Original Article

Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection

Jianxing Zeng[1,2,3#], Jinhua Zeng[1,2,4#], Kongying Lin[3#], Haitao Lin[3], Qionglan Wu[5], Pengfei Guo[3], Weiping Zhou[6], Jingfeng Liu[1,2,4]

[1]Department of Hepatic Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China; [2]The First Affiliated Hospital of Fujian Medical University, Fuzhou, China; [3]Southeast Big Data Institute of Hepatobiliary Health, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China; [4]The Liver Center of Fujian Province, Fujian Medical University, Fuzhou, China; [5]Department of Pathology, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China; [6]The Third Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China

Contributions: (I) Conception and design: JX Zeng, JH Zeng, K Lin, W Zhou, J Liu; (II) Administrative support: JH Zeng, P Guo, W Zhou, J Liu; (III) Provision of study materials: JH Zeng, P Guo, W Zhou, J Liu; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: JX Zeng, K Lin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jingfeng Liu, PhD. Department of Hepatic Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou 350025, China. Email: drjingfeng@126.com.

Background: Early recurrence is common for hepatocellular carcinoma (HCC) after surgical resection, being the leading cause of death. Traditionally, the COX proportional hazard (CPH) models based on linearity assumption have been used to predict early recurrence, but predictive performance is limited. Machine learning models offer a novel methodology and have several advantages over CPH models. Hence, the purpose of this study was to compare random survival forests (RSF) model with CPH models in prediction of early recurrence for HCC patients after curative resection.

Methods: A total of 4,758 patients undergoing curative resection from two medical centers were included. Fifteen features including age, gender, etiology, platelet count, albumin, total bilirubin, AFP, tumor size, tumor number, microvascular invasion, macrovascular invasion, Edmondson-Steiner grade, tumor capsular, satellite nodules and liver cirrhosis were used to construct the RSF model in training cohort. Discrimination, calibration, clinical usefulness and overall performance were assessed and compared with other models.

Results: Five hundred survival trees were used to generate the RFS model. The five highest Variable Importance (VIMP) were tumor size, macrovascular invasion, microvascular invasion, tumor number and AFP. In training, internal and external validation cohort, the C-index of RSF model were 0.725 [standard errors (SE) =0.005], 0.762 (SE =0.011) and 0.747 (SE =0.016), respectively; the Gönen & Heller's K of RSF model were 0.684 (SE =0.005), 0.711 (SE =0.008) and 0.697 (SE =0.014), respectively; the time-dependent AUC (2 years) of RSF model were 0.818 (SE =0.008), 0.823 (SE =0.014) and 0.785 (SE =0.025), respectively. The RSF model outperformed early recurrence after surgery for liver tumor (ERASL) model, Korean model, American Joint Committee on Cancer tumor-node-metastasis (AJCC TNM) stage, Barcelona Clinic Liver Cancer (BCLC) stage and Chinese stage. The RSF model is capable of stratifying patients into three different risk groups (low-risk, intermediate-risk, high-risk groups) in the training and two validation cohorts (all P<0.0001). A web-based prediction tool was built to facilitate clinical application (https://recurrenceprediction.shinyapps.io/surgery_predict/).

Conclusions: The RSF model is a reliable tool to predict early recurrence for patients with HCC after curative resection because it exhibited superior performance compared with other models. This novel model will be helpful to guide postoperative follow-up and adjuvant therapy.

Keywords: Hepatocellular carcinoma (HCC); liver resection; early recurrence; machine learning; individualized prediction.

Submitted May 05, 2020. Accepted for publication Jun 17, 2020.
doi: 10.21037/hbsn-20-466
View this article at: http://dx.doi.org/10.21037/hbsn-20-466

Introduction

Hepatocellular carcinoma (HCC) is the fifth most frequent malignancy and the third leading cause of cancer-related mortality worldwide (1). Currently, hepatic resection remains one of the most effective treatments with curative potential (2). However, long-term survival outcomes after resection remain unsatisfactory because of the high incidence of tumor recurrence, which exceeds 60% at 5 years even in patients with small tumors (3,4).

Hepatocellular carcinoma recurrence is commonly divided into early or late recurrence by using 2 years as the cut-off (5,6). Early recurrence represents metastasis from the initial HCC, whereas late recurrence is often of clonal origin (7,8). Early recurrence accounts for more than 70% of tumor recurrence (9). Therefore, identifying patients with HCC at high risk of early recurrence is important to enhance surveillance and to detect recurrence as early as possible.

Traditionally, the COX proportional hazard (CPH) models have been used in evaluating prognosis. The CPH models are used to identify the prognostic factors to predict early recurrence of individuals (10,11). However, the approaches make linearity assumption and thus cannot model the complicated, multidimensional and nonlinear relationships among different prognostic variables that may be present in biological systems, so the predictive performance is limited. Novel solutions that can deal with these potentially nonlinear variables are in great demand for accurate prognostic prediction.

Machine learning, an area of artificial intelligence that allows mining the relationships from complex datasets, has been used to make predictions about future outcomes (12). Machine learning models have several advantages over CPH models, which use nonlinear functions and consider all possible interactions between variables to improve the predictive performance (13,14). Previous studies applying machine learning models to HCC have reported good results. Singal *et al.* demonstrated that the machine learning model was better than the conventional regression model in predicting development of HCC (15). Kawaguchi *et al.* revealed that serum albumin level >3.7 g/dL was the best prognostic profile for nonalcoholic fatty liver disease (NAFLD)-HCC patients using data mining analysis (16). Cucchetti *et al.* reported that the artificial neural network (ANN) model could accurately predict tumor grade and microvascular invasion of HCC based on preoperative indicators (17). Qiao *et al.* also used ANN model to predict survival of patients with early HCC (18).

This study aimed to compare a machine learning model (Random Survival Forests model) with CPH models in prediction of early recurrence for patients with HCC after curative resection based on readily accessible clinical and pathological parameters. We present the following article in accordance with the TRIPOD reporting checklist (available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-20-466/rc) (19).

Methods

Patients

This study was conducted to the ethical guideline of the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Ethics Committee of the Mengchao Hepatobiliary Hospital of Fujian Medical University (No. 2020-092-01). Informed consent was obtained from each patient for their data to be used for research purposes. Data of patients with HCC who underwent primary hepatectomy at Eastern Hepatobiliary Surgery Hospital between January 2008 and December 2015, Mengchao Hepatobiliary Hospital of Fujian Medical University between January 2014 and December 2016 were prospectively collected and retrospectively analyzed.

The inclusion criteria were (I) Child-Pugh A or B7 liver function; (II) no extrahepatic metastasis; (III) R0 resection, defined as complete resection of macroscopic tumor nodules with tumor-free margins confirmed by histological examination (20). Patients who received palliative tumor resection, underwent preoperative anticancer treatments, had the history of other malignancies, had incomplete clinical data and lost to follow-up within 2 months of surgery were excluded from the analysis.

Eligible patients from Eastern Hepatobiliary Surgery Hospital between 2008 and 2013 formed the training cohort, whereas those patients between 2014 and 2015 formed the internal validation cohort. All eligible patients from Mengchao Hepatobiliary Hospital of Fujian Medical University were used as the external validation cohort in this study.

Clinicopathologic variables

Patient baseline characteristics included age, gender and liver cirrhosis. Routine serological examination included platelet count, albumin, total bilirubin, Alpha-fetoprotein (AFP), hepatitis B and hepatitis C virus immunology. Tumor characteristics included tumor size, tumor number, microvascular invasion, macrovascular invasion, Edmondson-Steiner grade, tumor capsular and satellite nodules.

