Annals of Hepatology xxx (2025) 102110

Contents lists available at ScienceDirect

Annals of Hepatology

journal homepage: www.elsevier.es/annalsofhepatology

Original article

MRI imaging and machine learning based radiomics for detection of mixed HCC and CCA tumors

Q1 Yuquan Qian[a,b,1], Qiao-Yuan Lu[c,1], Isaac Rodriguez[a], Michael Vácha[a], Xiangde Min[d], Muzaffer Reha Ümütlü[e], German A. Castrillon[f], Andreas Georg Schreyer[g], Michael Haimerl[h], Philipp Wiggermann[i], Stefan Schönberg[j], Matthias P. Ebert[k,l], Abhinay Vellala[j], Carlos Romero-Alaffita[m], Juan Alberto Garay Mora[n], Zhiqiang Guo[o], Jürgen Hesser[p], Christel Weiss[q], Matthias Froelich[j], Ying-Shi Sun[c,2], Andreas Teufel[a,l,2,*]

[a] Division of Hepatology, Division of Clinical Bioinformatics, Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[b] Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[c] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiology, Peking University Cancer Hospital and Institute, Beijing, China
[d] Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[e] Department of Radiology, University Hospital, LMU Munich, Munich, Germany
[f] Department of Radiology and Gastrohepatology, University of Antioquia, Medellin, Colombia
[g] Department of Radiology, Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Brandenburg an der Havel, Germany
[h] Department of Radiology, University Hospital Regensburg, Regensburg, Germany
[i] Department of Radiology and Nuclear Medicine, Hospital Braunschweig, Braunschweig, Germany
[j] Department of Radiology and Nuclear Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[k] Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[l] Clinical Cooperation Unit Healthy Metabolism, Center for Preventive Medicine and Digital Health (CPDBW), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[m] Universidad Autónoma de San Luis Potosí, San Luis Potosi, Mexico
[n] Instituto Nacional de Ciencias Médicas & Nutrición Salvador Zubiran, Mexico City, Mexico
[o] Department of Oncology, Shanxi Province Fenyang Hospital, Fenyang, China
[p] Department of Data Analysis and Modeling in Medicine, Mannheim Institute for Intelligent Systems in Medicine (MIISM), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[q] Department of Biomedical Informatics, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

ARTICLE INFO

Article History:
Received 20 January 2025
Accepted 1 July 2025
Available online xxx

Keywords:
Machine learning models
Diagnostic accuracy
Liver cancer
Radiomics;MRI scans

ABSTRACT

Introduction and objectives: Primary liver cancer (PLC), comprising hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA), is a leading cause of cancer mortality globally. The combined hepatocellular-cholangiocarcinoma (cHCC—CC) subtype may be less common but is relevant to treatment efficacy. We therefore evaluated the diagnostic accuracy of various approaches in distinguishing these liver cancers.
Materials and methods: Patients diagnosed with HCC, CCA, and cHCC—CC at Beijing University Cancer Hospital and Institute, China were included. Radiologists of varying expertise independently assessed MRI scans, and we measured their diagnostic consistency. Radiomic features were extracted from MRI scans, and machine learning was applied to differentiate the cancer types.
Results: Standard imaging was insufficient to reliably characterize cHCC—CC. Abdominal imaging experts (AIEs) had a higher mean sensitivity for HCC and CCA, 88% and 84% respectively, while non-experts (NIEs) had a lower sensitivity of 50% for HCC and 38% for CCA (HCC: $p = 0.03$, CCA: $p = 0.008$). Radiomic analysis found 'Sphericity' and 'ClusterShade' as the most relevant features. However, radiomics algorithms were also not sufficient to distinguish cHCC—CC from either HCC or CCA. Regarding sensitivity, the radiomic-based model was not better than radiologists for any of the three classes ($p = 0.065$ for HCC, $p = 0.426$ for CCA, and

Abbreviations: AIEs, Abdominal imaging experts; AUROC, Area under the receiver operating characteristic curve; CCA, Cholangiocarcinoma; cHCC-CC, Combined hepatocellular-cholangiocarcinoma; GLCM, Gray,level run length matrix; GLDM, Gray level dependence matrix; GLRLM, Gray level run length matrix; GLSZM, Gray level size zone matrix; HCC, Hepatocellular carcinoma; ICC, Interclass correlation coefficient; NGTDM, Neighboring gray-tone difference matrix; NIEs, Non-abdominal imaging experts; PLC, Primary liver cancer; ROI, Region of interest

* Corresponding author at: Division of Hepatology, Division of Clinical Bioinformatics, Department of Medicine II, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany,
E-mail address: Andreas.Teufel@medma.uni-heidelberg.de (A. Teufel).
[1] Contributed equally
[2] joined senior authors

https://doi.org/10.1016/j.aohep.2025.102110

Please cite this article as: Y. Qian, Q.-Y. Lu, I. Rodriguez et al., MRI imaging and machine learning based radiomics for detection of mixed HCC and CCA tumors, Annals of Hepatology (2025), https://doi.org/10.1016/j.aohep.2025.102110

$p = 1.0$ for cHCC—CC). The random forest algorithm yielded an accuracy of 76% in the test set, since it correctly classified most HCC and CCA, while only one quarter of cHCC—CC tumors.
Conclusions: Histopathological analysis, complemented by imaging as indicated, remains essential for accurate detection, diagnosis, and treatment of liver cancers.

1. Introduction

With approximately 900,000 new cases annually, primary liver cancer (PLC) ranks as the third most common cause of cancer-related death globally [1,2]. Hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA) are the two predominant types, constituting about 90 % of primary liver cancers [3]. Despite potential similarities in imaging, their treatments and survival rates differ significantly. Compared to CCA, HCC exhibits a much lower post-resectional relapse rate and advanced stages can be managed with systemic targeted therapies, such as VEGF inhibitors and tyrosine-kinase inhibitors as first-line options [4–6]. Combined hepatocellular-cholangiocarcinoma (cHCC—CC) constitutes 1.0 % to 14.2 % of primary liver cancers and is often viewed as an intermediary between these two entities in terms of diagnostic findings and prognosis [7–9]. Histologically, this tumor presents with both cancer cell types and cells displaying intermediate morphology [10]. Imaging findings are typically nonspecific and variable, often exhibiting features characteristic of both HCC and CCA [11,12].

Tumor biopsy significantly enhances diagnostic accuracy, guiding appropriate management strategies such as liver resection for resectable cases and immuno-/TKI or immuno/platinum-based therapy for unresectable cHCC—CCs [10,13,14]. However, current guidelines suggest that the diagnosis of HCC can rely solely on imaging [15–17]. This approach may lead to misdiagnosis, contradicting the principle of personalized treatment. Our objective is to evaluate the accuracy of radiologists and the usefulness of radiomic features in distinguishing between HCC, CCA, and cHCC—CC.

