Journal of Hepatocellular Carcinoma

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

A Machine Learning Guided System for Therapeutic Decision-Making in Intra-Arterial Therapy of Hepatocellular Carcinoma

Linling Tan[1], Rongshan Fan[1], Jun Zhang[1], Hongmei Fan[1], Shenglan Wu[1], Wang Li[2]

[1]Department of Hepatology, Shenzhen Hospital of Integrated Traditional Chinese and Western Medicine, Shenzhen, Guangdong, People's Republic of China; [2]Department of Minimal Invasive Intervention, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, 510060, People's Republic of China

Correspondence: Wang Li, Department of Minimal invasive intervention, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, 651, Dongfeng East Road, Guangzhou, 510060, People's Republic of China, Tel/Fax +86-20-87343272, Email liwang@sysucc.org.cn

Objective: To develop and validate a machine learning-guided system (MLGS) for stratifying prognostic risk of unresectable hepatocellular carcinoma (uHCC) based on clinical data before and after intra-arterial therapy (IAT).

Methods: Between April 2008 and June 2022, a total of 5646 eligible patients with uHCC who underwent initial IAT were consecutively identified in 15 hospitals. The 5 year mortality was used as primary predictive outcome. Thirty-five sets of clinical data, including 29 preoperative and 6 postoperative data, were input successively into five supervised ML models. The performance of the ML models was compared using the area under the receiver operating characteristic (AUC) curve with the DeLong test. Kaplan-Meier analysis was used to revealed overall survival (OS) risk stratification of the MLGS.

Results: These patients were divided into training datasets (TD, n=3387), internal testing datasets (ITD, n=1130), and external test datasets (ETD, n=1129), respectively.The CatBoost model yield the best discrimination using preoperative data in the ML models using 23 variables, the AUC of CatBoost model were 0.777–0.735 in three datasets. Meanwhile, the XGBoost model yield the best discrimination using postoperative 20 variables and the AUC of XGBoost model were 0.904–0.861 in three datasets, which outperform significantly the performance of the CatBoost model (DeLong test, $P < 0.001$). The MLGS based on the XGBoost model provide significantly different OS between three risk stratification in three datasets (all, $P < 0.001$).

Conclusion: The MLGS may guide radiologists in developing strategies of IAT for uHCC. Prospective studies are needed to evaluate its clinical utility.

Keywords: intra-arterial therapies, hepatocellular carcinoma, risk scoring scale model, machine learning, risk stratification

Introduction

Hepatocellular carcinoma (HCC) constitutes roughly 90% of all primary liver malignancies, ranking as the sixth most prevalent cancer diagnosis and the third leading cause of cancer-associated deaths globally.[1,2] For individuals diagnosed with unresectable HCC (uHCC), intra-arterial therapies (IATs), encompassing hepatic arterial infusion chemotherapy (HAIC) and transarterial chemoembolization (TACE), are established as standard therapeutic approaches in clinical practice. International clinical guidelines recommend TACE as the first-line intervention for intermediate-stage HCC,[3,4] whereas HAIC has become an essential treatment option for intermediate-advanced HCC in Asian populations5. In recent years, growing evidence has shown that combining IATs with molecular-targeted agents (MATs) and immune checkpoint inhibitors (ICIs) produces favorable therapeutic effects characterized by profound and long-lasting efficacy.[5,6] Furthermore, sequential local-regional therapies—such as surgical resection (SR), liver transplantation, and thermal ablation (TA)—administered following conversion therapy based on IAT-inclusive comprehensive treatment have substantially enhanced the long-term survival of uHCC patients.[7–10]

https://doi.org/10.2147/JHC.S613199
Received: 3 April 2026
Accepted: 7 May 2026
Published: 18 June 2026

Numerous previous studies have identified a variety of clinical factors closely linked to HCC prognosis, and multiple predictive models have been constructed based on these variables to assist clinicians in precisely assessing patient outcomes and conducting risk stratification.[11–14] However, The aforementioned model is limited by factors such as data deficiency and insufficient computing power, resulting in slightly inadequate predictive ability for the prognosis of IAT for uHCC. On the contrast, machine learning (ML) emerges is a promising technology that utilizes statistical, probabilistic, and optimization algorithms for model training, offering unique advantages such as standardized handling of large-scale multicenter data, imputation of missing values, and high computational efficiency.[15–19] By automating analytical models, ML algorithms can detect patterns and support decision-making with minimal human involvement, and they have been increasingly utilized in the development of recurrence prediction models within the field of oncology.

HCC displays substantial tumor heterogeneity at the histological, genomic,[20] and molecular levels. Additionally, therapeutic decisions vary depending on tumor response to IAT and disease progression, resulting in marked interindividual differences in the prognostic outcomes of uHCC patients. Consequently, accurate assessment of dynamic tumor progression during IAT is essential for improving long-term survival results. However, most existing predictive models are constrained by a limited range of clinical factors, small sample sizes for model development, inconsistent pathological evaluation criteria, and insufficient clinical applicability. To date, there remains an absence of a noninvasive, convenient, and accurate model to guide clinicians in IAT decision-making.

We hypothesize that ML models incorporating postoperative variables will outperform preoperative only models and existing staging systems in predicting 5-years mortality. The present study aimed to develop and validate an ML model based on clinical data from uHCC patients who received IAT, with the goal of providing guidance for clinicians in therapeutic decision-making and surveillance strategy selection.

Methods

This retrospective, multi-center study was carried out in line with the principles outlined in the 1975 Declaration of Helsinki. The study protocol was approved by the Institutional Review Board of XXX (No. XX). Owing to the retrospective nature of the study, written informed consent was exempted. To preserve patient privacy and data confidentiality, patient data have been deidentified before analysis. Clinical data were extracted from an internally developed database, and the study was reported in adherence to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline.[21]

Study Design

HCC diagnosis was confirmed in accordance with the guidelines established by the European Association for the Study of Liver (EASL) and the American Association for the Study of Liver Disease (AASLD).[22,23] From April 2008 to October 2022, clinical data from a total of 7829 uHCC patients who underwent initial IAT were retrospectively collected from 15 tertiary hospitals in China (eTable 1). We performed TACE, HAIC, and combined targeted immunotherapy at each participating hospital, with the specific steps. The detailed descriptions of IAT procedures, criteria for discontinuing the protocol, and combination with targeted-immunotherapy are available in the Supplemental information E 1.1–1.3. The inclusion criteria were as follows: *(a)* age between 18 and 75 years; *(b)* Eastern Cooperative Oncology Group (ECOG) performance status < 2; *(c)* Child-Turcotte-Pugh (CTP) class A or B liver function. The exclusion criteria included: *(a)* prior treatment before IAT; *(b)* concurrent other malignancies; *(c)* loss to follow-up for more than 6 months. The patient enrollment pathway is illustrated in Figure 1A.