According to previously described cut-offs, albumin-bilirubin (ALBI) grade divided into 3 grades (21). The pathological reviews of all resected specimens were carried out independently by two pathologists. Tumor size means the diameter of the largest tumor. The histologic grade of tumor cell differentiation was based on the Edmondson-Steiner grade (22). Satellite nodules are defined as tumor cell nests on microscopy or their sizes were less than 2 cm on macroscopy presenting within 2 cm of the main tumor (23).

Follow-up

Patients were followed up once every 3 months for the first 2 years after discharge from hospitals and every 3–6 months in subsequent years. The follow-up program included liver function, serum AFP level and an imaging study such as abdominal ultrasonography, contrast-enhanced computed tomography (CT) of abdomen, magnetic resonance imaging (MRI) of abdomen. The follow-up was censored on 31st December 2018.

The diagnosis of recurrent HCC was based on CT and/or MRI and elevated AFP levels. Once tumor recurrence was diagnosed, patients underwent further investigations. Appropriate treatments were given, which included percutaneous ethanol injection, radiofrequency ablation, transarterial chemoembolization, or liver re-resection, depending on the general condition of the patient, the liver functional reserve, the pattern of tumor recurrence, the patient's wish and the recommended treatment by the multidisciplinary team according to the EASL guideline (24).

Random survival forests (RSF) model

RSF model is used as a regression algorithm based on ensemble learning of decision trees using the techniques of random forests called feature and sample bragging which allows faster training process and less estimation bias. The RSF model can be used for censored survival data due to its modification of changing the Gini impurity which nodes split according to log-rank statistics to maximize the difference between the survival curves of Kaplan-Meier estimation after the cut-off.

The RSF can also estimate the individuals' cumulative hazard function (CHF) by integrating the Nelson-Allen estimator in the model (25). Besides, Variable Importance (VIMP) was obtained by measuring the decrease in prediction accuracy using out-of-bag data which were not used for building trees each time. The risk index was derived from the estimated CHF. In this study, a higher risk index implied a higher risk of recurrence. To assess the significance of the risk index, it was used as a continuous covariate into the Cox model. Risk groups were generated by the previously reported cutoffs (50th and 85th centile) of the risk index (26). Kaplan-Meier analysis of each risk group was plotted in each cohort.

Assessment and compassion of model performance

We used several complementary methods to assess different aspects of model performance, including model discrimination, model calibration, clinical usefulness and overall performance (27,28). Dynamic time-dependent measure was evaluated to be 2 years because we aimed to evaluate early recurrence.

Model discrimination was measured by the Harrell's C-index, Gönen & Heller's K, and time-dependent areas under the receiver operating characteristic curve (tdAUC). Model calibration was measured by the calibration plot. Estimates of predicted *vs.* actual 2-year recurrence probability were generated via bootstrapping (with 300 resampling). Clinical usefulness was measured by decision curve analysis (DCA) and net benefit at the threshold of 50%. Overall performance was measured by prediction error curves, time-dependent Brier score and time-dependent R^2 (29).

The RSF model was also compared to the early recurrence after surgery for liver tumor (ERASL) model (10), Korean model (11), American Joint Committee on Cancer tumor-node-metastasis (AJCC TNM) stage (30), Barcelona Clinic Liver Cancer (BCLC) stage (24), and Chinese stage (31) in each cohort. The diagnostic accuracy of the model was compared via category-based net reclassification improvement (NRI) and integrated discrimination improvement (IDI) (32,33). The category-based NRI was calculated by three risk categories (<50% risk, 50–85% risk, ≥85% risk).

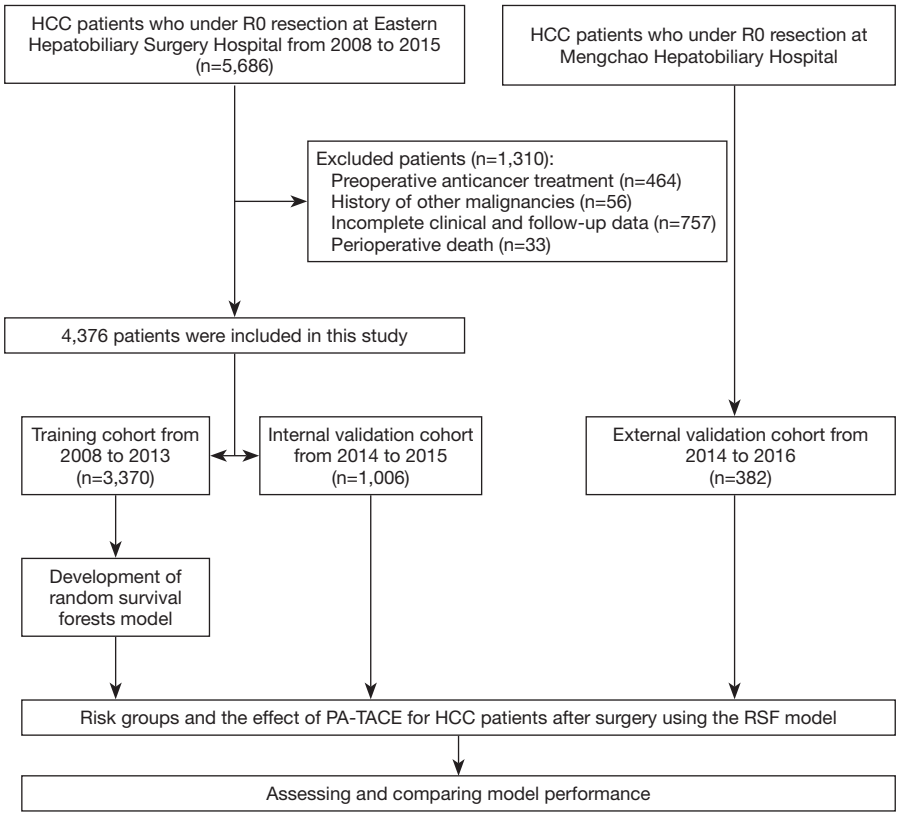

Figure 1 The flow chart for the three cohorts in the study. HCC, hepatocellular carcinoma; PA-TACE, postoperative adjuvant transcatheter arterial chemoembolization; RSF, random survival forests.

Statistical analysis

Categorical variables are presented as n (%) and compared using the chi-square test or Fisher exact test. Mean (standard deviation, SD) presented for normally distributed continuous variables and compared using the Student t-test, while median [interquartile range (IQR)] was given to those with non-normally distributed continuous variables and compared using the Mann-Whitney U test. All statistical tests were 2-tailed and a P value of less than 0.05 was considered statistically significant. All statistical analysis was performed with R version 3.5.2 (http://www.r-project.org/). These R packages were used in this study (Table S1).

Results

Baseline characteristics of patients

A total of the 5,686 HCC patients who underwent curative resection at Eastern Hepatobiliary Surgery Hospital between January 2008 and December 2015, 4,376 met the inclusion criteria. A total of 1,310 patients were excluded because of preoperative anticancer treatment (n=464), history of other malignancies (n=56), incomplete clinical and follow-up data (n=757) and perioperative death (n=33). Data collected from January 2008 to December 2013 including 3,370 HCC patients formed the training cohort. Recorded from January 2014 to December 2015, 1,006 HCC patients formed the internal validation cohort. The external validation cohort consisted of 382 patients from Mengchao Hepatobiliary Hospital of Fujian Medical University. The flow chart of these patients was shown in *Figure 1*.

