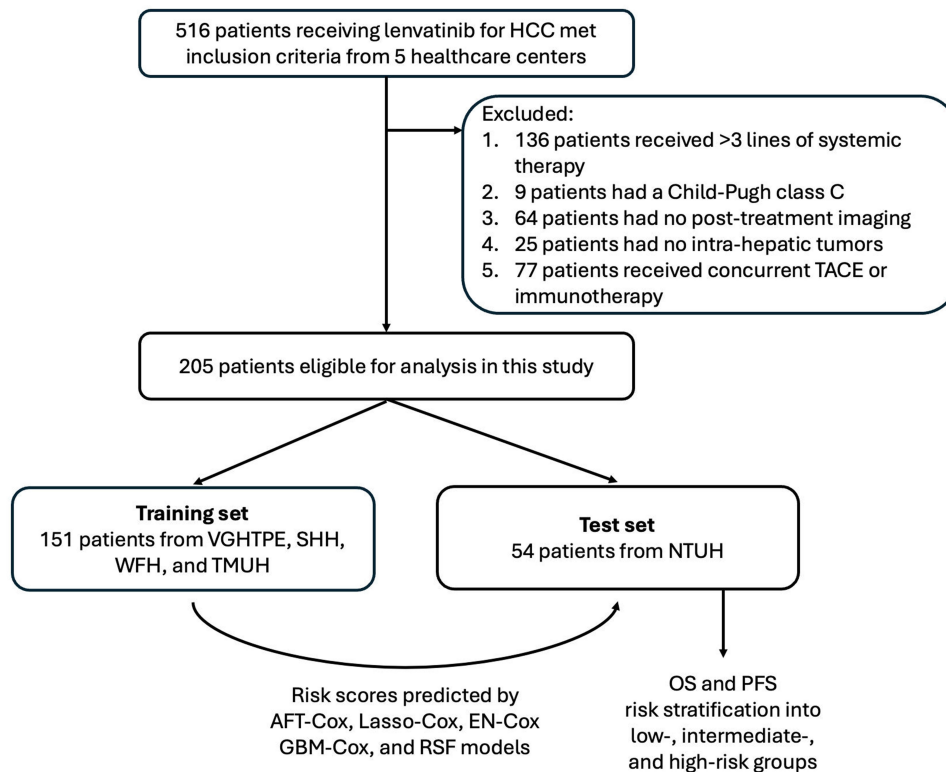


Figure 1 Data collection and splitting process for model training.

cross-validation within the training set for hyperparameter tuning ([Supplementary Table 1](#)). The best parameters were used to train the final models. The models were constructed via Python (version 3.11) with the xgboost, scikit-survival, scikit-learn, and lifelines packages.

Model Performance Evaluation and Comparison

The predictive performance of the ML models was evaluated using the Harrell's concordance index (C-index), integrated time-dependent area under the receiver operating characteristic curve (iAUC), and integrated Brier score (IBS). The C-index measures the concordance between the model's predicted risk scores and the observed survival outcomes across all comparable subject pairs,²⁹ and serves as our primary performance metric. The time-dependent area under the receiver operating characteristic curve (AUROC) quantifies the probability that, at a specific time point, a randomly selected patient who experiences the event by that time has a higher predicted risk score than a randomly selected patient who has not experienced the event by that time.^{29–31} The Brier score quantifies the accuracy of survival probability predictions by measuring the mean squared error between the predicted survival probability and the actual survival status at specific time points.²⁹ The iAUC and IBS were calculated as weighted averages of time-dependent AUROC values and Brier scores over time, respectively. To implement risk stratification and provide clinical granularity, the training set risk scores predicted by the best-performing ML model (with the highest C-index) were grouped into tertiles by selecting cutoffs that optimized the overall log-rank p-value of the survival curves across multiple percentile combinations in the training set. These cutoffs were then applied to the test set. SHAP (SHapley Additive exPlanations) values were calculated for feature importance analysis.³²

Survival Status Prediction at Discrete Time Points

We also applied the best-performing ML model with the highest C-index to predict survival status at discrete time points: from 6 to 36 months after starting lenvatinib therapy at 6-month intervals for OS prediction, and from 6 to 18 months for PFS prediction. Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and both binary and time-dependent AUROC were calculated. The optimal AUROC threshold at each time point was determined using the Youden index, which maximizes the sum of sensitivity and specificity.³³

Statistical Analysis

The statistical comparison of demographic data between patients in the training and test sets was performed using Student's *t*-test or the Mann–Whitney *U*-test for continuous variables. For categorical variables, the Pearson chi-square test or Fisher's exact test was used. The Kaplan–Meier method was used to estimate the survival curves via ML models-derived risk scores, and a Log rank test was used for comparison. All the statistical analyses were performed using SPSS (version 27.0.1.0; IBM, Armonk, NY, USA) and Python (version 3.11).

