

CT-Based 2.5D Deep Learning-Multi-Instance Learning for Predicting Early Recurrence of Hepatocellular Carcinoma and Correlating with Recurrence-Related Pathological Indicators

Yongyi Cen^{1,2,*}, Haiyang Nong^{1,2,*}, Dehui Du^{3,*}, Yingning Wu^{1,2}, Jianpeng Chen^{1,2}, Zhaolin Pan⁴, Yin Huang⁵, Ke Ding⁶, Deyou Huang^{1,2}

¹Department of Radiology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, Guangxi Zhuang Autonomous Region, 533000, People's Republic of China; ²Guangxi Clinical Medical Research Center for Hepatobiliary Diseases, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, Guangxi Zhuang Autonomous Region, 533000, People's Republic of China; ³Department of Radiology, The Affiliated Wuming Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, 530100, People's Republic of China; ⁴Department of Hepatobiliary Surgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, Guangxi Zhuang Autonomous Region, 533000, People's Republic of China; ⁵Pathology Department, Affiliated Hospital of Youjiang University of Nationalities, Baise, Guangxi Zhuang Autonomous Region, 533000, People's Republic of China; ⁶Department of Radiology, The Third Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, 530031, People's Republic of China

*These authors contributed equally to this work

Correspondence: Deyou Huang, Department of Radiology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, Guangxi Zhuang Autonomous Region, 533000, People's Republic of China, Email FZXYH2012@126.com; Ke Ding, Department of Radiology, The Third Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, 530031, People's Republic of China, Email 272480365@qq.com

Purpose: This study aims to evaluate the advantages of the 2.5D deep learning-multi-instance learning (2.5D DL-MIL) model, based on CT arterial phase images, in predicting early recurrence (ER) of hepatocellular carcinoma (HCC) and examining the biological significance of MIL features.

Patients and Methods: A total of 191 HCC patients were retrospectively included and categorized into ER (n=79) and non-early recurrence (NER, n=112) groups based on postoperative follow-up results. The patients were randomly divided to the training set (n=133) and validation set (n=58) in a 7:3 ratio. The predictive capabilities of the 2.5D DL-MIL model, Radiomics model, and Clinical model for ER of HCC were constructed and compared using CT arterial phase and clinical data. SHAP analysis was used to evaluate the contribution of MIL features in the model, and further analysis was conducted on the correlation between MIL features and microvascular invasion (MVI), Ki-67 expression, and pathological grading.

Results: The area under the curve (AUC) for the 2.5D DL-MIL model in the validation set was 0.840, surpassing that of the Radiomics model (AUC = 0.678, P = 0.047) and the Clinical model (AUC = 0.598, P = 0.009). Decision curve analyses indicated superior clinical utility for the 2.5D DL-MIL model. SHAP analysis revealed that bag-of-words features (eg, BoW_02 and BoW_09) were key contributors to the 2.5D DL-MIL model. Correlation analysis demonstrated that BoW_01, BoW_02, BoW_09, and BoW_1 were significantly correlated with MVI grade and Ki-67 expression (P < 0.05).

Conclusion: The 2.5D DL-MIL model demonstrates significant value in predicting ER of HCC, with its MIL features exhibiting strong associations with tumor invasiveness and proliferative activity.

Keywords: hepatocellular carcinoma, early recurrence, CT, deep learning, multi-instance learning

Introduction

Primary liver cancer is the sixth most common cancer worldwide and the third leading cause of cancer-related deaths. Among these, 75% to 85% are hepatocellular carcinoma (HCC). Postoperative recurrence is one of the key factors affecting



the prognosis of patients.^{1,2} The mechanisms underlying recurrence are closely associated with tumor invasiveness, including microvascular invasion (MVI),³ pathological grading,⁴ and proliferative activity, such as high Ki-67 expression.⁵ Although the Barcelona Clinic Liver Cancer (BCLC) staging system and serum tumor marker tests provide important information for HCC risk stratification, their predictive performance for postoperative recurrence risk is limited,^{6,7} and one possible reason could be their inability to fully capture the tumor's heterogeneity. Most existing HCC recurrence prediction models have failed to integrate the biological information contained in advanced imaging techniques, resulting in a lack of model interpretability. Therefore, developing a more precise and interpretable model for predicting early postoperative recurrence of HCC is crucial for optimizing clinical decisions and creating personalized follow-up strategies.

Traditional radiomics constructs predictive models by manually extracting features, such as texture and shape, from medical images. However, the predictive performance is often constrained due to the subjectivity and limitations inherent in manually designed features.⁸ Deep learning (DL) techniques demonstrate substantial advantages in medical image analysis by automatically learning multi-level features from images.⁹ Among these, the 2.5D deep learning model integrates the maximum cross-sectional tumor image and its adjacent multi-layer slice information, which more effectively captures the tumor's three-dimensional spatial features compared to traditional 2D models.^{10,11} Moreover, multi-instance learning (MIL) can effectively characterize tumor heterogeneity by aggregating prediction results from multiple slices.¹² However, there is limited research on the application of the 2.5D deep learning-multi-instance learning (2.5D DL-MIL) model in predicting early recurrence of HCC, and the correlation between extracted image features and tumor biological behavior requires further exploration.

This study aims to develop a 2.5D DL-MIL model using computed tomography (CT) arterial phase images and to compare its predictive performance for early postoperative recurrence of HCC with that of traditional radiomics (Radiomics features extracted based on manual delineation) and clinical models (Based on clinical indicators such as age, gender, and AFP). The decision-making mechanisms of the model were analyzed using Shapley Additive Explanations (SHAP), and the correlations between MIL features, microvascular invasion (MVI), Ki-67 expression, and pathological grading were explored to provide biological evidence supporting the clinical application of the model. The research process is illustrated in [Figure 1](#).

Materials and Methods

This study was approved by the Ethics Committee of the Affiliated Hospital of Youjiang Medical University for Nationalities (Approval No. YYFY-LL-2024-038) and conducted in accordance with the principles outlined in the Declaration of Helsinki for medical research involving human subjects. Given its retrospective design and the anonymization of all participant data, the ethics committee waived the requirement for obtaining written informed consent from the study participants.

Subjects

This retrospective study involved patients treated at the Affiliated Hospital of Youjiang Medical University for Nationalities between January 2019 and May 2024. Following systematic screening, a total of 191 HCC patients fulfilled the inclusion criteria. Based on postoperative follow-up, these patients were classified into two groups: early recurrence (ER, n=79) and non-early recurrence (NER, n=112). All patients were randomly assigned to a training set (n=133) and a validation set (n=58) in a 7:3 ratio.

Inclusion criteria: ① Patients who underwent radical resection and were confirmed as having HCC through postoperative pathology, with detailed clinical data and follow-up information; ② CT scans and relevant laboratory tests performed within one week prior to surgery. Exclusion criteria: ① Patients with artifacts in CT arterial phase images; ② Patients who had previously received radiotherapy, chemotherapy, interventional therapy, or other anti-tumor treatments (The inclusion and exclusion process is shown in [Figure 1](#)).

