



ORIGINAL ARTICLE

CEUS-Based Microvascular Invasion Predictor in HCC: Improving Prognostic Stratification Following Thermal Ablation

Keke Chen¹ | Shukang Zhang² | Tianjiao Huang³ | Yadan Xu¹ | Qi Zhang¹ | Yuli Zhu¹ | Peng Wan² | Wentao Kong^{3,4} | Wenping Wang¹

¹Department of Ultrasound, Zhongshan Hospital, Fudan University, Shanghai, P. R. China | ²College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China | ³Department of Ultrasound, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, P. R. China | ⁴Department of Ultrasound Medicine, Affiliated People's Hospital of Jiangsu University, Zhenjiang, P. R. China

Correspondence: Peng Wan (pengwan@nuaa.edu.cn) | Wentao Kong (breezewen@163.com) | Wenping Wang (puguang61@126.com)

Received: 26 August 2025 | **Revised:** 14 March 2026 | **Accepted:** 8 April 2026

Handling Editor: Dr. Alejandro Forner

Keywords: contrast-enhanced ultrasound | deep learning | hepatocellular carcinoma | microvascular invasion; thermal ablation

ABSTRACT

Background & Aims: Microvascular invasion (MVI) critically impacts hepatocellular carcinoma (HCC) management. We aimed to develop and validate a deep learning model integrating contrast-enhanced ultrasound (CEUS) and clinical features for assessing MVI risk preoperatively, and to explore its prognostic associations in the thermal ablation (TA) cohort.

Methods: We enrolled 688 patients with solitary HCC ≤ 5 cm undergoing CEUS before surgical resection (SR) or TA. CEUS features were extracted via a vision transformer and integrated with clinical features to build an MVI classifier for SR patients. SHAP plots visualized key variables. Recurrence-free survival (RFS) was then compared between model-stratified risk groups in the TA cohort. Cox regression analysis identified risk factors for RFS.

Results: The model achieved AUCs of 0.89 (internal validation) and 0.81 (external validation) for MVI prediction. The decision curve analysis further confirmed the clinical utility. SHAP plots identified AFP-L3%, PIVKA-II, and tumour size as the most influential clinical features. Among the CEUS features, the most significant feature was the arterial-phase dynamic sequence. In the TA cohort, patients with model-predicted high-risk MVI had significantly lower 1-, 2- and 3-year RFS rates (75.5%, 54.9%, 39.4%) compared to those with low-risk (83.8%, 80.0%, 78.0%) ($p < 0.001$). High MVI risk predicted by the model (HR = 2.90; 95% CI: 1.69–4.96) and AFP-L3% (HR = 1.34; 95% CI: 1.06–1.69) were independently associated with RFS in TA patients.

Conclusions: The model accurately predicted MVI in surgical candidates and showed exploratory prognostic value in TA patients. As a retrospective single-centre study, it warrants prospective validation.

Abbreviations: ACC, accuracy; AFP, alpha-fetoprotein; ALBI, albumin-bilirubin grading; AP, arterial phase; AUC, area under the curve; BMUS, B-mode ultrasound; CB, categorical boosting; CEUS, contrast-enhanced ultrasound; CNLC, China liver cancer staging; DCA, decision curve analysis; DL, deep learning; DP, delayed phase; HCC, hepatocellular carcinoma; HR, hazard ratio; LGBM, light gradient boosting machine; LR, logistic regression; ML, machine learning; MVI, microvascular invasion; MWA, microwave ablation; NN, Neural Network; NPV, negative predictive value; PPV, positive predictive value; PRE, precision; PVP, portal venous phase; RF, random forest; RFA, radiofrequency ablation; RFS, recurrence-free survival; ROC, receiver operating characteristic; SEN, sensitivity; SHAP, Shapley Additive Explanations; SPE, specificity; SR, surgical resection; SVM, support vector machine; TA, thermal ablation; ViT, Vision transformer; XGB, eXtreme gradient boosting.

Keke Chen and Shukang Zhang contributed equally to this work.

Lay Summary

What Is Already Known on This Topic

1. Microvascular invasion (MVI) is a critical prognostic factor for recurrence and metastasis in hepatocellular carcinoma (HCC) after radical treatment.
2. Previous studies have demonstrated the applicability of deep learning (DL) methods in predicting MVI using medical images. However, there are relatively few studies on multi-phase contrast-enhanced ultrasound (CEUS) that can provide temporal and spatial perfusion information of the microcirculation in HCC.
3. There is still a research gap in the development of the MVI DL architecture for thermal ablation (TA) candidates whose intact tumour tissue cannot be obtained.

What This Study Adds

1. An MVI prediction model for patients with China Liver Cancer Staging (CNLC) Ia-stage HCC was developed based on DL features from multi-phase CEUS data and clinical characteristics.
2. The MVI risk value obtained based on the model is an important prognostic factor for predicting recurrence after TA and can serve as a key variable to assist in evaluating the risks and benefits of TA.

How This Study Might Affect Research, Practice or Policy

1. Our model demonstrated robust performance across both internal and external validation cohorts in predicting MVI for early-stage HCC.
2. Patients stratified into the high-risk MVI may derive greater benefit from either choosing surgical resection or combining TA with postoperative adjuvant therapy.

when stratified by clinical-imaging risk models [7]. Moreover, adjuvant therapy improved survival outcomes in MVI-positive patients [8, 9]. Consequently, accurate preoperative MVI prediction is critical for treatment planning and optimizing prognostic outcomes.

To date, several studies have developed MVI prediction models using clinical and imaging features [10–13]. However, limited specificity/sensitivity of serological indicators and interobserver variability in imaging interpretation result in variable model performance and hinder model reproducibility. In recent years, the advancement of artificial intelligence has introduced a promising new strategy for radiological analysis, demonstrating encouraging results in the diagnosis and therapeutic efficacy evaluation of various diseases [14]. The most significant technology in this domain is deep learning (DL) algorithms, which use large-scale datasets to construct deep neural networks that reveal underlying imaging features beyond conventional radiomics [15]. Multiple studies have further validated its effectiveness and superiority in predicting MVI [15–19]. Notably, as the cornerstone of HCC surveillance, ultrasonography remains underexplored for MVI prediction, especially contrast-enhanced ultrasound (CEUS) with real-time microvascular perfusion imaging. Only a few studies have reported DL-based CEUS algorithms for predicting MVI [20–24]. Crucially, existing research neglected vital dynamic information and clinical parameter integration. Whether the application of multimodal CEUS-based DL can further enhance predictive performance remains unestablished.

The first aim of our study is to develop and validate a preoperative MVI prediction model integrating multimodal CEUS with clinical variables in patients with early-stage HCC undergoing SR. The second aim is to explore the association of the model-predicted MVI risk score with prognosis in patients receiving TA, generating hypotheses for future research on personalized treatment strategies.

1 | Introduction

Primary liver cancer ranks as the sixth most common cancer globally and the third leading cause of cancer-related deaths. Hepatocellular carcinoma (HCC) represents 75%–85% of cases [1]. For Barcelona Clinic Liver Cancer (BCLC) 0 (very early-stage) or BCLC A (early-stage) HCC, both ablation therapy and surgical resection (SR) can achieve curative outcomes [2]. Particularly in BCLC 0 stage, thermal ablation (TA) demonstrates results comparable to resection with less invasiveness, establishing it as a first-line therapy [3]. However, the high recurrence rate remains a global challenge in HCC management [1].

Microvascular invasion (MVI) is a critical prognostic factor for recurrence and metastasis in HCC. Higher MVI grades correlate with progressively worse clinical outcomes [4, 5]. It is histopathologically defined as the presence of tumour nests within endothelial-lined vascular channels. Thus, its assessment exclusively requires histopathological examination of surgically resected specimens [6]. However, obtaining intact tumour tissue is impractical following ablation therapy. Lee et al. demonstrated that MVI-positive patients undergoing radiofrequency ablation (RFA) exhibit higher recurrence rates than those receiving SR

2 | Materials and Methods

2.1 | Patients

From January 2018 to January 2022, 8368 patients who received SR and 1876 patients who received TA were retrieved and collected through the electronic case system of Centre A. Treatment decisions were made by the multidisciplinary recommendations of hepatobiliary surgeons, oncologists and radiologists, as well as patient preferences. The inclusion criteria of SR and TA patients were as follows: (1) Patients with pathologically or clinically confirmed China Liver Cancer Staging (CNLC) Ia HCC (solitary tumour ≤ 5 cm, performance status score 0–2, Child-Pugh class A or B, without macrovascular invasion or extrahepatic metastasis); (2) TA or SR as the initial treatment; (3) Achievement of complete TA or SR one-month post-procedure. (4) SR patients had MVI data; (5) Acquisition of CEUS within 2 weeks prior to SR or TA. The exclusion criteria were as follows: (1) Poor quality of CEUS images or incomplete clinical data; (2) A follow-up period after TA shorter than 6 months. SR patients from Centre A were divided into training and internal validation cohorts in an 8:2 ratio. From