Follow-Up Protocol and Endpoint Definition

All patients underwent regular follow-up after IAT, with serum alpha-fetoprotein (AFP) testing and imaging examinations conducted 1 month after IAT and every 3 to 6 months thereafter until death or the final follow-up date (October 31, 2025). The objective response (OR) to the first IAT was assessed by two radiologists in accordance with the modified Response Evaluation Criteria in Solid Tumours (mRECIST).[24] Early recurrence (ER) was defined as the presence of suspected lesions exhibiting arterial phase enhancement and portal venous phase washout on contrast-enhanced imaging within 1 year[25.] Overall survival (OS) was defined as the time from the date of initial IAT to the date of death or the final

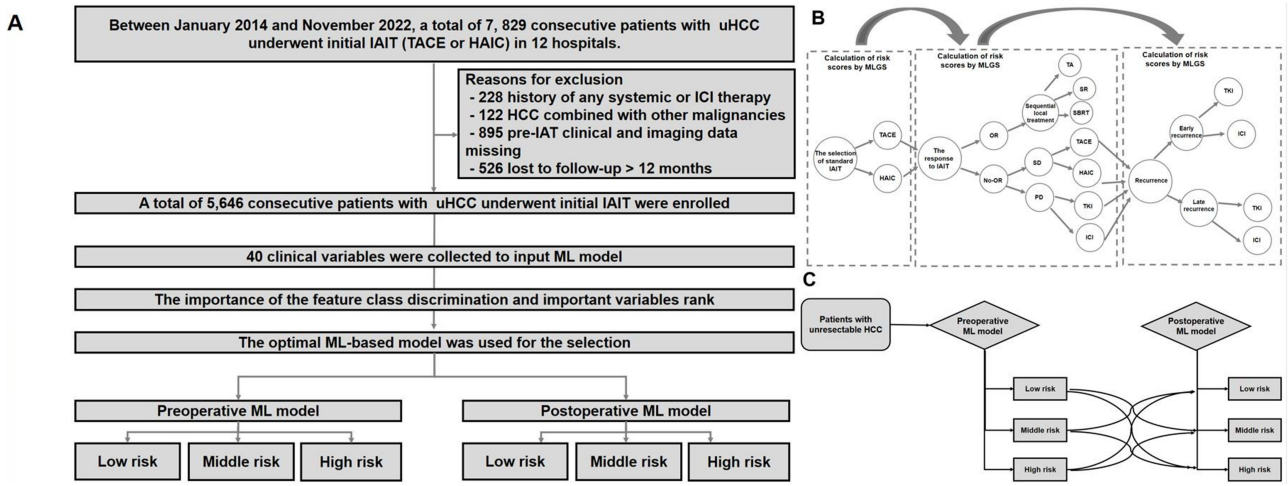

Figure 1 Enrollment pathway of unresectable HCC patients who underwent various IAT schemes. (**A**) HCC patients were enrolled to build an MLGS model; (**B**) both the preoperative and postoperative variables were input in the MLGS-pos model according to tumor progression and therapeutic procotols and the death risk score was calculated; (**C**) the clinical decision-making for IATIs based on death risk change from MLGS-pre model to MLGS-pos model.
Abbreviations: HCC, hepatocellular carcinoma; IAT, intra-arterial interventional therapy; ML, machine learning.

follow-up. The observation window for clinical indicators was set at the time of initial IAT, and the prediction window was defined as 1, 3, and 5 years after initial IAT. For patients with censored data within 5 years, their outcomes were classified as "nonevent" in all models.

Machine Learning-Guided System Construction

Data Collection

A total of 30 preoperative and 6 postoperative variables were collected for the development of ML models. The abbreviations and detailed descriptions of these 36 clinical variables are presented in the eTable 2. Five tumor staging systems—Barcelona Clinic Liver Cancer (BCLC), American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM), China Clinic Liver Cancer Staging (CNLC), Japan Society of Hepatology (JSH), and EASL—and several clinical evaluation scores, including albumin-bilirubin (ALBI) grade, platelet-albumin-bilirubin (PALBI), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and model for end-stage liver disease (MELD), were also collected to compare the predictive performance of the ML models.

Selection of Clinical Variables

Variables with a missing data rate exceeding 20% (including pathological variations) were excluded from the model. Five representative supervised ML algorithms—XGBoost, CatBoost, GBDT, LGBM, and random forest (RF)—were applied and compared in this study (Supplemental Information E1.4). The ML model with the optimal discriminative capability was utilized to develop a simplified restricted system for guiding IAT scheme selection (hereafter referred to as the ML-guided system [MLGS]), which integrated preoperative and postoperative clinical information. The code parameters for ML model development are provided in the eTables 3 and 4. To optimize model performance and evaluate its generalization capability, a random hyperparameter search combined with 5-fold cross-validation was employed to identify the optimal hyperparameter set. Subsequently, the final model was trained utilizing the entire training dataset based on the best-performing hyperparameters identified. Feature importance was then analyzed from this final model. In the cohort of the Sun Yat-sen University Cancer Center, HCC patients were divided into training and internal testing datasets at a ratio of 7:3, and the training datasets was optimized using 5-fold cross-validation. During the same period, data from another 14 tertiary hospitals were used as an external testing datasets.

Dynamic Prognostic Risk Based on the ML Model

In the current study, the MLGS was employed to assess the dynamic prognostic risk of HCC patients receiving IAT and to guide clinicians in the timely adjustment of treatment and follow-up monitoring schemes by stratifying the 5-year mortality risk as a long-term outcome after IAT. The guidance process of the MLGS primarily comprises two components: first, the preoperative ML model assists clinicians and patients in evaluating preoperatively the effectiveness of IAT for uHCC based on significant mortality risk stratification; second, the postoperative ML model helps clinicians determine whether to adjust sequential treatment schemes during follow-up based on IAT response and tumor recurrence.

The preoperative ML model assigns a clear death risk score to uHCC patients scheduled to receive IAT. However, when postoperative variables—such as IAT response, conversion therapy modalities, and tumor recurrence—are sequentially input into the preoperative ML model, the risk score of uHCC patients changes accordingly (Figure 1B). Concurrently, the original death risk stratification may shift (eg, from low risk to high risk), enabling clinicians to adjust therapeutic and monitoring strategies during this dynamic process (Figure 1C).

Statistical Analysis

All statistical analyses were performed using R software version 3.6.3 (http://www.r-project.org/). The performance of the ML models was evaluated using the area under the receiver operating characteristic curve (AUC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and F1 score. To improve the applicability of the predictive model by simplifying input variables, a stepwise Cox proportional hazards model was built based on the top important variables derived from the optimal ML model.

The discriminative ability of the models and different tumor staging systems was measured using the concordance index (C-index) and compared via the DeLong test. Time-dependent receiver operating characteristic curves and AUC values at each time point from 12 to 60 months were used to assess OS prediction accuracy at different time points. The prediction error of the models was evaluated using the integrated Brier score (IBS). To interpret the causal relationship of the MLGS, the Shapley Additive exPlanations (SHAP) method was used to explain the prediction results of each ML model.[26,27]

Survival curves were generated using the Kaplan-Meier method and compared using the Log rank test. All statistical tests were two-sided, and a P value < 0.05 was considered statistically significant.