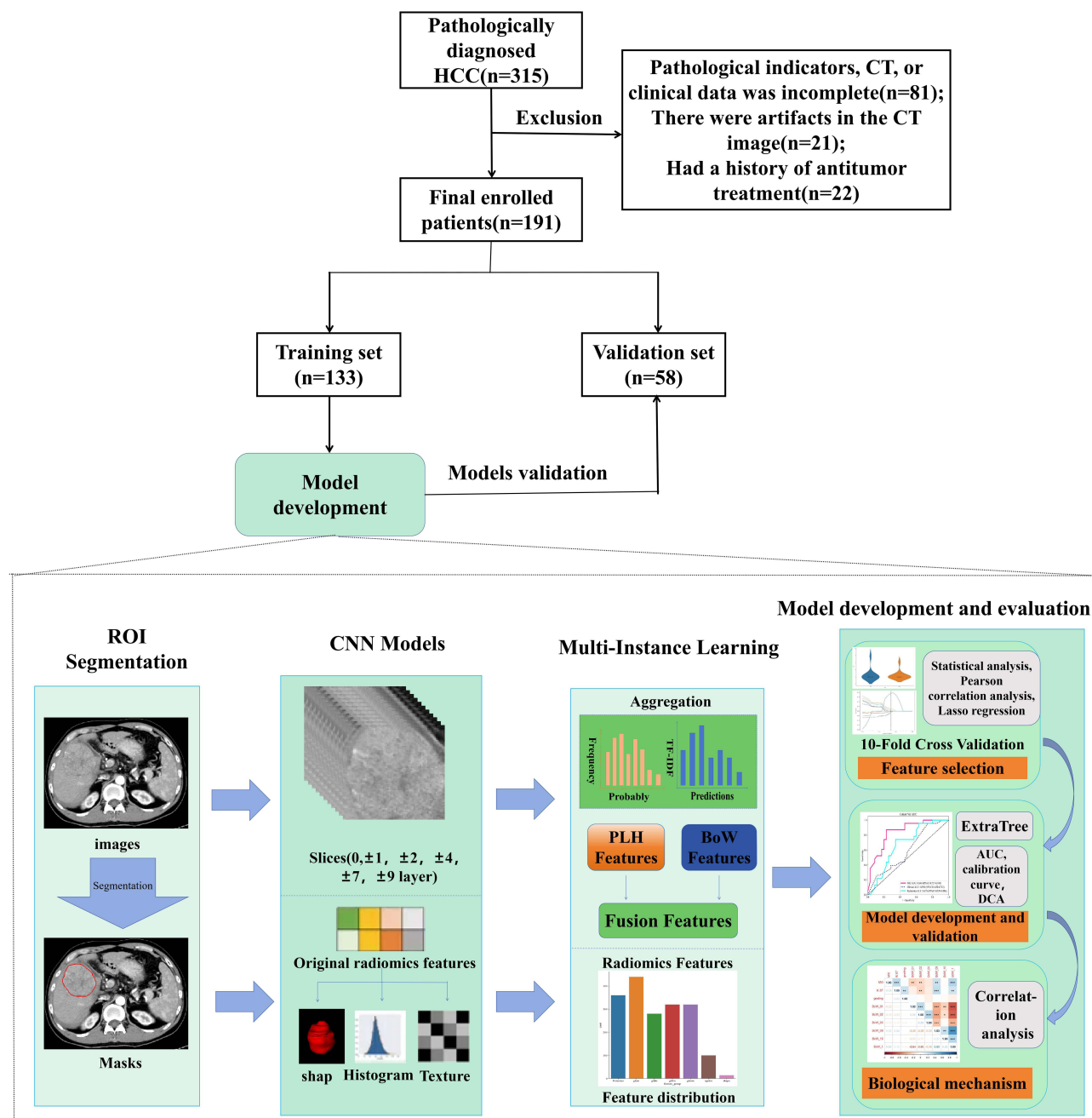


Figure 1 Study Population and Workflow.

Definition of Early Recurrence of HCC

In post-surgical follow-up for HCC, tumor recurrence is defined as the identification of a new tumor lesion in the liver or extrahepatic regions through imaging techniques such as ultrasound, CT, or MRI.¹³ Tumor recurrence diagnosis in all cases was assessed by one hepatobiliary surgeon and one radiologist, with both analyzing only the assigned clinical and imaging data while remaining unaware of other non-essential information. Based on follow-up results, cases with tumor recurrence within two years after surgery were classified into the ER group, whereas those without recurrence after two years were classified into the NER group.¹⁴

Clinical Features

This study implemented stringent protocols for data collection across all participants, with all relevant laboratory tests and CT scans performed within one week prior to surgery. The clinical indicators considered in this study were categorized into three main types. The first category comprises demographic baseline indicators, such as age, gender, and body mass index (BMI). The second category includes medical history-related indicators, including alcohol consumption history and hepatitis B surface antigen (HBsAg) test results. The third category comprises serological indicators encompassing four key biochemical parameters: alpha-fetoprotein (AFP), serum albumin (ALB), aspartate aminotransferase (AST), and alanine aminotransferase (ALT).

Pathological Indicators

This study incorporated three pathological indicators that are closely related to post-surgical recurrence of HCC, as identified in previous research: MVI,^{3,15} Ki-67,^{5,16} and pathological grading,^{4,17} in order to enhance the interpretability of the results. All pathological indicators were diagnosed by a certified pathologists in accordance with established procedures and guidelines.

MVI is defined as the presence of cancer cell nests within the lumen, which is lined by endothelial cells, as observed under a microscope. The grading criteria were as follows:¹⁸ M0 indicates no detected MVI; M1 (low-risk group) refers to ≤ 5 MVI lesions located within liver tissue near the tumor (distance to tumor ≤ 1 cm); M2 (high-risk group) is subdivided into two subtypes: M2a, with >5 MVI lesions near the tumor, and M2b, with MVI present in liver tissue distant from the tumor (distance >1 cm).

Ki-67 expression was detected following the conventional pathological process, where tissue samples were fixed with hematoxylin and eosin, embedded in paraffin, sectioned to a thickness of 3–5 μm , and then deparaffinized and hydrated. Immunohistochemical staining was performed using a mouse anti-human Ki-67 monoclonal antibody. Ki-67-positive cells exhibited brownish-yellow granules in the cell nucleus. The expression level was assessed by randomly selecting five high-power fields per slide for analysis. In each field, the number of Ki-67-positive cells among 100 tumor cells was counted, and the percentage of positive cells relative to the total number of tumor cells was calculated. The average value was used as the Ki-67 expression index.¹⁹

The HCC pathological grading followed the Edmondson-Steiner grading system.²⁰ Grade I tumor cells resemble normal liver cells and are arranged in thin trabeculae; Grade II cells exhibit increased cell volume and a higher nuclear-to-cytoplasm ratio, with darker nuclear staining and some atypia; Grade III cells demonstrate poor differentiation, further increased cell volume, nuclear-to-cytoplasm ratio, significant atypia, and frequent mitotic figures; Grade IV displays the poorest differentiation, scant cytoplasm, deeply stained nuclei, irregular cell morphology, poor adhesion, and loose arrangement.

CT Scanning Protocol

The range of the CT scan spanned from the apex of the diaphragm to the lower margin of the liver, with scanning parameters for each device provided in Table 1. The contrast-enhanced scan utilized a dual-syringe, high-pressure injector to administer iodixanol (Uvison 370) at a rate of 3.5 mL/s, with a dose calculated as 1.5 mL/kg, followed by a 40 mL saline flush at the same rate. A region of interest (ROI) was defined in the descending aorta, and a threshold-

Table 1 Scanning Parameters of Various CT Devices

| Devices | Revolution Aca (GE) | Ingenuity Core 64 (Philips) | Revolution (GE) |
|-----------------|---------------------|-----------------------------|-----------------|
| Slice thickness | 5mm | 5mm | 5mm |
| Slice spacing | 5mm | 5mm | 5mm |
| Tube voltage | 120KV | 120KV | 120KV |
| Tube current | 50mA | 30mA | 50mA |
| Matrix | 512×512 | 512×512 | 512×512 |
| ROI threshold | 100HU | 150HU | 120HU |

triggered scanning technique was employed. The arterial phase scan was initiated 8 seconds after the ROI CT value reached the preset threshold. All arterial phase contrast-enhanced CT images were exported in DICOM format and subsequently converted to NIfTI format for further image analysis.

CT Image Preprocessing and ROI Placement

To mitigate the influence of varying devices and scanning parameters on the results, the following standardization procedures were applied: First, the voxel spacing for all CT image data was standardized to $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$; second, windowing techniques were applied to standardize the window width and window level of the images (window width: 259 HU, window level: 40 HU). Tumor segmentation of CT arterial-phase images was conducted by two radiologists with over 5 years of abdominal imaging diagnostic experience using ITK-SNAP 3.8.0 software (www.itksnap.org). During the segmentation process, the radiologists remained blinded to the patients' recurrence status and manually delineated the tumor contours layer by layer (Figure 2).