The baseline characteristics of patients were shown in *Table 1*. Some clinicopathologic features such as tumor size, microvascular invasion, macrovascular invasion, Edmondson-Steiner grade, tumor capsular and satellite nodules were different among the three cohorts. The 2-year recurrence rate were 43.4% (95% CI: 41.7–45.1%), 37.6% (95% CI: 34.4–40.6%) and 50.2% (95% CI: 44.7–55.1%) in the three cohorts, respectively (Figure S1).

Table 1 Baseline characteristics of patients

Variables	Training cohort (n=3,370)	Internal validation cohort (n=1,006)	External validation cohort (n=382)
Patient factors/laboratory parameters			
Age [year, mean (SD)]	51.1 (10.8)	52.5 (10.5)	54.2 (10.9)
Gender, male, n (%)	2,927 (86.9)	863 (85.8)	318 (83.2)
Etiology, n (%)			
HBV	2,983 (88.5)	878 (87.3)	315 (82.5)
HCV	58 (1.7)	10 (1.0)	2 (0.5)
NBNC	329 (9.8)	118 (11.7)	65 (17.0)
PLT [10^9/L, mean (SD)]	165 (69.7)	166 (68.0)	175 (76.4)
ALB [g/L, mean (SD)]	42.1 (3.81)	42.1 (3.41)	40.3 (3.85)
TBIL [μmol/L, median (IQR)]	13.3 [10.6, 17.0]	13.3 [10.5, 16.8]	15.4 [11.6, 19.8]
AFP [ng/mL, median (IQR)]	80.3 [7.00, 1210]	84.9 [6.20, 1210]	54.6 [5.76, 842]
ALBI grade, n (%)			
1	2,598 (77.1)	807 (80.2)	206 (53.9)
2	771 (22.9)	198 (19.7)	176 (46.1)
3	1 (0.0003)	1 (0.1)	0 (0.0)
Tumor factors			
Tumor size [cm, mean (SD)]	6.38 (3.77)	5.89 (3.67)	5.93 (4.29)
Solitary tumor number, n (%)	2,747 (81.5)	811 (80.6)	314 (82.2)
Microvascular invasion, n (%)	1,251 (37.1)	452 (44.9)	233 (61.0)
Macrovascular invasion, n (%)	448 (13.3)	92 (9.1)	75 (19.6)
Edmondson-Steiner grade, n (%)			
I–II	514 (15.3)	68 (6.8)	109 (28.5)
III–IV	2,856 (84.7)	938 (93.2)	273 (71.5)
Tumor capsular, n (%)	2,681 (79.6)	825 (82.0)	174 (45.5)
Satellite nodules, n (%)	1,249 (37.1)	504 (50.1)	123 (32.2)
Liver cirrhosis, n (%)	2,454 (72.8)	699 (69.5)	297 (77.7)
BCLC stage, n (%)			
0/A	2,494 (74.0)	779 (77.5)	271 (70.9)
B	428 (12.7)	135 (13.4)	36 (9.5)
C	448 (13.3)	92 (9.1)	75 (19.6)
AJCC TNM stage, n (%)			
I	1,849 (54.9)	494 (49.1)	143 (37.4)
II	815 (24.1)	345 (34.3)	149 (39.1)
IIIA	258 (7.7)	75 (7.5)	15 (3.9)
IIIB	448 (13.3)	92 (9.1)	75 (19.6)

Categorical variables are presented as n (%). Mean (SD) was presented for normally distributed continuous variables, while median [IQR] was given to those with non-normally distributed continuous variables. HBV, hepatitis B virus; HCV, hepatitis C virus; NBNC, non-B non-C; PLT, platelet count; ALB, albumin; TBIL, total bilirubin; AFP, alpha-fetoprotein; ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; AJCC TNM, American Joint Committee on Cancer tumor-node-metastasis; SD, standard deviation; IQR, interquartile range; CI, confidence interval.

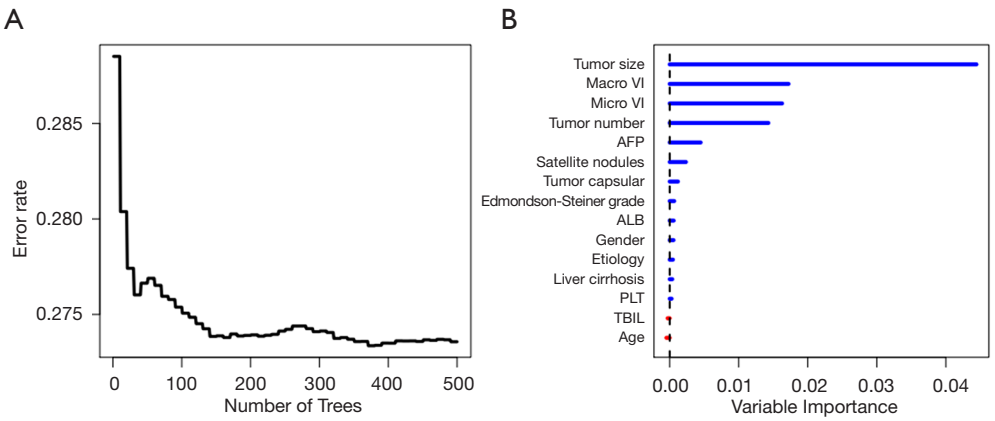

Figure 2 Construction of the RSF model in predicting early recurrence in the training cohort. (A) Prediction error rates. (B) The VIMP plot. Macro VI, macrovascular invasion; Micro VI, microvascular invasion; AFP, alpha-fetoprotein; ALB, albumin; PLT, platelet count; TBIL, total bilirubin; RSF, random survival forests; VIMP, variable importance.

Construction of the RSF model in predicting early recurrence in the training cohort

Fifteen features including age, gender, etiology, platelet count, albumin, total bilirubin, alpha-fetoprotein (AFP), tumor size, tumor number, microvascular invasion, macrovascular invasion, Edmondson-Steiner grade, tumor capsular, satellite nodules and liver cirrhosis were used to construct the RSF model. During the process of constructing 200 survival trees, the prediction error rate tended to be low and stable (*Figure 2A*). Variable importance (VIMP) for all the features used to grow trees was also generated after the complete construction of 500 trees. Higher VIMP indicated that the variable contributed more to the prediction of early recurrence. As shown in *Figure 2B*, the five highest-ranking variables were tumor size, macrovascular invasion, microvascular invasion, tumor number and AFP, which were aggressive tumor characteristics.

Assessing and comparing model performance

Model discrimination was compared via the Harrell's C-index, Gönen & Heller's K and time-dependent AUC (2 years). In training, internal and external validation cohort, the C-index of RSF model were 0.725 [standard errors (SE) =0.005], 0.762 (SE =0.011) and 0.747 (SE =0.016), respectively (*Table 2*). The Gönen & Heller's K of RSF model were 0.684 (SE =0.005), 0.711 (SE =0.008) and 0.697 (SE =0.014), respectively (*Table 2*). The time-dependent AUC (2 years) of RSF model were 0.818 (SE =0.008), 0.823 (SE =0.014) and 0.785 (SE =0.025), which were greater than ERASL model, Korean model, AJCC TNM stage, BCLC stage and Chinese stage in the three cohorts (*Table 2*; *Figure 3*).The Harrell's C-index and Gönen & Heller's K of the RSF model were also higher than other models in predicting early recurrence in the three cohorts (*Table 2*).