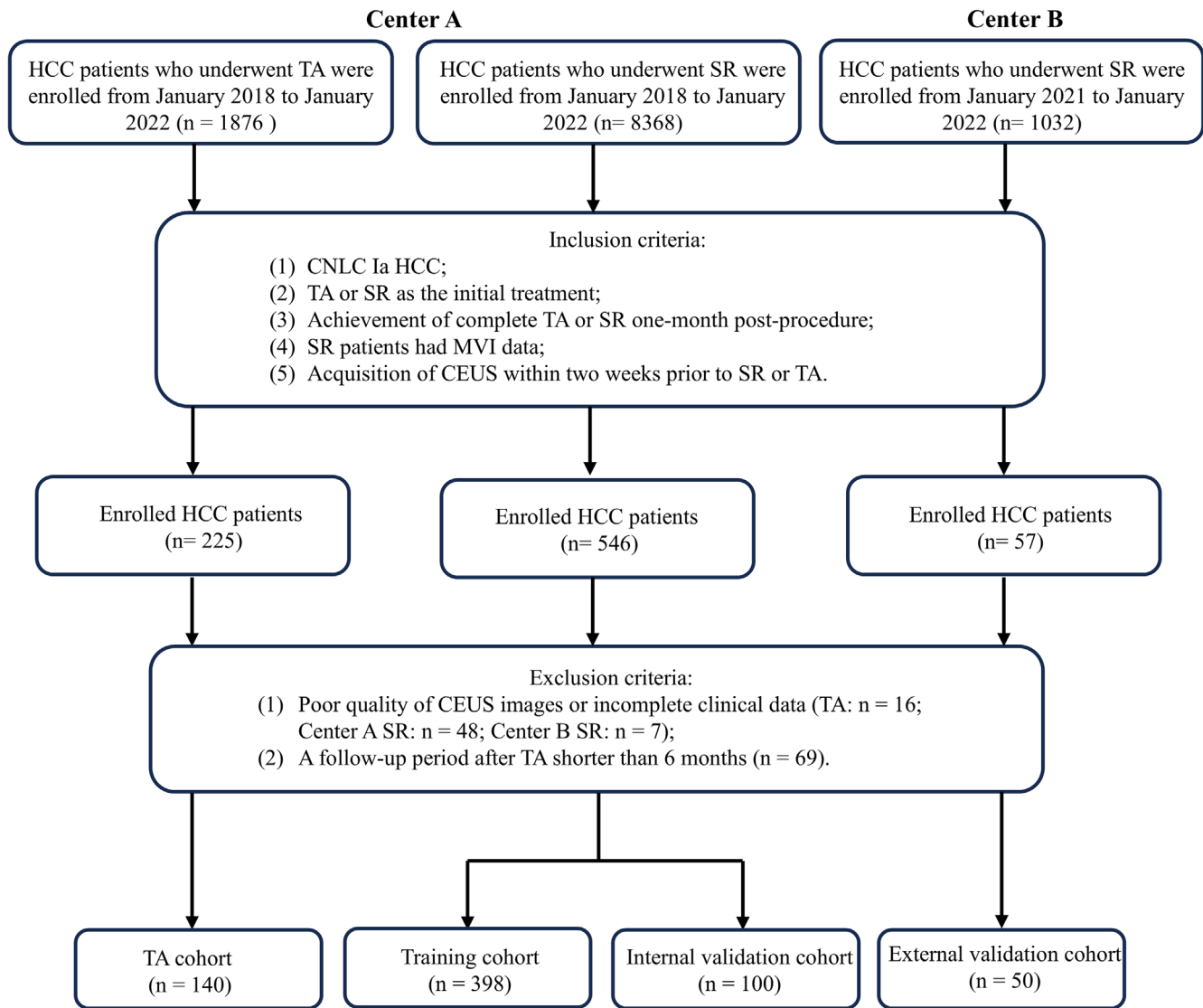


FIGURE 1 | Flowchart of patient inclusion and exclusion.

January 2021 to January 2022, HCC patients treated by SR meeting the above criteria at Centre B were retrospectively enrolled as the external validation cohort. Patient enrollment flow chart was presented in Figure 1. Baseline patient characteristics were recorded. Albumin-bilirubin grading (ALBI) scores were calculated using the previously established formula: $ALBI = (\log_{10} TBIL [\mu\text{mol/L}] \times 0.66) + (\text{albumin [g/L]} \times -0.085)$. Patients were divided into three ALBI grades: Grade 1 (ALBI score ≤ -2.60), Grade 2 ($-2.60 < \text{ALBI score} \leq -1.39$) and Grade 3 (ALBI score > -1.39).

2.2 | CEUS Protocol

All the ultrasound examinations were performed by senior ultrasound physicians. CEUS were performed by LOQIQ E9 (GE Healthcare, Milwaukee, WI, USA), Philips EPIQ7 (Philips, Bothell, WA, USA) and Resona R9 Super (Mindray, Shenzhen, GD, China) using the contrast agent of SonoVue (Bracco Imaging, Milan, Italy). The CEUS protocols were shown in [Supporting Information](#).

2.3 | Histopathological Assessment of MVI Status

We performed 1:1 sampling at the tumour margins (12, 3, 6 and 9 o'clock positions) where cancer meets non-cancerous liver tissue, following the 7-point sampling protocol [4]. Additionally, we obtained one tissue sample each from within the tumour itself and from liver tissue located ≤ 1 cm (proximal to the tumour margin) and > 1 cm (distal to the tumour margin) from the tumour edge. MVI refers to the microscopic detection of cancer cell clusters within vascular lumens lined by endothelial cells. All histopathological sections were independently assessed by two experienced pathologists, with discrepancies arbitrated by a third pathologist. This study classified MVI status as either negative or positive.

2.4 | Patient Follow-Up

CEUS or Contrast-enhanced magnetic resonance imaging (CEMRI) examinations were performed 1 month following TA to confirm the absence of abnormal enhancement and washout, thereby ensuring its effectiveness.

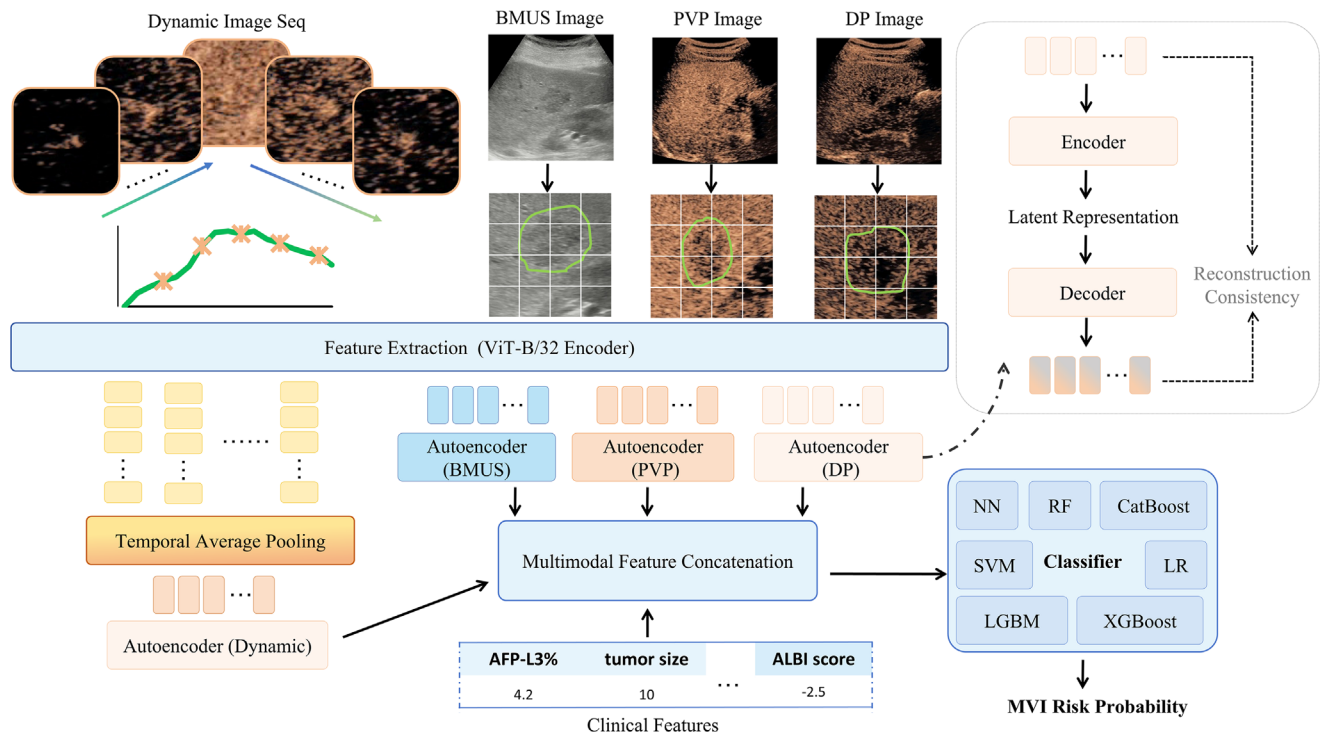


FIGURE 2 | The workflow of constructing the MVI classifier.

