



Artificial intelligence CT radiomics to predict early recurrence of intrahepatic cholangiocarcinoma: a multicenter study

Yangda Song^{1,2} · Guangyao Zhou^{1,3} · Yucheng Zhou⁴ · Yikai Xu⁵ · Jing Zhang⁵ · Ketao Zhang⁶ · Pengyuan He^{1,7} · Maowei Chen⁸ · Yanping Liu^{1,9} · Jiarun Sun^{1,2} · Chengguang Hu¹ · Meng Li^{1,2} · Minjun Liao^{1,2} · Yongyuan Zhang¹⁰ · Weijia Liao¹¹ · Yuanping Zhou^{1,2}

Received: 9 November 2022 / Accepted: 13 January 2023 / Published online: 23 February 2023
© Asian Pacific Association for the Study of the Liver 2023

Abstract

Objectives In this multicenter study, we sought to develop and validate a preoperative model for predicting early recurrence (ER) risk after curative resection of intrahepatic cholangiocarcinoma (ICC) through artificial intelligence (AI)-based CT radiomics approach.

Materials and methods A total of 311 patients (Derivation: 160; Internal and two external validations: 36, 74 and 61) from 8 medical centers who underwent curative resection were collected retrospectively. In derivation cohort, radiomics and clinical–radiomics models for ER prediction were constructed by LightGBM (a machine learning algorithm). A clinical model was also developed for comparison. Model performance was validated in internal and two external cohorts by ROC. In addition, we investigated the interpretability of the LightGBM model.

Results The combined clinical–radiomics model that included 15 radiomic features and 3 clinical features (CA19-9 > 1000 U/ml, vascular invasion and tumor margin), resulting in the area under the curves (AUCs) of 0.974 (95% CI 0.946–1.000) in the derivation cohort, and 0.871–0.882 (95% CI 0.672–0.962) in the internal and external validation cohorts, respectively, which are higher than the AJCC 8th TNM staging system (AUCs: 0.686–0.717, p all < 0.05). Especially, the sensitivity of this machine learning model could reach 94.6% on average for all the cohorts.

Conclusions This AI-driven combined radiomics model may provide as a useful tool to preoperatively predict ER and improve therapeutic management of ICC patients.

Keywords Machine learning · Intrahepatic cholangiocarcinoma · Recurrence · Radiomics · CT · Multicenter study · LightGBM · Prediction model · Biliary tract cancer · Recurrence-free survival

Abbreviations

| | |
|------|--|
| AUC | Area under the curve |
| ER | Early recurrence |
| ICC | Intrahepatic cholangiocarcinoma |
| mRMR | The maximum relevance minimum redundancy |
| RFS | Recurrence-free survival |

| | |
|------|-----------------------------------|
| ROC | Receiver operator characteristic |
| SHAP | The Shapley Additive exPlanations |
| VOI | Volume of interest |

Introduction

Cholangiocarcinoma comprises diverse groups of malignant tumors that appear in biliary system and can be anatomically classified into intrahepatic, perihilar and distal cholangiocarcinoma [1]. Although not as common as hepatocellular carcinoma, detected incidence of cholangiocarcinoma, especially intrahepatic cholangiocarcinoma (ICC), has been on rising in many countries, and has become the second most common liver malignancy [2]. Unlike extrahepatic cholangiocarcinoma, ICC rarely causes jaundice and is also usually asymptomatic at early stage [3], which makes early

Yangda Song and Guangyao Zhou have contributed equally to this work.

- ✉ Yongyuan Zhang
yongyuanzhang@hbvtech.com
- ✉ Weijia Liao
liaoweijia288@163.com
- ✉ Yuanping Zhou
yuanpingzhou@163.com

Extended author information available on the last page of the article

ICC diagnosis difficult. Consequently, only about 30–40% of patients can undergo curative surgical resection [4]. Furthermore, 40–80% of patients experienced recurrence with 44% of them detected in the first year after surgery [5]. In addition, ICC is usually more aggressive and frequently refractory to chemotherapy than HCC, making ICC a deadly cancer, resulting in 5-year survival rate below 30% after surgery [6]. Therefore, an accurate preoperative assessment of ICC patients' risk of recurrence will help to improve clinical management and prognosis of patients with ICC.

Currently, the AJCC 8th TNM staging system has been used for risk stratification of ICC [7]. However, in clinical practice, postoperative recurrence and prognosis differ markedly even among patients classified at the same stages [6, 8]. The use of existing clinical factors did play a positive role in patient stratification [9]. However, due to the high heterogeneity of ICC, the current clinical staging system still has many limitations, which prompted clinicians to seek more effective predictive tools.

Dynamic enhanced CT is routinely used for detection and evaluation of ICC. However, traditionally, the diagnosis and prognosis assessment are largely subjected to the physician's personal interpretation of images, which inevitably carries non-objectiveness and makes quantitative comparison difficult [10]. Moreover, with the busy workload of radiologists, relying solely on human's eyes sometimes inevitably leads to oversights. In contrast, radiomics is a novel and integrated discipline that combines artificial intelligence with imaging analysis, which has been increasingly used in oncology practice, such as tumor diagnosis, staging and assessment [11, 12], and is popular for its accessibility and non-invasiveness.

Despite the potential of radiomics, some investigators have urged the field to expand studying the reproducibility and generalizability of models generated, which is also a point that must be considered although radiomics has great potential, some researchers have called for the field to broaden its focus to include investigating the reproducibility and generalizability of the models it generates. This

is an important consideration for the practical application of radiomics studies [13, 14]. Since ICC is an uncommon tumor, most of the previous studies were single center based and reported a relatively small number of cases (for instance < 200) [15, 16]. Therefore, a multicenter study with a large sample is necessary. Moreover, different CT instruments and parameters utilized by different centers pose a great challenge to develop an accurate and reproducible prediction system. In this study, we collected large cohorts of ICC data from different regional centers in China and sought to develop and validate a preoperative model for predicting risk of early recurrence (ER) within 1 year after curative resection of ICC through artificial intelligence (AI)-based CT radiomics approach.

