


Development and Validation of an Interpretable Machine Learning Model for Predicting Tumor Recurrence After Microwave Ablation in Small Hepatocellular Carcinoma

Yongfen Ma*, Lei Zhao*, Qing Zhao 

Department of Gastroenterology, Zibo Central Hospital, Zibo, 255000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Qing Zhao, Department of Gastroenterology, Zibo Central Hospital, No. 54, Youth League West Road, Zhangdian District, Zibo, Shandong, 255000, People's Republic of China, Tel +86-18560292715, Email qingzh0801@163.com

Objective: Tumor recurrence remains a major clinical challenge following microwave ablation (MWA) for small hepatocellular carcinoma (sHCC). Accurate and interpretable prediction of post-ablation recurrence risk is essential for individualized surveillance and management strategies. This study aimed to develop and validate an interpretable machine learning (IML) model for predicting tumor recurrence after MWA in patients with sHCC.

Methods: This retrospective study included 536 consecutive patients with sHCC treated with ultrasound-guided MWA between July 2017 and July 2023. A two-stage feature selection strategy combining Boruta and LASSO regression was applied to identify the most robust predictors. Six ML algorithms were developed and compared, including support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN), decision tree (DT), extreme gradient boosting (XGBoost), and logistic regression (LR). Model discrimination, calibration, and clinical utility were evaluated using receiver operating characteristic (ROC) analysis, calibration curves, decision curve analysis (DCA), and confusion matrices. Model interpretability was assessed using SHapley Additive exPlanations (SHAP).

Results: Tumor recurrence occurred in 29.1% of patients within 1 year after MWA. Seven variables, including maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, platelet count (PLT), alpha-fetoprotein (AFP), and C-reactive protein (CRP), were selected as predictors. Among all models, XGBoost demonstrated the best performance, achieving an AUC of 0.889 in the validation set, with good calibration and favorable net clinical benefit on DCA. SHAP analysis provided transparent global and individualized explanations, identifying AFP and CRP as the most influential predictors and revealing nonlinear risk patterns.

Conclusion: An interpretable XGBoost-based model using routinely available clinical variables accurately predicts recurrence after MWA in patients with sHCC. This model may serve as a clinically accessible approach for individualized risk stratification and post-ablation management; however, further external validation in multicenter cohorts is required before broader clinical application.

Keywords: interpretable machine learning, small hepatocellular carcinoma, microwave ablation, tumor recurrence, SHapley Additive exPlanations

Introduction

Hepatocellular carcinoma (HCC) remains one of the leading causes of cancer-related mortality worldwide, particularly in regions with a high prevalence of chronic liver disease and viral hepatitis.¹ With the increasing implementation of surveillance programs, a growing proportion of patients are diagnosed at an early stage, commonly referred to as small hepatocellular carcinoma (sHCC), typically defined as tumors ≤ 3 cm in diameter. For these patients, curative-intent local therapies, including surgical resection, liver transplantation, and image-guided thermal ablation, are widely recommended by international guidelines.²⁻⁴ Among these, microwave ablation (MWA) has been widely adopted as an effective minimally invasive treatment, offering

favorable local control with reduced procedural morbidity, especially in patients who are not ideal surgical candidates.^{5,6} Despite technical refinements and standardized ablation protocols, post-treatment tumor recurrence remains a frequent and clinically consequential event following MWA. Recurrence may occur as local tumor progression due to incomplete ablation or as intrahepatic distant recurrence reflecting underlying tumor aggressiveness and field cancerization in cirrhotic liver tissue.^{7,8} Early identification of patients at high risk of recurrence is essential. It can help optimize follow-up strategies, guide adjuvant treatment decisions, and improve long-term outcomes.

Most existing studies for predicting tumor recurrence after MWA are primarily based on traditional statistical models or their visualized extensions in the form of nomograms.^{9–12} These models are generally favored for their structural simplicity and ease of implementation in clinical practice. However, they are inherently constrained by assumptions of linearity, additivity, and independence among predictors, which may inadequately represent the complex and heterogeneous mechanisms underlying post-MWA recurrence.¹³ In real-world clinical settings, recurrence risk is often driven by nonlinear effects and high-order interactions among tumor characteristics, host-related factors, inflammatory status, and ablation-related technical parameters—relationships that are difficult to capture within conventional regression-based frameworks. Moreover, continuous variables are frequently categorized or simplified to facilitate model construction and visualization, potentially leading to information loss and reduced discrimination at the individual patient level.¹⁴ Collectively, these limitations restrict the ability of traditional nomogram-based models to fully characterize recurrence heterogeneity after MWA and underscore the need for more flexible modeling approaches capable of learning complex patterns while supporting individualized risk assessment. In addition, recent studies have explored molecular biomarkers and bioinformatics-based approaches to better understand hepatocellular carcinoma progression and prognosis, including hub gene identification,¹⁵ microRNA-regulated pathways,¹⁶ and protein expression-based prognostic markers.¹⁷ These findings highlight the biological heterogeneity of HCC and suggest that integrating multidimensional information may improve risk stratification. However, many of these approaches rely on molecular data that are not routinely available in clinical practice, which may limit their applicability in real-world settings.

Recent advances in machine learning (ML) have offered new opportunities to improve prognostic modeling in oncology by leveraging high-dimensional clinical, laboratory, and imaging-derived data.¹⁸ ML algorithms are particularly well suited to model nonlinear relationships and complex feature interactions, and several studies have demonstrated their superior performance over conventional regression-based approaches in predicting outcomes of tumor treatments.¹⁹ Nevertheless, most existing ML models function as “black boxes”, providing limited insight into how individual predictors contribute to model outputs. This lack of interpretability represents a significant barrier to clinical adoption, especially in high-stakes decision-making scenarios such as cancer recurrence prediction.²⁰ Interpretable machine learning (IML) aims to bridge this gap by combining predictive accuracy with transparent and explainable decision mechanisms. IML frameworks allow the contribution of individual variables to be quantified and visualized, enabling clinicians to assess whether model behavior aligns with established clinical knowledge and to explore potential novel risk factors.²¹ Such interpretability is particularly relevant in post-ablation recurrence prediction, where therapeutic decisions may be influenced by nuanced risk stratification rather than binary outcomes.

To date, few studies have focused on developing IML models specifically for predicting tumor recurrence after MWA in patients with sHCC. Thus, the present study aimed to establish and validate an IML model for recurrence prediction following MWA in sHCC. By integrating routinely collected clinical and procedural variables with IML techniques, we sought to construct a robust and clinically meaningful tool capable of supporting individualized post-ablation management and surveillance strategies.

Materials and Methods

Study Population and Patient Selection

This retrospective study collected clinical data of patients with sHCC who underwent MWA at our institution between July 2017 and July 2023 using the hospital electronic medical record system. Inclusion criteria: (1) met the diagnostic criteria for sHCC;²² (2) met the indications for MWA and underwent MWA with complete ablation of the lesion; (3) Child-Pugh liver function class A or B; (4) China Liver Cancer (CNLC) staging system stage Ia;²³ (5) no prior treatment had been received at the time of the initial consultation; (6) complete clinical and follow-up data. Exclusion criteria: (1) concurrent other

malignant tumors (eg, cervical cancer, colorectal cancer); (2) refractory massive ascites or cachexia; (3) severe decompensated liver function; (4) concurrent active infection; (5) presence of severe coagulation disorders or a tendency to bleed; (6) postoperative complications such as biliary tract injury or intra-abdominal infection. After applying these criteria, a total of 536 eligible patients were included in the final analysis. To develop and internally validate the prediction model, the study cohort was randomly split at a ratio of 7:3 into a training set (n=375) and a validation set (n=161). The training set was used for model development and tuning, whereas the validation set was reserved for independent assessment of model performance. The detailed patient selection process, including reasons for exclusion at each step, is summarized in the patient screening flowchart (Figure 1).

Diagnosis of sHCC, Recurrence Definition, and Follow-Up

HCC was diagnosed in patients with an underlying liver disease history (eg, chronic hepatitis B/C or cirrhosis) based on serum biomarkers together with imaging and/or histopathology. Specifically, HCC was considered when serum alpha-fetoprotein (AFP) was ≥ 400 ng/mL and a hepatic solid lesion was detected on ultrasound and/or cross-sectional imaging; lesions demonstrating typical HCC imaging hallmarks on contrast-enhanced CT or MRI were diagnosed noninvasively, whereas cases without typical imaging features were confirmed by percutaneous biopsy and histopathology. For eligibility in this study, sHCC was defined as either a solitary tumor with a maximum diameter ≤ 3 cm or 2–3 tumors with a cumulative tumor diameter ≤ 3 cm.²²

All patients received standardized MWA after hospital admission; for those with residual tumor after the initial procedure, repeat ablation was performed. After complete ablation was confirmed, patients were followed for 1 year, with follow-up visits scheduled every 2–3 months. Follow-up assessments included serum tumor biomarkers, conventional ultrasound, and contrast-enhanced CT or multiparametric MRI (as well as contrast-enhanced ultrasound when indicated). Recurrence was defined as the detection, during scheduled surveillance, of a newly developed abnormally enhancing lesion outside the ablation zone or in other areas of the liver, with dynamic contrast-enhanced CT, MRI, or contrast-enhanced ultrasound demonstrating marked arterial-phase hyperenhancement. Patients who developed recurrence within 1 year after ablation were assigned to the recurrence group, whereas all remaining patients were assigned to the non-recurrence group.