Results

Baseline Characteristics

Finally, a total of 5646 eligible patients (4930 male, 716 female; median age, 57 years, IQR: 37–63 years) were included and divided into a derivation cohort consisted of training datasets (TD) (51%, N = 3367) and internal testing datasets (ITD) (13%, N = 1130) at the Sun Yat-sen University Cancer Center and an external testing datasets (ETD) (36%, N = 1129) at the another 14 hospitals. Overall, the mortality rate among patients who received IAT prior to the final follow-up endpoint was 61.5% (3360 out of 5646 patients). Among all deaths, 61.6% (2018/3367) occurred death in the TD, 62.0% (683/1130) in the ITD, and 60.7% (659/1129) in the ETD, with no statistically significant differences between these rates ($P = 0.913$). Table 1 presents a comparison of baseline characteristics across the three datasets. All preoperative variables were well-balanced and homogeneous among the datasets (all $P > 0.05$). A feature heatmap was generated through correlation analysis of 35 risk factors, resulting in a correlation coefficient matrix (eFigure 1). In the ITD and ETD cohorts, OR to the first IAT, sequential local treatment, targeted-immunotherapy, lymphocyte count, and creatinine level were strongly negatively correlated with patient outcomes. Additionally, we compared the performance of the five ML models in predicting 1- and 3-year mortality across the three datasets (eTables 5 and 6), and the predictive capability of the models for 1- and 3-year mortality was consistent with that for 5-year mortality.

Comparison of the Predictive Performance of the ML Models

Table 2 summarizes the number of predictors and the AUC, SEN, SPE, PPV, NPV, and F1 score of each ML model based on clinical variables for predicting 5-year mortality. For preoperative ML models, the CatBoost model achieved the

Table 1 The Baseline Characteristics of uHCC Patients Underwent IATs in Three Datasets

Variables	TD (n= 3386)	ITD (n=1130)	ETD (n=1129)	*P* value
Demographic and history (n =10)				
Age (years)[a]	53.19 ± 11.81	53.73 ± 11.56	52.69 ±12.46	0.116
BMI (kg/cm^2)[a]	22.58 ± 4.62	22.69 ± 3.58	22.54 ± 4.30	0.658
Gender[b]				0.971
Male	2851 (84.20%)	948 (83.89%)	950 (84.15%)	
Female	535 (15.80%)	182 (16.11%)	179 (15.85%)	
ECOG[b]				0.360
PS 0	3195 (94.36%)	1059 (93.72%)	1074 (95.13%)	
PS 1	191 (5.64%)	71 (6.29%)	55 (4.87%)	
Smoking[b]				<0.001
Absence	2871 (84.79%)	976 (86.37%)	1014 (89.81%)	
Presence	515 (15.21%)	154 (13.63%)	115 (10.19%)	
Drinking[b]				0.083
Absence	3251 (96.01%)	1086 (96.11%)	1100 (97.43%)	
Presence	135 (3.99%)	44 (3.89%)	29 (2.57%)	
Comorbidities[b]				0.240
Absence	2963 (87.51%)	982 (86.90%)	1006 (89.11%)	
Presence	423 (12.49%)	148 (13.10%)	123 (10.89%)	
HBV[b]				0.256
Absence	184 (5.43%)	70 (6.19%)	67 (5.93%)	
Presence	3202 (94.57%)	1060 (93.81%)	1062 (94.07%)	
Cirrhosis[b]				0.818
Absence	273 (8.06%)	88 (7.79%)	96 (8.50%)	
Presence	3113 (91.94%)	1042 (92.21%)	1033 (91.50%)	
Ascites[b]				0.982
Absence	3062 (90.43%)	1024 (90.62%)	1021 (90.43%)	
Presence	324 (9.57%)	106 (9.38%)	108 (9.57%)	
Child-Pugh class[b]				0.806
A	3002 (88.66%)	997 (88.23%)	1006 (89.11%)	
B	384 (11.34%)	133 (11.77%)	123 (10.89%)	
Tumor characteristics (n = 5)				
Pseudo-capsulated				0.944
Absence	2780 (82.10%)	929 (82.21%)	932 (82.55%)	
Presence	606 (17.90%)	201 (17.79%)	197 (17.45%)	
Maximum diameter of tumors (mean (SD)) (cm)[c]	9.5 ±1.2	9.5 ± 1.5	9.6 ±1.2	0.975
Number of tumors[b]				0.613
1–3	1698 (50.15%)	583 (51.59%)	580 (51.37%)	
>3	1688 (49.85%)	547 (48.41%)	549 (48.63%)	
Vascular invasion[b]				0.411
Absence	2156 (63.67%)	704 (62.30%)	724 (64.13%)	
Presence	1230 (36.33%)	426 (37.70%)	405 (35.87%)	
Metastasis[b]				0.774
Absence	2585 (76.34%)	873 (77.26%)	859 (76.09%)	
Presence	801 (23.66%)	257 (22.74%)	270 (23.91%)	
Laboratory findings (n =15)				
AFP (ug/mL)[b]				0.518
<400	1475 (43.56%)	510 (45.13%)	509 (45.08%)	
>400	1911 (56.44%)	620 (54.87%)	620 (54.92%)	
DCP (ug/mL)[b]				0.185
<400	1651 (48.76%)	364 (32.21%)	347 (30.74%)	
>400	2035 (51.24%)	766 (67.79%)	782 (69.26%)	

(*Continued*)

Table 1 (Continued).