Materials and methods

Patients

The study was reported according to the STROBE guidelines. This multicenter study included ICC patients who underwent radical surgical resection between January 2008 and December 2020 at eight hospitals in China. Two inclusion criteria applied for enrollment: (1) contrast-enhanced CT examination was performed within 1 month before surgery; (2) patients underwent surgery with curative intent and ICC was confirmed by postoperative pathology. Patients with following characteristics were excluded: (1) lack of preoperative CT data; (2) previous history of other ICC treatments (including interventional and ablation, etc.); (3) history of other malignancies; (4) lost to follow-up, and whether recurrence occurs within 1 year after surgery was unknown. A total of 311 cases were included in the study (Fig. 1). Patients were divided into three cohorts and each of which was located within each province territory (detailed comprising hospitals were listed in Supplementary Materials). Finally, a derivation cohort was established within

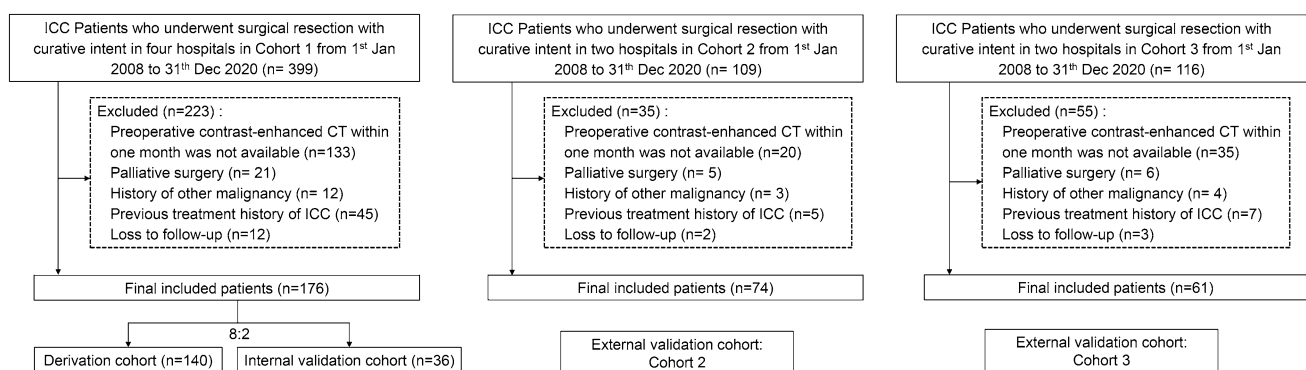


Fig. 1 Flowchart of the enrolled study patients

the Cohort 1 and randomly divided into a training group ($n = 140$) and an internal validation group ($n = 36$) at a ratio of 8:2, and both Cohort 2 ($n = 74$) and Cohort 3 ($n = 61$) served as two independent external validation cohorts.

Clinical data and assessments

Following clinical and laboratory data before surgery were collected: basic demographic information, serum levels of alanine transaminase (ALT), aspartate transaminase (AST), albumin, total bilirubin, TNM tumor staging (8th edition), concurrent hepatobiliary system disease (viral hepatitis, cirrhosis, biliary fluke infection, primary sclerosing cholangitis and cholelithiasis), alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9). The threshold values for ALT, AST, AFP, CA19-9, and CEA used here were 35, 40 U/L, 20 ng/ml, 1000 u/ml, and 5 ng/L, respectively.

Radiographic data and assessments

Preoperative dynamic-enhanced CT (both unenhanced and arterial enhancement phases) images were collected. Details of CT acquisition and parameters are described in (Supplementary Table S1). Two experienced radiologists (JZ and YKX) independently documented tumor number, size, margins, vascular invasion, and lymph node metastases (see supplementary materials for detailed definitions). When disagreement presented, a consensus was made through discussion with a third radiologist.

Image segmentation and feature extraction

Two radiologists (JZ and YKX, both with more than 15 years of experience in abdominal imaging), who were blinded to clinical outcomes, independently performed segmentation of the tumor layer-by-layer, resulting in 3D volume of interest (VOI) using 3D Slicer (version 4.10.2). Then, a peri-tumoral area through a uniform expansion of 5 mm outward from an entire VOI edge was calculated and generated by 3D Slicer (Supplementary Fig S1). The largest one was selected for analysis if there were multiple tumors. To ensure the stability of the extracted features, the inter-observer reproducibility of the extracted features between two readers was evaluated. And images from 50 cases in the derivation cohort were then randomly selected and the same segmentation procedure was repeated 2 weeks later by the same two radiologists to assess the intra-observer reproducibility. Inter-observer and intra-observer reproducibility were evaluated by calculating the intraclass correlation coefficient.

All voxel of images were resampled at a size of $1 \times 1 \times 1 \text{ mm}^3$ for standardization to establish comparable and reproducible radiomics analysis. Then, CT features were

extracted from the normalized images and filtered images under Wavelet and Laplacian of Gaussian (LoG) filters using Pyradiomics (version 3.0). The detailed parameters were listed in Supplementary materials. A total of 6296 features covering tumoral and peri-tumoral areas for each case were finally obtained. The features were classified to include seven categories: (1) Shape; (2) First-order statistics; (3) Gray Level Cooccurrence Matrix; (4) Gray Level Run Length Matrix; (5) Gray Level Size Zone Matrix; (6) Neighboring Gray Tone Difference Matrix; (7) Gray Level Dependence Matrix [17].

Feature selection

The features were first normalized by the z-score method to eliminate differences in scale. Then, the procedure of feature selection in the derivation cohort consists of three steps: (1) robust features with the intraclass correlation coefficients (both inter-observer and intra-observer) higher than 0.75 were retained; (2) features were ranked by the maximum relevance minimum redundancy (mRMR) algorithm, and the top 100 features were retained to select the features with the highest correlation and minimum redundancy; (3) recursive feature elimination and cross-validation were performed to determine the best feature subset.

Clinical features (including laboratory tests and conventional imaging features) were selected through univariable and multivariable logistic regression for subsequent model building.

Model construction and validation

The selected features were analyzed by the LightGBM machine learning algorithm to construct two models: (1) Radiomics model; and (2) Combined clinical–radiomics model. A clinical model was also developed by the clinical factors selected by the above univariable and multivariable analysis. Discrimination was quantified with the area under the curve (AUC). Prediction accuracy was evaluated by calibration curves, confusion matrix and Hosmer–Lemeshow test [18]. Net reclassification index and integrated discrimination improvement were used to evaluate the incremental differences of the combined clinical–radiomics models. The LightGBM model was demonstrated using the Shapley Additive exPlanations (SHAP) to visualize the correlation between variables and ER. Patients in each cohort were stratified according to the risk predicted by the combined radiomics model and plotted by Kaplan–Meier analysis. Additionally, we also tried several different machine learning classifiers and found that LightGBM performed the best (Supplementary Table S2). Finally, we compared the combined clinical–radiomics model with AJCC 8th TNM system in predicting ER. Evaluation of TNM tumor staging model

was derived using the classical logistic regression model based on different staging grade variables.

Follow-up

Patients were followed up postoperatively at each center per the guideline recommendations [19–22]. Clinical assessment, and CT or MRI examinations were performed every 3 months during the first year after surgery, and at 6 months interval thereafter. Considering > 40% of ICC patients may experience recurrence in the first year after surgery [5], ER in this study was defined as recurrence within 1 year after surgery. Locations of the recurrence were classified as intrahepatic only, extrahepatic only, or concurrent intra- and extrahepatic for the purposes of analysis. Recurrence-free survival (RFS) was defined as interval between curative-intent surgery and the date of tumor recurrence or the last follow-up date (without recurrence).

Statistical analysis

Statistical analyses were performed by SPSS (version 26), R software (version 3.6.2) and Python (version 3.8). For normal distribution, data are expressed as means \pm SD and analyzed by one-way ANOVA. For abnormal distribution, data are described as medians [interquartile ranges, IQR] and analyzed by the Kruskal–Wallis *H* test. Comparison between different ROC curves was performed by Delong test [23]. mRMR analysis was performed using the ‘mRMRe’ package of R software. The rest of the feature selection and model construction were performed using the Python-based ‘scikit-learn’ machine learning framework [24]. ROC curves, confusion matrices and radar plots were plotted by the Python-based ‘Matplotlib’ library. For all tests, $p < 0.05$ was considered statistically significant.

