Journal of Hepatocellular Carcinoma

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Integrating CEUS Imaging Features and LI-RADS Classification for Postoperative Early Recurrence Prediction in Solitary Hepatocellular Carcinoma: A Machine Learning-Based Prognostic Approach

Li Liang[1,2,*], Jinshu Pang[1,*], Bulin Zhang[2], Qiao Que[1], Ruizhi Gao[1], Yuquan Wu[1], Jinbo Peng[1], Wei Zhang[2], Xiumei Bai[1], Rong Wen[1], Yun He[1], Hong Yang[1]

[1]Department of Medical Ultrasound, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, People's Republic of China; [2]Department of Medical Ultrasound, Liuzhou People's Hospital, Liuzhou, Guangxi Zhuang Autonomous Region, People's Republic of China

*These authors contributed equally to this work

Correspondence: Hong Yang; Yun He, Department of Medical Ultrasound, The First Affiliated Hospital of Guangxi Medical University, No. 6 Shuangyong Road, Nanning, Guangxi Zhuang Autonomous Region, People's Republic of China, Tel +8607715356706, Email yanghong@gxmu.edu.cn; heyun@gxmu.edu.cn

---

Purpose: To develop and validate a machine learning (ML) model for predicting early postoperative recurrence in hepatocellular carcinoma (HCC) patients by integrating contrast-enhanced ultrasound (CEUS) features with Liver Imaging Reporting and Data System (LI-RADS) classification.

Materials and Methods: A retrospective analysis was conducted on data from 279 patients who underwent surgical resection for HCC. CEUS-derived features, including the LI-RADS classification, were integrated with clinical and pathological variables to construct predictive models. Patients were randomly assigned to training (n = 196) and validation (n = 83) cohorts in a 7:3 ratio. Feature selection was performed using univariate Cox regression ($p \leq 0.05$), and four ML algorithms—Random Survival Forest (RSF), Gradient Boosting Machine (GBM), CoxBoost, and XGBoost—were applied to develop recurrence prediction models. Model performance was evaluated using the concordance index (C-index), area under the curve (AUC), calibration curves, decision curve analysis (DCA), and Kaplan–Meier (KM) survival analysis.

Results: Five significant features identified by univariate Cox regression were included in model development: microvascular invasion (MVI), tumor size, LI-RADS classification, tumor necrosis, and arterial enhancement patterns. Among the four ML algorithms, GBM achieved the best overall performance, with the following results. The C-index for 1-year and 2-year recurrence prediction was 0.802 and 0.735 in the training cohort, and 0.804 and 0.710 in the validation cohort, respectively. The corresponding AUCs were 0.820 and 0.764 in the training cohort, and 0.817 and 0.716 in the validation cohort. Feature importance analysis identified LI-RADS classification, MVI, and tumor size as the top three prognostic indicators, while KM survival analysis confirmed the model's ability to stratify patients into distinct risk groups (training cohort: $p < 0.001$; validation cohort: $p = 0.003$).

Conclusion: The GBM-based ML model integrating CEUS imaging features and LI-RADS classification demonstrates potential for predicting early postoperative recurrence of HCC, which may assist in guiding follow-up strategies.

Keywords: hepatocellular carcinoma, CEUS, LI-RADS, early recurrence, machine learning, prognostic modeling

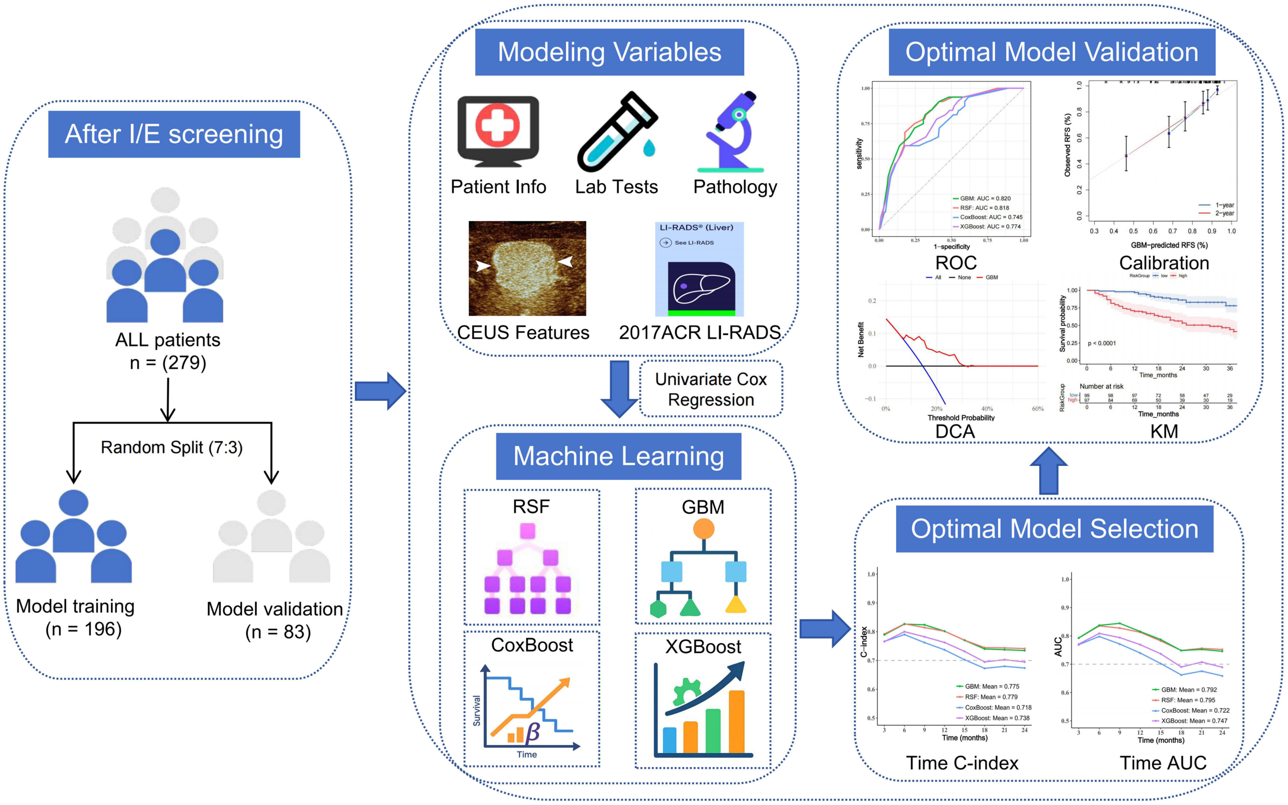
---

Background

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and remains a leading cause of cancer-related mortality worldwide.[1–3] Despite significant advancements in early detection and treatment strategies, the recurrence rate of HCC after curative resection remains high, with up to 70% of patients experiencing recurrence within five years.[4] This high recurrence rate underscores the importance of accurately identifying patients at high risk of recurrence and

Received: 9 April 2025
Accepted: 26 June 2025
Published: 3 July 2025

Graphical Abstract



implementing personalized surveillance and treatment strategies post-surgery.[5] Therefore, the ability to predict postoperative recurrence in HCC patients is critical for improving patient outcomes and guiding clinical decision-making.

Contrast-enhanced ultrasound (CEUS) is a valuable non-invasive imaging modality for assessing liver lesions, particularly in patients with chronic liver disease or cirrhosis. The Liver Imaging Reporting and Data System (LI-RADS), developed by the American College of Radiology, provides a standardized approach for categorizing liver lesions and assessing malignancy risk (American College of Radiology, 2017).[6] The integration of CEUS with LI-RADS has shown promising potential in identifying lesions with high malignancy risk, and more recently, it has been explored as a tool for predicting recurrence in HCC patients post-surgery.[7–10]

However, predicting HCC recurrence involves multiple factors, including tumor size, histopathological features, and serum biomarkers.[11–13] These factors alone are often insufficient for accurate prediction. Recent studies suggest that incorporating machine learning (ML) techniques can enhance predictive accuracy by analyzing complex, high-dimensional data from clinical, imaging, and laboratory sources.[14,15] These techniques enable the development of more robust models that account for intricate relationships within the data, ultimately improving prognostic predictions.