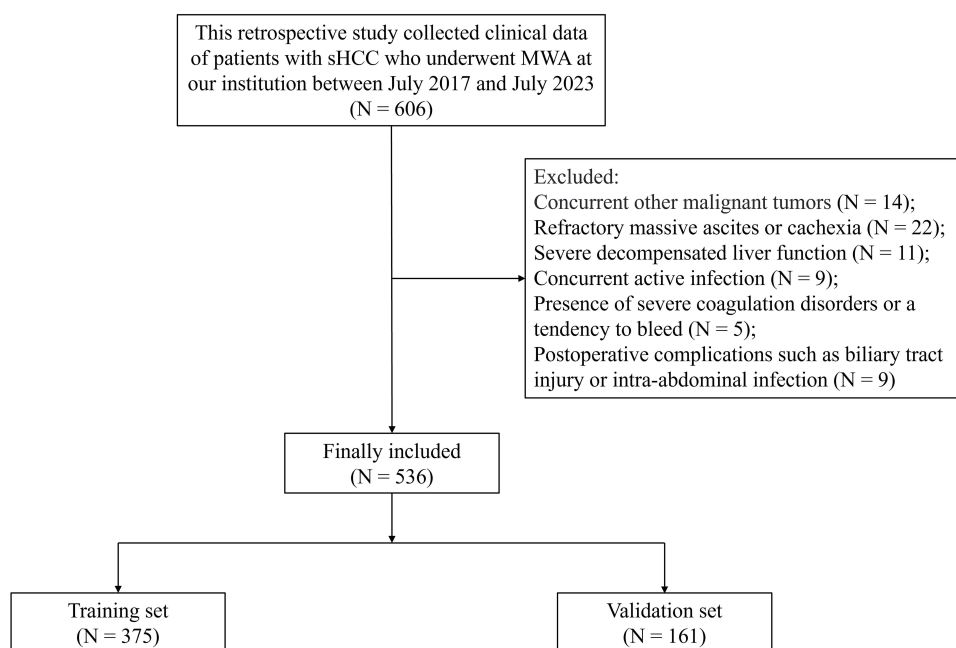


Figure 1 Flowchart illustrating the patient selection process and cohort allocation.
Abbreviations: sHCC, small hepatocellular carcinoma; MWA, microwave ablation.

Microwave Ablation Procedure

The KY-2000 microwave therapy device was employed, equipped with a 14G water-cooled MWA needle (Nanjing Kangyou Medical Technology Co., Ltd., frequency: 2450 MHz). Patients were positioned in the supine or lateral decubitus position according to tumor location. After skin disinfection at the puncture site and administration of local anesthesia, the microwave antenna was percutaneously advanced into the predefined target within the tumor under real-time ultrasound guidance. Continuous ultrasonographic monitoring was performed during needle insertion to avoid injury to adjacent major bile ducts, blood vessels, or other vital organs. Once accurate antenna placement was confirmed, microwave energy was delivered at a power output of 60–80 W for 15–20 min. Single-antenna ablation was applied for tumors <2 cm in diameter, whereas multipoint overlapping ablation was performed for tumors measuring 2–3 cm. Throughout the procedure, real-time ultrasound monitoring ensured that the ablation zone extended 0.5–1.0 cm beyond the tumor margin. At the end of the procedure, track ablation was routinely performed to prevent needle-track bleeding and tumor seeding. Postoperatively, patients received standard supportive treatment, including hepatoprotective therapy, anti-infective treatment, and maintenance of fluid and electrolyte balance. Contrast-enhanced CT was performed 3–4 weeks after ablation, and repeat MWA was carried out if residual tumor was detected at the original lesion site.

Clinical Data Collection

A standardized case report form for baseline data collection was designed. Clinical data were extracted from the electronic medical records by trained investigators and recorded using a double-check procedure. Collected variables included gender (male/female), age, body mass index (BMI), maximum tumor diameter, number of tumors (single or 2–3 lesions), Child-Pugh liver function class (A or B), hypertension, diabetes mellitus, hyperlipidemia, family history of cancer, smoking history, drinking history, history of hepatitis B virus infection, liver cirrhosis, portal hypertension, and ablation frequency (once or twice). Laboratory parameters at admission included white blood cell count (WBC), platelet count (PLT), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), AFP, gamma-glutamyl transpeptidase (GGT), fibrinogen (FIB), Des- γ -carboxy prothrombin (DCP), C-reactive protein (CRP), and serum albumin level.

Feature Selection, Model Development, and Interpretable Machine Learning Analysis

To reduce dimensionality and minimize overfitting, a two-step feature selection strategy was adopted. First, the Boruta algorithm, a wrapper method based on random forests, was applied to identify all relevant features by iteratively comparing original variables with their randomized shadow counterparts. Second, least absolute shrinkage and selection operator (LASSO) regression was performed, which applies L1 regularization to shrink regression coefficients and select variables with non-zero coefficients under optimal penalization. Only variables consistently identified by both Boruta and LASSO were retained, and the intersection of the two feature sets was used as the final input feature set for model development. Based on the selected features, six ML models were developed, including support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN), decision tree (DT), extreme gradient boosting (XGBoost), and logistic regression (LR). All models were trained on the training set, with hyperparameters optimized through cross-validation to achieve optimal predictive performance.

The predictive performance of the six models was comprehensively evaluated and compared. Discrimination ability was primarily assessed using the area under the receiver operating characteristic (ROC) curve (AUC), which served as the main criterion for model comparison. In addition, other performance metrics, including F1 score, accuracy, precision, recall, sensitivity and specificity were calculated to provide a more complete assessment of model performance. The model demonstrating the highest AUC with balanced overall performance was selected as the final prediction model. The calibration performance of the selected model was assessed using calibration plots, which examined the consistency between predicted probabilities and observed outcomes. To evaluate the potential clinical value of the model, decision curve analysis (DCA) was conducted to estimate the net benefit across a range of clinically relevant threshold probabilities. Furthermore, a confusion matrix was generated to present detailed classification outcomes and error distributions.

To further enhance model transparency and facilitate clinical interpretability, Shapley Additive Explanations (SHAP) were applied to the final prediction model. SHAP is a game theory-based approach that attributes each model prediction to individual feature contributions by estimating the marginal contribution of each variable across all possible feature combinations. Global SHAP summary plots were generated to rank features according to their overall importance and to visualize the direction and magnitude of their effects on model output. In addition, SHAP dependence plots were used to explore potential nonlinear relationships and interaction effects between key predictors and the predicted outcome. At the individual level, SHAP force plots were employed to illustrate how specific features collectively influenced risk estimation for individual patients.

Quality Control Measures

This study was conducted in strict accordance with principles of good clinical research practice. A multidisciplinary research team with extensive professional expertise and clinical experience was established, and the final study objectives and methodological details were determined after comprehensive literature review and rigorous protocol development. For all patients with small hepatocellular carcinoma who underwent MWA and were included in the analysis, baseline clinical characteristics and serum biochemical parameters were complete and readily available. With respect to data quality control, dedicated personnel were responsible for follow-up management, with precise definition and accurate documentation of follow-up time points and study endpoints to ensure the reliability of the original data. Finally, standardized information workflows were applied for data collection, organization, and statistical analysis, thereby ensuring methodological rigor throughout data processing and enhancing the clinical relevance and interpretability of the study findings.

Statistical Analysis

All analyses were conducted in R software (version 4.3.1). Continuous variables were assessed for distributional characteristics using the Shapiro–Wilk test and summarized accordingly. Variables following a normal distribution were expressed as mean \pm standard deviation (SD) and compared using the Student's *t*-test. Non-normally distributed variables were summarized as median and interquartile range (IQR) and compared using the Mann–Whitney *U*-test. Categorical variables were presented as counts and percentages. Group comparisons were performed using the chi-square test, or Fisher's exact test when the expected cell count was less than five.

To assess multicollinearity among candidate predictors, variance inflation factor (VIF) analysis was conducted prior to model construction. Variables with a VIF exceeding 5 were considered to exhibit significant collinearity and were carefully evaluated or excluded to ensure model stability. For machine learning model development, hyperparameter optimization was performed using a grid search strategy combined with *k*-fold cross-validation (*k* = 5) within the training dataset. For each algorithm, a predefined grid of candidate hyperparameter values was evaluated, and the optimal combination was selected based on cross-validated performance (primarily AUC). To improve reproducibility and transparency, the main hyperparameters and their corresponding search ranges are provided in [Supplementary Table S1](#).

Data preprocessing and visualization were performed using “tidyverse”. Feature selection was implemented using “glmnet” (LASSO) and “Boruta”. Model training and tuning were performed using “caret” and “tidymodels”. Discrimination was assessed using “pROC”; clinical utility was evaluated using “rmda”; confusion matrices were produced using “caret”. IML analyses were performed using “shapviz” and “fastshap” packages. To ensure reproducibility, a fixed random seed (set to 4321) was used throughout all stochastic procedures, including cohort splitting, cross-validation, and hyperparameter optimization. All statistical tests were two-sided, and statistical significance was defined as a *P* value < 0.05.

Results

Baseline Profile of the Study Cohort

The study cohort consisted of 536 patients, of whom 375 were assigned to the training set and 161 to the validation set. Tumor recurrence was observed in 156 patients (29.1%) in the overall cohort, with similar recurrence rates in the training and validation sets (28.3% vs. 31.1%, *P* = 0.515). The mean age of the overall population was 59.21 \pm 6.04 years, and 55.6% of patients were male. No statistically significant differences were identified between the training and validation

sets with respect to baseline demographic characteristics, liver function status, comorbidities, tumor burden, treatment characteristics, or laboratory parameters (all $P > 0.05$), supporting the appropriateness of the cohort split for model development and validation (Table 1).