Results

Patient characteristics

A total of 311 ICC patients from 8 different hospitals covering the southern and eastern China were included in this study. Summary of the demographics and clinical baseline characteristics of all patients is listed in Table 1. The incidence of cholelithiasis was relatively more common in the Cohort 3 (65.6% (40/61)), accompanied by relatively high bilirubin levels. The remaining clinical features were not statistically different among centers. In addition, owing to the lack of data on liver flukes and primary sclerosing cholangitis in more than half of the retrospective cases, they were not further included in the analysis. Most patients (87.1% (271/311)) received R0 resection with no significant

differences between cohorts. After a median follow-up time of 50 weeks [IQR 28–106 weeks], the median RFS time was approximately 45–50 weeks, and there was no difference in the ER rates (ranging from 47.2 to 54.1%) among the four cohorts.

Feature selection and development of radiomics model

Of the 6296 radiomic features, 528 of them were deemed unreliable (intra- or inter-class correlation coefficient < 0.75) and excluded, and then the remaining features were ranked by the mRMR algorithm to retain the top 100 features (Supplementary Fig S2A). The final 15 radiomic features were generated after subjecting to recursive feature elimination + cross-validation (Supplementary Fig S2B). Radiomics model was developed using the tree kernel-based LightGBM algorithm, showing AUCs values of 0.877 (95% CI 0.816–0.938), 0.780 (95% CI 0.627–0.934), 0.751 (95% CI 0.636–0.866) and 0.739 (95% CI 0.611–0.866) in the derivation cohort, internal and two external validation cohorts, respectively.

Development of combined radiomics model

Univariable and multivariable analyses showed that CA19-9 (> 1000 U/ml), vascular invasion and tumor margin were independently associated with ER (Supplementary Fig S3) with variance inflation factors (1.8, 1.7 and 1.5, respectively) all < 10 , suggesting no multicollinearity between the variables. A combined radiomics model was developed through incorporating the above three clinical features into the radiomics model using the LightGBM algorithm (see Supplementary Materials for detailed parameters). The categorical net reclassification index and integrated discrimination improvement of the combined radiomics model were improved in all the four cohorts with the addition of radiomics features (Supplementary Table S3). The inclusion of SHAP enables to perform an interpretative analysis of the combined radiomics model through visualizing specific impact of each variable on ER prediction (Fig. 2). Furthermore, each patient can be analyzed individually (Supplementary Fig S4).

Internal and external validation of combined radiomics model

As shown by ROC curves (Fig. 3A–D), the combined radiomics model yielded the highest AUCs of 0.974 (95% CI 0.946–1.000), 0.882 (95% CI 0.672–0.962), 0.871 (95% CI 0.786–0.955) and 0.878 (95% CI 0.797–0.960) in the derivation cohort, internal and two external validation cohorts, respectively. All the four cohorts showed good agreement on

Table 1 Baseline characteristics of ICC patients in the derivation cohort and 3 validation cohorts ($n=311$)

| Characteristics | Derivation cohort | Internal validation cohort | External validation cohorts | | <i>p</i> value |
|--|----------------------|----------------------------|-----------------------------|---------------------|----------------|
| | Cohort 1 ($n=140$) | Cohort 1 ($n=36$) | Cohort 2 ($n=74$) | Cohort 3 ($n=61$) | |
| Age | 59.06 ± 11.85 | 55.44 ± 10.55 | 56.53 ± 10.50 | 60.15 ± 10.79 | 0.092 |
| Gender | | | | | 0.166 |
| Male | 53 (37.9) | 14 (38.9) | 37 (54.4) | 32 (52.5) | |
| Female | 87 (62.1) | 22 (61.1) | 39 (52.7) | 29 (47.5) | |
| Total bilirubin ($\mu\text{mol/L}$) [†] | 17.15 [13.00–25.73] | 15.20 [8.85–21.93] | 11.92 [7.98–27.95] | 17.40 [10.90–50.90] | 0.015 |
| Albumin (g/L) [†] | 37.70 [34.15–40.20] | 37.10 [34.38–39.00] | 36.84 [33.12–40.05] | 37.50 [34.20–40.80] | 0.459 |
| PLT ($10^9/\text{L}$) | 218.75 (84.34) | 219.75 (82.03) | 256.96 (92.93) | 225.03 (87.07) | 0.018 |
| ALT (U/L) [†] | 32.00 [23.00–44.50] | 31.00 [19.50–75.50] | 25.15 [15.35–71.08] | 32.00 [17.10–94.00] | 0.442 |
| AST (U/L) [†] | 31.00 [24.00–42.75] | 32.00 [23.50–54.00] | 30.27 [21.23–53.80] | 37.00 [23.20–94.00] | 0.479 |
| AFP (ng/ml) | | | | | 0.166 |
| ≤ 20 | 130 (92.9) | 33 (91.7) | 59 (86.8) | 58 (95.1) | |
| > 20 | 10 (7.1) | 3 (8.3) | 9 (13.2) | 3 (4.9) | |
| CEA (ng/l) | | | | | 0.151 |
| ≤ 5 | 100 (71.4) | 26 (72.2) | 40 (58.8) | 33 (54.1) | |
| > 5 | 40 (28.6) | 10 (27.8) | 28 (41.2) | 28 (45.9) | |
| CA-199 (u/ml) | | | | | 0.137 |
| ≤ 1000 | 110 (78.6) | 29 (80.6) | 55 (74.3) | 39 (63.9) | |
| > 1000 | 30 (21.4) | 7 (19.4) | 19 (25.7) | 22 (36.1) | |
| Tumor number | | | | | 0.518 |
| Single | 108 (77.1) | 29 (80.6) | 62 (83.8) | 45 (73.8) | |
| Multiple | 32 (22.9) | 7 (19.4) | 12 (16.2) | 16 (26.2) | |
| Tumor size (cm) | 4.80 ± 2.04 | 4.65 ± 1.83 | 5.33 ± 2.19 | 4.78 ± 1.98 | 0.230 |
| Viral hepatitis | | | | | 0.066 |
| No | 95 (67.9) | 23 (63.9) | 49 (66.2) | 49 (80.3) | |
| Yes | 45 (32.1) | 13 (36.1) | 25 (33.8) | 12 (19.7) | |
| Cirrhosis | | | | | 0.129 |
| No | 106 (75.7) | 27 (75.0) | 50 (67.6) | 52 (85.2) | |
| Yes | 34 (24.3) | 9 (25.0) | 24 (32.4) | 9 (14.8) | |
| Cholelithiasis | | | | | 0.001 |
| No | 99 (70.7) | 24 (66.7) | 54 (73.0) | 21 (34.4) | |
| Yes | 41 (29.3) | 12 (33.3) | 20 (27.0) | 40 (65.6) | |
| CT-reported lymphatic metastasis | | | | | 0.556 |
| No | 87 (62.1) | 20 (55.6) | 49 (66.2) | 42 (68.9) | |
| Yes | 53 (37.9) | 16 (44.4) | 25 (33.8) | 19 (31.1) | |
| CT-reported vascular invasion | | | | | 0.082 |
| No | 84 (60.0) | 26 (72.2) | 54 (73.0) | 46 (75.4) | |
| Yes | 56 (40.0) | 10 (27.8) | 20 (27.0) | 15 (24.6) | |
| Tumor margin | | | | | 0.560 |
| Smooth | 53 (37.9) | 10 (27.8) | 25 (33.8) | 25 (41.0) | |
| Non-smooth | 87 (62.1) | 26 (72.2) | 49 (66.2) | 36 (59.0) | |
| AJCC 8th TNM | | | | | 0.797 |
| IA | 35 (25.0) | 11 (30.6) | 15 (20.3) | 20 (32.8) | |
| IB | 17 (12.1) | 3 (8.3) | 12 (16.2) | 10 (16.4) | |
| II | 29 (20.7) | 6 (16.7) | 14 (18.9) | 10 (16.4) | |
| IIIA | 13 (9.3) | 1 (2.8) | 8 (10.8) | 2 (3.3) | |
| IIIB | 42 (30.0) | 13 (36.1) | 22 (29.7) | 18 (29.5) | |
| IV | 35 (25.0) | 11 (30.6) | 15 (20.3) | 20 (32.8) | |
| Recurrence | | | | | 0.895 |

