Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multimodal model for diagnosis of tumor-associated macrophages with M2 polarization and prognosis of targeted therapy in hepatocellular carcinoma: A multicenter retrospective study

Jiaojiao Xu
Shanxi Provincial People's Hospital
Wei Liu
Shanxi Provincial People's Hospital
Kunming Pu
Chengdu Second People's Hospital affiliated Sichuan University
Qian Feng
General Hospital of Western Theater Command
Yuhong Xue
Shanxi Provincial People's Hospital
Liang Mu
Shanxi Provincial People's Hospital
Xin Li
lixin301@hotmail.com

Chinese PLA General Hospital

Research Article

Keywords: Hepatocellular carcinoma, Artificial intelligence, Prognosis, Tumor microenvironment

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10256007/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background:

Tumor-associated macrophages with M2 polarization (TAMs-M2) was a key immunohistochemical phenomenon in hepatocellular carcinoma (HCC), which associated with therapy resistance and poor prognosis. However, diagnosis of TAMs-M2 highly relies on invasive biopsy, which reduced its clinical value.

Objective:

To develop a multimodal model combining ultrasound (US), magnetic resonance imaging (MRI) and clinical information for TAMs-M2 diagnosis non-invasively and evaluate model value in prognosis.

Methods:

US image, MRI image, clinical information, immunohistochemical result on TAMs-M2, and prognostic information were retrospectively collected from six centers. Multimodal model based on US deep learning feature, MRI radiomics feature and clinical information was developed to non-invasively predict the TAMs-M2 in HCC. Patients with targeted therapy were classified as high-risk and low-risk groups by multimodal model, and the treatment response and overall survival (OS) were compared in two groups.

Results:

Since January 2018 and December 2024, 552 eligible HCCs were enrolled and divided into training (n=222), validation (n=111), internal test (n=111), and external test (n=108) sets. Multimodal model demonstrated well diagnostic performance in internal (AUC=0.802, 95%CI: 0.689-0.901) and external (AUC=0.797, 95%CI: =0.689-0.888) test sets. Since January 2020 and December 2024, 225 eligible HCCs were collected from other three centers as prognosis test set, and HCCs in low-risk group showed significantly higher objective response rate (11.2% vs 5.5%, p=0.032), disease control rate (57.3% vs 23.3%, p<0.001), and longer OS (p=0.007) than high-risk group.

Conclusion:

Multimodal model can accurately and non-invasively predict TAMs-M2 in HCC and effectively stratify the prognosis of patient received targeted therapy, holding promise for guiding personalized treatment strategies.

Key points

1. Multimodal model integrated ResNet-50 ultrasound features, MRI radiomics features, and clinical variables delivered an AUC of 0.802 (95% CI: 0.689–0.901) in internal test set and 0.797 (0.689–0.888) in external test set, outperforming single-modality US model or MR model.

2. Patients classified into Low TAMs-M2 versus High TAMs-M2 groups exhibited markedly divergent outcomes: objective response rate 11.2% vs 5.5%, disease-control rate 57.3% vs 23.3%, and significantly longer overall survival (log-rank p = 0.007).

3. This study suggests that conventional imaging, when combined with artificial intelligence, may serve as a potential tool for risk stratification in HCC patients receiving targeted therapy.

Introduction

Hepatocellular carcinoma (HCC) is the six most frequent primary malignant tumor and the third most common cause of all cancer-related mortalities(1). The recurrence rate of HCC is quite high, but most patients with tumor recurrence lost the surgery opportunity, due to the physical conditions and the tumor state. Targeted therapy brings new hope to these patients and plays a crucial role in improving their life quality and prolonging survival. However, the effectiveness of targeted therapy varies greatly among individuals, with only approximately 30% of patients can benefit from it. Therefore, accurately identifying the benefiting population is of great significance for targeted therapy.

Previous studies have proved that tumor microenvironment play multifunctional roles in cancer progression and metastasis, and have the potential to influence the effectiveness of targeted therapy(2). Tumor-associated macrophages (TAMs) are the essential component of tumor microenvironment and represent one of the main tumor-infiltrating immune cell types(3, 4). TAMs could be categorized into two subtypes, the classical activated M1 macrophages and the alternatively activated M2 macrophages. M1 macrophages exert anti-tumor functions, including directly mediate cytotoxicity and antibody-dependent cell-mediated cytotoxicity to kill tumor cells(4), while M2 macrophages can promote the occurrence and metastasis of tumor cells, inhibit T cell-mediated anti-tumor immune response, promote tumor angiogenesis, and lead to tumor progression. M2 macrophages are usually associated with targeted therapy resistance and poor prognosis (5), so identifying tumor-associated macrophages-2 polarized (TAMs-M2) is the crucial step in analyzing the effectiveness of targeted therapy (6). However, the current identification method of TAMs-M2 relies on biopsy, so the complication risk and the inadequate sampling inevitably limit the application value of TAMs-M2 for HCC management (7). Therefore, it is clinically necessary to develop a non-invasively method for TAMs-M2.

Ultrasound (US) and magnetic resonance imaging (MRI) are the main image examinations in HCC diagnosis (8, 9), with the ability to reflect tumor texture, blood perfusion and vascular distribution, which can indirectly reflect the state of the immune microenvironment and provide assistances for the diagnosis of TAMs-M2 (10, 11). However, the complex microscopic information reflected by these imaging modalities is difficult for doctors to recognize and utilize.

Artificial intelligence (AI) can provide significant assistance in analyzing the image features that beyond the recognition capabilities of radiologists (12, 13). Deep learning has demonstrated superior performance in analyzing US images, achieving area-under-the-curve (AUC) values exceeding 0.90 in distinguishing HCC (14). Radiomics based on MRI can effectively stratify the prognosis of HCC patients

(15, 16). Therefore, we believe that the AI technology had the ability to extract high-dimensional features from images (17), and identifying the imaging biomarkers of TAMs-M2.

The aim of this study is to develop a multimodal model based on clinical data, US, and MRI to predict TAMs-M2 and stratify the prognosis of patients received targeted therapy.

Methods

Patients

This retrospective, multi-center study was conducted with the approval of the Institutional Review Boards of all participating centers [S2025-23-71]. In model building phase, the inclusion criteria were: (1) Pathologically confirmed HCC; (2) With TAMs-M2 status assessed by immunohistochemistry; (3) Received US and MRI within 2 weeks before pathology; (4) Tumor size more than 1cm. The exclusion criteria were: (1) Age less than 18 years; (2) With poor image quality; (3) Incomplete clinical information. In prognosis assessment phase, the inclusion criteria were: (1) Patients with HCC history and clinically diagnosed HCC (based on imaging and serum α-fetoprotein levels) who received targeted therapy; (2) Received US and MRI within 2 weeks before targeted therapy; (3) Tumor size more than 1cm. The exclusion criteria were: (1) Age less than 18 years; (2) With poor image quality; (3) Incomplete clinical information; (4) Poor image quality was defined as the presence of severe motion artifacts, significant signal dropout, or insufficient anatomic coverage.

