SUPPLEMENTARY MATERIALS

Development of radiomics models

Feature extraction: A total of 1834 radiomic features were extracted from the contrast-enhanced arterial phase CT images using Pyradiomics (version 3.0.1), including tumor shape features, first-order features, texture features, and wavelet features. The texture features were derived from the Gray Level Cooccurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), Neighbouring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM) (Figure S1a).

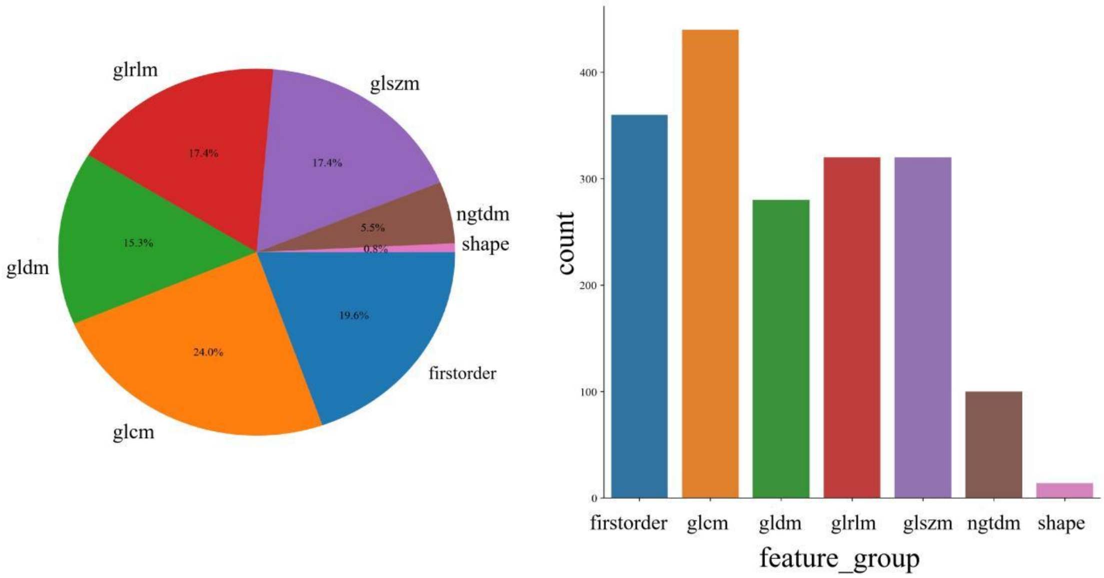
Feature Selection: To avoid overfitting and enhance the robustness of the model, we first randomly selected 30 samples to perform intraclass correlation coefficient (ICC) analysis of radiomics features extracted from the ROI placed by two doctors, removing features with ICC < 0.75.Secondly, after standardizing all features using Z-score normalization, we used t/u tests to identify features with statistical significance (p-value < 0.05).To reduce multicollinearity, we selectively excluded features with a Pearson correlation coefficient greater than 0.9.Finally, we improved the feature set using Lasso regression within a 10-fold cross-validation framework, optimized the regularization parameter λ, and selected the most relevant features for our model.

Model Development: We constructed radiomics models (including ExtraTrees, LightGBM, and MLP) using dimension-reduced robust features, incorporating the SMOTE method in the training phase, and ensuring the robustness of the model through 5-fold cross-validation and hyperparameter optimization via grid search.We further compared the performance of these models and selected the one with the best performance as the final Radiomics Signature.

Results: After conducting ICC analysis, a total of 1342 radiomic features were retained in the CT arterial phase. After performing t/Mann-Whitney U tests, Pearson correlation analysis, and Lasso regression, dimensionality reduction was conducted, and we selected 10 radiomic features to build a radiomic model(Figure S1b).The results showed that the MLP model had the best overall predictive performance, with accuracy, AUC, sensitivity, and specificity of 0.672, 0.727, 0.882, and 0.585, respectively, in the validation set (Table S1a and Figure S1b).