Table 1 Baseline Characteristics of the Training and Validation Sets

Variables	Total (n=536)	Training Set (n=375)	Validation Set (n=161)	$\chi^2/t/Z$	P
Recurrence				0.425	0.515
No	380 (70.9%)	269 (71.7%)	111 (68.9%)		
Yes	156 (29.1%)	106 (28.3%)	50 (31.1%)		
Gender				0.009	0.926
Male	298 (55.6%)	208 (55.5%)	90 (55.9%)		
Female	238 (44.4%)	167 (44.5%)	71 (44.1%)		
Age (years)	59.21±6.04	59.16±5.95	59.31±6.27	0.263	0.793
BMI (kg/m ²)	22.48±2.37	22.52±2.31	22.38±2.52	0.626	0.532
Maximum tumor diameter (cm)	1.56±0.67	1.56±0.68	1.57±0.66	0.157	0.875
Number of tumors				0.157	0.692
Single	270 (50.4%)	191 (50.9%)	79 (49.1%)		
2–3 tumors	266 (49.6%)	184 (49.1%)	82 (50.9%)		
Child-Pugh grade				0.975	0.323
Grade A	382 (71.3%)	272 (72.5%)	110 (68.3%)		
Grade B	154 (28.7%)	103 (27.5%)	51 (31.7%)		
Hypertension				0.019	0.891
No	384 (71.6%)	268 (71.5%)	116 (72.0%)		
Yes	152 (28.4%)	107 (28.5%)	45 (28.0%)		
Diabetes mellitus				0.302	0.583
No	437 (81.5%)	308 (82.1%)	129 (80.1%)		
Yes	99 (18.5%)	67 (17.9%)	32 (19.9%)		
Hyperlipidemia				0.125	0.724
No	394 (73.5%)	274 (73.1%)	120 (74.5%)		
Yes	142 (26.5%)	101 (26.9%)	41 (25.5%)		
Family history of cancer				0.129	0.720
No	467 (87.1%)	328 (87.5%)	139 (86.3%)		
Yes	69 (12.9%)	47 (12.5%)	22 (13.7%)		
Smoking				0.047	0.829
No	416 (77.6%)	292 (77.9%)	124 (77.0%)		
Yes	120 (22.4%)	83 (22.1%)	37 (23.0%)		
Drinking				0.081	0.776
No	385 (71.8%)	268 (71.5%)	117 (72.7%)		
Yes	151 (28.2%)	107 (28.5%)	44 (27.3%)		
Hepatitis B virus infection				0.039	0.843
No	106 (19.8%)	75 (20.0%)	31 (19.3%)		
Yes	430 (80.2%)	300 (80.0%)	130 (80.7%)		
Cirrhosis				0.634	0.426
No	392 (73.1%)	278 (74.1%)	114 (70.8%)		
Yes	144 (26.9%)	97 (25.9%)	47 (29.2%)		
Portal hypertension				1.338	0.247
No	426 (79.5%)	303 (80.8%)	123 (76.4%)		
Yes	110 (20.5%)	72 (19.2%)	38 (23.6%)		
Ablation frequency				0.149	0.700
Once	380 (70.9%)	264 (70.4%)	116 (72.0%)		
Twice	156 (29.1%)	111 (29.6%)	45 (28.0%)		

(Continued)

Table 1 (Continued).

Variables	Total (n=536)	Training Set (n=375)	Validation Set (n=161)	$\chi^2/t/Z$	P
WBC ($\times 10^9/L$)	6.31 \pm 1.42	6.33 \pm 1.35	6.28 \pm 1.57	0.374	0.709
PLT ($\times 10^9/L$)	113 (76, 153)	113 (77, 153)	113 (74, 152)	0.360	0.718
AST (U/L)	44.21 \pm 6.22	44.14 \pm 6.39	44.37 \pm 5.83	0.392	0.695
ALT (U/L)	39.89 \pm 6.33	39.80 \pm 6.45	40.11 \pm 6.06	0.519	0.604
TBIL (μ mol/L)	17.2 (11.7, 25.5)	17.4 (11.9, 25.7)	16.8 (11.2, 25.1)	0.621	0.530
GGT (U/L)	198.90 \pm 36.14	198.33 \pm 35.52	200.24 \pm 37.61	0.561	0.575
FIB (g/L)	4.23 \pm 0.75	4.21 \pm 0.78	4.28 \pm 0.69	0.985	0.325
AFP (ng/mL)	566.16 \pm 55.88	567.63 \pm 53.82	562.73 \pm 60.45	0.931	0.353
DCP (mAU/mL)	165 (106, 249)	164 (105, 248)	168 (110, 253)	0.402	0.693
CRP (mg/L)	11.2 (5.8, 16.1)	10.7 (5.8, 16.3)	11.4 (5.8, 15.6)	0.034	0.973
Albumin (g/L)	38.04 \pm 4.76	38.09 \pm 4.65	37.93 \pm 5.01	0.357	0.722

Abbreviations: BMI, body mass index; WBC, white blood cell count; PLT, platelet count; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TBIL, total bilirubin; GGT, gamma-glutamyl transferase; FIB, fibrinogen; AFP, alpha-fetoprotein; DCP, Des- γ -carboxy prothrombin; CRP, C-reactive protein.

Comparison of Baseline Characteristics by Recurrence Status in the Training Set

As shown in Table 2, within the training set, 106 patients developed tumor recurrence, whereas 269 patients remained recurrence-free. Baseline demographic characteristics, including age, sex, and BMI, were comparable between the recurrence and non-recurrence groups (all $P > 0.05$). Patients in the recurrence group had a larger maximum tumor diameter than those in the non-recurrence group (1.75 ± 0.62 cm vs. 1.48 ± 0.69 cm, $P < 0.001$). The distribution of Child-Pugh grade differed significantly, with a higher proportion of Child-Pugh grade B observed among patients with

Table 2 Clinical and Laboratory Profiles Stratified by Recurrence Status in the Training Set

Variables	Recurrence Group (n=106)	Non-Recurrence Group (n=269)	$\chi^2/t/Z$	P
Gender			0.171	0.679
Male	57 (53.8%)	151 (56.1%)		
Female	49 (46.2%)	118 (43.9%)		
Age (years)	59.68 \pm 6.12	58.96 \pm 5.88	1.055	0.292
BMI (kg/m^2)	22.48 \pm 2.18	22.53 \pm 2.37	0.188	0.851
Maximum tumor diameter (cm)	1.75 \pm 0.62	1.48 \pm 0.69	3.509	<0.001
Number of tumors			0.470	0.493
Single	51 (48.1%)	140 (52.0%)		
2–3 tumors	55 (51.9%)	129 (48.0%)		
Child-Pugh grade			12.728	<0.001
Grade A	63 (59.4%)	209 (77.7%)		
Grade B	43 (40.6%)	60 (22.3%)		
Hypertension			0.037	0.848
No	75 (70.8%)	193 (71.7%)		
Yes	31 (29.2%)	76 (28.3%)		
Diabetes mellitus			0.101	0.751
No	86 (81.1%)	222 (82.5%)		
Yes	20 (18.9%)	47 (17.5%)		
Hyperlipidemia			0.020	0.887
No	78 (73.6%)	196 (72.9%)		
Yes	28 (26.4%)	73 (27.1%)		

(Continued)

Table 2 (Continued).

Variables	Recurrence Group (n=106)	Non-Recurrence Group (n=269)	$\chi^2/t/Z$	P
Family history of cancer			0.198	0.656
No	94 (88.7%)	234 (87.0%)		
Yes	12 (11.3%)	35 (13.0%)		
Smoking			0.016	0.899
No	83 (78.3%)	209 (77.7%)		
Yes	23 (21.7%)	60 (22.3%)		
Drinking			0.489	0.484
No	73 (68.9%)	195 (72.5%)		
Yes	33 (31.1%)	74 (27.5%)		
Hepatitis B virus infection			0.003	0.954
No	21 (19.8%)	54 (20.1%)		
Yes	85 (80.2%)	215 (79.9%)		
Cirrhosis			12.650	<0.001
No	65 (61.3%)	213 (79.2%)		
Yes	41 (38.7%)	56 (20.8%)		
Portal hypertension			7.907	0.005
No	74 (69.8%)	229 (85.1%)		
Yes	32 (30.2%)	40 (14.9%)		
Ablation frequency			0.025	0.875
Once	74 (69.8%)	190 (70.6%)		
Twice	32 (30.2%)	79 (29.4%)		
WBC ($\times 10^9/L$)	6.17 \pm 1.26	6.39 \pm 1.38	1.424	0.155
PLT ($\times 10^9/L$)	98 (69, 137)	117 (80, 162)	3.159	0.002
AST (U/L)	45.18 \pm 6.12	43.73 \pm 6.46	1.986	0.048
ALT (U/L)	40.37 \pm 6.11	39.58 \pm 6.58	1.068	0.286
TBIL (μ mol/L)	17.1 (10.9, 29.2)	17.6 (12.5, 24.6)	0.403	0.687
GGT (U/L)	201.34 \pm 38.57	197.15 \pm 34.25	1.029	0.304
FIB (g/L)	4.26 \pm 0.86	4.19 \pm 0.74	0.787	0.432
AFP (ng/mL)	598.48 \pm 55.88	555.47 \pm 47.89	7.461	<0.001
DCP (mAU/mL)	163 (102, 257)	165 (106, 244)	0.036	0.971
CRP (mg/L)	11.2 (7.9, 20.0)	9.5 (5.2, 15.2)	3.537	<0.001
Albumin (g/L)	37.85 \pm 3.81	38.18 \pm 4.94	0.619	0.536

Abbreviations: BMI, body mass index; WBC, white blood cell count; PLT, platelet count; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TBIL, total bilirubin; GGT, gamma-glutamyl transferase; FIB, fibrinogen; AFP, alpha-fetoprotein; DCP, Des- γ -carboxy prothrombin; CRP, C-reactive protein.

recurrence (40.6% vs. 22.3%, $P < 0.001$). In addition, patients in the recurrence group exhibited a higher prevalence of cirrhosis (38.7% vs. 20.8%, $P < 0.001$) and portal hypertension (30.2% vs. 14.9%, $P = 0.005$). Comorbid conditions, lifestyle factors, hepatitis B virus infection status, tumor number, and ablation frequency were comparable between the recurrence and non-recurrence groups (all $P > 0.05$). Regarding laboratory parameters, patients with recurrence showed lower PLT and higher levels of AST, AFP, and CRP compared with the non-recurrence group (all $P < 0.05$). No significant differences were observed in WBC, ALT, TBIL, GGT, FIB, DCP, or albumin levels between groups.