Variables	TD (n= 3386)	ITD (n=1130)	ETD (n=1129)	*P* value
ALT (U/L)[c]	45.6	52.6	55.1	0.548
AST (U/L)[c]	65.2	91.5	90.6	0.812
ALB (g/L)[a]	37.5 ± 5.5	37.7 ± 4.3	37.6 ± 4.5	0.825
TBIL (μmol/L)[c]	21.2	18.6	20.3	0.478
PT(s)[a]	12.6 ± 1.4	12.5± 1.3	12.3 ± 1.8	0.432
INR[a]	1.09 ± 0.13	1.10±0.11	1.09±0.16	0.167
Platelet counts ($\times 10^9$)[c]	159	178	157	0.900
Creatinine (μmol/L)[a]	71.3 ± 8.36	71.6 ± 7.77	72.4 ± 5.93	0.756
CRP (μmol/L)[c]	15.2	15.6	13.8	0.447
Lymphocyte[a]	1.5 ± 0.2	1.5 ± 0.3	1.6 ± 0.2	0.992
Neutrophils[a]	4.5 ± 1.1	4.6 ± 1.4	4.3 ± 1.1	0.209
Treatment parameter (n =6)				
IAIT modalities[b]				0.705
TACE	2151 (63.53%)	728 (64.42%)	709 (62.73%)	
HAIC	1235 (36.47%)	402 (35.58%)	420 (37.27%)	
Combination with TKI and ICI[b]				0.740
Absence	2130 (62.91%)	688 (60.88%)	700 (62.00%)	
Presence				
TKI or ICI	589 (17.40%)	202 (17.87%)	194 (17.18%)	
TKI plus ICI	667 (19.69%)	240 (21.25%)	235 (20.81%)	
Sequential local therapy[b]				0.082
Absence	2230 (65.86%)	769 (68.05%)	749 (66.34%)	
Presence				
SR	389 (11.49%)	115 (10.18%)	154 (13.64%)	
TA	647 (19.11%)	208 (18.41%)	182 (16.12%)	
SBRT	120 (3.54%)	38 (3.36%)	44 (3.90%)	
The response of first IAIT[b]				<0.001
OR	1035 (30.57%)	169 (14.96%)	203 (17.98%)	
Non OR	2351 (69.43%)	961 (85.04%)	926 (82.02%)	
Early recurrence[b]				0.844
Absence	1137 (33.65%)	383 (34.07%)	382 (33.90%)	
Presence	2242 (66.35%)	741 (65.93%)	745 (66.10%)	
Recurrence type[b]				0.949
Absence	884 (26.11%)	295 (26.11%)	291 (25.78%)	
Presence	2501 (73.86%)	835 (73.89%)	838 (74.22%)	
Biomarkers and stages (n=6)				
ALBI grade[b]				0.437
1	1533 (45.29%)	507 (44.87%)	527 (46.68%)	
2	1795 (54.71%)	598 (55.23%)	588 (53.32%)	
PALBI grade[b]				0.899
1	1201 (35.48%)	399 (35.31%)	416 (36.85%)	
2	1383 (40.86%)	467 (41.33%)	446 (39.50%)	
3	801 (23.66%)	264 (23.36%)	267 (23.65%)	
NLR[b]				0.076
≤ 4	3369 (99.50%)	1120 (99.12%)	1124 (99.56%)	
> 4	17 (0.50%)	10 (0.88%)	5 (0.44%)	
PLR[b]				0.993
≤ 112	2138 (63.14%)	712 (63.01%)	711 (62.98%)	
> 112	1248 (36.86%)	418 (36.99%)	418 (37.02%)	

(*Continued*)

Table 1 (Continued).

Variables	TD (n= 3386)	ITD (n=1130)	ETD (n=1129)	*P* value
MELD[b]				0.274
≤ 25	3369 (99.50%)	1120 (99.12%)	1124 (99.56%)	
> 25	17 (0.50%)	10 (0.88%)	5 (0.44%)	
BCLC grade[b]				0.692
A	649 (19.17%)	210 (18.58%)	235 (20.81%)	
B	1142 (33.73%)	378 (33.45%)	367 (32.51%)	
C	1595 (47.11%)	542 (47.96%)	527 (46.68%)	

Notes: Data are number of patients; data in parentheses are percentage unless otherwise indicated. Data in bracket was percent of patients. [a]Data are mean ± standard deviation; [b]Data are frequencies with percentages in parentheses; [c]Date are median value; P-value < 0.05 indicated a significant difference.

Abbreviations: ITA, intra-arterial therapy; TACE, transarterial chemoembolization; HAIC, hepatic arterial infusion chemotherapy; TD, training datasets; ITD, internal test datasets; ETD, external test datasets; uHCC, unresectable hepatocellular carcinoma; HBV, hepatitis B virus;ALT, alanine aminotransferase; AST:aspartate aminotransferase; AFP, α-fetoprotein; DCP, des-γ-carboxy prothrombin; ALB:Albumin; TBIL:Total bilirubin; BIL:Bilirubin; Cr:Creatinine; CRP, C reactive protein PT:Prothrombin time; INR, international normalized ratio; SR, surgical resection; TA, thermal ablation; SBRT, stereotactic body radiotherapy; TKI, Tyrosine kinase inhibitors; ICI, immune checkpoint inhibitors.

Table 2 The Performance Comparison of Five ML Models Based on Preoperative and Postoperative Variables

Models	Cohorts	No. of Predictors in Model	AUC	NPV	PPV	SENS	SPEC	F1 Score
CatBoost[a]	TD	23	0.7771	0.8159	0.5616	0.7432	0.6623	0.6397
	ITD		0.7429	0.8251	0.5258	0.7855	0.5882	0.6300
	ETD		0.7346	0.8010	0.5453	0.7236	0.6485	0.6219
GBDT[a]	TD	22	0.7885	0.8009	0.6023	0.6854	0.7366	0.6411
	ITD		0.7418	0.7737	0.5784	0.6313	0.7325	0.6037
	ETD		0.7310	0.8182	0.5316	0.7692	0.6050	0.6287
LGBM[a]	TD	25	0.7969	0.8613	0.5483	0.8339	0.6002	0.6616
	ITD		0.7364	0.8201	0.5054	0.7928	0.5490	0.6173
	ETD		0.7328	0.8081	0.5306	0.7500	0.6134	0.6215
RF[a]	TD	26	0.7502	0.8121	0.5264	0.7608	0.6016	0.6223
	ITD		0.7215	0.8143	0.4940	0.7928	0.5280	0.6087
	ETD		0.7180	0.8110	0.4962	0.7861	0.5350	0.6084
XGBoost[a]	TD	21	0.7979	0.8334	0.5793	0.7681	0.6754	0.6605
	ITD		0.7427	0.8123	0.5313	0.7566	0.6120	0.6243
	ETD		0.7322	0.8077	0.5325	0.7476	0.6176	0.6220
CatBoost[b]	TD	19	0.8955	0.8894	0.7183	0.8266	0.8113	0.7687
	ITD		0.8838	0.9215	0.6403	0.8964	0.7073	0.747
	ETD		0.8561	0.8502	0.6993	0.7548	0.8109	0.726
GBDT[b]	TD	25	0.9089	0.9102	0.7254	0.8628	0.8099	0.7881
	ITD		0.8868	0.9291	0.6358	0.9084	0.6975	0.748
	ETD		0.8588	0.8895	0.6388	0.8462	0.7213	0.728
LGBM[b]	TD	34	0.9087	0.9151	0.7226	0.8716	0.8052	0.7901
	ITD		0.8866	0.933	0.6402	0.9133	0.7017	0.7527
	ETD		0.8590	0.8667	0.664	0.7981	0.7647	0.7249
RF[b]	TD	16	0.8850	0.8874	0.6905	0.8291	0.7837	0.7535
	ITD		0.8786	0.8774	0.6836	0.812	0.7815	0.7423
	ETD		0.8492	0.8746	0.6489	0.8173	0.7423	0.7234
XGBoost[b]	TD	20	0.9042	0.8985	0.7382	0.8395	0.8267	0.7856
	ITD		0.8865	0.9340	0.6344	0.9157	0.6933	0.7495
	ETD		**0.8619**	0.8973	0.6336	0.8606	0.7101	0.7299

Notes: The bold values represent the optimal AUC value in the external testing cohort. [a]Represent preoperative ML models and [b]Represent postoperative ML. models.

