

RESEARCH

Open Access



# An MRI-based fusion model for preoperative prediction of perineural invasion status in patients with intrahepatic cholangiocarcinoma

Zuochao Qi<sup>1</sup>, Hao Yuan<sup>1</sup>, Qingshan Li<sup>1</sup>, Pengyu Chen<sup>2</sup>, Dongxiao Li<sup>3</sup>, Kunlun Chen<sup>4</sup>, Bo Meng<sup>5</sup>, Peigang Ning<sup>6</sup>, Haibo Yu<sup>1\*</sup> and Deyu Li<sup>1\*</sup>

## Abstract

**Background** To develop and validate an MRI-based fusion model for preoperative prediction of perineural invasion (PNI) status in patients with intrahepatic cholangiocarcinoma (ICC).

**Methods** A retrospective collection of 192 ICC patients from three medical centers (training set:  $n = 147$ ; external test set:  $n = 45$ ) was performed. Patients were classified into the PNI-positive and PNI-negative groups based on postoperative pathological results. After image preprocessing, a total of 1,197 features were extracted from T2-weighted imaging (T2WI). Feature selection was performed, and a radiomics model was constructed using machine learning algorithms, followed by SHapley Additive exPlanations (SHAP) visualization. Subsequently, a deep learning model was constructed based on the pre-trained ResNet101, with Gradient-weighted Class Activation Mapping (Grad-CAM) used for visualization. Finally, a fusion model incorporating deep learning, radiomics, and clinical features was developed using logistic regression, and visualization was performed with a nomogram. The predictive performance of the model was evaluated based on the area under the curve (AUC), calibration curves, and decision curve analysis (DCA).

**Results** The fusion model, which integrates deep learning signature, radiomics signature, and two clinical features, demonstrated strong discrimination for PNI status. In the training set, the AUC was 0.905, with an accuracy of 0.823; in the external test set, the AUC was 0.760, with an accuracy of 0.778. Visualization methods provided support for the practical application of the model.

**Conclusion** The fusion model aids in the preoperative identification of PNI status in patients with ICC, and may help guide clinical decision-making regarding preoperative staging and adjuvant therapy.

**Keywords** Intrahepatic cholangiocarcinoma, Perineural invasion, Deep learning, Radiomics, Magnetic resonance imaging

\*Correspondence:

Haibo Yu  
yhb2101661@zzu.edu.cn  
Deyu Li  
lidy0408@sohu.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Intrahepatic cholangiocarcinoma (ICC) is a highly aggressive malignant tumor originating from the secondary intrahepatic bile ducts to the smallest branches [1]. It is the second most common primary liver cancer, accounting for about 10–15% of primary liver cancers [2]. The incidence of ICC is rising worldwide, with Asia having the highest rates, and mortality has also increased over the past decades [3, 4]. Radical surgical resection remains the optimal treatment option for achieving long-term survival in patients with ICC [5]. However, ICC has a high postoperative recurrence rate of 60–70%, and even after curative resection, the five-year survival rate for ICC patients is only around 20–35% [6, 7].

Perineural invasion (PNI) is defined as cancer cells infiltrating the nerve sheath or surrounding nerve fibers up to 33% of their circumference [8]. It is an independent pathological feature, distinct from vascular and lymphatic metastasis, and not an extension of tumor spread through low-resistance pathways [9, 10]. A recent study has shown that PNI is a robust, independent predictor of tumor recurrence and long-term survival following radical resection of ICC, underscoring the need for assessing PNI status [11].

Magnetic resonance imaging (MRI) is non-invasive and free from ionizing radiation, offering multi-parameter, multi-angle, and multi-sequence imaging with high soft-tissue resolution and clear depiction of anatomical structures and lesions [12]. It is thus critical in clinical and research applications. The study by Granata et al. found that T2WI signal intensity differences between ICC and control groups were statistically significant [13]. Considering clinical applicability, this study selected MRI T2WI for further investigation. Radiomics transforms medical images into high-dimensional, mineable data, which can then be analyzed alongside clinical information to diagnose and predict disease characteristics [14, 15]. With the rapid development of radiomics, numerous studies have confirmed the outstanding performance of MRI-based radiomics in ICC diagnosis and differentiation [16, 17]. Deep learning, an important component of machine learning first proposed by Hinton, has since been widely applied in various fields, with convolutional neural networks (CNNs) being the most commonly used model in image analysis [18, 19]. Deep learning based on medical imaging data has made significant progress in liver disease research, providing powerful tools for diagnosis and treatment [20–22].

This study aims to construct a fusion model integrating deep learning and radiomics using MRI data to preoperatively predict PNI status in patients with ICC, enabling clinicians to better assess patient conditions and develop survival-oriented treatment strategies.

## Methods

### Patients

This study was approved by the institutional ethical review boards of all the hospitals involved and adhered to the Declaration of Helsinki. All patients enrolled in the study were informed and provided informed consent.

Inclusion criteria were as follows: (1) curative surgical resection treatment, (2) MRI examination within one week before surgery, and (3) postoperative pathologic confirmation of ICC. Exclusion criteria included (1) preoperative adjuvant tumor therapy such as radiotherapy, (2), inability to identify the primary tumor on MRI or poor image quality (e.g., severe artifacts), and (3), incomplete clinical and imaging data. A retrospective collection of 99 ICC cases treated at Zhengzhou University People's Hospital (Center I) from January 2018 to October 2023, 48 ICC cases at Henan Cancer Hospital (Center II) from January 2020 to January 2023, and 45 ICC cases at the First Affiliated Hospital of Zhengzhou University (Center III) from January 2021 to January 2023 was conducted. Centers I and II served as the training set, while Center III was used as the external test set (Fig. 1).

### Perineural invasion

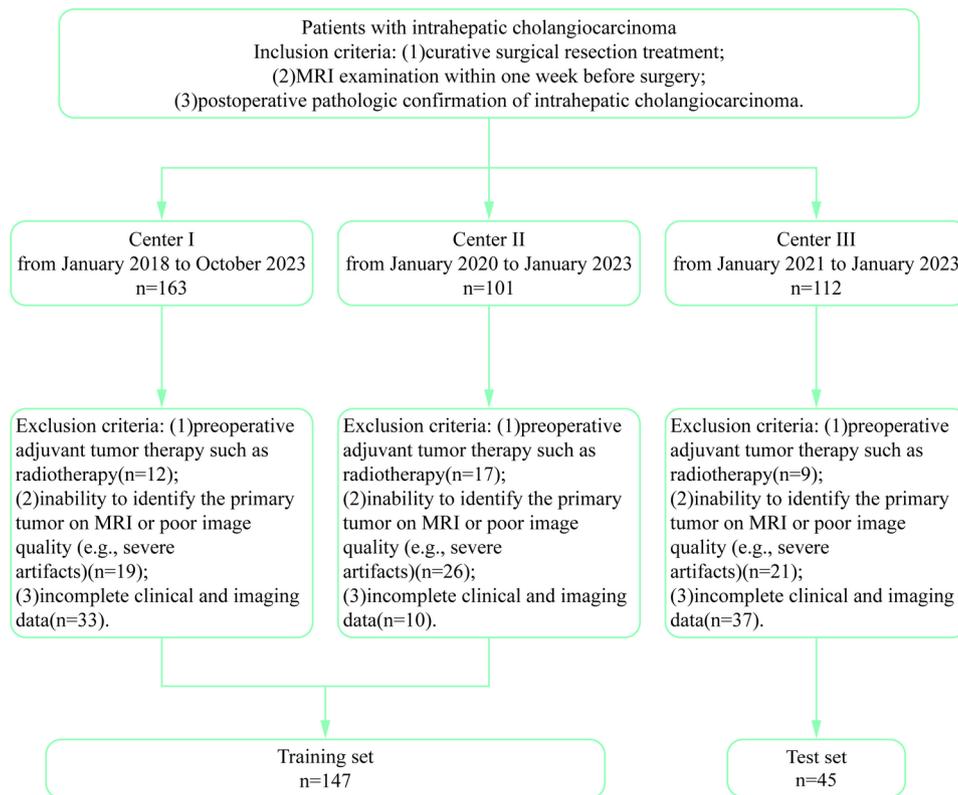
The outcome index of this study was the PNI status in patients with ICC, determined by hepatic pathology experts. The diagnostic criteria required microscopic observation of cancer cell infiltration in any layer of the nerve fiber (including the epineurium, perineurium, and endoneurium) or cancer encircling  $\geq 33\%$  of the nerve circumference [8].