2. Materials and methods

2.1. Patients enrollment

Patients ($n = 68$) diagnosed with HCC, CCA, or cHCC—CC at Beijing University Cancer Hospital and Institute, China between June 2010 and September 2020 were recruited for this study. Inclusion criteria required an MRI scan within four weeks before or after the pathological diagnosis. The majority of pathological diagnoses (75 %, 51/68) were based on resection specimens and the rest were from multiple core biopsies. Patients with low-quality images, respiratory artifacts, or those who only underwent CT scans were excluded.

2.2. Qualitative image analysis

Seven radiologists from Asia, Europe, North America, and South America, comprising abdominal imaging experts (AIEs) and non-abdominal imaging experts (NIEs) or trainees, conducted independent and blinded assessments of the MRI scans. They were tasked with providing a diagnosis of either HCC, CCA, or cHCC—CC after evaluating predominantly qualitative radiological findings including biliary dilation, capsular retraction, cirrhosis, tumor diameter, number of tumors, number of segments infiltrated, intralesional fat, hemorrhage, peripheral rim enhancement, progressive enhancement, arterial enhancement, tumor thrombus, washout. They also evaluated subjective features including level of confidence in spotting the diagnosis, quality and contrast of the image, as well as its trustability.

MRI quality, contrast, and diagnostic confidence were rated on a scale of 1 to 5, where 1 = non-diagnostic, 2 = severely impaired, 3 = impaired, 4 = minor artifacts, and 5 = excellent. The confidence in diagnosis was rated as 1 = low, 2 = medium, or 3 = high. For T1 characteristics, the categories were 1 = hypointense, 2 = heterogeneous, 3 = isointense / not seen. For T2 characteristics, 1 meant homogeneously intermediate/ hyperintense, 2 peripheral hyperintense, central hypointense, 3 = heterogeneous, 4 = isointense / not seen. For the presence of cirrhosis, the categories were 0 = no, 1 = yes, 3 = not clear. For arterial enhancement, the categories were 0 = hypoenhancement, 1 = mild enhancement, and 2 = strong enhancement. The remaining categorical variables were binary (the feature is present/ absent).

2.3. Radiomic feature extraction

From the portal venous phase of contrast-enhanced MRI, the region of interest (ROI) was delineated as the biggest tumor within the liver. Segmentation was performed by a physician (I.R., with two years of experience in segmentation) and reviewed and refined by a clinical radiologist (M.F., with five years of experience in oncologic imaging) using a semi-automated approach with 3D Slicer (version 4.10.2) [18].

Automatic preprocessing was standardized for each case, involving resampling (downsampling to voxel size $1 \times 1 \times 1$ mm to mitigate the influence of varying layer thicknesses and employing linear interpolation), intensity normalization (z-score), and discretization (with a binwidth set to 20). Radiomics feature extraction was conducted in Python using the pyradiomics framework [19]. From each ROI, a total of 107 radiomic features were extracted following the Image Biomarker Standardisation Initiative guidelines [20]. These encompassed 18 first-order metrics, 14 shape features, and 85 texture features including 24 gray level run length matrix (GLCM), 16 gray level run length matrix (GLRLM), 16 gray level size zone matrix (GLSZM), 5 neighboring gray-tone difference matrix (NGTDM), 14 gray level dependence matrix (GLDM).

2.4. Model training and testing

We used Python version 3.12. and the scikit-learn library version 1.2. to develop a model for classifying the three tumor types. We evaluated accuracy and area under the receiver operating characteristic curve (AUROC) to compare the performance of different models. First, all features were scaled using the Min-Max method. Feature selection was conducted using the univariate SelectKBest method to rank the features by their F-values (from analysis of variance). These features were then combined in a Random Forest model using stratified 75 % of the dataset for training and validation while the rest was for testing (hold-out set). The class weights were adjusted according to their frequency. We used the random forest algorithm, one of the most popular and precise algorithms in conventional machine learning [21]. The 5 main hyperparameters (criterion, max depth, min samples leaf, min samples split, number of estimators), that we optimized by comparing the AUROCs in grid search. The hyperparameter grid was defined rationally, concerning the number of samples (2–50 estimators, max depth 2–10, min samples leaf 1–10, samples split 1–10). The optimal number of features (k) was based on comparing the AUROCs in 5-fold cross-validation and limited to 6 due to the sample size. The selected model's performance on the test dataset was evaluated with optimized hyperparameters.

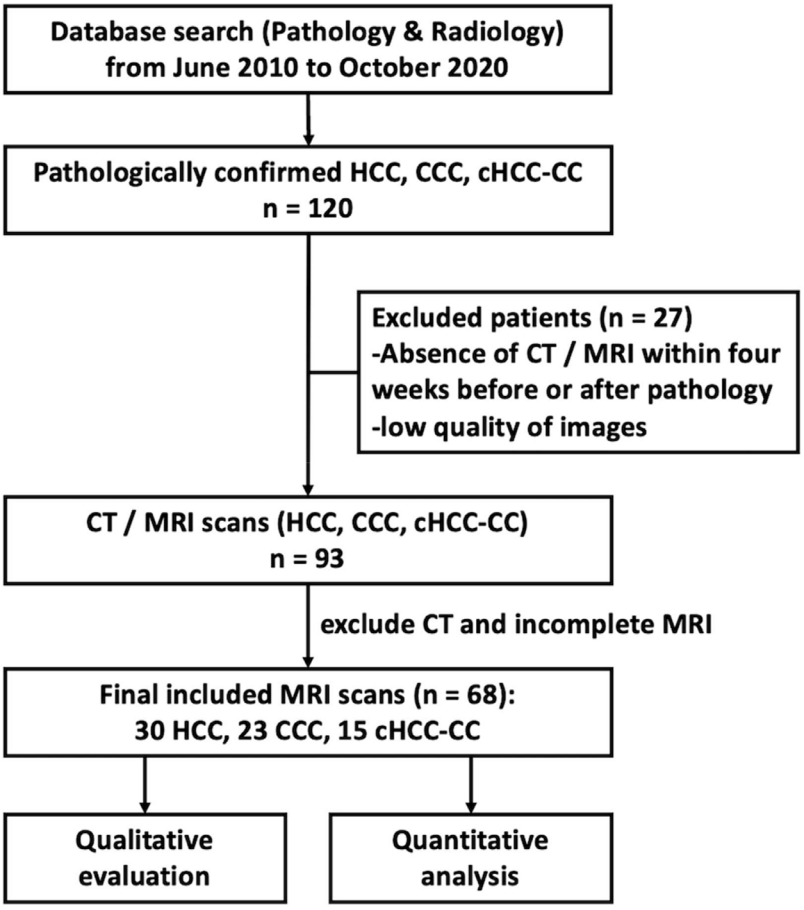

Fig. 1. Flowchart depicting patient cohort construction.