Abbreviations: ML, machine learning; TD, training datasets; ITD, internal datasets; ETD, external datasets; AUC, areas under receiver operating characteristic curve; ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value; SENS, sensitivity; SPEC, specificity.

optimal predictive performance, with an AUC of 0.777 (95% CI, 0.712–0.815) in the TD, 0.743 (95% CI, 0.705–0.811) in the ITD, and 0.735 (95% CI, 0.68–0.782) in the ETD when trained using the 23 most important variables identified by the CatBoost importance score. Among postoperative ML models, the XGBoost model demonstrated the best predictive performance. When trained using the 20 most important variables identified by the XGBoost importance score, the XGBoost model achieved an AUC of 0.904 (95% CI, 0.876–0.963) in the TD, 0.886 (95% CI, 0.845–0.927) in the ITD, and 0.862 (95% CI, 0.818–0.932) in the ETD (Figure 2A).

Interpretation Methods for the ML Models

As depicted in the SHAP summary plots (Figure 2B), the contribution of each feature to the ML model was evaluated using average SHAP values and presented in descending order. Additionally, local explanation was employed to analyze how a specific prediction was made for an individual by incorporating personalized input data. Figure 2C illustrates the impact of clinical variables on the 5-year mortality rate with a specific probability. Furthermore, a heatmap interpreting the outcomes of uHCC patients who received IAT in the ETD is shown in Figure 2D, and the mean SHAP values were compared among the three mortality risk subgroups (Figure 2E). Our results demonstrated that ER was most strongly associated with poor prognosis (eFigure 2). Non-OR to the first IAT was the second most significant risk factor for mortality and was linked to a higher risk of poor prognosis than OR (eFigure 3). Patients who received IAT alone had a significantly higher risk of adverse outcomes than those who received IAT combined with sequential local treatment (eFigure 4) or targeted-immunotherapy (eFigure 5). Patients with an AFP concentration > 400 ng/mL had a higher risk of poor prognosis than those with an AFP concentration ≤ 400 ng/mL (eFigure 6). Patients with a tumor diameter > 10 cm or more than 3 lesions had a poorer prognosis than those with a tumor diameter ≤ 10 cm or 1–3 lesions, respectively (eFigures 7 and 8). The relationship between laboratory findings (including creatinine) and SHAP values exhibited an

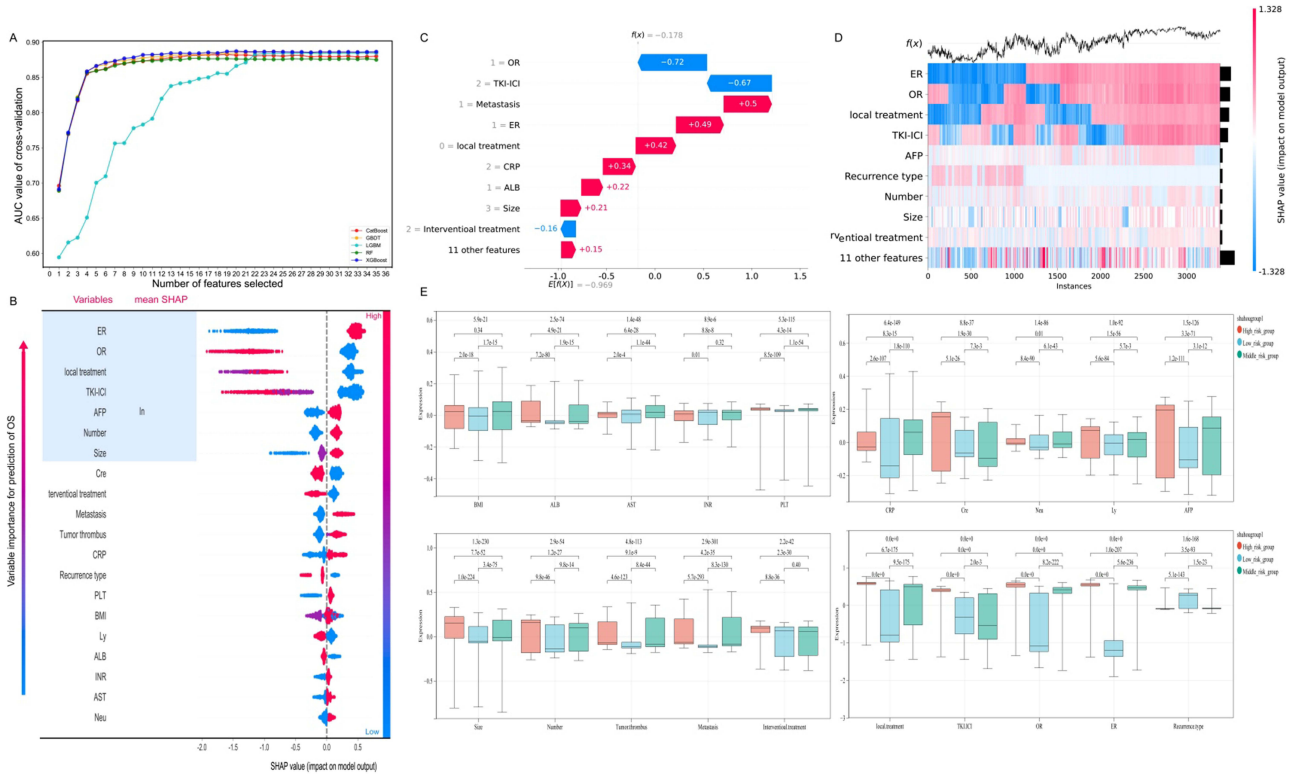

Figure 2 The death risk possibility of prediction based on ML models were explained using SHAP algorithm. (**A**) the XGBoost model trained using the 20 most important variables measured by the AUC values; (**B**) SHAP summary plots show the contributions of the feature to the ML model; (**C**) SHAP bar chart represented this clinical variables towards the 5-years death rate with a probability of specific percentage; (**D**) a heatmap of interpretation for outcomes of uHCC patients underwent IAT; (**E**) the mean SHAP value were compared between three death risk subgroups.

S-shaped curve with distinct turning points (eFigure 9). The HAIC protocol without metastasis was associated with better OS than the TACE protocol or HAIC with metastasis (eFigures 10 and 11).