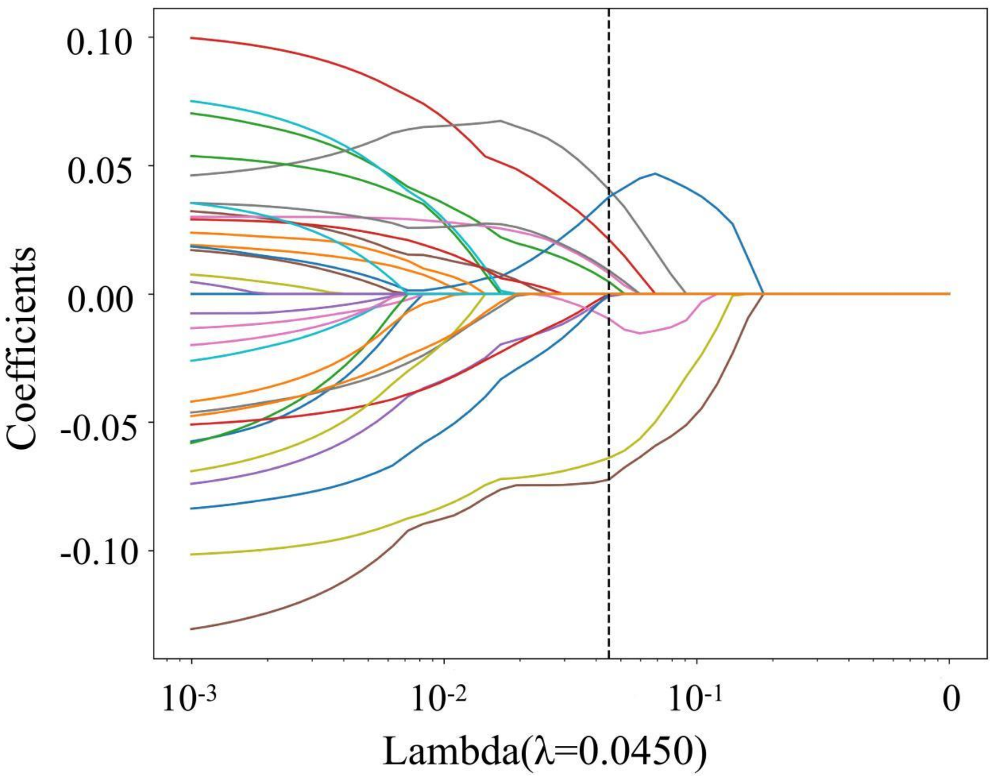
Baseline clinical feature analysis and results of the clinical model

The results of the baseline clinical characteristics analysis are presented in Table S2a. As shown in Table S2b and Figure S2a, the overall performance of the ExtraTrees, LightGBM, and MLP models constructed with Age and AFP is generally moderate, with MLP showing the best predictive performance (test accuracy = 0.500, AUC = 0.631, sensitivity = 0.882, specificity = 0.341).



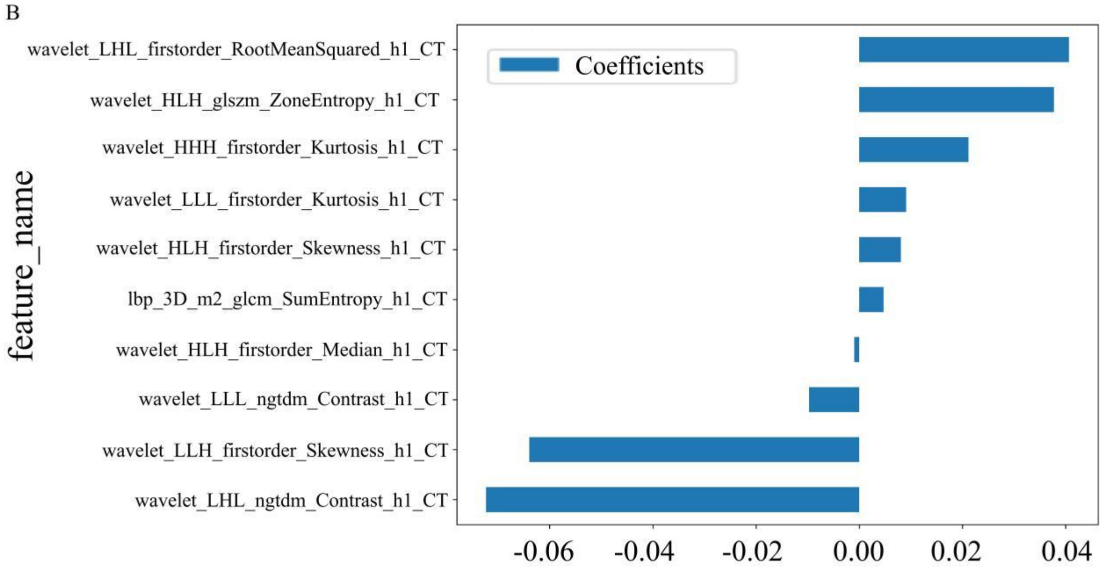
Supplementary Figure 1 Distribution of radiomic features.