Data from Centers A and B were used as the training, validation, and internal test set to establish the model. Data from Center C were constituted as the independent external test to evaluate the model diagnostic performance on TAMs-M2. In additional, a prognostic test set, including patients with targeted therapy, was collected from Centers D, E and F to assess the model performance on prognostic outcomes.

Data Collection

US and MRI examinations were performed by mid-level or senior radiologists, with more than eight-year experience in liver imaging. For US image, regions of interest (ROI)s were annotated through the LabelMe software (version 4.5.13) to facilitate deep learning. The annotation of US images was performed using tightly fitted bounding boxes that closely adhered to the tumor margins, which were determined through consultation between two radiologists with more than five years of experience. For MRI image, ROIs were annotated through the 3Dslicer (version 5.7.0) to facilitate radiomics analysis. The annotation of MRI images was performed using contour delineation, which was conducted independently by two radiologists with more than five years of experience, followed by a consistency assessment. The radiologists involved in the annotation were blinded to the tumors TAMs-M2.

Collected clinical data included demographic information (age, sex and etiology), tumor characteristics (primary or tumor size), Child-Pugh, BCLC and serological indicators, including: alanine aminotransferase (ALT), aspartate aminotransferase (AST), total protein (TP), serum albumin (ALB), total bilirubin (TB), direct bilirubin (DB) and alpha fetoprotein (AFP).

TAMs-M2 status was determined by immunohistochemistry (IHC) staining. HCC tissue specimens obtained from surgical resection were stained with antibodies against CD163. The slides were independently interpreted by two pathologists using the ≥ 10% threshold (18). In case of disagreement, a consensus was reached through consultation. The IHC assay was completed within two weeks after sample collection to ensure antigen stability.

Prognostic data were collected in prognostic test set. Targeted therapies included sorafenib and lenvatinib. Sorafenib was administered at a standard dose of 400 mg twice daily, and lenvatinib was dosed according to body weight (≥ 60 kg: 12 mg once daily; <60 kg: 8 mg once daily). No patient received immune checkpoint inhibitor combination therapy in this cohort. Treatment response was evaluated based on the Response Evaluation Criteria in Solid Tumors (RECIST), categorizing outcomes as complete remission (CR), partial remission (PR), stable disease (SD), or progressive disease (PD). The objective response rate (ORR) was calculated as the proportion of patients achieving CR or PR, while the disease control rate (DCR) included patients with CR, PR, or SD. Overall survival (OS) was defined as the time from treatment initiation to death from any cause. Patients were followed at 3- to 6-month intervals with combined imaging and clinical assessments to monitor for disease process. The assessments were independently performed by two radiologists (each with > 5 years of experience) in blinded manner. In cases of disagreement, a third radiologist (> 15 years of experience) would be invited as the adjudicator.

Model Construction

For US images, we utilized and compared multiple deep learning algorithms to extract image features and construct the US-based model. The deep learning algorithms included VGG-19, ResNet-50, and DenseNet-121, implemented with TensorFlow (v2.9). Training details were provided in the supplementary materials, and visualization of model was achieved through the class activation mapping.

For MRI images, we first performed image normalization. Then, we utilized pyradiomics (version 2.27) to extract image features from T1 MRI sequences. Subsequently, consistency assessment was performed, features with intraclass correlation coefficient (ICC) > 0.8 were retained. Then, these features were selected through LASSO (Least Absolute Shrinkage and Selection Operator). Multiple machine learning algorithms were then applied to build the MRI-based model. The machine algorithms included Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB). Visualization of machine learning models through SHAP.

The prediction results of US and MRI models were continuous scores, representing the probability of TAMs-M2 positivity derived from the respective imaging data. For clinical data, feature screening was performed using univariate and multivariate analysis. US score and MRI score were integrated with the selected clinical features through logistic regression to form the multimodal TAMs-M2 model.

Model Evaluation

The diagnostic performance of multimodal model on TAMs-M2 was assessed in the training, validation, internal and external test sets. The evaluation methods included the area under the receiver operating characteristic (ROC) curve, confusion matrices, calibration curves, and decision curve analysis (DCA). The evaluation indexes included AUC, accuracy, specificity, sensitivity and so on.

The prognostic impact of multimodal model was assessed in the prognostic test set. Patients were stratified into high-risk group (TAMs-M2 positive predicted by the model) and low-risk group (TAMs-M2 negative predicted by the model). The ORR, DCR and OS were compared between two groups.

Statistical analysis

Continuous variables were expressed as mean ± standard deviation and compared using either the independent samples t-test or analysis of variance, while categorical variables were analyzed using the chi-square test or Fisher's exact test as appropriate. The diagnostic performance of the model was comprehensively evaluated using multiple metrics, including the area under the receiver operating characteristic curve (AUC), accuracy, specificity, sensitivity, precision, recall, F1 score, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), Cohen’s kappa coefficient, Matthews correlation coefficient (MCC), diagnostic odds ratio (DOR), Youden’s index, Brier score, and balanced accuracy. Prognostic outcomes were assessed through Kaplan-Meier analysis, with between-group comparisons performed using the log-rank test. All statistical estimates were reported with 95% confidence intervals (CI) derived from 1000 bootstrap samples. The entire statistical analysis was performed using R statistical software (version 4.0.0).

Results

Patients

Since January 2018 to December 2024, 444 cases from Centers A and B were enrolled and were randomly allocated into training set (n = 222), validation set (n = 111), and internal test set (n = 111). The baseline demographic and clinical characteristics were comparable across the training, validation, and internal test sets (Table 1). Since January 2018 to December 2024, an additional 108 cases from Center C were collected as an independent external test set to assess the model diagnostic performance on TAMs-M2. Since January 2020 to December 2024, 225 patients received targeted therapy from Center D,

E and F were collected as the prognostic test set, and the median follow-up time was 14.2 months (Fig. 1).