### Clinical data collection

Patient clinical data from the week before surgery, including age, sex, alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA), carbohydrate antigen 19–9 (CA19-9), albumin (ALB), total bilirubin (TB), prothrombin time (PT), and activated partial thromboplastin time (APTT), were collected through the hospital information system.

### MRI acquisition, preprocessing, and image segmentation

Routine abdominal MRI scans were performed at different clinical MRI systems across the three centers, all within one week before surgery, using 3.0T magnetic field strength. The MRI protocols included conventional sequences such as T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI). Further details are provided in Supplementary Information 1. T2WI images in Digital Imaging and Communications in Medicine (DICOM) format were downloaded for all patients using the hospital information system. Preprocessing steps were applied to eliminate batch effects from different centers. First, N4 bias field correction was performed using the SimpleITK library to reduce artifacts caused by



**Fig. 1** Flowchart of patient recruitment and grouping. MRI, magnetic resonance imaging; Center I, Zhengzhou University People's Hospital; Center II, Henan Cancer Hospital; Center III, the First Affiliated Hospital of Zhengzhou University

scanner variations or inconsistencies during the imaging process, enhancing subsequent image analysis [23]. Next, to minimize voxel variability due to different scanners or acquisition protocols, images were resampled to a fixed resolution of 1 mm × 1 mm × 1 mm using the nearest-neighbor interpolation method. All MRI images were retrospectively evaluated by two researchers (NPG, YHB), each with over 10 years of experience in abdominal MRI interpretation. They were blinded to the clinical and pathological outcomes. Initially, Researcher 1 manually segmented the region of interest (ROI) on the axial plane using ITKSNAP ([www.itk-snap.org](http://www.itk-snap.org)). After segmenting all ROIs, 30 patients were randomly selected for repeated segmentation by both Researcher 1 and Researcher 2 to assess intra- and inter-observer reliability of radiomics features. In cases with multiple lesions, the largest lesion was selected for evaluation.

#### Radiomics feature extraction

Radiomics features were extracted from the MRI images using the Pyradiomics package based on the Python 3.7 platform (<https://github.com/Radiomics/pyradiomics>). Filters such as Laplacian of Gaussian and wavelet transforms were applied to generate additional derived images. Seven categories of features were extracted from both the original and derived images, including

shape-based features, first-order statistical features, gray-level co-occurrence matrix (GLCM) features, gray-level run-length matrix (GLRLM) features, gray-level size zone matrix (GLSZM) features, gray-level dependence matrix (GLDM) features, and neighboring gray-tone difference matrix (NGTDM) features. Feature extraction followed the guidelines proposed by the Imaging Biomarker Standardization Initiative (IBSI) [24].

#### Radiomics feature selection and radiomics signature construction

Feature selection was conducted following these steps to minimize bias and potential overfitting: (i) Z-score normalization was applied to the extracted features to adjust them to 0 mean and 1 variance; (ii) intra- and interclass correlation coefficients were calculated to assess feature stability and reproducibility. Features with intra- and interclass correlation coefficients > 0.80 were considered acceptable for reproducibility and selected for further analysis; (iii) features most relevant to PNI status were selected using t-test or Mann-Whitney U test to identify significant differences ( $P < 0.05$ ) between PNI-positive and PNI-negative groups; (iv) Pearson correlation analysis was performed to assess multicollinearity, retaining only features with coefficients < 0.7 or > -0.7 that demonstrated superior diagnostic performance; (v) the Least

Absolute Shrinkage and Selection Operator (LASSO) regression model was employed to further reduce feature dimensionality, using 10-fold cross-validation to select the most predictive features with non-zero coefficients by adjusting the penalty parameter.

After LASSO, the final features were input into five machine learning models—logistic regression (LR), Naive Bayes, k-nearest neighbors (KNN), extremely randomized trees (ExtraTrees), and multilayer perceptron (MLP) (Supplementary Information 2)—to construct risk models. A five-fold cross-validation strategy was used to select the optimal model configurations and hyperparameter settings. Ultimately, the model with the best area under the curve (AUC) was chosen to construct the radiomics signature for each patient.

#### **Deep transfer learning model and deep learning signature construction**

The rectangular ROI encompassing the largest cross-section of the tumor was cropped, and an additional 10 voxels were added outward to include surrounding tissue as much as possible in the original image. To improve model generalization, data augmentation techniques were applied, including random horizontal and vertical flips, and random cropping to a size of  $224 \times 224$  pixels. The input image size was set to  $224 \times 224$  pixels. Training deep learning models is computationally expensive and requires a large number of images due to the millions of parameters that need to be adjusted. To address the issue of limited data, transfer learning techniques can enable knowledge transfer to the deep learning model. ImageNet (<http://www.image-net.org>) is a large-scale image database containing millions of annotated images across thousands of categories. We selected the ResNet101 network, pre-trained on the ImageNet dataset, to apply transfer learning for the PNI status prediction task [25]. A global fine-tuning strategy was adopted to update parameters, with an initial learning rate set at 0.01, dynamically adjusted using a cosine annealing algorithm. Training consisted of 50 epochs with a batch size of 32, using a stochastic gradient descent (SGD) optimizer to update model parameters. Model performance evaluation relied on the loss function. The trained ResNet101 model can predict each patient's probability of PNI, with this probability used as each patient's deep learning signature.

#### **Construction of the fusion model**

The significance of clinical factors in predicting PNI status was assessed using univariate and multivariate logistic regression analyses. Clinical variables with  $P < 0.05$  were selected and combined with radiomics signature and deep learning signature to construct the final model using a logistic regression algorithm.

#### **Evaluation of model performance**

The predictive performance of each model for PNI risk classification was evaluated by assessing the AUC, accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). The DeLong test was used to compare the AUCs of different models. Calibration curves were plotted to evaluate consistency between observed and predicted values, and decision curve analysis (DCA) was applied to assess the clinical utility of the predictive model.

#### **Model visualization**

To improve transparency in the model's decision-making process, SHapley Additive exPlanations (SHAP) was used to visualize the radiomics model, a game-theoretic approach for model interpretation [26]. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the deep learning model. Gradients from the last convolutional layer of the convolutional neural network were weighted and combined to generate a class activation map, highlighting important regions in target images [27]. A nomogram of the fusion model was also presented for straightforward and efficient use of the model.

The overall workflow of this study is illustrated in Fig. 2.