A

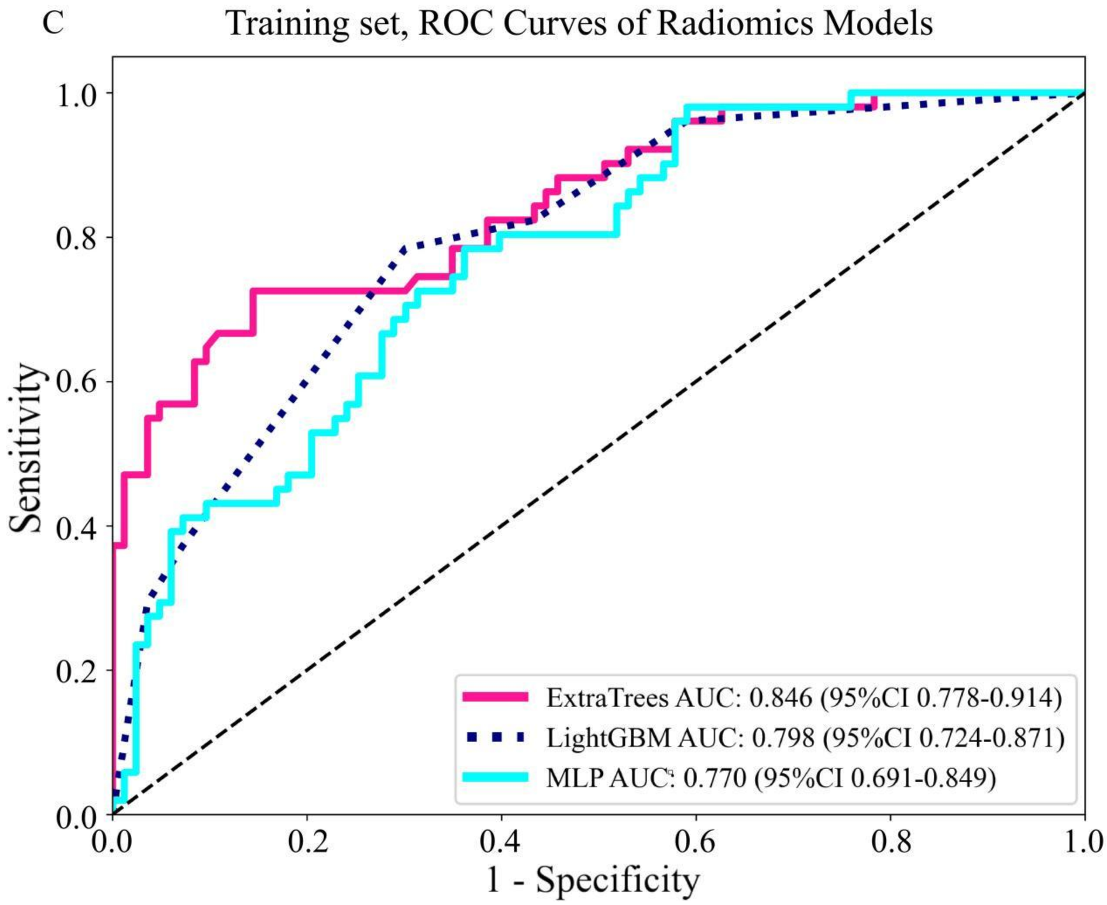


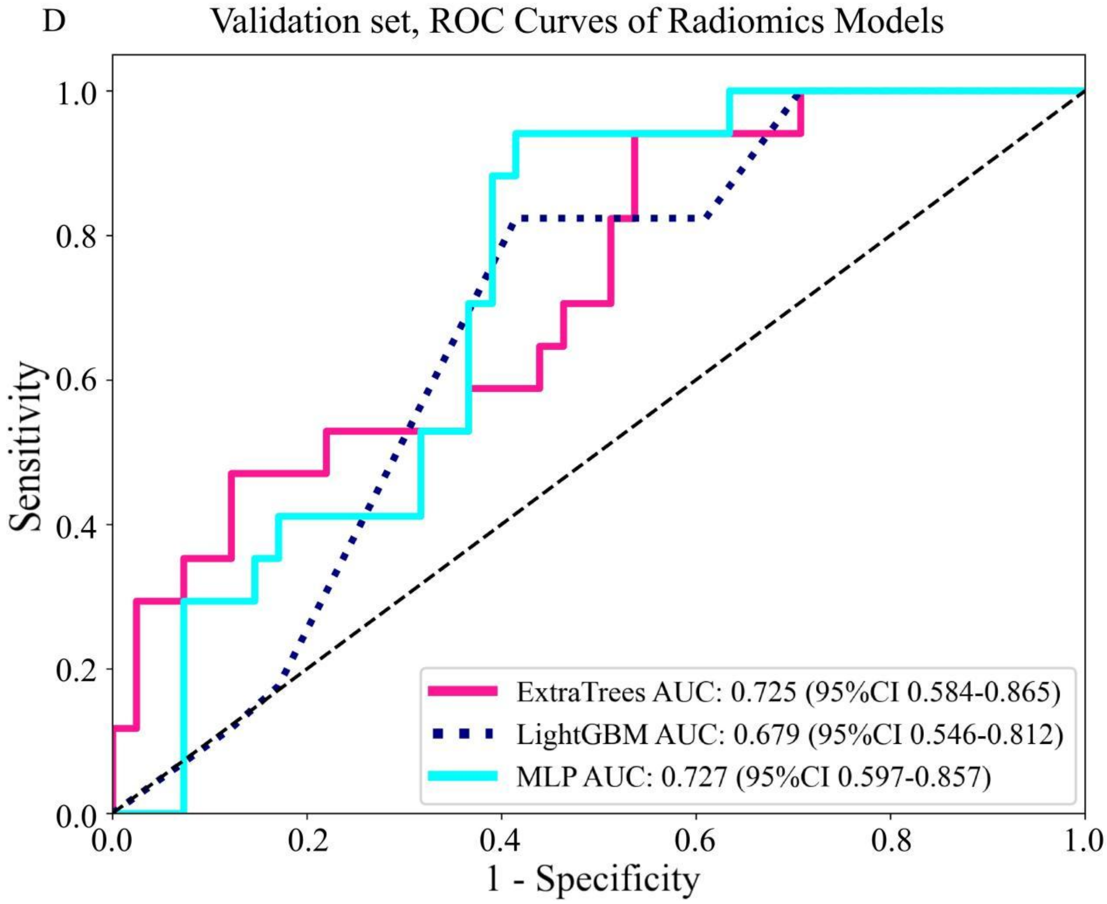
0.10
0.05
0.00
-0.05
-0.10
Coefficients
10-3
10-2
10-1
0
Lambda(λ=0.0450)