Table 1
Baseline of training, validation and internal test sets.

n		Training set	Validation set	Internal test set	External test set	p
		222	**111**	**111**	**108**	
Sex (%)	Man	153 (68.9%)	82 (73.9%)	79 (71.2%)	79 (73.1%)	0.765
	Woman	69 (31.1%)	29 (26.1%)	32 (28.8%)	29 (26.9%)	
Age (year)		60.89 ± 12.27	61.44 ± 11.14	59.35 ± 10.76	59.58 ± 11.41	0.436
Etiology (%)	Viral hepatitis	141 (63.5%)	79 (71.2%)	82 (73.9%)	32 (29.6%)	0.205
	Other	81 (36.5%)	32 (28.8%)	29 (26.1%)	76 (70.4%)	
Child-Pugh	A	198 (89.2%)	100 (90.1%)	101 (91.0%)	100 (92.6%)	0.794
	B	24 (10.8%)	11 (9.9%)	10 (9.0%)	8 (7.4%)	
BCLC	0-A	203 (91.4%)	104 (93.7%)	103 (92.8%)	95 (88.0%)	0.777
	B	16 (7.2%)	6 (5.4%)	7 (6.3%)	12 (11.1%)	
	C	3 (1.4%)	1 (0.9%)	1 (0.9%)	1 (0.9%)	
Tumor type (%)	Primary	146 (65.8%)	68 (61.3%)	76 (68.5%)	76 (70.4%)	0.508
	Recurrence	76 (34.2%)	43 (38.7%)	35 (31.5%)	32 (29.6%)	
Tumor size (cm)		3.76 ± 1.65	3.97 ± 1.67	3.86 ± 1.61	3.77 ± 1.58	0.701
ALT (U/L)		35.14 ± 16.33	33.69 ± 15.77	35.87 ± 15.89	38.71 ± 15.93	0.122
AST (U/L)		35.28 ± 15.80	33.88 ± 15.27	35.99 ± 15.59	37.21 ± 15.05	0.446

Note:

ALT=Alanine aminotransferase; AST=Aspartate aminotransferase; TP=Total protein; ALB=Serum albumin; TB=Total bilirubin; DB=Direct bilirubin; AFP=Alpha fetoprotein; TAMs-M2 = Tumor-Associated Macrophages - M2 phenotype

n		Training set	Validation set	Internal test set	External test set	p
		222	111	111	108	
TP (g/L)		67.46 ± 5.89	66.66 ± 6.42	66.41 ± 6.91	67.15 ± 6.54	0.48
ALB (g/L)		39.32 ± 4.57	39.17 ± 4.72	39.40 ± 5.42	39.04 ± 4.54	0.939
TB (μmol/L)		16.45 ± 9.87	15.45 ± 7.32	16.44 ± 9.50	16.80 ± 10.63	0.735
DB (μmol/L)		6.30 ± 4.86	5.94 ± 2.50	6.03 ± 4.84	6.46 ± 4.91	0.795
AFP (%)	< 20 U/L	101 (45.5%)	49 (44.1%)	43 (38.7%)	56 (51.9%)	0.277
	> 20 U/L	121 (54.5%)	62 (55.9%)	68 (61.3%)	52 (48.1%)	
TAMs-M2 (%)	Low	163 (73.4%)	84 (75.7%)	84 (75.7%)	80 (74.1%)	0.959
	High	59 (26.6%)	27 (24.3%)	27 (24.3%)	28 (25.9%)	
Note:						
ALT=Alanine aminotransferase; AST=Aspartate aminotransferase; TP=Total protein; ALB=Serum albumin; TB=Total bilirubin; DB=Direct bilirubin; AFP=Alpha fetoprotein; TAMs-M2 = Tumor-Associated Macrophages - M2 phenotype						

Model Development

To assess the inter-observer reproducibility of the US bounding-box annotations, we randomly selected 20% of the US images from the training set for re-annotation by two independent radiologists who were blinded to the previous annotations. The Dice similarity coefficient (DSC) was calculated to quantify the spatial overlap between the original and re-annotated bounding boxes. The mean DSC across the evaluated samples was 0.91, indicating excellent agreement and ensuring the robustness of the input data for subsequent deep-learning feature extraction.

In the process of building US-based model, we trained models through three deep learning algorithms (VGG-19, ResNet-50, and DenseNet-121), and compared model performance in training set and validation set. In the three models, ResNet-50 model demonstrated the highest performance, achieving an AUC of 0.831 (95% CI: 0.778–0.882) in training set and 0.780 (95% CI: 0.670–0.877) in validation set (Fig. 2, A and B). Therefore, we chose the ResNet-50 algorithm as the US-based model. Visualization of the US-based model was showed in Figure S1.

In the process of building MRI-based model, we extracted a total of 1310 features from the MRI image through pyradiomics. Among them, 249 features had an ICC of less than 0.80 and were excluded, and 1061 features were selected for feature engineering. Then, 12 features closely related to TAMs-M2 were chosen through LASSO and thus were used to establish the MRI-based model (Figure S2, features in supplementary materials). We trained three model through three machine learning algorithms (RF, SVM, and XGB), and compared model performance in training set and validation set. In the three models, XGB model demonstrated the highest performance, achieving an AUC of 0.864 (95% CI: 0.805–0.915) in training set and 0.794 (95% CI: 0.701–0.875) in validation set (Fig. 2, C and D). Therefore, we chose the XGB algorithm as the MRI-based model. Visualization of the MRI-based model was showed in Figure S3.

Concurrently, univariate logistic regression identified viral hepatitis etiology (OR: 0.329, 95% CI: 0.176–0.605), recurrence tumor type (OR: 0.329, 95% CI: 0.205–0.701), and larger tumor size (OR: 0.807, 95% CI: 0.667–0.970) as factors significantly associated with high TAMs-M2 status. In the subsequent multivariate analysis, all three variables retained independence: viral hepatitis etiology (OR: 0.311, 95% CI: 0.161–0.587), recurrence tumor type (OR: 0.349, 95% CI: 0.181–0.666), and larger tumor size (OR: 0.783, 95% CI: 0.639–0.952) (Table 2).

Table 2
Univariate/Multivariate analysis on clinical information

	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	**p**	**HR**	**95% CI**	**p**
Sex	0.611	0.300–1.185	0.664			
Age	1.001	0.976–1.025	0.751			
Etiology	0.329	0.176–0.605	0.003	0.311	0.161–0.587	0.002
Child-Pugh	0.702	0.224–1.846	0.502			
BCLC	0.637	0.142–2.068	0.494			
Tumor type	0.329	0.205–0.701	0.012	0.349	0.181–0.666	0.017
Tumor size	0.807	0.667–0.970	0.026	0.783	0.639–0.952	0.031
ALT	1.004	0.986–1.023	0.856			
AST	1.006	0.987–1.025	0.762			
TP	1.003	0.954–1.056	0.463			
ALB	0.993	0.931–1.061	0.469			
TB	0.981	0.944–1.013	0.798			
DB	1.014	0.952–1.074	0.369			
AFP	0.619	0.338–1.126	0.227			
Note:						
ALT=Alanine aminotransferase; AST=Aspartate aminotransferase; TP=Total protein; ALB=Serum albumin; TB=Total bilirubin; DB=Direct bilirubin; AFP=Alpha fetoprotein; TAMs-M2 = Tumor-Associated Macrophages - M2 phenotype; CI = Confidence Interval.						