MIL Feature Extraction Method

2.5D Data Acquisition

This study presents a 2.5D data generation method in which CT arterial phase image data are utilized. The central slice represents the maximum cross-section of the ROI, while adjacent slices are selected at layer intervals of ± 1 , ± 2 , ± 4 , ± 7 , and ± 9 , thus generating a dataset consisting of 11 CT slices per patient. The data were standardized and cropped using the OKT-crop_max_roi tool available on the OnekeyAI platform.

Slice-Level Model Training

In the slice-level model training, three DL architectures—ResNet18, ResNet101, and DenseNet121—were employed for training on 2.5D image slices. The model's performance was assessed using metrics, including accuracy, area under the curve (AUC), sensitivity, and specificity, with the specific training protocol outlined in [Supplementary Material 1A](#).

Multi-Instance Learning Feature Extraction

As demonstrated in [Supplementary Material 1B](#), this study utilized a MIL framework that integrates predicted likelihood histograms (PLH) and bag-of-words (BoW) approaches, coupled with a term frequency-inverse document frequency (TF-IDF) weighting strategy to aggregate slice predictive labels, ultimately constructing MIL features.

Feature Selection and Model Construction

2.5D DL-MIL: To mitigate the risk of overfitting, we applied Z-score normalization to standardize the MIL features. The *t*-test or Mann-Whitney *U*-test was employed to identify significant features ($P < 0.05$), and redundant features, exhibiting Pearson correlation coefficients greater than 0.9, were selectively removed to mitigate collinearity. In the context of a 10-fold cross-validation framework, the Least Absolute Shrinkage and Selection Operator (Lasso) regression

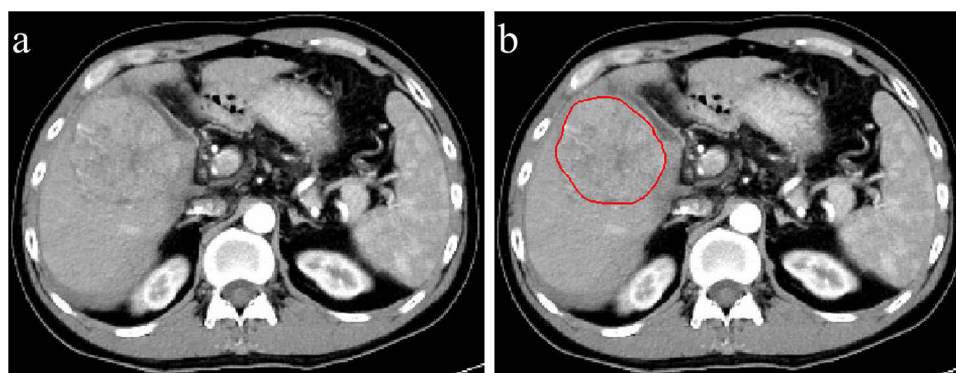


Figure 2 Tumor segmentation in a representative case. (a) Original CT arterial phase image, (b) Tumor boundary delineation performed using ITK-SNAP software. The red circular outline in the image represents the manually labeled tumor boundary, which is used for subsequent image cropping and radiomics feature extraction.

was utilized to optimize the regularization parameter λ and conduct feature selection. Finally, the ExtraTrees algorithm was implemented to construct the predictive model, while the Synthetic Minority Oversampling Technique (SMOTE) was employed to rectify the sample imbalance issue. Hyperparameter tuning was conducted using 5-fold cross-validation in combination with grid search to ensure the robustness of the model.

Radiomics: Manually extracted radiomics features through contouring served as the baseline, and a modeling approach analogous to the 2.5D DL-MIL model was employed to assess the predictive performance of traditional radiomics methods, thereby providing a reference for the comparative analysis between the 2.5D DL-MIL and traditional Radiomics models. The detailed modeling process and results are provided in [Supplementary Material 2A](#), and the types of radiomics features are shown in [Figure S1a](#).

Clinical: The Shapiro–Wilk test was employed to assess the normality of clinical features, and the *t*-test, Mann–Whitney *U*-test, or χ^2 -test was applied to identify significant clinical features ($P < 0.05$) based on the data type. The 2.5D DL-MIL model's machine learning algorithm was adopted, and a predictive model was constructed using the selected clinical features.

Statistical Analysis

Sample size estimation was conducted using MedCalc software (<https://www.medcalc.org>). The power of the test was set at 80%, the two-sided significance level was set to 0.05, and the alternative hypothesis for AUC was 0.800 (null hypothesis: 0.500). The samples in the ER and NER groups were allocated in a 1:1 ratio, and the minimum required sample size for both the training and validation sets was 26 cases (13 ER and 13 NER samples). A total of 133 cases were included in the training set (56 ER cases, 77 NER cases), and 58 cases in the validation set (23 ER cases, 35 NER cases). This sample size permits the effective detection of differences in AUC from 0.500 to ≥ 0.800 with 80% power.

The research was conducted in the Python 3.7.12 environment for data processing and model construction, employing Statsmodels 0.13.2 for statistical analysis, PyRadiomics 3.0.1 for radiomic feature extraction, Scikit-learn 1.0.2 for machine learning algorithms, PyTorch 1.11.0 for DL model development, and CUDA 11.3.1 with cuDNN 8.2.1 for acceleration and optimization. Intergroup comparisons of continuous variables were performed using either the *t*-test or the Mann–Whitney *U*-test, depending on the normality of the data, and categorical variables were analyzed using the chi-square test. The diagnostic performance of the model was evaluated using the receiver operating characteristic (ROC) curve, clinical applicability was verified through decision curve analysis (DCA), and AUC comparisons between models were conducted using the Delong test.

For the optimal prediction model, the Shapley Additive Explanation (SHAP) method was employed using Python 3.7 to quantify the contribution of each feature to the model's decision-making process. Meanwhile, Pearson correlation analysis was conducted using R software (version 4.2.2, <http://www.Rproject.org>) to investigate the relationship between the selected MIL features and HCC pathological indicators. All statistical analyses were deemed statistically significant at $P < 0.05$.

Results

Baseline Characteristics

The baseline clinical characteristics of the patients are presented in [Table 2](#). The analysis of the training set revealed that AFP levels ($P = 0.012$) and gender ($P = 0.016$) demonstrated significant differences between the ER and NER groups, while differences in other clinical features were not statistically significant ($P > 0.05$). Consequently, AFP and gender were utilized in the construction of the Clinical model.

Concerning the pathological characteristics, among 191 HCC patients, the mean Ki-67 expression index was 0.23 (standard deviation = 0.21), and the incidence of MVI was 35.08% (67/191), with 37 cases of M1 and 30 cases of M2. The pathological grade distribution was as follows: Edmondson–Steiner grade I in 13 cases (6.81%), grade II in 82 cases (42.93%), grade III in 73 cases (38.22%), and grade IV in 23 cases (12.04%).

Table 2 Analysis of the Baseline Clinical Characteristics

| Clinical feature | Training Set (n=133) | | | Validation Set (n=58) | | |
|-----------------------------|----------------------|----------------|-------|-----------------------|----------------|-------|
| | NER (n=77) | ER (n=56) | p | NER (n=35) | ER (n=23) | p |
| Age | 54.49±10.24 | 51.04±10.88 | 0.063 | 50.46±10.31 | 50.17±9.73 | 0.917 |
| BMI | 22.26±3.46 | 22.00±2.54 | 0.632 | 21.71±3.41 | 22.80±2.29 | 0.183 |
| HBsAg | 967.49±869.13 | 1019.11±726.57 | 0.496 | 1141.26±1046.98 | 1141.35±762.88 | 0.893 |
| AFP | 369.27±494.96 | 539.03±519.80 | 0.012 | 523.78±532.39 | 473.20±539.57 | 0.975 |
| ALB | 39.35±5.21 | 39.18±4.60 | 0.842 | 41.55±7.36 | 38.83±4.59 | 0.069 |
| AST | 82.31±135.98 | 51.94±27.49 | 0.758 | 64.54±69.10 | 68.47±71.76 | 0.169 |
| ALT | 75.34±118.98 | 50.77±37.14 | 0.531 | 62.21±77.61 | 74.23±98.27 | 0.474 |
| Gender | | | 0.016 | | | 0.08 |
| Female | 17(22.08%) | 3(5.36%) | | 9(25.71%) | 1(4.35%) | |
| Male | 60(77.92%) | 53(94.64%) | | 26(74.29%) | 22(95.65%) | |
| Alcohol consumption history | | | 0.173 | | | 1 |
| No | 54(70.13%) | 32(57.14%) | | 24(68.57%) | 15(65.22%) | |
| Yes | 23(29.87%) | 24(42.86%) | | 11(31.43%) | 8(34.78%) | |

Abbreviations: ER, early recurrence; NER, non-early recurrence; BMI, Body mass index; HBsAg, Hepatitis B surface antigen; AFP, Alpha-fetoprotein; ALB, Albumin; AST, Aspartate aminotransferase; ALT, Alanine aminotransferase.