Results

Patient Baseline Characteristics

A total of 205 patients were included in this study, with 151 patients from four healthcare centers in the training set, and 54 patients from one large-volume healthcare center in the test set. The detailed baseline characteristics are summarized in [Table 1](#). The mean age of the entire cohort was 67.4 years, and the majority of patients (75.1%) were male. Hepatitis B virus infection was the most common underlying liver disease, affecting 62.9% of the participants. The disease was classified as BCLC stage C in 80.5% of the patients, indicating significantly more advanced-stage HCC at the time of treatment initiation. In terms of tumor characteristics, 22.9% of patients presented with main portal vein thrombosis, 57.1% presented with major vascular invasion, and 44.9% had evidence of metastatic disease. According to ALBI grading, 36.1% of patients were categorized as ALBI grade 1, whereas 60.0% and 2.4% were classified as grades 2 and 3, respectively. Additionally, 78.5% of patients had a tumor radiologic burden exceeding the Up-to-7 criteria, and 41.0% had received prior systemic therapy. Comparisons of patient characteristics between the training and test sets revealed no significant differences in demographics, laboratory tests, or radiological assessments.

Table 1 Baseline Characteristics of the Entire Cohort, Training Set, and Test Set

Variable	All Cohort (N = 205)	Training (N=151)	Test (N=54)	p value
Mean age (SD), years	67.4±12.2	66.4±12.5	70.1±11.1	0.06
Sex, male	154 (75.1)	112 (74.2)	42 (77.8)	0.60
HBsAg, positive	129 (62.9)	93 (61.6)	36 (66.7)	0.51
Anti-HCV, positive	35 (17.1)	22 (14.6)	13 (24.1)	0.11
Prior systemic therapy, yes	84 (41.0)	57 (37.7)	27 (50.0)	0.12
AFP, ng/mL	21075.2±147,421.3	27,292.6±172,257.6	4380.3±13,360.3	0.33
Cre, mg/dl	1.0±0.7	1.0±0.6	1.2±0.9	0.19
PLT, ×10 ⁹ /L	198.2±119.2	202.9±126.7	185.2±94.8	0.35
ALT, U/L	54.3±68.6	56.3±70.7	49.1±62.4	0.51
Metastasis, positive	92 (44.9)	70 (46.4)	22 (40.7)	0.48
MVI, positive	117 (57.1)	89 (58.9)	28 (51.9)	0.37
Vp4, positive	47 (22.9)	38 (25.2)	19 (16.7)	0.20
BCLC stage, B	40 (19.5)	29 (19.2)	11 (20.4)	0.85
BCLC stage, C	165 (80.5)	122 (80.8)	43 (79.6)	
ALBI grade, 1	74 (36.1)	50 (33.8)	24 (44.4)	0.38
ALBI grade, 2	123 (60.0)	94 (62.3)	29 (53.7)	
ALBI grade, 3	5 (2.4)	4 (2.6)	1 (1.9)	
ALBI score	-2.4±0.5	-2.4±0.5	-2.5±0.5	0.11
Up-to-7 criteria, in	161 (78.5)	123 (81.5)	38 (70.3)	0.09
Up-to-7 criteria, out	44 (21.5)	28 (18.5)	16 (29.6)	
Median OS	12.2	11.5	13.4	0.76
Median PFS	7.3	7.6	6.6	0.37

Notes: The data are presented as the means ± SDs or N values (% or number).

Abbreviations: SD, standard deviation; HBsAg, hepatitis B surface antigen; Anti-HCV, anti-hepatitis C antibody; AFP, alpha-fetoprotein; Cre, creatinine; PLT, platelet; ALT, alanine transaminase; MVI, major vascular invasion; Vp4, main portal vein thrombosis; BCLC, Barcelona Clinic Liver Cancer; ALBI, albumin–bilirubin; OS, overall survival; PFS, progression-free survival.

Observed Survival Outcomes

The treatment response, evaluated via the RECIST 1.1 criteria, demonstrated an ORR of 13.2% for the entire cohort. The median OS of the entire cohort was 12.2 months, whereas the median PFS was 7.3 months. No significant differences were observed between the training and test sets for OS (11.5 vs 13.4 months, $p = 0.76$) or PFS (7.6 vs 6.6 months, $p = 0.37$). During the follow-up period, 143 patients experienced disease progression or death, among whom 130 patients died.

Model Performance Evaluation and Comparison

For OS prediction, all the ML-based survival models outperformed the CoxPH model, with the GBM-Cox model achieving the highest C-index of 0.617 and iAUC of 0.732. The Lasso-Cox model achieved the lowest IBS of 0.197. For PFS prediction, the GBM-Cox model also achieved the highest C-index of 0.645 and iAUC of 0.709. The Lasso-Cox model achieved the lowest IBS of 0.203. The model performances are detailed in Table 2.

Risk Stratification for OS and PFS

Using the predicted risk score threshold derived from the training set, survival curves for OS risk stratification into low-, intermediate-, and high-risk groups differed significantly in both the training ($p < 0.001$) and test sets ($p = 0.004$). The median OS of the low-, intermediate-, and high-risk groups in the test set were 18.7, 13.6, and 8.8 months, respectively (Figure 2). For PFS risk stratification, the survival curves also significantly differed when patients were stratified into low-, intermediate-, and high-risk groups in both the training ($p < 0.001$) and test sets ($p = 0.017$). The median PFS of the low-, intermediate-, and high-risk groups in the test set were 8.2, 4.0, and 3.7 months, respectively (Figure 3). This ML-based risk-stratification tool has been publicly available as an interactive application at <https://hcc-survival-predictor.onrender.com>.