2.5. Statistical analysis

Statistical analysis was performed in Python 3.12. Quantitative variables were analyzed either by *t*-test/ analysis of variance (ANOVA) or by their non-parametric variants in case of non-normal distribution. For categorical variables (high vs. low feature value, radiologists vs. radiomics), we calculated p-values by Fisher's exact test, and odds ratios from a contingency table (in case of a zero value, Haldane-Anscombe correction was applied). Optimal feature cutoffs were calculated by a function maximizing the Youden Index. Cohen's kappa values were employed to assess the consistency of the radiologists' qualitative (categorical) findings of the MRI images. A kappa value of <0.01 indicated no agreement, 0.01 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and >0.80 almost perfect agreement [22]. For quantitative continuous features (diameter, number of lesions, number of segments infiltrated), we calculated an interclass correlation coefficient (ICC), where <0.50 indicated poor agreement, 0.50 to 0.75 fair agreement, 0.75 to 0.90 good agreement, and >0.90 excellent agreement [23].

2.6. Ethical statement

Ethical review and approval were waived for this study, due to the retrospective nature of the study. Participant consent was waived due to this study was conducted retrospectively from data obtained for clinical purposes.

3. Results

3.1. Patient collective

A total of 120 patients were initially identified for the study. After excluding those with inadequate MRI scans, 68 patients remained eligible for inclusion. This remaining patient cohort for further analysis comprised 30 patients with HCC, 23 with CCA, and 15 with cHCC—CC. A flowchart outlining this selection process is summarized in Fig. 1.

Regarding gender distribution, both HCC and cHCC—CC exhibited a predominance of male patients (23 males for HCC and 12 for cHCC—CC) compared to females (7 females for HCC and 3 for cHCC—CC). In contrast, CCA presented an almost equal gender

Table 1
Patient characteristics.

	HCC	CCA	cHCC—CC	P-value
Number of patients	30	23	15	
Gender				0.1018
Male	23	12	12	
Female	7	11	3	
Age (Mean SD)	60±9.9	60±11.6	55.8 ± 12.9	0.4431
Nodule diameter (Maximum, cm)	4.9 ± 3.6	6.4 ± 3.1	4.7 ± 2.2	0.0003
Cirrhosis				<0.0001
Yes	22	0	10	
No	8	23	5	

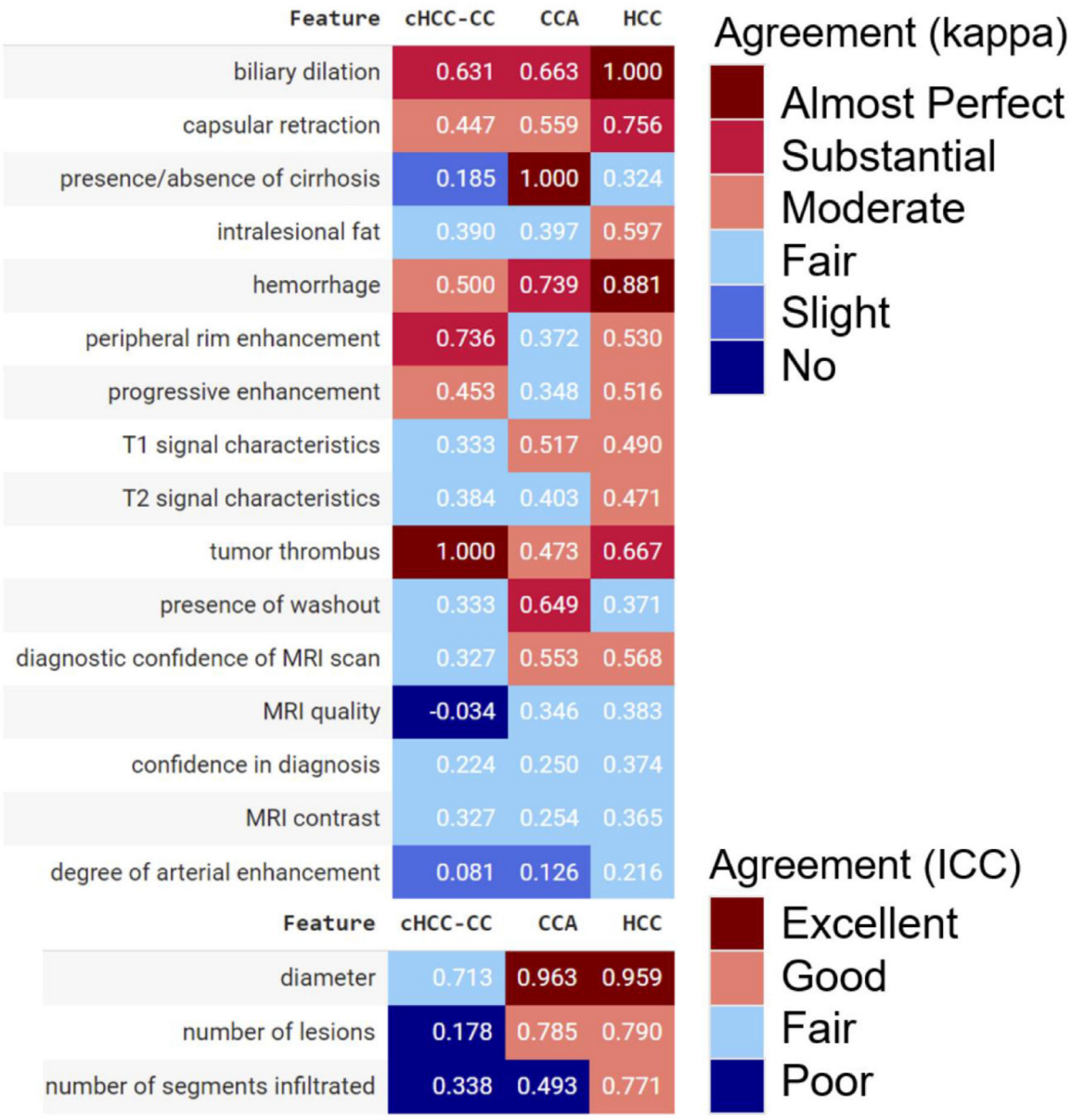

Fig. 2. Correlation matrix of radiological characteristics and MRI reliability evaluation depicting Cohen's kappa values or interclass correlation coefficients (ICC) among radiologists. For categorical variables, a kappa value of <0.01 indicated no agreement, 0.01 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and >0.80 almost perfect agreement. For continuous quantitative variables, an ICC of <0.50 indicated poor agreement, 0.50 and 0.75 fair agreement, 0.75 and 0.90 good agreement, and >0.90 excellent agreement.

distribution (12 males and 11 females). However, this did not represent any statistical difference ($p = 0.10$). The age of the patients was fairly consistent across the three tumor types, although those with cHCC—CC were, on average, slightly younger (55.8 years) compared to 60 years for the other tumor types ($p = 0.44$). Nodule size was similar between HCC and cHCC—CC, averaging 4.9 cm and 4.7 cm, respectively, whereas CCA tumors were larger, averaging 6.4 cm ($p = 0.0003$; post hoc HCC vs CCA

Table 2A
Sensitivity of radiologists and the tuned radiomics-based model in classifying the three tumor types.