Table 1 (continued)

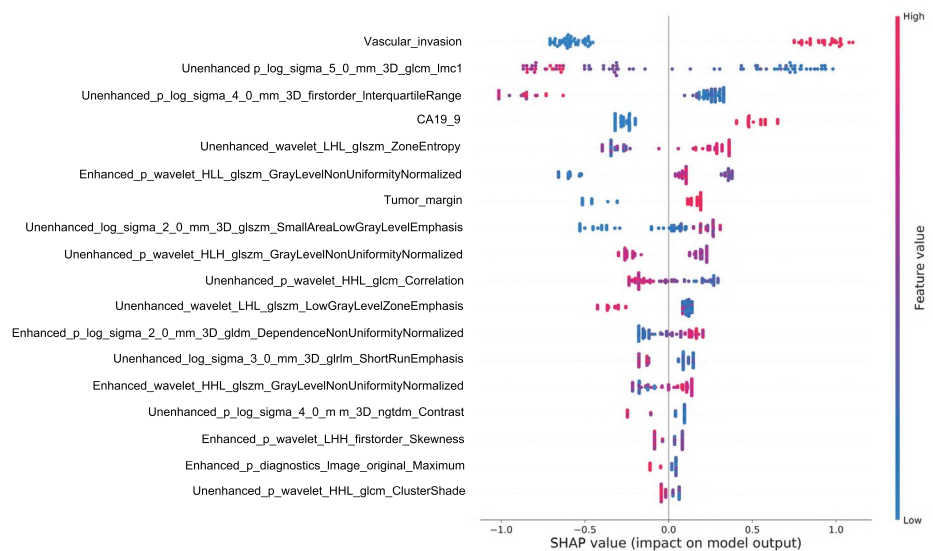
| Characteristics | Derivation cohort | Internal validation cohort | External validation cohorts | | p value |
|-------------------------------|--------------------|----------------------------|-----------------------------|-------------------|---------|
| | Cohort 1 (n = 140) | Cohort 1 (n = 36) | Cohort 2 (n = 74) | Cohort 3 (n = 61) | |
| No | 65 (46.4) | 17 (47.2) | 34 (45.9) | 30 (49.2) | 0.988 |
| Yes | 75 (53.6) | 19 (52.8) | 40 (54.1) | 31 (50.8) | |
| Sites of recurrence | | | | | 0.988 |
| Intrahepatic | 47 (62.7) | 11 (57.8) | 26 (65.0) | 20 (64.5) | |
| Extrahepatic | 8 (10.7) | 4 (21.1) | 6 (15.0) | 4 (12.9) | |
| Intrahepatic and extrahepatic | 20 (26.6) | 4 (21.1) | 8 (20.0) | 7 (22.6) | |
| RFS (weeks) | 50.0 [27.0–94.5] | 49.0 [29.3–123.5] | 49.5 [26.0–116.5] | 45.0 [34.0–109.0] | 0.877 |

Values in parentheses are percentages

ALT alanine transaminase, AST aspartate transaminase, AFP alpha-fetoprotein, CA19-9 carbohydrate antigen 19-9, CEA, carcinoembryonic antigen

†For abnormal distribution, data are shown as medians [interquartile ranges] and were analyzed by the Kruskal–Wallis H test

Fig. 2 Contribution analysis of each feature by using the SHapley Additive exPlanations (SHAP) in the LightGBM model. The features are ranked according to the sum of the SHAP values for all patients. A positive/negative SHAP value indicate a positive/negative correlation between a variable and outcome



the calibration curves (Fig. 4A–D), and the Hosmer–Lemeshow test suggested no significant differences between the predicted probabilities and the actual observed recurrence probabilities ($p=0.988, 0.541, 0.241$ and 0.330), supporting the accurate performance by this model. The confusion matrix for each model is shown in Supplementary Fig S5.

In addition, comparisons between each ROC curve in the three validation cohorts all showed p value greater than 0.05 (Supplementary Table S4), which also suggested good stability and reproducibility of the combined radiomics model.

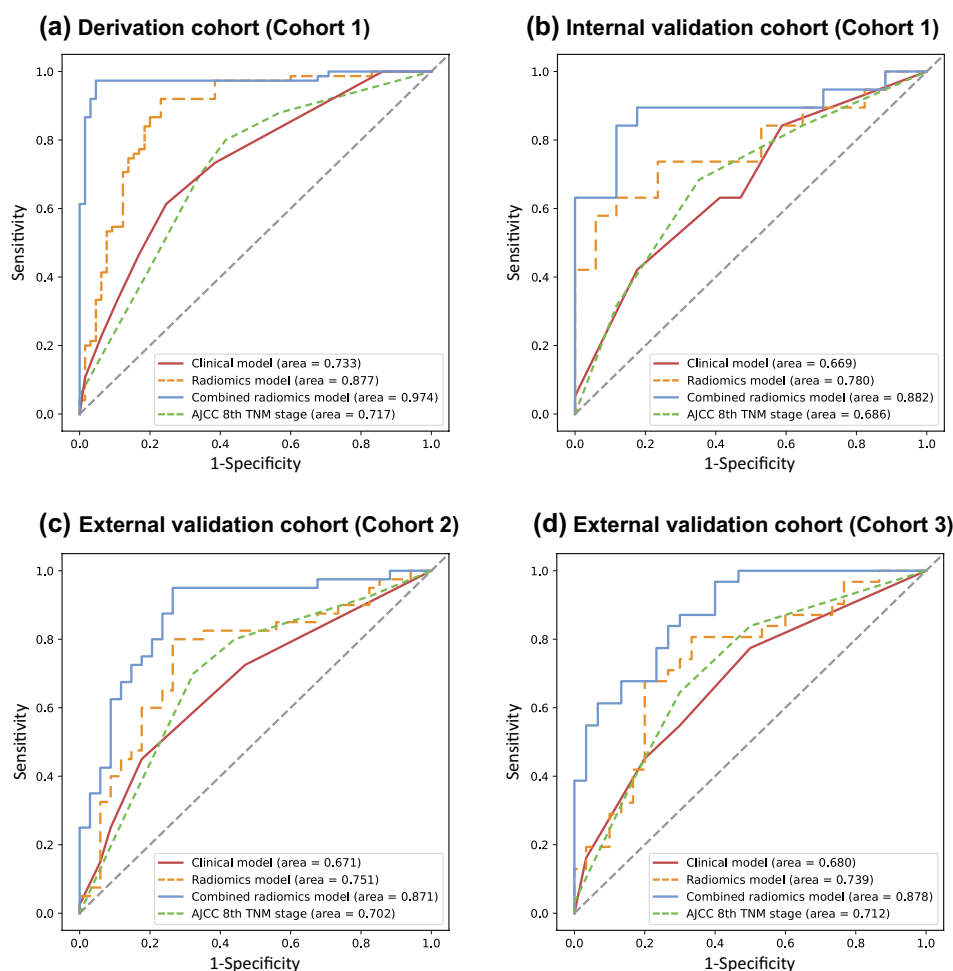
Patients in each cohort were preoperatively stratified according to the risk predicted by the combined radiomics model and Kaplan–Meier analysis all suggested good discrimination (Supplementary Fig S6).