This study aims to develop and validate a ML model for predicting postoperative recurrence of HCC. We will integrate clinical, laboratory, histopathological, and CEUS features, including LI-RADS classification, into the model. The performance of four ML algorithms—Random Survival Forest (RSF), Gradient Boosting Machine (GBM), CoxBoost, and XGBoost—will be compared for their ability in survival analysis and risk stratification.

Materials and Methods

Study Population

This retrospective study analyzed adult patients who underwent surgical resection for HCC at the First Affiliated Hospital of Guangxi Medical University and Liuzhou People's Hospital between April 2019 and August 2023. Inclusion criteria were: (1) surgical resection of a solitary HCC with pathological confirmation, without vascular invasion, lymph node metastasis, or distant metastasis on imaging; (2) complete pathological evaluation; (3) preoperative CEUS within 4 weeks prior to surgery, meeting the 2017 ACR LI-RADS criteria; (4) liver function classified as Child-Pugh A or B; and (5) no history of other malignancies or active severe comorbidities (eg, liver failure or infections). Exclusion criteria were: (1) history of prior anti-tumor treatments (radiotherapy, chemotherapy, or other therapies) or concurrent malignancies; (2) incomplete or poor-quality CEUS data; and (3) loss to follow-up or follow-up duration < 12 months.

Clinical, Laboratory, and Histopathologic Characteristics

Relevant clinical and laboratory data were extracted from patients' medical records. Patient characteristics included gender, age, time of surgical resection, and chronic liver disease risk factors. Baseline laboratory data focused on key indicators, including total bilirubin (TB), albumin (ALB), aspartate aminotransferase (AST), alanine aminotransferase (ALT), alpha-fetoprotein (AFP), and des-gamma-carboxy prothrombin (DCP).

All resected liver tumor specimens and H&E-stained sections were reviewed by experienced hepatopathologists at each center. Tumor differentiation was assessed according to the WHO classification system.[16] Microvascular invasion (MVI) was defined as tumor cell clusters within the vasculature of the tumor capsule or adjacent non-tumor liver tissue, undetectable on gross examination.[17] Liver cirrhosis was diagnosed based on surgical pathology, with hepatic inflammation and fibrosis graded using the Galea and Ishak scoring systems.

Imaging Procedures and Image Evaluation

Imaging Procedures

In this study, all participants underwent CEUS within four weeks before surgery using the following equipment: (1) GE LOGIQ E9 (GE Healthcare, USA, C1-6VN probe), (2) Mindray Resona R9 (Mindray, China, SC6-1U probe), and (3) FUJIFILM ARIETTA 850 (FUJIFILM, Japan, C252 probe). CEUS parameters followed standard guidelines, regardless of the device. Examinations were performed by an experienced ultrasound specialist with over five years of liver imaging expertise. The tumor was initially located and assessed using grayscale imaging, measuring its largest diameter. After injection of 2.4 mL SonoVue (Bracco, Italy) and a 5-mL saline flush, CEUS imaging started immediately, capturing dynamic images for at least 60 seconds, followed by intermittent 5–10 second clips every 30 seconds for up to 5 minutes or until microbubbles dissipated. All data were saved in DICOM format for offline analysis.

Image Evaluation

All grayscale ultrasound (US) and CEUS images were independently reviewed by two ultrasound specialists with over five years of liver CEUS experience. The reviewers were blinded to clinical and pathological data. Disagreements were resolved through discussion, and if consensus could not be reached, two senior specialists with over ten years of liver ultrasound and CEUS experience made the final decision. Lesion size and echogenic patterns (hypoechoic, isoechoic, or hyperechoic) were recorded from grayscale images. During CEUS, enhancement patterns in the arterial, portal venous, and delayed phases were classified as hypo-, iso-, or hyper-enhancement. Arterial phase hyper-enhancement (APHE) was defined as complete or partial hyper-echogenicity, excluding peripheral nodular enhancement. Enhancement patterns were further classified as homogeneous, inhomogeneous, iso-enhancement, or rim enhancement, with rim APHE indicating stronger peripheral enhancement.[6] The presence of feeding arteries and proliferative arteries was documented.[18] Tumor margins were described as clear or obscure, and vascular abnormalities as irregular or proliferative vessel patterns.[19] Tumor necrosis was indicated by non-enhanced regions.[20] According to the 2017 ACR CEUS LI-RADS criteria,[6] "washout" refers to reduced lesion enhancement relative to the surrounding liver parenchyma during the portal venous or delayed phase, classified as early (within 60 seconds) or late (after 60 seconds). Washout intensity is further classified as "marked" (significant reduction or complete disappearance) or "mild" (less pronounced reduction). The LI-RADS classification is as follows:

- LR-M: Rim APHE (excluding peripheral nodular enhancement), or early washout, or marked washout.
- LR-5: Lesions ≥10 mm with APHE and mild late-phase washout.
- LR-4: Lesions <10 mm with LR-5 characteristics, or lesions ≥10 mm with APHE but no washout, or lesions ≥20 mm with mild late-phase washout but no APHE.
- LR-3: Lesions with arterial phase hypoenhancement or isoenhancement but no washout, regardless of nodule size, or lesions >20 mm with arterial phase hypo or isoenhancement and mild delayed washout.

Postoperative Follow-up

All patients were followed per a standardized protocol. One month after resection, contrast-enhanced CT or MRI was performed, followed by abdominal ultrasound or CT/MRI every 3–6 months. AFP and other laboratory tests were monitored. Intrahepatic recurrence was classified as local (within the resection area) or distant (remote from the margin). Extrahepatic recurrence was assessed using chest CT, brain MRI, whole-body bone scans, or other imaging. The primary endpoint, recurrence-free survival (RFS), was defined as the time from surgery to recurrence or tumor-related death. Patients were followed until recurrence, death, or the study cutoff (October 10, 2024). For missed follow-ups, the research team contacted family members to confirm survival status and cause of death.

Machine Learning Model Construction and Validation

Feature Selection and Model Development

In this study, four ML methods—GBM, RSF, CoxBoost, and XGBoost—were applied to develop relapse prediction models and identify key features associated with relapse. Univariate Cox regression analysis was first conducted to select variables with $p \leq 0.05$, which were then incorporated into the modeling. Multicollinearity analysis was performed to eliminate highly correlated variables, ensuring minimal collinearity and reducing potential bias.

For the GBM model, the maximum tree depth was set to 5, the minimum node sample size to 3, and the learning rate to 0.01. 10-fold cross-validation was used to determine the optimal parameters, with the number of trees set to 1000. Feature importance was calculated and visualized to identify key predictors for survival. In the RSF model, 1000 trees were constructed with a minimum node sample size of 3. At each split, two features were randomly selected (mtry = 2), and the Log rank test was used for splitting. Feature importance was standardized, and proximity analysis was performed to examine feature relationships. The CoxBoost model employed Cox regression boosting with L2 regularization. The penalty coefficient (initially set to 500) and the number of steps were optimized via 5-fold cross-validation, with parallel computation improving efficiency. Feature selection was based on the absolute values of the model coefficients. For the XGBoost model, Cox regression was used as the objective function, with the maximum tree depth set to 3 and the learning rate to 0.03. Column sampling and subsample ratios were set to 1.0, and a gamma value of 0.6 was applied. Early stopping after 50 rounds was used to prevent overfitting, and the contributions of different features were analyzed post-training.

Model Validation

To evaluate the performance of the prediction models, we employed a comprehensive validation strategy. The predictive ability of each model was first assessed using time-dependent area under the curve (time-AUC) and time-dependent concordance index (time-C-index), as well as the concordance index (C-index) and receiver operating characteristic (ROC) curves at the 1-year and 2-year time points. Calibration curves were utilized to assess model reliability, and decision curve analysis (DCA) was applied to evaluate clinical utility. The model with the best performance was then selected for further validation. Using the predicted values from this model, patients were stratified into high-risk and low-risk groups based on the median cutoff. Kaplan–Meier (KM) survival curves were generated to compare survival outcomes between the groups, with statistical significance determined by the Log rank test. All validation steps were conducted on both the training and test datasets to ensure robustness and generalizability. The entire study workflow is shown in Figure 1.