Cancer types	Number of cases	Sensitivity				
		Radiologists			Radiomics – Tuned Model	
		Non-abdominal imaging experts (NIE)/Trainees	Abdominal imaging experts (AIE)	Weighted average	Training set	Testing set
HCC	30	50 %	88 %	71 %	91 %	100 %
CCA	23	38 %	84 %	64 %	88 %	83 %
cHCC—CC	15	40 %	25 %	31 %	91 %	25 %

Radiologists were categorized as abdominal imaging experts (AIE) or non-abdominal imaging experts (NIE)/trainees.

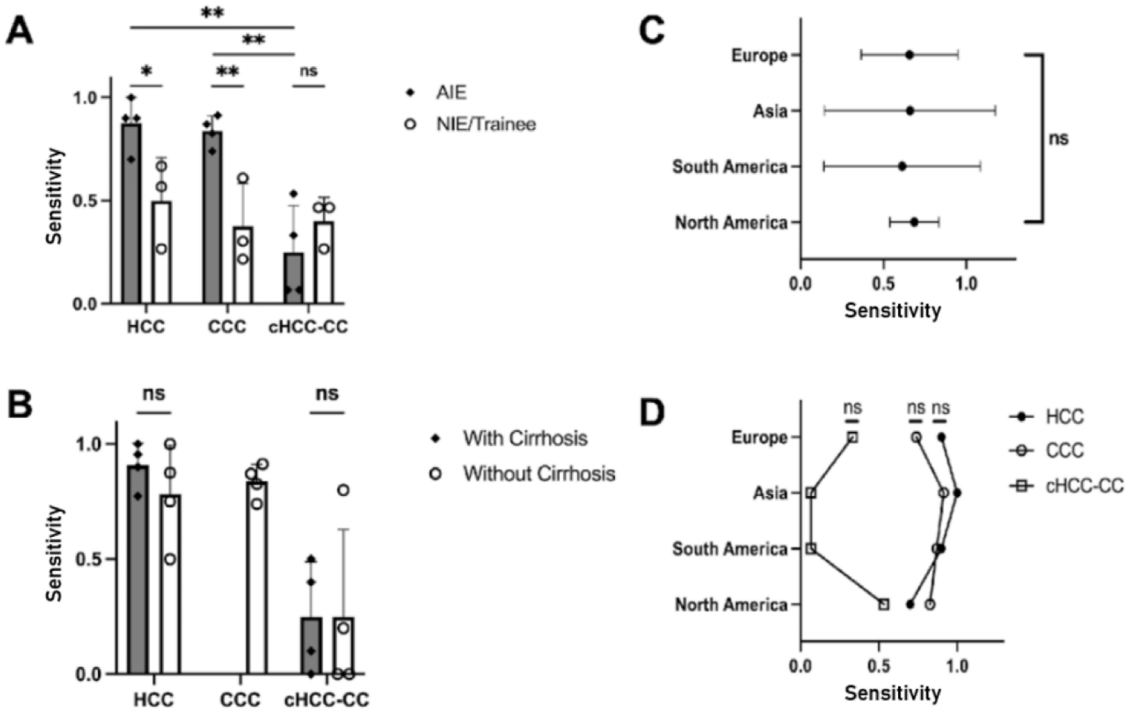

Fig. 3. Graphical result display of diagnostic accuracy rate for HCC, CCA, cHCC—CC depending on training (A), cirrhosis (B), and reviewers' country of origin (C, D).

$p = 0.0001$, HCC vs cHCC—CC $p = 0.29$, CCA vs cHCC—CC $p = 0.08$). Cirrhosis was more common among patients with HCC (22 patients) and cHCC—CC (10 patients), while none of the CCA patients had cirrhosis ($p < 0.0001$). These findings were summarized in Table 1.

3.2. Blinded radiological diagnosis of liver cancer

The qualitative radiological assessment of MRI scans achieved almost perfect agreement (kappa >0.80) for biliary dilation, tumor diameter, and hemorrhage in HCC; for tumor diameter and absence of cirrhosis in CCA; and for the presence of tumor thrombus in cHCC—CC. Despite their importance in distinguishing HCC from CCA, characteristics such as the presence of intralesional fat and progressive enhancement demonstrated only fair to moderate agreement. For HCC and CCA, the radiologists agreed excellently on the tumor diameter, while the agreement was only fair for cHCC—CC. Similarly, the agreement was poor for the number of lesions and infiltrated segments in the cHCC—CC class (Fig. 2).

AIEs demonstrated the highest proficiency in diagnosing HCC and CCA, achieving a sensitivity of 88 % and 84 %, respectively. However, their performance in detecting cHCC—CC was notably less effective, with sensitivity ranging between 7 % and 53 %, resulting in an average of 25 % (Table 2A).

On the other hand, NIEs displayed significantly lower detection rates of HCC and CCA. For HCC, their average sensitivity was 50 %, indicating a need for further training in the detection of this common liver cancer (AIE vs NIE $p = 0.03$). The average sensitivity for CCA was even lower among NIEs, 38 % (AIE vs NIE $p = 0.0076$). The detection of cHCC—CC by NIEs also showed limited success, with rates between

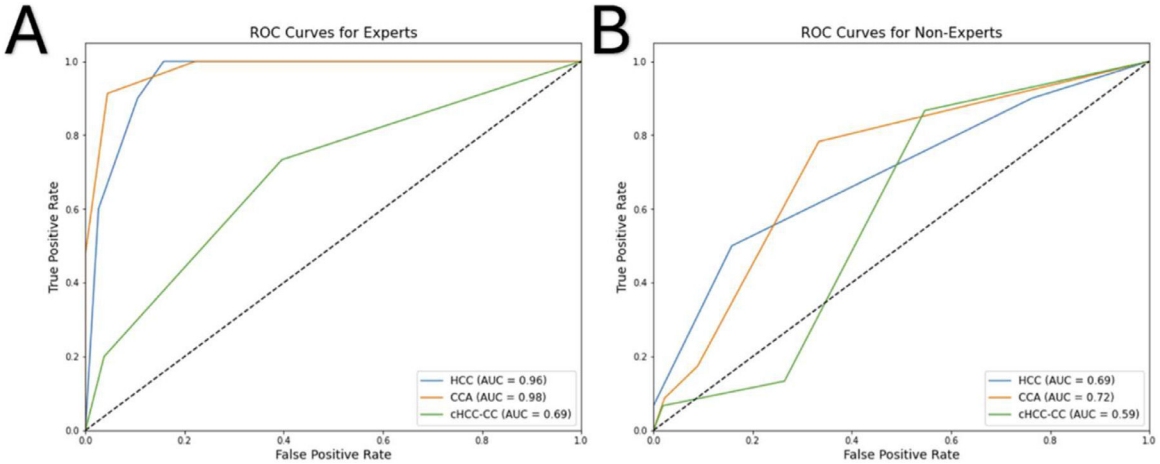

Fig. 4. Performance of AIEs (A) and NIEs (B) clearly highlighting the higher ability of AIEs to correctly diagnose HCC and CCA, nevertheless, both groups showed a low performance diagnosing cHCC—CC.