B

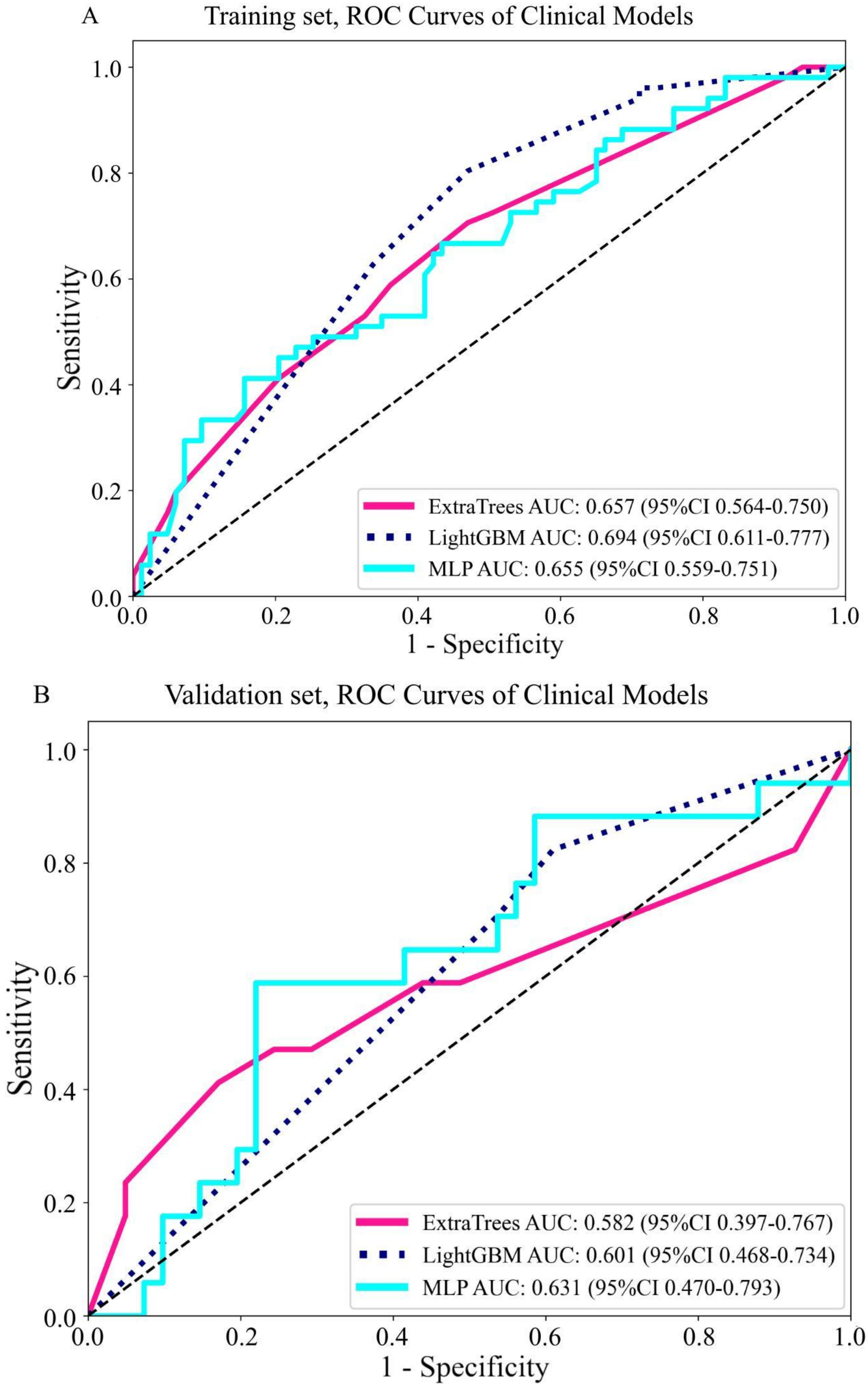


Coefficients
feature_name
wavelet_LHL_firstorder_RootMeanSquared_h1_CT
wavelet_HLH_glszm_ZoneEntropy_h1_CT
wavelet_HHH_firstorder_Kurtosis_h1_CT
wavelet_LLL_firstorder_Kurtosis_h1_CT
wavelet_HLH_firstorder_Skewness_h1_CT
lbp_3D_m2_glcm_SumEntropy_h1_CT
wavelet_HLH_firstorder_Median_h1_CT
wavelet_LLL_ngtdm_Contrast_h1_CT
wavelet_LLH_firstorder_Skewness_h1_CT
wavelet_LHL_ngtdm_Contrast_h1_CT
-0.06
-0.04
-0.02
0.00
0.02
0.04


C
Training set, ROC Curves of Radiomics Models
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
ExtraTrees AUC: 0.846 (95%CI 0.778-0.914)
LightGBM AUC: 0.798 (95%CI 0.724-0.871)
MLP AUC: 0.770 (95%CI 0.691-0.849)


D
Validation set, ROC Curves of Radiomics Models
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
ExtraTrees AUC: 0.725 (95%CI 0.584-0.865)
LightGBM AUC: 0.679 (95%CI 0.546-0.812)
MLP AUC: 0.727 (95%CI 0.597-0.857)

Supplementary Figure 2 Radiomic feature selection and modeling results. A: Visualization of Lasso regression dimensionality reduction, B: The respective weight distributions of the radiomic features, C: ROC curves of three radiomic models—ExtraTrees (red), LightGBM (cyan), and MLP (dark blue)—in the training set, D: ROC curves of three radiomic models—ExtraTrees (red), LightGBM (cyan), and MLP (dark blue)—in the validation set. ROC: Receiver operating characteristic; MLP：Multi-layer Perception.