Statistical Analysis

Statistical analyses were conducted using R software (version 4.4.0) and IBM SPSS Statistics (version 27.0). Continuous variables were reported as the mean ± standard deviation (SD) for normally distributed data or the median (interquartile range,

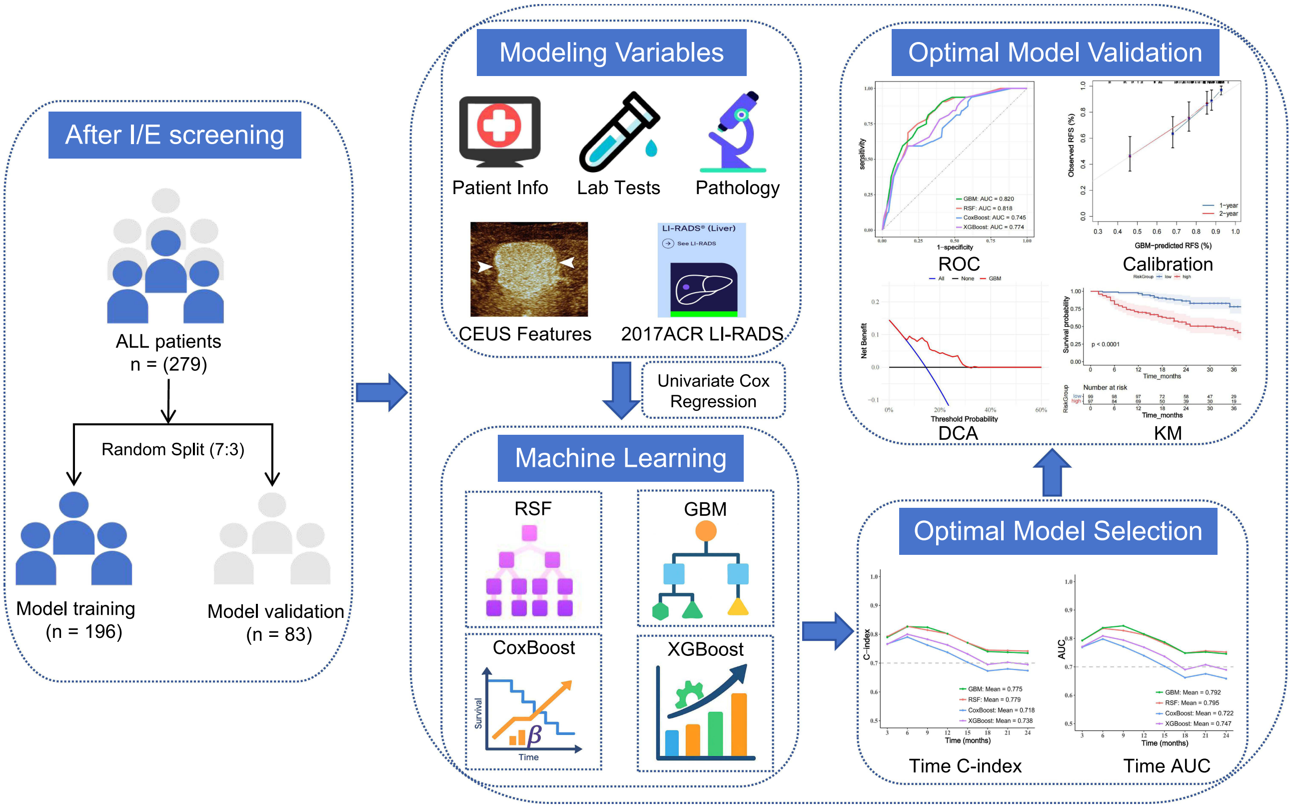


Figure 1 Flowchart of machine learning model development and validation for recurrence prediction.
Abbreviation: I/E, inclusion/exclusion criteria.

IQR) for skewed distributions. Normality was assessed before selecting appropriate tests. The independent *t*-test was applied for normally distributed variables, while the Mann–Whitney *U*-test was used for non-normally distributed data. Categorical variables were analyzed using the chi-square test (χ^2) or Fisher's exact test when expected frequencies were below 5.

Cox proportional hazards regression assessed factors influencing RFS. In univariate analysis, hazard ratios (HR), 95% confidence intervals (CI), and *p*-values were calculated, with $p \leq 0.05$ considered significant. Variables meeting this threshold proceeded to further analysis. Multicollinearity was examined using the generalized variance inflation factor (GVIF), where values exceeding 5 indicated moderate collinearity.

KM curves estimated RFS, with Log rank tests used for group comparisons. Median follow-up time was determined via the reverse Kaplan-Meier method. All statistical tests were two-tailed, with *p*-values < 0.05 deemed statistically significant.

Results

Baseline Characteristics

A total of 279 HCC cases were included in this study and divided into a training cohort (196 cases) and a validation cohort (83 cases) at a 7:3 ratio. The baseline characteristics of the two cohorts were comparable (Table 1). Overall, 88.53% of the patients were male and 11.47% were female. The median age was 52 years, with no significant difference between the training and validation cohorts ($p = 0.726$). The proportion of cases with tumor diameters between 2 cm and 5 cm was 63.80%, with no significant difference between the two cohorts ($p = 0.081$). The prevalence of liver cirrhosis was 49.82% ($p = 0.628$), and the majority of cases (96.06%) had HBV infection. Regarding pathological grading, moderate differentiation accounted for 81.00%, with no significant difference between the cohorts ($p = 0.453$). Based on the LI-RADS classification, LR-5 lesions comprised 61.65%, LR-M 34.77%, and LR-4 3.58%, with no significant difference between the cohorts ($p = 0.860$). Additionally, tumor grayscale echogenicity and contrast-enhanced ultrasound patterns showed no significant differences between the two cohorts ($p > 0.05$).

https://doi.org/10.2147/JHC.S530848

Table 1 Baseline Patient Characteristics by Cohort

Characteristics	ALL (n=279)	Training Cohort (n=196)	Validation Cohort (n=83)	*p* value
Recurrence Occurrence	97 (34.77%)	70 (35.71%)	27 (32.53%)	0.709
Gender of Male	247 (88.53%)	175 (89.29%)	72 (86.75%)	0.687
Age (y)[a]	52.00 [44.00; 60.00]	52.00 [44.75; 60.00]	53.00 [44.00; 60.50]	0.726
MVI Present	80 (28.67%)	56 (28.57%)	24 (28.92%)	1.000
WHO Grade				0.453
Well	22 (7.89%)	17 (8.67%)	5 (6.02%)	
Moderate	226 (81.00%)	155 (79.08%)	71 (85.54%)	
Poor	31 (11.11%)	24 (12.24%)	7 (8.43%)	
Cirrhosis Present	139 (49.82%)	100 (51.02%)	39 (46.99%)	0.628
Chronic HBV Present	268 (96.06%)	189 (96.43%)	79 (95.18%)	0.738
Tumor size				0.081
≤ 2cm	26 (9.32%)	23 (11.73%)	3 (3.61%)	
> 2 cm, ≤ 5 cm	178 (63.80%)	124 (63.27%)	54 (65.06%)	
> 5cm	75 (26.88%)	49 (25.00%)	26 (31.33%)	
AFP				0.431
≤ 100 ng/mL	191 (68.46%)	134 (68.37%)	57 (68.67%)	
> 100 ng/mL, ≤ 200 ng/mL	10 (3.58%)	9 (4.59%)	1 (1.20%)	
> 200 ng/mL	78 (27.96%)	53 (27.04%)	25 (30.12%)	
DCP				0.260
≤ 40 AU/mL	62 (22.22%)	48 (24.49%)	14 (16.87%)	
> 40 AU/mL, ≤100 AU/mL	58 (20.79%)	37 (18.88%)	21 (25.30%)	
> 100 AU/mL	159 (56.99%)	111 (56.63%)	48 (57.83%)	
TB > 34 μmol/L	7 (2.51%)	3 (1.53%)	4 (4.82%)	0.202
ALB > 36 g/L	236 (84.59%)	166 (84.69%)	70 (84.34%)	1.000
AST > 40 U/L	88 (31.54%)	68 (34.69%)	20 (24.10%)	0.109
ALT > 40 U/L	98 (35.13%)	67 (34.18%)	31 (37.35%)	0.712
US Echogenicity				0.596
Hypoechoic	229 (82.08%)	162 (82.65%)	67 (80.72%)	
Isoechoic	14 (5.02%)	11 (5.61%)	3 (3.61%)	
Hyperechoic	36 (12.90%)	23 (11.73%)	13 (15.66%)	
CEUS				
Arterial phase enhancement patterns				1.000
Homogeneous Hyperenhancement	166 (59.50%)	117 (59.69%)	49 (59.04%)	
Inhomogeneous Hyperenhancement	110 (39.43%)	77 (39.29%)	33 (39.76%)	
Isoenhancement	3 (1.08%)	2 (1.02%)	1 (1.20%)	
Feeding artery Present	209 (74.91%)	146 (74.49%)	63 (75.90%)	0.922
Chaotic vessels Present	114 (40.86%)	78 (39.80%)	36 (43.37%)	0.673
Boundary of the tumor enhancement				0.274
Clear	252 (90.32%)	180 (91.84%)	72 (86.75%)	
Obscure	27 (9.68%)	16 (8.16%)	11 (13.25%)	
Tumor necrosis Present	74 (26.52%)	53 (27.04%)	21 (25.30%)	0.879
ACR LIRADS 2017				0.860
LR-4	10 (3.58%)	8 (4.08%)	2 (2.41%)	
LR-5	172 (61.65%)	121 (61.73%)	51 (61.45%)	
LR-M	97 (34.77%)	67 (34.18%)	30 (36.14%)	