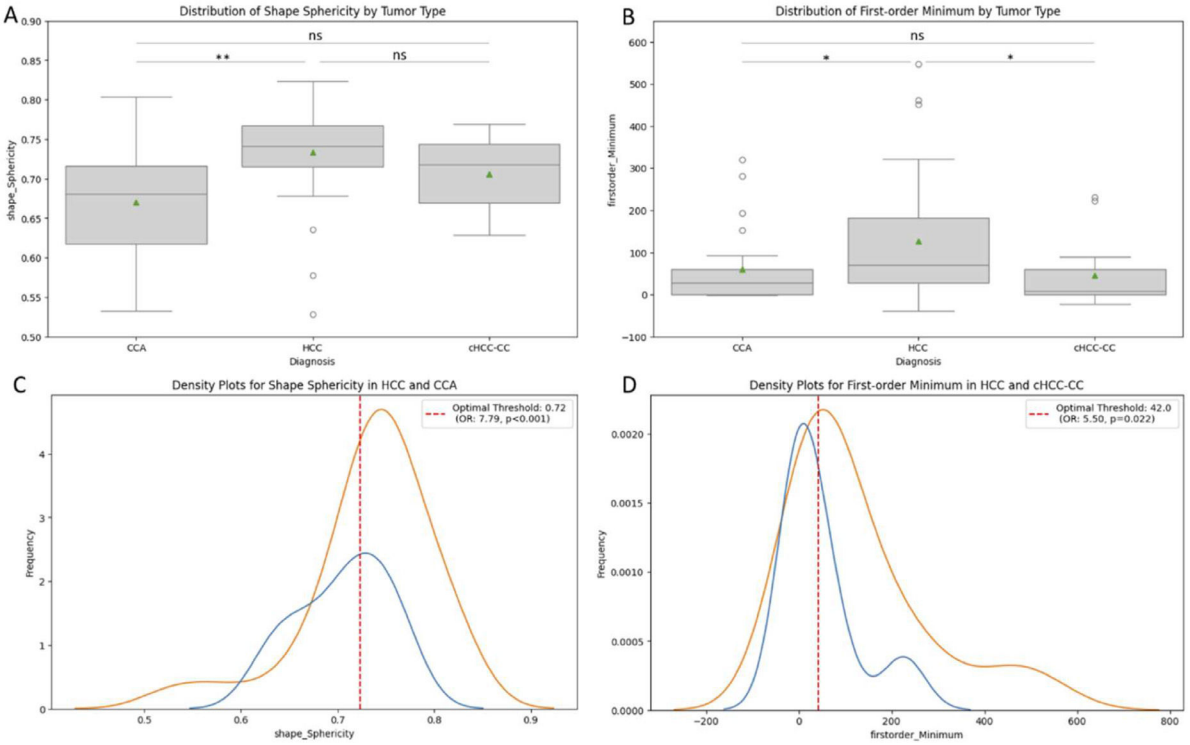

Fig. 5. Boxplots for 'Shape Sphericity'(A) and 'First-order Minimum'(B) according to tumor type (HCC, CCA, or cHCC—CC). These two features showed a significant difference in a univariate analysis (Kruskal-Wallis test) of these three groups. In subgroup analyses, Shape Sphericity displayed a significant difference between HCC vs. CCA with an optimal threshold of 0.72 (C). First-order Minimum differed significantly between HCC vs. cHCC—CC, with an optimal threshold of 42.0 (D), and between HCC vs. CCA.

27 % and 47 %, nevertheless, this did not reach statistical significance (AIE vs NIE $p = 0.59$).

Among the AIEs, the sensitivity in diagnosing HCC or CCA compared to cHCC—CC was significantly better (HCC vs cHCC—CC $p = 0.0029$, CCA vs cHCC—CC $p = 0.0026$) (Fig. 3A). The presence of cirrhosis did not have any significant influence on the proportion of correct diagnoses between the different liver tumor types ($p = 0.79$ for HCC and $p > 0.99$ for cHCC—CC) (Fig. 3B). There was no significant difference between the different geographical locations, neither as a whole ($p = 0.99$) nor when stratified according to the different types of liver tumors (all $p > 0.05$) (Fig. 3C-D).

NIEs achieved a slightly higher sensitivity of 40 % compared to AIEs at 25 %. This suggests that while AIEs are generally more effective, cHCC—CC tumors remain a challenging area for both groups (Fig. 4).

3.3. Radiomics performance to differentiate the different types of tumors

Given the inability of radiologists to effectively distinguish cHCC—CC from HCC and CCA, standard machine learning techniques were employed to further characterize radiomic features for HCC, CCA, and cHCC—CC. In a univariate analysis of all these three classes, 'Shape Sphericity', 'First-order Minimum', 'Shape Maximum2DDiameterSlice', and 'Shape Maximum3DDiameter' were the only four features with a significant difference ($p = 0.003$, 0.027, 0035, and 0.050, respectively), however, the last two were influenced by different tumor sizes (see Table 1). 'Shape sphericity' was significantly higher in the HCCs than in the CCAs ($p = 0.001$), with an AUROC of 0.76, and an odds ratio of 7.79 for the cut-off value of 0.72 (Fig. 5A, C). 'FirstOrder Minimum' was significantly higher in the HCCs than in the cHCCs-CCs ($p = 0.027$), with an AUROC of 0.70, and an odds ratio of 5.50 for the cut-off value of 42 (Fig. 5B, D), and it was also higher in the HCCs than the CCAs ($p = 0.031$). No feature differed significantly between CCA and cHCC—CC.

The pipeline for feature selection based on tuned model performance (multivariate analysis) repeatedly displayed that two features, namely 'Shape Sphericity' and 'GLCM ClusterShade', lead to the best model performance. Both features contributed to the model similarly (46 % vs. 54 %, respectively). The initial model with default hyperparameters was strongly overfitting, displaying an accuracy of 100 % and 65 % and AUROC of 1.00 and 0.82 in the training and testing set, respectively (Fig. 6). The model with optimized hyperparameters (criterion: entropy, max depth: 10, min samples leaf: 1, min samples split: 5, number of estimators: 25) showed an accuracy of 90 % and 76 % and AUROC of 0.98 and 0.91 in the training and testing set, respectively (Fig. 6). The recall and precision in the testing set were 100 % and 83 % for HCC, 83 % and 70 % for CCA, but only 25 % and 100 % for cHCC—CC.

Our radiomic-based model demonstrated sensitivity for detecting HCC comparable to the AIEs ($p = 0.604$), but significantly better than NIEs ($p = 0.002$). For CCA, the sensitivity of AIEs and the radiomic model was similar ($p = 1.0$), whereas NIEs showed significantly lower sensitivity ($p = 0.010$). In the case of cHCC—CC, our model did not achieve superior sensitivity compared to radiologists in either group ($p = 1.0$ for both AIEs and NIEs). Overall, when considering radiologists as a whole, our model did not outperform them in any of the three classes ($p = 0.065$ for HCC, $p = 0.426$ for CCA, and $p = 1.0$ for cHCC—CC) (see Table 2B).