Based on these results, the multimodal model integrated the US-based model, MRI-based model and clinical factors (etiology, tumor type, and tumor size) (Fig. 2, E). In internal test set, the multimodal model (AUC: 0.802 (95% CI: 0.689–0.901)) achieved slightly better performance than the US-based model (AUC: 0.740, 95% CI: 0.630–0.827, p = 0.076) and the MRI-based model (AUC: 0. 744, 95% CI: 0.661–0.835, p = 0.088) (Fig. 2, F). In internal test set, multimodal model also achieved satisfactory performance accuracy of 0.829 (95% CI: 0.757–0.901), sensitivity of 0.815, and specificity of 0.833 (Table S1).

Model performance on TAMs-M2 status

In external test set, multimodal model achieved AUC of 0.797 (95% CI: 0.689–0.888), accuracy of 0.806 (95% CI: 0.731–0.880), sensitivity of 0.750 (95% CI: 0.600-0.909), and specificity of 0.825 (95% CI: 0.733–0.909) (Fig. 3, Table S2). The confusion matrix detailed the classification performance, with 66 true negatives and 21 true positives, against 14 false positives and 7 false negatives. The calibration curve indicated the agreement between the model's predicted probabilities and the observed outcomes. Furthermore, decision curve analysis confirmed the model's clinical utility across a wide range of threshold probabilities (Fig. 3A).

Subgroup analyses were performed to evaluate the model's robustness. The model demonstrated consistently high AUC and accuracy values in males (AUC: 0.787, Accuracy: 0.792), females (AUC: 0.812, Accuracy: 0.824), tumor size ≤ 3 cm (AUC: 0.788, Accuracy: 0.795), tumor size > 3 cm (AUC: 0.809, Accuracy: 0.820), patients over 60 years (AUC: 0.821, Accuracy: 0.818), and patients less 60 years (AUC: 0.781, Accuracy: 0.794) (Fig. 3B and C).

Model performance on prognosis

225 patients from three centers were included as the prognostic test set (Table 3), and were stratified into low-risk and high-risk groups based on the multimodal model prediction. Stratification by multimodal model revealed significant disparities in treatment response and survival outcomes. The low-risk group exhibited a markedly higher ORR of 11.2%, compared to 5.5% in high-risk group ($p = 0.032$). Correspondingly, the DCR was 57.3% in low-risk group versus 23.3% in high-risk group ($p < 0.001$). Kaplan-Meier survival analysis further demonstrated a significant survival advantage for patients in the low-risk group compared to those in the high-risk group ($p = 0.007$) (Fig. 4). Univariate and multivariate Cox regression analyses showed that the model-based risk group (OR = 1.601, 95%CI (1.105–2.318)), Child-Pugh (OR = 1.952, 95%CI (1.063–3.584)), and BCLC (OR = 1.592, 95%CI (1.085–2.336)) were independent prognostic factors for OS (Table 4).

Table 3
Baseline of prognostic test set

n		**Prognostic test set**
		225
Sex (%)	Man	165 (73.3%)
	Woman	60 (26.7%)
Age (year)		59.35 ± 11.33
Etiology (%)	Viral hepatitis	159 (70.7%)
	Other	66 (29.3%)
Child-Pugh	A	209 (92.9%)
	B	16 (7.1%)
BCLC	B	103 (45.8%)
	C	122 (54.2%)
Tumor type (%)	Primary	67 (29.8%)
	Recurrence	158 (70.2%)
Tumor size (cm)		3.74 ± 1.58
ALT (U/L)		38.40 ± 15.75
AST (U/L)		36.95 ± 14.87
TP (g/L)		66.98 ± 6.41
ALB (g/L)		38.92 ± 4.53
TB (μmol/L)		17.03 ± 10.90
DB (μmol/L)		6.67 ± 5.38
AFP (%)	< 20 U/L	111 (49.3%)
	> 20 U/L	114 (50.7%)
Note:		
ALT=Alanine aminotransferase; AST=Aspartate aminotransferase; TP=Total protein; ALB=Serum albumin; TB=Total bilirubin; DB=Direct bilirubin; AFP=Alpha fetoprotein;		

Table 4
Univariate/Multivariate Cox analysis on prognosis

	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	**p**	**HR**	**95% CI**	**p**
Model-based group	1.657	1.1–2.4	0.007	1.601	1.105–2.318	0.013
Sex	0.998	0.98–1	0.802			
Age	0.998	0.65–1.5	0.991			
Etiology	0.692	0.47–1	0.057			
Child-Pugh	1.989	1.1–3.6	0.024	1.952	1.063–3.584	0.031
BCLC	1.662	1.1–2.4	0.008	1.592	1.085–2.336	0.018
Tumor type	0.911	0.61–1.4	0.645			
Tumor size	1.122	1–1.3	0.042	1.1	0.978–1.236	0.112
ALT	1.002	0.99–1	0.783			
AST	0.999	0.99–1	0.99			
TP	1.006	0.98–1	0.676			
ALB	0.993	0.95–1	0.737			
TB	0.984	0.96–1	0.126			
DB	0.965	0.92–1	0.114			
AFP	1.557	1.1–2.3	0.019	1.383	0.949–2.015	0.091
Note:						
ALT=Alanine aminotransferase; AST=Aspartate aminotransferase; TP=Total protein; ALB=Serum albumin; TB=Total bilirubin; DB=Direct bilirubin; AFP=Alpha fetoprotein; TAMs-M2 = Tumor-Associated Macrophages - M2 phenotype; CI = Confidence Interval.						

Discussion

The high recurrence rates and the limited long-term survival of HCC patients was urgently needed an effective method reflected the tumor microenvironment (19, 20). TAMs-M2 played a pivotal role in promoting tumor progression, angiogenesis, and immunosuppression, and are associated with poor prognosis and targeted therapy resistance(21, 22). Currently, TAMs-M2 status was assessed through immunohistochemistry on tissue samples obtained through biopsy, which was invasive, subject to sampling error, and not easily repeatable. In this study, we developed a multimodal model integrating US, MRI and clinical data, for the non-invasive diagnosis of TAMs-M2 in HCC. The multimodal model

demonstrated well diagnostic performance, with AUC 0.797 in external test set. Furthermore, the model effectively stratified patients into high-risk and low-risk groups, which showed significant differences in treatment response and OS. As far as we know, this is the first study to non-invasively predict TAMs-M2 in HCC through medical imaging.