Notes: [a]Data are median values, and data in parentheses represent the interquartile range. Unless otherwise specified, data in parentheses are percentages.

Abbreviations: MVI, microvascular invasion; WHO, World Health Organization; HBV, Hepatitis B Virus; AFP, alpha-fetoprotein; DCP, des-γ-carboxy-prothrombin; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TB, total bilirubin; US, ultrasound; CEUS, contrast-enhanced ultrasound; ACR LIRADS 2017, 2017 American College of Radiology Liver Imaging Reporting and Data System.

Feature Selection and Model Development

The median follow-up time in this study was 33 months (95% CI: 31–35). In univariate Cox regression analysis, variables with $p \leq 0.05$ were identified as follows: MVI (HR = 2.16, 95% CI: 1.34–3.47, $p = 0.002$), tumor size > 5 cm (HR = 2.63, 95% CI: 1.00–6.89, $p = 0.050$), inhomogeneous hyperenhancement (HR = 1.91, 95% CI: 1.19–3.07, $p = 0.007$), tumor necrosis (HR = 1.98, 95% CI: 1.23–3.21, $p = 0.005$), and LR-M (HR = 1.97, 95% CI: 1.23–3.17, $p = 0.005$) (Figure 2). Additionally, the GVIF analysis showed that all variables had values below 5 (MVI: 1.0718, tumor size: 1.7935, arterial phase enhancement patterns: 2.1604, tumor necrosis: 2.2306, LIRADS: 1.2237), indicating minimal multicollinearity. These results confirm the independence of the selected features and support the stability of the model.

Based on these selected features, four predictive models were developed: GBM, RSF, CoxBoost, and XGBoost. Model performance in the training and validation cohorts was evaluated using the C-index and ROC curves. The C-index for RFS is shown in Table 2. In the training cohort, the C-index for predictions were as follows: GBM (1-year: 0.802, 2-year: 0.735, average: 0.775), RSF (1-year: 0.802, 2-year: 0.741, average: 0.779), CoxBoost (1-year: 0.737, 2-year: 0.674, average: 0.718), and XGBoost (1-year: 0.762, 2-year: 0.695, average: 0.738). In the validation cohort, the C-index for predictions

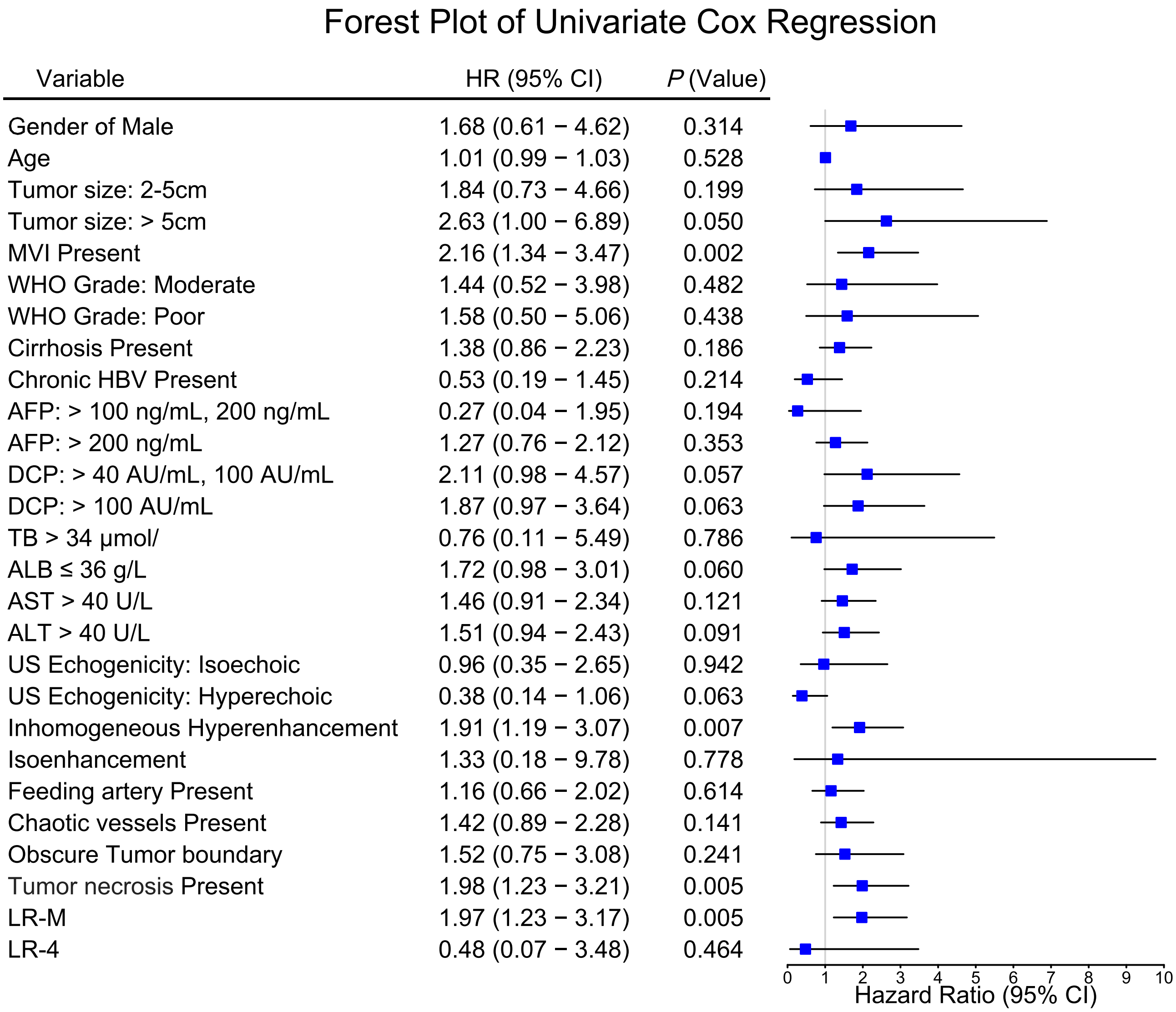


Figure 2 Forest plot of univariate Cox regression analysis.

Notes: Reference categories—tumor size: ≤2 cm; WHO grade: well-differentiated; AFP: ≤100 ng/mL; DCP: ≤40 AU/mL; US echogenicity: hypoechoic; arterial phase enhancement patterns: homogeneous hyperenhancement; LR: LR-5.