4. Discussion

Radiological imaging plays an essential role in the diagnosis and staging of HCC. In fact, it is likely the only cancer to be diagnosed solely on the basis of radiological imaging. This is mostly due to the typical characteristics of HCC on multiphase CT and Gd-enhanced MRI, which led to the development of the widely accepted LI-RADS scoring system [24]. Although some clinicians have controversially discussed the approach of a solely imaging-based diagnosis of HCC [25], current guidelines have widely accepted non-invasive diagnosis

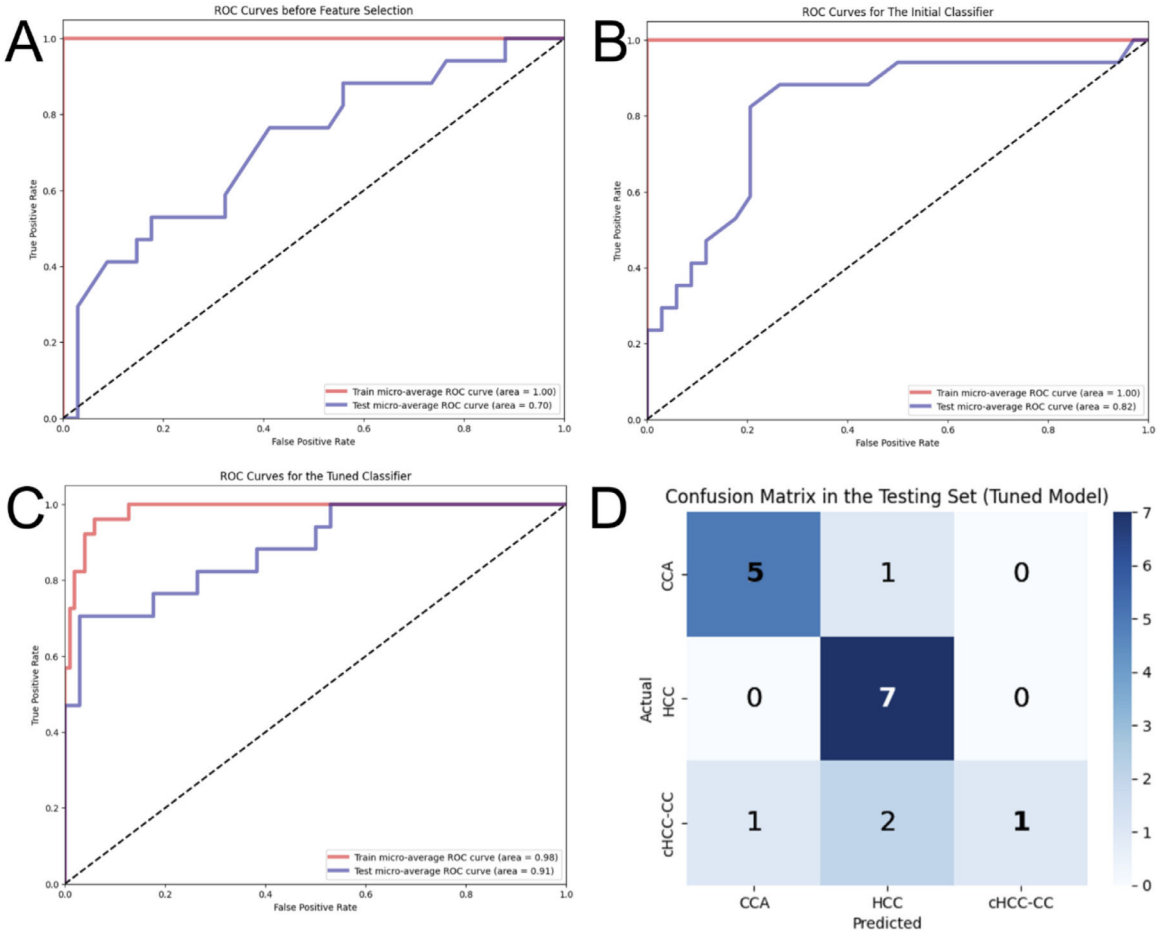

Fig. 6. Performance of three models built by random forest: Receiver Operating Characteristic (ROC) curves in the training and testing set before feature selection (A), after feature selection (B), and after hyperparameter tuning (C). The confusion matrix (D) shows predictions made by the tuned model on the samples in the testing set.

of HCC in high-risk patients [15–17]. However, the definition of high-risk patients varies. For example, the EASL considers high-risk only patients with cirrhosis [15] and APASL guidelines also include patients with chronic viral hepatitis B or C (HBV, HBC) [17]. The recent AASLD guidelines present a compromised solution, incorporating PAGE-B risk score for HBV patients and excluding patients with cirrhosis of vascular etiology [16]. Even though in most cases imaging detects HCC with high accuracy, radiology-based liver cancer diagnosis still has its limitations. Sensitivity is relatively low in nodules <20 mm, and the recommendations cannot be applied to nodules <10 mm at all [26]. A diagnostic biopsy should be performed in low-risk patients or when the findings are inconclusive on both MRI and CT scans (LI-RADS categories 4 and M) [15–17].

In contrast, in CCA tumor biopsy was still considered necessary either from biopsy or resection [5,6,27]. This allows not only histological verification but also genetic testing as more than 40 % of advanced CCAs may harbor druggable molecular targets [28].

Combined HCC—CC may represent only a small fraction of primary liver cancers, which still corresponds to thousands of new patients worldwide every year [29]. Furthermore, the disease may be underdiagnosed, as it may resemble other liver malignancies, especially HCC, where the diagnosis can be made without histological assessment.

However, evidence shows that around 40 % of histologically proven cHCC—CCs were radiologically assigned to LI-RADS category 4 or 5 [30,31]. In another study, 54 % met all the major criteria for HCC although 88 % of those had at least one ancillary feature favoring non-HCC malignancy [32]. Overall, there is no reliable method for cHCC—CC recognition and as a result, the sensitivity by contrast-enhanced CT or MRI is reported to be only 33 % [11]. A similar number was achieved by our international group of radiologists, where the mean sensitivity reached 31 %. Interestingly, the low diagnostic accuracy was consistent between all countries and levels of expertise. Thus overall, cHCC—CC might be easily misclassified as HCC and standard imaging seems insufficient for cHCC—CC diagnosis.