Table 2 Comparison of C-Index, iAUC, and IBS Among the Machine Learning Models

	C-index	iAUC	IBS
OS prediction			
CoxPH	0.565	0.695	0.216
AFT-XGB	0.588	0.678	0.443
Lasso-Cox	0.597	0.717	0.197
EN-Cox	0.566	0.699	0.319
GBM-Cox	0.617	0.732	0.332
RSF	0.608	0.704	0.317
PFS prediction			
CoxPH	0.591	0.644	0.210
AFT-XGB	0.594	0.647	0.333
Lasso-Cox	0.595	0.649	0.203
EN-Cox	0.588	0.644	0.299
GBM-Cox	0.645	0.709	0.294
RSF	0.610	0.666	0.285

Abbreviations: iAUC, integrated time-dependent area under the receiver operating characteristic curve; IBS, integrated Brier score; CoxPH, Cox proportional hazards model; AFT-XGB, accelerated failure time-XGBoost model; Lasso-Cox, Lasso-regularized Cox proportional hazards model; EN-Cox, elastic net-regularized Cox proportional hazards model; GBM-Cox, gradient-boosting machine Cox proportional hazards model; RSF, random survival forest; OS, overall survival; PFS, progression-free survival.

Survival Status Prediction at Discrete Time Points

The survival status prediction performance of the GBM-Cox model at discrete time points is detailed in Table 3. For OS prediction, it achieved binary AUROCs ranging from 0.59 to 0.93 and time-dependent AUROCs from 0.58 to 0.94 with the highest values (0.75 to 0.94) observed at 18 to 36 months from the start of lenvatinib therapy. For PFS prediction, it achieved binary AUROCs ranging from 0.63 to 0.79 and time-dependent AUROCs from 0.57 to 0.78 at 6 to 18 months.

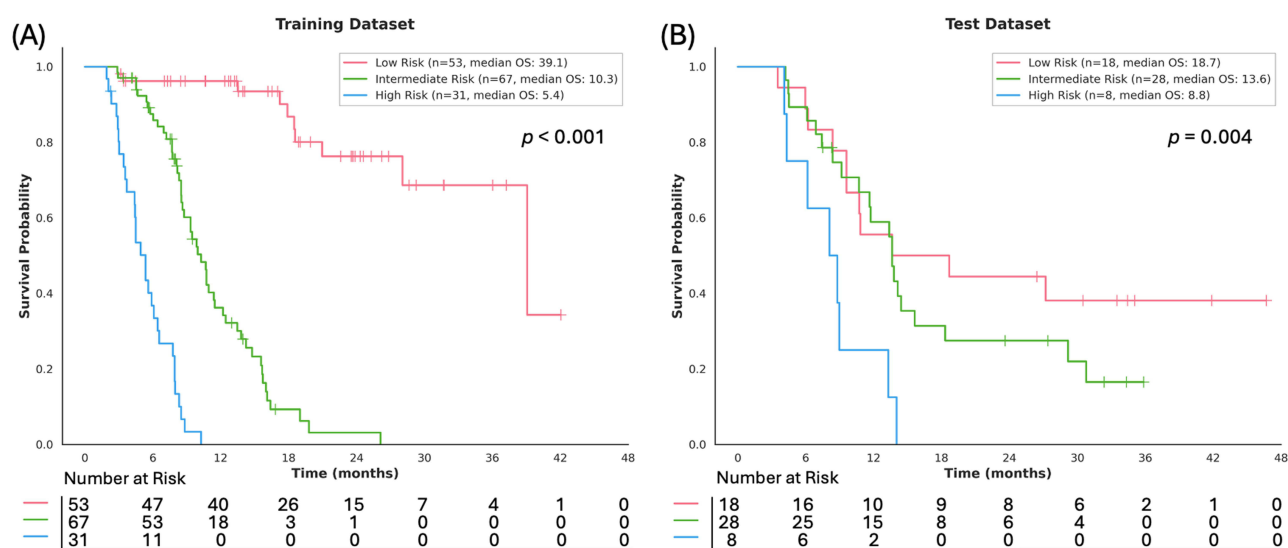


Figure 2 OS survival curves of the low-, intermediate-, and high-risk groups of the (A) training set and (B) test set.
Abbreviation: OS, overall survival.