Table 3 Slice-Level Prediction Results

| Dataset | DL Model | Accuracy | AUC(95% CI) | Sensitivity | Specificity |
|------------|-------------|----------|--------------------|-------------|-------------|
| Training | Resnet101 | 0.545 | 0.590(0.561–0.619) | 0.726 | 0.413 |
| Validation | Resnet101 | 0.541 | 0.588(0.544–0.633) | 0.846 | 0.340 |
| Training | Densenet121 | 0.698 | 0.733(0.707–0.759) | 0.482 | 0.855 |
| Validation | Densenet121 | 0.718 | 0.675(0.628–0.722) | 0.593 | 0.800 |
| Training | Resnet18 | 0.705 | 0.746(0.720–0.772) | 0.648 | 0.746 |
| Validation | Resnet18 | 0.702 | 0.716(0.674–0.758) | 0.688 | 0.712 |

Abbreviation: DL, deep learning.

Slice-Level Prediction Results

In a comparative study of the ResNet18, ResNet101, and DenseNet121 models, performance evaluation results indicated that ResNet18 exhibited a distinct advantage on the validation set. The AUC of ResNet18 was 0.716, which is higher than the AUCs of ResNet101 (0.588) and DenseNet121 (0.675), as shown in [Table 3](#), [Figure 3a](#) and [b](#). Additionally, ResNet18 demonstrated stable sensitivity (0.688) and specificity (0.712) on the validation set, thereby establishing it as the preferred model for subsequent MIL application development. [Figure 3c](#) and [d](#) illustrate visualizations of the final convolutional layer activations related to class predictions, utilizing the Gradient-weighted Class Activation Mapping (Grad-CAM) technique, thereby offering an intuitive explanation of the model's decision-making process.

Prediction Performance of Different Models

This study aggregated 13 PLH features and 13 BoW features using MIL. Following feature selection, feature distribution analysis and Lasso regression were applied, resulting in the selection of six features (BoW_01, BoW_02, BoW_03, BoW_09, BoW_10, and BoW_16) to construct the predictive model (details on feature distribution, Lasso regression results, and feature weights are provided in [Figure 4](#)). The performance evaluation revealed that the 2.5D DL-MIL model achieved accuracy, AUC, sensitivity, and specificity of 0.741, 0.840, 0.870, and 0.657, respectively, in the validation set ([Table 4](#)).

The radiomics features were analyzed using intra-group correlation coefficient (ICC) analysis, *t*-tests or Mann-Whitney *U*-tests, Pearson correlation analysis, and Lasso regression dimensionality reduction, resulting in the selection of eight features to construct the Radiomics model ([Figure S1b](#) and [c](#)). The model achieved an accuracy of 0.672, AUC of 0.678, sensitivity of 0.696, and specificity of 0.657 in the validation set ([Table 4](#) and [Figure S1d](#)).

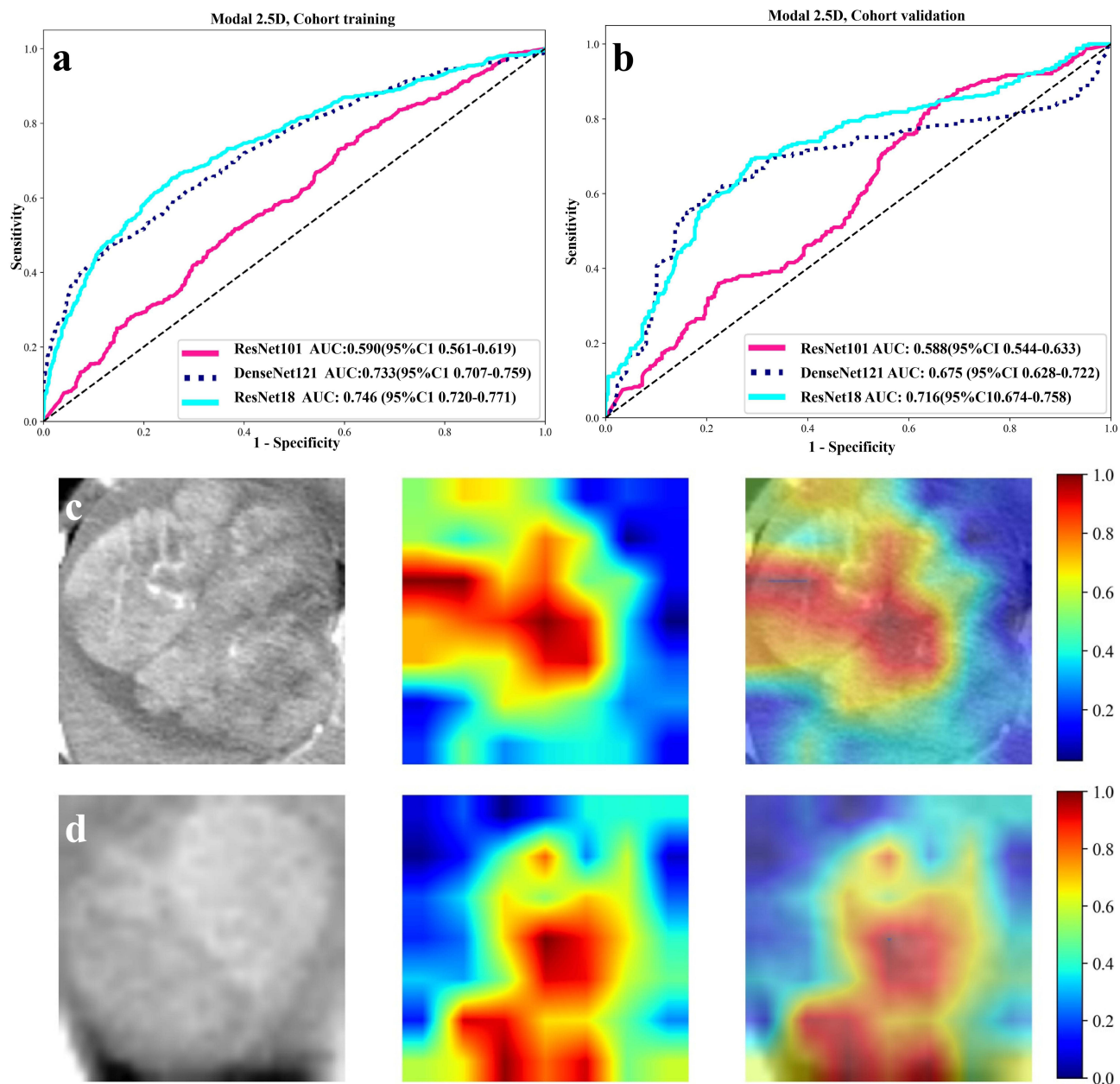


Figure 3 ROC curve and Grad-CAM visualization at the slice-level prediction. (a) illustrates the ROC curves for ResNet18 (blue), ResNet101 (red), and DenseNet121 (cyan) in the training set. (b) presents the ROC curves for ResNet18 (blue), ResNet101 (red), and DenseNet121 (cyan) in the validation set. (c and d) demonstrate the Grad-CAM visualizations of two representative samples, highlighting how the model selectively attends to different regions of the image for prediction. This facilitates a deeper understanding of the model's attention mechanism in practical applications.