Then, we compared the combined radiomics model with the AJCC 8th TNM system, showing that the

combined radiomics model performed better in both internal and external validation cohorts (Table 2 and Fig. 5). Especially, the combined radiomics model had a significantly higher sensitivity reaching up to 94.6% on average for all the cohorts. The ROC curves comparison between with the combined radiomics and the AJCC 8th TNM staging showed significant difference ($p < 0.05$) in all the cohorts by Delong test (Supplementary Table S5).

In terms of practical applications, as shown in Fig. 6, it helps clinicians to identify a subset of patients preoperatively who are at high ER risk and tailor better treatment and follow-up plans, such as early combination of neoadjuvant chemotherapy or shortening the follow-up interval, which would help reduce the risk of ER and improve the prognosis of patients.

Fig. 3 Comparison of ROC curves between different models for predicting the risk of early recurrence after curative resection of ICC. ICC cholangiocarcinoma, ROC receiver operating characteristic



Discussion

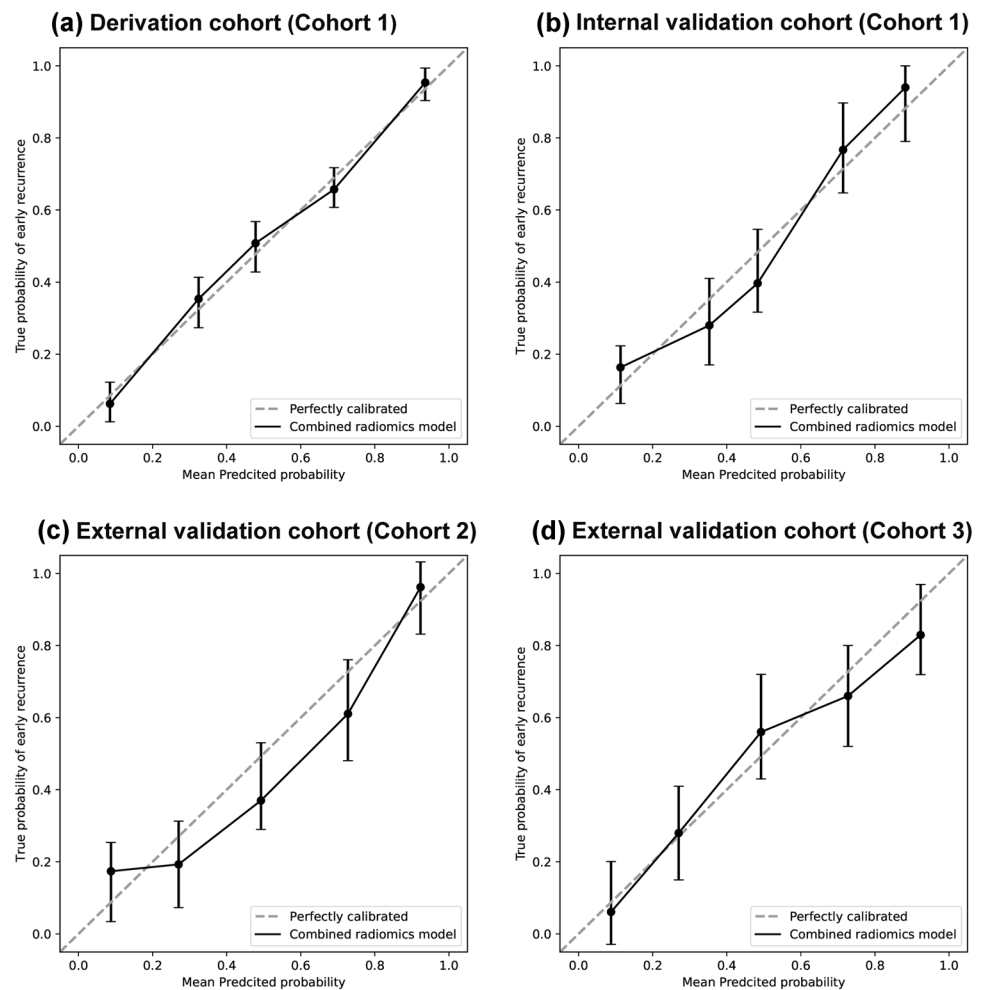
In this multicenter study, we established and evaluated a combined radiomics model with LightGBM machine learning algorithm for predicting ER after curative resection in ICC. Using this robust model, a total of 149 patients were identified as having high risk of recurrence in all cohorts, accounting for 94.6% of the actual ER patients and only 5.1% patients were missed by this model, which is of great importance for ICC, a tumor with high malignancy.

Several radiomics studies have been reported on the prediction of the prognosis of ICC. Liang et al. [15] developed a nomogram based on preoperative MR images to predict ER of ICC in a single center study with AUCs of 0.90 and 0.85 in the training and validation groups. In the present study, the AUCs reached 0.97 and 0.88 in the training and internal validation cohorts, and still reached more than 0.87 in two independent external validation centers, indicating that both CT and MR image-based radiomics models can achieve comparable predictive efficacy, and CT has more economical and temporal advantages. Also as a preoperative CT image-based study, Park et al. [25] segmented the tumor

on the single transverse section and developed a model by LASSO-COX regression, yielding the C-index with 0.75 and 0.71 in the training and validation groups, which was not significantly different from the existing TNM staging system (C-index: 0.73). In our study, the combined radiomics model outperformed the TNM system in both internal and external validation, highlighting two factors for improving the ability for ER prediction: one is the 3D VOI segmentation and peri-tumoral area included, and the other is the use of LightGBM-based machine learning algorithm to construct the model. Recently, Zhu Y et al. also predicted early recurrence of ICC based on a CT radiomics approach, with the AUCs of 0.844 (95% CI 0.751–0.912) and 0.793 (95% CI 0.617–0.914) in the training and validation groups [16]. Admittedly the model performed well, but its AUCs still slightly lower than our model. And the study by Zhu Y et al. was also based solely on the transverse section of tumor maximum diameter for segmentation and its relatively small number of cases (only 125) should also be noted.