Decision curve analysis (DCA) was used to facilitate the comparison between the RSF model and 5 other models in the three cohorts. As shown in *Figure 4*, DCA has graphed the clinical usefulness of each model based on probability thresholds of recurrent risk (X-axis) and the net benefit of using the model (Y-axis). DCA revealed that the RSF model had a better net benefit than 5 other models.

In addition, as shown in *Figure 5*, the RSF model displayed a lower prediction error rate than other models. Time-dependent Brier score and R^2 (2 years) were also better than other models (*Table 2*).

Calibration plots displayed an overall good agreement between the prediction of the RSF model and actual outcome in the probability of 2-year recurrence in the three cohorts (*Figure 6*).

The diagnostic accuracy of the model was compared via net reclassification improvement (NRI) and integrated discrimination improvement (IDI). The RSF model improved diagnostic accuracy when compared to the ERASL model (NRI =0.135, P<0.001; IDI =0.054, P<0.001; Table S2; Figure S2).

Risk stratification

Based on the risk index of the RSF model, using 32.524

Table 2 Comparison of model performance between RSF model and 5 other models in predicting early recurrence

Performance	Cohort	RSF	ERASL	Korean	AJCC TNM	BCLC	Chinese
Discrimination							
Harrell's C-index	Training	0.725 (0.005)	0.706 (0.006)	0.658 (0.006)	0.674 (0.006)	0.635 (0.006)	0.684 (0.006)
	Internal	0.762 (0.011)	0.726 (0.012)	0.672 (0.013)	0.711 (0.012)	0.646 (0.012)	0.709 (0.012)
	External	0.747 (0.016)	0.727 (0.017)	0.722 (0.017)	0.711 (0.017)	0.658 (0.018)	0.696 (0.018)
Gönen & Heller's K	Training	0.684 (0.005)	0.672 (0.005)	0.638 (0.006)	0.647 (0.005)	0.616 (0.004)	0.642 (0.004)
	Internal	0.711 (0.008)	0.694 (0.010)	0.654 (0.011)	0.667 (0.008)	0.617 (0.008)	0.651 (0.008)
	External	0.697 (0.014)	0.689 (0.015)	0.688 (0.015)	0.657 (0.014)	0.619 (0.013)	0.632 (0.013)
Time-dependent AUC (2 years)	Training	0.818 (0.008)	0.791 (0.008)	0.721 (0.009)	0.747 (0.008)	0.689 (0.008)	0.757 (0.008)
	Internal	0.823 (0.014)	0.784 (0.016)	0.727 (0.017)	0.757 (0.015)	0.676 (0.014)	0.758 (0.016)
	External	0.785 (0.025)	0.783 (0.025)	0.780 (0.025)	0.749 (0.026)	0.678 (0.025)	0.717 (0.027)
Clinical usefulness							
Net benefit at threshold 50%	Training	0.166	0.154	0.093	0.139	0.137	0.137
	Internal	0.121	0.092	0.041	0.095	0.073	0.073
	External	0.206	0.190	0.222	0.185	0.154	0.154
Overall performance							
Time-dependent Brier (2 years)	Training	0.147	0.156	0.174	0.160	0.167	0.161
	Internal	0.129	0.143	0.159	0.144	0.154	0.146
	External	0.156	0.162	0.161	0.169	0.180	0.176
Time-dependent R^2 (2 years)	Training	0.287	0.239	0.142	0.220	0.175	0.214
	Internal	0.306	0.233	0.145	0.220	0.150	0.206
	External	0.235	0.225	0.230	0.187	0.125	0.140

The parentheses are standard errors. AUC, areas under receiver operating characteristic curve; RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor; AJCC TNM, American Joint Committee on Cancer tumor-node-metastasis; BCLC, Barcelona Clinic Liver Cancer.

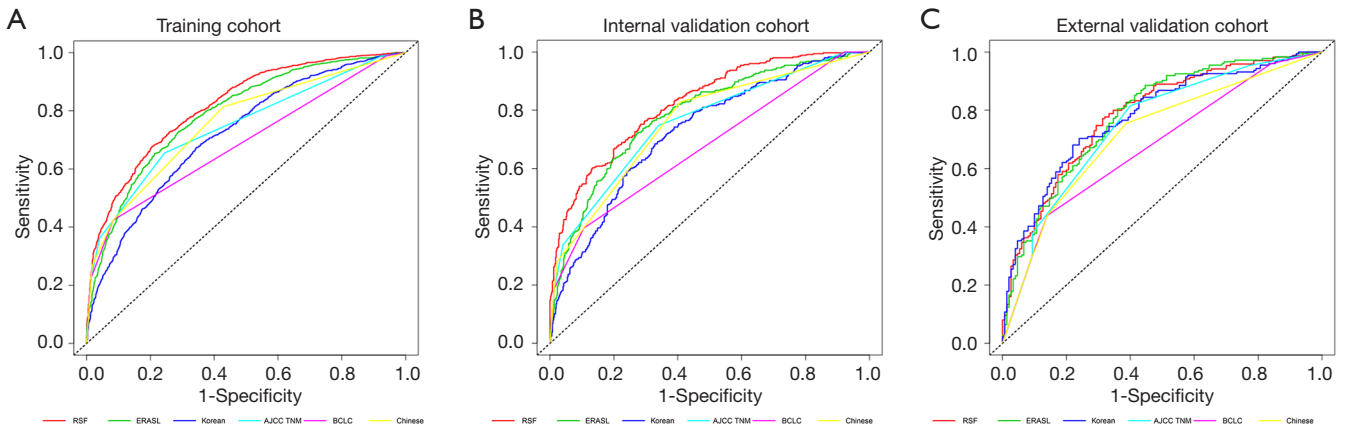

Figure 3 Comparison of time-dependent ROC (2 years) between the RSF model and 5 other models. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort. ROC, receiver operating characteristic curve; RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor; AJCC TNM, American Joint Committee on Cancer tumor-node-metastasis; BCLC, Barcelona Clinic Liver Cancer.

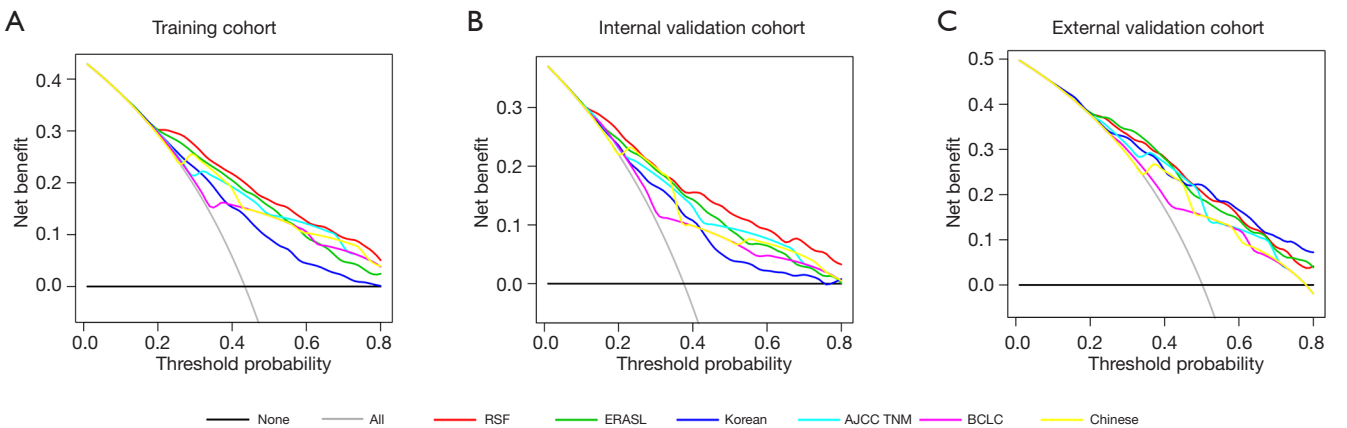

Figure 4 Comparison of decision curve analysis between the RSF model and 5 other models in predicting early recurrence. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort. RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor; AJCC TNM, American Joint Committee on Cancer tumor-node-metastasis; BCLC, Barcelona Clinic Liver Cancer.