The Clinical model, based on APF and gender, demonstrated suboptimal predictive performance, with an accuracy of 0.517, AUC of 0.598, sensitivity of 0.478, and specificity of 0.543 in the validation dataset (Table 4).

Model Comparison and SHAP Analysis

The ROC curves of various models are presented in Figure 5a and b, with the results demonstrating that the 2.5D DL-MIL model displayed the most superior overall predictive performance. The results of the Delong test on the validation set showed that the 2.5D DL-MIL model's AUC (0.840) significantly outperformed both the Radiomics model (AUC = 0.678, $P = 0.047$) and the Clinical model (AUC = 0.598, $P = 0.009$). The DCA curves (Figure 5c and d) further substantiated that the decision curves of the 2.5D DL-MIL model were substantially higher than the reference line in both

Table 4 Prediction Metrics of Different Models

| Dataset | Model | Accuracy | AUC(95% CI) | Sensitivity | Specificity |
|------------|-------------|----------|--------------------|-------------|-------------|
| Training | 2.5D DL-MIL | 0.812 | 0.861(0.800–0.923) | 0.714 | 0.883 |
| Validation | 2.5D DL-MIL | 0.741 | 0.840(0.737–0.943) | 0.870 | 0.657 |
| Training | Radiomics | 0.677 | 0.764(0.684–0.843) | 0.714 | 0.649 |
| Validation | Radiomics | 0.672 | 0.678(0.539–0.816) | 0.696 | 0.657 |
| training | Clinical | 0.632 | 0.674(0.585–0.764) | 0.589 | 0.662 |
| Validation | Clinical | 0.517 | 0.598(0.452–0.743) | 0.478 | 0.543 |

Abbreviations: 2.5D DL-MIL, 2.5D deep learning-multi-instance learning.

the training and validation sets, suggesting its potential for significant clinical net benefit, with this benefit being more pronounced compared to other models. For the top-performing 2.5D DL-MIL model, SHAP value analysis was conducted to evaluate the contribution and direction of influence of each MIL feature in the model. The results revealed that BoW_01, BoW_02, and BoW_03 negatively influenced the prediction of ER in HCC, while BoW_09, BoW_10, and BoW_1 exerted a positive effect, with BoW_02 making the most significant contribution to the prediction model (Figure 5e). Using sample 149 as an example, the waterfall plot (Figure 5f) and force plot (Figure 5g) visualizations demonstrated that BoW_01, BoW_02, and BoW_10 exerted a negative modulation effect in the prediction process, while BoW_03, BoW_09, and BoW_1 exhibited a positive modulation. The final predicted probability for this sample was 0.495, indicating a low likelihood of ER.

Correlation Between MIL Features and Pathological Characteristics

To further elucidate the underlying mechanisms of the optimal 2.5D DL-MIL model, we conducted a comprehensive analysis of the correlations between the selected MIL features and MVI, Ki-67 expression, and pathological grading of HCC, as presented in Table 5 and Figure 6. The correlation analysis revealed that BoW_01 ($r = -0.22$, $P = 0.0021$) and BoW_02 ($r = -0.20$, $P = 0.0066$) were significantly negatively correlated with MVI grade, whereas BoW_09 ($r = 0.22$, $P = 0.0022$) and BoW_1 ($r = 0.30$, $P < 0.0001$) exhibited a significant positive correlation with MVI grade. Regarding Ki-67 expression, BoW_02 exhibited a significant negative correlation with the Ki-67 expression index ($r = -0.22$, $P = 0.0026$), whereas BoW_09 ($r = 0.24$, $P = 0.0007$) and BoW_1 ($r = 0.19$, $P = 0.0076$) demonstrated significant positive correlations. Moreover, BoW_01, BoW_02, BoW_03, BoW_09, BoW_10, and BoW_1 exhibited no statistically significant correlation with HCC pathological grading ($P > 0.05$ for all).

Discussion

In current DL models for predicting HCC recurrence, 2D CNN, while efficient, is prone to losing tumor three-dimensional spatial correlation information.²¹ 3D CNN, though capable of capturing complete spatial features, is limited

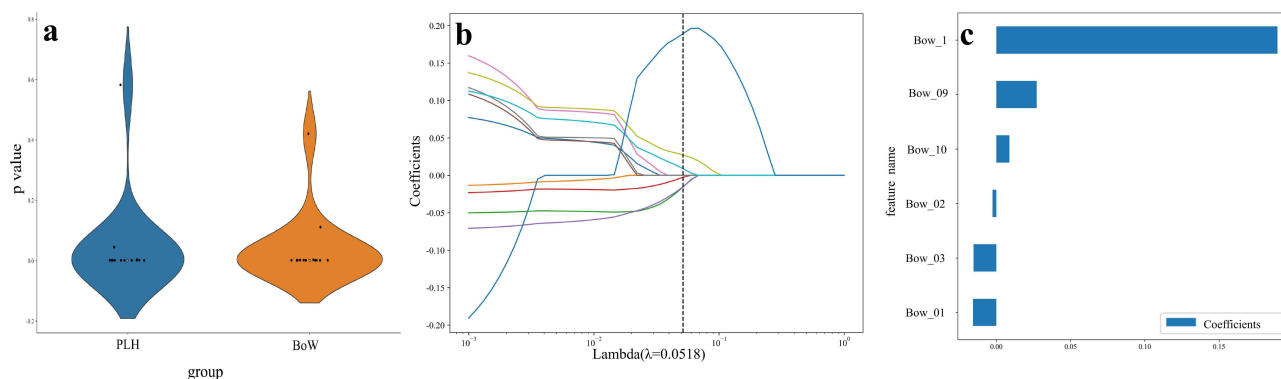


Figure 4 Visualization of the dimensionality reduction of MIL features: (a) distribution of differences in MIL features between the early recurrence and non-early recurrence groups in HCC, (b) lasso-based dimensionality reduction analysis with 10-fold cross-validation, and (c) distribution of selected feature weights.