The variables included in the model of this study included 15 radiomic features and 3 clinical features. It is worth noting that 93.3% (14/15) of the radiomics features were 3D and

Fig. 4 Calibration curves in each cohort. The dotted line in the middle diagonal is the ideal perfect situation. The solid line is the predicted performance of the current model, and the vertical bars represent confidence interval. When the solid line is closer to the dotted line, the goodness of fit of the model is better



wavelet features, which is consistent with the results of previous studies [26, 27] and indicates that tumor heterogeneity can be reflected as differences in the spatial distribution of voxel intensities in images, which is one of the assumptions of radiomics [28]. After univariate and multifactorial regression analyses, three variables were finally selected in terms of clinical features: CA19-9 (> 1000 U/ml), tumor vascular invasion and tumor margin, where the latter two variables were consistent with several previous studies and were considered as one of the independent risk factors for postoperative recurrence [29]. Whether CA19-9 can serve as an independent risk factor used for the prediction of recurrence is controversial. Park et al. [25] and Jeong et al. [9] concluded that tumor markers including CA19-9 and CEA were not significantly associated with postoperative recurrence, while the studies of Wang et al. [29] and Ji et al. [30] found that CA19-9 (> 1000 U/ml) was significantly associated with RFS ($p < 0.05$). In addition, Roongruedee et al. [31] also found that CA19-9 > 1000 U/ml was the optimal threshold for predicting the prognosis after resection by comparing different thresholds of CA19-9 ($\geq 100, \geq 500, \geq 1000$ U/

ml), and these findings are also consistent with the present study. Finally, it is of interest that although tumor size was associated with ER in univariate analysis, it did not show significant statistical differences in multivariate analysis ($p = 0.146$). In AJCC 8th staging, T1 stage was divided into T1a and T1b stages using tumor diameter > 5 cm as the cut-off value, while several studies also suggested that tumor diameter > 5 cm was an independent risk factor for tumor recurrence and survival [32, 33]. However, the aforementioned studies mainly focused on patients with solitary ICC without vascular invasion, i.e., in stage T1, whereas in our study, the staging of the included ICC patients varied and was not limited to stage T1, suggesting that tumor diameter size may not be a major influencing factor associated with ER after covering all stages of ICC.

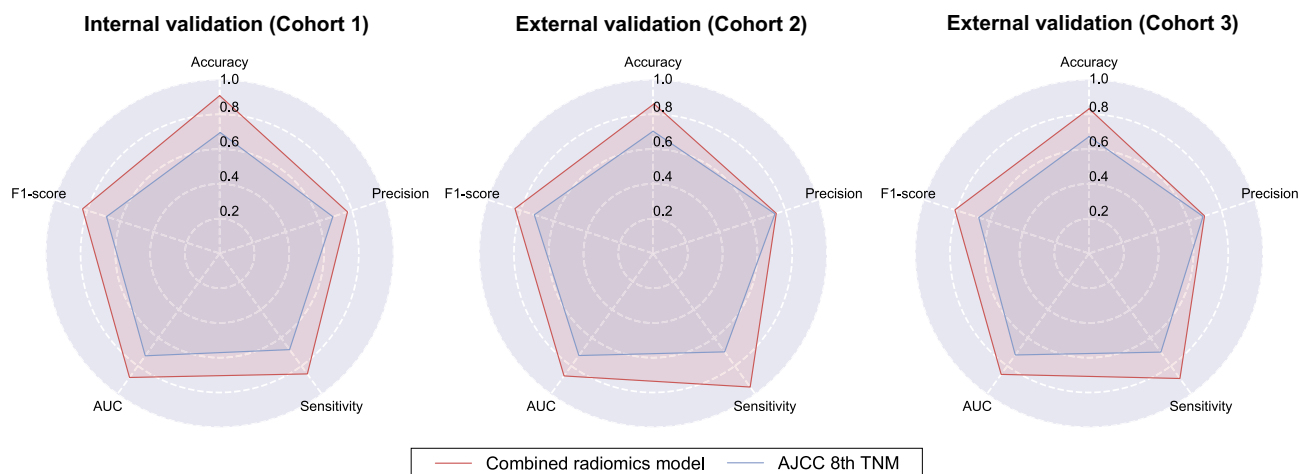
In real world, variables may interact in both linear and non-linear fashions. Traditional linear models used in previous studies often yielded poor performance in real-world testing. However, machine learning can capture complex linear and non-linear relationships between variables. In this study, the LightGBM algorithm was selected for its superior

Table 2 Comparison of prediction performance between Combined radiomics model and AJCC 8th TNM staging

| | AUC (95% CI) | ACC | SEN | SPE | PPV | NPV | <i>p</i> value [†] |
|---------------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|-----------------------------|
| Derivation cohort (cohort 1) | | | | | | | 0.010 |
| Combined radiomics model | 0.974 (0.946–1.000) | 0.964 (0.929–0.993) | 0.973 (0.920–1.000) | 0.954 (0.908–1.000) | 0.961 (0.923–1.000) | 0.969 (0.913–1.000) | |
| AJCC 8th TNM staging | 0.717 (0.631–0.803) | 0.700 (0.629–0.771) | 0.747 (0.627–0.867) | 0.646 (0.508–0.753) | 0.704 (0.635–0.782) | 0.688 (0.590–0.800) | |
| Internal validation cohort (cohort 1) | | | | | | | 0.046 |
| Combined radiomics model | 0.882 (0.672–0.962) | 0.889 (0.778–0.972) | 0.895 (0.732–1.000) | 0.941 (0.765–1.000) | 0.938 (0.792–1.000) | 0.850 (0.700–1.000) | |
| AJCC 8th TNM staging | 0.686 (0.510–0.862) | 0.667 (0.528–0.806) | 0.684 (0.421–0.895) | 0.647 (0.412–0.882) | 0.688 (0.546–0.875) | 0.650 (0.500–0.842) | |
| External validation cohort (cohort 2) | | | | | | | 0.028 |
| Combined radiomics model | 0.871 (0.786–0.955) | 0.851 (0.770–0.919) | 0.950 (0.725–1.000) | 0.764 (0.618–0.941) | 0.830 (0.745–0.943) | 0.920 (0.744–1.000) | |
| AJCC 8th TNM staging | 0.702 (0.580–0.824) | 0.703 (0.595–0.797) | 0.725 (0.600–0.875) | 0.677 (0.471–0.824) | 0.718 (0.622–0.829) | 0.677 (0.567–0.813) | |
| External validation cohort (cohort 3) | | | | | | | 0.018 |
| Combined radiomics model | 0.878 (0.797–0.960) | 0.820 (0.738–0.902) | 0.968 (0.748–1.000) | 0.800 (0.533–1.000) | 0.821 (0.689–1.000) | 0.852 (0.682–1.000) | |
| AJCC 8th TNM staging | 0.712 (0.581–0.843) | 0.689 (0.574–0.803) | 0.710 (0.516–0.903) | 0.667 (0.467–0.867) | 0.688 (0.581–0.826) | 0.688 (0.568–0.857) | |

ACC Accuracy, AUC Area Under the Curve, NPV Negative predictive value, PPV Positive predictive value, SEN Sensitivity, SPE Specificity, CI confidence interval

[†]Comparison of the ROC curves of the combined radiomics model and the TNM system by Delong test

**Fig. 5** The radar chart visualization of prediction performance for Combined radiomics model and AJCC 8th TNM staging

performance after comparing different algorithms. As an emerging algorithm born in 2017, LightGBM is also an algorithm based on ensemble learning. In addition to its high accuracy, LightGBM also has the advantages of supporting distributed learning, faster training speed and lower memory usage. It has won the Kaggle data science competitions several times and has been practically applied in the finance and power industries [34, 35]. In this study, LightGBM delivered efficient performance, the average training time takes < 30 s

even after 100,000 times of cross-validation, which greatly enhances the possibility of practical clinical applications in the future.

Derived from Gdbt, LightGBM has better accuracy in prediction compared to traditional linear models, but it also loses the interpretability of linear models, and is usually considered a black box model. In 2017, Lundberg and Lee [36] proposed SHAP values, a unified approach to interpreting model predictions, thus making it possible to

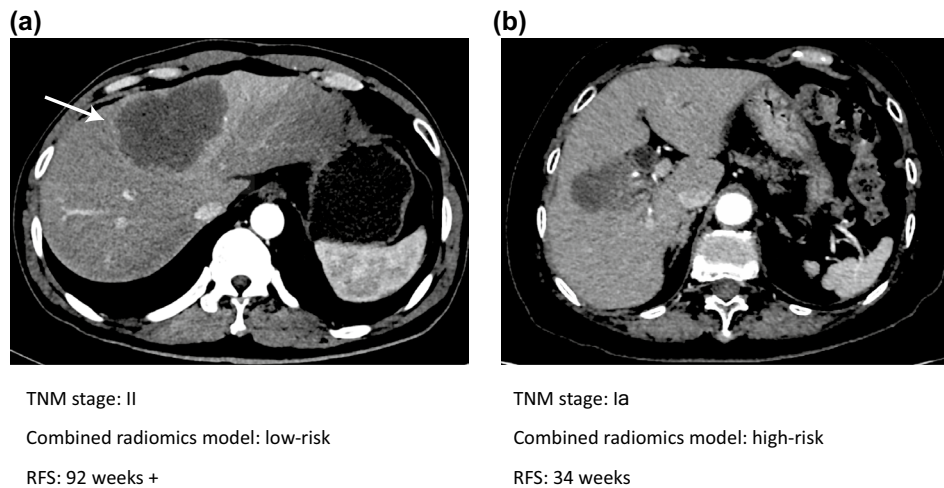


Fig. 6 Two representative contrast-enhanced CT image examples of ICC. **A** CT image of a 48-year-old man in arterial phase shows a 7.9×4.8 cm nodular lesion with arterial peri-tumoral enhancement (arrow) in hepatic segment IV. Postoperative pathology showed positive vascular invasion and the TNM staging of this patient is II. Using the combined radiomics model, this patient was stratified as low risk and he survived without recurrence until the last follow-up

(92 weeks); **B** CT image of a 74-year-old woman in arterial phase shows a 4.3×2.6 cm nodular lesion without obvious arterial enhancement in hepatic segment V/VIII. Postoperative pathology showed negative vascular invasion, and his TNM staging was stage Ia. However, using the combined radiomics model, this patient was stratified as high risk and the tumor recurred 34 weeks after surgery. ICC intra-hepatic cholangiocarcinoma, RFS recurrence free survival

interpret black box models such as LightGBM and neural networks possible. This new approach not only allows us to understand the intricate relationships between the predictors and the target variables, but also helps to explore in depth the relationship between these large number of mathematically transformed features in radiomics and gene expression in life science.

There are several limitations in this study. First, considering the long appointment time for MR and the convenience available for CT, we chose CT images for analysis, but the use of MR has also attracted increasing attention in hepatobiliary diseases, especially MRCP and MR-specific sequences [37, 38]. Second, we selected non-contrast and arterial phase-based images for analysis in this study and, theoretically, although there might be a decrease in predictive performance compared to the full-phase CT image model. Surprisingly, the predictive model we developed also achieved relatively superior performance compared to related studies (AUC = 0.871–0.974), which can be attributed to the relatively large sample size in our study and the machine learning algorithm. In the future, we also plan to perform feature extraction and analysis based on the full-phase CT. Third, considering the accuracy of image segmentation, we used manual segmentation of ROI, but it inevitably takes more time. With the development of AI technology, the inclusion of automatic segmentation may further improve the practical clinical application in the future. Fourth, the present study only evaluated 311 Chinese ICC patients and was unable to evaluate different

ethnic patients. Further validations of the main findings by international multicenter are required.

In conclusion, we used the LightGBM machine learning algorithm to construct the combined radiomics model for ICC ER prediction. The combined radiomics model showed highly sensitivity and detected 94.6% of the ICC patients who actually developed ER for all the cohorts and delivered a better performance to TNM staging system. This model, once further validated, may provide a new clinical tool to improve clinical management of ICC patients.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12072-023-10487-z>.

Acknowledgements Thanks to Professor Shixing Tang from the Department of Epidemiology, Southern Medical University for his help in the statistical analysis in this study.

Author contributors (I) Conception and design: YPZ, YDS; (II) administrative support: YPZ, WJL and Y-YZ; (III) data acquisition: GYZ, YCZ, KTZ, PYH, MWC, YPL; (IV) collection and assembly of data: YDS, GYZ, JRS, CGH, ML, MJL; (V) data analysis and interpretation: YDS, YKX, JZ, JRS; (VI) manuscript writing: all authors; (VII) manuscript revised: YPZ, Y-YZ, WJL and YDS; (VIII) final approval of manuscript: all authors.

Funding This study was supported by the grants from the National Natural Science Foundation of China (81772923 and 81773148). The funding agencies has no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability All experiments and implementation details are described thoroughly in the Methods section. Requests for access to

data including codes can be obtained after asking for the corresponding authors and clarifying purpose of use.

Declarations

Conflict of interest Yangda Song, Guangyao Zhou, Yucheng Zhou, Yikai Xu, Jing Zhang, Ketao Zhang, Pengyuan He, Maowei Chen, Yanping Liu, Jiarun Sun, Chengguang Hu, Meng Li, Minjun Liao, Yongyuan Zhang, Weijia Liao and Yuanping Zhou declare that they have no conflict of interest.

Ethical approval This study was approved by the ethics committee of each of participating centers, and the requirement for informed consent was waived since this was a retrospective study. All procedures involving human participants were followed in accordance with the Declaration of Helsinki and its subsequent amendments.

References

- Bridgewater J, Galle PR, Khan SA, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol*. 2014;60:1268–1289
- Saha SK, Zhu AX, Fuchs CS, Brooks GA. Forty-year trends in cholangiocarcinoma incidence in the US: intrahepatic disease on the rise. *Oncologist*. 2016;21:594–599
- Endo I, Gonen M, Yopp AC, et al. Intrahepatic cholangiocarcinoma: rising frequency, improved survival, and determinants of outcome after resection. *Ann Surg*. 2008;248:84–96
- Weber SM, Jarnagin WR, Klimstra D, et al. Intrahepatic cholangiocarcinoma: resectability, recurrence pattern, and outcomes. *J Am Coll Surg*. 2001;193:384–391
- Zhang XF, Beal EW, Bagante F, et al. Early versus late recurrence of intrahepatic cholangiocarcinoma after resection with curative intent. *Br J Surg*. 2018;105:848–856
- Mavros MN, Economopoulos KP, Alexiou VG, Pawlik TM. Treatment and prognosis for patients with intrahepatic cholangiocarcinoma: systematic review and meta-analysis. *JAMA Surg*. 2014;149:565–574
- Lee AJ, Chun YS. Intrahepatic cholangiocarcinoma: the AJCC/UICC 8th edition updates. *Chin Clin Oncol*. 2018;7:52
- Sirica AE, Gores GJ, Groopman JD, et al. Intrahepatic cholangiocarcinoma: continuing challenges and translational advances. *Hepatology*. 2019;69:1803–1815
- Jeong S, Cheng Q, Huang L, et al. Risk stratification system to predict recurrence of intrahepatic cholangiocarcinoma after hepatic resection. *BMC Cancer*. 2017;17:464
- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446
- Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34:2157–2164
- Nie K, Shi L, Chen Q, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res*. 2016;22:5256–5264
- Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging*. 2019;46:2638–2655
- Baeßler B, Weiss K, Pinto DSD. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol*. 2019;54:221–228
- Liang W, Xu L, Yang P, et al. Novel nomogram for preoperative prediction of early recurrence in intrahepatic cholangiocarcinoma. *Front Oncol*. 2018;8:360
- Zhu Y, Mao Y, Chen J, et al. Radiomics-based model for predicting early recurrence of intrahepatic mass-forming cholangiocarcinoma after curative tumor resection. *Sci Rep*. 2021;11:18347
- van Griethuysen J, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107
- Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med*. 2007;35:2052–2056
- Weber SM, Ribero D, O'Reilly EM, et al. Intrahepatic cholangiocarcinoma: expert consensus statement. *HPB (Oxford)*. 2015;17:669–680
- Zhang H, Yang T, Wu M, Shen F. Intrahepatic cholangiocarcinoma: epidemiology, risk factors, diagnosis and surgical management. *Cancer Lett*. 2016;379:198–205
- Valle JW, Borbath I, Khan SA, et al. Biliary cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27:v28–v37
- Banales JM, Marin J, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol*. 2020;17:557–588
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2012;12:2825
- Park HJ, Park B, Park SY, et al. Preoperative prediction of post-surgical outcomes in mass-forming intrahepatic cholangiocarcinoma based on clinical, radiologic, and radiomics features. *Eur Radiol*. 2021;11:8638–8648
- Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *AJNR Am J Neuroradiol*. 2010;31:809–816
- Xu L, Yang P, Liang W, et al. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. *Theranostics*. 2019;9:5374–5385
- Campbell PJ, Yachida S, Mudie LJ, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. 2010;467:1109–1113
- Wang Y, Li J, Xia Y, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J Clin Oncol*. 2013;31:1188–1195
- Ji GW, Zhang YD, Zhang H, et al. Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes. *Radiology*. 2019;290:90–98
- Chaiteerakij R, Harmsen WS, Marrero CR, et al. A new clinically based staging system for perihilar cholangiocarcinoma. *Am J Gastroenterol*. 2014;109:1881–1890
- Hwang S, Lee YJ, Song GW, et al. Prognostic impact of tumor growth type on 7th AJCC staging system for intrahepatic cholangiocarcinoma: a single-center experience of 659 cases. *J Gastrointest Surg*. 2015;19:1291–1304
- Ruiyang W, ZhiMing Y, Jiao F, Liang Z, Gang Z. Evaluation and recommendation of the 8th edition of american joint committee on cancer (AJCC) staging system for intrahepatic cholangiocarcinoma (ICC) in 820 patients from the surveillance epidemiology and end results (SEER) database. *J Gastrointest Surg*. 2021;25:145–154
- Ma X, Sha J, Wang D, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and

- XGboost algorithms according to different high dimensional data cleaning. *Electron Commer Res Appl.* 2018;31:24–39
35. Ju Y, Sun G, Chen Q, et al. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access.* 2019;7:28309–28318
 36. Lundberg SM, Lee S. A unified approach to interpreting model predictions. *Neural Inf Process Syst (NIPS 2017).* 2017;2017:4768–4777
 37. Jhaveri KS, Hosseini-Nik H. MRI of cholangiocarcinoma. *J Magn Reson Imaging.* 2015;42:1165–1179
 38. Zou X, Luo Y, Li Z, et al. Volumetric apparent diffusion coefficient histogram analysis in differentiating intrahepatic mass-forming

cholangiocarcinoma from hepatocellular carcinoma. *J Magn Reson Imaging.* 2019;49:975–983

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Yangda Song^{1,2} · Guangyao Zhou^{1,3} · Yucheng Zhou⁴ · Yikai Xu⁵ · Jing Zhang⁵ · Ketao Zhang⁶ · Pengyuan He^{1,7} · Maowei Chen⁸ · Yanping Liu^{1,9} · Jiarun Sun^{1,2} · Chengguang Hu¹ · Meng Li^{1,2} · Minjun Liao^{1,2} · Yongyuan Zhang¹⁰ · Weijia Liao¹¹ · Yuanping Zhou^{1,2} 

¹ Department of Gastroenterology, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Ave, Guangzhou 510515, China

² Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

³ Department of Infectious Diseases, the Second Affiliated Hospital of Wenzhou Medical University, Wenzhou 325027, China

⁴ Department of General Surgery, Hospital of Integrated TCM and Western Medicine, Southern Medical University, Guangzhou 510315, China

⁵ Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

⁶ Department of Hepatobiliary Surgery, Shunde Hospital of Southern Medical University, Foshan 528308, Guangdong, China

⁷ Department of Infectious Diseases, The Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai, Guangdong, China

⁸ Department of Infectious Diseases, Wuming Hospital of Guangxi Medical University, Nanning 530199, Guangxi, China

⁹ Department of Gastroenterology, Second Affiliated Hospital, University of South China, Hengyang 421001, Hunan, China

¹⁰ HBVtech, Germantown, MD 20874, USA

¹¹ Laboratory of Hepatobiliary and Pancreatic Surgery, Affiliated Hospital of Guilin Medical University, Guilin 541001, Guangxi, China