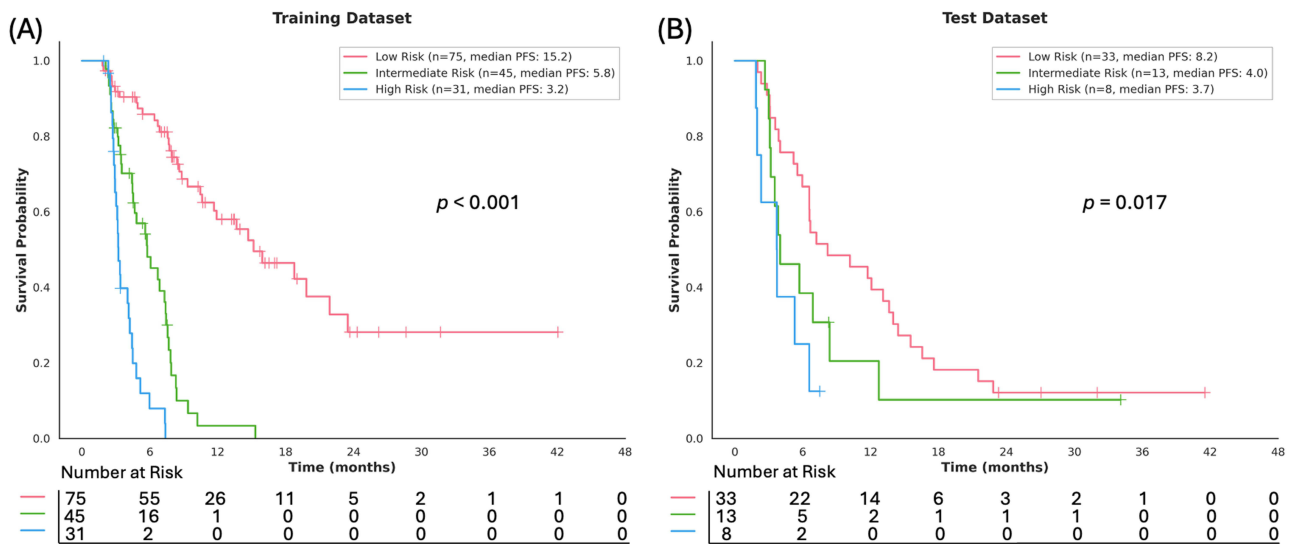


Figure 3 PFS survival curves of the low-, intermediate-, and high-risk groups of the (A) training set and (B) test set. **Abbreviation:** PFS, progression-free survival.

Features of Importance Analysis

To assess the predictive importance of features, we calculated SHAP values for the best-performing models. For OS prediction, the five features with the highest SHAP values for the GBM-Cox model were the ALBI score, ALT level, age, creatinine level, and presence of metastasis (Figure 4A). For PFS prediction, the top five features for the GBM-Cox model were presence of macrovascular invasion, the ALBI score, AFP level, ALT level, and creatinine level (Supplementary Figure 1A). Heatmaps of the feature SHAP values for the test set patients are shown in Figure 4B and Supplementary Figure 1B for OS and PFS, respectively.

Table 3 Performance Metrics of the GBM-Cox Model for Survival Status Prediction in the Test Set (N=54) at Specific Time Points

Time (Months)	Sen (%)	Spe (%)	PPV (%)	NPV (%)	Acc (%)	Binary AUROC	Time-Dependent AUROC
Overall survival							
6	83.0	57.1	92.9	33.3	79.6	0.62	0.62
12	88.9	36.0	60.0	75.0	63.5	0.59	0.58
18	94.1	42.9	44.4	93.8	59.6	0.75	0.75
24	42.9	97.3	85.7	81.8	82.4	0.76	0.76
30	50.0	97.4	83.3	88.4	87.8	0.78	0.78
36	100.0	85.0	25.0	100.0	85.7	0.93	0.94
Progression-free survival							
6	79.3	56.0	67.7	70.0	68.5	0.71	0.70
12	81.3	69.4	54.2	89.3	73.1	0.79	0.78
18	85.7	48.9	20.7	95.7	53.9	0.63	0.57

Notes: Survivals and non-progression status are labeled as positive.
Abbreviations: Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value; Acc, accuracy; AUROC, area under the receiver operating characteristic curve.

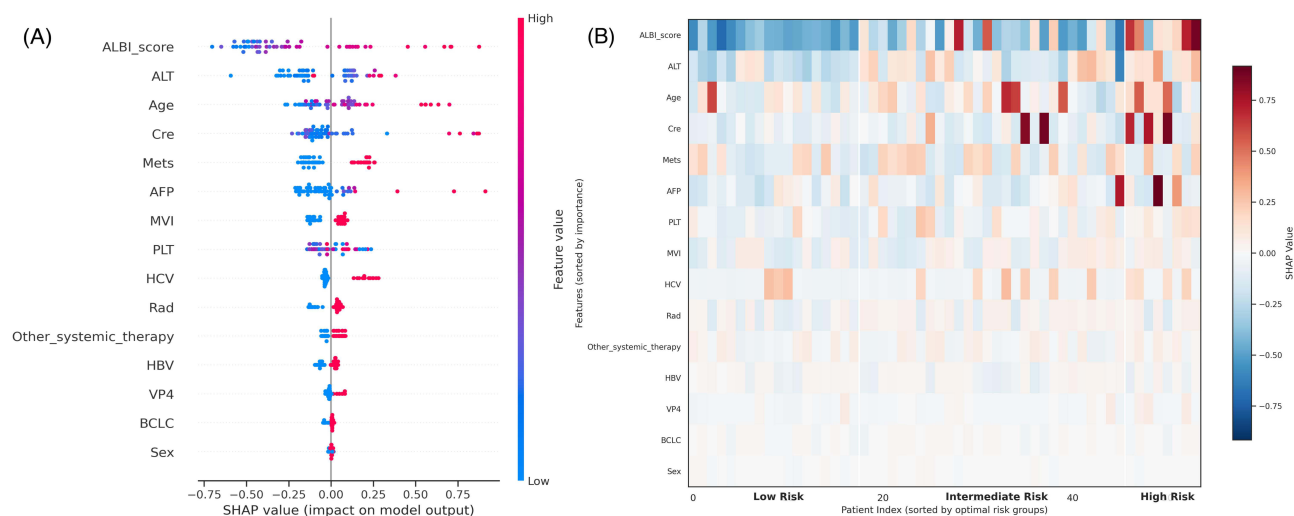


Figure 4 Feature importance analysis based on SHAP values for the GBM-Cox model in OS prediction. **(A)** Beeswarm plot of SHAP values. **(B)** Heatmap of SHAP values for patients in the test set. Each point or panel represents the SHAP value of a feature to the predicted risk for a patient.