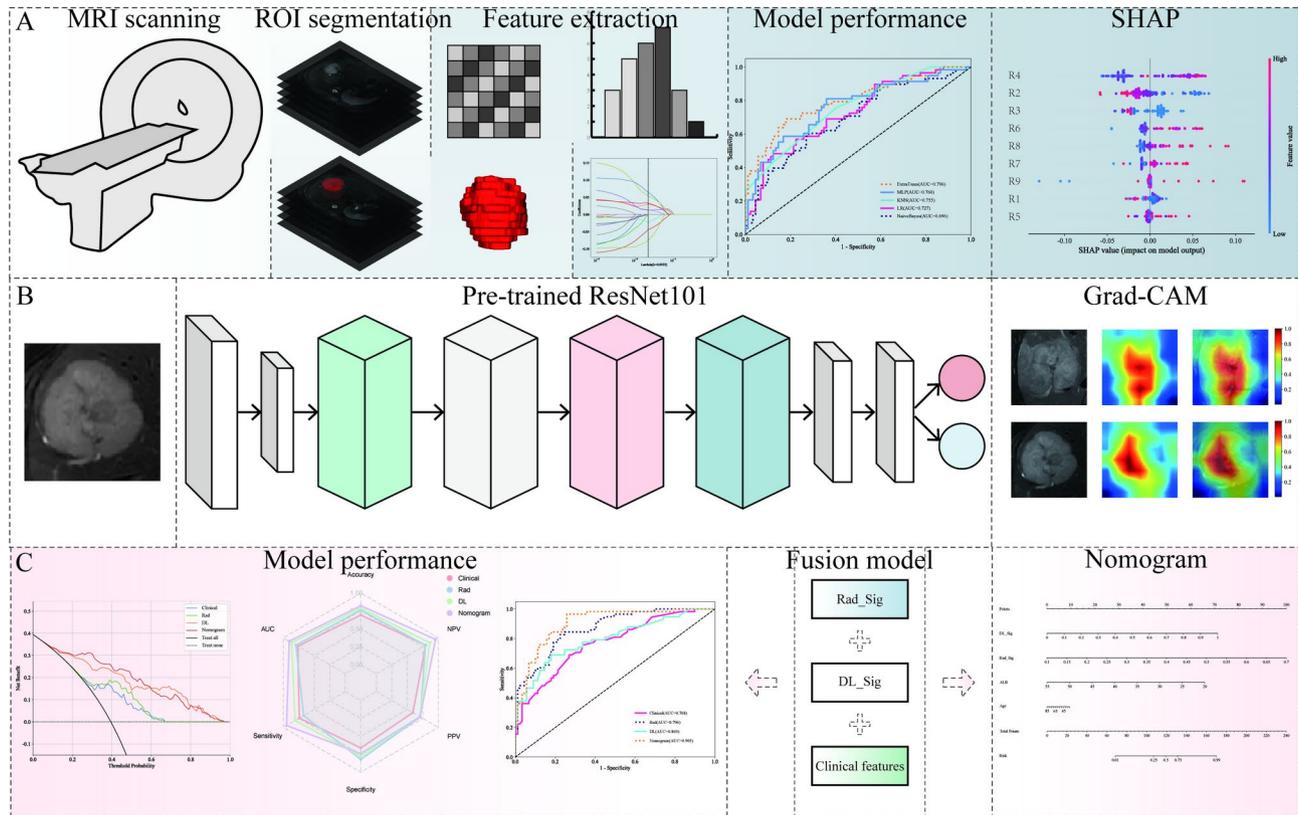
#### **Statistical analysis**

Model construction and evaluation were performed using the pyRadiomics package (version 3.0) and scikit-learn package (version 1.0.2) in Python 3.7.12. Data analysis was conducted in R (version 4.3.2). Continuous variables are described as mean  $\pm$  standard deviation or median (interquartile range), and categorical variables are presented as frequency and percentage. Normal distribution of continuous variables was assessed using the Shapiro–Wilk test. For comparisons between two groups, an independent samples t-test was used for normally distributed variables; otherwise, the Mann–Whitney U test was applied. Comparisons of categorical variables between groups used Fisher's exact test or the chi-square test. A two-sided  $P < 0.05$  was considered statistically significant.

## **Results**

#### **Baseline characteristics**

A total of 192 ICC patients from three centers were included, divided into the PNI-positive group ( $n = 72$ , including 39 cases of center I, 19 of center II, and 14 of center III) and the PNI-negative group ( $n = 120$ , including 60 cases of center I, 29 of center II, and 31 of center III). All included ICC patients were then divided into two cohorts. The training set consisted of 147 patients from Centers I and II, including 58 in the PNI-positive group (39 cases of center I and 19 of center II) and 89 in



**Fig. 2** Workflow diagram of this study. Model development and visualization include: a machine learning model based on MRI radiomics (A), a deep transfer learning model utilizing a pre-trained ResNet101 (B), and a fusion model integrating deep learning with radiomics (C). MRI, magnetic resonance imaging; ROI, region of interest; SHAP, shapley additive explanations; Grad-CAM, gradient-weighted class activation mapping

**Table 1** Baseline characteristics of study sets

Characteristics	Training set <i>n</i> = 147	Test set <i>n</i> = 45	<i>P</i> value
Age (years)	60.0 [53.0;68.0]	62.0 [51.0;69.0]	0.929
Sex			0.689
Female	68 (46.3%)	23 (51.1%)	
Male	79 (53.7%)	22 (48.9%)	
AFP(ng/mL)	3.66 [2.40;6.06]	2.94 [2.15;5.33]	0.239
CEA(ng/mL)	2.85 [1.83;12.2]	4.10 [2.10;9.27]	0.318
CA19-9(U/mL)	99.2 [12.7;1000]	70.3 [14.1;316]	0.701
ALB(g/L)	40.0 [36.9;43.9]	39.1 [34.6;44.0]	0.397
TB(μmol/L)	11.4 [8.75;17.6]	12.0 [9.70;15.9]	0.566
PT(S)	12.0 [11.2;12.9]	12.2 [11.3;13.0]	0.424
APTT(S)	30.3 [26.9;35.2]	33.8 [28.3;37.0]	0.082
PNI			0.403
Yes	58 (39.5%)	14 (31.1%)	
No	89 (60.5%)	31 (68.9%)	

AFP, alpha-fetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; ALB, albumin; TB, total bilirubin; PT, prothrombin time; APTT, activated partial thromboplastin time; PNI, perineural invasion

the PNI-negative group (60 cases of center I and 29 of center II), used for model training. The external test set consisted of 45 patients from Center III, divided into the PNI-positive group (*n* = 14) and the PNI-negative group (*n* = 31), used for external validation of the model.

Table 1 presents the detailed clinical characteristics of both cohorts. No significant differences were observed between the two cohorts in terms of age, sex, AFP, CEA, CA19-9, ALB, TB, PT, or APTT (*P* > 0.05). In the two

cohorts, 58 patients (39.5%) and 14 patients (31.1%) had PNI, respectively ( $P > 0.05$ ).

### Radiomics model and visualization

We extracted a total of 1,197 radiomics features from preoperative MRI images (Fig. 3A, Supplementary Table 1). After removing features with low reproducibility (intra- and interclass correlation coefficients  $< 0.8$ ), a total of 971 radiomics features were selected for t-test or Mann-Whitney U test. Subsequently, correlation analysis was conducted on the remaining 269 features to remove redundancy, resulting in 19 features for LASSO analysis (Fig. 3B, Supplementary Fig. 1). Ultimately, LASSO selected 9 radiomics features, including 3 first-order statistics, 1 shape feature, 4 GLSZM features, and 1 GLRLM feature (Fig. 3C, Supplementary Information 3). Pearson correlation analysis indicated that all correlation coefficients were below 0.7 (Fig. 3D).