Given the difficulties of radiologists in detecting cHCC—CC by means of conventional radiology procedures, we applied machine learning and radiomics algorithms in order to potentially enhance diagnostic accuracy. Those radiomics approaches certainly offer new avenues for classifying liver cancer and represent the current cutting edge of oncology research, with rapidly growing popularity [21]. Over the past several years, multiple models were developed to classify HCC and CCA from standard imaging methods, outperforming radiologic evaluation in diagnostic accuracy [33–36]. Li et al. included cHCC—CC cases in a contrast-enhanced ultrasound-based model [12]. Unfortunately, it was only trained to discriminate between HCC and non-HCC (including CCA + cHCC—CC) tumors. Guo et al. developed a logistic regression model using clinical and radiomic features from a large cohort and various MRI sequences [37]. Their hybrid model, based on the portal venous phase, achieved an accuracy of 80 % and an AUROC of 0.88. However, it only differentiated between HCC and cHCC—CC, excluding the significant class of CCA. Furthermore, a study by Zhen et al. integrated radiomics features with clinical data and demonstrated the potential for comprehensive diagnostic algorithms to be capable of accurately classifying different tumor types [38]. Also, Xia et al. highlight the utility of radiomics in predicting microvascular invasion, a key determinant of HCC prognosis, and its association with recurrence-free and overall survival [39].

Table 2B
Correct diagnoses and sensitivity for each radiologist.

Cancer types	Number of cases	Sensitivity								
		Non-abdominal imaging expert (NIE) / Trainee				Abdominal imaging expert (AIE)				Total weighted average
		Reviewer 1 (Germany)	Reviewer 2 (Mexico)	Reviewer 3 (China)	Group Average	Reviewer 4 (Germany)	Reviewer 5 (Colombia)	Reviewer 6 (Mexico)	Reviewer 7 (China)	Group Average	
HCC	30	8 (27 %)	17 (57 %)	20 (67 %)	45 (50 %)	27 (90 %)	27 (90 %)	21 (70 %)	30 (100 %)	105 (88 %)	150 (71 %)
With cirrhosis	22	6 (27 %)	15 (68 %)	16 (73 %)	37 (56 %)	20 (90 %)	21 (96 %)	17 (77 %)	22 (100 %)	80 (91 %)	80 (76 %)
Without cirrhosis	8	2 (25 %)	2 (25 %)	4 (50 %)	8 (33 %)	7 (88 %)	6 (75 %)	4 (50 %)	8 (100 %)	25 (78 %)	33 (59 %)
CCA	23	14 (61 %)	5 (22 %)	7 (30 %)	26 (38 %)	17 (74 %)	20 (87 %)	19 (83 %)	21 (91 %)	77 (84 %)	103 (64 %)
With cirrhosis	0										
Without cirrhosis	23	14 (61 %)	5 (22 %)	7 (30 %)	26 (38 %)	17 (74 %)	20 (87 %)	19 (83 %)	21 (91 %)	77 (84 %)	103 (64 %)
cHCC−CC	15	7 (47 %)	7 (47 %)	4 (27 %)	18 (40 %)	5 (33 %)	1 (7 %)	8 (53 %)	1 (7 %)	15 (25 %)	33 (31 %)
With cirrhosis	10	5 (50 %)	4 (40 %)	2 (20 %)	11 (37 %)	5 (50 %)	1 (10 %)	4 (40 %)	0 (0 %)	10 (25 %)	21 (30 %)
Without cirrhosis	5	2 (40 %)	3 (60 %)	2 (40 %)	7 (47 %)	0 (0 %)	0 (0 %)	4 (80 %)	1 (20 %)	5 (25 %)	12 (34 %)

Detailed table showing the number of correct diagnoses and sensitivity for each radiologist and tumor type as a whole, with, and without cirrhosis.

In an univariate analysis, only one feature (first-order minimum) was significantly different between HCC and cHCC—CC and no one between CCA and cHCC—CC. In a multivariate analysis, our optimized random forest model successfully classified most HCC and CCA cases while the sensitivity for cHCC—CC was low. These results suggest that current standard machine learning-based radiomic approaches will not even come close to substantial clinical support in imaging-based diagnosis of cHCC—CC when both CCA and HCC are considered differential diagnoses. This underscores the need for further refinement of machine learning approaches or the exploration of advanced algorithms that might improve the diagnostic accuracy of radiomics-based characterization in liver cancers.

Our study encountered several limitations, including the relatively small sample size of cHCC—CC tumors, potential bias from imaging resampling, and overfitting risk suggested by pre-tuning 100 % training accuracy. Further exploration and validation in multicenter and larger cohorts is warranted. Despite these limitations, we consider that liver biopsy remains indispensable for the diagnosis and management of liver cancer, particularly for cHCC—CC, until more advanced radiomics algorithms and biomarkers become available. While radiomics techniques hold promise for non-invasive characterization of liver lesions [40], they may currently not provide sufficient information to guide clinical decision-making, especially in complex cases such as cHCC—CC cancers, in order to support radiologist and increase diagnostic accuracy. However, as the treatment of HCC and CCA and thus also cHCC—CC significantly differs, efforts to further improve radiology and machine learning-based diagnosis of liver cancer should continue to improve the accuracy while also exploring novel non-invasive diagnostic approaches to enhance its urgently needed clinical utility.

5. Conclusions

Current standard imaging and machine learning-based radiomics analysis algorithms were insufficient to reliably characterize cHCC—CC. Therefore, combing imaging with biopsy of liver cancer remains critical to the detection, diagnosis, and effective treatment of these cancers.

Funding

AT received grants from the Sino-German Center for Research Promotion (grant numbers: GZ-1546 and C-0012), the State Ministry of Baden-Wuerttemberg for Sciences, Research and Arts supporting the Clinical Cooperation Unit Healthy Metabolism at the Center for Preventive Medicine and Digital Health (grant identifier: CCU Healthy Metabolism), the Foundation for Biomedical Alcohol Research (grant identifier: N/A), the Baden-Wuerttemberg Center for Digital Early Disease Detection and Prevention (grant identifier: BW- ZDFP), and the Federal Ministry of Science and Education of Germany (identifier: Q-HCC, 01KD2214).

Authors contributions

Image Acquisition: YQ, QL, YS. Data analysis: YQ, IR, XM, MÜ, GC, AS, MH, PW, AV, CRA, AGM, ZG, CW, MF, AT. Expert advice: SS, ME. Manuscript writing and revision: IR, MV, YQ, AT
Final approval: all authors

Declaration of interests

None.