Abbreviations: HR, hazard ratio; CI, confidence interval.

Table 2 Performance of Prognostic Models Built by Machine Learning Algorithms in the Training and Validation Cohorts

Model	Time-AUC		Mean AUC	Time-C-Index		Mean C-Index
	1-Year RFS AUC	2-Year RFS AUC		1-Year RFS C-Index	2-Year RFS C-Index	
Training cohort						
GBM	0.820	0.764	0.792	0.802	0.735	0.775
RSF	0.818	0.772	0.795	0.802	0.741	0.779
CoxBoost	0.745	0.699	0.722	0.737	0.674	0.718
XGBoost	0.774	0.720	0.747	0.762	0.695	0.738
Validation cohort						
GBM	0.817	0.716	0.766	0.804	0.710	0.759
RSF	0.808	0.712	0.762	0.792	0.704	0.750
CoxBoost	0.747	0.680	0.714	0.743	0.673	0.726
XGBoost	0.763	0.689	0.726	0.751	0.679	0.724

Abbreviations: RFS, recurrence-free survival; C-index, concordance index; AUC, area under the receiver operating characteristic curve; GBM, gradient boosting machine; RSF, random survival forest; CoxBoost, Cox proportional hazards boosting; XGBoost, extreme gradient boosting.

were as follows: GBM (1-year: 0.804, 2-year: 0.710, average: 0.759), RSF (1-year: 0.792, 2-year: 0.704, average: 0.750), CoxBoost (1-year: 0.743, 2-year: 0.673, average: 0.726), and XGBoost (1-year: 0.751, 2-year: 0.679, average: 0.724).

The time-dependent AUC and C-index curves for all four models provide a comprehensive evaluation of performance over time (Figure 3A–D), while the corresponding 1-year and 2-year ROC curves illustrate discriminative ability for early recurrence prediction (Figure 3E–H). The AUC values for RFS are shown in Table 2, with the specific values as follows: In the training cohort, the AUC values were: GBM (1-year: 0.820, 95% CI: 0.745–0.894; 2-year: 0.764, 95% CI: 0.690–0.840; average: 0.792), RSF (1-year: 0.818, 95% CI: 0.746–0.890; 2-year: 0.772, 95% CI: 0.700–0.845; average: 0.795), CoxBoost (1-year: 0.745, 95% CI: 0.654–0.836; 2-year: 0.699, 95% CI: 0.616–0.782; average: 0.722), and XGBoost (1-year: 0.774, 95% CI: 0.691–0.858; 2-year: 0.720, 95% CI: 0.638–0.801; average: 0.747). In the validation cohort, the AUC values were: GBM (1-year: 0.817, 95% CI: 0.721–0.914; 2-year: 0.716, 95% CI: 0.600–0.835; average: 0.766), RSF (1-year: 0.808, 95% CI: 0.708–0.908; 2-year: 0.712, 95% CI: 0.594–0.830; average: 0.762), CoxBoost (1-year: 0.747, 95% CI: 0.620–0.874; 2-year: 0.680, 95% CI: 0.554–0.806; average: 0.714), and XGBoost (1-year: 0.763, 95% CI: 0.640–0.886; 2-year: 0.689, 95% CI: 0.564–0.813; average: 0.726).

Performance Validation of the Optimal Model

Given these findings, the GBM model outperformed the other models. Feature importance analysis using the Permutation Importance method identified MVI, tumor size, and LIRADS as the most influential factors, underscoring their critical role in predicting early recurrence (Figure 4).

To assess the GBM model's predictive performance over time, the 1-year and 2-year calibration curves demonstrated a strong alignment between predicted and observed recurrence probabilities, indicating good model calibration and reliability (Figure 5A and B). Additionally, DCA for both 1-year and 2-year predictions showed that the GBM model provides a modest net benefit across a range of decision thresholds, supporting its potential clinical utility (Figure 5C–F).

Kaplan-Meier Analysis and Risk Stratification

In this study, the GBM model stratified patients into high- and low-risk groups based on the median predicted value. KM survival analysis revealed significant differences in recurrence times between the two groups in both the training and validation cohorts, with the high-risk group showing shorter recurrence times (training cohort: Log rank test, $p < 0.001$; validation cohort: Log rank test, $p = 0.003$) (Figure 5G and H). These results confirm that the GBM model effectively predicts early recurrence and provides valuable risk stratification for hepatocellular carcinoma.

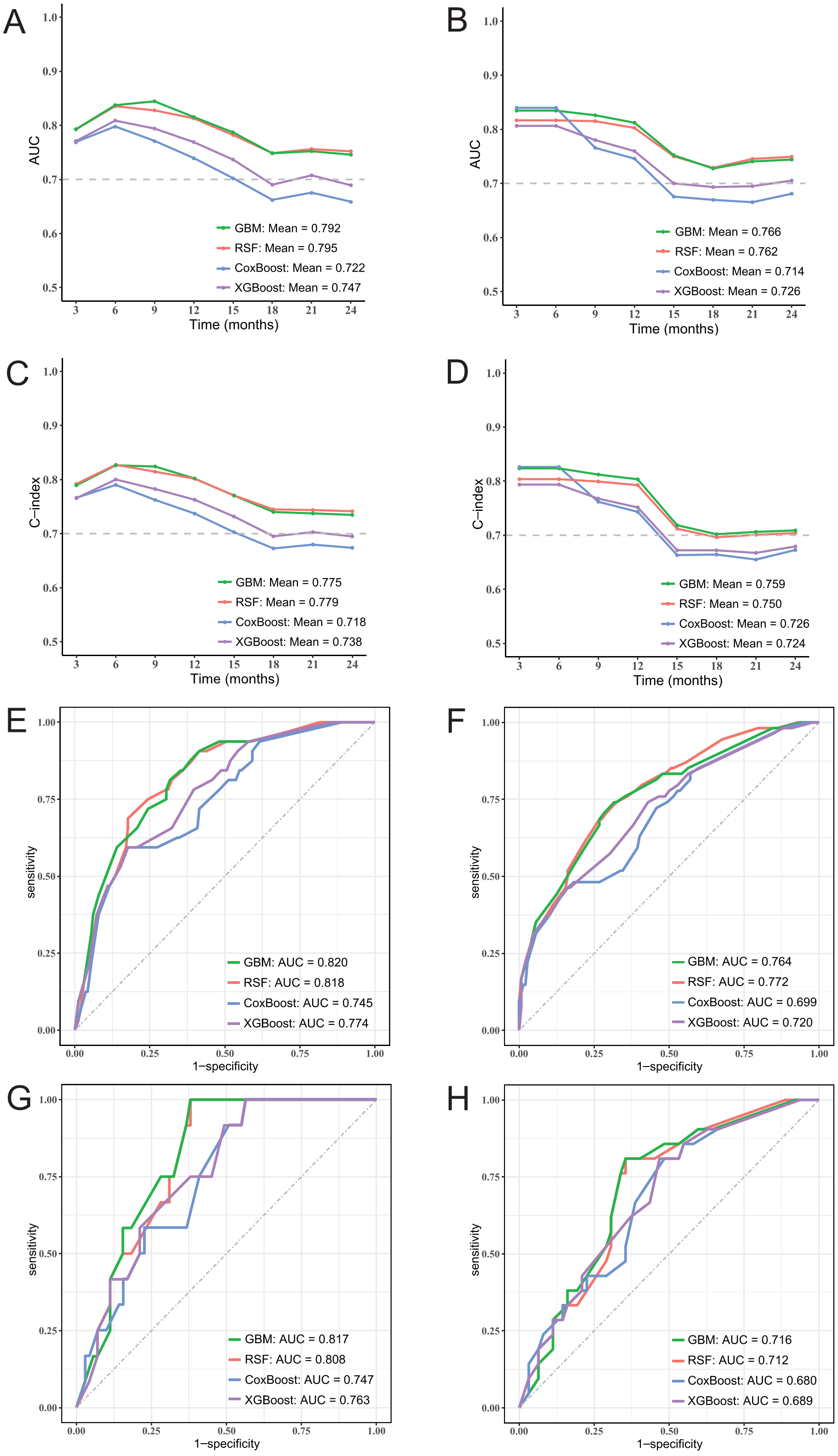


Figure 3 Predictive performance of machine learning models. (**A** and **B**) Time-dependent AUC curves in the training and validation cohorts, (**C** and **D**) time-dependent C-index curves in the training and validation cohorts, (**E** and **F**) 1-year and 2-year ROC curves in the training cohort, (**G** and **H**) 1-year and 2-year ROC curves in the validation cohort.