and 66.511 as the cut-off values (which correspond to the 50th and 85th centile of risk index in training cohort), the patients were classified into low-risk, intermediate-risk, high-risk groups. Kaplan-Meier analysis showed that recurrence rates were stratified among three risk groups in the training and two validation cohorts (all P<0.0001) (Table S3; Figure S3).

We implemented a web-based prediction tool for clinicians to use the RSF model. This tool could output the risk index, risk groups, the recurrence-free probability at 3, 6, 9, 12, 18, 24 months, was available at (Figure S4; https://recurrenceprediction.shinyapps.io/surgery_predict/).

Discussion

Tumor recurrence within 2 years, which accounts for 30-50% of patients, is a main cause of mortality (24). Therefore, identification of HCC patients after resection who are at high risk of early recurrence is important to facilitate screening and decision on adjuvant therapy. The COX proportional hazard (CPH) models have been commonly used to evaluate early recurrence based on

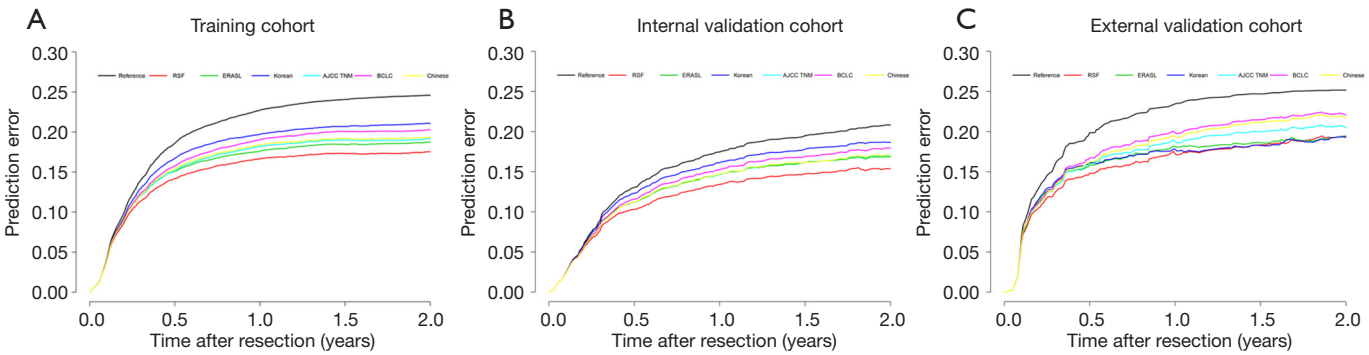

Figure 5 Comparison of prediction error curve (2 years) between the RSF model and 5 other models in predicting early recurrence. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort. RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor; AJCC TNM, American Joint Committee on Cancer tumor-node-metastasis; BCLC, Barcelona Clinic Liver Cancer.

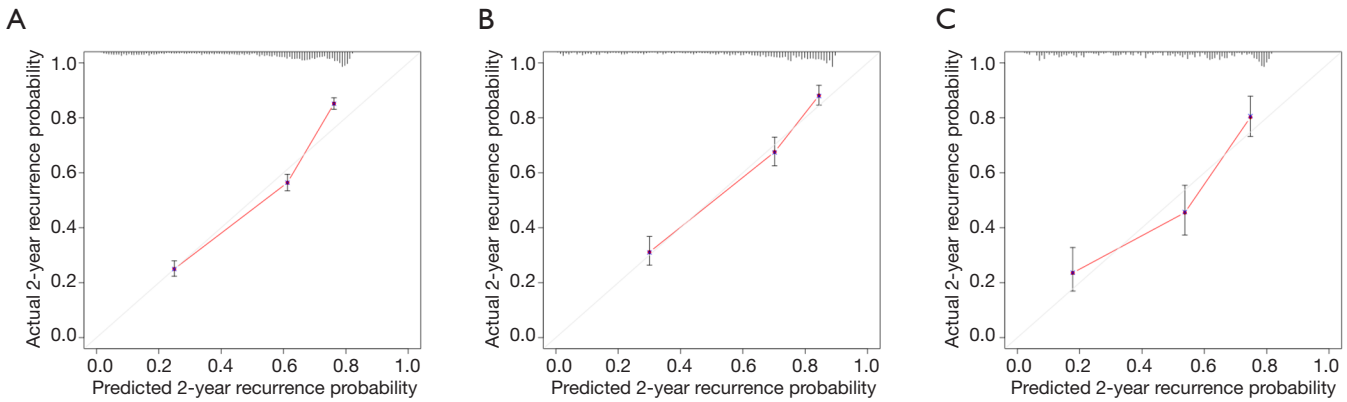

Figure 6 Calibration plots for the RSF model in predicting early recurrence. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort. RSF, random survival forests.

an assumption of linear association, but the predictive performance is limited. Machine learning models offer a novel methodology and have several advantages over CPH models, which use nonlinear functions and consider all possible interactions between variables to improve the predictive performance. Toward this goal, a machine learning model, RSF model, was developed and compared with CPH models to predict early recurrence for HCC patients who underwent curative resection based on readily accessible clinical and pathological parameters.

To our knowledge, our study is the first to report and validate a machine learning model for predicting early recurrence in HCC patients treated with curative resection. The results found that the machine learning model was superior to conventional statistical regression methods by assessing different indexes of model performance such as model discrimination, clinical usefulness and overall performance. The RSF model is a novel nonlinear machine learning model for survival analysis (25,34). The core elements of the RSF model are generating the survival tree and constructing the ensemble cumulative hazard function. The main advantage of the RSF model is that it exhibits an improvement for all variables with the use of nonlinear risk functions and does not use required assumptions such as the CPH model.

According to VIMP analysis, our findings (*Figure 2B*) echo numerous previous studies in that early recurrence is mainly associated with aggressive tumor characteristics such as tumor size, vascular invasion, tumor multiplicity and higher AFP (7-9). These results demonstrated that the RSF model also has the function of finding out the important factors for predicting early recurrence (according to VIMP)

like the Cox model (according to P value).

In addition, several novel measures are employed to assess model performance, including reclassification tables, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) (32,33). These measures also demonstrated that the RSF model outperformed other models in predicting early recurrence. Moreover, net benefit (NB), with visualization in DCA, is a simple summary measure to quantify clinical usefulness when decisions are to be supported by a prediction model (35,36). The DCA showed that the RSF model provided superior net benefit when compared to other models. For instance, calculating NB at a single threshold 50%, the RSF model could improve NB by 0.012 compared to the ERASL model in the training cohort, equivalent to 1.1 more detected early recurrence per 100 patients at no additional cost (*Table 2*).