In the process of model development, we utilized and compared various artificial intelligence methods tailored to the characteristics of different imaging modalities, and selected the appropriate methods for US and MRI image. For US image, we used the deep learning algorithm due to its maturity and success in handling two-dimensional image. However, for three-dimensional image, it was lack of highly effective and universally applicable methods in deep learning algorithm. On the contrary, we used the radiomics to process the three-dimensional MRI image, because radiomics was more efficient in capturing the spatial characteristics of tumors, which could supplement the limitation of two-dimensional images that only contain textural features. In the US-based model training, we used three deep learning methods, namely VGG-19, ResNet-50, and DenseNet-121, and ResNet-50 achieved the best performance. Compared to VGG-19, ResNet-50 was significantly deeper yet more parameter-efficient and easier to train. While DenseNet-121 promoted feature reuse, it was computationally expensive and unsteadiness. ResNet-50 achieved an optimal balance: its residual blocks exceled at propagating gradients and preserving crucial features through noisy, low-contrast ultrasound data, leading to robust performance crucial for our task (23–25). In the MRI-based model training, we used pyradiomics to extract features, LASSO to select effective features and three machine learning methods, namely RF, SVM and XGB, to build the model. XGB achieved the best performance. Compared to the parallel trees of RF, the gradient-boosted trees of XGB could sequentially correct errors, and capture complex textual patterns more effectively. And unlike SVM, which struggled with high-dimensional sparse text data and requires careful kernel tuning, XGBoost natively handled feature sparsity, scales efficiently, and automates hyperparameter optimization.

In this study, the individual US-based model and MRI-based model both achieved good performance on TAMs-M2 diagnosis, highlighting the value of image in decoding tumor microenvironment. However, given the value demonstrated in the univariate/multivariate analysis of clinical factors, the complementary nature of two-dimensional and three-dimensional features, and the fact that most recurrent patients undergo both US and MRI examinations, we further developed the multimodal model integrating these three types of information. In the validation set, the multimodal model indeed demonstrated better performance to models based on single-modality information. The multimodal model demonstrated satisfactory and comparable performance in both internal and external test sets, indicating its strong generalizability. In subgroup analyses based on gender, age, and tumor size, the model consistently maintained robust performance, highlighting its excellent robustness. The calibration and decision curve analyses further confirmed that multimodal model was not only accurate but also clinically useful across a wide range of decision thresholds, supporting its potential integration into clinical workflows. In addition, the inherent class imbalance in our cohort yielded a high negative predictive value but a modest positive predictive value for the multimodal model. Consequently, the model was better positioned as a rule-out screening tool to identify patients with a high probability of low

TAMs-M2 burden who are more likely to benefit from targeted therapy, rather than a definitive rule-in diagnostic test for high-risk status.

Building upon the multimodal model's accurate diagnosis of TAMs-M2, we further explored its application for prognostic stratification in HCC patients undergoing targeted therapy, given TAMs-M2's established role as the influence factor on targeted therapy (26, 27). In the prognosis test set, we divided the patients into low- and high- TAMs-M2 groups through the multimodal model, and compared them in terms of the solid tumors therapeutic effect and the overall survival. The comparison results were as we expected. The high- TAMs-M2 group predicted by the model had a lower ORR and DCR, and a shorter OS. Therefore, we believed that the well performance of multimodal model to TAMs-M2 could help doctors identify patients who may have resistance to targeted therapy as early as possible, thereby assisting them in formulating other more effective treatment strategies.

There were some limitations in this study. First, this study was a retrospective study. Although data were collected from multiple centers, prospective test was still required to evaluate the clinical value of the model. Second, the fusion method for multimodal information was relatively simple, which might lead to incomplete integration of cross-modal features, and we would explore more advanced fusion approaches in future. Third, due to the limited sample size, the performance superiority of multimodal model over the unimodal model did not reach statistical significance. More HCCs should be collected in future to evaluated the advantages of multimodal.

Conclusion

In this study, we developed a multimodal model for non-invasively prediction of TAMs-M2 in HCC. By integrating US, MRI and clinical information, the multimodal model achieved the well diagnostic performance on TAMs-M2, and the effectively stratification ability of the prognosis of patients receiving targeted therapy. With the widespread application of targeted therapy, the non-invasive diagnostic model might play an increasingly important role in optimizing therapeutic strategies and improving outcomes.

Declarations

Ethics approval and consent to participate:

The study was approved by the institutional review board of the Shaanxi Provincial People's Hospital [S2022-056-06] and the requirement for informed consent from patients was waived. This study complied with the Helsinki Declaration. All human participants in this study provided informed consent prior to their involvement. Participants were explicitly informed about the purpose of the research, procedures, potential risks/benefits, and their right to withdraw at any stage without penalty. Written consent was obtained from all individuals, and confidentiality of personal data is protected in accordance.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data, images, or clinical details that would compromise anonymity, and therefore no specific consent for publication is required

Competing interests:

All authors disclosed no relevant relationships

Funding:

Young Talent Fund of Association for Science and Technology in Shaanxi, China (No: SWYY202216); Science and Technology Talent Support Plan for Provincial People's Hospital of Shaanxi Province (No: 2022JY-30).

Author Contribution

LX and ML conceived and designed the study. XJJ and LW acquired the data. XJJ did the statistical analyses. LW verified the underlying raw data. All authors had access to the data resented in the manuscript. All authors interpreted the data and contributed to manuscript preparation. All authors were responsible for the decision to submit the manuscript for publication.

Availability of data and materials:

The datasets generated and/or analyzed in the current study are not publicly available due to patient privacy protection but are available from the corresponding author on reasonable request.

References

1. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. Hepatology. 2018;68(2):723–50.
2. Jinhuan W, Wenqi W, Tian Y, Lili W, Li Z, Qing L et al. Tumor-associated macrophages and PD-L1 in prostate cancer: a possible key to unlocking immunotherapy efficacy. Aging. 2024;16(1).
3. Xiaonan X, Jianguo W, Di L, Xiao X. Targeting tumor-associated macrophages to synergize tumor immunotherapy. Signal Transduct Target Ther. 2021;6(1).
4. Jessica WP, Alexandra H-M, William HT, Lucas PW, Selena JL, Daniel SW et al. Antigen presentation by tumor-associated macrophages drives T cells from a progenitor exhaustion state to terminal exhaustion. Immunity. 2024;58(1).