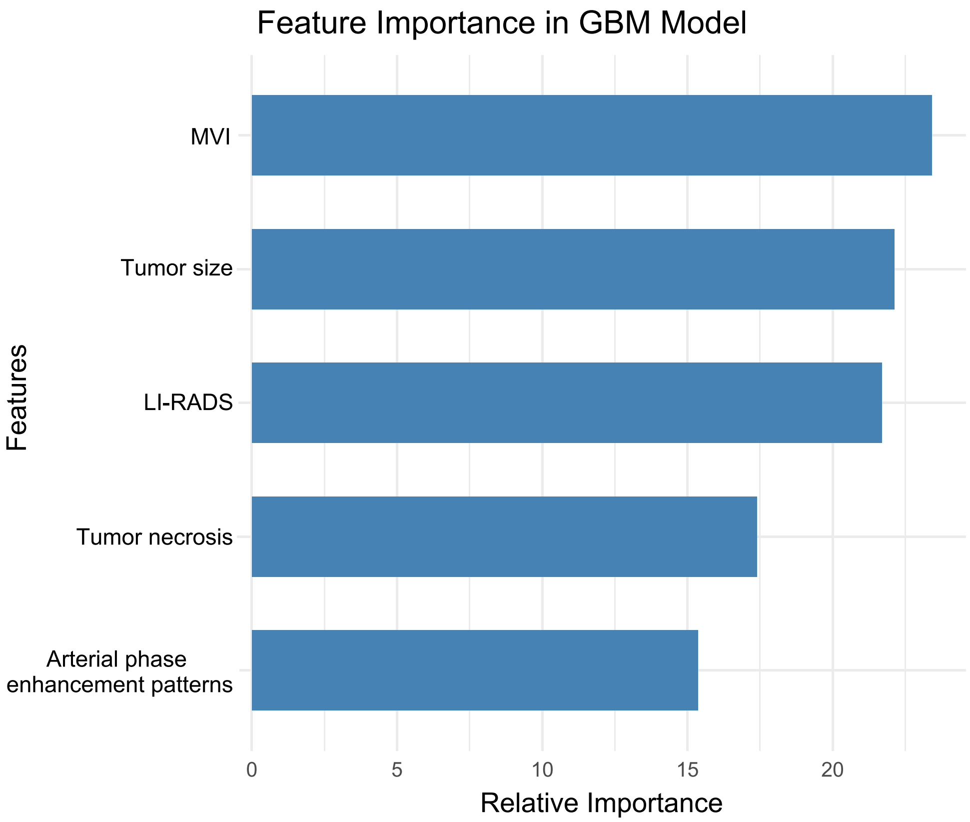


Figure 4 Feature importance analysis in the GBM model, highlighting the contribution of five key variables in model construction: Microvascular invasion (MVI), tumor size, LI-RADS classification, tumor necrosis, and arterial phase enhancement patterns.

Discussion

In recent years, ML has been increasingly applied to tumor diagnosis and prognosis prediction, particularly in the integration of multi-omics data. For instance, Chen et al[21] developed diagnostic (AUC = 0.967) and prognostic models (AUC = 0.832) for gastric cancer by combining metabolomics with ML, enabling precise identification of high-risk populations. Wang et al[22] further advanced this field by integrating 101 machine learning algorithms to construct the AIDPI prediction model, which uncovered the metabolic vulnerabilities of high-risk osteosarcoma patients and surpassed traditional clinical features in survival probability prediction. In the domain of radiomics, Lin et al[23] utilized T2WI and CE-T1WI imaging features in conjunction with the XGBoost algorithm to develop a recurrence prediction model for endometrial cancer, demonstrating that incorporating imaging features led to superior predictive performance compared to conventional clinical models. Furthermore, a multi-center study[24] established a 5-year survival prediction model for HCC patients following surgery, leveraging preoperative MRI radiomics features and random forest methods. This model achieved a mean AUC of 0.9804 in the training set and 0.7578 in the validation set, reinforcing the significant role of radiomics in HCC prognosis prediction.

Compared to MRI imaging studies, the application of CEUS in predicting postoperative recurrence of HCC remains relatively limited. Although previous studies have shown that preoperative CEUS arterial-phase "fast washout" enhancement patterns can serve as an independent predictor of early HCC recurrence,[25] research involving CEUS combined with ML to construct HCC prognosis models is still rare. This study develops and validates an early postoperative recurrence prediction model for HCC based on CEUS imaging features and clinical-pathological data. The results demonstrate that combining CEUS imaging features with machine learning methods improves the accuracy of recurrence risk assessment. Among various machine learning models, the GBM model performed best in both training and validation datasets, emphasizing the importance of CEUS imaging features, particularly the LI-RADS classification, in postoperative recurrence prediction and highlighting the value of ML in enhancing model prediction capability.

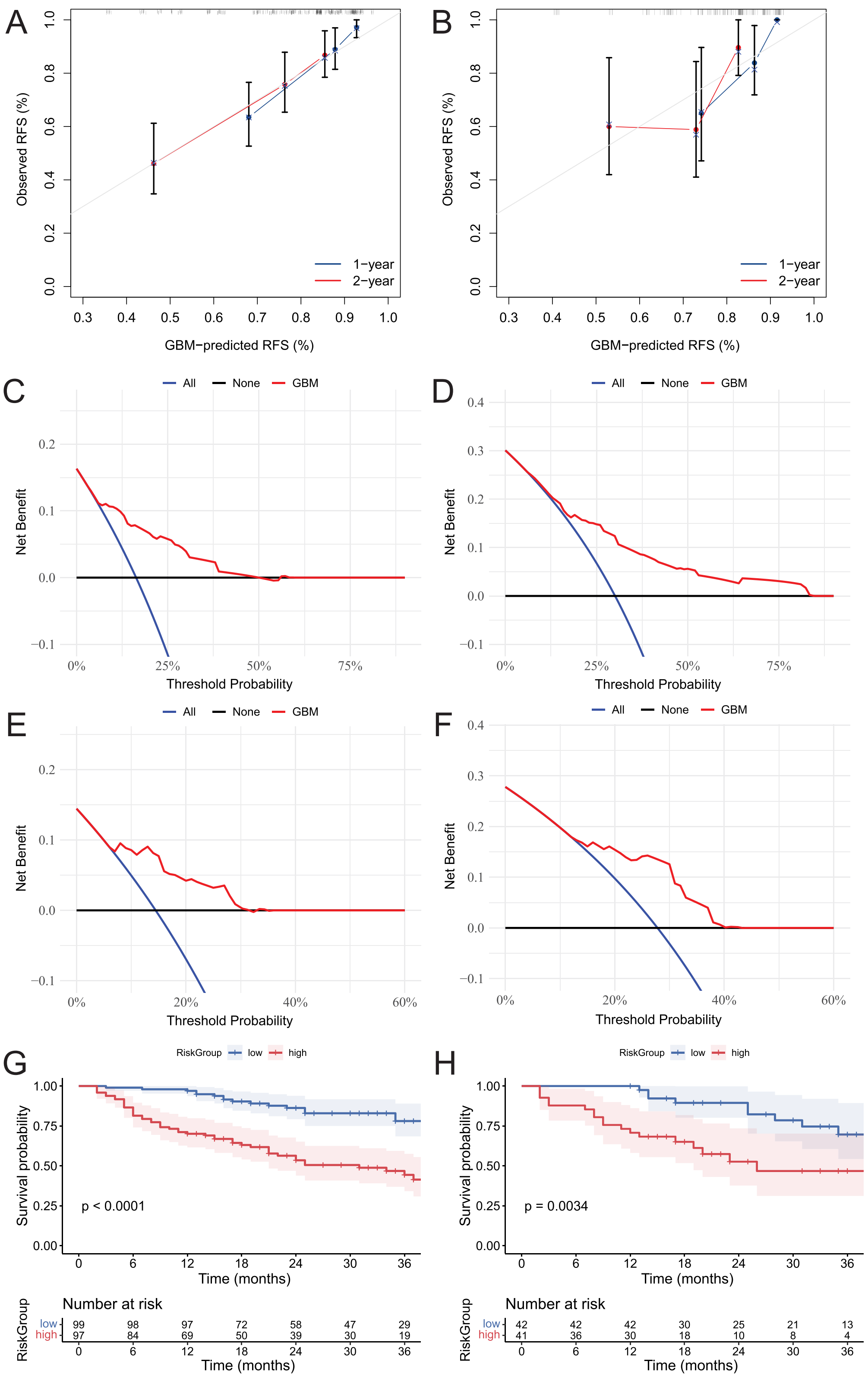


Figure 5 Performance evaluation of the GBM model for recurrence-free survival (RFS) prediction. (**A** and **B**) Calibration curves of the GBM model in the training (**A**) and validation (**B**) cohorts for 1-year (blue) and 2-year (red) RFS. (**C**–**F**) Decision curve analysis (DCA) for 1-year (**C** and **E**) and 2-year (**D** and **F**) RFS in the training (**C** and **D**) and validation (**E** and **F**) cohorts. The red line represents the GBM model. (**G** and **H**) Kaplan-Meier survival curves for RFS in the training (**G**) and validation (**H**) cohorts, stratified by GBM-predicted risk groups. Log-rank *p*-values are indicated.

Feature importance analysis revealed that MVI, tumor size, and LI-RADS classification made substantial contributions to predicting postoperative recurrence of HCC in the ML model. Previous studies have confirmed that MVI is a key determinant of HCC recurrence, closely linked to the ability of tumor cells to spread through blood vessels and significantly associated with poorer prognosis.[26,27] Additionally, research has demonstrated that tumor size plays a critical role in HCC prognosis. Larger tumors are more likely to experience MVI, higher recurrence risks, and poorer long-term survival rates. A retrospective study analyzing 813 HCC patients found that larger tumors were associated with a higher recurrence risk and lower long-term survival, with multivariate analysis confirming that tumor size > 5 cm is an independent risk factor, likely related to increased tumor load, vascular invasion, and intrahepatic micro-metastasis.[28] The inclusion of imaging features further enhanced the model's predictive ability. Previous studies have shown that LR-M lesions are independent risk factors for HCC prognosis, often associated with poorly differentiated, highly invasive tumors.[16,29,30] Features such as inhomogeneous hyperenhancement and tumor necrosis, identified through CEUS imaging, also showed high importance, suggesting that CEUS imaging can partially reflect the biological characteristics and heterogeneity of the tumor.[31–33] Therefore, imaging features not only hold significant value in predicting postoperative recurrence of HCC but are also key variables in constructing ML models. The combined analysis of CEUS imaging features and clinical factors can help improve predictive accuracy, providing more precise information for postoperative recurrence risk assessment in HCC.

Compared to traditional Cox regression models, ML methods offer greater flexibility and generalizability in predicting postoperative recurrence of HCC. While Cox regression models primarily rely on linear assumptions, ML methods can handle high-dimensional, nonlinear data, automatically identifying complex relationships between variables and optimizing model performance.[34,35] Additionally, ML techniques can use feature importance analysis to pinpoint key predictors, which helps avoid selection bias and improves clinical applicability. However, the "black-box" nature of machine learning models remains a significant limitation, as it hinders interpretability. Recent advancements in explainable artificial intelligence (XAI) methods, such as SHapley Additive Explanations (SHAP), allow for the quantification of individual feature contributions to predictions.[36–39] Future studies should incorporate SHAP or other interpretability tools to improve model transparency, enhance clinicians' understanding of the decision-making process, and facilitate the integration of machine learning models into clinical practice.

Despite the promising predictive results, several limitations exist in this study. Firstly, being a retrospective analysis, the study is subject to selection bias, as it relies on past data that may not be representative of the broader patient population. The findings could be influenced by biases in data collection, patient selection, and clinical practices. To address this, future prospective, multi-center studies are needed to validate the stability and generalizability of the model in diverse clinical settings. Secondly, although this study incorporated CEUS imaging features and laboratory test results, it did not consider other factors that could influence recurrence, such as genomic information, the immune microenvironment, and treatment-related variables. Genomic data and immune characteristics could offer valuable insights into recurrence risk. Future research should explore the integration of multi-omics data, combining radiomics, biomarkers, laboratory results, and clinical variables, to further enhance predictive accuracy. Thirdly, due to the retrospective design and exploratory nature of this study, a formal power calculation was not performed. To mitigate potential overfitting and enhance internal reliability, several model-specific strategies were applied, including univariate Cox regression for preliminary feature selection, regularization techniques (eg, CoxBoost and XGBoost), and internal validation methods such as 10-fold cross-validation and early stopping. Nevertheless, the relatively limited number of recurrence events may have constrained the statistical power of the analysis. Future prospective multi-center studies with predefined power calculations are necessary to validate and extend these findings.

Conclusion

This study developed and internally validated a machine learning-based model for predicting early postoperative recurrence in HCC, with the GBM model demonstrating the best performance. By integrating CEUS imaging features, particularly LI-RADS classification, the model offers a promising non-invasive approach for recurrence risk stratification. While the current findings are encouraging, further prospective, multi-center studies are warranted to support broader clinical application.

Ethical Declarations

This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No. 2025-E0188) and conducted in accordance with the Declaration of Helsinki and its later amendments. Given the retrospective design and the use of fully anonymized clinical data, including ultrasound imaging, laboratory test results, and pathological findings, the requirement for informed consent was waived by the ethics committee.

Acknowledgments

We thank the Key Laboratory of Ultrasonic Molecular Imaging and Artificial Intelligence, Guangxi Zhuang Autonomous Region Engineering Research Center for Artificial Intelligence Analysis of Multimodal Tumor Images, and Guangxi Key Laboratory of Early Prevention and Treatment for Regional High Frequency Tumor/Key Laboratory of Early Prevention and Treatment for Regional High Frequency Tumor (Guangxi Medical University), Ministry of Education for their support of this study.