The RSF model is capable of stratifying patients into three different risk groups. The high-risk groups accounted for 14.6% of the patients among the entire cohort but 86.2% of those occurred early recurrence, whereas the low-risk and intermediate-risk groups consisted of 48.9% and 36.5% of patients but only 21.6% and 56.5% of that developed early recurrence, respectively (Figure S3). The model can identify a small subset of patients with a high risk of early recurrence. While it may not be reasonable to exclude these patients with a high risk of early recurrence from surgical treatment, they would be candidates for postoperative adjuvant therapy.

There are some limitations to our study. Firstly, selection bias was hard to avoid in this study. However, this bias has been minimized by two large independent cohorts. Secondly, this study was conducted in China and most HCC patients had a background of HBV infection, but aetiological factors and liver background contributed less to early recurrence in the previous study (7-9). Moreover, aetiology and liver cirrhosis were not identified as important predictors in this study. Still, it should be admitted that further external validation in different geographic regions and aetiology is of necessity. Thirdly, the machine learning model may appear complex and hard to apply in clinical practice, but our simple online web-based tool overcomes this problem.

In summary, the RSF model is a robust tool to predict early recurrence for patients with HCC after curative resection because it exhibited better performance compared with other models. The model is able to stratify patients into three different groups (low-risk, intermediate-risk, high-risk groups). This novel approach may provide clinicians with useful guidance for postoperative follow-up and treatments.

Acknowledgments

Funding: This study was supported by the Special Fund of Fujian Development and Reform Commission (31010308) and the Natural Science Foundation of Fujian Province (2018J01140).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-20-466/rc

Data Sharing Statement: Available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-20-466/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-20-466/coif). All authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted to the ethical guideline of the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Ethics Committee of the Mengchao Hepatobiliary Hospital of Fujian Medical University (No. 2020-092-01). Informed consent was obtained from each patient for their data to be used for research purposes.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics,

2012. CA Cancer J Clin 2015;65:87-108.
2. Zhang X, Li C, Wen T, et al. Appropriate treatment strategies for intrahepatic recurrence after curative resection of hepatocellular carcinoma initially within the Milan criteria: according to the recurrence pattern. Eur J Gastroenterol Hepatol 2015;27:933-40.
3. Dhir M, Melin AA, Douaiher J, et al. A Review and Update of Treatment Options and Controversies in the Management of Hepatocellular Carcinoma. Ann Surg 2016;263:1112-25.
4. Poon RT, Sheung T, Chung M, et al. Long-term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function: Implications for a strategy of salvage transplantation. Ann Surg 2002;235:373-82.
5. Imamura H, Matsuyama Y, Tanaka E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. J Hepatol 2003;38:200-7.
6. Portolani N, Coniglio A, Ghidoni S, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. Ann Surg 2006;243:229-35.
7. Du ZG, Wei YG, Chen KF, et al. Risk factors associated with early and late recurrence after curative resection of hepatocellular carcinoma: a single institution's experience with 398 consecutive patients. Hepatobiliary Pancreat Dis Int 2014;13:153-61.
8. Poon RTP. Differentiating Early and Late Recurrences After Resection of HCC in Cirrhotic Patients: Implications on Surveillance, Prevention, and Treatment Strategies. Ann Surg Oncol 2009;16:792.
9. Chan AWH, Chan SL, Wong GLH, et al. Prognostic Nutritional Index (PNI) Predicts Tumor Recurrence of Very Early/Early Stage Hepatocellular Carcinoma After Surgical Resection. Ann Surg Oncol 2015;22:4138-48.
10. Chan AWH, Zhong J, Berhane S, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. J Hepatol 2018;69:1284-93.
11. Shim JH, Jun MJ, Han S, et al. Prognostic Nomograms for Prediction of Recurrence and Survival After Curative Liver Resection for Hepatocellular Carcinoma. Ann Surg 2015;261:939-46.
12. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-73.
13. Waljee AK, Higgins PDR. Machine learning in medicine: a primer for physicians. Am J Gastroenterol 2010;105:1224-6.
14. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019;380:1347-58.
15. Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108:1723-30.
16. Kawaguchi T, Tokushige K, Hyogo H, et al. A Data Mining-based Prognostic Algorithm for NAFLD-related Hepatoma Patients: A Nationwide Study by the Japan Study Group of NAFLD. Sci Rep 2018;8:10434.
17. Cucchetti A, Piscaglia F, Grigioni ADE, et al. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study. J Hepatol 2010;52:880-8.
18. Qiao G, Li J, Huang A, et al. Artificial neural networking model for the prediction of post-hepatectomy survival of patients with early hepatocellular carcinoma. J Gastroenterol Hepatol 2014;29:2014-20.
19. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.
20. Wang K, Liu J, Yan ZL, et al. Overexpression of aspartyl-(asparaginyl)-beta-hydroxylase in hepatocellular carcinoma is associated with worse surgical outcome. Hepatology (Baltimore, Md) 2010;52:164-73.
21. Johnson PJ, Sarah B, Chiaki K, et al. Assessment of Liver Function in Patients With Hepatocellular Carcinoma: A New Evidence-Based Approach-The ALBI Grade. J Clin Oncol 2015;33:550-8.
22. Edmondson HA, Steiner PE. Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies. Cancer 1954;7:462-503.
23. Yang T, Lu JH, Lau WY, et al. Perioperative blood transfusion does not influence recurrence-free and overall survivals after curative resection for hepatocellular carcinoma: A Propensity Score Matching Analysis. J Hepatol 2016;64:583-93.
24. Galle PR, Forner A, Llovet JM, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol 2018;69:182-236.
25. Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. The Annals of Applied Statistics 2008;2:841-60.
26. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res

Methodol 2013;13:33.
27. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology 2010;21:128-38.
28. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162:W1-W73.
29. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. J Stat Softw 2012;50:1-23.
30. Chun YS, Pawlik TM, Vauthey JN. 8th Edition of the AJCC Cancer Staging Manual: Pancreas and Hepatobiliary Cancers. Ann Surg Oncol 2018;25:845-7.
31. Zhou J, Sun HC, Wang Z, et al. Guidelines for Diagnosis and Treatment of Primary Liver Cancer in China (2017 Edition). Liver Cancer 2018;7:235-60.
32. Pencina MJ, D'Agostino Sr RB, D' Agostino Jr RB, et al. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72; discussion 207-12.
33. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA 2017;318:1377-84.
34. Matsuo K, Purushotham S, Jiang B, et al. Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. Am J Obstet Gynecol 2019;220:381.e1-381.e14.
35. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Med Decis Making 2006;26:565-74.
36. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol 2018;74:796-804.