5. Yinghua Z, Ziwei Z, Xin D, Xiaorong L, Zhi-Mei L, Si C et al. Cancer cell-derived arginine fuels polyamine biosynthesis in tumor-associated macrophages to promote immune evasion. Cancer Cell. 2025;43(6).
6. Muhammad T, Nadia H, Kanwal R, Muhammad Sajid Hamid A, Amal H I AH, Amira SAS et al. Macrophages M2 polarization is involved in lapatinib-mediated chemopreventive effects in the lung cancer. Biomed Pharmacother. 2023;161(0).
7. Luca DT, Marco S, Matteo D, Nicola P, Abubaker E, Alessio A et al. Role of liver biopsy in hepatocellular carcinoma. World J Gastroenterol. 2019;25(40).
8. Wendi K, Xiang Q, Yingen L, Jianwei L, Yang L, Junqing X et al. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. J Transl Med. 2023;21(1).
9. Xuanyi W, Tiansong X, Jurui L, Zhengrong Z, Xiaoli Y, Xiaomao G. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. Breast Cancer Res. 2022;24(1).
10. Haiqing F, Yilin L, Fang G, Bin T, Yongqin X, Guipeng W et al. Artificial intelligence-based MRI radiomics and radiogenomics in glioma. Cancer Imaging. 2024;24(1).
11. Marie-Caline ZA, Barry W, Carolyn E. J A. Imaging the Tumor Microenvironment. Adv Exp Med Biol. 2017;1036(0).
12. Julien C, Tobias Paul S, Tom L, Tracey GS. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol. 2022;76(6).
13. Ding W, Wu J, Dou J, Liu F, Han Z, Yu J et al. Biologically explicable multimodal model predicting local tumor progression and tumor invasiveness of hepatocellular carcinoma. Hepatology. 2026.
14. Ding W, Meng Y, Ma J, Pang C, Wu J, Tian J, et al. Contrast-enhanced ultrasound-based AI model for multi-classification of focal liver lesions. J Hepatol. 2025;83(2):426–39.
15. Yu-Ting S, Liang C, Wen-Wen Y, Hui-Xiong X. Artificial intelligence in ultrasound. Eur J Radiol. 2021;139(0).
16. Tianyi X, Ben Z, Binrong L, Ying L, Yang S, Yuancheng W et al. MRI-Based Radiomics and Deep Learning in Biological Characteristics and Prognosis of Hepatocellular Carcinoma: Opportunities and Challenges. J Magn Reson Imaging. 2023;59(3).
17. Ahmed H, Chintan P, John Q, Lawrence HS, Hugo. J W L A. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:8.
18. Yeung OWH, Lo C-M, Ling C-C, Qi X, Geng W, Li C-X, et al. Alternatively activated (M2) macrophages promote tumour growth and invasiveness in hepatocellular carcinoma. J Hepatol. 2015;62(3):607–16.
19. Lichun M, Maria OH, Yongmei Z, Monika M, Bao T, Michael K et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. Cancer Cell. 2019;36(4).

20. Senlin Z, Yushuai M, Binbin Z, Ping W, Yanzi G, Zhengxiang Z et al. Highly-metastatic colorectal cancer cell released miR-181a-5p-rich extracellular vesicles promote liver metastasis by activating hepatic stellate cells and remodelling the tumour microenvironment. J Extracell Vesicles. 2022;11(1).
21. Kun C, Ning C, Jinghan Z, Xing Y, Huifang L, Wanguang Z. Tumor-associated macrophages in liver cancer: From mechanisms to therapy. Cancer Commun (Lond). 2022;42(11).
22. Chun W, Cuimin C, Wenting H, Lili T, Jiakang C. Revealing the role of necroptosis microenvironment: FCGBP + tumor-associated macrophages drive primary liver cancer differentiation towards cHCC-CCA or iCCA. Apoptosis. 2023;29(0).
23. Yongfeng Y, Yangdan C, Tian'an J, Jianhua Z, Li Z, Baowen Q et al. Ultrasound identification of hepatic echinococcosis using a deep convolutional neural network model in China: a retrospective, large-scale, multicentre, diagnostic accuracy study. Lancet Digit Health. 2023;5(8).
24. Zhe X, Wenlong Z, Jianyang G, Guirong T, Xiaopeng C, Jianpeng S et al. Artificial Intelligence in the Evolution of Interventional Therapy for Liver Cancer: Navigating the Tumor Microenvironment. Front Biosci (Landmark Ed). 2025;30(10).
25. Wei W, Weicai H, Xianqi Y, Cheng F, Hongkun W, Zhe L et al. Multimodal radiopathomics approach for predictions of prognosis and immunotherapy response in patients with GC: a multicohort retrospective study. Int J Surg. 2025;111(11).
26. Benjamin R, Matthias B, Sepideh B, Noemi K, Lichun M, Mahler R et al. Tumor-associated macrophages trigger MAIT cell dysfunction at the HCC invasive margin. Cell. 2023;186(17).
27. Ankur S, Justine Jia Wen S, Charles-Antoine D, Rhea P, Camille B, Archita M et al. Onco-fetal Reprogramming of Endothelial Cells Drives Immunosuppressive Macrophages in Hepatocellular Carcinoma. Cell. 2020;183(2).

Figures

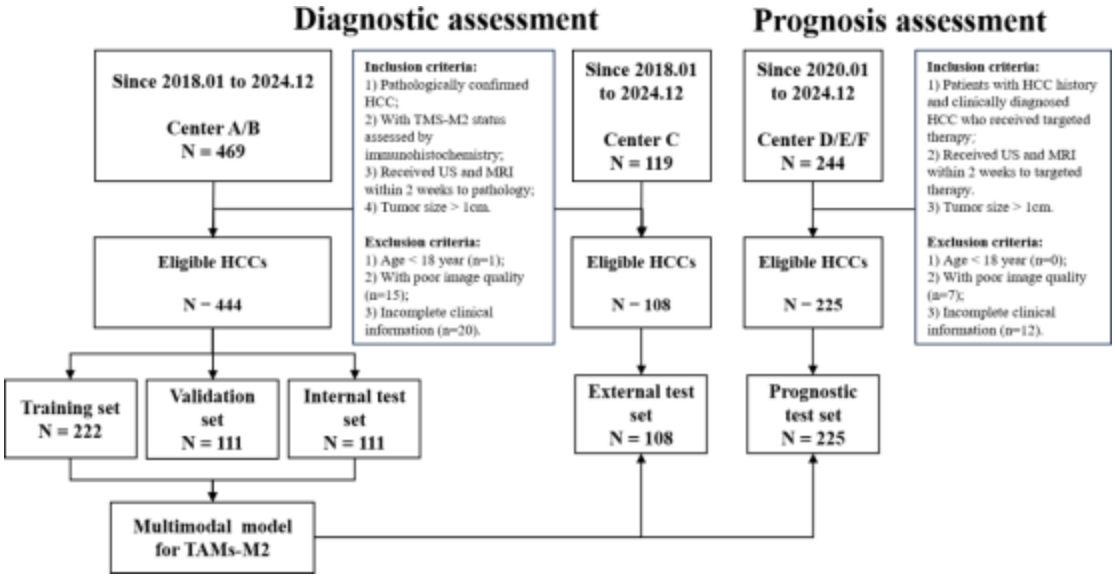


Figure 1

The Flowchart

Since January 2018 to December 2024, patients from Centers A and B were collected to establish the multimodal model for TAMs-M2, and patients from Center C were collected as external test set to evaluate the model diagnostic ability. Since January 2020 to December 2024, patients from Centers D, E and F were collected as prognostic test set to assess the impact of the model's prediction results on prognosis.