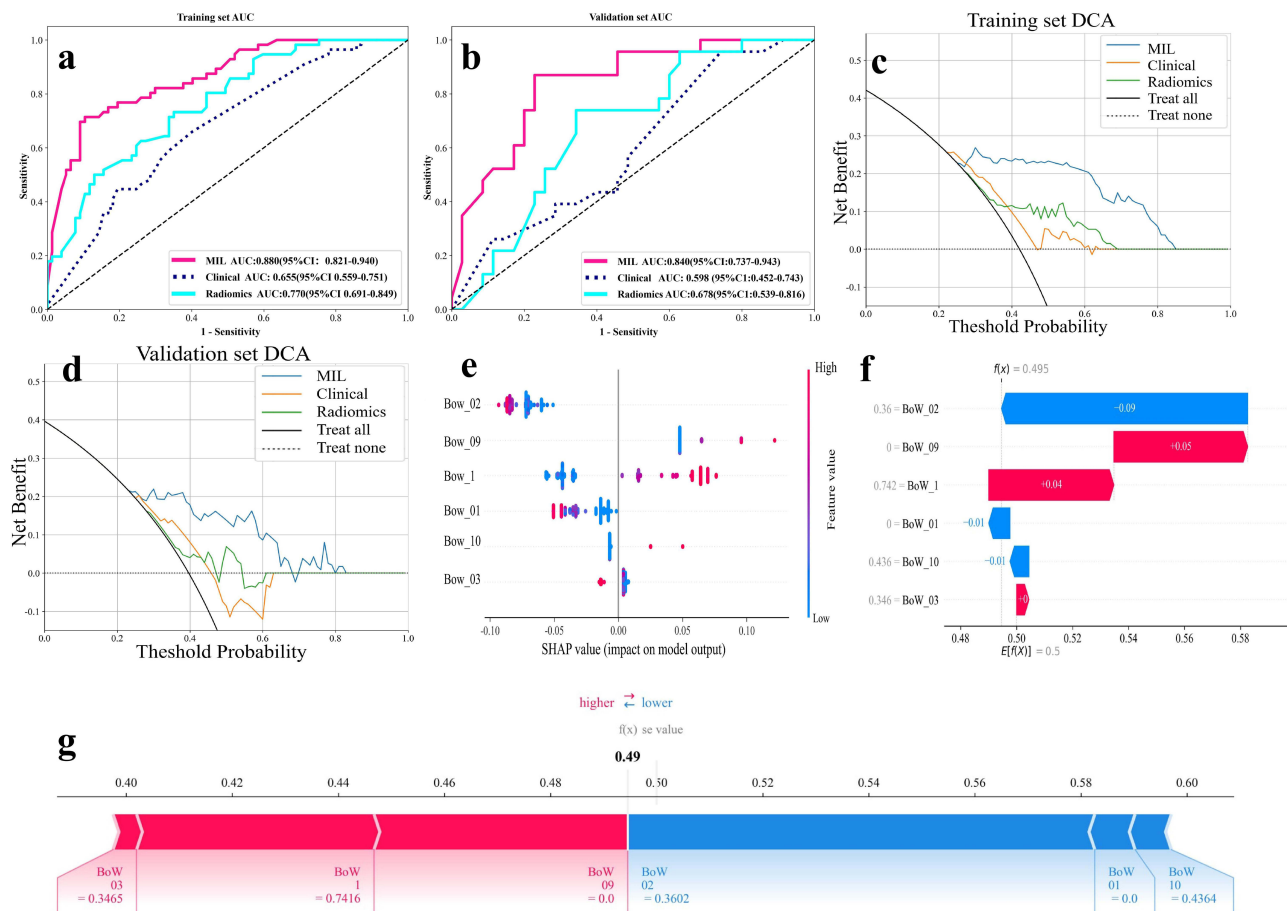


Figure 5 Model Comparison and SHAP Analysis. (a and b) present the ROC curves for each model in both the training and validation sets, where the red curve corresponds to the 2.5D DL-MIL model, the cyan curve represents the clinical model, and the blue curve indicates the radiomics model. It is evident that the 2.5D DL-MIL model achieves the highest AUC. (c and d) illustrate the DCA curves for the training and validation sets, with blue, yellow, and green again representing the 2.5D DL-MIL, clinical, and radiomics models, respectively. Notably, the clinical net benefit is highest for the 2.5D DL-MIL model. (e) presents the bee swarm plot, where the contributions and influence directions of various MIL features within the model are clearly shown. Finally, (f and g) depict the waterfall and force plots, which visually represent the contribution levels, directions, and final prediction probabilities of different features in individual sample decisions.

by its high computational cost.²² 2.5D DL precisely addresses the shortcomings in the collaboration between “spatial integrity - heterogeneity representation” in the above studies. In this study, we developed a CT-based 2.5D DL-MIL model, a Radiomics model, and a Clinical model, comparing their predictive performance for early postoperative recurrence of HCC. Additionally, we evaluated the correlation between the selected MIL features and pathological parameters. The results indicated that the 2.5D DL-MIL model demonstrated superior predictive performance, and

Table 5 Correlation Between MIL Features and Pathological Characteristics

| | MVI | | Ki-67 | | Pathological Grading | |
|--------|--------------------|---------|--------------------|-------|----------------------|-------|
| | r(95% CI) | p | r(95% CI) | p | r(95% CI) | p |
| BoW_01 | -0.22(-0.36~-0.08) | 0.002 | -0.08(-0.22~-0.06) | 0.263 | 0.03(-0.11~0.17) | 0.686 |
| BoW_02 | -0.20(-0.33~-0.06) | 0.007 | -0.22(-0.35~-0.08) | 0.003 | -0.02(-0.16~0.12) | 0.784 |
| BoW_03 | -0.13(-0.27~-0.01) | 0.070 | -0.10(-0.24~-0.04) | 0.160 | -0.09(-0.23~0.06) | 0.229 |
| BoW_09 | 0.22(0.08~0.35) | 0.002 | 0.24(0.10~0.37) | 0.001 | 0.01(-0.13~0.16) | 0.843 |
| BoW_10 | -0.02(-0.16~-0.12) | 0.793 | 0.01(-0.13~0.15) | 0.884 | -0.03(-0.17~0.12) | 0.722 |
| BoW_1 | 0.30(0.16~0.42) | <0.0001 | 0.19(0.05~0.33) | 0.008 | -0.08(-0.22~0.06) | 0.258 |

Notes: |r|<0.3 indicates a weak correlation, and 0.3≤|r|<0.5 indicates a moderate correlation.

Abbreviations: MIL, Multiple Instance Learning; BoW, Bag of Words; MVI, Microvascular Invasion.

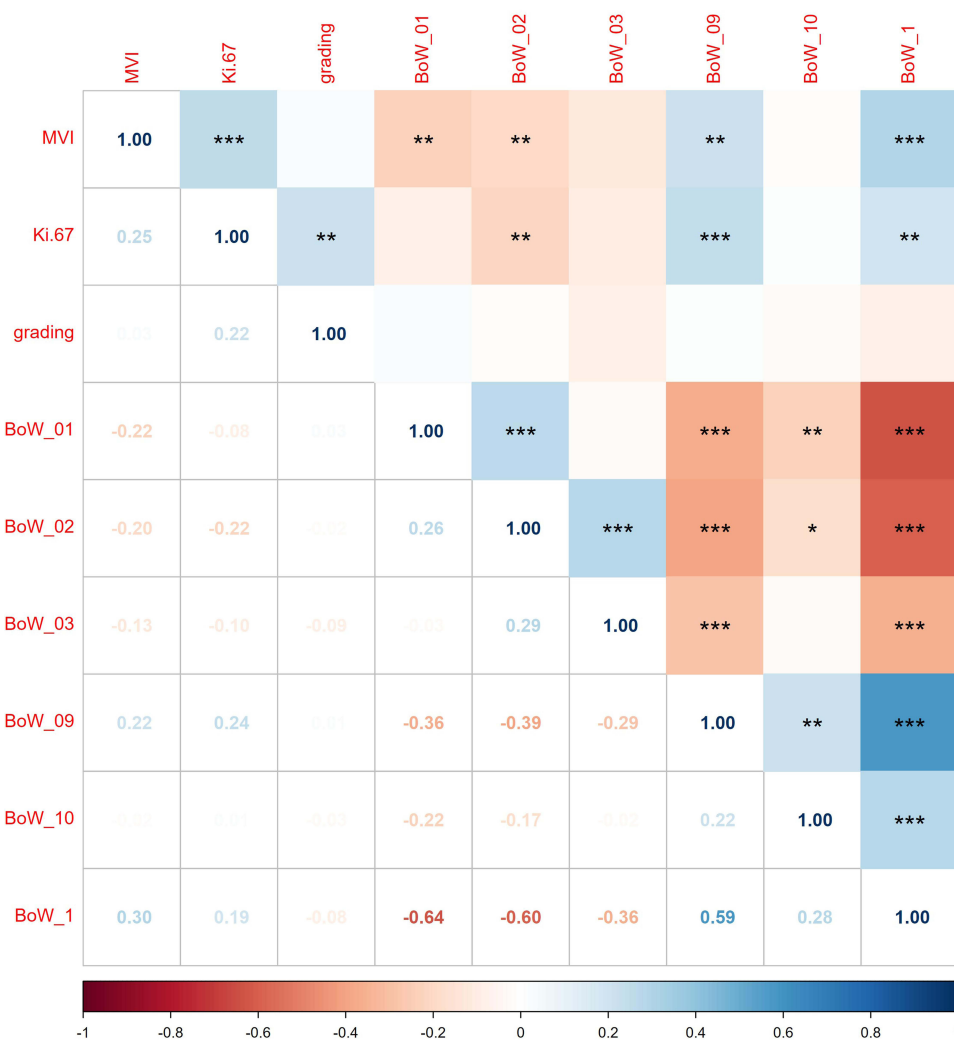


Figure 6 Correlation heatmap of MIL features with MVI, Ki-67 expression, and pathological grading in HCC. The values and colors within the grid indicate the correlation coefficients between variables, with warmer tones (eg, red) indicating negative correlations and cooler tones (eg, blue) representing positive correlations. The intensity of the color corresponds to the strength of the correlation. *in the grid denotes $p < 0.05$, **indicates $p < 0.01$, and ***represents $p < 0.001$.

certain MIL features were significantly correlated with MVI and Ki-67 expression of HCC. These findings highlight substantial innovation and clinical relevance.