Subsequently, patients underwent follow-up every 3 months for the first 2 years and every 6 months thereafter, including abdomen-pelvis ultrasound or contrast-enhanced imaging and serum alpha-fetoprotein (AFP) levels. The endpoint of the study was recurrence-free survival (RFS), defined as the time interval from the date of TA to the date of first recurrence or last visit. The last follow-up date was March 2025. All recurrences were determined by at least two different imaging features or histopathological examination.

2.5 | Image Annotation and Preprocessing

This study utilized multi-phase ultrasound images, including static images from the B-mode ultrasound (BMUS), portal venous phase (PVP) and delayed phase (DP), as well as dynamic arterial phase (AP) sequences. The data preprocessing steps were as follows: (1) Annotation and Expansion of Static Images: Experienced radiologists annotated lesion boundaries on all static images using 3D Slicer software (version 5.7.0, <https://www.slicer.org/>). Based on the physician annotations, morphological dilation was applied to expand the lesion region, ensuring the input images encompassed surrounding tissue areas to enhance the model's perception of tumour margins and peritumoral features. (2) Keyframe Sampling from dynamic AP sequences: Physicians first annotated lesions in dynamic AP sequences using rectangular bounding boxes with the aid of auxiliary annotation and tracking software, generating time-intensity curves (TIC). Ten key frames were distributedly selected based on TIC features. (3) Image Size and Pixel Standardization: All images were uniformly resized to a resolution of 224×224 pixels to accommodate DL model input requirements. Pixel values were standardized to the $[0, 1]$ range to minimize imaging parameter variations. (4) Clinical Feature Preprocessing: All

clinical feature data underwent consistency checks. Missing values were median-imputed (Table S1). We also tested alternative imputation strategies, including K-Nearest Neighbours imputation (KNN) and multiple imputation (MI), and observed no meaningful differences in model performance (Table S2). (5) Label Annotation and Data Augmentation: MVI labels, serving as the gold standard, were provided by postoperative pathological examination results. Due to the imbalance of original dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was applied during model training to balance the classes (1:1 ratio). Internal and external validation cohorts were kept unchanged to preserve the original distribution and ensure unbiased performance evaluation. To ensure that SMOTE did not artificially inflate model performance, we conducted additional comparisons with two alternative strategies: random undersampling and class-weighted training. These findings indicate that SMOTE effectively mitigated majority-class dominance while avoiding overestimation (Table S3).

2.6 | Model Development

This study employed a Vision Transformer (ViT-B/32)-based deep neural network to extract features from multimodal ultrasound images and achieved feature dimensionality reduction through autoencoders to generate compact representations for classification. This process comprised three stages: ViT-B/32 feature extraction, dynamic AP sequence processing, and autoencoder dimensionality reduction (Figure 2). Each input image was first linearly embedded, segmented into fixed-size patches, and mapped into a high-dimensional space. After processing through multiple transformer encoder layers, global visual features were obtained. Average pooling was applied to dynamic AP sequences along the temporal

dimension to compress the features, yielding static representations. This preserved key enhancement information while reducing feature redundancy. A symmetrically structured autoencoder was designed to reduce the dimensionality of high-dimensional features from each modality into one-dimensional latent representations. The autoencoder, featuring an encoder $\phi(\cdot)$ and a decoder $\psi(\cdot)$, was trained for 20 epochs using the Adam (learning rate: 1×10^{-3}). Finally, the above features were concatenated with clinical features to form the final input feature vector. This feature vector was input into various classification models, including Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM) and Categorical Boosting (CB), to predict MVI risk probability. Hyperparameters for each classifier were optimized using cross-validation and grid search. The full hyperparameter settings for all downstream classifiers were reported in the [Supporting Information](#). Subsequently, the optimal classifier was used to generate the MVI risk score for the TA cohort. To ensure fairness and clarity, we confirm that all models were trained using the same set of clinical variables, which are listed in Table 1, and no model incorporated additional or model-specific clinical features.

2.7 | Statistical Analysis

All statistical analyses were performed within the Python 3.8 environment (<https://www.python.org/>) and related scientific computing libraries. Stratified random sampling was employed to partition the dataset into training and validation cohorts. The MVI classification model was trained on the training cohort and evaluated on the validation cohort. The hyperparameters were determined through grid search combined with five-fold cross-validation. Model performance was quantified using the area under the curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), positive predictive value (PPV), negative predictive value (NPV), and F1 score. For each fold, repeated experiments were conducted and the AUC together with its 95% confidence interval (95% CI) was estimated from the 2.5th and 97.5th percentiles of the AUC distribution. The overall model performance across folds was summarized as mean \pm standard deviation (SD). The clinical utility of the model was assessed using decision curve analysis (DCA) to evaluate net benefit across varying probability thresholds. Shapley Additive Explanations (SHAP) values were used to analyse the importance of variables in predicting MVI. The deep imaging features and clinical characteristics of patients undergoing TA were input into the MVI classification model to generate the MVI risk probability for each patient. The optimal cutoff for the MVI risk probability was determined in the training cohort only according to the maximum Youden index on the receiver operating characteristic (ROC) curve. The internal validation, external validation, and TA cohorts were not involved in threshold determination or optimization. Survival curves were plotted using the Kaplan–Meier method with 95% CI bands estimated using the Greenwood formula, and the differences in RFS among different risk groups were compared using the log-rank test. Univariate and multivariate Cox regression analyses were used to identify the factors influencing RFS, reporting the hazard

ratio (HR) and its 95% CI. The inter-group comparison for continuous variables was conducted using the Mann–Whitney U test, while categorical variables were compared using the Chi-square test or Fisher's exact test, as appropriate. All statistical tests were two-sided, and $p < 0.05$ was used to determine statistical significance.

3 | Results

3.1 | Patient Characteristics

This study retrospectively enrolled 688 patients with solitary HCC ≤ 5 cm from two institutions. The SR dataset from Centre A was randomly divided into the training cohort ($n = 398$) and the internal validation cohort ($n = 100$) in an 8:2 ratio. The SR dataset from Centre B was assigned to the external validation cohort ($n = 50$). The prevalence of MVI positivity was 26.4% (105/398) in the training cohort, 27% (27/100) in the internal validation cohort, and 34% (17/50) in the external validation cohort. Among the 140 patients who underwent TA, 18 patients (12.9%) underwent intraoperative approach (TA treatment performed via laparoscopic or open surgical approach). The baseline clinical characteristics of the training, internal validation, and external test cohorts were comprehensively summarized in Table 1.

3.2 | Construction of the MVI Classifier

We integrated the DL features with the clinical characteristics of the patients and input them into multiple machine learning (ML) classifiers to establish MVI prediction models. Model performance was evaluated using five-fold cross-validation (Tables S4 and S5). The results showed that the model based on XGB (ViT-XGB model) demonstrated the highest prediction accuracy for MVI and exhibited satisfactory MVI prediction ability in all folds. The ViT-XGB model achieved mean AUC values of 0.89 and 0.81 in the internal and external validation cohorts, respectively. The corresponding fold-specific 95% CIs for the AUC values were provided in the [Supporting Information](#). Other diagnostic indicators of all prediction models, including ACC, SEN, SPE, PRE, PPV, NPV and F1 Score, were shown in Table 2 and Table S6. Notably, when compared to existing non-DL MVI prediction models based on clinical and imaging data (AUC range: 0.71–0.83), our multimodal DL framework (ViT-XGB) demonstrated a superior performance (Table S7). Additionally, we compared the performance of imaging-only models that used BMUS alone, CEUS alone, or the combination of BMUS and CEUS under the ViT-XGB framework. The results showed that the CEUS-only model achieved the highest AUC (0.71), outperforming both the BMUS-only model (AUC = 0.67) and the combined BMUS+CEUS model (AUC = 0.69). While the addition of BMUS increased accuracy under a fixed threshold, it did not improve AUC performance. When compared with our ViT-XGB model that integrated imaging and clinical variables, the all imaging-only models showed relatively inferior performance (Table S8). To clarify the specific contribution of the ViT-XGB architecture, we compared it with recent transformer-based models. We first evaluated several backbone architectures—ResNet-50, ViT-ImageNet and MedCLIP (Swin Transformer)—by concatenating their extracted multimodal

TABLE 1 | Baseline patient characteristics.

Characteristics	Training cohort (<i>n</i> = 398)	Internal validation cohort (<i>n</i> = 100)	External validation cohort (<i>n</i> = 50)	<i>p</i> ^a	TA cohort (<i>n</i> = 140)	<i>p</i> ^b
Gender (male/female)	318/80	88/12	40/10	0.171	102/38	0.025
Age (y)	58.2 ± 10.7	56.1 ± 10.9	57.2 ± 9.6	0.207	60.3 ± 11.3	0.018
Cirrhosis (present/ absent)	223/175	54/46	34/16	0.226	116/24	<0.001
Fatty liver (present/ absent)	82/316	19/81	12/38	0.775	14/126	0.004
Tumour size (mm)	27.0 (19.8, 38.0)	29.0 (18.0, 36.8)	24.0 (19.8, 35.3)	0.789	18.0 (14.0, 21.0)	<0.001
PLT (× 10 ⁹ /L)	152.0 (115.8, 192.3)	156.0 (122.3, 193.5)	144.0 (117.0, 183.0)	0.545	118.0 (69.0, 170.0)	<0.001
TBil (μmol/L)	13.0 (9.7, 17.0)	12.9 (9.9, 17.1)	13.3 (10.5, 18.0)	0.970	14.2 (10.8, 19.7)	0.003
TP (g/L)	68.0 (65.0, 73.0)	68.0 (65.2, 72.0)	68.1 (64.0, 71.5)	0.663	73.0 (68.0, 75.8)	<0.001
ALB (g/L)	43.0 (41.0, 46.0)	43.0 (41.0, 46.0)	42.5 (39, 44.0)	0.885	42.5 (37.2, 46.0)	0.029
ALT (U/L)	23.0 (17.0, 34.0)	24.0 (18.0, 33.8)	27.5 (19.7, 34.8)	0.636	30.0 (20.0, 43.8)	<0.001
AST (U/L)	23.5 (19.0, 32.0)	24.5 (19.0, 30.0)	24.0 (19.9, 30.5)	0.969	33.0 (24.0, 42.0)	<0.001
ALP (U/L)	75.0 (62.0, 93.0)	76.0 (63.0, 89.0)	81.4 (67.2, 100.3)	0.737	88.5 (65.3, 108.0)	<0.001
LDH (U/L)	172.5 (154.0, 192.0)	170.5 (156.0, 191.5)	178.0 (151.8, 194.3)	0.946	188.1 (176.3, 203.3)	<0.001
PT (s)	12.0 (11.5, 12.5)	12.0 (11.5, 12.4)	11.9 (11.2, 12.6)	0.757	12.5 (11.7, 13.4)	<0.001
HA (ng/mL)	89.6 (60.1, 143.8)	82.2 (60.0, 135.5)	101.4 (78.9, 152.6)	0.242	197.5 (87.6, 292.4)	<0.001
HBsAg (positive/ negative)	308/90	78/22	39/11	0.988	101/39	0.178
Anti-HCV (positive/ negative)	12/386	2/98	2/48	0.772	6/134	0.412
AFP-L3%	9.5 (0.8, 23.6)	11.9 (0.7, 23.6)	20.1 (8.0, 25.6)	0.897	18.6 (8.9, 28.1)	<0.001
PIVKA-II (mAU/ml)	96.5 (30.0, 433.8)	54.5 (27.0, 367.2)	65.7 (16.8, 592.4)	0.226	52.0 (26.0, 108)	<0.001
AFP (ng/ml)	14.7 (3.3, 194.3)	10.8 (3.7, 207.8)	30.7 (3.3, 303.1)	0.955	19.8 (4.7, 111.0)	0.770
CEA (ng/ml)	2.4 (1.7, 3.5)	2.7 (1.6, 3.7)	1.5 (0.8, 2.2)	0.746	2.8 (1.9, 3.8)	0.003
CA199 (ng/ml)	14.1 (8.5, 21.7)	13.8 (7.7, 21.7)	12.9 (8.7, 20.1)	0.584	21.1 (10.0, 37.1)	<0.001
Child-Pugh (A/B)	397/1	100/0	49/1	0.201	132/8	<0.001

(Continues)

TABLE 1 | (Continued)

Characteristics	Training cohort (<i>n</i> = 398)	Internal validation cohort (<i>n</i> = 100)	External validation cohort (<i>n</i> = 50)	<i>p</i> ^a	TA cohort (<i>n</i> = 140)	<i>p</i> ^b
ALBI grade (1/2/3)	344/54/0	89/11/0	43/7/0	0.780	94/46/0	<0.001
MVI (positive/ negative)	105/293	27/73	17/33	0.521	/	/

Note: Data are numbers, means \pm SD, or medians (ranges).

Abbreviations: AFP, alpha-fetoprotein; AFP-L3%, alpha-fetoprotein heterogeneity; ALB, albumin; ALBI, albumin-bilirubin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; anti-HCV, hepatitis C antibody; AST, aspartate aminotransferase; CA199, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; HA, hyaluronic acid; HBsAg, hepatitis B surface antigen; LDH, lactic dehydrogenase; MVI, microvascular invasion; IVKA-II, protein induced by vitamin K absence-II; PLT, platelet; PT, prothrombin time; TBil, total bilirubin; TP, total protein.

^a*p* indicates the statistical significance between the internal training cohort, internal validation cohort, and external validation cohort.

^b*p* value indicates statistical significance between the SR and TA cohort.

ultrasound embeddings with clinical variables, followed by training a fully connected classifier. Then, we replaced the ViT backbone with a MedCLIP-pretrained Swin Transformer while keeping the XGBoost classifier unchanged. The results demonstrated that traditional end-to-end models achieved only moderate performance (AUC ranging from 0.73 to 0.83), whereas hybrid models like ViT-XGB and SwinViT-XGB achieved significantly higher AUC of 0.89 (Table S9). The ROC curve demonstrated the discriminatory ability of the ViT-XGB model in the internal and external validation cohort (Figure 3A,B). Additionally, the DCA showed that the net benefit of the ViT-XGB model was consistent across a range of probability thresholds, thereby establishing its clinical utility of this model (Figure 3C).

3.3 | Feature Selection and Correlation Heatmap for the Model

We employed the SHAP framework to visualize the key variables contributing to the predicted MVI probability. The horizontal position represents the magnitude and direction of each feature's impact on MVI prediction. The colour gradient reflects the feature value (low to high). The relative importance of each feature was assessed by its mean absolute SHAP value and displayed in descending order (Figure 4A). The results revealed that the three most influential variables for the model were AFP-L3%, PIVKA-II and tumour size. Among the CEUS features, the dynamic sequence in the arterial phase demonstrated the highest predictive contribution. Given that AFP-L3% and PIVKA-II are not routinely available in all clinical settings, we developed a simplified model (Model 1) excluding these markers and compared it with the complete model (Model 2). The results demonstrated a significant decline in the performance of Model 1, with the AUC decreasing from 0.89 ± 0.04 to 0.71 ± 0.04 and the ACC dropping from 0.85 ± 0.02 to 0.73 ± 0.02 (Table S10 and Figure S1). These findings indicated that although the simplified model remained feasible to a certain extent, its predictive performance was inferior to that of the complete model. To mitigate multicollinearity and ensure model stability, we used a Spearman's correlation heatmap to analyse the pairwise correlations among the top 15 selected features (Figure 4B). Two representative individual-level SHAP decision profiles that

demonstrate how the model forms clinically consistent predictions were presented in Figure S2.

3.4 | Exploratory Analysis of Prognostic Stratification After TA

In the absence of pathological MVI confirmation in the TA cohort, we investigated the association between the model-predicted MVI risk score and RFS as an exploratory analysis of its prognostic utility. During the follow-up period until March 2025, 53.6% (75/140) of the patients experienced recurrence after TA, with a median follow-up time of 4.9 years. The 1-, 2- and 3-year RFS rates for patients undergoing TA were 82%, 63% and 54%, respectively. The MVI risk scores in the TA cohort ranged from 0.028 to 0.118, with a relatively uniform distribution and close to a normal distribution (Figure 5A). Box plot showed that the MVI risk score of the high-risk MVI group was significantly higher than those of the low-risk MVI group ($p < 0.05$, Figure 5B), and showed a negative linear correlation with time to RFS ($r = -0.28$, $p < 0.05$, Figure 5C). Among the 140 patients in the TA cohort, 107 (76.4%) underwent RFA and 33 (23.6%) underwent microwave ablation (MWA). The model showed similar MVI risk scores in different subgroups of ablation methods (Figure S3). Using the minimum distance method, the TA cohort was divided into the MVI high-risk ($n = 75$) and the MVI low-risk groups ($n = 65$). The MVI high-risk group exhibited significantly lower RFS rates (1-year: 75.5%; 2-year: 54.9%; 3-year: 39.4%) compared to the low-risk group (1-year: 83.8%; 2-year: 80.0%; 3-year: 78.0%), with statistically significant intergroup differences ($p < 0.001$, Figure 3D). Multivariate Cox analysis showed that the MVI risk score (HR = 2.90, 95% CI: 1.69–4.96, $p < 0.001$), AFP-L3% (HR = 1.34, 95% CI: 1.06–1.69, $p = 0.016$) were independent risk factors for RFS rate (Table 3).

4 | Discussion

In this study, we innovatively developed a high-precision MVI risk prediction model integrating multimodal CEUS and clinical characteristics. Utilizing a Vision Transformer (ViT-B/32) architecture, we extracted the global DL features from CEUS and finally identified the optimal predictive model through eight ML classifiers. Subsequently, we validated these models in both internal and external validation cohorts to assess its universality

TABLE 2 | The average performance of the 8 combination models in the internal validation cohort.

Models	AUC	ACC	SEN	SPE	PRE	PPV	NPV	F1 score
Vit-NN	0.54 ± 0.06	0.73 ± 0.01	0.26 ± 0.09	0.90 ± 0.04	0.45 ± 0.07	0.45 ± 0.07	0.77 ± 0.02	0.30 ± 0.08
Vit-SVM	0.73 ± 0.03	0.76 ± 0.02	0.46 ± 0.07	0.87 ± 0.02	0.55 ± 0.05	0.55 ± 0.05	0.82 ± 0.02	0.50 ± 0.06
Vit-LR	0.74 ± 0.04	0.70 ± 0.02	0.64 ± 0.06	0.72 ± 0.04	0.45 ± 0.03	0.45 ± 0.03	0.85 ± 0.02	0.53 ± 0.02
Vit-GB	0.84 ± 0.03	0.82 ± 0.02	0.48 ± 0.05	0.95 ± 0.03	0.78 ± 0.10	0.78 ± 0.10	0.84 ± 0.01	0.59 ± 0.04
Vit-RF	0.84 ± 0.03	0.82 ± 0.02	0.41 ± 0.02	0.97 ± 0.02	0.84 ± 0.11	0.84 ± 0.11	0.82 ± 0.01	0.55 ± 0.04
Vit-CB	0.86 ± 0.03	0.85 ± 0.02	0.55 ± 0.08	0.96 ± 0.02	0.86 ± 0.08	0.86 ± 0.08	0.86 ± 0.02	0.66 ± 0.06
Vit-LGBM	0.87 ± 0.04	0.85 ± 0.03	0.47 ± 0.10	0.99 ± 0.02	0.96 ± 0.08	0.96 ± 0.08	0.84 ± 0.02	0.62 ± 0.09
Vit-XGB	0.89 ± 0.04	0.85 ± 0.02	0.58 ± 0.08	0.95 ± 0.02	0.80 ± 0.07	0.80 ± 0.07	0.86 ± 0.02	0.66 ± 0.06

Note: Data are means ± SD.

Abbreviations: ACC, accuracy; AUC, area under the curve; CB, categorical boosting; GB, gradient boosting; LGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; NPV, negative predictive value; PPV, positive predictive value; PRE, precision; RF, random forest; SEN, sensitivity; SPE, specificity; SVM, support vector machine; XGB, eXtreme gradient boosting.

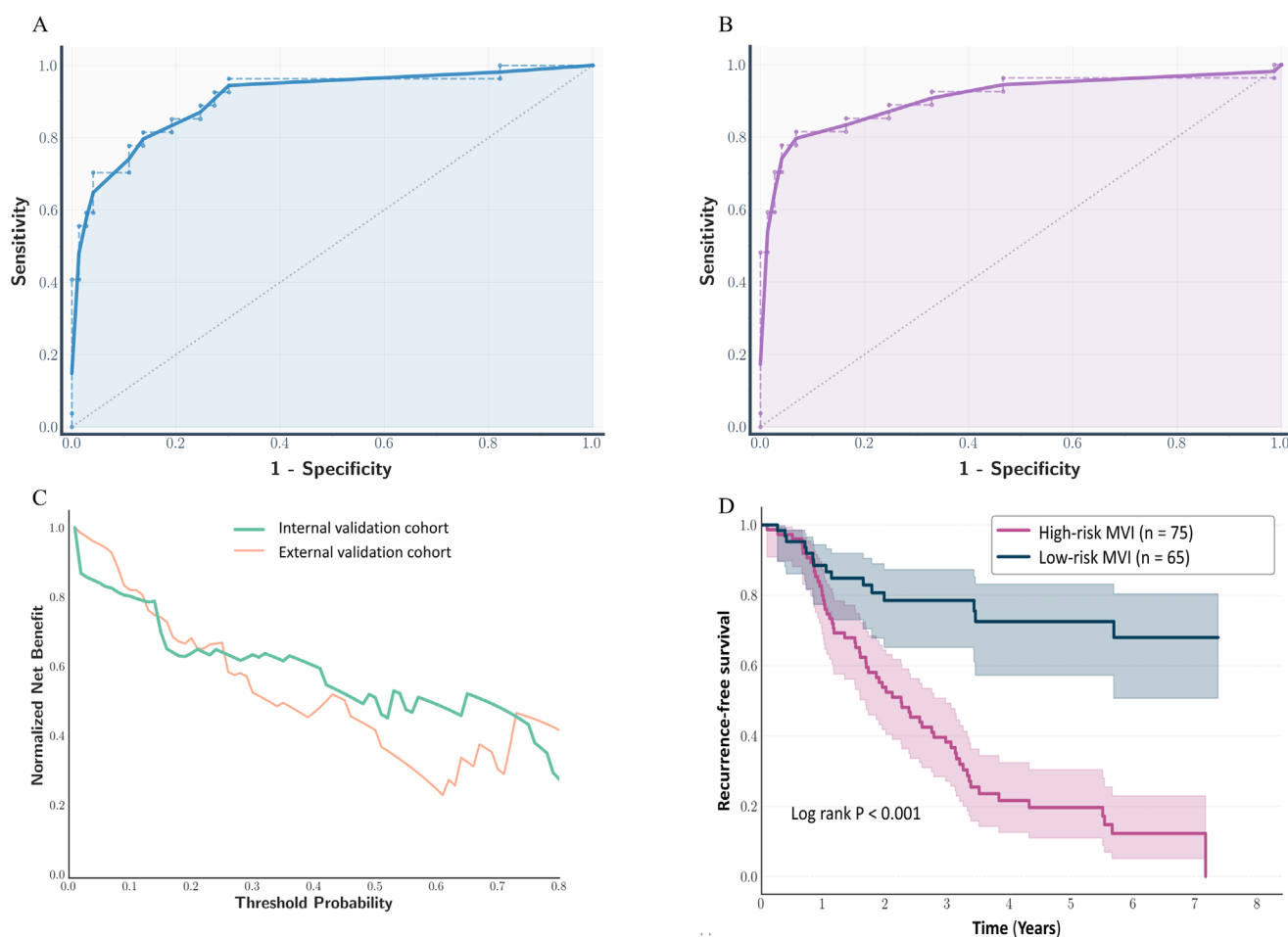


FIGURE 3 | Model performance and stratified RFS after TA. (A) ROC curve of the ViT-XGB model in the internal validation cohort; (B) ROC curve of the ViT-XGB model in the external validation cohort; (C) DCA curves of the ViT-XGB model in internal and external validation cohorts; (D) Comparison of RFS between the high-risk and low-risk groups of MVI after TA. The 95% CI bands were presented as blue and pink shaded areas on the Kaplan–Meier survival curve.

across different patient populations. Our research indicated that the optimal model (ViT-XGB model) exhibited excellent MVI prediction capabilities in both internal and external validation

cohorts, achieving AUC of 0.89 and 0.81, as well as ACC of 0.85 and 0.78. Furthermore, we applied the predictive model to the TA dataset where MVI results were routinely unavailable. The

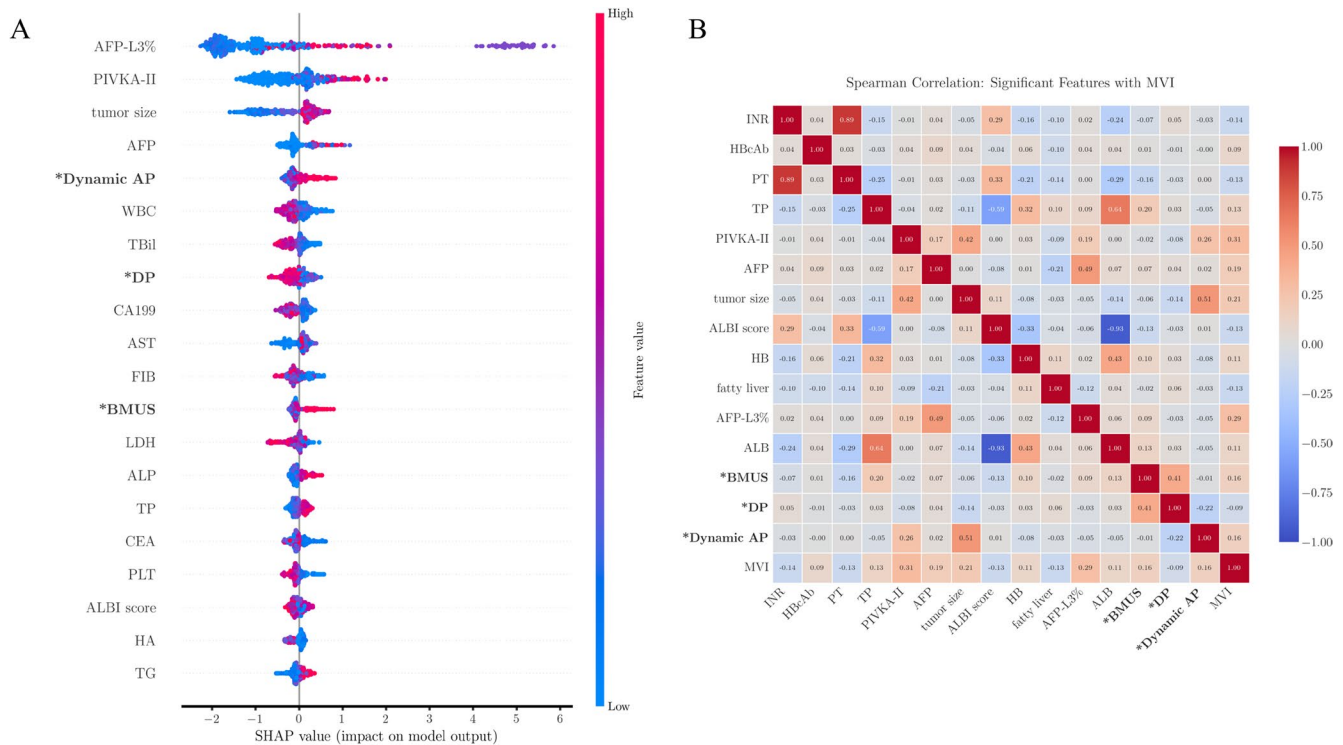


FIGURE 4 | Interpretation of the ViT-XGB model for MVI prediction. (A) SHAP plot of the ViT-XGB model; (B) Spearman correlation heatmap.

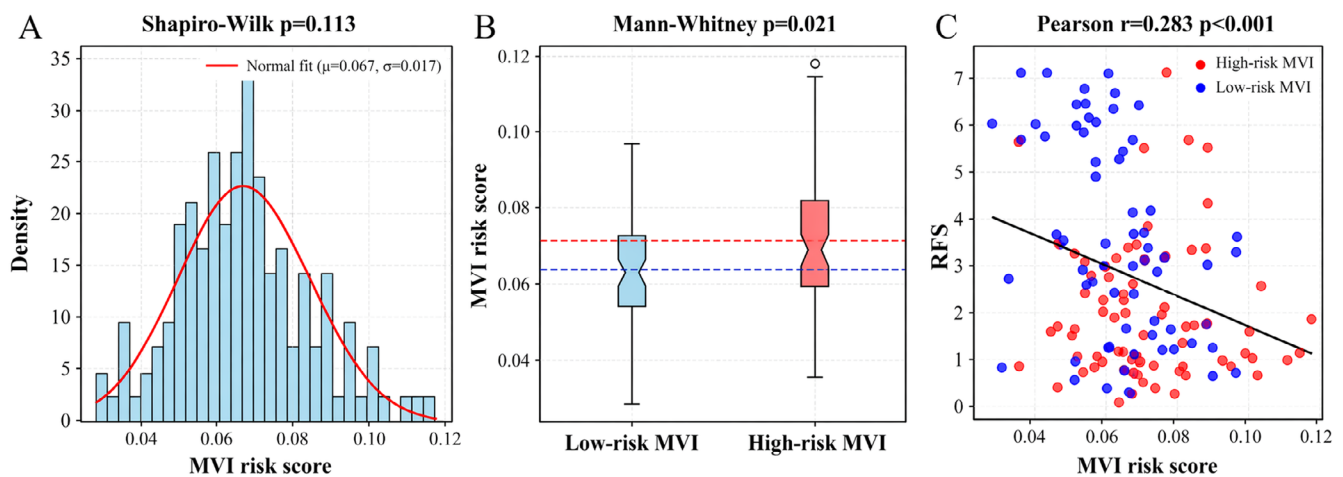


FIGURE 5 | Evaluation of MVI risk in the TA cohort using the ViT-XGB model; (A) Distribution characteristics of MVI risk score; (B) Comparison of MVI risk scores between groups; (C) Correlation between MVI risk score and RFS.

results showed that the RFS in the MVI high-risk group was significantly lower than that in the MVI low-risk group, suggesting that the model-predicted MVI risk score was associated with RFS after TA.

According to the Chinese guidelines for the diagnosis and treatment of primary liver cancer, radical SR and ablation are recommended as first-line treatment options for patients diagnosed with CNLC Ia stage HCC. TA serves as a critical alternative therapy, particularly for patients deemed unsuitable for surgery [25]. However, the current choice between SR and ablation mainly depends on the preoperatively accessible characteristics including tumour size, tumour number, tumour location, liver function and overall health status. Numerous studies have shown that

MVI is an important histological marker of HCC aggressiveness and is closely related to the prognosis [26, 27]. Yamashita et al. demonstrated significantly superior disease-free survival (DFS) and overall survival (OS) with hepatic resection, especially anatomical resection, compared to RFA in solitary HCC ≤ 3 cm patients with preoperative prediction of high MVI risk [28]. A recent double-centre study on HCC patients with MVI showed that adjuvant transarterial chemoembolization (TACE) plus lenvatinib could significantly improve survival outcomes [29]. Chen et al. reported that adjuvant TACE combined with targeted therapy or targeted immunotherapy could significantly enhance the DFS of MVI-positive HCC patients [30]. Therefore, developing a non-invasive preoperative approach for predicting MVI is imperative to inform prognostic stratification and

TABLE 3 | Cox regression univariate and multivariate analyses of RFS after TA.

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	<i>p</i>	HR (95% CI)	<i>p</i>
Age	1.10 (0.88–1.38)	0.391	NA	NA
MVI high-risk	2.87 (1.77–4.66)	<0.001	2.90 (1.69–4.96)	<0.001
Tumour size	1.32 (1.07–1.61)	0.008	1.30 (1.00–1.69)	0.053
AFP-L3	1.22 (1.00–1.50)	0.049	1.34 (1.06–1.69)	0.016
AFP	0.78 (0.57–1.07)	0.121	NA	NA
CA199	0.98 (0.77–1.24)	0.846	NA	NA
PIVKA-II	1.01 (0.83–1.23)	0.902	NA	NA
AST	0.96 (0.78–1.18)	0.669	NA	NA
TBil	1.35 (1.06–1.71)	0.016	1.21 (0.87–1.69)	0.258
ALP	1.23 (1.02–1.48)	0.027	1.13 (0.86–1.48)	0.390
LDH	1.17 (0.99–1.38)	0.067	NA	NA
TP	0.84 (0.68–1.04)	0.101	NA	NA
SCR	1.01 (0.81–1.26)	0.928	NA	NA
TG	0.92 (0.71–1.18)	0.502	NA	NA
HA	1.15 (0.96–1.38)	0.139	NA	NA
WBC	0.85 (0.67–1.07)	0.169	NA	NA
PLT	0.89 (0.71–1.11)	0.307	NA	NA
ALBI score	1.38 (1.11–1.71)	0.004	1.39 (0.86–2.24)	0.18

Note: Bold values are variables with $p < 0.05$.

Abbreviations: AFP, alpha-fetoprotein; AFP-L3%, alpha-fetoprotein heterogeneity; ALBI, albumin-bilirubin; ALP, alkaline phosphatase; AST, aspartate aminotransferase; CA199, carbohydrate antigen 19–9; HA, hyaluronic acid; LDH, lactic dehydrogenase; MVI, microvascular invasion; PIVKA-II, protein induced by vitamin K absence-II; PLT, platelet; SCR, serum creatinine; TBil, total bilirubin; TG, triglyceride; TP, total protein; WBC, white blood cell count.

facilitate early therapeutic planning. This holds particular clinical significance for ablation candidates, whose MVI status is unavailable prior to treatment.

While DL models for MVI prediction have been increasingly reported, CEUS-based approaches remain scarce, largely due to limited sample sizes hindering deep neural network training. Our study addresses these limitations by including a substantial patient cohort. To extract high-level semantic features from CEUS images, we employed the pretrained Vision Transformer (ViT-B/32) model from the Contrastive Language-Image Pretraining (CLIP) model. This model has been trained on large-scale image-text alignment data and possesses strong general image comprehension capabilities. In this study, we used average pooling operations to compress redundant information and combined an autoencoder to generate compact low-dimensional representations. This approach mitigated the challenges of noise sensitivity and dynamic sequence spatiotemporal modelling of ultrasound images. Moreover, the unique self-attention mechanism of the ViT architecture can effectively integrate complex spatial features and contextual information within and surrounding tumours, enabling the extraction of rich global features from multimodal ultrasound images. The XGB algorithm based on decision tree ensemble is well-suited for combining these image embeddings with structured clinical variables. Through

ensemble learning and rigorous regularization, we effectively reduced bias and variance, resulting in higher accuracy and robustness and enhancing the generalization ability of the model. Compared to purely end-to-end DL models, this complementary integration enables more effective modelling of heterogeneous information, which explains the observed performance gain.

Existing DL models for MVI assessment mainly relied on imaging-based training. For instance, Wang et al. used the Swin Transformer network to extract multiphasic MRI features to predict MVI and enhance the prognostic stratification of HCC after RFA [31]. Qin et al. demonstrated the effectiveness of a CEUS-based DL model in predicting MVI [23]. However, considering that certain clinical features are closely related to MVI, our work focused on developing a combined model of clinical and multimodal imaging features for MVI prediction. Our model integrates multi-phase CEUS with a comprehensive clinical feature set that included AFP-L3% and PIVKA-II. We found that combining CEUS with clinical variables significantly improved model performance. This confirmed that although CEUS-only prediction was feasible, optimal risk stratification required integrating imaging with clinical context—highlighting why clinical variables remain indispensable for reliable decision support. Furthermore, in contrast to the previous CEUS-based DL model developed by Zhang et al. that relied only on static

arterial or portal-phase frames, our study incorporates dynamic arterial-phase sequences using TIC-guided keyframe extraction [22]. This allows the model to learn microvascular perfusion patterns that cannot be captured by static-image frameworks. Moreover, we directly compared our model with several representative non-DL MVI prediction models reported in previous studies. The results suggested that the added complexity of DL was justified when accurate preoperative MVI risk stratification is required, as our model may capture subtle features beyond conventional clinical or radiological assessments. However, we acknowledge that simpler models may remain adequate in resource-limited settings, whereas our approach offers added value in centres capable of standardized CEUS.

Prior CEUS studies typically present black-box models; we provide a more transparent interpretability framework by combining global SHAP rankings, individual-level decision explanations, and collinearity analyses [20]. These results consistently highlight biologically plausible contributors such as AFP-L3%, PIVKA-II, tumour size, AFP and arterial-phase CEUS patterns, which are consistent with previous studies. Pote et al. found that serum PIVKA-II > 90 mAU/mL or AFP > 100 IU/mL were independent risk factors for MVI [32]. Araki et al. reported that AFP-L3%, PIVKA-II and metabolic tumour volume (MTV) were independent predictors of MVI in HCC ≤ 3 cm [33]. And tumour size is a standard clinical indicator of HCC progression and tumour burden [34]. Our model fully considered these factors and had higher reliability and clinical applicability in predicting MVI. DCA further substantiated that the model may have valuable clinical utility. Beyond the association with MVI confirmed by prior studies, our research further quantified the role of AFP-L3 and PIVKA-II within the multimodal prediction model. We found that excluding these two biomarkers led to a significant decline in the model's discriminative ability, with the AUC dropping from 0.89 to 0.71. This indicated that they provided important and non-redundant predictive information. AFP-L3 is a key isoform of AFP that is specific to HCC. PIVKA-II is produced due to impaired synthesis and incomplete carboxylation of prothrombin precursor during hepatocarcinogenesis. Both are likely to provide valuable insights into tumour invasiveness and molecular-level biological activity. Whereas, CEUS enables the visualization of hemodynamic patterns and tumour heterogeneity at the imaging level. The complementary combination of these features may produce a synergistic effect, helping to more comprehensively reflect the characteristics of the tumour and more effectively assess the risk of MVI. However, it is important to recognize that AFP-L3% and PIVKA-II are not routinely available in all medical centres. Although these biomarkers significantly enhance model performance, their limited accessibility must be acknowledged. Therefore, in resource-constrained environments, the simplified model may serve as a pragmatic and reasonable alternative, whereas the full model is expected to achieve optimal performance in specialized centres where such biomarker testing is accessible. We conducted an exploratory analysis by applying the optimal model to the TA cohort to assess its prognostic stratification value. A significant difference in RFS was observed between the different MVI risk groups predicted by this model. We also evaluated the factors related to RFS after TA. Our study indicated that high-risk MVI and AFP-L3% were independent predictors of RFS. These findings highlighted the exploratory prognostic value of the DL model in

patients undergoing TA, suggesting its potential as a risk stratification tool for recurrence. However, these should be considered hypothesis-generating and warrant prospective validation. Additionally, we subdivided the ablation cohort into RFA and MWA subgroups for further analysis. Our subgroup analysis did not show a significant difference in preoperative MVI risk distribution between the two groups. Nevertheless, larger prospective datasets are still required to further investigate potential interactions between different ablation modalities and risk scores.

This study had several limitations that warrant acknowledgment. First, a major limitation was that while our model was developed and validated for MVI prediction using SR patients with pathological confirmation, its diagnostic accuracy in TA patients remains unvalidated. Since TA destroys tumour tissue, pathological MVI verification is unavailable in this cohort. Therefore, the correlation between model-predicted MVI risk and RFS, while clinically interesting, represented exploratory prognostic analysis rather than diagnostic validation. Larger-scale prospective data are needed to further validate our model. Second, our model's strong performance is substantially dependent on AFP-L3% and PIVKA-II biomarkers, which are not routinely measured in many clinical settings globally. Performance without these markers (AUC = 0.71) is considerably reduced, limiting generalizability to centres where these tests are unavailable. Future work should explore whether alternative biomarker combinations could maintain performance while improving accessibility. Third, with approximately 76% of enrolled patients having HBV-related aetiology, the model's applicability in patients with other causes requires further verification. Fourth, the small sample size of the external validation cohort is a significant constraint. Future larger multicentre cohorts, particularly those including routine CEUS acquisition, will be crucial for confirming the model's stability and generalizability. Fifth, we did not further classify MVI-positive patients into M1 and M2 categories. Although the proportion of M2 patients is small, identifying these high-risk patients preoperatively could further aid in formulating treatment strategies. Future research requires accumulating a sufficient number of M2 cases to develop and validate models capable of predicting MVI severity grading. Sixth, this study utilized images captured by ultrasound devices from three major manufacturers. Although all examinations followed a uniform protocol, including patient preparation, scanning planes, dynamic contrast imaging, and the use of ultrasound contrast agent (SonoVue), the use of multiple devices may have introduced potential heterogeneity. Future studies could further control the impact of device heterogeneity by conducting cross-device comparison studies in the same patient population or by developing advanced image standardization algorithms. Finally, due to the lack of follow-up data for some patients, the stratified role of the model in the long-term survival of patients after TA was not further investigated.

In summary, this study constructed an MVI prediction model for patients with CNLC Ia-stage HCC based on DL features from CEUS data and clinical characteristics. The model demonstrated promising performance in surgical cohorts where pathological confirmation was available. However, the generalizability of its performance and its ultimate clinical utility require validation through larger-scale studies. Furthermore, the

model demonstrated exploratory prognostic value in patients undergoing TA, suggesting its potential utility as an adjunct tool for recurrence risk assessment, but this requires prospective validation.

Author Contributions

Conceptualization: Wentao Kong and Wenping Wang; Data curation: Keke Chen, Tianjiao Huang, Yadan Xu, Qi Zhang and Yuli Zhu; Data analysis: Keke Chen, Shukang Zhang, Peng Wan, Wentao Kong; Administrative support: Wentao Kong and Wenping Wang. Manuscript drafting: Keke Chen, Shukang Zhang. Manuscript review: Peng Wan, Wentao Kong, Wenping Wang. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors have nothing to report.

Funding

This work was supported by National Natural Science Foundation of China (Grant No. 82472004 and No. 82272013), Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (2022-LCYJ-MS-24), Social Development Foundation of Zhenjiang (SH2025068) and Departmental Research Fund from Affiliated People's Hospital of Jiangsu University (KFB2025007).

Ethics Statement

This study was approved by the ethics committees (No. 2022-140-01 and B2022-223R), and was conducted in accordance with the Declaration of Helsinki. Given the retrospective observational nature of the study without involving any direct patient contact or intervention, the requirement for informed consent has been waived.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. H. Sung, J. Ferlay, R. L. Siegel, et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians* 71, no. 3 (2021): 209–249.
2. A. G. Singal, J. M. Llovet, M. Yarrow, et al., "AASLD Practice Guidance on Prevention, Diagnosis, and Treatment of Hepatocellular Carcinoma," *Hepatology* 78, no. 6 (2023): 1922–1965.
3. A. Vogel, S. L. Chan, L. A. Dawson, et al., "Hepatocellular carcinoma: ESMO clinical practice guideline for diagnosis, treatment and follow-up," *Annals of Oncology* 36, no. 5 (2025): 491–506.
4. W. M. Cong, H. Bu, J. Chen, et al., "Practice Guidelines for the Pathological Diagnosis of Primary Liver Cancer: 2015 Update," *World Journal of Gastroenterology* 22, no. 42 (2016): 9279–9287.
5. Z. Zheng, R. Guan, W. Jianxi, et al., "Microvascular Invasion in Hepatocellular Carcinoma: A Review of Its Definition, Clinical Significance, and Comprehensive Management," *Journal of Oncology* 2022 (2022): 9567041.

6. X. Sheng, Y. Ji, G. P. Ren, et al., "A Standardized Pathological Proposal for Evaluating Microvascular Invasion of Hepatocellular Carcinoma: A Multicenter Study by LCPGC," *Hepatology International* 14, no. 6 (2020): 1034–1047.
7. S. Lee, T. W. Kang, K. D. Song, et al., "Effect of Microvascular Invasion Risk on Early Recurrence of Hepatocellular Carcinoma After Surgery and Radiofrequency Ablation," *Annals of Surgery* 273, no. 3 (2021): 564–571.
8. J. Ouyang, Y. Yang, Y. Zhou, et al., "Adjuvant Camrelizumab Plus Apatinib in Resected Hepatocellular Carcinoma With Microvascular Invasion: A Multi-Center Real World Study," *Hepatobiliary Surgery and Nutrition* 13, no. 4 (2024): 616–631.
9. J. Qian, Y. Shen, L. Cui, et al., "Survival Effects of Postoperative Adjuvant TACE in Early-HCC Patients With Microvascular Invasion: A Multicenter Propensity Score Matching," *Journal of Cancer* 15, no. 1 (2024): 68–78.
10. Y. Sun, H. Yang, S. Li, et al., "An Accurate Model for Microvascular Invasion Prediction in Solitary Hepatocellular Carcinoma ≤ 5 Cm Based on CEUS and EOB-MRI: A Retrospective Study With External Validation," *Academic Radiology* S1076-6332, no. 25 (2025): 361–367.
11. J. Li, X. Su, X. Xu, et al., "Preoperative Prediction and Risk Assessment of Microvascular Invasion in Hepatocellular Carcinoma," *Critical Reviews in Oncology/Hematology* 190 (2023): 104107.
12. N. Fujita, Y. Ushijima, K. Ishimatsu, et al., "Multiparametric Assessment of Microvascular Invasion in Hepatocellular Carcinoma Using Gadoteric Acid-Enhanced MRI," *Abdominal Radiology (New York)* 49, no. 5 (2024): 1467–1478.
13. H. Jiang, B. Li, T. Zheng, et al., "MRI-Based Prediction of Microvascular Invasion/High Tumor Grade and Adjuvant Therapy Benefit for Solitary HCC ≤ 5 Cm: A Multicenter Cohort Study," *European Radiology* 35, no. 6 (2025): 3223–3237.
14. G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (2017): 60–88.
15. T. Zheng, Y. Zhu, H. Jiang, et al., "MRI-Based Topology Deep Learning Model for Noninvasive Prediction of Microvascular Invasion and Assisting Prognostic Stratification in HCC," *Liver International* 45, no. 3 (2025): e16205.
16. Z. Huang, W. Huang, L. Jiang, et al., "Decision Fusion Model for Predicting Microvascular Invasion in Hepatocellular Carcinoma Based on Multi-MR Habitat Imaging and Machine-Learning Classifiers," *Academic Radiology* 32, no. 4 (2025): 1971–1980.
17. G. Miao, X. Qian, Y. Zhang, et al., "An MRI-Based Radiomics Model for Preoperative Prediction of Microvascular Invasion and Outcome in Intrahepatic Cholangiocarcinoma," *European Journal of Radiology* 183 (2025): 111896.
18. X. Dong, X. Jia, W. Zhang, et al., "Interpretable and Generalizable Deep Learning Model for Preoperative Assessment of Microvascular Invasion and Outcome in Hepatocellular Carcinoma Based on MRI: A Multicenter Study," *Insights Into Imaging* 16, no. 1 (2025): 151.
19. H. Ma, L. Wang, L. Sun, et al., "Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma From Multi-Sequence Magnetic Resonance Imaging Based on Deep Fusion Representation Learning," *IEEE Journal of Biomedical and Health Informatics* 29, no. 5 (2025): 3259–3271.
20. Q. Qin, J. Pang, J. Li, et al., "Transformer Model Based on Sonazoid Contrast-Enhanced Ultrasound for Microvascular Invasion Prediction in Hepatocellular Carcinoma," *Medical Physics* 52, no. 7 (2025): e17895.
21. Y. Wang, W. Xie, C. Li, et al., "Automated Microvascular Invasion Prediction of Hepatocellular Carcinoma via Deep Relation Reasoning From Dynamic Contrast-Enhanced Ultrasound," *Computerized Medical Imaging and Graphics* 124 (2025): 102606.

22. W. Zhang, Q. Guo, Y. Zhu, et al., "Cross-Institutional Evaluation of Deep Learning and Radiomics Models in Predicting Microvascular Invasion in Hepatocellular Carcinoma: Validity, Robustness, and Ultrasound Modality Efficacy Comparison," *Cancer Imaging* 24, no. 1 (2024): 142.
23. X. Qin, J. Zhu, Z. Tu, Q. Ma, J. Tang, and C. Zhang, "Contrast-Enhanced Ultrasound With Deep Learning With Attention Mechanisms for Predicting Microvascular Invasion in Single Hepatocellular Carcinoma," *Academic Radiology* 30, no. Suppl 1 (2023): S73–S80.
24. Y. Zhang, Q. Wei, Y. Huang, et al., "Deep Learning of Liver Contrast-Enhanced Ultrasound to Predict Microvascular Invasion and Prognosis in Hepatocellular Carcinoma," *Frontiers in Oncology* 12 (2022): 878061.
25. J. Zhou, H. Sun, Z. Wang, et al., "Guidelines for the Diagnosis and Treatment of Primary Liver Cancer (2022 Edition)," *Liver Cancer* 12, no. 5 (2023): 405–444.
26. S. Yang, H. Ni, A. Zhang, et al., "Grading Severity of MVI Impacts Long-Term Outcomes After Laparoscopic Liver Resection for Early-Stage Hepatocellular Carcinoma: A Multicenter Study," *American Journal of Surgery* 238 (2024): 115988.
27. C. Fuster-Anglada, E. Mauro, J. Ferrer-Fàbrega, et al., "Histological Predictors of Aggressive Recurrence of Hepatocellular Carcinoma After Liver Resection," *Journal of Hepatology* 81, no. 6 (2024): 995–1004.
28. Y. I. Yamashita, K. Imai, T. Yusa, et al., "Microvascular Invasion of Single Small Hepatocellular Carcinoma ≤ 3 Cm: Predictors and Optimal Treatments," *Annals of Gastroenterological Surgery* 2, no. 3 (2018): 197–203.
29. Y. C. Hou, J. K. Feng, K. Wang, et al., "Adjuvant Transarterial Chemoembolization Plus Lenvatinib for Patients With HCC With MVI After Resection: A Multicenter Retrospective Study," *Oncologist* 30, no. 6 (2025): oyaf139.
30. X. Chen, X. Wu, W. Peng, et al., "Combined TACE With Targeted and Immunotherapy Versus TACE Alone Improves DFS in HCC With MVI: A Multicenter Propensity Score Matching Study," *Journal of Hepatocellular Carcinoma* 12 (2025): 561–577.
31. W. Wang, Y. Wang, D. Song, et al., "A Transformer-Based Microvascular Invasion Classifier Enhances Prognostic Stratification in HCC Following Radiofrequency Ablation," *Liver International* 44, no. 4 (2024): 894–906.
32. N. Poté, F. Cauchy, M. Albuquerque, et al., "Performance of PIVKA-II for Early Hepatocellular Carcinoma Diagnosis and Prediction of Microvascular Invasion," *Journal of Hepatology* 62, no. 4 (2015): 848–854.
33. R. Fukushima, N. Harimoto, T. Okuyama, et al., "New Predictors of Microvascular Invasion for Small Hepatocellular Carcinoma ≤ 3 cm," *International Journal of Clinical Oncology* 29, no. 8 (2024): 1182–1190.
34. G. B. Levi Sandri, G. Spoletini, G. Vennarecci, E. Francone, M. Abu Hilal, and G. M. Ettorre, "Laparoscopic Liver Resection for Large HCC: Short- and Long-Term Outcomes in Relation to Tumor Size," *Surgical Endoscopy* 32, no. 12 (2018): 4772–4779.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Percentage of missing data in each group. **Table S2:** Pairwise comparison of different imputation strategies. **Table S3:** Performance comparison of different imbalance-handling strategies, including random undersampling, class-weighted training, and SMOTE. **Table S4:** The performance of the combined model in the internal validation cohort. **Table S5:** The performance of the combined model in the external validation cohort. **Table S6:** The average performance of the 8 combination models in the external validation cohort. **Table S7:** Comparison of performance among several MVI prediction models. **Table S8:** Performance comparison of imaging-only classifiers using BMUS, CEUS, and BMUS+CEUS modalities under the

ViT-XGB framework. **Table S9:** Performance comparison of end-to-end deep models and hybrid XGB-based models using different visual encoders. **Table S10:** Performance comparison of ViT-XGB models without (Model 1) versus with (Model 2) AFP-L3% and PIVKA-II. **Figure S1:** SHAP value comparison for ViT-XGB models without (Panel A) and with (Panel B) AFP-L3% and PIVKA-II, illustrating their additional discriminative contribution. **Figure S2:** Individual-level SHAP decision plots illustrating how clinical and imaging features contribute to model predictions for representative low-risk and high-risk cases. (A) A low-risk case in which markedly low AFP-L3%, low PIVKA-II, and low-risk CEUS enhancement patterns during arterial-phase generate strong negative SHAP values, collectively driving the prediction toward a low probability of MVI; (B) A high-risk case where elevated AFP-L3%, high PIVKA-II, larger tumour size, and high-risk CEUS/BMUS patterns yield substantial positive SHAP values, pushing the prediction toward high MVI likelihood. **Figure S3:** Comparison of predicted MVI risk scores between radiofrequency ablation (RFA) and microwave ablation (MWA). (Left panel) Violin-box plots showing no significant difference in MVI risk scores between the two modalities ($p = 0.161$); (Right panel) Cumulative distribution curves further confirm the similar riskscore distributions for RFA and MWA.