Feature Selection Using Boruta and LASSO

Feature selection was performed in the training set using a combined Boruta and LASSO strategy to identify robust predictors associated with tumor recurrence after MWA. As illustrated in [Figure 2A](#) and [B](#), the Boruta algorithm was first applied to rank the importance of candidate variables by comparing original features with their corresponding shadow features. Variables with importance values consistently exceeding the maximum shadow importance were confirmed as

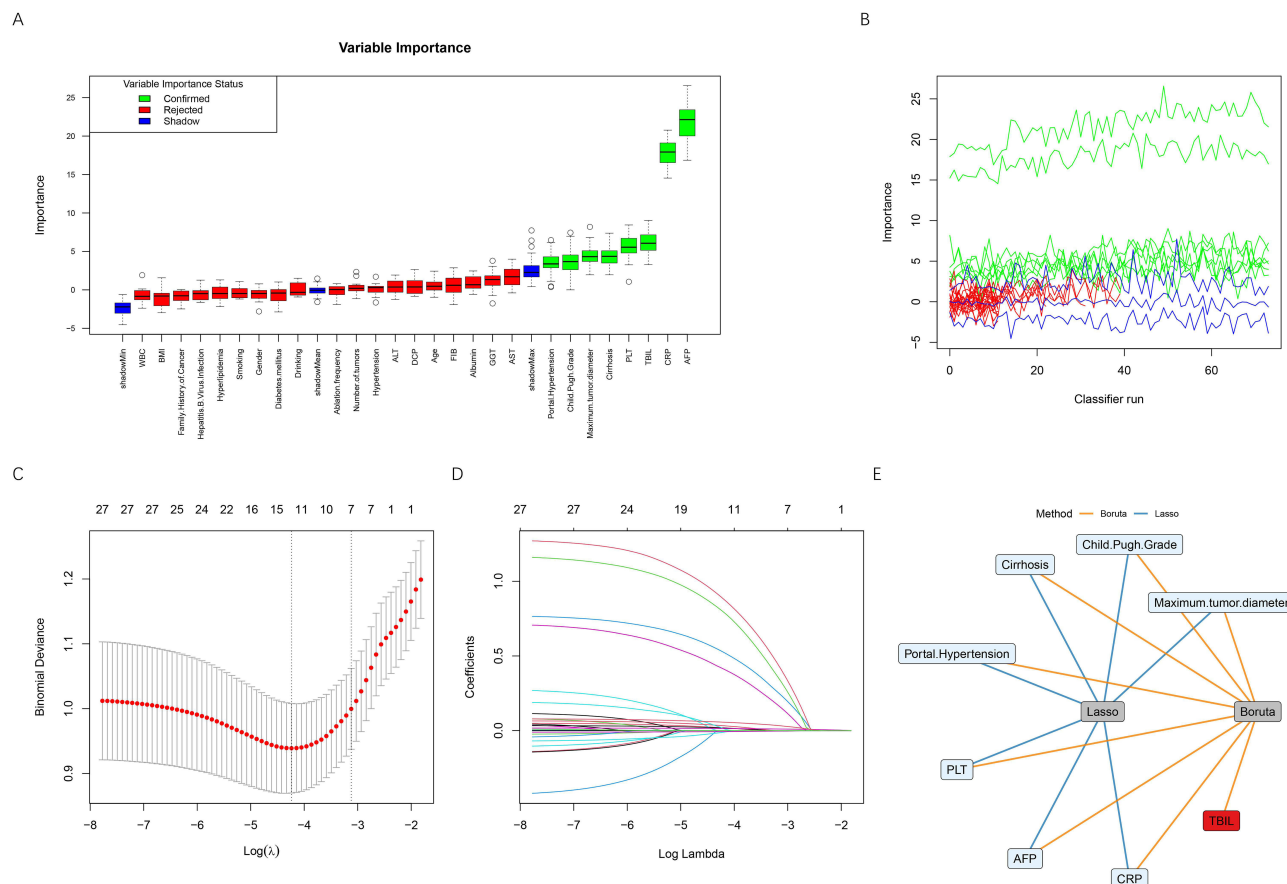


Figure 2 Feature selection using Boruta and LASSO regression. **(A)** Variable importance distribution obtained from the Boruta algorithm. Green boxes represent confirmed important features, red boxes indicate rejected features, and blue boxes denote shadow features used as references. **(B)** Evolution of variable importance across Boruta iterations, illustrating the stability of confirmed, rejected, and shadow features during repeated classifier runs. **(C)** Ten-fold cross-validation curve of the LASSO regression model showing the relationship between $\log(\lambda)$ and binomial deviance. The left dotted vertical line corresponds to the λ value that minimized the binomial deviance (λ_{\min}), while the right dotted line represents the more regularized λ within one standard error of the minimum deviance (λ_{1se}). The optimal λ was selected based on the minimum deviance criterion, balancing model fit and complexity. **(D)** Coefficient trajectories of candidate variables as a function of $\log(\lambda)$. As λ increased, coefficients were progressively shrunk toward zero, and only variables with non-zero coefficients at the optimal λ were retained. **(E)** Overlap of variables selected by Boruta and LASSO methods. Features consistently identified by both approaches were selected as the final predictors for subsequent machine learning model construction.

Abbreviations: BMI, body mass index; WBC, white blood cell count; PLT, platelet count; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TBIL, total bilirubin; GGT, gamma-glutamyl transferase; FIB, fibrinogen; AFP, alpha-fetoprotein; DCP, Des- γ -carboxy prothrombin; CRP, C-reactive protein.

relevant. Using this approach, Boruta identified maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, PLT, AFP, CRP, and TBIL as important features.

Subsequently, LASSO regression was used to further constrain the feature space. The optimal regularization parameter was selected based on cross-validation (Figure 2C), and the corresponding coefficient shrinkage paths are illustrated in Figure 2D. At the optimal λ value, LASSO retained maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, PLT, AFP, and CRP. To ensure robustness and reduce redundancy, only features selected by both Boruta and LASSO were carried forward for model construction. The intersection of the two methods is summarized in Figure 2E. Consequently, seven variables, including maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, PLT, AFP, and CRP, were ultimately selected as input features for subsequent ML analyses.

Variable Encoding and Collinearity Assessment

Before model construction, selected variables were appropriately encoded according to their data types. Continuous variables were retained in their original numerical form, while categorical variables were transformed into binary or ordinal variables based on clinical relevance to ensure compatibility with machine learning algorithms. All variables were processed in a consistent manner across the training and validation datasets (Table 3).

Table 3 Definitions and Coding of Study Variables

Variable type	Variable Name	Value Assignment
Dependent	Recurrence	No =0, Yes =1
Independent	Maximum tumor diameter	Continuous
Independent	Child-Pugh grade	No =0, Yes =1
Independent	Cirrhosis	No =0, Yes =1
Independent	Portal hypertension	No =0, Yes =1
Independent	PLT	Continuous
Independent	AFP	Continuous
Independent	CRP	Continuous

Abbreviations: PLT, platelet count; AFP, alpha-fetoprotein; CRP, C-reactive protein.

Collinearity among the selected variables was evaluated using VIF analysis. The calculated VIF values were 1.032 for maximum tumor diameter, 1.040 for Child-Pugh grade, 1.047 for cirrhosis, 1.019 for portal hypertension, 1.017 for PLT, 1.045 for AFP, and 1.063 for CRP. All predictors demonstrated VIF values below the predefined threshold of 5, indicating no evidence of significant multicollinearity. Consequently, all variables were retained for subsequent ML model construction.

Model Development and Performance Comparison

Using the selected features, six ML models, including SVM, RF, KNN, DT, XGBoost, and LR, were developed and their predictive performance was systematically evaluated. As shown in [Figure 3A](#), in the training set, all models demonstrated discriminatory ability for predicting tumor recurrence. Among them, XGBoost achieved the highest performance with an AUC of 0.926 (95% CI, 0.897–0.954), followed closely by RF with an AUC of 0.924 (95% CI, 0.899–0.949). The SVM, KNN, and LR models showed moderate discrimination, whereas the DT model exhibited comparatively lower performance. The performance pattern was largely preserved in the validation set ([Figure 3B](#)), indicating good model generalizability. XGBoost maintained the highest discriminative ability (AUC = 0.889), while SVM and LR achieved comparable AUC values. Although RF and KNN showed a slight decline in performance, their overall discrimination remained acceptable. DT consistently demonstrated the lowest performance across both datasets.

In addition to AUC, multiple classification metrics were assessed to provide a comprehensive evaluation of model behavior ([Figure 3C and D](#)). Across these metrics, XGBoost and RF generally exhibited more stable and balanced performance, whereas DT showed greater variability, particularly in measures related to agreement and predictive consistency. Based on its superior AUC performance, robust generalization across datasets, and balanced classification metrics, the XGBoost model was selected as the final prediction model for subsequent calibration, clinical utility assessment, and interpretability analysis.

Model Calibration and Clinical Performance

The calibration performance of the final XGBoost model was evaluated in both the training and validation sets. As shown in [Figure 4A and B](#), the apparent and bias-corrected calibration curves closely followed the ideal reference line, indicating good agreement between predicted probabilities and observed recurrence rates across a wide range of risk levels. The clinical utility of the model was evaluated using DCA ([Figure 4C and D](#)). In both datasets, the XGBoost model provided a consistently higher net benefit than the “treat-all” and “treat-none” strategies across a broad range of threshold probabilities, suggesting meaningful clinical usefulness for guiding risk-based decision-making.

The classification performance of the XGBoost model was summarized using confusion matrices in [Figure 4E](#) (training set) and [Figure 4F](#) (validation set). In the training set, the model correctly identified 225 non-recurrence cases and 95 recurrence cases, with 11 false positives and 44 false negatives. In the validation set, 89 non-recurrence cases and 41 recurrence cases were correctly classified, while 9 false positives and 22 false negatives were observed. These results indicate balanced discrimination with acceptable error rates in both datasets. Overall, the XGBoost model