The 2.5D DL-MIL model demonstrated distinct advantages in predicting ER in HCC, achieving an AUC of 0.840 in the validation set, significantly outperforming the Radiomics model (AUC = 0.678) and the Clinical model (AUC = 0.598). These advantages align with prior research findings based on the 2.5D DL-MIL model.^{10,23,24} The decision curve analyses further confirmed the superiority of the 2.5D DL-MIL model in terms of clinical applicability. These advantages may stem from several factors. First, at the data feature extraction stage, the 2.5D DL technique integrates multi-plane slice information surrounding the tumor's largest cross-section. In contrast to traditional 2D models, which overlook spatial information between slices, it can capture tumor features in three-dimensional space more comprehensively.^{10,11,25,26} Second, the MIL framework overcomes the limitations of single-slice information by aggregating the prediction results from multiple image slices, capturing the tumor's heterogeneity as a whole, and mitigating the one-sidedness of local features.^{23,27} In model construction, the 2.5D DL-MIL model undergoes slice-level training using an advanced DL architecture (ultimately selecting ResNet18), combined with PLH and BoW methods for feature fusion, enabling the automatic learning of complex, effective feature representations. Traditional Radiomics models, however, rely on manual delineation for feature extraction, which is influenced by human factors.^{8,28} Although baseline clinical features showed significant differences in AFP and gender between the ER and NER groups of HCC in the

training set ($P < 0.05$), the Clinical model was constructed based on limited clinical indicators, which fail to fully capture the heterogeneity and complexity of the tumor.⁸ Furthermore, the AUC of our 2.5D deep learning model is comparable to that of the 3D convolutional neural network model (AUC=0.846).²² This may be related to our model's ability to effectively capture the deep features and spatial information of the tumor.

SHAP analysis is a crucial method for elucidating the decision-making mechanisms of the 2.5D DL-MIL model. The results of this study showed that different MIL features played different roles in model decision-making. In terms of the correlation between imaging features and tumor recurrence risk, negative-impact features may represent imaging characteristics associated with relatively benign biological behavior of the tumor, such as higher ADC values and a complete capsule.²⁹ Conversely, positive-impact features may reflect imaging characteristics indicative of higher invasiveness or recurrence tendency of the tumor, such as larger tumor size and non-smooth tumor margins.^{29,30} Visual analysis of waterfall and force plots allows for the intuitive identification of the contribution and direction of each feature in predicting a single sample, thereby providing strong support for clinicians in interpreting the model's predictions. This enhances the model's interpretability and clinical trustworthiness.³¹

The correlation between MIL features and pathological indicators offers valuable insights into the biological mechanisms underlying the 2.5D DL-MIL model. Previous studies have established that MVI is a crucial pathological marker for evaluating tumor invasiveness and metastatic potential,³² while Ki-67 expression levels are closely associated with tumor cell proliferative activity.³³ Moreover, higher MVI grades and Ki-67 expression levels in HCC patients correlate with a markedly increased risk of postoperative recurrence.^{3,15,34,35} In our study, the association between MIL features and MVI grade, as well as Ki-67 expression level, suggests that HCC with low BoW_02 value and high BoW_09 and BoW_1 value has higher invasiveness and active tumor cell proliferation, making these patients more prone to short-term recurrence, which is consistent with the direction of influence of these features on the model in the SHAP analysis. Additionally, the correlation between BoW_01 and MVI grade, as well as Ki-67 expression level, suggests that BoW_01 reflects the invasiveness of HCC but does not characterize the proliferative activity of tumor cells. The aforementioned correlation results between MIL features and pathological indicators suggest that the 2.5D DL-MIL model effectively captures imaging information related to postoperative recurrence of HCC. The features extracted by the model may indirectly reflect the characteristics of tumor molecular biological processes in imaging, thereby establishing an intrinsic relationship with pathological indicators. Moreover, this study found that the Grad-CAM visualization showed that the DL model can automatically focus on the tumor regions with rich blood supply and the junction areas between tumor and normal liver tissue, which are consistent with the areas where pathological occurrences such as HCC MVI take place. Although most of the significantly correlated BoW features show weak correlation, they still demonstrate good clinical value in predicting the high-risk population for early recurrence of HCC preoperatively and non-invasively.

The 2.5D DL-MIL model in this study provides precise guidance for post-operative personalized treatment and follow-up of HCC. For patients with a high recurrence risk of HCC suggested by the model, we can shorten the follow-up intervals and consider post-operative adjuvant therapies (such as post-operative transcatheter arterial chemoembolization, targeted drug therapy, etc). For low-risk patients, we can appropriately extend the follow-up intervals and reduce unnecessary medical expenses and radiation exposure. However, this study presents several limitations. First, retrospective studies are prone to potential selection bias, which may impact the generalizability of the study's findings. Secondly, the study sample is derived from a single center and includes a relatively small sample size, thus requiring further validation of the model's generalizability. In the future, we plan to conduct a multi-center, large-sample prospective study at three medical centers in China, and perform effective external data validation to more precisely evaluate the performance of the 2.5D DL-MIL model and its clinical application value. Finally, This study confirmed the correlation between MIL features and pathological indicators (MVI, Ki-67), but it has not yet explored its association with the tumor's molecular mechanisms. In the future, we will clarify the molecular mechanisms behind imaging features through imaging genomics association analysis and further validate the stability of the imaging-genomics association in a multi-center cohort.

Conclusion

The 2.5D DL-MIL model developed in this study exhibited considerable advantages in predicting early recurrence of HCC, while SHAP analysis, alongside correlation studies, offered key insights into the model's decision-making mechanism and its biological implications. Although this study has certain limitations, it establishes a robust foundation for future research and clinical applications and is anticipated to propel advancements in radiology within the realm of precision diagnosis and treatment of HCC.

Acknowledgments

We would like to acknowledge all funding organizations and patients of this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the “Summit Plan (New Departure)” project for the development of doctoral degree authorization points and professional disciplines at the Affiliated Hospital of Youjiang Medical University for Nationalities (DF20244433); Self-funded research project by the Guangxi Health and Wellness Committee (Z-L20240824, Z-L20240834); and The Project to Enhance the Research Foundations of Young and Mid-career Faculty in Guangxi Universities (2024KY0562, 2024KY0559).

Disclosure

The author(s) report no conflicts of interest in this work.