Based on the aforementioned 9 radiomics features, we developed five robust supervised models to predict PNI status and compared their performance to determine the optimal model (Table 2). Then, the ExtraTrees model was chosen to construct the radiomics signature, as it outperformed other classifiers in both the training and test sets. In the training set, the ExtraTrees model achieved an AUC of 0.796 (Fig. 3E), with a prediction sample histogram shown in Fig. 3F. In the external test set, the model achieved an AUC of 0.713.

SHAP-based visualization links feature values with their contributions to the model, providing both global and local interpretations of the ExtraTrees model's internal mechanisms. From a global perspective, R4 (wavelet\_LHL\_firstorder\_RootMeanSquared) is the most significant feature in differentiating PNI status, with model output increasing alongside this feature value. The next most important feature is R2 (original\_shape\_Sphericity) (Fig. 4A). At a local level, in a PNI-positive patient, the R4 feature (wavelet\_LHL\_firstorder\_RootMeanSquared) with a value of 3.8598 contributed substantially to the positive assessment (Fig. 4B). For another PNI-negative patient, the R4 feature (wavelet\_LHL\_firstorder\_RootMeanSquared) with a value of -1.9473 contributed negatively (Fig. 4C).

### Deep learning model and visualization

Through transfer learning, we developed a deep learning model based on the ResNet101 framework. The model achieved an AUC of 0.869 on the training set and 0.736 on the external test set (Fig. 5A). Grad-CAM showed that the deep learning model had clearly defined areas of focus during sample classification, primarily concentrated within the tumor's internal region and boundary, with no activation in normal liver tissue areas (Fig. 5B and C, Supplementary Information 4).

### Fusion model and visualization

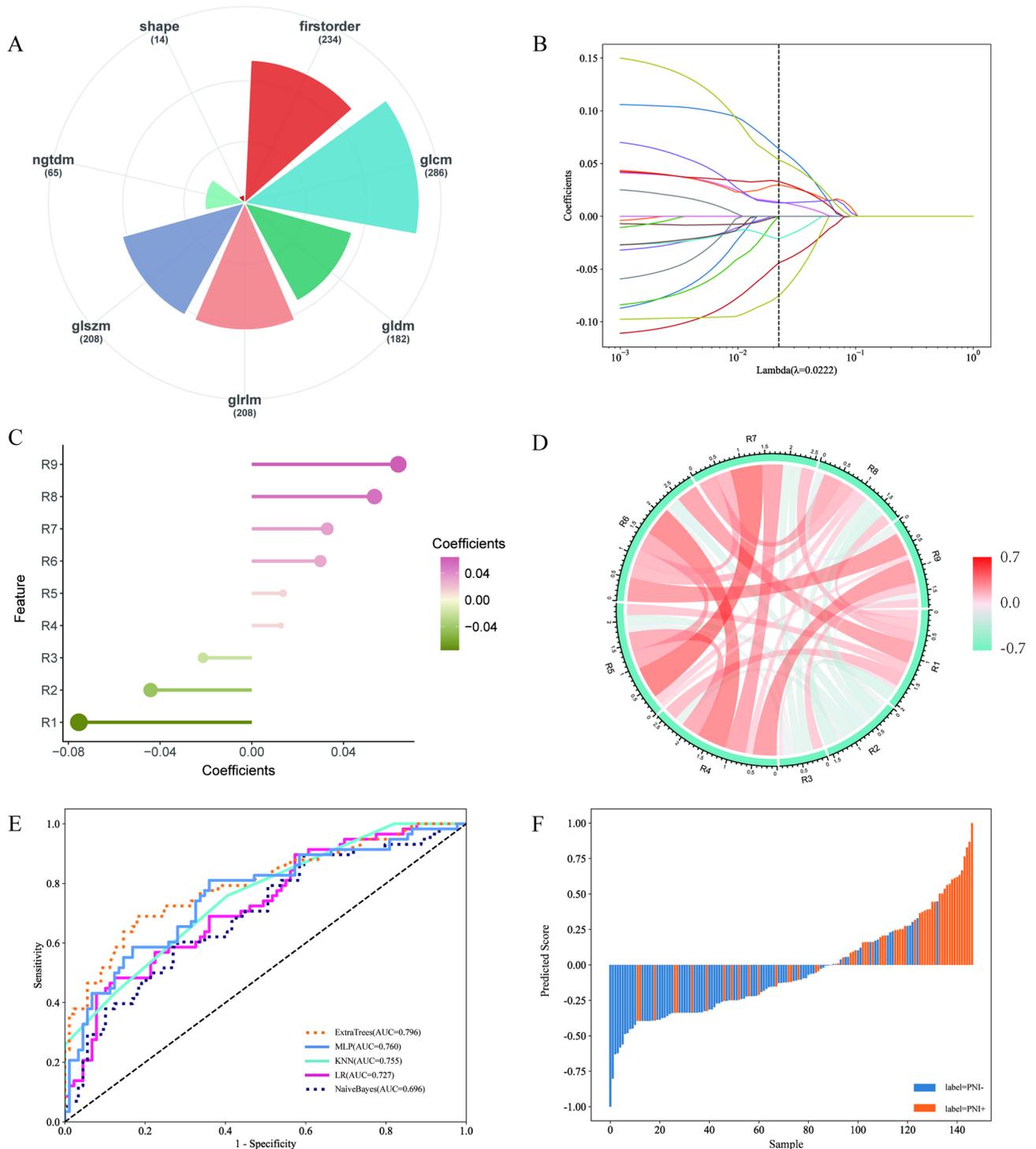
Univariate and multivariate logistic regression analyses of clinical characteristics are presented in Table 3, showing statistically significant effects of age and ALB on PNI outcomes. We constructed a clinical model using the same machine learning model (ExtraTrees) as employed in radiomics signature construction for model comparison (Supplementary Fig. 2).

Subsequently, a fusion model was developed by integrating radiomics and deep learning signatures with age and ALB, visualized through a nomogram (Fig. 6A). In the training set, the fusion model achieved an AUC of 0.905, while in the external test set, the AUC was 0.760 (Fig. 6B). The fusion model demonstrated strong performance across various metrics in both the training set (Fig. 6C) and the test set (Fig. 6D). Compared to the deep learning, radiomics, and clinical models, the fusion model demonstrated significantly superior performance (Supplementary Fig. 3). Calibration curves indicated strong consistency between predicted and observed probabilities for the fusion model (Fig. 6E). Decision curves suggested that the fusion model yielded a favorable net benefit in both the training set (Fig. 6F) and external test set (Fig. 6G).

### Discussion

In this study, we developed a novel fusion model that integrates clinical features, radiomics signature, and deep learning signature based on MRI T2WI to preoperatively predict PNI status in patients with ICC. The model's predictive performance was validated using an independent external cohort, demonstrating its robustness and translational potential.