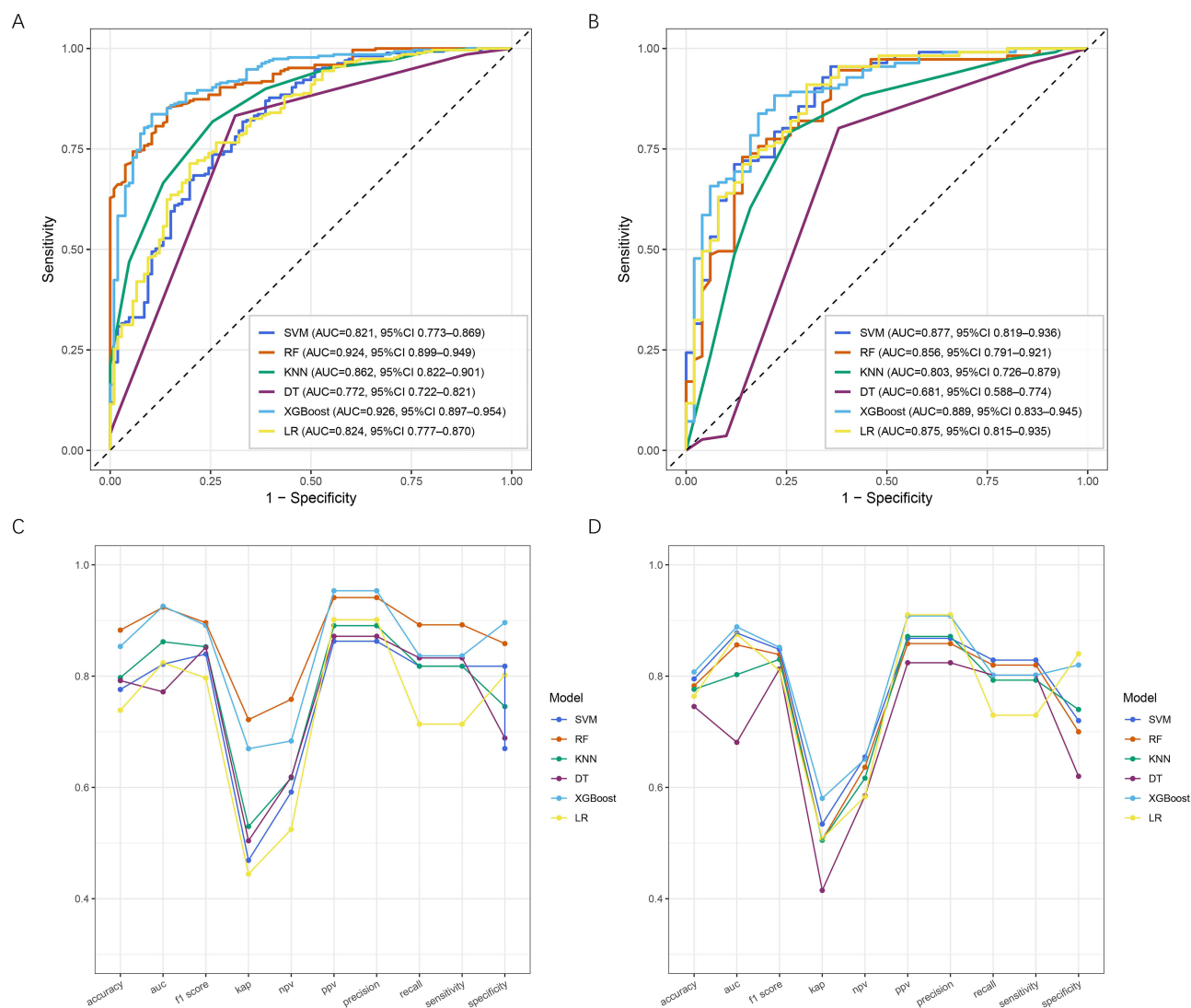


Figure 3 Performance of five ML models in predicting post-sepsis cognitive impairment. **(A)** ROC curves of the six models in the training set, with corresponding AUC values and 95% confidence intervals. **(B)** ROC curves of the six models in the validation set, with corresponding AUC values and 95% confidence intervals. **(C)** Comparison of multiple performance metrics, including accuracy, AUC, F1 score, kappa, NPV, PPV, precision, recall, sensitivity, and specificity, in the training set. **(D)** Comparison of multiple performance metrics in the validation set.

Abbreviations: ML, machine learning; SVM, support vector machine; RF, random forest; KNN, k-nearest neighbor; DT, decision tree; XGBoost, extreme gradient boosting; LR, logistic regression; AUC, Area under the curve; NPV, negative predictive value; PPV, positive predictive value.

demonstrated good calibration, favorable clinical net benefit, and stable classification performance, supporting its robustness and potential applicability for predicting tumor recurrence after MWA.

Model Interpretability Analysis Using SHAP

The interpretability of the final XGBoost model was examined using SHAP to quantify the contribution of individual predictors to recurrence risk after MWA. Global feature importance based on the mean absolute SHAP values is presented in [Figure 5A](#). AFP and CRP emerged as the most influential predictors, followed by cirrhosis, maximum tumor diameter, PLT, Child-Pugh grade, and portal hypertension. The SHAP summary plot ([Figure 5B](#)) further illustrated both the magnitude and direction of feature effects, demonstrating that higher AFP and CRP values were generally associated with increased recurrence risk, whereas higher PLT values tended to reduce the predicted risk. Local interpretability at the individual level is illustrated in [Figure 5C–E](#) using representative SHAP force plots. These

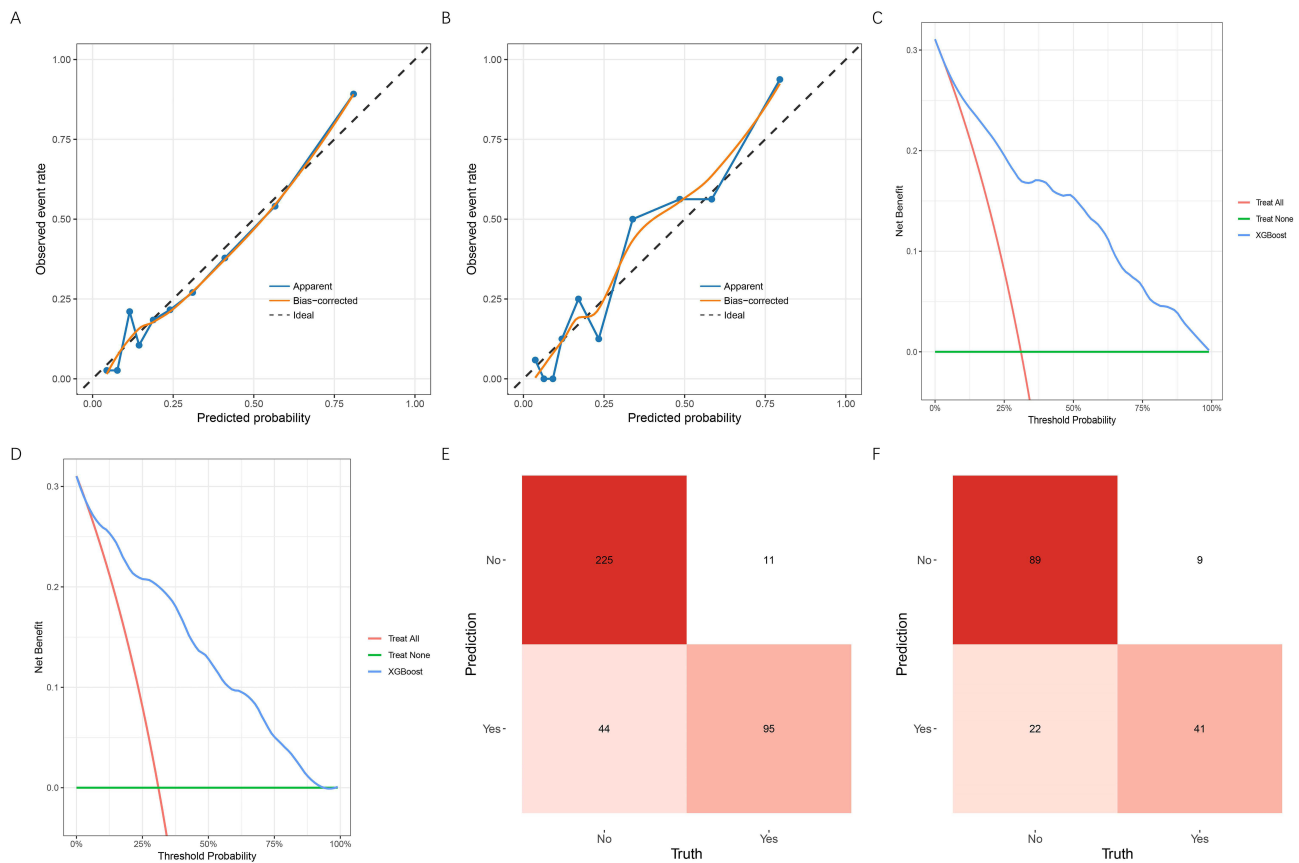


Figure 4 Calibration, Clinical utility, and confusion matrix evaluation of the XGBoost model. **(A)** Calibration curve of the XGBoost model in the training set, showing the agreement between predicted probabilities and observed recurrence rates. **(B)** Calibration curve of the XGBoost model in the validation set. **(C)** DCA showing the net benefit of the model across a range of threshold probabilities in the training cohort. **(D)** DCA in the validation cohort, demonstrating consistency of clinical net benefit. **(E)** Confusion matrix of the XGBoost model in the training set, summarizing classification results for recurrence and non-recurrence. **(F)** Confusion matrix of the XGBoost model in the validation set.

Abbreviations: XGBoost, extreme gradient boosting; DCA, decision curve analysis.

examples showed how different combinations of clinical features contributed to individual recurrence risk predictions, highlighting the additive nature of feature effects and the heterogeneity of risk profiles among patients.

The relationships between predictors and model output were further explored using SHAP dependence plots (Figure 6A–G). AFP and CRP displayed pronounced nonlinear effects, with SHAP values increasing sharply beyond certain concentration ranges. Maximum tumor diameter showed a progressive positive association with recurrence risk, whereas PLT exhibited an inverse relationship. For categorical variables, including cirrhosis, Child-Pugh grade, and portal hypertension, clear stepwise differences in SHAP values were observed, indicating distinct risk contributions between categories. Overall, SHAP analysis provided both global and individualized explanations for model predictions, enhancing the transparency and clinical interpretability of the XGBoost model for predicting tumor recurrence after MWA.