Funding

This study was funded by the National Natural Science Foundation of China (82160336), the Joint Project on Regional High-Incidence Diseases Research of Guangxi Natural Science Foundation (2023GXNSFDA026013) and Guangxi Graduate Education Innovation Project (YCBZ2024119).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301–1314. doi:10.1016/S0140-6736(18)30010-2
2. Brown ZJ, Tsilimigras DI, Ruff SM, et al. Management of hepatocellular carcinoma: a review. *JAMA Surg*. 2023;158(4):410–420. doi:10.1001/jamasurg.2022.7989
3. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019;16(10):589–604. doi:10.1038/s41575-019-0186-y
4. Yao L, Chen Z, Feng Z, et al. Clinical features of recurrence after hepatic resection for early-stage hepatocellular carcinoma and long-term survival outcomes of patients with recurrence: a multi-institutional analysis. *Ann Surg Oncol*. 2022;29(7):4291–4303. doi:10.1245/s10434-022-11454-y
5. Singal AG, Kanwal F, Llovet JM. Global trends in hepatocellular carcinoma epidemiology: implications for screening, prevention and therapy. *Nat Rev Clin Oncol*. 2023;20(12):864–884. doi:10.1038/s41571-023-00825-3
6. American College of Radiology. CEUS LI-RADS® v2017 CORE. Available from: https://www.acr.org/media/ACR/Files/RADS/LIRADS/CEUS-LI-RADS-2017-Core.pdf. Accessed October 19, 2022.
7. Bartolotta TV, Terranova MC, Gagliardo C, Taibbi A. CEUS LI-RADS: a pictorial review. *Insights Imaging*. 2020;11(1):1–3. doi:10.1186/s13244-019-0819-2
8. Cunha GM, Sirlin CB, Fowler KJ. Imaging diagnosis of hepatocellular carcinoma: LI-RADS. *Chin Clin Oncol*. 2021;10(1):3. doi:10.21037/cco-20-107
9. Shin J, Lee S, Bae H, et al. Contrast-enhanced ultrasound liver imaging reporting and data system for diagnosing hepatocellular carcinoma: a meta-analysis. *Liver Int*. 2020;40(10):2345–2352. doi:10.1111/liv.14617
10. Schellhaas B, Bernatik T, Bohle W, et al. Contrast-enhanced ultrasound algorithms (CEUS-LIRADS/ESCULAP) for the noninvasive diagnosis of hepatocellular carcinoma: a prospective multicenter DEGUM study. *Ultraschall Med*. 2021;42(2):178–186. doi:10.1055/a-1198-4874
11. Fan RTPS, Wong J. Risk factors, prevention, and management of postoperative recurrence after resection of hepatocellular carcinoma. *Ann Surg*. 2000;232(1):10–24. doi:10.1097/00000658-200007000-00003
12. Xu X, Xing H, Han J, et al. Risk factors, patterns, and outcomes of late recurrence after liver resection for hepatocellular carcinoma. *JAMA Surg*. 2019;154(3):209. doi:10.1001/jamasurg.2018.4334
13. Cai Y, Xie K, Adeeb Alhmoud MN, et al. Effect of PIVKA-II and AFP secretion status on early recurrence of hepatocellular carcinoma after open and laparoscopic surgery. *Cancer Med*. 2023;12(17):17866–17877. doi:10.1002/cam4.6422
14. Wang F, Zhan G, Chen QQ, et al. Multitask deep learning for prediction of microvascular invasion and recurrence-free survival in hepatocellular carcinoma based on MRI images. *Liver Int*. 2024;44(6):1351–1362. doi:10.1111/liv.15870
15. Gross M, Haider SP, Ze Evi T, et al. Automated graded prognostic assessment for patients with hepatocellular carcinoma using machine learning. *Eur Radiol*. 2024;34(10):6940–6952. doi:10.1007/s00330-024-10624-8
16. Cannella R, Matteini F, Dioguardi Burgio M, et al. Association of LI-RADS and histopathologic features with survival in patients with solitary resected hepatocellular carcinoma. *Radiology*. 2024;310(2):e231160. doi:10.1148/radiol.231160
17. Cong W, Bu H, Chen J, et al. Practice guidelines for the pathological diagnosis of primary liver cancer: 2015 update. *World J Gastroenterol*. 2016;22(42):9279. doi:10.3748/wjg.v22.i42.9279
18. Wu J, Zhao Q, Wang Y, et al. Feeding artery: a valuable feature for differentiation of regenerative nodule, dysplastic nodules and small hepatocellular carcinoma in CEUS LI-RADS. *Eur Radiol*. 2024;34(2):745–754. doi:10.1007/s00330-023-10006-6

19. Xian MF, Huang Y, Xie WX, et al. LR-M observations on contrast-enhanced ultrasound: detection of hepatocellular carcinoma using additional features in comparison with current LI-RADS criteria. *AJR Am J Roentgenol*. 2022;219(1):76–85. doi:10.2214/AJR.21.26837
20. Fei X, Zhu L, Han P, et al. Value of high frame rate contrast-enhanced ultrasound in predicting microvascular invasion of hepatocellular carcinoma. *Insights Imaging*. 2024;15(1):273. doi:10.1186/s13244-024-01821-6
21. Chen Y, Wang B, Zhao Y, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun*. 2024;15(1):1657. doi:10.1038/s41467-024-46043-y
22. Wang Y, Ma X, Xu E, et al. Identifying squalene epoxidase as a metabolic vulnerability in high-risk osteosarcoma using an artificial intelligence-derived prognostic index. *Clin Transl Med*. 2024;14(2):e1586. doi:10.1002/ctm2.1586
23. Lin Z, Wang T, Li Q, et al. Development and validation of MRI-based radiomics model to predict recurrence risk in patients with endometrial cancer: a multicenter study. *Eur Radiol*. 2023;33(8):5814–5824. doi:10.1007/s00330-023-09685-y
24. Wang X, Long L, Cui Y, et al. MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma. *Br J Cancer*. 2020;122(5):695–703. doi:10.1038/s41416-019-0706-0
25. Wang Y, Liao J, Qi W, Xie L, Li Y. Predictive value of conventional ultrasound and contrast-enhanced ultrasound in early recurrence of hepatocellular carcinoma after surgical resection. *Ultrasound Med Biol*. 2016;42(5):1042–1048. doi:10.1016/j.ultrasmedbio.2015.12.010
26. Wu F, Sun H, Zhou C, et al. Prognostic factors for long-term outcome in bifocal hepatocellular carcinoma after resection. *Eur Radiol*. 2023;33(5):3604–3616. doi:10.1007/s00330-023-09398-2
27. Xu X, Diao Y, Zeng Y, et al. Association of severity in the grading of microvascular invasion with long-term oncological prognosis after liver resection for early-stage hepatocellular carcinoma: a multicenter retrospective cohort study from a hepatitis B virus-endemic area. *Int J Surg*. 2023;109(4):841–849. doi:10.1097/JS9.0000000000000325
28. Liang B, Gu J, Xiong M, et al. Tumor size may influence the prognosis of solitary hepatocellular carcinoma patients with cirrhosis and without macrovascular invasion after hepatectomy. *Sci Rep*. 2021;11(1):16343. doi:10.1038/s41598-021-95835-5
29. Heo S, Kang HJ, Choi SH, et al. Proliferative hepatocellular carcinomas in cirrhosis: patient outcomes of LI-RADS category 4/5 and category M. *Eur Radiol*. 2024;34(5):2974–2985. doi:10.1007/s00330-023-10305-y
30. Min JH, Kim SH, Hwang JA, et al. Prognostic value of LI-RADS category on gadoxetic acid-enhanced MRI and 18F-FDG PET-CT in patients with primary liver carcinomas. *Eur Radiol*. 2021;31(6):3649–3660. doi:10.1007/s00330-020-07378-4
31. Yen Y, Kuo F, Eng H, et al. Tumor necrosis as a predictor of early tumor recurrence after resection in patients with hepatoma. *PLoS One*. 2023;18(11):e292144. doi:10.1371/journal.pone.0292144
32. Wei T, Zhang X, Bagante F, et al. Tumor necrosis impacts prognosis of patients undergoing curative-intent hepatocellular carcinoma. *Ann Surg Oncol*. 2021;28(2):797–805. doi:10.1245/s10434-020-09390-w
33. Ling Y, Chen J, Wen S, et al. Tumor necrosis as a poor prognostic predictor on postoperative survival of patients with solitary small hepatocellular carcinoma. *BMC Cancer*. 2020;20(1):1–9. doi:10.1186/s12885-020-07097-5
34. Wu X, Yuan X, Wang W, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension. *Hypertension*. 2020;75(5):1271–1278. doi:10.1161/HYPERTENSIONAHA.119.13404
35. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. 2021;11(1):6968. doi:10.1038/s41598-021-86327-7
36. Wang Y, Lang J, Zuo JZ, et al. The radiomic-clinical model using the SHAP method for assessing the treatment response of whole-brain radiotherapy: a multicentric study. *Eur Radiol*. 2022;32(12):8737–8747. doi:10.1007/s00330-022-08887-0
37. Mastropietro A, Feldmann C, Bajorath J. Calculation of exact Shapley values for explaining support vector machine models using the radial basis function kernel. *Sci Rep*. 2023;13(1):19561. doi:10.1038/s41598-023-46930-2
38. Lieslehto J, Tiihonen J, Lähteenvuo M, et al. Development and validation of a machine learning-based model of mortality risk in first-episode psychosis. *JAMA Network Open*. 2024;7(3):e240640. doi:10.1001/jamanetworkopen.2024.0640
39. Ding R, Deng M, Wei H, et al. Machine learning-based prediction of clinical outcomes after traumatic brain injury: hidden information of early physiological time series. *CNS Neurosci Ther*. 2024;30(7):e14848. doi:10.1111/cns.14848