Abbreviations: ALBI, albumin-bilirubin; ALT, alanine aminotransferase; Cre, creatinine; Mets, metastasis; AFP, alpha-fetoprotein; PLT, platelet; HCV, hepatitis C; Rad, radiologic burden evaluated by the Up-to-7 criteria; HBV, hepatitis B; VP4, main portal vein thrombosis; BCLC, Barcelona Clinic Liver Cancer stage.

Discussion

The present study conducted a comprehensive survival analysis using five ML-based survival models in patients receiving lenvatinib for unresectable HCC. Among these models, the GBM-Cox model achieved the best performance, with the highest C-indices and iAUCs for both OS and PFS prediction. Based on the risk scores predicted by the GBM-Cox model, patients were stratified into low-, intermediate-, and high-risk groups.

Lenvatinib marked a new era of treatment, which showed noninferiority to sorafenib in patients with HCC in a Phase 3 clinical trial in 2018.² However, the treatment response and OS are still modest, with a real-world median OS of approximately 11.4 months and an ORR of approximately 36.0% according to the mRECIST criteria.^{3,4,34} Although treatment response is strongly associated with OS, other factors, such as preserved liver function, AFP level, and tumor size, have also been identified as independent prognostic factors.^{3,35} Predicting OS in patients with HCC treated with lenvatinib is important because it reflects the overall evaluation of both treatment effects and nontreatment effects. It can guide clinicians' treatment plans with early shifts to or the addition of other therapies if a poor prognosis is anticipated.

Previous studies have investigated the use of ML algorithms for predicting OS and PFS in patients with HCC. However, owing to the relatively new advent of lenvatinib, many of them do not include patients treated with lenvatinib.^{7,36–38} Some studies have used ML algorithms to predict treatment response in patients with HCC treated with lenvatinib. Bo et al achieved excellent performance in predicting treatment response to lenvatinib.⁹ However, by applying unsupervised ML algorithms, they reported that two radiomics subtypes were associated with different PFS but not OS. Ma et al used clinical data to predict treatment response to lenvatinib combined with transarterial chemoembolization, achieving a high AUROC of 0.91 via a random forest model.¹¹ Although the predictive accuracy is high, their study did not reveal OS or PFS predictions. Hua et al used clinical data and radiomics features to predict treatment response to lenvatinib plus PD-1 inhibitors and interventional therapy, and stratified patients into high- and low-risk groups for OS and PFS.¹² Other studies have used ML models to predict survival status at discrete time points.^{39–41} Han et al used XGBoost to predict survival status at specific time points and stratified patients into risk groups based on the predicted survival probabilities.⁴⁰ Simsek et al applied a LightGBM model to predict survival status at specific time points but did not perform an overall prognostic evaluation.³⁹ To the best of our knowledge, this is the first study to apply ML-based survival models in patients with unresectable HCC treated with lenvatinib. By applying ML-based survival models, censored survival data can be leveraged, and the predicted risk scores can be used to stratify patients into risk groups for overall OS and PFS evaluation.

Among the ML models, the GBM-Cox model exhibited the highest C-indices and iAUCs for both OS and PFS prediction, while the Lasso-Cox model had the lowest IBS. These findings suggest that the GBM-Cox model has the best discriminative ability for risk prediction, but is less effective than the Lasso-Cox model in estimating survival probabilities. The GBM-Cox model also demonstrated robust prediction of survival status at discrete time points, particularly between 18 and 36 months after initiating lenvatinib therapy (AUROCs: 0.75–0.94) for OS and between 6 and 12 months (AUROCs: 0.70–0.79) for PFS. Although our study aimed to perform an overall evaluation of OS and PFS, these results also demonstrate the ability of ML-based survival models to predict survival status at discrete time points.

By investigating SHAP values, feature importance analysis identified key prognostic factors. For OS prediction with the GBM-Cox model, the five features with the highest SHAP values were the ALBI score, ALT level, age, creatinine level, and presence of metastasis. For PFS prediction, the most influential features were presence of macrovascular invasion, the ALBI score, AFP, ALT, and creatinine levels. Importantly, the identified prognostic factors, such as vascular invasion, tumor size, and metastasis, have been reported to be associated with gene expression profiles involved in the development and prognosis of HCC.^{42,43} The ALBI score, calculated from serum albumin and total bilirubin levels, has been validated as an independent prognostic factor for OS in patients with HCC or chronic liver disease.^{44,45} A higher ALBI score is associated with poorer outcomes and may provide better prognostic value than the Child-Pugh class.⁴⁵ Macrovascular invasion has also been found to be associated with shorter PFS or OS.^{3,46} However, studies have also shown inconsistent results on prognostic factors in patients with HCC treated with lenvatinib.^{3,35,47–50} For example, Welland et al and Kudo et al reported that AFP \geq 200 ng/mL was an independent negative prognostic factor for OS,^{3,35} whereas Hiraoka et al reported that only the ALBI grade was independently prognostic.⁴⁷ Our models reinforce the prognostic significance of these key variables and may offer insights into potential nonlinear effects among variables that were not identified previously using traditional statistical approaches.⁵¹

This study has several limitations. First, the sample size was relatively small. Although regularized and tree-based ML algorithms were employed to mitigate the effects of less predictive variables and reduce the risk of overfitting, the limited sample size may still constrain statistical power and affect the model robustness. However, compared with most studies that applied ML models to patients treated with lenvatinib, our sample size was slightly larger. Future studies with larger, prospectively collected datasets are warranted. Second, our input features did not include radiomic features or raw images from pretreatment CT or MRI scans, which may have limited the model's predictive performance. Incorporating such data could potentially enhance the predictive power of the model. However, radiomic features are often not standardized, and including numerous radiomic features in a relatively small dataset increases the risk of overfitting and unreliable predictions.^{52–54} Third, body weight, body height, and Child–Pugh class were excluded from the training variables because each had more than 10% missing data, potentially introducing bias. Future studies incorporating these factors may better mitigate this limitation. Finally, although we employed five commonly used and efficient ML-based survival models, we did not include all available models. Other models may have the potential to achieve better predictive performance.

Conclusion

Our study demonstrated that ML-based survival models effectively stratified patients into low-, intermediate-, and high-risk groups for OS and PFS prediction by using the predicted risk scores for patients with HCC treated with lenvatinib. Unlike traditional ML algorithms, our models manage censored survival data and provide an overall evaluation instead of survival status prediction at discrete time points. Among the models, the GBM-Cox model is the best with the highest C-indices and iAUCs. By using baseline patient demographics, laboratory tests, and tumor characteristics, these ML-based survival models enable effective risk stratification for OS and PFS, offering prognostic insights that may aid clinicians in treatment planning and improve clinical decision-making. Future research should validate these findings in larger prospective cohort studies and explore integration with imaging biomarkers to optimize predictive models.

Abbreviations

AFP, alpha-fetoprotein; AFT, accelerated failure time; ALBI, albumin–bilirubin; ALT, alanine aminotransferase; Anti-HCV, anti-hepatitis C antibody; AUROC, area under the receiver operating characteristic; BCLC, Barcelona Clinic Liver Cancer; C-index, concordance index; CoxPH, Cox proportional hazards model; Cre, creatinine; CT, computed tomography;

EN, elastic net; GBM, gradient boosting machine; HBsAg, hepatitis B surface antigen; Lasso, least absolute shrinkage and selection operator; HCC, hepatocellular carcinoma; iAUC, integrated time-dependent area under the receiver operating characteristic; ML, machine learning; mRECIST, modified Response Evaluation Criteria in Solid Tumors; MRI, magnetic resonance imaging; NPV, negative predictive value; TACE, transarterial chemoembolization; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PLT, platelet; PPV, positive predictive value; RECIST, Response Evaluation Criteria in Solid Tumors; RSF, random survival forest; SD, standard deviation.

Data Sharing Statement

The GBM-Cox model has been deployed as a publicly available interactive web-tool (<https://hcc-survival-predictor.onrender.com>). Access to the patient data underlying this study's findings requires approval from the Institutional Review Boards. For additional information, please reach out to the corresponding author.

Ethics Approval and Informed Consent

The study received approval from the institutional review boards and was conducted in compliance with the Declaration of Helsinki (approval numbers: 2023-09-012BC for Taipei Veterans General Hospital, 201805070RIND for National Taiwan University Hospital, and N202308023 for the Taipei Medical University-affiliated hospitals, including Shuang Ho Hospital, Wan Fang Hospital, and Taipei Medical University Hospital). All procedures adhered to the applicable guidelines and regulations. The requirement for informed consent was waived by the institutional review boards of Taipei Veterans General Hospital, National Taiwan University Hospital, Shuang Ho Hospital, Wan Fang Hospital, and Taipei Medical University Hospital due to the retrospective nature of the study. All data were handled in accordance with institutional guidelines, the Medical Care Act, and the Personal Data Protection Act of Taiwan to ensure participant privacy and confidentiality. Patient information was anonymized and securely maintained throughout the study. The participants were not provided with study-related information.

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

This study was funded by grants from the Ministry of Health and Welfare (MOHW 114-TDU-B-221-144003 and MOHW 113-TDU-B-221-134003), Taipei Veterans General Hospital (VGHTPE V113C-148), the Taiwan Clinical Oncology Research Foundation awarded to San-Chi Chen, and National Taiwan University Hospital (107-PC1204 and 113-A170).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet*. 2018;391(10126):1163–1173. doi:10.1016/s0140-6736(18)30207-1
3. Welland S, Leyh C, Finkelmeier F, et al. Real-world data for lenvatinib in hepatocellular carcinoma (ELEVATOR): a retrospective multicenter study. *Liver Cancer*. 2022;11(3):219–232. doi:10.1159/000521746
4. Finn RS, Qin S, Piscaglia F, et al. Characterization of tumor responses in patients with unresectable hepatocellular carcinoma treated with lenvatinib in the phase 3 randomized trial: REFLECT. *Liver Cancer*. 2024;13(5):537–547. doi:10.1159/000537947
5. Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med*. 2020;382(20):1894–1905. doi:10.1056/NEJMoa1915745