References

- Hyuna S, Jacques F, Rebecca LS, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249. doi:10.3322/caac.21660
- Jonggi C, Chanyoung J, Young-Suk L. Tenofovir versus entecavir on recurrence of hepatitis B virus-related hepatocellular carcinoma after surgical resection. *Hepatology.* 2020;73(2):661–673. doi:10.1002/hep.31289
- Xu XF, Diao YK, Zeng YY, et al. Association of severity in the grading of microvascular invasion with long-term oncological prognosis after liver resection for early-stage hepatocellular carcinoma: a multicenter retrospective cohort study from a hepatitis B virus-endemic area. *Int J Surg.* 2023;109(4):841–849. doi:10.1097/js9.0000000000000325
- Fuster-Anglada C, Mauro E, Ferrer-Fàbrega J, et al. Histological predictors of aggressive recurrence of hepatocellular carcinoma after liver resection. *J Hepatol.* 2024;81(6):11. doi:10.1016/j.jhep.2024.06.018
- Zhang LN, Xiao YQ, Dong MS, Li MS, Chen HM, Wang J. Three-dimensional MR elastography-based stiffness for assessing the status of Ki67 proliferation index and Cytokeratin-19 in hepatocellular carcinoma. *Eur Radiol.* 2025;14. Early Access. doi:10.1007/s00330-025-11375-w
- Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol.* 2022;76(3):681–693. Review. doi:10.1016/j.jhep.2021.11.018
- Walaa A, Mohamed EK. Hepatocellular carcinoma recurrence: predictors and management. *Liver Res.* 2023;7(4):321–332. doi:10.1016/j.livres.2023.11.004
- Huynh BN, Groendahl AR, Tomic O, et al. Head and neck cancer treatment outcome prediction: a comparison between machine learning with conventional radiomics features and deep learning radiomics. *Front Med.* 2023;10:20. doi:10.3389/fmed.2023.1217037
- Chengyan W, Shuo W, Sha H, et al. A protocol for body MRI/CT and extraction of Imaging-Derived Phenotypes (IDPs) from the China phenobank project. *Phenomics.* 2025;4(6):594–616. doi:10.1007/s43657-023-00141-x
- Zhang YB, Chen ZQ, Bu Y, Lei P, Yang W, Zhang W. Construction of a 2.5D deep learning model for predicting early postoperative recurrence of hepatocellular carcinoma using multi-view and multi-phase CT images. *J Hepatocell Carcinoma.* 2024;11:2223–2239. doi:10.2147/jhc.S493478
- Lin C, Cao T, Tang MW, Pu W, Lei PG. Predicting hepatocellular carcinoma response to TACE: a machine learning study based on 2.5D CT imaging and deep features analysis. *Eur J Radiol.* 2025;187:11. doi:10.1016/j.ejrad.2025.112060
- Su ZY, Rezapour M, Sajjad U, Gurcan MN, Khan MK. Attention2Minority: a salient instance inference-based multiple instance learning for classifying small lesions in whole slide images. *Comput Biol Med.* 2023;167:11. doi:10.1016/j.compbimed.2023.107607
- Brandao ABD, Rodriguez S, Marroni CA, Fleck AD Jr, Fernandes MV, Mucenic M. Performance of eight predictive models for hepatocellular carcinoma recurrence after liver transplantation: a comparative study. *Ann Hepatol.* 2024;29(2):6. doi:10.1016/j.aohep.2023.101184

14. Nevola R, Ruocco R, Criscuolo L, et al. Predictors of early and late hepatocellular carcinoma recurrence. *World J Gastroenterol.* 2023;29(8):1243–1260. Review. doi:10.3748/wjg.v29.i8.1243
15. Xia TY, Zhou ZH, Meng XP, et al. Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. *Radiology.* 2023;307(4):11. doi:10.1148/radiol.222729
16. Zhang HM, Li SP, Yu Y, et al. Bi-directional roles of IRF-1 on autophagy diminish its prognostic value as compared with Ki67 in liver transplantation for hepatocellular carcinoma. *Oncotarget.* 2016;7(25):37979–37992. doi:10.18632/oncotarget.9365
17. Zhou ZY, Cao SY, Chen CB, et al. A novel nomogram for the preoperative prediction of edmondson-steiner grade III-IV in hepatocellular carcinoma patients. *J Hepatocell Carcinoma.* 2023;10:1399–1409. doi:10.2147/jhc.S417878
18. Cong WM, Bu H, Chen J, et al. Practice guidelines for the pathological diagnosis of primary liver cancer: 2015 update. *World J Gastroenterol.* 2016;22(42):9279–9287. doi:10.3748/wjg.v22.i42.9279
19. Wu HZ, Han XR, Wang ZH, et al. Prediction of the Ki-67 marker index in hepatocellular carcinoma based on CT radiomics features. *Phys Med Biol.* 2020;65(23):11. doi:10.1088/1361-6560/abac9c
20. Pfeiffenberger J, Mogler C, Gotthardt DN, et al. Hepatobiliary malignancies in Wilson disease. *Liver Int.* 2015;35(5):1615–1622. doi:10.1111/liv.12727
21. Hui Z, Fanding H. Prediction of early recurrence of HCC after hepatectomy by contrast-enhanced ultrasound-based deep learning radiomics. *Front Oncol.* 2022;12:930458. doi:10.3389/fonc.2022.930458
22. Jie P, Jiaren W, Hongbo Z, et al. Three-dimensional multimodal imaging for predicting early recurrence of hepatocellular carcinoma after surgical resection. *J Adv Res.* 2025. doi:10.1016/j.jare.2025.06.031
23. Chang BW, Geng Z, Mei JM, et al. Application of multimodal deep learning and multi-instance learning fusion techniques in predicting STN-DBS outcomes for Parkinson's disease patients. *Neurotherapeutics.* 2024;21(6):9. doi:10.1016/j.neurot.2024.e00471
24. Kim Y, Kim YG, Park JW, et al. A CT-based deep learning model for predicting subsequent fracture risk in patients with hip fracture. *Radiology.* 2024;310(1):9. doi:10.1148/radiol.230614
25. Zeng YW, Zhang XY, Kawasumi Y, et al. A 2.5D deep learning-based method for drowning diagnosis using post-mortem computed tomography. *IEEE J Biomed Health Inform.* 2023;27(2):1026–1035. doi:10.1109/jbhi.2022.3225416
26. Zhu JL, Zou L, Xie X, Xu RZ, Tian Y, Zhang B. 2.5D deep learning based on multi-parameter MRI to differentiate primary lung cancer pathological subtypes in patients with brain metastases. *Eur J Radiol.* 2024;180:8. doi:10.1016/j.ejrad.2024.111712
27. Wang ZK, Bi Y, Pan T, et al. Targeting tumor heterogeneity: multiplex-detection-based multiple instance learning for whole slide image classification. *Bioinformatics.* 2023;39(3):7. doi:10.1093/bioinformatics/btad114
28. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From handcrafted to deep-learning-based cancer radiomics challenges and opportunities. *IEEE Signal Process Mag.* 2019;36(4):132–160. doi:10.1109/msp.2019.2900993
29. Wu YY, Ye Z, Yang T, et al. Preoperative prediction of early recurrence in hepatocellular carcinoma using simultaneous multislice diffusion kurtosis imaging. *Eur Radiol.* 2025;12. Early Access. doi:10.1007/s00330-025-11633-x
30. Wei YY, Pei W, Qin YY, Su DK, Liao H. Preoperative MR imaging for predicting early recurrence of solitary hepatocellular carcinoma without microvascular invasion. *Eur J Radiol.* 2021;138:7. doi:10.1016/j.ejrad.2021.109663
31. Zhong X, Salahuddin Z, Chen Y, et al. An interpretable radiomics model based on two-dimensional shear wave elastography for predicting symptomatic post-hepatectomy liver failure in patients with hepatocellular carcinoma. *Cancers.* 2023;15(21):15. doi:10.3390/cancers15215303
32. Zhang ZH, Jiang C, Qiang ZY, et al. Role of microvascular invasion in early recurrence of hepatocellular carcinoma after liver resection: a literature review. *Asian J Surg.* 2024;47(5):2138–2143. doi:10.1016/j.asjsur.2024.02.115
33. Andrés-Sánchez N, Fisher D, Krasinska L. Physiological functions and roles in cancer of the proliferation marker Ki-67. *J Cell Sci.* 2022;135(11):13. Review. doi:10.1242/jcs.258932
34. Li HH, Qi LN, Ma L, Chen ZS, Xiang BD, Li LQ. Effect of KI-67 positive cellular index on prognosis after hepatectomy in barcelona clinic liver cancer stage A and B hepatocellular carcinoma with microvascular invasion. *OncoTargets Ther.* 2018;11:4747–4754. doi:10.2147/ott.S165244
35. Liu ZW, Yang SM, Chen XJ, et al. Nomogram development and validation to predict Ki-67 expression of hepatocellular carcinoma derived from Gd-EOB-DTPA-enhanced MRI combined with T1 mapping. *Front Oncol.* 2022;12:15. doi:10.3389/fonc.2022.954445

Journal of Hepatocellular Carcinoma

Publish your work in this journal

The Journal of Hepatocellular Carcinoma is an international, peer-reviewed, open access journal that offers a platform for the dissemination and study of clinical, translational and basic research findings in this rapidly developing field. Development in areas including, but not limited to, epidemiology, vaccination, hepatitis therapy, pathology and molecular tumor classification and prognostication are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-hepatocellular-carcinoma-journal>

Dovepress
Taylor & Francis Group