Current prognostic evaluation of ICC relies on tumor size, number, vascular invasion, and lymph node metastasis, which have been incorporated into the American Joint Committee on Cancer (AJCC) staging system [28]. However, these factors remain insufficient for adequately predicting overall outcomes. Studies indicate that post-operative recurrence rates in ICC patients can reach 60–80% even without vascular invasion or lymph node metastasis, underscoring the need to incorporate additional prognostic factors into ICC staging [29, 30]. As a common pathological feature, PNI is closely associated with prognosis in gastrointestinal malignancies. Emerging evidence suggests PNI serves as a critical prognostic marker for ICC, correlating with poorer outcomes and increased recurrence risk [31]. Zou et al. identified PNI as a robust prognostic indicator for ICC, advocating its inclusion in the AJCC staging system to refine risk stratification [32]. Surgical resection remains the optimal treatment for long-term survival in ICC patients, with PNI and narrow resection margins identified as independent risk factors for survival [5, 33]. This highlights the

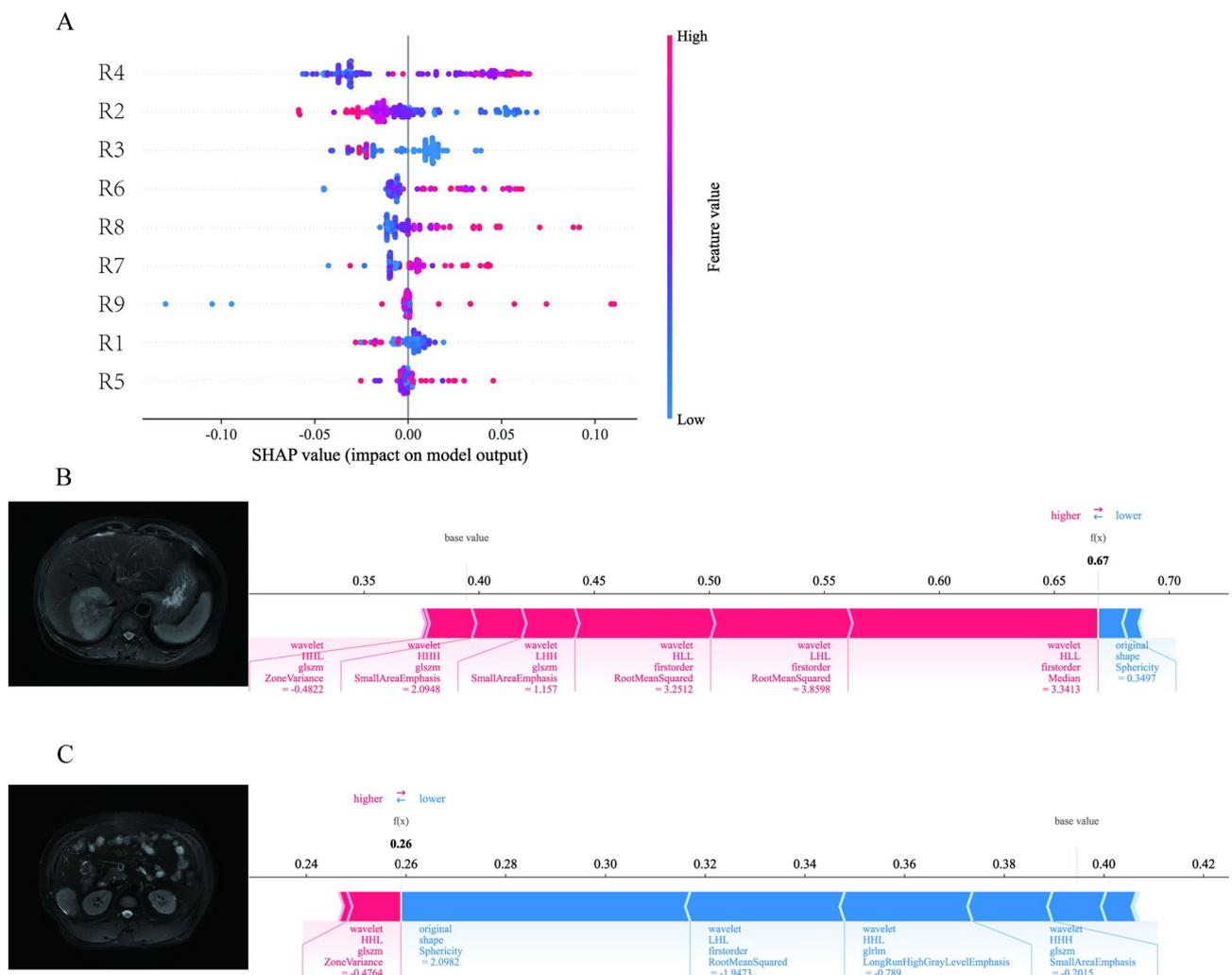


**Fig. 3** Classification of extracted radiomics features (A). Feature selection using LASSO (B). LASSO coefficients (C) and correlation coefficients (D) for the selected 9 features. Comparison of receiver operating characteristic curves across different machine learning models (E). Histogram of sample predictions for the ExtraTrees model (F). *GLCM*, gray-level co-occurrence matrix; *GLDM*, gray-level dependence matrix; *GLRLM*, gray-level run-length matrix; *GLSZM*, gray-level size zone matrix; *NGTDM*, neighboring gray-tone difference matrix; *ExtraTrees*, extremely randomized trees; *MLP*, multilayer perceptron; *KNN*, k-nearest neighbors; *LR*, logistic regression; *AUC*, area under the curve; *PNI*, perineural invasion; *LASSO*, least absolute shrinkage and selection operator

**Table 2** The performance comparison of different machine learning models

Model		AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
LR	Training set	0.727[0.644–0.810]	0.714	0.466	0.876	0.711	0.716
	Test set	0.673[0.512–0.834]	0.511	0.929	0.323	0.382	0.909
NaiveBayes	Training set	0.696[0.609–0.784]	0.667	0.586	0.719	0.576	0.727
	Test set	0.673[0.486–0.859]	0.689	0.643	0.710	0.500	0.815
KNN	Training set	0.755[0.680–0.830]	0.701	0.431	0.876	0.694	0.703
	Test set	0.634[0.467–0.800]	0.667	0.357	0.806	0.455	0.735
ExtraTrees	Training set	0.796[0.720–0.873]	0.762	0.672	0.820	0.709	0.793
	Test set	0.713[0.547–0.879]	0.689	0.643	0.710	0.500	0.815
MLP	Training set	0.760[0.679–0.841]	0.701	0.793	0.640	0.590	0.826
	Test set	0.705[0.543–0.867]	0.600	0.857	0.484	0.429	0.882

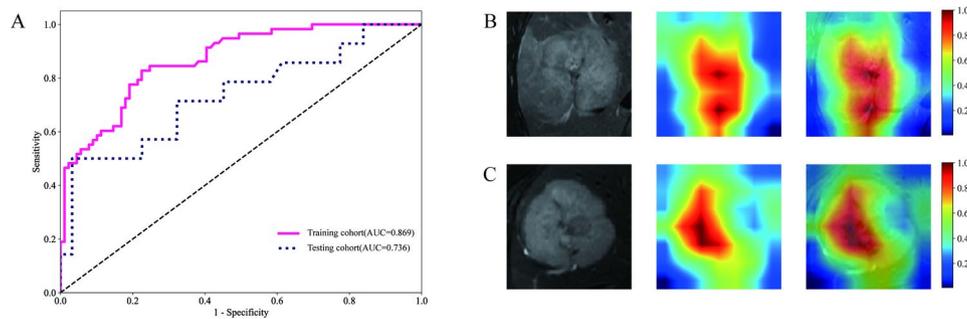
AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; LR, logistic regression; KNN, k-nearest neighbors; ExtraTrees, extremely randomized trees; MLP, multilayer perceptron



**Fig. 4** SHAP summary plot for the ExtraTrees model (A). SHAP force plot illustrating feature contributions for a PNI-positive patient (B) and a PNI-negative patient (C). SHAP, SHapley Additive exPlanations; ExtraTrees, extremely randomized trees; PNI, perineural invasion

importance of preoperative PNI identification for surgical planning. Furthermore, Meng et al. demonstrated that PNI-positive patients exhibited significantly prolonged overall survival following adjuvant chemotherapy

(e.g., capecitabine, gemcitabine, and oxaliplatin), whereas PNI-negative patients showed no significant benefit [34]. These findings emphasize the clinical value of early PNI detection for personalized treatment [34, 35].