Discussion

In this retrospective cohort of 536 patients with sHCC treated with ultrasound-guided MWA, we developed and internally validated an IML model to predict 1-year post-ablation recurrence. Tumor recurrence occurred in 29.1% of the overall cohort, underscoring that early relapse remains a clinically relevant challenge even in small tumors treated with curative-intent ablation. Using a two-step Boruta-LASSO feature selection strategy, we consistently identified seven routinely available variables, including maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, PLT, AFP, and CRP, which were subsequently used as inputs for model development. Among six candidate algorithms, the XGBoost

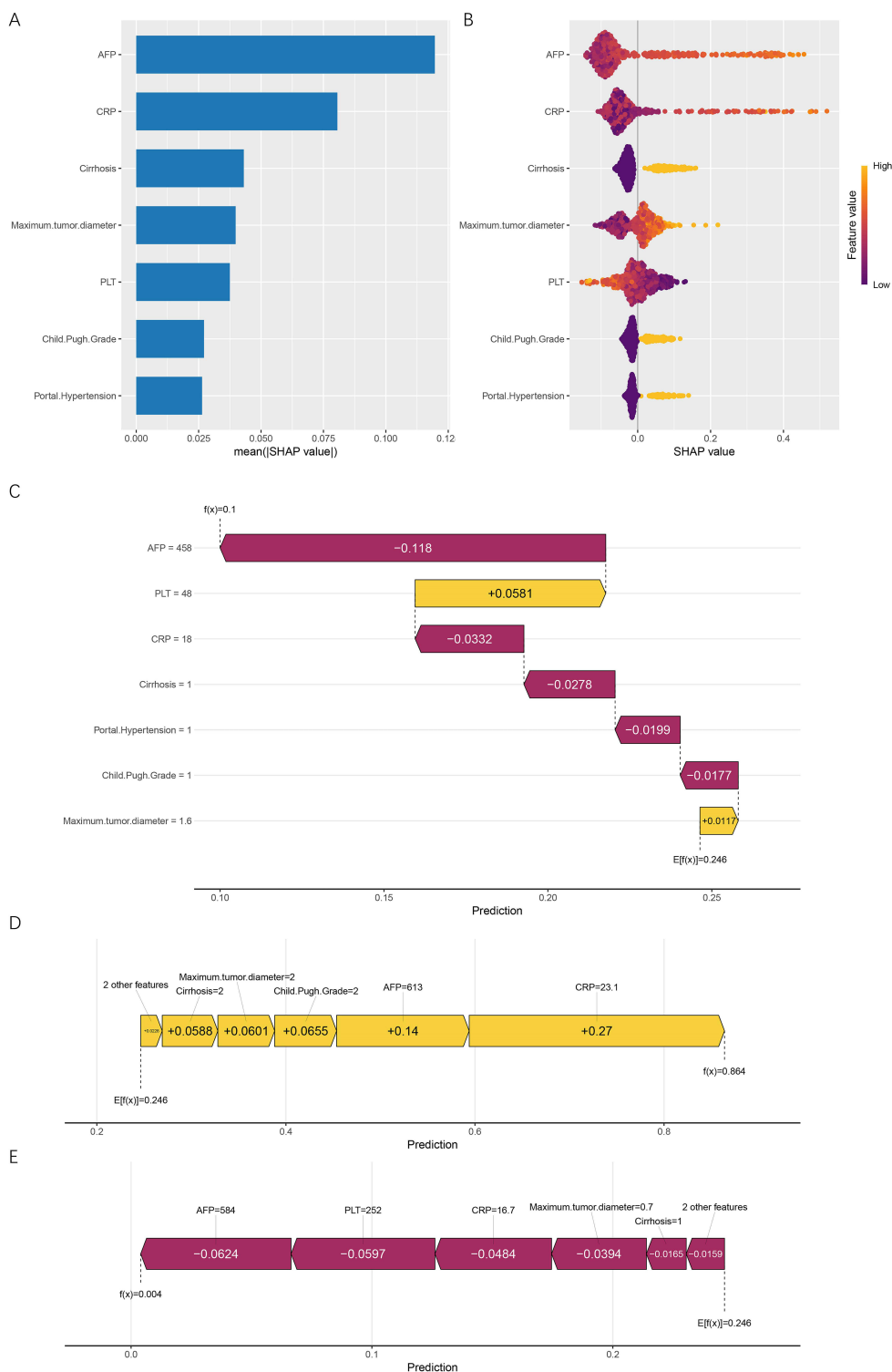


Figure 5 SHAP based interpretation of the XGBoost model. **(A)** Global feature importance ranked by mean absolute SHAP values, showing the relative contribution of each predictor to the model output. **(B)** SHAP summary plot illustrating the distribution and direction of feature effects on recurrence risk, where each point represents an individual sample and colors indicate feature values from low to high. **(C)** SHAP force plot of the first patient in the dataset, demonstrating how individual features contributed to decreasing or increasing the predicted probability relative to the baseline. **(D)** SHAP force plot for a high-risk patient, showing cumulative positive contributions that increased the predicted recurrence probability. **(E)** SHAP force plot for a low-risk patient, illustrating feature contributions that collectively reduced the predicted recurrence probability. $E[f(x)]$ represents the baseline prediction of the model, corresponding to the mean predicted recurrence risk, whereas $f(x)$ denotes the final predicted recurrence probability for an individual patient after incorporating all feature contributions.

Abbreviations: SHAP, SHapley Additive exPlanations; PLT, platelet count; AFP, alpha-fetoprotein; CRP, C-reactive protein.

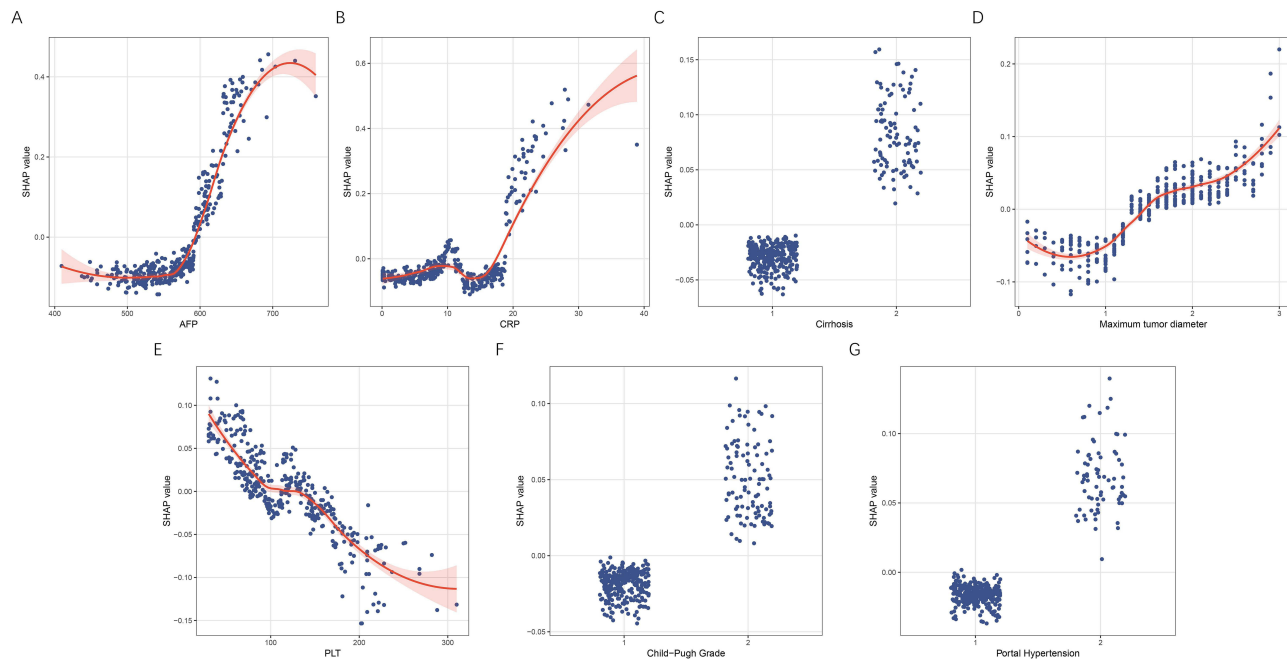


Figure 6 SHAP dependence plots illustrating the relationships between key predictors and model output. **(A)** AFP. **(B)** CRP. **(C)** Cirrhosis. **(D)** Maximum tumor diameter. **(E)** PLT. **(F)** Child-Pugh grade. **(G)** Portal hypertension. Each dot represents an individual patient. The x-axis shows the feature value and the y-axis shows the corresponding SHAP value, with positive values indicating increased predicted recurrence risk and negative values indicating decreased risk. The red curve indicates the smoothed SHAP dependence, and the shaded band represents the 95% confidence interval.

Abbreviations: SHAP, SHapley Additive exPlanations; PLT, platelet count; AFP, alpha-fetoprotein; CRP, C-reactive protein.

model provided the best overall discrimination and generalizability (training AUC 0.926; validation AUC 0.889), accompanied by favorable calibration and meaningful net benefit across a broad range of thresholds on DCA. Importantly, SHAP analyses provided transparent explanations at both the global and individual levels, revealing that AFP and CRP contributed most to recurrence risk, followed by cirrhosis and tumor diameter, and demonstrating nonlinear risk patterns for key continuous variables. Collectively, our findings suggest that an interpretable XGBoost model based on routine clinical variables can facilitate individualized recurrence risk stratification after MWA for sHCC.

Recurrence after curative-intent therapy for sHCC is common and represents a major determinant of long-term outcomes; therefore, accurate risk stratification after ablation is an active area of investigation. Prior prediction efforts in this space have largely relied on regression-based models and nomograms. For instance, Wang et al retrospectively developed a logistic regression-based nomogram to predict 1-year recurrence after ultrasound-guided MWA for HCC in a cohort of 119 patients, achieving an AUC of 0.86 in the validation cohort.¹¹ Zhang et al established a multi-parameter model integrating preoperative MRI features to predict early recurrence within one year after MWA in patients with hepatitis B-associated HCC. The model demonstrated good external performance, with an AUC of 0.842 in the validation cohort.¹² Despite these encouraging results, several limitations remain in the existing predictive literature. Most available models were derived from retrospective cohorts of small sample sizes, which may restrict statistical stability and generalizability. In addition, regression-based approaches and nomograms inherently rely on assumptions of linearity and additivity among predictors, potentially limiting their ability to capture complex nonlinear relationships and high-order interactions that underlie tumor recurrence after ablation. Imaging-driven models, although promising, often require advanced imaging protocols and feature extraction pipelines that may not be uniformly available across clinical settings, thereby constraining real-world applicability. In addition, recent ML studies have also explored clinically relevant outcomes after MWA in HCC. For example, Ren et al developed ML-based models to predict local tumor progression after initial MWA in patients with early-stage HCC and reported that the best-performing CatBoost model achieved an AUC of 0.898 in the training cohort.²⁴ Although local tumor progression and overall recurrence represent different clinical endpoints, this study further supports the potential value of ML approaches in post-ablation risk assessment. Against this background, the present study leverages a substantially larger cohort of patients with sHCC and adopts a ML-based framework

with rigorous feature selection to model recurrence risk without prespecified linear assumptions. By systematically comparing multiple algorithms and incorporating SHAP-based interpretability, our approach enables both robust discrimination and transparent understanding of individualized risk profiles. Notably, the final XGBoost model demonstrated excellent and consistent discriminative performance, achieving an AUC of 0.926 in the training cohort and 0.889 in the validation cohort, which compares favorably with previously reported models. These findings suggest that an IML model based on routinely available clinical variables may offer a more flexible and effective strategy for recurrence risk stratification after MWA in patients with sHCC.

The seven predictors retained in the final model are clinically coherent and reflect the interplay between tumor aggressiveness, systemic inflammatory status, and the underlying hepatic microenvironment, all of which are central to recurrence risk after curative-intent MWA. Tumor size emerged as a fundamental determinant of recurrence risk after MWA. Even within the spectrum of sHCC, increasing tumor diameter has been consistently associated with more aggressive tumor biology, including a higher likelihood of microscopic satellite nodules and microvascular invasion, which are not readily detectable on preoperative imaging but predispose to early relapse.^{25,26} From a technical standpoint, larger tumors pose greater challenges in achieving complete and homogeneous ablation with adequate safety margins, particularly in lesions approaching 3 cm or located near major vessels, where heat-sink effects and anatomical constraints may compromise energy deposition.^{27,28} These biological and procedural factors jointly explain the monotonic increase in recurrence risk with increasing tumor size observed in our SHAP analyses. Child-Pugh grade serves as a composite marker of hepatic functional reserve and chronic liver injury severity. Poorer Child-Pugh status may impair post-ablation liver regeneration and create a permissive microenvironment for residual tumor growth.^{29,30} Consistently, our findings indicate that higher Child-Pugh grade is associated with an increased risk of recurrence after MWA. Cirrhosis is a pathological state that develops after prolonged liver damage, often leading to chronic inflammation and fibrosis within the liver. Pinter et al³¹ found that most hepatocellular carcinomas arise against a cirrhotic background. In cirrhotic patients, disrupted liver architecture and abnormal blood circulation create favorable conditions for tumor cell growth and metastasis. Following microwave ablation therapy for small hepatocellular carcinomas, the presence of cirrhosis can impair the liver's repair and regenerative capacity. This impairs the liver's ability to eliminate residual tumor cells, allowing them to evade the body's immune surveillance and killing mechanisms, thereby increasing the risk of tumor recurrence.³² Our research also observed that patients with cirrhosis are more prone to recurrence after ablation. Portal hypertension is a clinical syndrome caused by elevated pressure within the portal venous system, primarily resulting from increased resistance to portal blood flow or augmented portal inflow.³³ Jang et al reported that portal hypertension is an important factor influencing long-term survival and tumor recurrence after local ablative therapies.³⁴ In patients with sHCC, portal hypertension readily induces alterations in portal hemodynamics, characterized by increased portal venous pressure and reduced blood flow velocity. Under such conditions, even after effective MWA, residual tumor cells may disseminate intrahepatically via the portal venous system, thereby increasing the risk of tumor recurrence. In addition, portal hypertension promotes the development of portosystemic collateral circulation, which may allow tumor cells to bypass normal hepatic metabolism and immune surveillance, further elevating the risk of distant dissemination and post-treatment recurrence.³⁵

PLT also contributed meaningfully to recurrence prediction. In a retrospective cohort of 172 patients, Pang et al demonstrated that PLT and multiple platelet-based indices were independently associated with postoperative recurrence in HCC.³⁶ In addition, a large meta-analysis of 5545 patients demonstrated that a low preoperative PLT significantly increased the risk of recurrence and poor recurrence-free survival in HCC.³⁷ In chronic liver disease, thrombocytopenia commonly reflects portal hypertension and advanced hepatic fibrosis and may additionally influence tumor progression through angiogenic and immunomodulatory pathways.³⁸ Consistently, lower PLT in our cohort was associated with an increased risk of post-ablation recurrence, highlighting the interplay between portal hemodynamics and host hematologic status. AFP is an important biomarker for the diagnosis and surveillance of hepatocellular carcinoma, and elevated AFP levels may indicate increased tumor activity or tumor burden.³⁹ In patients with sHCC, high serum AFP levels may promote tumor angiogenesis by modulating the peritumoral microenvironment, thereby providing essential nutrients and oxygen to support tumor cell proliferation and dissemination. In addition, elevated AFP has been shown to exert immunosuppressive effects on the host immune system, including inhibition of T-lymphocyte activity and attenuation of natural killer cell cytotoxicity, which

weakens immune surveillance and clearance of malignant cells and consequently increases the risk of tumor recurrence.⁴⁰ CRP, a marker of systemic inflammation, further highlights the role of host inflammatory status in recurrence risk.⁴¹ Chronic inflammation is a recognized driver of hepatocarcinogenesis and tumor progression, particularly in the cirrhotic liver.⁴² Elevated CRP levels may indicate a pro-tumorigenic inflammatory milieu that facilitates tumor cell survival, angiogenesis, and immune evasion.⁴³ The association between higher CRP levels and increased recurrence risk observed in our model supports the concept that systemic inflammation contributes to an unfavorable post-ablation tumor microenvironment. It is worth noting that TBIL was identified as a relevant feature in the initial Boruta screening but was not retained in the final model after LASSO selection. This discrepancy may be attributed to differences in the underlying principles of the two methods. Boruta aims to identify all potentially relevant variables, including those with weak or indirect associations, whereas LASSO favors a more parsimonious model by selecting variables with stronger and more stable predictive contributions. In the presence of other liver function-related variables, the effect of TBIL may have been relatively redundant or less informative, leading to its exclusion in the final model.

The current study presents several methodological strengths that enhance the robustness and clinical applicability of the predictive model. The use of a two-step feature selection strategy, incorporating both Boruta and LASSO regression, is a major strength. The Boruta algorithm effectively identified relevant variables by comparing original features to their randomized shadow counterparts, ensuring that only truly important features were retained. LASSO regression further refined the feature selection process, minimizing overfitting and ensuring that only variables with consistent predictive value were included in the final model. This combined approach minimized redundancy and ensured the stability of the selected predictors. Another strength of the study is its use of multiple ML models for comparison (SVM, RF, KNN, DT, XGBoost, and LR), which enabled a comprehensive evaluation of the most appropriate model. Among the six models, XGBoost demonstrated superior performance in terms of discrimination, achieving an AUC of 0.926 in the training set and 0.889 in the validation set. This robust performance was coupled with favorable calibration and clinical utility, as assessed through DCA, further supporting its potential for real-world application. The integration of SHAP for model interpretability is another important strength. SHAP provides a transparent framework for understanding how individual features contribute to predictions, both globally and at the patient level. This is particularly important for clinical adoption, as it allows clinicians to interpret the model's decisions and gain insights into the underlying risk factors contributing to recurrence, thus fostering trust in the model's output. The predictive model developed in this study has direct clinical implications for managing patients with sHCC following MWA. Tumor recurrence remains a major challenge, and the ability to predict recurrence risk early can guide clinicians in personalizing patient care. By identifying patients at higher risk for recurrence, clinicians can modify follow-up plans, increase the intensity of surveillance, and implement timely interventions, potentially improving long-term outcomes. By utilizing routinely collected clinical data, including tumor size, Child-Pugh grade, and biomarkers such as AFP and CRP, the model offers a practical tool that can be easily integrated into daily clinical workflows without the need for complex or expensive tests. This makes it a cost-effective and efficient solution for risk stratification in real-world settings. The inclusion of SHAP-based interpretability further strengthens the model's clinical value. SHAP allows clinicians to understand how each feature contributes to the prediction of recurrence, providing transparent, individualized explanations. This transparency fosters greater trust and confidence in the model, making it easier for healthcare providers to use the model as a supportive tool in clinical decision-making. While this study demonstrates promising results, there are several limitations that should be addressed in future research. First, the main limitation of this study is that all data were derived from a single center and no external validation was performed. Although internal validation was conducted, the absence of external validation substantially limits the generalizability of the model across different institutions and patient populations. Therefore, multicenter prospective studies with independent external validation are needed before this model can be more broadly applied in clinical practice. Second, although the study included a wide range of clinical and procedural variables, it did not consider all potential factors that might influence recurrence risk, such as genetic or molecular markers. Incorporating genomic or proteomic data into the model could further improve its accuracy and predictive power, particularly for patients with specific molecular subtypes of HCC. Future studies could explore the integration of these data to enhance the model's precision. Third, the follow-up period in this study was limited to one year, and the model was specifically developed to predict early recurrence after MWA. However, early and late recurrence of HCC may differ in their

underlying biological mechanisms, with early recurrence more likely related to residual microscopic disease or aggressive tumor biology, whereas late recurrence may be more closely associated with the carcinogenic background of the diseased liver. Therefore, predicting recurrence within one year may not fully reflect the long-term prognostic trajectory of patients. Future studies with longer follow-up are needed to assess the model's performance for predicting long-term recurrence patterns and outcomes. Fourth, an imbalance between recurrence and non-recurrence cases was present in the dataset, which may introduce potential bias in model training and prediction. Although multiple performance metrics were used to provide a comprehensive evaluation, class imbalance may still affect model performance. Future studies with more balanced datasets or the use of resampling techniques may further improve the robustness and generalizability of the model. Finally, in the present study, recurrence was defined as the presence of newly enhanced lesions either outside the ablation zone or elsewhere in the liver, without further differentiation between local tumor progression and intrahepatic distant recurrence. These recurrence patterns may be associated with different biological mechanisms and clinical implications. However, due to limitations in the available retrospective data, detailed classification of recurrence patterns was not consistently available. Therefore, we were unable to evaluate whether the predictive performance of the model differs between these subtypes. Future studies incorporating standardized definitions of recurrence patterns are needed to further refine and validate the model.

Conclusion

In conclusion, our study presents an IML model developed to predict tumor recurrence after MWA in patients with sHCC. By incorporating seven key predictors, including maximum tumor diameter, Child-Pugh grade, cirrhosis, portal hypertension, PLT, AFP, and CRP, the model demonstrated strong predictive performance with an AUC of 0.926 in the training set and 0.889 in the validation set. The application of SHAP analysis further enhanced the model's clinical value by providing clear explanations of how individual variables contribute to recurrence risk. These findings suggest that the model may serve as a useful and clinically accessible approach for individualized risk stratification and post-ablation management. However, because the model was developed and internally validated in a single-center cohort, further external validation in larger multicenter populations and longer follow-up are required to confirm its generalizability and long-term clinical utility. Future studies should focus on external validation and on exploring whether the integration of additional molecular or genomic data could further enhance its accuracy and clinical utility.