Cite this article as: Zeng J, Zeng J, Lin K, Lin H, Wu Q, Guo P, Zhou W, Liu J. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection. HepatoBiliary Surg Nutr 2022;11(2):176-187. doi: 10.21037/hbsn-20-466

Supplementary

Table S1 R packages

R package	Source	Function
table1	CRAN Repository	table1
randomForestSRC	CRAN Repository	rfsrc.fast, predict
survminer	CRAN Repository	survfit, survival
survival	CRAN Repository	coxph
timeROC	CRAN Repository	timeROC, plot
Hmisc	CRAN Repository	rcorrcens
CPE	CRAN Repository	phcpe
stdca	MSKCC website	stdca
pec	CRAN Repository	pec, ibs, R2
rms	CRAN Repository	cph, calibrate
nricens	CRAN Repository	nricens
survIDINRI	CRAN Repository	IDI.INF, IDI.INF.OUT
shiny	CRAN Repository	EventReactive, shinyServer

http://dx.doi.org/10.21037/hbsn-20-466

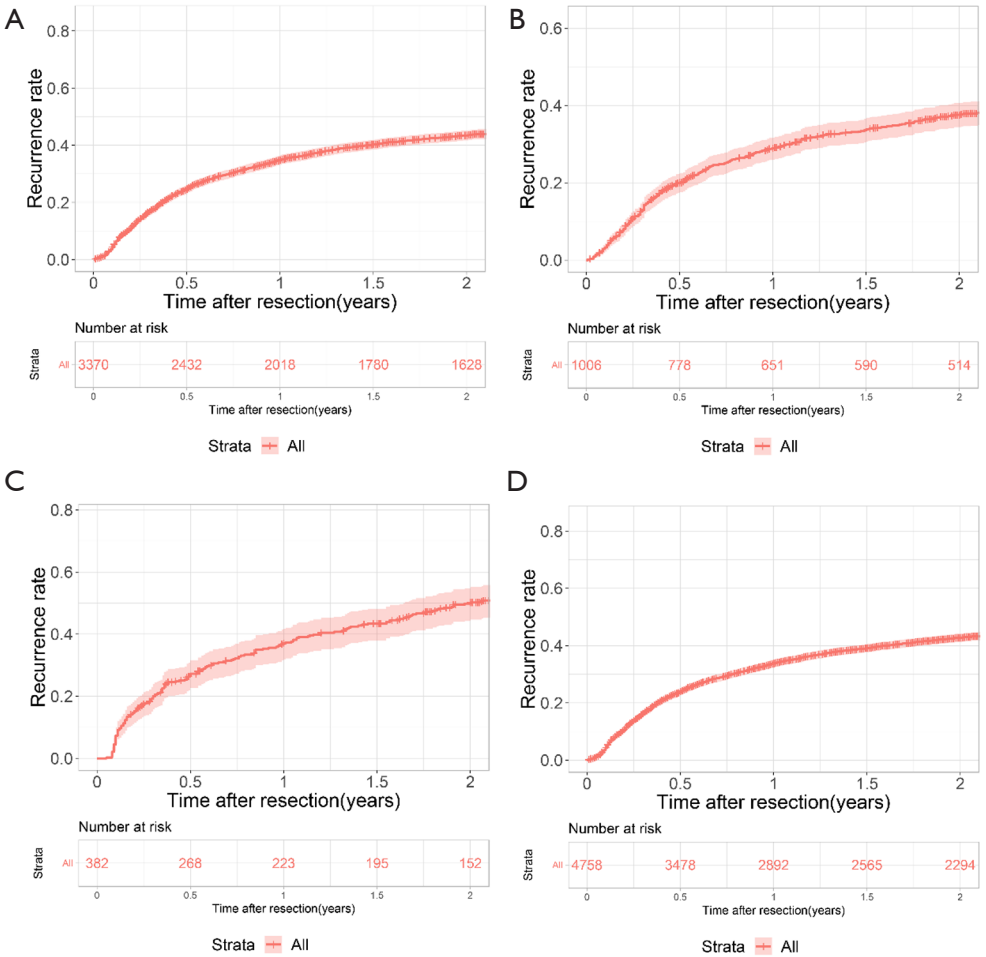

Figure S1 Kaplan-Meier curves for early recurrence. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort, (D) the entire cohort.

http://dx.doi.org/10.21037/hbsn-20-466

Table S2 Risk reclassification table for all patients in predicting early recurrence probabilities by the RSF and ERASL model

ERASL model	RSF model			Total
	<50% risk	50–85% risk	≥85% risk	
Patients with early recurrence (n=1,935)				
<50% risk	592	103	23	718
50–85% risk	207	423	320	950
≥85% risk	0	53	214	267
Total	799	579	557	1,935
Patients without early recurrence (n=2,287)*				
<50% risk	1,756	73	2	1,831
50–85% risk	180	214	33	427
≥85% risk	0	17	12	29
Total	1,936	304	47	2,287

*, lost to follow-up or dead within 2 years of surgery were excluded. RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor.

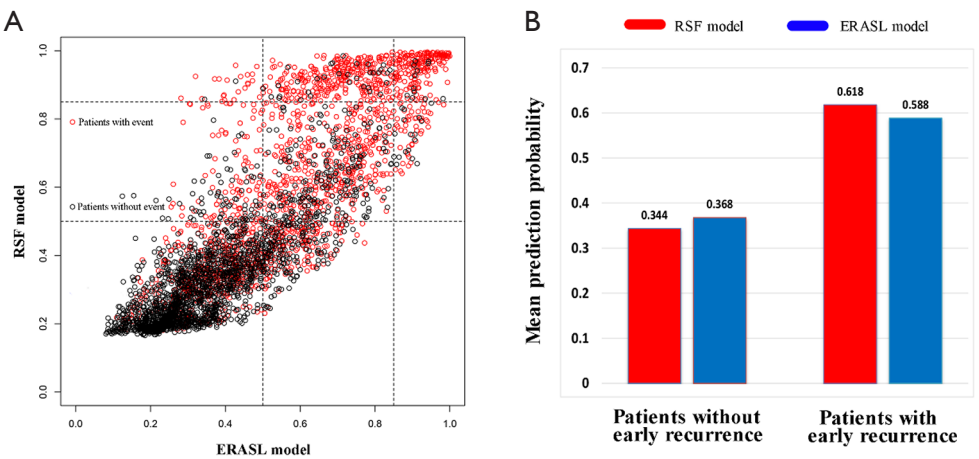

Figure S2 NRI and IDI. (A) Scatter plot for the entire cohort in predicting early recurrence probabilities by the RSF and ERASL models. (Red circle, early recurrence; black circle, without early recurrence). (B) Histogram for the entire cohort with early recurrence and without early recurrence in mean prediction probability by the RSF model (red) and the ERASL model (blue). NRI, net reclassification improvement; IDI, integrated discrimination improvement; RSF, random survival forests; ERASL, early recurrence after surgery for liver tumor.

http://dx.doi.org/10.21037/hbsn-20-466

Table S3 2-year recurrence rates, hazard ratio and P-value according to the three risk groups as defined by the RSF model

Cohort	Risk groups	n	2-year recurrence rate, % (95% CI)	Hazard radio (95% CI)	P value
Training	Low	1,685	22.5 (20.5, 24.5)	1	
	Intermediate	1,179	58.7 (55.6, 61.5)	2.527 (2.423, 2.632)	<0.0001
	High	506	86.5 (82.1, 89.8)	6.699 (6.566, 6.831)	<0.0001
Internal	Low	484	17.2 (13.7, 20.7)	1	
	Intermediate	402	48.3 (42.9, 53.1)	3.084 (2.851, 3.318)	<0.0001
	High	120	85.2 (76.6, 90.6)	11.655 (11.371, 11.938)	<0.0001
External	Low	158	24.0 (16.7, 30.6)	1	
	Intermediate	157	61.4 (52.6, 68.6)	3.366 (3.012, 3.720)	<0.0001
	High	67	86.7 (74.0, 93.2)	7.712 (7.308, 8.115)	<0.0001
Entire	Low	2,327	21.6 (19.8, 23.2)	1	
	Intermediate	1,738	56.5 (54.0, 58.9)	2.638 (2.547, 2.730)	<0.0001
	High	693	86.2 (82.7, 89.0)	7.423 (7.309, 7.537)	<0.0001