**Fig. 5** Receiver operating characteristic curves of the deep learning model for the training and external test sets (A). Attention regions highlighted by the deep learning model for a PNI-positive (B) and a PNI-negative patient (C). *AUC*, area under the curve; *PNI*, perineural invasion

**Table 3** Univariable and multivariable analyses of the training set to assess characteristics associated with PNI in ICC

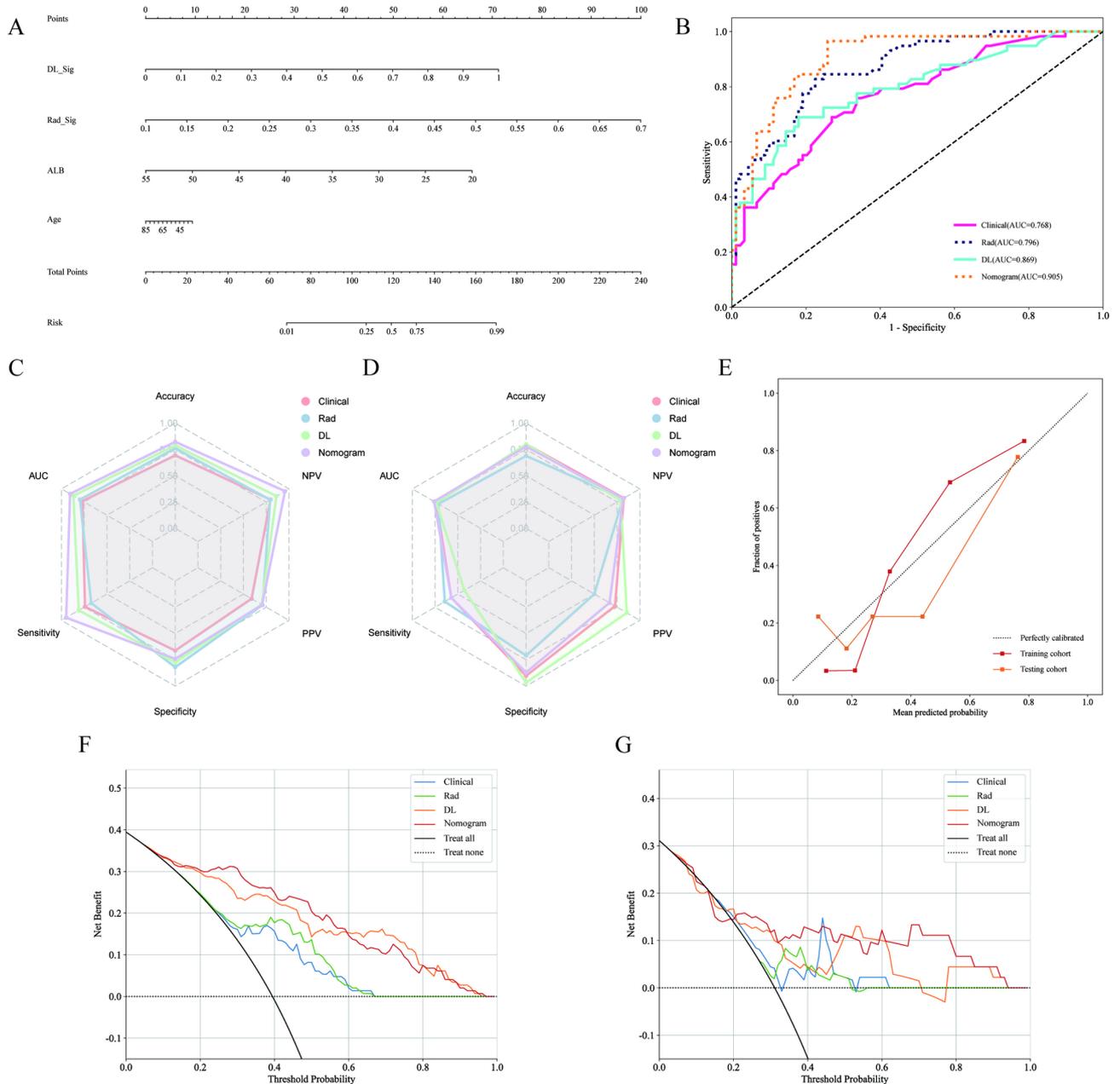
Characteristics	Univariate analysis	<i>P</i> value	Multivariable analysis	<i>P</i> value
	OR (95% CI)		OR (95% CI)	
Age	0.994[0.989–0.998]	0.025	1.043[1.017–1.070]	0.006
Sex	0.580[0.395–0.851]	0.020	0.762[0.424–1.372]	0.447
AFP	1.001[0.998–1.004]	0.476		
CEA	0.999[0.994–1.003]	0.640		
CA19-9	1.000[1.000–1.000]	0.212		
ALB	0.987[0.980–0.994]	0.003	0.929[0.895–0.964]	0.001
TB	1.002[0.998–1.007]	0.376		
PT	0.976[0.956–0.996]	0.047	0.998[0.957–1.040]	0.923
APTT	0.987[0.978–0.995]	0.012	1.003[0.961–1.047]	0.918

*PNI*, perineural invasion; *ICC*, intrahepatic cholangiocarcinoma; *OR*, odds ratio; *CI*, confidence interval; *AFP*, alpha-fetoprotein; *CEA*, carcinoembryonic antigen; *CA19-9*, carbohydrate antigen 19–9; *ALB*, albumin; *TB*, total bilirubin; *PT*, prothrombin time; *APTT*, activated partial thromboplastin time

Nevertheless, current PNI diagnosis still relies on post-operative histopathology, revealing an urgent need for noninvasive preoperative prediction.

We utilized MRI data from ICC patients to develop both individual and fusion models capable of radiologically identifying patients with histological evidence of PNI. A previous study reported that ICC patients with PNI often exhibit distinctive tumor microenvironment characteristics, including decreased natural killer cell infiltration, increased neutrophil infiltration, elevated immune checkpoint ligand expression, and higher KRAS mutation frequency [34]. Owing to its superior soft tissue resolution, MRI has significant value in differentiating liver tumors and predicting pathological subtypes [36]. Meanwhile, radiomics analysis can extract quantitative features from imaging data, thereby revealing potential tumor-related biological information. In this study, we first constructed a model based on clinical features, which achieved an AUC of 0.768, similar to the ICC PNI prediction model built by Wang et al. using clinical data [37]. This outcome may be attributed to the omission of imaging features. Subsequently, we developed five radiomics-based machine learning models using MRI data, among which the ExtraTrees model yielded an AUC of 0.796 in the training set, surpassing the model built solely on clinical features. As an ensemble learning

method, ExtraTrees enhances the model's generalization performance by introducing additional randomness, thereby reducing overfitting risk and effectively handling high-dimensional data and complex nonlinear relationships [38]. Interestingly, the primary radiomics features we identified were derived from wavelet transformations, suggesting that such features may possess unique value in characterizing tumor heterogeneity and warrant further exploration regarding their potential links to the tumor microenvironment. Consistent with our findings, a previous ICC PNI prediction model that integrated clinical features with CT-based radiomics also demonstrated favorable performance, with an AUC reaching 0.884 [39]. Although radiomics techniques reduce human measurement errors to a certain extent, they rely on manually designed feature extraction, which may be insufficient to fully capture the complex heterogeneity of tumors, thus limiting predictive performance and clinical applicability [40]. In contrast, deep learning employs end-to-end feature learning that does not require manually designed features, enabling it to automatically extract complex patterns directly from raw images [40, 41]. As an essential branch of machine learning, deep learning leverages multiple layers of nonlinear transformations to learn data features, endowing it with robust representational capabilities [41]. In the field of medical image analysis, CNNs