Prognostic Risk Stratification of the MLGS

Stratifying prognoses for specific patient populations is vital for establishing criteria for their inclusion in clinical trials and therapeutic interventions (Figure 3A). Based on the cutoff values (22.34, 76.25) from both the preoperative MLGS (MLGS-pre) and postoperative MLGS (MLGS-post) models, HCC patients were divided into three risk groups: low-risk, middle-risk, and high-risk. The person-months in each risk group are shown in Figure 3B, indicating that patients in the low-risk group had better OS than those in the high-risk group. According to the MLGS-pre model, the low-risk group in the TD had significantly higher 1-, 3-, and 5-year cumulative OS rates (87.8%, 62.1%, and 49.1%, respectively) compared to the middle-risk group (63.2%, 33.4%, and 25.4%, respectively) and the high-risk group (35.9%, 13.0%, and 8.9%, respectively) ($P < 0.001$). Similar results were observed in the ITD and ETD cohorts using the preoperative model (Figure 3C–E). According to the MLGS-post model, the 1-, 3-, and 5-year cumulative OS rates in the low-risk group in the TD were 98.3%, 81.7%, and 65.8%, respectively, while those in the middle-risk group were 68.0%, 22.1%, and 13.5%, respectively—both significantly higher than those in the high-risk group (20.5%, 1.7%, and 0%) ($P < 0.001$). Similar results were obtained in the ITD and ETD cohorts using the postoperative model (Figure 3F–H).

Comparison of the Predictive Accuracy Between MLGS, Tumor Stages, and Biomarkers

To evaluate the discriminative capability of the ML models, we compared their prognostic performance with that of ALBI grade, PALBI grade, PLR, NLR, MELD grade, and other staging systems (BCLC, JSH, AJCC TNM, EASL, and CNLC). Time-dependent AUC analysis revealed that the MLGS had superior OS prediction performance compared to the staging systems at various time points in both study cohorts (all Log rank test, $P < 0.001$) (Figures 3I–K). The prediction error curves of all models and staging systems are shown, and the IBS values for MLGS-post were 0.17 in the TD, 0.14 in the ITD, and 0.22 in the ETD (Figures 3L–N). The bootstrapped calibration curves plotted with 5-years mortality was well matched with the idealized 45°line for the MLGS in the three datasets (eFigure 12). Decision curve analysis (DCA) graphically indicated that the MLGS provided a larger benefit across the range of reasonable threshold probabilities in the three datasets (eFigure 12).

Clinical Value of the MLGS

The MLGS developed in this study comprises two ML models: the preoperative and postoperative models. The preoperative model provides a prognostic risk score for individual uHCC patients before IAT and classifies them into low-risk, middle-risk, and high-risk groups. Subsequently, the six postoperative variables are input into the preoperative model, and the risk coefficient changes with the input of positively and negatively correlated variables. During follow-up after IAT, inputting the time and status of tumor recurrence into the preoperative ML model leads to a significant increase in the mortality hazard. Furthermore, the IAT scheme plays a crucial role in improving survival benefits—for instance, selecting between HAIC and TACE, or combining IAT with targeted-immunotherapy and sequential local treatment. Based on the dynamic changes in mortality risk following the input of these variables, the MLGS can guide clinicians in continuously adjusting treatment schemes during the IAT process.

Discussion

IAT possesses considerable potential as an effective treatment for uHCC, with growing evidence highlighting its significant role in improving survival and quality of life for HCC patients.[4,28,29] Accurate and rational IAT decision-making is critical for achieving long-term survival benefits. Therefore, we developed the MLGS to assist clinicians in formulating comprehensive IAT plans throughout the treatment process. The MLGS can effectively classify HCC patients who received IAT into three distinct risk groups with significantly different long-term OS durations. Additionally, it can visually display the dynamic changes in mortality risk over time, taking into account variations in