Data Sharing Statement

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

Ethics Approval and Consent to Participate

The study was approved by the Ethics Committee of The Zibo Central Hospital (No. 2024ZCH-10-163217), and informed consent was obtained from all patients. This study was conducted in accordance with the Declaration of Helsinki.

Author Contributions

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

Funding

There is no funding to report.

Disclosure

The authors declare that they have no competing interests.

References

- Alawiyia B, Constantinou C. Hepatocellular carcinoma: a narrative review on current knowledge and future prospects. *Curr Treat Options Oncol.* 2023;24(7):711–724. doi:10.1007/s11864-023-01098-9
- Wu X, Lokken RP, Mehta N. Optimal treatment for small HCC (<3 cm): resection, liver transplantation, or locoregional therapy? *JHEP Rep.* 2023;5(8):100781. doi:10.1016/j.jhepr.2023.100781
- Ahn KS, Kang KJ. Appropriate treatment modality for solitary small hepatocellular carcinoma: radiofrequency ablation vs. resection vs. transplantation? *Clin Mol Hepatol.* 2019;25(4):354–359. doi:10.3350/cmh.2018.0096
- Sensi B, Angelico R, Toti L, et al. Mechanism, potential, and concerns of immunotherapy for hepatocellular carcinoma and liver transplantation. *Curr Mol Pharmacol.* 2024;17:e18761429310703. doi:10.2174/0118761429310703240823045808
- Liang W, Hao W, Shao G, et al. Safety and feasibility of microwave ablation for hepatocellular carcinomas in the elderly: a systematic review. *Front Oncol.* 2022;12:855909. doi:10.3389/fonc.2022.855909
- Lachenmayer A, Tinguely P, Maurer MH, et al. Stereotactic image-guided microwave ablation of hepatocellular carcinoma using a computer-assisted navigation system. *Liver Int.* 2019;39(10):1975–1985. doi:10.1111/liv.14187
- Huang Z, Guo Z, Ni J, et al. Four types of tumor progression after microwave ablation of single hepatocellular carcinoma of ≤5 cm: incidence, risk factors and clinical significance. *Int J Hyperthermia.* 2021;38(1):1164–1173. doi:10.1080/02656736.2021.1962548
- Swan RZ, Sindram D, Martinie JB, Iannitti DA. Operative microwave ablation for hepatocellular carcinoma: complications, recurrence, and long-term outcomes. *J Gastrointest Surg.* 2013;17(4):719–729. doi:10.1007/s11605-013-2164-y
- Moteleub H, Moety A, Baddour NM, Shendidi AAE. Predictors for hepatocellular carcinoma recurrence after microwave ablation. *Clin Exp Hepatol.* 2020;6(2):77–84. doi:10.5114/ceh.2019.95115
- Groeschl RT, Wong RK, Quebbeman EJ, et al. Recurrence after microwave ablation of liver malignancies: a single institution experience. *HPB.* 2013;15(5):365–371. doi:10.1111/j.1477-2574.2012.00585.x
- Wang X, Gu D, Qin H, et al. Prediction model and nomogram of early recurrence of hepatocellular carcinoma after ultrasound-guided microwave ablation. *Curr Med Imaging.* 2024;20:e15734056291428. doi:10.2174/0115734056291428240523060933
- Zhang Y, Yu JJ, Chen W, et al. Multi-parameter MRI-based model for the prediction of early recurrence of hepatitis B-associated hepatocellular carcinoma after microwave ablation. *Front Cell Infect Microbiol.* 2025;15:1638779. doi:10.3389/fcimb.2025.1638779
- Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* 2015;16(4):e173–80. doi:10.1016/S1470-2045(14)71116-7
- Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008;26(8):1364–1370. doi:10.1200/JCO.2007.12.9791
- Gudivada IP, Amajala KC. Integrative bioinformatics analysis for targeting hub genes in hepatocellular carcinoma treatment. *Curr Genomics.* 2025;26(1):48–80. doi:10.2174/0113892029308243240709073945
- Xu W, Liao S, Hu Y, Huang Y, Zhou J. Upregulation of miR-3130-5p enhances hepatocellular carcinoma growth by suppressing ferredoxin 1: miR-3130-5p enhances HCC growth via inhibiting FDX1. *Curr Mol Pharmacol.* 2024;17:e18761429358008. doi:10.2174/0118761429358008250305070518
- Ye W, Wang J, Zheng Y, Jiang M, Zhou Y, Wu Z. Association between higher expression of Vav1 in hepatocellular carcinoma and unfavourable clinicopathological features and prognosis. *Protein Pept Lett.* 2024;31(9):706–713. doi:10.2174/0109298665330781240830042601
- Tseng HH, Wei L, Cui S, Luo Y, Ten Haken RK, El Naqa I. Machine learning and imaging informatics in oncology. *Oncology.* 2020;98(6):344–362. doi:10.1159/000493575
- Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 2021;13(1):152. doi:10.1186/s13073-021-00968-x
- Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
- Azodi CB, Tang J, Shiu SH. Opening the black box: interpretable machine learning for geneticists. *Trends Genet.* 2020;36(6):442–455. doi:10.1016/j.tig.2020.03.005
- Galle PR, Forner A, Llovet JM, et al. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. *J Hepatol.* 2018;69(1):182–236.
- Xie DY, Ren ZG, Zhou J, Fan J, Gao Q. 2019 Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr.* 2020;9(4):452–463. doi:10.21037/hbsn-20-480
- Ren H, An C, Fu W, et al. Prediction of local tumor progression after microwave ablation for early-stage hepatocellular carcinoma with machine learning. *J Cancer Res Ther.* 2023;19(4):978–987. doi:10.4103/jert.jert_319_23
- Hirokawa F, Hayashi M, Miyamoto Y, et al. Outcomes and predictors of microvascular invasion of solitary hepatocellular carcinoma. *Hepatol Res.* 2014;44(8):846–853. doi:10.1111/hepr.12196
- Okusaka T, Okada S, Ueno H, et al. Satellite lesions in patients with small hepatocellular carcinoma with reference to clinicopathologic features. *Cancer.* 2002;95(9):1931–1937. doi:10.1002/cncr.10892
- Lee MW, Han S, Gu K, Rhim H. Local ablation therapy for hepatocellular carcinoma: clinical significance of tumor size, location, and biology. *Invest Radiol.* 2025;60(1):53–59. doi:10.1097/RLI.0000000000001100
- Lencioni R, Crocetti L. Local-regional treatment of hepatocellular carcinoma. *Radiology.* 2012;262(1):43–58. doi:10.1148/radiol.11110144
- Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* 2018;391(10127):1301–1314. doi:10.1016/S0140-6736(18)30010-2
- Huang F, Gao J. Modified Child-Pugh grade vs albumin-bilirubin grade for predicting prognosis of hepatocellular carcinoma patients after hepatectomy. *World J Gastroenterol.* 2020;26(7):749–758. doi:10.3748/wjg.v26.i7.749

31. Pinter M, Trauner M, Peck-Radosavljevic M, Sieghart W. Cancer and liver cirrhosis: implications on prognosis and management. *ESMO Open*. 2016;1(2):e000042. doi:10.1136/esmoopen-2016-000042
32. Zanutti S, Boot GF, Coto-Llerena M, et al. The role of chronic liver diseases in the emergence and recurrence of hepatocellular carcinoma: an omics perspective. *Front Med*. 2022;9:888850. doi:10.3389/fmed.2022.888850
33. Ramkissoon R, Cao S, Shah VH. The pathophysiology of portal hypertension. *Clin Liver Dis*. 2024;28(3):369–381. doi:10.1016/j.cld.2024.03.001
34. Jang SW, Cho YK, Kim JW, Gil JR, Kim MY, Lee Y. Tumor recurrence in hepatocellular carcinoma patients after radiofrequency ablation: portal hypertension as an indicator of recurrence of hepatocellular carcinoma. *J Korean Soc Radiol*. 2018;79(5):264. doi:10.3348/jksr.2018.79.5.264
35. Allaire M, Rudler M, Thabut D. Portal hypertension and hepatocellular carcinoma: des liaisons dangereuses.... *Liver Int*. 2021;41(8):1734–1743. doi:10.1111/liv.14977
36. Pang Q, Zhang JY, Xu XS, et al. Significance of platelet count and platelet-based models for hepatocellular carcinoma recurrence. *World J Gastroenterol*. 2015;21(18):5607–5621. doi:10.3748/wjg.v21.i18.5607
37. Pang Q, Qu K, Zhang JY, et al. The prognostic value of platelet count in patients with hepatocellular carcinoma: a systematic review and meta-analysis. *Medicine*. 2015;94(37):e1431. doi:10.1097/MD.0000000000001431
38. Ramadori P, Klag T, Malek NP, Heikenwalder M. Platelets in chronic liver disease, from bench to bedside. *JHEP Rep*. 2019;1(6):448–459. doi:10.1016/j.jhepr.2019.10.001
39. Hu X, Chen R, Wei Q, Xu X. The landscape of alpha fetoprotein in hepatocellular carcinoma: where are we? *Int J Biol Sci*. 2022;18(2):536–551. doi:10.7150/ijbs.64537
40. Wang X, Wang Q. Alpha-fetoprotein and hepatocellular carcinoma immunity. *Can J Gastroenterol Hepatol*. 2018;2018:9049252. doi:10.1155/2018/9049252
41. Ren Y, Fan X, Chen G, Zhou D, Lin H, Cai X. Preoperative C-reactive protein/albumin ratio to predict mortality and recurrence of patients with hepatocellular carcinoma after curative resection. *Med Clin*. 2019;153(5):183–190. doi:10.1016/j.medcli.2018.11.010
42. Leone V, Ali A, Weber A, Tschaharganeh DF, Heikenwalder M. Liver inflammation and hepatobiliary cancers. *Trends Cancer*. 2021;7(7):606–623. doi:10.1016/j.trecan.2021.01.012
43. Yang C, Liang Z, Zhao L, Li R, Ma P. Prediction of microvascular invasion in hepatocellular carcinoma using a preoperative serum C-reactive protein-based nomogram. *Sci Rep*. 2025;15(1):522. doi:10.1038/s41598-024-84835-w

Journal of Hepatocellular Carcinoma

Publish your work in this journal

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

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

Dovepress
Taylor & Francis Group