A
Training set, ROC Curves of Clinical Models
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
ExtraTrees AUC: 0.657 (95%CI 0.564-0.750)
LightGBM AUC: 0.694 (95%CI 0.611-0.777)
MLP AUC: 0.655 (95%CI 0.559-0.751)
B
Validation set, ROC Curves of Clinical Models
ExtraTrees AUC: 0.582 (95%CI 0.397-0.767)
LightGBM AUC: 0.601 (95%CI 0.468-0.734)
MLP AUC: 0.631 (95%CI 0.470-0.793)

Supplementary Figure 3 ROC curves of different machine learning methods in the Clinical model. A: ROC curves of three radiomics models: ExtraTrees (red), LightGBM (cyan), and MLP (dark blue) in the training set. B: ROC curves of the same three radiomics models: ExtraTrees (red), LightGBM (cyan), and MLP (dark blue) in the validation set. ROC: Receiver operating characteristic; MLP：Multi-layer Perception.

Supplementary Table 1 Predictive metrics of different machine learning methods in the radiomics model

Dataset	Model	Accuracy	AUC(95% CI)	Sensitivity	Specificity
training	ExtraTrees	0.791	0.846(0.778 - 0.914)	0.686	0.855
validation	ExtraTrees	0.569	0.725(0.584 - 0.865)	0.824	0.463
training	LightGBM	0.716	0.798(0.724 - 0.871)	0.373	0.928
validation	LightGBM	0.638	0.679(0.546 - 0.812)	0.176	0.829
training	MLP	0.687	0.770(0.691 - 0.849)	0.765	0.639
validation	MLP	0.672	0.727(0.597 - 0.857)	0.882	0.585

MLP：Multi-layer Perception

Supplementary Table 2 Baseline clinical features analysis

Clinical feature	Training set(n=134)			Validation set(n=58)			Tace treatment cohort(n=45)
	MVI-Negative	MVI-Positive	P value	MVI-Negative	MVI-Positive	P value	

Age	53.63±9.92	49.45±8.71	0.015	53.54±12.17	50.41±11.82	0.373	56.96±10.19
BMI	35.74±5.20	36.39±5.46	0.492	35.81±4.79	37.05±6.01	0.412	35.99±5.36
HBsAg	1074.63±944.73	1009.77±756.37	0.989	962.55±773.81	1090.64±843.00	0.765	933.52±1048.39
AFP	380.89±489.91	606.46±538.15	0.005	343.06±473.23	702.29±525.86	0.029	597.81±560.66
ALB	39.37±5.00	40.97±6.64	0.331	39.00±4.73	38.31±5.11	0.626	36.55±5.36
AST	65.08±72.61	57.61±53.91	0.812	65.63±69.56	122.28±240.69	0.074	133.39±396.86
ALT	63.95±80.87	58.94±69.90	0.922	68.03±91.37	86.66±170.69	0.966	75.83±182.29
sex			0.383			0.28	
female	12(14.46%)	4(7.84%)		12(29.27%)	2(11.76%)		5(11.11%)
male	71(85.54%)	47(92.16%)		29(70.73%)	15(88.24%)		40(88.89%)
Dringking history			0.099			0.045	

no	60(72.29%)	29(56.86%)	30(73.17%)	7(41.18%)	21(46.67%)
yes	23(27.71%)	22(43.14%)	11(26.83%)	10(58.82%)	24(53.33%)

BMI: body mass index, HBsAg: hepatitis B surface antigen, AFP: alpha-fetoprotein, ALB: albumin, AST: aspartate aminotransferase , ALT: alanine aminotransferase.

Supplementary Table 3 Predictive metrics of different machine learning methods in the clinical model

Dataset	**Dodel**	**Accuracy**	**AUC(95% CI)**	**Sensitivity**	**Specificity**
training	ExtraTrees	0.619	0.657(0.564 - 0.750)	0.588	0.639
validation	ExtraTrees	0.638	0.582(0.397 - 0.767)	0.471	0.707
training	LightGBM	0.634	0.694(0.611 - 0.777)	0.804	0.530
validation	LightGBM	0.517	0.601(0.468 - 0.734)	0.824	0.390
training	MLP	0.530	0.655(0.559 - 0.751)	0.882	0.313
validation	MLP	0.500	0.631(0.470 - 0.793)	0.882	0.341