HCC = Hepatocellular carcinoma; TAMs-M2 = Tumor-associated macrophages with M2 polarization.

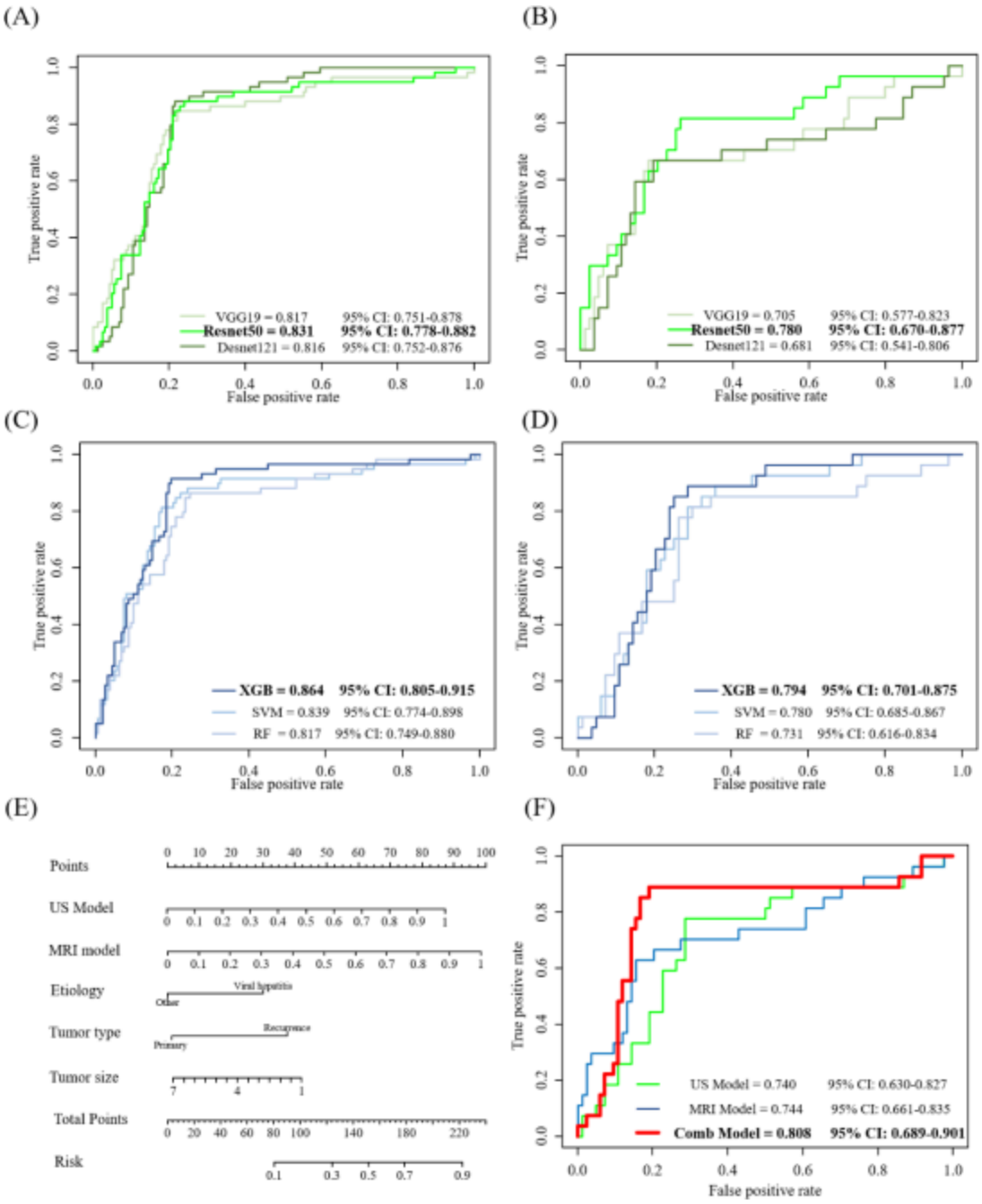


Figure 2

Model performance in training and validation sets.

The performance of US model in training set (A) and validation set (B); the performance of MRI model in training set (C) and validation set (D); the nomogram of combined model (E); the comparison of US model, MRI model and multimodal model in validation set (F).

XGB = XGBoost; SVM = Support Vector Machine; RF = Random Forest; CI = Confidence interval.