**Fig. 6** Nomogram of the fusion model (A). Comparison of receiver operating characteristic curves for the clinical, radiomics, deep learning, and fusion models (B). Performance metrics for each model on the training (C) and test sets (D), respectively. Calibration curves for the fusion model (E). Decision curves for the training (F) and test sets (G), respectively. *DL\_Sig*, deep learning signature; *Rad\_Sig*, radiomics signature; *ALB*, albumin; *DL*, deep learning; *Rad*, radiomics; *AUC*, area under the curve; *PPV*, positive predictive value; *NPV*, negative predictive value

have been extensively applied to detect and classify various diseases [19]. In this study, we selected a ResNet-101 model and performed pretraining on the ImageNet dataset to leverage transfer learning. The results showed that the deep learning model achieved an AUC of 0.869, outperforming the radiomics model and further underscoring the potential of deep learning in PNI prediction. Notably, fusion models often compensate for the limitations of individual models and yield superior performance, aligning with our findings [42]. Our fusion model,

which integrated radiomics signature, deep learning signature, and two selected clinical features, ultimately achieved the best performance in PNI prediction.

As machine learning models become increasingly complex, the “black box” effect becomes more pronounced, making it difficult to understand the underlying mechanisms and decision rationales. This issue is particularly critical in the high-risk field of medicine, where merely demonstrating strong performance metrics does not guarantee clinical utility [43]. Providing reasonable

explanations for model outputs not only facilitates an understanding of the decision-making process but also helps identify potential biases and errors [44]. In this study, we employed the SHAP method to analyze feature differences between patients with positive versus negative PNI, thereby assisting clinicians in more intuitively understanding the model's decision logic. SHAP analysis not only revealed the impact of individual features on the prediction outcomes but also evaluated the importance distribution of these features at a global level. The results indicated that R4 (Wavelet\_LHL\_firstorder\_Root-MeanSquared) was identified as the most predictive radiomics feature. In addition, we utilized Grad-CAM to visualize the decision-making process of the CNN. Compared with other visualization methods, Grad-CAM retains spatial resolution while more accurately capturing high-level semantic information, offering robust support for medical image analysis [27]. In our study, we analyzed the final convolutional feature maps of the deep learning model. The Grad-CAM visualization indicated that the model primarily focused on the internal and boundary regions of the tumor, both of which are crucial for distinguishing patients with positive or negative PNI. To further enhance clinical applicability, we also constructed a nomogram for the fusion model, thereby facilitating the incorporation of these visualization and interpretability findings into real-world clinical decision-making.

Drawing on our clinical experience, the fusion model developed in this study can assist clinicians in accurately predicting PNI status in ICC patients prior to surgery. The implementation of this approach may change ICC treatment strategies and potentially improve long-term patient outcomes. However, the clinical application of the current model remains relatively complex. In the future, developing user-friendly clinical applications could facilitate its broader adoption in real-world practice. Simultaneously, further prospective studies are warranted to compare various treatment regimens based on preoperative assessments of PNI status in ICC patients, thereby providing evidence-based guidance for individualized treatment strategies.

Our study has several limitations. First, although we used multicenter data for model training and validation, the number of enrolled cases remained limited due to the low incidence of ICC. Second, the retrospective design of this study may introduce selection bias. Third, the potential relationship between radiomics features and the tumor microenvironment has not been fully elucidated, requiring further mechanistic investigation. Moreover, future studies could incorporate multi-modal imaging data such as ultrasound and CT to enhance the predictive accuracy of the models.

## Conclusion

This study developed a fusion model based on MRI images for preoperative prediction of PNI status in patients with ICC. This model, integrating radiomics and deep learning signatures along with clinical features, achieved the highest performance among all models. It may aid in guiding clinical decisions related to preoperative staging and adjuvant treatment, supporting personalized and precise therapy for ICC patients and enhancing patient outcomes.

## Abbreviations

AFP	Alpha-fetoprotein
AJCC	American Joint Committee on Cancer
ALB	Albumin
APTT	Activated partial thromboplastin time
AUC	Area under the curve
CA19-9	Carbohydrate antigen 19–9
CEA	Carcinoembryonic antigen
CI	Confidence interval
CNNs	Convolutional neural networks
DCA	Decision curve analysis
DICOM	Digital Imaging and Communications in Medicine
ExtraTrees	Extremely randomized trees
GLCM	Gray-level co-occurrence matrix
GLDM	Gray-level dependence matrix
GLRLM	Gray-Level run-length matrix
GLSZM	Gray-level size zone matrix
GRAD-CAM	Gradient-weighted class activation mapping
IBSI	Imaging Biomarker Standardization Initiative
ICC	Intrahepatic cholangiocarcinoma
KNN	K-nearest neighbors
LASSO	Least absolute shrinkage and selection operator
LR	Logistic regression
MLP	Multilayer perceptron
MRI	Magnetic resonance imaging
NGTDM	Neighboring gray-tone difference matrix
NPV	Negative predictive value
OR	Odds ratio
PNI	Perineural invasion
PPV	Positive predictive value
PT	Prothrombin time
ROI	Region of interest
SGD	Stochastic gradient descent
SHAP	SHapley Additive exPlanations
T2WI	T2-weighted imaging
TB	Total bilirubin

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-025-03819-w>.

Supplementary Material 1

## Author contributions

Zc Q and Dy L conceived and designed the study. Zc Q, H Y, Py C, Kl C, B M, and Pg N collected and collated the data. Qs L and Dx L performed the statistical analysis. Zc Q drafted the manuscript. Hb Y and Dy L contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content. All authors reviewed the paper and approved the final version.

## Funding

This work was supported by the Henan Province Key Research and Development and Promotion Special Project (222102310709) and the Henan Province Middle-aged and Young Health and Medical Science and Technology

Innovation Leading Talent Cultivation Program (YXKC2022002). Additionally, it was supported by the National Natural Science Foundation of China (Grant Nos. 82103618 and 82103617).

#### Data availability

The datasets analysed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

The study protocol was conducted in accordance with the Declaration of Helsinki Ethical Guidelines and was approved by the Institutional Review Boards of Zhengzhou University People's Hospital (Ref No. 2023–012), Henan Cancer Hospital (Ref No. 2023–203), and the First Affiliated Hospital of Zhengzhou University (Ref No. 2021-KY-1137–002). All patients enrolled in the study were informed and provided informed consent.

##### Consent for publication

All authors and participants have agreed to the publication of the results of this study.