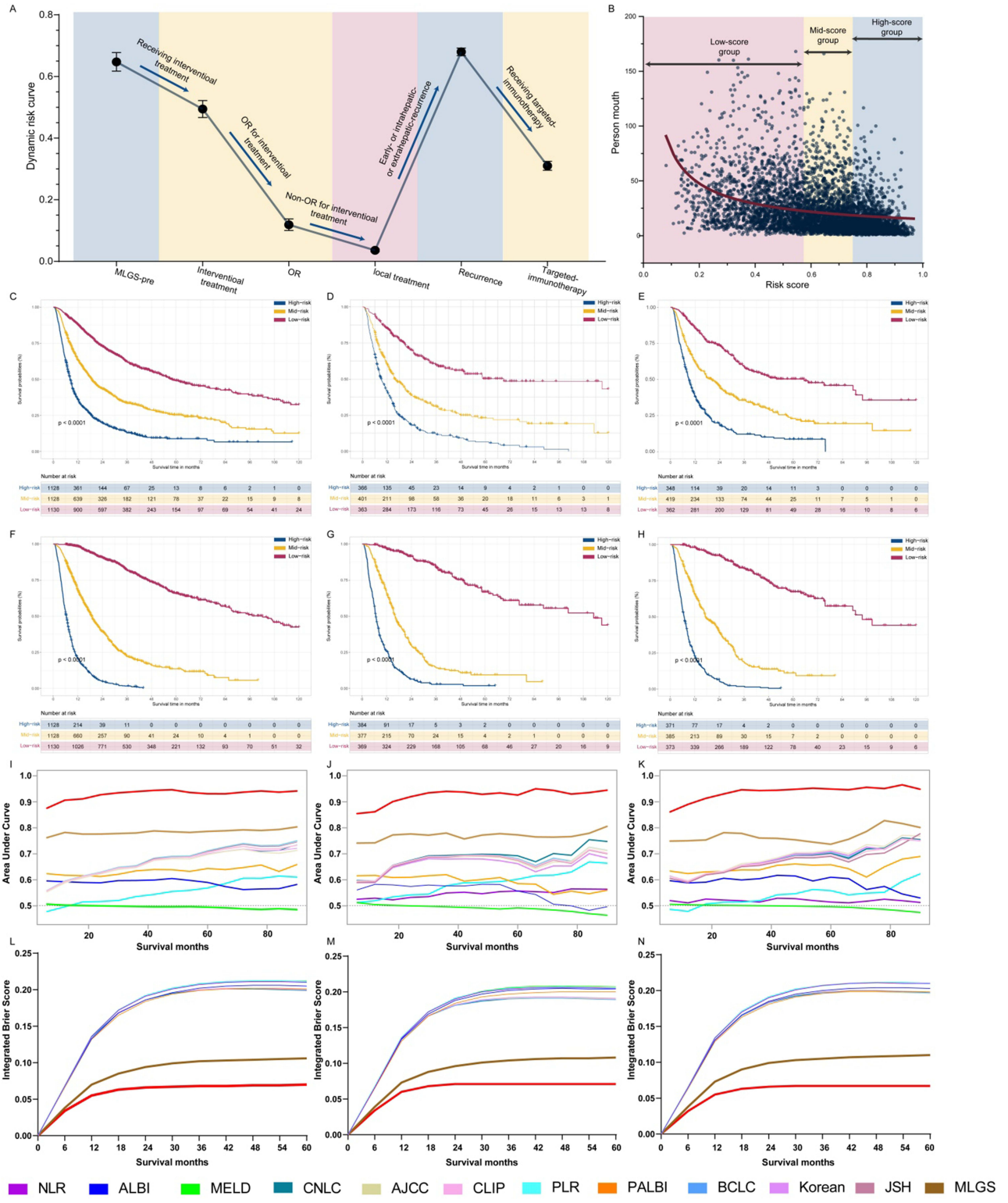

Figure 3 Risk strata for predicting the prognosis of HCC patients who underwent IATs. (**A**) Segmenting prognoses for particular patient populations were shown using line chart; (**B**). The uHCC patients were categorized into three risk group and the person-months were showed in different risk group; the survival outcomes of HCC patients received the IAT scheme between three risk subgroups based on MLGS-pre model (**C**) OS, (**D**) PFS, and (**E**) EPFS the survival outcomes of HCC patients received the IAT scheme between three risk subgroups based on MLGS-post model (**F**) OS, (**G**) PFS, and (**H**) EPFS. Model discrimination measured with time-dependent areas under the receiver operating characteristic curves at various time points for the training (**I**), internal testing (**J**) and external testing (**K**) cohorts. The integrated Brier score of models for the training (**L**), internal testing (**M**) and external testing (**N**) cohorts.

tumor status and treatment strategies. Thus, the MLGS may facilitate personalized decision-making and reduce future mortality rates, treatment-related side effects, and economic burden for a larger number of uHCC patients receiving IAT.

This study has several strengths and innovations. First, the MLGS was developed and externally validated using clinical data from 5646 patients (from 15 hospitals in China) and 35 available clinical variables, with a relatively long median follow-up time (4.2 years). This ensures the comprehensiveness, robustness, and accuracy of the MLGS from multiple perspectives. Second, the MLGS was developed using a wide range of variables, including demographic characteristics, tumor features, laboratory test results, and treatment schemes. Preoperative and postoperative variables were sequentially input based on the timeline before and after IAT, enabling visualization of the dynamic trajectory of mortality risk. Third, five representative ML algorithms were compared, and the XGBoost model—with the most reasonable variable input—was selected to develop the MLGS for prognostic risk stratification. Fourth, the predictive performance of the MLGS was significantly superior to that of other internationally recognized clinical and biological markers, which are considered important in clinical practice.

Given the lack of interpretability of ML processes, the SHAP method was utilized to explain the relationship between clinical variables and mortality risk. These variables mainly fell into two categories: those associated with increased mortality risk (eg, high AFP levels, non-OR to the first IAT, ER) and those associated with reduced mortality risk or improved survival benefits (eg, local treatment after IAT, combination of IAT with targeted-immunotherapy). These findings are consistent with the recent global Phase III randomized clinical trial (EMERALD-1), which demonstrated better progression-free survival with the combination of bevacizumab, durvalumab, and TACE compared to TACE monotherapy in patients with intermediate HCC.[29] Previous studies have also clarified that the tumor microenvironment, including proteoglycans, immune cells, and hypoxia, plays a critical role in tumor initiation and progression during the invasion-metastasis cascade.[30–32]

Previous research has focused solely on decision-making for a single treatment option. For example, Liu et al constructed two models using semantic information and ultrasound image data to predict ER in HCC patients undergoing TA or SR and identify suitable candidates.[33] Ding et al developed a hybrid model to accurately predict the probability of ER after TA or SR, providing reliable evidence for optimal treatment decision-making in patients with single 3- to 5-cm HCCs.[34] These decision support models were individual-based and cross-stratified twice according to treatment type. In contrast, the MLGS developed in this study is based on a dynamic risk trajectory formed by the input of preoperative clinical data, postoperative tumor status, and IAT schemes. Our results showed that tumor burden, tumor immunity-related biomarkers, and liver function affect the prognosis of uHCC patients before IAT.However, with the selection of IAT modality, achievement of initial OR, and application of subsequent treatment schemes, the ability to predict mortality risk was significantly improved compared to the preoperative model. By leveraging the dynamic changes in risk scores generated by the MLGS, clinicians can more accurately formulate personalized IAT regimens.

The MLGS incorporates simple, easily accessible clinical and radiological variables that can be used to identify high-risk HCC patients. High-risk patients require prolonged and intensive monitoring due to their elevated mortality risk, along with additional systemic medications. The MLGS effectively classifies HCC patients into different subgroups based on risk levels and helps identify the most suitable candidates for IAT. In the TD, 33% of uHCC patients were classified as high-risk, with a 5-year OS rate of 0%, while the low-risk group had a 5-year OS rate of 49.1%. The MLGS is clinically significant as it enables the identification of a subset of IAT-treated patients at high risk of mortality who may benefit from targeted interventions. For example, if patients receiving HAIC are classified as high-risk by the MLGS, clinicians may modify the treatment plan to TACE or implement a proactive targeted-immunotherapy and monitoring strategy with shorter intervals.

This study possesses several inherent limitations. First, participants were enrolled across 15 Chinese medical institutions, each with distinct protocols, surgical techniques, and medication regimens for TACE and HAIC. These institutional variations may have introduced heterogeneity, potentially impacting the generalizability of the findings. Second, the pharmacological diversity in TACE and HAIC regimens across centers may have introduced confounding variables affecting the final outcomes. Furthermore, the clinical characteristics of the study cohort differ substantially from Western populations, where patients typically present with smaller tumor burdens and alcoholic cirrhosis as the primary etiology of HCC, limiting the broad applicability of our results to Western settings. Finally, given the established

role of radiomics in predicting prognosis for patients undergoing IAT, future investigations should integrate radiomic features into our ML models to enhance predictive precision.

In conclusion, the MLGS accurately stratifies prognostic risk and may assist in identifying high-risk patients who could benefit from more intensive surveillance or adjunctive therapies. Prospective studies are needed to evaluate its clinical utility.

Abbreviations

HCC, hepatocellular carcinoma; ML, machine learning; MLGS, machine learning guided system; TACE, Transarterial chemoembolization; HAIC, hepatic arterial infusion chemotherapy; IATs, intra-arterial interventional therapies; RF, Random Forest; XGBoost, eXtreme Gradient Boosting; CatBoost, Categorical Gradient Boosting; GBDT, Gradient Boosting Decision Tree; LGBM, Light Gradient Boosting Machine; SHAP, SHapley Additive exPlanations; OS, overall survival; TKI, tyrosine kinase inhibitors; ICI, immune checkpoint inhibitors; HR, hazard ratios; CI, confidence intervals.

Data Sharing Statement

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

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.