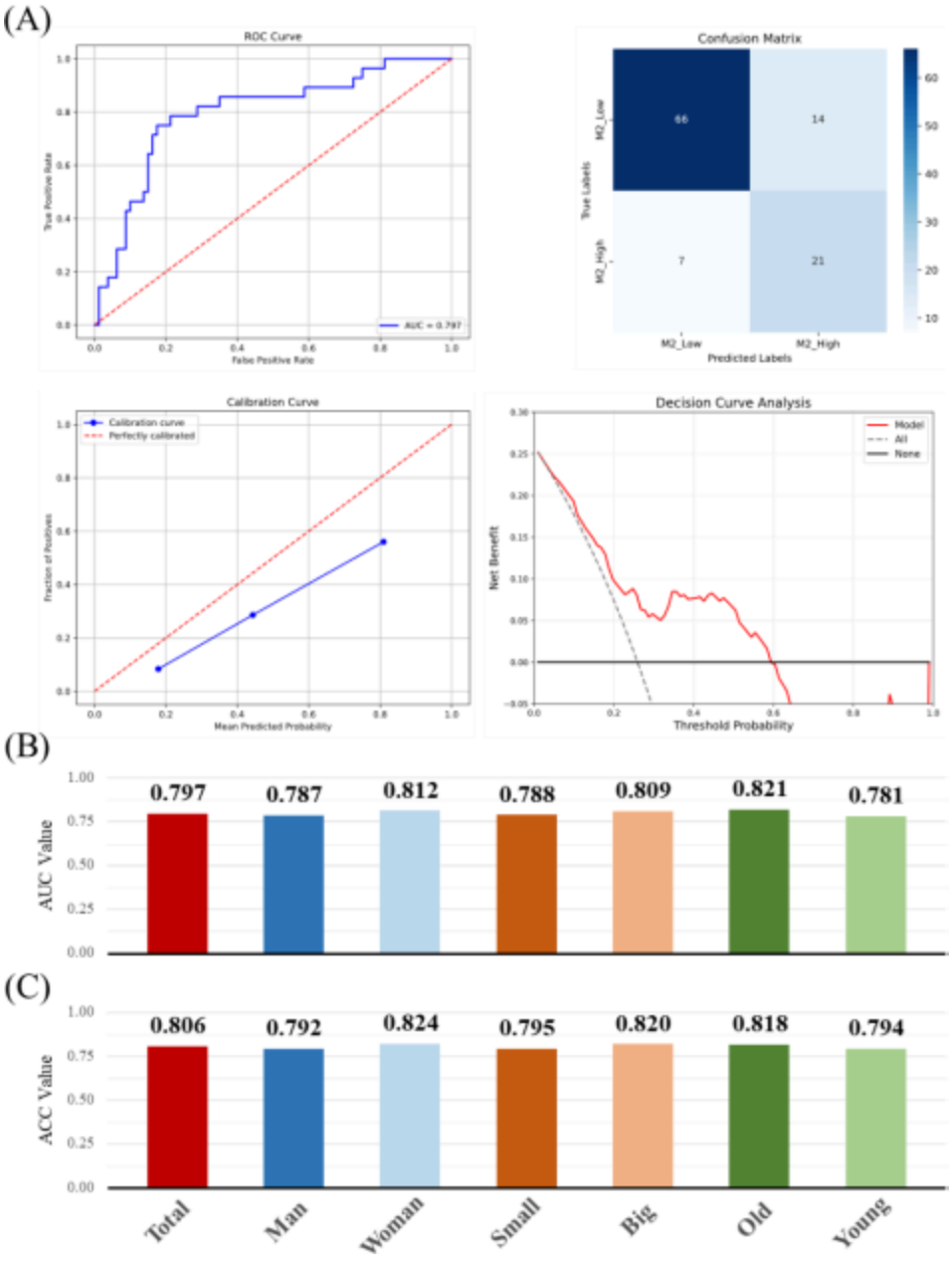


Figure 3

Multimodal model performance on TAMs-M2 diagnosis in external test set.

(A) ROC curve, confusion matrix, calibration curve and decision curve of multimodal model in external test set; (B) and (C) AUC and ACC of multimodal model in subgroup analyses based on different

genders, tumor sizes, and patient ages.

TAMs-M2 = Tumor-associated macrophages with M2 polarization; ROC = Receiver operating characteristic; AUC = Area under the receiver operating characteristic curve; ACC = Accuracy.

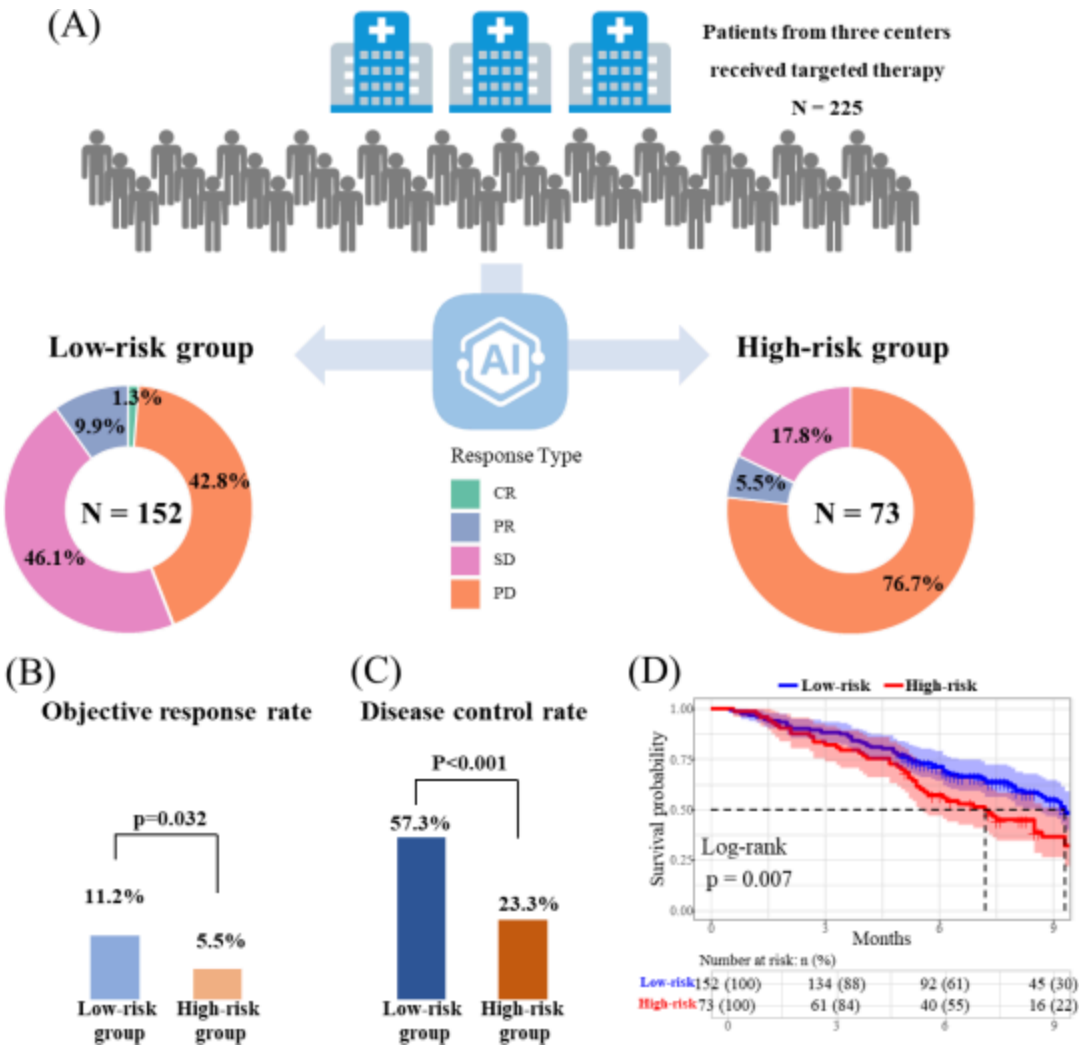


Figure 4

Multimodal model performance on prognosis of patients with targeted therapy.

225 Patients from three centers were classified into Low-risk group and High-risk. We compared the proportions of CR, PR, SD and PD (A); the Objective response rate (B); the Disease control rate (C); and the OS (D).

CR = complete remission; PR = partial remission; SD = stable disease; PD = progress disease; K-M curve = Kaplan-Meier curve; OS = Overall survival.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx