scientific reports

Check for updates

OPEN

Machine learning integrating MRI and clinical features predicts early recurrence of hepatocellular carcinoma after resection

Lijuan Feng[1,5], Ningbin Luo[2,5], Fengqiu Ruan[1], Xihuan Zheng[2], Xiaoyu Pan[1], Xuan Li[1], Liang Fu[1] & Liling Long[1,3,4]✉

This study aims to construct a robust artificial intelligence (AI) model to predict early recurrence of hepatocellular carcinoma (HCC) following surgical resection, leveraging clinical blood biomarkers, pathological parameters, and MRI-derived features. We included 240 hepatectomy patients from two medical centers, collecting clinical blood biomarkers, MRI features, and postoperative pathological data. Feature reduction was conducted using Spearman correlation and the least absolute shrinkage and selection operator (LASSO) regression. Predictive models were constructed using five machine learning algorithms and validated on an external dataset. The models were subsequently compared. The ExtraTrees, XGBoost, and LightGBM models exhibited high predictive performance in the training set, with AUCs of 0.816 (95% CI 0.748–0.884), 0.978 (95% CI 0.958–0.998), and 0.898 (95% CI 0.846–0.950), respectively. In the validation set, their AUC values were 0.759 (95% CI 0.641–0.876), 0.789 (95% CI 0.684–0.894), and 0.760 (95% CI 0.650–0.869). Decision curve analysis indicated favorable net benefits for predicting early recurrence across all three models. Tumor margin and age were identified as significant factors, showing strong associations with early recurrence. This study developed AI model utilizing clinical blood biomarkers, MRI features, and pathological information to predict early recurrence of HCC after surgery. The models demonstrated good predictive performance and showed clinical applicability in predicting early recurrence, potentially assisting clinicians in identifying high-risk patients, guiding individualized surveillance, and optimizing postoperative management. However, inherent biases in this retrospective study necessitate further research for validation and refinement.

Keywords Hepatocellular carcinoma, Early recurrence, MRI, Machine learning, AI

Abbreviations

AI	Artificial intelligence
HCC	Hepatocellular carcinoma
LASSO	The least absolute shrinkage and selection operator
AFP	Alpha-fetoprotein levels
NEUT	Neutrophi
LY	Lymphocyte
NLR	Neutrophil-to-lymphocyte ratio
PLT	Platelet count
ALB	Albumin
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase

[1]Department of Radiology, First Affiliated Hospital of Guangxi Medical University, No. 6, Shuangyong Road, Zhuangautonomous region, Nanning 530021, Guangxi, People's Republic of China. [2]Department of Radiology, Guangxi Medical University Cancer Hospital, Nanning, China. [3]Key Laboratory of Early Prevention and Treatment for Regional High Frequency Tumor, Gaungxi Medical University, Ministry of Education, Nanning, Guangxi, China. [4]Guangxi Key Laboratory of Immunology and Metabolism for Liver Diseases, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China. [5]These authors contributed equally: Lijuan Feng and Ningbin Luo. ✉email: cjr.longliling@vip.163.com

PA	Prealbumin
GGT	Gamma-glutamyl transferase
TBIL	Total bilirubin
DBIL	Direct bilirubin
PT	Prothrombin time
INR	International normalized ratio
HBP	Hepatobiliary phase
BCLC	Barcelona clinical liver cancer
MVI	Microvascular invasion
ES	Edmondson–Steiner
AUC	Area under the receiver operating characteristic curve
DCA	Decision curve analysis

Hepatocellular carcinoma (HCC) is the most prevalent form of primary liver cancer and a leading cause of cancer-related deaths worldwide[1]. Despite advances in treatment, the long-term prognosis remains suboptimal due to high postoperative recurrence rates, ranging from 61.4 to 83.3%[2]. Early recurrence (within 2 years postoperative) is particularly critical, as it typically originates from pre-existing intrahepatic metastases and indicates aggressive tumor biology and poor survival[3]. Therefore, preoperative identification of recurrence risk factors is crucial for improving postoperative outcomes in HCC and ensuring timely follow-up.

Current predictive methods face several challenges. Traditional predictors, such as Alpha-fetoprotein (AFP) and tumor size, rely on linear assumptions that fail to capture the complex, non-linear interactions of prognostic variables[4–6]. Radiomics has been used to predict postoperative recurrence of HCC[7–10], but the segmentation process is complex and requires improvements in automatic lesion detection accuracy, model generalizability, and robustness[11]. Genomics explores disease pathogenesis and recurrence mechanisms at the molecular level, offering precise intervention strategies. High sequencing costs and specific technical platform requirements, however, place heavy economic burdens on patients, limiting widespread application in resource-limited settings[12–14]. Therefore, developing an economical, accurate, and widely applicable model is crucial.

Machine learning (ML) has emerged as a powerful approach to address these limitations by integrating multidimensional data. Previous studies have demonstrated that ML models utilizing clinical and imaging features outperform traditional statistical methods in predicting HCC recurrence[6,15,16]. However, robust models that comprehensively integrate clinical blood biomarkers, MRI features, and pathological parameters are still needed to enhance predictive accuracy. This study aims to develop and validate ML models using multimodal data to predict early recurrence of HCC after resection, providing a practical tool for individualized risk assessment.

Materials and methods

Patients' population

The study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University and the Ethics Committee of the Affiliated Cancer Hospital of Guangxi Medical University, in accordance with the Declaration of Helsinki. The requirement for informed consent was waived by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University and the Ethics Committee of the Affiliated Cancer Hospital of Guangxi Medical University. The inclusion criteria were: (1) patients with HCC who underwent hepatectomy without distant metastasis; (2) preoperative MRI with contrast enhancement performed within 2 months before surgery; (3) BCLC stage 0, A or B[17]: as these stages are guideline-recommended for curative hepatectomy due to localized or moderately progressive intrahepatic tumors without distant metastasis or uncontrollable portal hypertension; and (4) no preoperative chemotherapy, radiotherapy, embolization, radiofrequency ablation, or other anticancer treatments. The exclusion criteria were: (1) presence of concurrent malignancies at other sites; (2) incomplete clinical laboratory data; (3) incomplete or poor-quality MRI images; (4) incomplete pathological data, including histological grade and microvascular invasion; and (5) patients lost to follow-up after surgery.

169 eligible patients from the First Affiliated Hospital of Guangxi Medical University (December 2015 to December 2018) formed the training set. An additional 79 eligible patients from the Affiliated Cancer Hospital of Guangxi Medical University were used as the testing set. The selection process and study design are illustrated in Fig. 1.

Clinical laboratory examinations

Preoperative clinical laboratory indicators were extracted from the electronic medical record systems of two hospitals, including sex, age, neutrophil (NEUT), lymphocyte (LY), neutrophil-to-lymphocyte ratio (NLR), platelet count (PLT), albumin (ALB), alpha-fetoprotein (AFP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), prealbumin (PA), gamma-glutamyl transferase (GGT), total bilirubin (TBIL), direct bilirubin (DBIL), prothrombin time (PT), international normalized ratio (INR), CA199, CA125, CEA, CA153, HBsAg, diabetes, hypertension, intraoperative hemorrhage, and ascites.

MRI imaging analysis

We collected MRI images from four machines across two hospitals, including axial dual-echo (in-phase and opposed-phase) T1-weighted imaging, fat-suppressed T2-weighted imaging, arterial phase, venous phase, and hepatobiliary phase (HBP). Scanning parameters are available in the Supplementary parameters. All images were obtained from the Digital Imaging and Communications in Medicine format within the Picture Archiving and Communication System. Two independent radiologists with 5 and 10 years of experience in abdominal imaging, respectively, qualitatively assessed the imaging features of liver lesions without knowledge of clinical

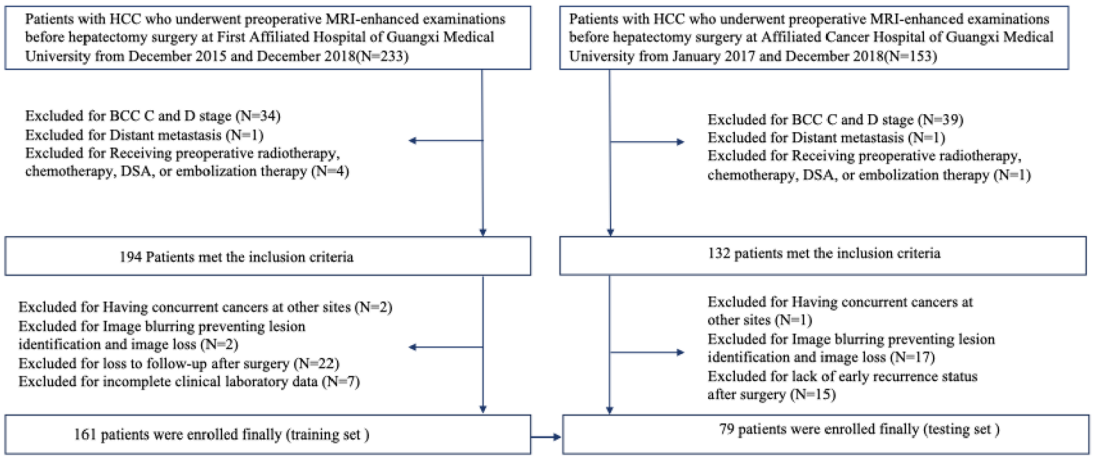

Fig. 1. Flowchart showing patient selection.

pathology results or image interpretations. The assessment was based on principles outlined in the Liver Imaging Reporting and Data System (LI-RADS v2018). The assessed features included: (1) tumor size; (2) tumor number; (3) Tumor margin; (4) capsule: (a) no capsule, (b) incomplete capsule and complete capsule; (5) arterial phase hyperenhancement: (a) rim arterial phase hyperenhancement, (b) nonrim arterial phase hyperenhancement; (6) peritumoral HBP hypointensity, defined as a wedge-shaped or flame-like hypointense area of hepatic parenchyma located outside the tumor margin on HBP images; (7) intratumoral artery; (8) washout: (a) non washout, (b) non-rim washout; (9) mosaic appearance; (10) intratumoral necrosis; (11) intratumoral hemorrhage; (12) intratumoral fat; (13) Arterial peritumoral enhancement.

Pathologic variables

Microvascular invasion (MVI) was diagnosed from postoperative tumor sections stained with hematoxylin. MVI is defined as clusters of more than 50 cancerous endothelial cells in the microvascular lumen[18].Barcelona Clinical Liver Cancer (BCLC) stage. In our study, HCC is classified into low (grades I or II) and high (grades III or IV) pathological Edmondson-Steiner (ES) grades[19].

Follow-up

Patients were followed every 3 months for the first 2 years and every 3–6 months thereafter. Follow-up data were primarily retrieved from electronic medical records, with telephone interviews conducted for patients with incomplete records to verify recurrence status. Follow-up included liver function tests, serum AFP levels, and imaging such as abdominal ultrasound, contrast-enhanced CT, and MRI. Recurrent HCC was diagnosed based on imaging findings (CT and/or MRI) and elevated AFP levels.

Feature extraction and selection

The features included clinical laboratory variables as previously described, MRI qualitative characteristics, and pathological variables, with some being binary and others continuous. Continuous features were first standardized using z-score normalization (mean = 0, standard deviation = 1) to conform to a standard normal distribution. Spearman's rank correlation coefficient was then used to evaluate the correlation between all features. When the Spearman correlation coefficient between two features was > 0.9, one of the highly correlated features was retained. Finally, least absolute shrinkage and selection operator (LASSO) regression with L1 regularization was applied for feature selection.

Development and validation of models

This study developed binary outcome models for early recurrence of HCC using ExtraTrees, XGBoost, LightGBM, and GradientBoosting machine learning algorithms. To minimize overfitting, hyperparameters were optimized using automated Grid Search. The final configurations, selected based on optimal AUC, are detailed in Supplementary Table 3. Model performance was evaluated using metrics such as the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. The DeLong test was used to compare the differences in AUC values among the machine learning models. Subsequently, clinical decision curve analysis (DCA) was performed to reflect the net benefit of different threshold probabilities in both the training and testing sets, assessing the clinical utility of the models. The calibration curve was used to validate the consistency between the observed outcomes and predicted probabilities.

Statistical analysis

In terms of basic characteristics, categorical variables are presented as counts (n) and percentages (%). Chi-square test or Fisher's exact test is used to compare categorical variables between the training and validation sets, while Mann–Whitney U test is used for continuous variables. Continuous variables with normal distribution

are expressed as mean ± standard deviation, and those with non-normal distribution are expressed as median (range). Inter-rater reliability of MRI feature assessments between the two radiologists was quantified using Cohen's kappa coefficient. All statistical analyses are performed using SPSS version 26 and R version 4.0.3 (R Project for Statistical Computing). A p-value < 0.05 is considered statistically significant.

Results

Patient characteristics

This study included 240 patients with HCC and the baseline characteristics of these patients are summarized in Table 1. A total of 89 cases (37.1%) experienced early recurrence. Among them, in the training set ($n = 161$), 59 cases (36.6%) had early recurrence; in the validation set ($n = 79$), 30 cases (38.0%) experienced early recurrence.

In both the training and validation sets, Size, CA153, and Edmondson were found to be significantly associated with early recurrence ($p < 0.05$). Washout, blood products in mass, size, Edmondson grade, TBiL, DBiL, ALB, INR, and PT showed statistically significant differences between the two cohorts. A comprehensive comparison of all variables included in the analysis between the training and validation sets is provided in Supplementary Table 1. The Cohen's kappa values for MRI feature assessment were all greater than 0.85, indicating excellent inter-rater agreement (see Supplementary Table 2 for details).

Feature selection

After removing highly correlated features via Spearman correlation analysis (coefficient > 0.9), we applied LASSO regression with an optimal λ value of 0.0256 (Fig. 2A,B) to identify 14 key features with non-zero coefficients for model construction (Fig. 2C).

Construction and validation of AI models

Models including ExtraTrees, XGBoost, LightGBM, and GradientBoosting were constructed using the selected features, with performance parameters listed in Table 2. The ExtraTrees, XGBoost, and LightGBM models demonstrated strong predictive performance in the training set, with AUCs of 0.816 (95% CI 0.748–0.884), 0.978 (95% CI 0.958–0.998), and 0.898 (95% CI: 0.846–0.950), respectively. Similarly, in the validation set, their AUCs were 0.759 (95% CI 0.641–0.876), 0.789 (95% CI 0.684–0.894), and 0.760 (95% CI 0.650–0.869), demonstrating relatively good predictive performance. The ROC curves are shown in Fig. 3A,B. Other performance parameters are detailed in Table 2. The DeLong test results indicated that the AUC values of the five machine learning models in the validation set were comparable, with p-values all greater than 0.05 (Supplementary Fig. 1A,B). The DCA curves for these 4 AI models in both the training and validation cohorts are presented in Fig. 4A,B. DCA curves indicates that the four machine learning models demonstrate greater net benefits across a majority of reasonable threshold probability ranges in both the training and validation sets. The calibration curve for the validation set (Fig. 4D) shows an overall consistency between the predicted probabilities and the observed outcomes, but it is slightly inferior to that of the training set (Fig. 4C).

Feature importance analysis for the models

Feature analysis identified the critical determinants of early recurrence. CA153 and AFP were identified as universally important predictors, ranking within the top five across the ExtraTrees, XGBoost, LightGBM, and GradientBoosting models. The specific importance scores are illustrated in Supplementary Fig. 2A–D, while the directional impact of these features is visualized in the SHAP summary plots in Supplementary Fig. 3A–D. Subsequently, univariate and multivariate regression analyses were conducted on the selected features, with odds ratios (OR) and p values presented in Table 3. Nonsmooth tumor margin and age were significantly associated with HCC, as evidenced by p values < 0.05 in both univariate and multivariate regression analyses.

Discussion

Recurrence within 2 years after HCC surgery is common and is a major cause of poor long-term survival outcomes[3,20]. In this study, we developed predictive models for early recurrence after HCC resection using machine learning algorithms including LightGBM, XGBoost, GradientBoosting, and ExtraTrees, and systematically evaluated their performance. Integrating clinical blood biomarkers, MRI, and pathological multi-modal data, our models demonstrated superior predictive performance and significant clinical utility.

Regarding the performance difference observed between the training and external validation cohorts, several baseline variables, such as washout, blood products in mass, size, Edmondson grade, TBiL, DBiL, ALB, INR, and PT, differed significantly, which may introduce potential bias; nevertheless, the model still demonstrated good performance, suggesting acceptable robustness and reflecting real-world clinical heterogeneity that supports its generalizability.

In this study, ExtraTrees, XGBoost, and LightGBM models achieved AUCs of 0.816, 0.978, and 0.898 in the training cohort, and 0.759, 0.789, and 0.760 in the external validation cohort, demonstrating robust prediction of early HCC recurrence. Zeng et al.'s Random Survival Forest reported AUCs of 0.818, 0.823, and 0.785 in training, internal, and external cohorts, respectively[21]. Wei et al. used preoperative MR features for solitary HCC without microvascular invasion, with AUCs of 0.788 and 0.783[22]. Kawashima et al. developed a clinical risk model, achieving an external AUC of 0.80 and C-index 0.71[23]. Overall, our multimodal models, integrating blood biomarkers, MRI, and pathological features, provide comparable or superior predictive performance, with multimodal data adding incremental prognostic value. The moderate decline from training to validation highlights the need for further optimization and prospective, multicenter validation.

MRI provides high-resolution imaging of soft tissues, offering detailed anatomical and functional information about HCC tumors. Gd-EOB-DTPA-enhanced MRI clearly delineates the contrast between tumors and normal

Characteristics	Training cohort ($n=161$)			Test cohort ($n=79$)		
	Nonearly recurrence	Early recurrence	P value	Nonearly recurrence	Early recurrence	P value
Size (cm)	4.09 ± 3.44	5.50 ± 4.33	**0.03**	4.69 ± 3.07	7.94 ± 5.73	**0.02**
Age (years)	51.59 ± 10.40	47.76 ± 10.75	**0.03**	51.00 ± 9.77	51.73 ± 11.33	0.76
NEUT (10^9/L)	3.67 ± 2.42	4.01 ± 1.88	0.09	3.78 ± 1.49	4.55 ± 2.64	0.24
LY (10^9/L)	1.76 ± 0.65	1.94 ± 0.74	0.20	1.88 ± 0.59	1.61 ± 0.55	0.04
NLR	2.71 ± 6.11	2.69 ± 3.23	0.84	2.28 ± 1.83	3.07 ± 1.98	**0.01**
PLT (10^9/L)	180.12 ± 75.75	206.68 ± 73.84	**0.03**	193.35 ± 71.71	210.57 ± 77.86	0.31
TBiL (μmol/L)	11.25 ± 5.00	14.72 ± 11.54	0.10	17.20 ± 6.73	17.94 ± 7.46	0.71
DBiL (μmol/L)	3.79 ± 1.71	6.02 ± 7.96	0.07	6.10 ± 2.98	7.07 ± 3.71	0.12
ALB (g/L)	39.65 ± 3.88	39.61 ± 5.20	0.27	37.45 ± 4.21	36.40 ± 4.26	0.28
PA (mg/L)	196.38 ± 57.92	190.15 ± 69.35	0.60	185.54 ± 46.05	168.27 ± 50.21	0.120
ALT (U/L)	36.57 ± 21.76	49.15 ± 39.40	**0.02**	46.67 ± 61.60	41.70 ± 33.92	0.940
AST (U/L)	35.21 ± 18.51	47.15 ± 39.42	0.13	42.04 ± 24.78	52.43 ± 31.12	**0.04**
GGT (U/L)	71.82 ± 65.12	101.02 ± 113.17	0.13	80.61 ± 138.90	120.57 ± 105.77	**< 0.001**
CA199 (U/mL)	120.36 ± 1047.12	32.03 ± 55.25	**0.01**	14.38 ± 12.85	15.84 ± 13.00	0.56
CA125 (U/mL)	15.59 ± 16.57	43.18 ± 155.26	0.07	15.36 ± 24.38	19.78 ± 21.17	**< 0.001**
CEA (ng/mL)	865.90 ± 6084.25	257.17 ± 1952.45	0.66	2.57 ± 2.56	4.25 ± 11.95	0.64
CA153 (U/mL)	10.42 ± 5.93	14.35 ± 8.30	**< 0.001**	10.53 ± 6.48	18.10 ± 11.67	**< 0.001**
INR	1.11 ± 1.28	1.01 ± 0.11	0.15	1.08 ± 0.10	1.11 ± 0.09	0.11
PT (seconds)	11.35 ± 1.73	11.91 ± 1.29	0.07	12.76 ± 1.22	13.16 ± 1.03	0.13
AFP (ng/mL)	1467.40 ± 6366.82	13312.81 ± 37654.88	0.09	1470.47 ± 7702.42	10204.87 ± 34268.62	**0.00**
Intraoperative blood loss (mL)	387.21 ± 679.42	532.54 ± 557.08	**0.02**	244.69 ± 265.19	359.33 ± 362.83	0.05
Sex, n (%)			0.77			0.91
Male	85(83.33)	51(86.44)		39(79.59)	25(83.33)	
Female	17(16.67)	8(13.56)		10(20.41)	5(16.67)	
Washout, n (%)			0.49			0.70
No	23(22.55)	17(28.81)		28(57.14)	15(50.00)	
Yes	79(77.45)	42(71.19)		21(42.86)	15(50.00)	
Capsule, n (%)			0.55			0.17
No	53(51.96)	27(45.76)		27(55.10)	11(36.67)	
Yes	49(48.04)	32(54.24)		22(44.90)	19(63.33)	
Intratumor fat, n (%)			0.87			1
No	93(91.18)	55(93.22)		47(95.92)	29(96.67)	
Yes	9(8.82)	4(6.78)		2(4.08)	1(3.33)	
APHE, n (%)			0.22			0.70
None rim APHE	18(17.65)	16(27.12)		13(26.53)	10(33.33)	
None & non-rim APHE	84(82.35)	43(72.88)		36(73.47)	20(66.67)	
Corona enhancement, n (%)			0.12			0.74
No	93(91.18)	48(81.36)		38(77.55)	25(83.33)	
Yes	9(8.82)	11(18.64)		11(22.45)	5(16.67)	
Intratumoral artery, n (%)			0.25			**0.02**
No	88(86.27)	46(77.97)		41(83.67)	17(56.67)	
Yes	14(13.73)	13(22.03)		8(16.33)	13(43.33)	
Nonsmooth tumor margin, n (%)			**0.02**			1
No	65(63.73)	26(44.07)		21(42.86)	13(43.33)	
Yes	37(36.27)	33(55.93)		28(57.14)	17(56.67)	
Nodule_in_nodule, n (%)			0.28			0.90
No	92(90.20)	49(83.05)		34(69.39)	22(73.33)	
Yes	10(9.80)	10(16.95)		15(30.61)	8(26.67)	
Mosaic_architecture			**0.01**			0.49
Absent	91(89.22)	43(72.88)		39(79.59)	21(70.00)	
Present	11(10.78)	16(27.12)		10(20.41)	9(30.00)	
HBP peritumoral_Hypo_intensity, n (%)			0.26			0.06
No	85(83.33)	44(74.58)		40(81.63)	18(60.00)	
Yes	17(16.67)	15(25.42)		9(18.37)	12(40.00)	
Intratumor hemorrhage, n (%)			0.28			0.07

Continued

Characteristics	Training cohort ($n=161$)			Test cohort ($n=79$)		
	Nonearly recurrence	Early recurrence	P value	Nonearly recurrence	Early recurrence	P value
No	92(90.20)	49(83.05)		41(83.67)	19(63.33)	
Yes	10(9.80)	10(16.95)		8(16.33)	11(36.67)	
Intratumor necrosis, n (%)			0.40			**<0.001**
No	80(78.43)	42(71.19)		38(77.55)	12(40.00)	
Yes	22(21.57)	17(28.81)		11(22.45)	18(60.00)	
BCLC, n (%)			0.820			**<0.001**
0&A	79(77.45)	44(74.58)		47(95.92)	20(66.67)	
B	23(22.55)	15(25.42)		2(4.08)	10(33.33)	
Edmondson, n (%)			**0.03**			**0.03**
Ⅰ&Ⅱ	80(78.43)	36(61.02)		25(51.02)	7(23.33)	
Ⅲ&Ⅳ	22(21.57)	23(38.98)		24(48.98)	23(76.67)	
MVI, n (%)			0.08			**0.01**
No	77(75.49)	36(61.02)		34(69.39)	11(36.67)	
Yes	25(24.51)	23(38.98)		15(30.61)	19(63.33)	
HBsAg, n (%)			0.83			0.81
No	13(12.75)	9(15.25)		9(18.37)	7(23.33)	
Yes	89(87.25)	50(84.75)		40(81.63)	23(76.67)	
Diabetes, n (%)			0.89			0.16
No	92(90.20)	52(88.14)		41(83.67)	29(96.67)	
Yes	10(9.80)	7(11.86)		8(16.33)	1(3.33)	
Hypertension, n (%)			0.10			0.43
No	80(78.43)	53(89.83)		45(91.84)	25(83.33)	
Yes	22(21.57)	6(10.17)		4(8.16)	5(16.67)	
Cirrhosis, n (%)			0.34			1
No	31(30.39)	13(22.03)		9(18.37)	5(16.67)	
Yes	71(69.61)	46(77.97)		40(81.63)	25(83.33)	
Ascites, n (%)			0.77			0.13
No	85(83.33)	51(86.44)		46(93.88)	24(80.00)	
Yes	17(16.67)	8(13.56)		3(6.12)	6(20.00)	
Tumor number, n (%)			0.160			0.43
1	88(86.27)	45(76.27)		45(91.84)	25(83.33)	
≥2	14(13.73)	14(23.73)		4(8.16)	5(16.67)	

Table 1. Baseline characteristics. NEUT, neutrophils; LY, lymphocytes; NLR, neutrophil-to-lymphocyte ratio; PLT, platelet; TBiL, total bilirubin; DBiL, direct bilirubin; ALB, Albumin-Bilirubin; PA, Prealbumin; ALT, alanine aminotransferase; AST, aspartate transaminase; GGT, γ-glutamyl transferase; CA199, Carbohydrate Antigen 199; CA125, Carbohydrate Antigen125; CEA, Carcinoembryonic Antigen; CA153, Carbohydrate Antigen153; INR, International Normalized Ratio; PT, prothrombin time; AFP, α-fetoprotein; APHE, arterial phase hyperenhancement; HBP, hepatobiliary phase; BCLC, Barcelona clinical liver cancer; MVI, microvascular invasion; HBsAg, Hepatitis B virus surface antigen.

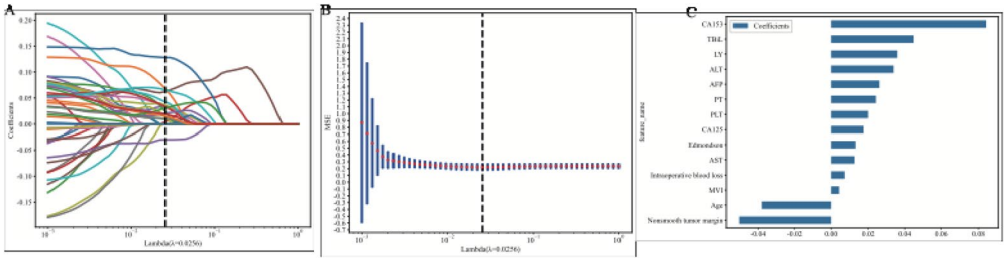

Fig. 2. Feature selection process. (**a**) LASSO coefficients corresponding to different λ values, with the vertical dashed line indicating the number of features at the optimal λ value. (**b**) The optimal λ value, selected based on tenfold cross-validation and minimum MSE, is represented by a vertical dashed line. (**c**) Non-zero coefficient features obtained after feature selection using the LASSO regression. MSE, mean squared error; LASSO, Least absolute shrinkage and selection operator.

Model	Cohort	Acc	AUC (95% Cl)	Sen	Spe	PPV	NPV	Precision	Recall	F1	Threshold
ExtraTrees	train	0.739	0.816 (0.748–0.884)	0.678	0.775	0.635	0.806	0.635	0.678	0.656	0.370
	test	0.759	0.759 (0.641–0.876)	0.667	0.816	0.690	0.80	0.690	0.667	0.678	0.450
XGBoost	train	0.938	0.978 (0.958–0.998)	0.915	0.951	0.915	0.951	0.915	0.915	0.915	0.382
	test	0.759	0.789 (0.684–0.894)	0.800	0.735	0.649	0.857	0.649	0.800	0.716	0.514
LightGBM	train	0.839	0.898 (0.746–0.950)	0.814	0.853	0.762	0.888	0.762	0.814	0.787	0.377
	test	0.722	0.760 (0.650–869)	0.567	0.816	0.654	0.755	0.654	0.567	0.607	0.499
GradientBoosting	train	0.888	0.941 (0.905–0.977)	0.763	0.961	0.918	0.875	0.918	0.763	0.833	0.391
	test	0.684	0.723 (0.609–0.837)	0.667	0.694	0.571	0.773	0.571	0.667	0.615	0.418

Table 2. Performance of AI models in predicting early recurrence after HCC resection in the training and testing cohorts. Acc, Accuracy; AUC, Area under the curve; Sen, Sensitivity; Spe, Specificity; PPV, Positive predictive value; NPV, Negative predictive value; F1, F1.

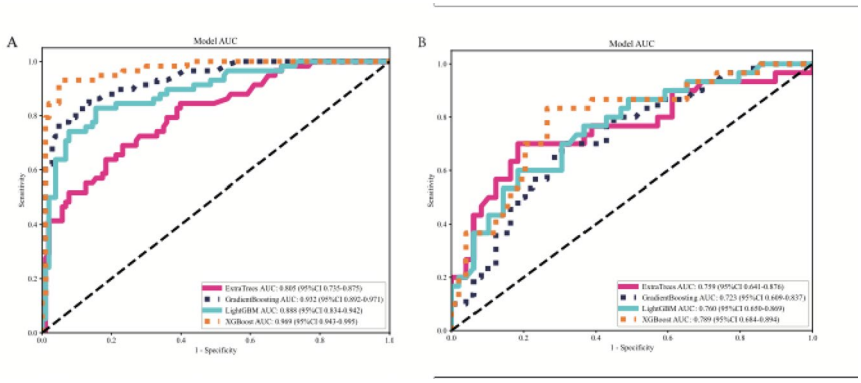

Fig. 3. Receiver operating characteristic curves evaluating ExtraTrees, XGBoost, LightGBM, and GradientBoosting models in the training (**A**) and Test (**B**) Cohorts.

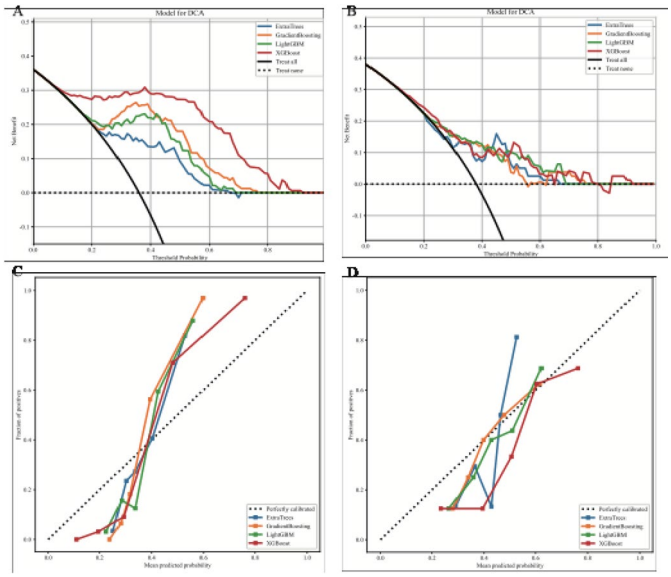

Fig. 4. Decision curve analysis and calibration curves comparing the performance of four machine learning models—ExtraTrees, XGBoost, LightGBM, and GradientBoosting—in predicting early recurrence. (**A**,**C**) Training cohort, (**B**,**D**) Validation cohort.

Variables	Univariate analysis		Multivariate analysis	
	OR (95%CI)	*P* value	OR (95%CI)	*P* value
Nonsmooth tumor margin	0.4 (0.273–0.586)	**<0.001**	0.344 (0.189–0.627)	**0.003**
LY (10^9/L)	0.806 (0.702–0.924)	**0.01**	1.318 (0.875–1.986)	0.267
MVI	0.92 (0.572–1.480)	0.773		
PT (seconds)	0.958 (0.937–0.981)	0.002	1.151 (1.018–1.301)	0.059
TBiL (μmol/L)	0.984 (0.967–1.002)	0.145		
CA153 (U/mL)	0.986 (0.968–1.005)	0.219		
Age (years)	0.988 (0.983–0.993)	**<0.001**	0.953 (0.931–0.976)	**0.001**
ALT (U/L)	0.996 (0.991–1.001)	0.203		
AST (U/L)	0.996 (0.990–1.001)	0.196		
PLT (10^9/L)	0.998 (0.997–0.999)	**0.018**	1.002 (0.998–1.005)	0.522
AFP (ng/mL)	1 (1–1)	0.096		
Intraoperative blood loss (mL)	1 (0.999–1)	0.419		
CA125 (U/mL)	1.002 (0.998–1.005)	0.494		
Edmondson	1.046 (0.64–1.707)	0.882		

Table 3. Univariate and multivariate logistic regression analyses of features for machine learning model construction in the training cohort. LY, lymphocytes; PLT, platelet; TBiL, total bilirubin; ALT, alanine aminotransferase; AST, aspartate transaminase; CA125, Carbohydrate Antigen125; CA153, Carbohydrate Antigen153; PT, prothrombin time; AFP, α-fetoprotein; MVI, microvascular invasion.

liver tissue, aiding in the detection of smaller or hidden lesions. This study found that an irregular tumor margin is an independent risk factor for early recurrence, consistent with previous studies[22,24,25]. Irregular tumor margins likely reflect an aggressive HCC phenotype. Pathologically, they are strongly associated with microvascular invasion, satellite nodules, and loss of capsule integrity, all of which promote intrahepatic spread[24,26]. Imaging–pathology studies further confirm that nonsmooth margins correspond to sites of extranodular extension with frequent MVI[26]. Moreover, such growth patterns have been linked to proliferative molecular subtypes and stem-cell marker expression (e.g., CK19), which are known to predict poor outcomes[27].

AFP has been reported to be associated with prognosis and overall survival[16,28,29]. In this study, AFP ranked among the top in importance across all machine learning models (Supplementary Fig. 2A–D, Supplementary Fig. 3A–D). Yen et al. reported that AFP ≥ 400 ng/mL significantly increased the risk of early recurrence within two years after liver resection (HR = 1.755)[30], and a nomogram integrating preoperative imaging and AFP developed by Giuffrè et al. also highlighted AFP as an important predictor[31]. These findings are consistent with our models, further supporting the clinical value of AFP in preoperative risk assessment. Due to ongoing debate regarding the optimal cutoff value for AFP[16], we analyzed it as a continuous variable instead. This approach yielded results consistent with previous studies[16,28,29]. Similarly, we applied the same method to other clinical laboratory variables included in the model, such as age, CA153, CA125, TBiL, ALT, PT, PLT, LY, AST, and intraoperative blood loss. Although CA15-3 is primarily used in breast cancer monitoring, it was also identified as important in our models. Studies in breast cancer have shown that elevated preoperative CA15-3 is associated with increased recurrence risk (HR = 6.87, 95% CI 5.81–8.11), and rising levels during follow-up may indicate distant metastasis[32]. As a carbohydrate antigen, CA15-3 elevation may reflect tumor-related systemic changes, liver dysfunction, or alterations in tumor cell glycosylation, suggesting its potential predictive value for early recurrence in HCC. Wang et al. found that a nomogram constructed using age, AFP, and PT can predict postoperative recurrence of HCC with considerable accuracy[16]. Liu et al. identified PLT as a positive prognostic factor for HCC early recurrence through principal component analysis and incorporated it into a preoperative KNN machine learning model to predict early recurrence[28]. Zheng et al. found that AST is an independent factor influencing early recurrence in combined hepatocellular-cholangiocarcinoma[33]. Studies have also shown that AST[34,35] and TBIL[35] are independent risk factors affecting postoperative prognosis in HCC. Qin et al. found that elevated baseline CA125 levels may be associated with poor prognosis in HBV-related HCC patients[36].

Our finding that younger age predicts early recurrence aligns with recent studies on HBV-related HCC[37,38]. This is likely due to more aggressive tumor biology in younger patients, characterized by larger tumors, higher AFP, and vascular invasion[37,38]. Additionally, surgical selection bias may contribute; younger patients often have better liver function (e.g., better ALBI grade), allowing resection of aggressive tumors that might be considered inoperable in older patients[38].

In this study, incorporating Edmondson and MVI into the model construction is consistent with previous research. Wang et al. demonstrated through multivariate logistic regression analysis that MVI is an independent predictor of early recurrence in HCC[39]. Zeng et al. developed an RSF machine learning model using 15 features, including Edmondson and MVI, to predict early recurrence, with MVI ranking in the top 3 for variable importance in external validation[21]. This aligns with the emerging trend of Explainable AI, emphasizing that interpretable models are essential to enhance clinical trust and adoption[40,41]. Consistent with most previous studies, we employed the lasso regression for feature selection. Compared with methods such as recursive feature elimination (RFE) or mutual information, lasso enables simultaneous variable selection and parameter estimation, yielding sparse and stable models that improve interpretability and reduce overfitting[42,43]. In addition,

lasso is computationally efficient and accounts for correlations among variables, making it well suited for clinical prediction modeling. It has also been widely applied in oncological prognostic research, demonstrating robustness and reliability[44–46].

In this study, the XGBoost model achieved an AUC of 0.978 in the training set and 0.789 in the validation set, indicating a notable performance drop. This may be due to the high complexity of XGBoost, which constructs multiple deep trees and can overfit the training data, the presence of numerous features or noise causing the model to capture patterns specific to the training set that do not generalize, and differences in data distribution between training and validation sets from different centers[47]. We assessed the model's generalizability using an independent validation set and plan to further improve model robustness in future studies via parameter tuning, regularization, or early stopping strategies.

This study has several limitations, as follows: First, the retrospective design introduces inherent selection bias and data constraints. Specific prognostic variables, such as hemodynamic portal hypertension and detailed surgical extent (major vs. minor), were not consistently recorded in historical data. Additionally, Performance Status was not analyzed as a standalone variable due to its low variance in this surgical cohort, although it is inherently captured by the BCLC staging system. Second, the recruitment from two centers in the same region may limit generalizability, and the sample size may pose a risk of overfitting for complex algorithms. To address these issues, we are currently conducting a large-scale, multicenter prospective study to rigorously validate our findings with comprehensive clinical data. Second, the recruitment from two centers in the same region may limit generalizability, and the sample size may pose a risk of overfitting for complex algorithms. To address these issues, we are currently conducting a large-scale, multicenter prospective study to rigorously validate our findings with comprehensive clinical data. Second, the study population was recruited from two centers in the same region, which may limit the diversity of patient characteristics and thereby reduce the generalizability of the findings. Although the present sample size was adequate for exploratory modeling, it may be underpowered for highly complex algorithms such as XGBoost, potentially increasing the risk of overfitting. Further validation in large-scale, multicenter prospective studies is still needed to confirm the robustness of the model. Finally, despite the adoption of rigorous feature selection methods, the inherent complexity of certain machine learning algorithms may still lead to model overfitting and reduced generalizability.

This study developed and validated AI models using clinical blood biomarkers, MRI features, and pathological data to predict early recurrence after liver resection for HCC. These models serve as effective tools for individualized risk stratification and guiding postoperative surveillance intensity, helping identify high-risk patients who may benefit from adjuvant therapies to improve prognosis. Future studies should focus on validating these models in larger, prospective, and multicenter cohorts to confirm their generalizability and clinical utility.

Data availability

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

Received: 25 October 2024; Accepted: 12 January 2026
Published online: 18 January 2026

References

1. Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Primers.* **7** (1), 6. https://doi.org/10.1038/s41572-020-00240-3 (2021).
2. Zhang, Z. H. et al. Role of microvascular invasion in early recurrence of hepatocellular carcinoma after liver resection: A literature review. *Asian J. Surg.* **47** (5), 2138–2143. https://doi.org/10.1016/j.asjsur.2024.02.115 (2024).
3. Shi, S., Zhao, Y. X., Fan, J. L., Chang, L. Y. & Yu, D. X. Development and external validation of a nomogram including body composition parameters for predicting early recurrence of hepatocellular carcinoma after hepatectomy. *Acad. Radiol.* **30** (12), 2940–2953. https://doi.org/10.1016/j.acra.2023.05.022 (2023).
4. Cheng, Z. et al. Risk factors and management for early and late intrahepatic recurrence of solitary hepatocellular carcinoma after curative resection. *HPB* **17** (5), 422–427. https://doi.org/10.1111/hpb.12367 (2015).
5. Nevola, R. et al. Predictors of early and late hepatocellular carcinoma recurrence. *World J. Gastroenterol.* **29** (8), 1243–1260. https://doi.org/10.3748/wjg.v29.i8.1243 (2023).
6. Zeng, J. et al. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection. *Hepatobiliary Surg. Nutr.* **11** (2), 176–187. https://doi.org/10.21037/hbsn-20-466 (2022).
7. Gao, W. et al. A predictive model integrating deep and radiomics features based on gadobenate dimeglumine-enhanced MRI for postoperative early recurrence of hepatocellular carcinoma. *Radiol. Med.* **127** (3), 259–271. https://doi.org/10.1007/s11547-021-01445-6 (2022).
8. Zhang, Z. et al. Hepatocellular carcinoma: radiomics nomogram on Gadoxetic acid-enhanced MR imaging for early postoperative recurrence prediction. *Cancer Imaging.* **19** (1), 22–32. https://doi.org/10.1186/s40644-019-0209-5 (2019).
9. Ji, G. W. et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: a multi-institutional study. *EBioMedicine* **50**, 156–165. https://doi.org/10.1016/j.ebiom.2019.10.057 (2019).
10. Yan, M. et al. Deep learning nomogram based on gd-EOB-DTPA MRI for predicting early recurrence in hepatocellular carcinoma after hepatectomy. *Eur. Radiol.* **33** (7), 4949–4961. https://doi.org/10.1007/s00330-023-09419-0 (2023).
11. Miranda, J. et al. Current status and future perspectives of radiomics in hepatocellular carcinoma. *World J. Gastroenterol.* **29** (1), 43–60. https://doi.org/10.3748/wjg.v29.i1.43 (2023).
12. Zheng, C. et al. Gut Microbiome as a biomarker for predicting early recurrence of HBV-related hepatocellular carcinoma. *Cancer Sci.* **114** (12), 4717–4731. https://doi.org/10.1111/cas.15983 (2023).
13. Zhou, S. L. et al. Genomic sequencing identifies WNK2 as a driver in hepatocellular carcinoma and a risk factor for early recurrence. *J. Hepatol.* **71** (6), 1152–1163. https://doi.org/10.1016/j.jhep.2019.07.014 (2019).
14. Fu, Y. et al. Combining a machine-learning derived 4-lncRNA signature with AFP and TNM stages in predicting early recurrence of hepatocellular carcinoma. *BMC Genom.* **24** (1), 89–104. https://doi.org/10.1186/s12864-023-09194-8 (2023).
15. Dong, B. et al. Development of a machine learning-based model to predict prognosis of alpha-fetoprotein-positive hepatocellular carcinoma. *J. Transl. Med.* **22** (1), 455–465. https://doi.org/10.1186/s12967-024-05203-w (2024).

16. Wang, X. et al. Clinical-radiological characteristic for predicting ultra-early recurrence after liver resection in solitary hepatocellular carcinoma patients. *J. Hepatocell Carcinoma*. **10**, 2323–2335. https://doi.org/10.2147/JHC.S434955 (2023).
17. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.* **69** (1), 182–236. https://doi.org/10.1016/j.jhep.2018.03.019 (2018).
18. Kang, I. et al. Subclassification of microscopic vascular invasion in hepatocellular carcinoma. *Ann. Surg.* **274** (6), e1170–e1178. https://doi.org/10.1097/SLA.0000000000003781 (2021).
19. Yan, Y. et al. Radiomic analysis based on Gd-EOB-DTPA enhanced MRI for the preoperative prediction of Ki-67 expression in hepatocellular carcinoma. *Acad. Radiol.* **31** (3), 859–869. https://doi.org/10.1016/j.acra.2023.07.019 (2024).
20. Shah, S. A. et al. Factors associated with early recurrence after resection for hepatocellular carcinoma and outcomes. *J. Am. Coll. Surg.* **202** (2), 275–283. https://doi.org/10.1016/j.jamcollsurg.2005.10.005 (2006).
21. Zeng, J. et al. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection. *HepatoBiliary Surg. Nutr.* **11** (2). https://doi.org/10.21037/hbsn-20-466 (2022).
22. Wei, Y., Pei, W., Qin, Y., Su, D. & Liao, H. Preoperative MR imaging for predicting early recurrence of solitary hepatocellular carcinoma without microvascular invasion. *Eur. J. Radiol.* **138**, 109663. https://doi.org/10.1016/j.ejrad.2021.109663 (2021).
23. Kawashima, J. et al. Development and validation of a preoperative risk prediction model for severe complications and very early recurrence after liver resection for hepatocellular carcinoma. *Surgery* **185**, 109527. https://doi.org/10.1016/j.surg.2025.109527 (2025).
24. Ariizumi, S. et al. A non-smooth tumor margin in the hepatobiliary phase of Gadoxetic acid disodium (Gd-EOB-DTPA) - enhanced magnetic resonance imaging predicts microscopic portal vein invasion, intrahepatic metastasis, and early recurrence after hepatectomy in patients with hepatocellular carcinoma. *J. Hepato Biliary Pancreat.* **18** (4), 575–585. https://doi.org/10.1007/s00534-010-0369-y (2011).
25. Zhang, L. et al. The role of preoperative dynamic contrast-enhanced 3.0-T MR imaging in predicting early recurrence in patients with early-Stage hepatocellular carcinomas after curative resection. *Front. Oncol.* **9**, 1336. https://doi.org/10.3389/fonc.2019.01336 (2019).
26. Chou, C. T. et al. Prediction of microvascular invasion of hepatocellular carcinoma: preoperative CT and histopathologic correlation. *Am. J. Roentgenol.* **203** (3), W253–W259. https://doi.org/10.2214/AJR.13.10595 (2014).
27. Chartampilas, E. et al. Current imaging diagnosis of hepatocellular carcinoma. *Cancers* **14** (16), 3997. https://doi.org/10.3390/cancers14163997 (2022).
28. Liu, C. et al. A K-nearest neighbor model to predict early recurrence of hepatocellular carcinoma after resection. *J. Clin. Transl. Hepatol.* **10** (4), 600–607. https://doi.org/10.14218/JCTH.2021.00348 (2022).
29. Wang, H. et al. A novel nomogram predicting the early recurrence of hepatocellular carcinoma patients after R0 resection. *Front. Oncol.* **13**, 1133807. https://doi.org/10.3389/fonc.2023.1133807 (2023).
30. Yen, Y. H. et al. A simple model to predict early recurrence of hepatocellular carcinoma after liver resection. *Langenbecks Arch. Surg.* **409** (1), 261. https://doi.org/10.1007/s00423-024-03449-y (2024).
31. Giuffrè, M. et al. Predictors of hepatocellular carcinoma early recurrence in patients treated with surgical resection or ablation treatment: a single-center experience. *Diagnostics* **12** (10), 2517. https://doi.org/10.3390/diagnostics12102517 (2022).
32. Ryu, J. M. et al. Prognostic impact of elevation of cancer antigen 15 – 3 (CA15-3) in patients with early breast cancer with normal serum CA15-3 level. *J. Breast Cancer*. **26** (2), 126–136. https://doi.org/10.4048/jbc.2023.26.e17 (2023).
33. Zheng, C. et al. Nomogram based on clinical and preoperative CT features for predicting the early recurrence of combined hepatocellular-cholangiocarcinoma: a multicenter study. *Radiol. med.* **128** (12), 1460–1471. https://doi.org/10.1007/s11547-023-01726-2 (2023).
34. Wang, J. et al. 3D MR elastography of hepatocellular carcinomas as a potential biomarker for predicting tumor recurrence. *Magn. Reson. Imaging*. **49** (3), 719–730. https://doi.org/10.1002/jmri.26250 (2019).
35. Zang, Y., Long, P., Wang, M., Huang, S. & Chen, C. Development and validation of prognostic nomograms in patients with hepatocellular carcinoma: a population-based study. *Future Oncol.* **17** (36), 5053–5066. https://doi.org/10.2217/fon-2020-1065 (2021).
36. Qin, C., Gao, Y., Li, J., Huang, C. & He, S. Predictive effects of preoperative serum CA125 and AFP levels on post-hepatectomy survival in patients with hepatitis B-related hepatocellular carcinoma. *Oncol. Lett.* **21** (6), 487–500. https://doi.org/10.3892/ol.2021.12748 (2021).
37. Zhang, Z. Y., Guan, J., Wang, X. P., Hao, D. S. & Zhou, Z. Q. Outcomes of adolescent and young patients with hepatocellular carcinoma after curative liver resection: a retrospective study. *World J. Surg. Oncol.* **20** (1), 210–217. https://doi.org/10.1186/s12957-022-02658-3 (2022).
38. Zeng, J. et al. Prognosis factors of young patients undergoing curative resection for hepatitis B virus-related hepatocellular carcinoma: a multicenter study. *Cancer Manag. Res.* **12**, 6597–6606. https://doi.org/10.2147/CMAR.S261368 (2020).
39. Chen, J. P. et al. Pre-operative enhanced magnetic resonance imaging combined with clinical features predict early recurrence of hepatocellular carcinoma after radical resection. *World J. Gastrointest. Oncol.* **16** (4), 1192–1203. https://doi.org/10.4251/wjgo.v16.i4.1192 (2024).
40. Romeo, M. et al. Clinical applications of artificial intelligence (AI) in human cancer: is it time to update the diagnostic and predictive models in managing hepatocellular carcinoma (HCC)? *Diagnostics* **15** (3), 252–274. https://doi.org/10.3390/diagnostics15030252 (2025).
41. Seven, İ. et al. Predicting hepatocellular carcinoma survival with artificial intelligence. *Sci. Rep.* **15** (1), 6226. https://doi.org/10.1038/s41598-025-90884-6 (2025).
42. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B*. **58** (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x (1996).
43. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** (1), 1–22 (2010).
44. Feng, L. et al. Predicting overall survival in hepatocellular carcinoma patients via a combined MRI radiomics and pathomics signature. *Transl. Oncol.* **51**, 102174. https://doi.org/10.1016/j.tranon.2024.102174 (2025).
45. Jiang, W. et al. Pathomics signature for prognosis and chemotherapy benefits in stage III colon cancer. *JAMA Surg.* **159** (5), 519–529. https://doi.org/10.1001/jamasurg.2023.8015 (2024).
46. Chen, D. et al. Prognostic and predictive value of a pathomics signature in gastric cancer. *Nat. Commun.* **13** (1), 6903. https://doi.org/10.1038/s41467-022-34703-w (2022).
47. Chen, T., Guestrin, C. & XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785 (2016).

Author contributions

L.F.: Conceptualization, formal analysis, visualization, writing original draft, investigation, software. N.L.: Conceptualization, data curation, formal analysis, investigation, validation, visualization, software, writing original draft. F.R.: Investigation, data curation, validation. X.Z.: Data curation, validation. X.P.: Data curation, validation. X.L.: Data curation, validation. L.F.: Data curation, validation. L.L.: Methodology, project administration,

writing review and editing, funding acquisition.

Funding
This study was supported by the National Natural Science Foundation of China, Grant/Award Number: 82060310.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-36261-3.

Correspondence and requests for materials should be addressed to L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.