Disclosure

The authors declare no conflicts of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30. doi:10.3322/caac.21590
2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301–1314. doi:10.1016/S0140-6736(18)30010-2
3. Vitale A, Trevisani F, Farinati F, Cillo U. Treatment of hepatocellular carcinoma in the Precision Medicine era: from treatment stage migration to therapeutic hierarchy. *Hepatology*. 2020. doi:10.1002/hep.31187
4. Lyu N, Kong Y, Mu L, et al. Hepatic arterial infusion of oxaliplatin plus fluorouracil/leucovorin vs. sorafenib for advanced hepatocellular carcinoma. *J Hepatol*. 2018;69(1):60–69. doi:10.1016/j.jhep.2018.02.008
5. Shen L, Xi M, Zhao L, et al. Combination therapy after TACE for hepatocellular carcinoma with macroscopic vascular invasion: stereotactic body radiotherapy versus sorafenib. *Cancers*. 2018;10(12). doi:10.3390/cancers10120516
6. Shimose S, Iwamoto H, Tanaka M, et al. Alternating lenvatinib and trans-arterial therapy prolongs overall survival in patients with inter-mediate stage hepatocellular carcinoma: a propensity score matching study. *Cancers*. 2021;13(1). doi:10.3390/cancers13010160
7. Yuan H, Lan Y, Li X, Tang J, Liu F. Large hepatocellular carcinoma with local remnants after transarterial chemoembolization: treatment by sorafenib combined with radiofrequency ablation or sorafenib alone. *Am J Cancer Res*. 2019;9(4):791–799.
8. English K, Brodin NP, Shankar V, et al. Association of addition of ablative therapy following transarterial chemoembolization with survival rates in patients with hepatocellular carcinoma. *JAMA Network Open*. 2020;3(11):e2023942. doi:10.1001/jamanetworkopen.2020.23942
9. She WH, Cheung TT. Bridging and downstaging therapy in patients suffering from hepatocellular carcinoma waiting on the list of liver transplantation. *Transl Gastroenterol Hepatol*. 2016;1:34. doi:10.21037/tgh.2016.03.04
10. Fan J, Tang ZY, Yu YQ, et al. Improved survival with resection after transcatheter arterial chemoembolization (TACE) for unresectable hepatocellular carcinoma. *Dig Surg*. 1998;15(6):674–678. doi:10.1159/000018676
11. He YZ, He K, Huang RQ, et al. A clinical scoring system for predicting tumor recurrence after percutaneous radiofrequency ablation for 3 cm or less hepatocellular carcinoma. *Sci Rep*. 2021;11(1):8275. doi:10.1038/s41598-021-87782-y
12. Yang Y, Chen Y, Ye F, et al. Late recurrence of hepatocellular carcinoma after radiofrequency ablation: a multicenter study of risk factors, patterns, and survival. *Eur Radiol*. 2021;31(5):3053–3064. doi:10.1007/s00330-020-07460-x
13. Hu C, Song Y, Zhang J, et al. Preoperative gadoxetic acid-enhanced mri based nomogram improves prediction of early HCC recurrence after ablation therapy. *Front Oncol*. 2021;11:649682. doi:10.3389/fonc.2021.649682
14. Yoo J, Lee MW, Lee DH, Lee JH, Han JK. Evaluation of a serum tumour marker-based recurrence prediction model after radiofrequency ablation for hepatocellular carcinoma. *Liver Int*. 2020;40(5):1189–1200. doi:10.1111/liv.14406
15. Dagliati A, Marini S, Sacchi L, et al. Machine Learning Methods to Predict Diabetes Complications. *J Diabetes Sci Technol*. 2018;12(2):295–302. doi:10.1177/1932296817706375

16. Chang W, Liu Y, Xiao Y, et al. A machine-learning-based prediction method for hypertension outcomes based on medical data. *Diagnostics*. 2019;9(4). doi:10.3390/diagnostics9040178
17. Oh S, Park Y, Cho KJ, Kim SJ. Explainable machine learning model for glaucoma diagnosis and its interpretation. *Diagnostics*. 2021;11(3). doi:10.3390/diagnostics11030510
18. An C, Yang H, Yu X, et al. A machine learning model based on health records for predicting recurrence after microwave ablation of hepatocellular carcinoma. *J Hepatocell Carcinoma*. 2022;9:671–684. doi:10.2147/JHC.S358197
19. Li XJ, Chang L, Mi Y, et al. Integrated-omics analysis defines subtypes of hepatocellular carcinoma based on circadian rhythm. *J Integr Med*. 2025;23(4):445–456.
20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148–158.
21. Benson AB, D'Angelica MI, Abbott DE, et al. Guidelines insights: hepatobiliary cancers, version 2.2019. *J Natl Compr Canc Netw*. 2019;17(4):302–310. doi:10.6004/jnccn.2019.0019
22. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67(1):358–380. doi:10.1002/hep.29086
23. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182–236. doi:10.1016/j.jhep.2018.03.019
24. Yu H, Bai Y, Xie X, Feng Y, Yang Y, Zhu Q. RECIST 1.1 versus mRECIST for assessment of tumour response to molecular targeted therapies and disease outcomes in patients with hepatocellular carcinoma: a systematic review and meta-analysis. *BMJ Open*. 2022;12(6):e052294. doi:10.1136/bmjopen-2021-052294
25. Yang Y, Xin Y, Ye F, et al. Early recurrence after radiofrequency ablation for hepatocellular carcinoma: a multicenter retrospective study on definition, patterns and risk factors. *Int J Hyperthermia*. 2021;38(1):437–446. doi:10.1080/02656736.2020.1849828
26. Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause death in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med*. 2021;137:104813. doi:10.1016/j.compbiomed.2021.104813
27. Ma M, Liu R, Wen C, et al. Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms. *Eur Radiol*. 2022;32(3):1652–1662. doi:10.1007/s00330-021-08271-4
28. Xu L, Peng ZW, Chen MS, et al. Prognostic nomogram for patients with unresectable hepatocellular carcinoma after transcatheter arterial chemoembolization. *J Hepatol*. 2015;63(1):122–130. doi:10.1016/j.jhep.2015.02.034
29. Ueshima K, Ogasawara S, Ikeda M, et al. Hepatic Arterial Infusion Chemotherapy versus Sorafenib in Patients with Advanced Hepatocellular Carcinoma. *Liver Cancer*. 2020;9(5):583–595. doi:10.1159/000508724
30. Voron T, Colussi O, Marcheteau E, et al. VEGF-A modulates expression of inhibitory checkpoints on CD8+ T cells in tumors. *J Exp Med*. 2015;212(2):139–148. doi:10.1084/jem.20140559
31. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018;24(5):541–550. doi:10.1038/s41591-018-0014-x
32. Villanueva A, Hoshida Y, Battiston C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology*. 2011;140(5):1501–1512.e2. doi:10.1053/j.gastro.2011.02.006
33. Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liver Cancer*. 2020;9(4):397–413. doi:10.1159/000505694
34. Ding W, Wang Z, Liu FY, et al. A hybrid machine learning model based on semantic information can optimize treatment decision for naïve single 3–5-cm HCC patients. *Liver Cancer*. 2022;11(3):256–267. doi:10.1159/000522123