6. Ahn JC, Ng WH, Yeo YH, et al. Comparative effectiveness of immunotherapy versus lenvatinib in advanced hepatocellular carcinoma: a real-world analysis using target trial emulation. *Hepatology*. 2025. doi:10.1097/hep.0000000000001328
7. Zhang Q, Yu X, Zheng Q, He Y, Guo W. A molecular subtype model for liver HBV-related hepatocellular carcinoma patients based on immune-related genes. *Front Oncol*. 2020;10:560229. doi:10.3389/fonc.2020.560229
8. Xue C, Gu X, Zhao Y, et al. Prediction of hepatocellular carcinoma prognosis and immunotherapeutic effects based on tryptophan metabolism-related genes. *Cancer Cell Int*. 2022;22(1):308. doi:10.1186/s12935-022-02730-8
9. Bo Z, Chen B, Zhao Z, et al. Prediction of response to lenvatinib monotherapy for unresectable hepatocellular carcinoma by machine learning radiomics: a multicenter cohort study. *Clin Cancer Res*. 2023;29(9):1730–1740. doi:10.1158/1078-0432.Ccr-22-2784
10. Dong W, Ji Y, Pi S, Chen Q-F. Noninvasive imaging-based machine learning algorithm to identify progressive disease in advanced hepatocellular carcinoma receiving second-line systemic therapy. *Sci Rep*. 2023;13(1):10690. doi:10.1038/s41598-023-37862-y
11. Ma J, Bo Z, Zhao Z, et al. Machine learning to predict the response to lenvatinib combined with transarterial chemoembolization for unresectable hepatocellular carcinoma. *Cancers*. 2023;15(3):625. doi:10.3390/cancers15030625
12. Hua Y, Sun Z, Xiao Y, et al. Pretreatment CT-based machine learning radiomics model predicts response in unresectable hepatocellular carcinoma treated with lenvatinib plus PD-1 inhibitors and interventional therapy. *J Immunother Cancer*. 2024;12(7):e008953. doi:10.1136/jitc-2024-008953
13. Kok P-S, Cho D, Yoon W-H, et al. Validation of progression-free survival rate at 6 months and objective response for estimating overall survival in immune checkpoint inhibitor trials: a systematic review and meta-analysis. *JAMA Network Open*. 2020;3(9):e2011809–e2011809. doi:10.1001/jamanetworkopen.2020.11809
14. Cammarota A, Zanuso V, Pession T, Personeni N, Rimassa L. Assessment and monitoring of response to systemic treatment in advanced hepatocellular carcinoma: current insights. *J Hepatocell Carcinoma*. 2022;9:1011–1027. doi:10.2147/jhc.S268293
15. Vock DM, Wolfson J, Bandyopadhyay S, et al. Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. *J Biomed Inform*. 2016;61:119–131. doi:10.1016/j.jbi.2016.03.009
16. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc Series B*. 2005;67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x
17. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc*. 2018;58(1):267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
18. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841–860,20.
19. Pölsterl S. scikit-survival: a library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res*. 2020;21(1):Article212.
20. Barnwal A, Cho H, Hocking T. Survival regression with accelerated failure time model in XGBoost. *J Comput Graph Stat*. 2022;31(4):1292–1302. doi:10.1080/10618600.2022.2067548
21. Mazzaferro V, Llovet JM, Miceli R, et al. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol*. 2009;10(1):35–43. doi:10.1016/s1470-2045(08)70284-5
22. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276–1296. doi:10.1002/sim.7992
23. Ding X, Li X, Liu M, Wang J, Li W, Chen J. Development of a multivariate prognostic model for lenvatinib treatment in hepatocellular carcinoma. *Oncologist*. 2023;28(10):e942–e949. doi:10.1093/oncolo/oyad107
24. Friedrich S, Groll A, Ickstadt K, et al. Regularization approaches in clinical biostatistics: a review of methods and their applications. *Stat Methods Med Res*. 2023;32(2):425–440. doi:10.1177/09622802221133557
25. Wang H, Li G. A selective review on random survival forests for high dimensional data. *Quant Biosci*. 2017;36(2):85–96. doi:10.22283/qbs.2017.36.2.85
26. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J Royal Stat Soc*. 2011;73(3):273–282. doi:10.1111/j.1467-9868.2011.00771.x
27. Chen Y, Jia Z, Mercola D, Xie X. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Comput Math Methods Med*. 2013;2013:873595. doi:10.1155/2013/873595
28. Crowther MJ, Royston P, Clements M. A flexible parametric accelerated failure time model and the extension to time-dependent acceleration factors. *Biostatistics*. 2023;24(3):811–831. doi:10.1093/biostatistics/kxac009
29. Park SY, Park JE, Kim H, Park SH. Review of statistical methods for evaluating the performance of survival or other time-to-event prediction models (from Conventional to Deep Learning Approaches). *Korean J Radiol*. 2021;22(10):1697–1707. doi:10.3348/kjr.2021.0223
30. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17–18):2529–2545. doi:10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5
31. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61(1):92–105. doi:10.1111/j.0006-341X.2005.030814.x
32. Ponce-Bobadilla AV, Schmitt V, Maier CS, Stodtmann S, Mensing S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. *Clin Transl Sci*. 2024;17(11):e70056. doi:10.1111/cts.70056
33. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cnrc2820030106>3.0.co;2-3
34. Wang S, Wang Y, Yu J, Wu H, Zhou Y. Lenvatinib as first-line treatment for unresectable hepatocellular carcinoma: a systematic review and meta-analysis. *Cancers*. 2022;14(22). doi:10.3390/cancers14225525
35. Kudo M, Finn RS, Qin S, et al. Overall survival and objective response in advanced unresectable hepatocellular carcinoma: a subanalysis of the REFLECT study. *J Hepatol*. 2023;78(1):133–141. doi:10.1016/j.jhep.2022.09.006
36. Lin H, Zeng L, Yang J, Hu W, Zhu Y. A machine learning-based model to predict survival after transarterial chemoembolization for BCLC stage B hepatocellular carcinoma. *Front Oncol*. 2021;11:608260. doi:10.3389/fonc.2021.608260
37. Li X, Bao H, Shi Y, et al. Machine learning methods for accurately predicting survival and guiding treatment in stage I and II hepatocellular carcinoma. *Medicine*. 2023;102(45):e35892. doi:10.1097/md.00000000000035892
38. Zhang M, Kuang B, Zhang J, et al. Enhancing prognostic prediction in hepatocellular carcinoma post-TACE: a machine learning approach integrating radiomics and clinical features. *Front Med Lausanne*. 2024;11:1419058. doi:10.3389/fmed.2024.1419058
39. Simsek C, Can Guven D, Koray Sahin T, et al. Artificial intelligence method to predict overall survival of hepatocellular carcinoma. *Hepatol Forum*. 2021;2(2):64–68. doi:10.14744/hf.2021.2021.0017