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Department of Hepatobiliary and Pancreatic Surgery, Zhengzhou University People's Hospital, Zhengzhou 450003, China

<sup>2</sup>Department of Hepatobiliary and Pancreatic Surgery, Henan University People's Hospital, Zhengzhou 450003, China

<sup>3</sup>Department of Gastroenterology, Zhengzhou University People's Hospital, Zhengzhou 450003, China

<sup>4</sup>Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

<sup>5</sup>Department of Hepatobiliary and Pancreatic Surgery, Henan Cancer Hospital, Zhengzhou 450003, China

<sup>6</sup>Department of Radiology, Zhengzhou University People's Hospital, Zhengzhou 450003, China

Received: 31 January 2025 / Accepted: 15 April 2025

Published online: 26 April 2025

#### References

1. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet*. 2014;383:2168–79.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
3. Florio AA, Ferlay J, Znaor A, Ruggieri D, Alvarez CS, Laversanne M, et al. Global trends in intrahepatic and extrahepatic cholangiocarcinoma incidence from 1993 to 2012. *Cancer*. 2020;126:2666–78.
4. Khan SA, Genus T, Morement H, Murphy A, Rous B, Tataru D. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *J Hepatol*. 2019;71:1261–2.
5. DeOliveira ML, Cunningham SC, Cameron JL, Kamangar F, Winter JM, Lillemoe KD, et al. Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. *Ann Surg*. 2007;245:755–62.
6. Chun YS, Javle M. Systemic and adjuvant therapies for intrahepatic cholangiocarcinoma. *Cancer Control*. 2017;24:594–9.
7. Ribero D. Surgical approach for long-term survival of patients with intrahepatic cholangiocarcinoma: a multi-institutional analysis of 434 patients. *Arch Surg*. 2012;147:1107.
8. Liebig C, Ayala G, Wilks JA, Berger DH, Albo D. Perineural invasion in cancer: A review of the literature. *Cancer*. 2009;115:3379–91.
9. Meng F, Jing X, Song G, Jie L, Shen F. Prox1 induces new lymphatic vessel formation and promotes nerve reconstruction in a mouse model of sciatic nerve crush injury. *J Anat*. 2020;237:933–40.
10. Hassan MO, Maksem J. The prostatic perineural space and its relation to tumor spread: an ultrastructural study. *Am J Surg Pathol*. 1980;4:143–8.
11. Wei T, Zhang X-F, He J, Popescu I, Marques HP, Aldrighetti L, et al. Prognostic impact of perineural invasion in intrahepatic cholangiocarcinoma: multicentre study. *Br J Surg*. 2022;109:610–6.
12. Harisinghani MG, O'Shea A, Weissleder R. Advances in clinical MRI technology. *Sci Transl Med*. 2019;11:eaba2591.
13. Granata V, Grassi R, Fusco R, Setola SV, Belli A, Ottiano A, et al. Intrahepatic cholangiocarcinoma and its differential diagnosis at MRI: how radiologist should assess MR features. *Radiol Med*. 2021;126:1584–600.
14. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout R, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
15. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30:1234–48.
16. Xu Y, Ye F, Li L, Yang Y, Ouyang J, Zhou Y, et al. MRI-based radiomics nomogram for preoperatively differentiating intrahepatic mass-forming cholangiocarcinoma from resectable colorectal liver metastases. *Acad Radiol*. 2023;30:2010–20.
17. Zhao L, Ma X, Liang M, Li D, Ma P, Wang S, et al. Prediction for early recurrence of intrahepatic mass-forming cholangiocarcinoma: quantitative magnetic resonance imaging combined with prognostic immunohistochemical markers. *Cancer Imaging*. 2019;19:49.
18. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313:504–7.
19. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology*. 2019;290:590–606.
20. Nakatsuka T, Tateishi R, Sato M, Hashizume N, Kamada A, Nakano H, et al. Deep learning and digital pathology powers prediction of HCC development in steatotic liver disease. *Hepatology*. 2025;81:976–89.
21. Xi IL, Wu J, Guan J, Zhang PJ, Horii SC, Soulen MC, et al. Deep learning for differentiation of benign and malignant solid liver lesions on ultrasonography. *Abdom Radiol*. 2021;46:534–43.
22. Yin L, Liu R, Li W, Li S, Hou X. Deep learning-based CT radiomics predicts prognosis of unresectable hepatocellular carcinoma treated with TACE-HAIC combined with PD-1 inhibitors and tyrosine kinase inhibitors. *BMC Gastroenterol*. 2025;25:24.
23. Beare R, Lowekamp B, Yaniv Z. Image segmentation, registration and characterization in R with simpleitk. *J Stat Softw*. 2018;86:1–35.
24. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:328–38.
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;770–778.
26. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In *Adv Neural Inf Process Syst*. 2017;30:4768–77.
27. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2020;128:336–59.
28. Kim Y, Moris DP, Zhang X, Bagante F, Spolverato G, Schmidt C, et al. Evaluation of the 8th edition American joint commission on cancer (AJCC) staging system for patients with intrahepatic cholangiocarcinoma: a surveillance, epidemiology, and end results (SEER) analysis. *J Surg Oncol*. 2017;116:643–50.
29. Meng Z-W, Pan W, Hong H-J, Chen J-Z, Chen Y-L. Macroscopic types of intrahepatic cholangiocarcinoma and the eighth edition of AJCC/UICC TNM staging system. *Oncotarget*. 2017;8:101165–74.
30. Cheng Z, Lei Z, Jin X, Zhang Q, Si A, Yang P, et al. Postoperative adjuvant transarterial chemoembolization for intrahepatic cholangiocarcinoma patients with microvascular invasion: a propensity score analysis. *J Gastrointest Oncol*. 2021;12:819–30.
31. Shirai K, Ebata T, Oda K, Nishio H, Nagasaka T, Nimura Y, et al. Perineural invasion is a prognostic factor in intrahepatic cholangiocarcinoma. *World J Surg*. 2008;32:2395–402.
32. Zou M, Sheng J, Ruan M, Zhou W, Ye F, Yang G, et al. Perineural invasion confers poorer clinical outcomes in patients with T1/T2 intrahepatic cholangiocarcinoma: a single center, retrospective cohort study. *J Gastrointest Oncol*. 2023;14:2500–10.
33. Yoo T, Park S-J, Han S-S, Kim SH, Lee SD, Kim Y-K, et al. Postoperative CA19-9 change is a useful predictor of intrahepatic cholangiocarcinoma survival following liver resection. *Dis Markers*. 2015;2015:1–7.

34. Meng X-L, Lu J-C, Zeng H-Y, Chen Z, Guo X-J, Gao C, et al. The clinical implications and molecular features of intrahepatic cholangiocarcinoma with perineural invasion. *Hepatol Int.* 2023;17:63–76.
35. Fisher SB, Patel SH, Kooby DA, Weber S, Bloomston M, Cho C, et al. Lymphovascular and perineural invasion as selection criteria for adjuvant therapy in intrahepatic cholangiocarcinoma: a multi-institution analysis. *Hpb.* 2012;14:514–22.
36. Park S, Lee Y, Kim H, Yu MH, Lee ES, Yoon JH, et al. Subtype classification of intrahepatic cholangiocarcinoma using liver MR imaging features and its prognostic value. *Liver Cancer.* 2022;11:233–46.
37. Wang M, Sun Z, Wang Q, Guo J, Li X, Chen J, et al. Establishment and validation of a nomogram model for preoperative prediction of the risk of cholangiocarcinoma with perineural invasion. *Am J Cancer Res.* 2023;13:5082–93.
38. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* 2006;63:3–42.
39. Liu Z, Luo C, Chen X, Feng Y, Feng J, Zhang R, et al. Non-invasive prediction of perineural invasion in intrahepatic cholangiocarcinoma by clinicoradiological features and computed tomography radiomics based on interpretable machine learning: a multicenter cohort study. *Int J Surg.* 2024;110:1039–51.
40. Zhou Z. Artificial intelligence on MRI for molecular subtyping of diffuse gliomas: feature comparison, visualization, and correlation between radiomics and deep learning. *Eur Radio.* 2022;32:745–6.
41. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–44.
42. Wang W, Liang H, Zhang Z, Xu C, Wei D, Li W, et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. *Eclinicalmedicine.* 2024;67:102385.
43. Yoon CH, Torrance R, Scheinerman N. Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned? *J Med Ethics.* 2022;48:581–5.
44. Yamasawa D, Ozawa H, Goto S. The importance of interpretability and validations of Machine-Learning models. *Circ J.* 2023;88:157–8.

### Publisher's note

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