RSF, random survival forests; CI, confidence interval.

http://dx.doi.org/10.21037/hbsn-20-466

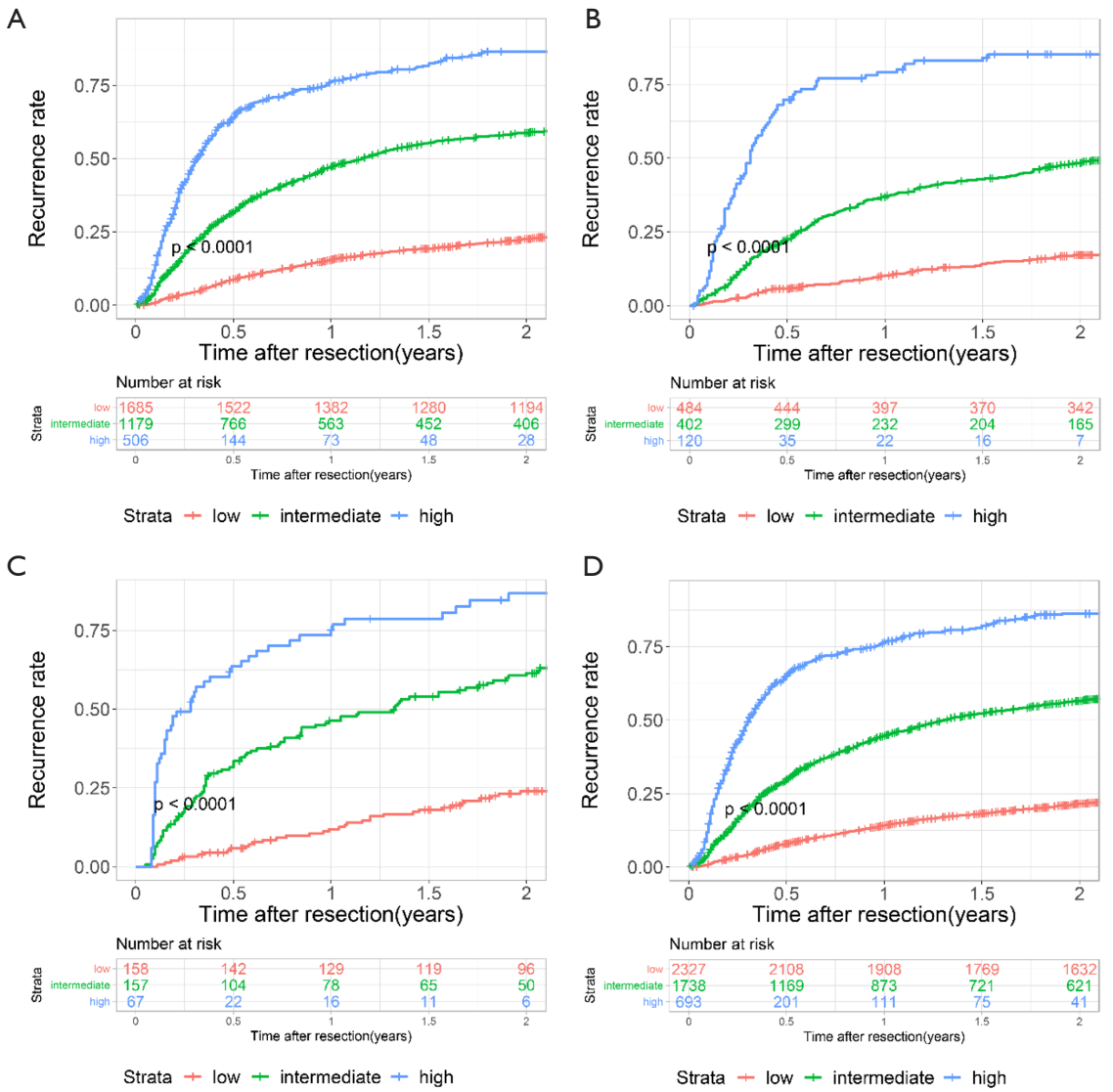

Figure S3 Kaplan-Meier curves for early recurrence in the low-risk, intermediate-risk, and high-risk groups defined by the RSF model. (A) Training cohort, (B) internal validation cohort, (C) external validation cohort, (D) the entire cohort.

http://dx.doi.org/10.21037/hbsn-20-466

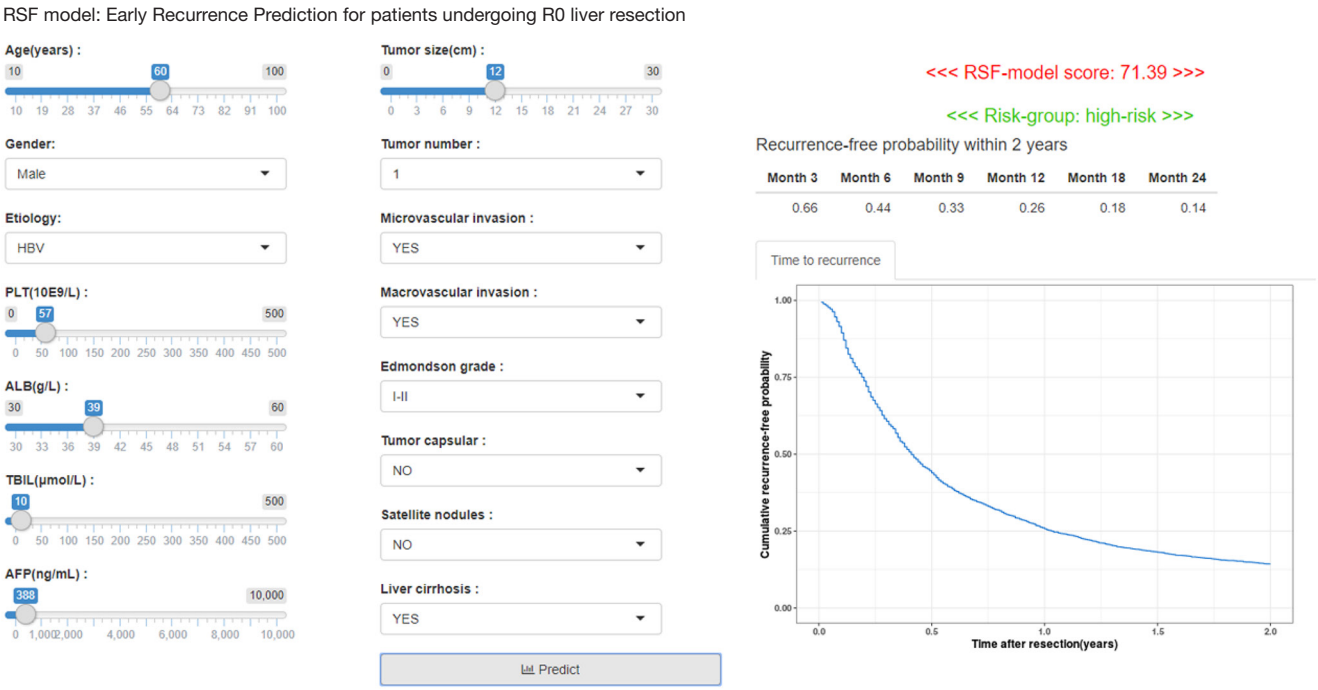

Figure S4 Screenshot of the online prediction tool based on the RSF model.

http://dx.doi.org/10.21037/hbsn-20-466