40. Han JW, Lee SK, Kwon JH, et al. A machine learning algorithm facilitates prognosis prediction and treatment selection for barcelona clinic liver cancer stage C hepatocellular carcinoma. *Clin Cancer Res.* 2024;30(13):2812–2821. doi:10.1158/1078-0432.Ccr-23-3978
41. Seven İ, Bayram D, Arslan H, et al. Predicting hepatocellular carcinoma survival with artificial intelligence. *Sci Rep.* 2025;15(1):6226. doi:10.1038/s41598-025-90884-6
42. Zhou Y, Gu J, Yu H, et al. Screening and identification of ESR1 as a target of icaritin in hepatocellular carcinoma: evidence from bibliometrics and bioinformatic analysis. *Curr Mol Pharmacol.* 2024;17:e18761429260902. doi:10.2174/0118761429260902230925044009
43. Ye W, Wang J, Zheng J, Jiang M, Zhou Y, Wu Z. Association between higher expression of Vav1 in hepatocellular carcinoma and unfavourable clinicopathological features and prognosis. *Protein Pept Lett.* 2024;31(9):706–713. doi:10.2174/0109298665330781240830042601
44. Toyoda H, Johnson PJ. The ALBI score: from liver function in patients with HCC to a general measure of liver function. *JHEP Rep.* 2022;4(10):100557. doi:10.1016/j.jhepr.2022.100557
45. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach—the ALBI grade. *J Clin Oncol.* 2015;33(6):550–558. doi:10.1200/jco.2014.57.9151
46. Casadei-Gardini A, Rimini M, Kudo M, et al. Real life study of lenvatinib therapy for hepatocellular carcinoma: RELEVANT study. *Liver Cancer.* 2022;11(6):527–539. doi:10.1159/000525145
47. Hiraoka A, Kumada T, Atsukawa M, et al. Prognostic factor of lenvatinib for unresectable hepatocellular carcinoma in real-world conditions—Multicenter analysis. *Cancer Med.* 2019;8(8):3719–3728. doi:10.1002/cam4.2241
48. Hiraoka A, Kumada T, Tada T, et al. Nutritional index as prognostic indicator in patients receiving lenvatinib treatment for unresectable hepatocellular carcinoma. *Oncology.* 2020;98(5):295–302. doi:10.1159/000506293
49. Lu CH, Kao WY, Wu CH, et al. Predicting survival outcomes in patients with hepatocellular carcinoma receiving lenvatinib by using the up7-ALBI score. *Liver Cancer.* 2025:1–14. doi:10.1159/000546185.
50. Yang X, Chen B, Wang Y, et al. Real-world efficacy and prognostic factors of lenvatinib plus PD-1 inhibitors in 378 unresectable hepatocellular carcinoma patients. *Hepatol Int.* 2023;17(3):709–719. doi:10.1007/s12072-022-10480-y
51. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci.* 2021;2(3):160. doi:10.1007/s42979-021-00592-x
52. An C, Park YW, Ahn SS, Han K, Kim H, Lee SK. Radiomics machine learning study with a small sample size: single random training-test set split may lead to unreliable results. *PLoS One.* 2021;16(8):e0256152. doi:10.1371/journal.pone.0256152
53. Zhang X, Zhang Y, Zhang G, et al. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. *Front Oncol.* 2022;12:773840. doi:10.3389/fonc.2022.773840
54. Cobo M, Menéndez Fernández-Miranda P, Bastarrika G, Lloret Iglesias L. Enhancing radiomics and deep learning systems through the standardization of medical imaging workflows. *Sci Data.* 2023;10(1):732. doi:10.1038/s41597-023-02641-x

Journal of Hepatocellular Carcinoma

Publish your work in this journal

The Journal of Hepatocellular Carcinoma is an international, peer-reviewed, open access journal that offers a platform for the dissemination and study of clinical, translational and basic research findings in this rapidly developing field. Development in areas including, but not limited to, epidemiology, vaccination, hepatitis therapy, pathology and molecular tumor classification and prognostication are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-hepatocellular-carcinoma-journal>

Dovepress
Taylor & Francis Group