References

[1] Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021;149:778–89. https://doi.org/10.1002/ijc.33588.
[2] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin 2024:21834 caac. https://doi.org/10.3322/caac.21834.
[3] Altekruse Sean F, Kleiner David E. Histological classification of liver and intrahepatic bile duct cancers in SEER registries. J Registry Manag 2011;38:201–5.
[4] Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado Á, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. J Hepatol 2022;76:681–93. https://doi.org/10.1016/j.jhep.2021.11.018.
[5] Alvaro D, Gores GJ, Walicki J, Hassan C, Sapisochin G, Komuta M, et al. EASL-ILCA Clinical Practice Guidelines on the management of intrahepatic cholangiocarcinoma. J. Hepatol. 2023;79:181–208. https://doi.org/10.1016/j.jhep.2023.03.010.
[6] Vogel A, Bridgewater J, Edeline J, Kelley RK, Klümpen HJ, Malka D, et al. Biliary tract cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2023;34:127–40. https://doi.org/10.1016/j.annonc.2022.10.506.
[7] Jarnagin WR, Weber S, Tickoo SK, Koea JB, Obiekwe S, Fong Y, et al. Combined hepatocellular and cholangiocarcinoma: demographic, clinical, and prognostic factors. Cancer 2002;94:2040–6. https://doi.org/10.1002/cncr.10392.
[8] Lee C-C, Wu C-Y, Chen J-T, Chen G-H. Comparing combined hepatocellular-cholangiocarcinoma and cholangiocarcinoma: a clinicopathological study. Hepatogastroenterology 2002;49:1487–90.
[9] Garancini M, Goffredo P, Pagni F, Romano F, Roman S, Sosa JA, et al. Combined hepatocellular-cholangiocarcinoma: a population-level analysis of an uncommon primary liver tumor: combined hepatocellular-cholangiocarcinoma. Liver Transpl 2014;20:952–9. https://doi.org/10.1002/lt.23897.
[10] Stavraka C, Rush H, Ross P. Combined hepatocellular cholangiocarcinoma (cHCC-CC): an update of genetics, molecular biology, and therapeutic interventions. J Hepatocell Carcinoma 2018;6:11–21. https://doi.org/10.2147/JHC.S159805.
[11] Fowler KJ, Sheybani A, Parker RA, Doherty S, M, Brunt E, Chapman WC, et al. Combined hepatocellular and cholangiocarcinoma (Biphenotypic) tumors: imaging features and diagnostic accuracy of contrast-enhanced CT and MRI. Am J Roentgenol 2013;201:332–9. https://doi.org/10.2214/AJR.12.9488.
[12] Li R, Yang D, Tang C-L, Cai P, Ma K, Ding S-Y, et al. Combined hepatocellular carcinoma and cholangiocarcinoma (biphenotypic) tumors: clinical characteristics, imaging features of contrast-enhanced ultrasound and computed tomography. BMC Cancer 2016;16:158. https://doi.org/10.1186/s12885-016-2156-x.
[13] Gigante E, Ronot M, Bertin C, Ciolina M, Bouattour M, Dondero F, et al. Combining imaging and tumour biopsy improves the diagnosis of combined hepatocellular-cholangiocarcinoma. Liver Int 2019;39:2386–96. https://doi.org/10.1111/liv.14261.
[14] Kobayashi S, Terashima T, Shiba S, Yoshida Y, Yamada I, Iwadou S, et al. Multicenter retrospective analysis of systemic chemotherapy for unresectable combined hepatocellular and cholangiocarcinoma. Cancer Sci 2018;109:2549–57. https://doi.org/10.1111/cas.13656.
[15] Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul J-L, et al. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. J Hepatol 2018;69:182–236. https://doi.org/10.1016/j.jhep.2018.03.019.
[16] Singal AG, Llovet JM, Yarchoan M, Mehta N, Heimbach JK, Dawson LA, et al. AASLD Practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology 2023;78:1922–65. https://doi.org/10.1097/HEP.0000000000000466.
[17] Omata M, Cheng A-L, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia–Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatol Int 2017;11:317–70. https://doi.org/10.1007/s12072-017-9799-9.
[18] Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323–41. https://doi.org/10.1016/j.mri.2012.05.001.
[19] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339.
[20] 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. https://doi.org/10.1148/radiol.2020191145.
[21] Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, et al. Radiomics in oncology: a practical guide. RadioGraphics 2021;41:1717–32. https://doi.org/10.1148/rg.2021210037.
[22] McHugh ML. Interrater reliability: the kappa statistic. Biochem Medica 2012;22:276–82.
[23] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
[24] Santillan C, Chernyak V, Sirlin C. LI-RADS categories: concepts, definitions, and criteria. Abdominal Radiology 2018;43:101–10. https://doi.org/10.1007/s00261-017-1334-x.
[25] Schirmacher P, Bedossa P, Roskams T, Tiniakos DG, Brunt EM, Zucman-Rossi J, et al. Fighting the bushfire in HCC trials. J Hepatol 2011;55:276–7. https://doi.org/10.1016/j.jhep.2011.03.004.
[26] Lee YJ, Lee JM, Lee JS, Lee HY, Park BH, Kim YH, et al. Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging—A systematic review and meta-analysis. Radiology 2015;275:97–109. https://doi.org/10.1148/radiol.14140690.
[27] Bowlus CL, Arrivé L, Bergquist A, Deneau M, Forman L, Ilyas SI, et al. AASLD practice guidance on primary sclerosing cholangitis and cholangiocarcinoma. Hepatology (Baltimore, Md) 2023;77:659–702. https://doi.org/10.1002/hep.32771.
[28] Silverman IM, Murugesan K, Lihou CF, Féliz L, Frampton GM, Newton RC, et al. Comprehensive genomic profiling in FIGHT-202 reveals the landscape of actionable alterations in advanced cholangiocarcinoma. J Clin Oncol 2019;37:4080.–4080. https://doi.org/10.1200/JCO.2019.37.15_suppl.4080.
[29] Teufel A, Rodriguez I, Winzler C, Kokh D, Ebert MP, Surovtsova I, et al. Clinical characterization of HCC/CCA mixed cancers in a population-based cohort. J Gastrointestin Liver Dis 2023;32:190–6. https://doi.org/10.15403/jgld-4893.
[30] Jeon SK, Joo I, Lee DH, Lee SM, Kang H-J, Lee K-B, et al. Combined hepatocellular cholangiocarcinoma: LI-RADS v2017 categorisation for differential diagnosis and prognostication on gadoxetic acid-enhanced MR imaging. Eur Radiol 2019;29:373–82. https://doi.org/10.1007/s00330-018-5605-x.
[31] Kim MY, Joo I, Kang HJ, Bae JS, Jeon SK, Lee JM. LI-RADS M (LR-M) criteria and reporting algorithm of v2018: diagnostic values in the assessment of primary liver cancers on gadoxetic acid-enhanced MRI. Abdominal Radiology 2020;45:2440–8. https://doi.org/10.1007/s00261-020-02545-z.
[32] Potretzke TA, Tan BR, Doyle MB, Brunt EM, Heiken JP, Fowler KJ. Imaging features of biphenotypic primary liver carcinoma (Hepatocholangiocarcinoma) and the potential to mimic hepatocellular carcinoma: LI-RADS analysis of CT and MRI features in 61 cases. Am J Roentgenol 2016;207:25–31. https://doi.org/10.2214/AJR.15.14997.
[33] Xu X, Mao Y, Tang Y, Liu Y, Xue C, Yue Q, et al. Classification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma based on radiomic analysis. Comput Math Methods Med 2022;2022:1–9. https://doi.org/10.1155/2022/5334095.
[34] Liu N, Wu Y, Tao Y, Zheng J, Huang X, Yang L, et al. Differentiation of hepatocellular carcinoma from intrahepatic cholangiocarcinoma through MRI radiomics. Cancers (Basel) 2023;15:5373. https://doi.org/10.3390/cancers15225373.
[35] Mahmoudi S, Bernatz S, Ackermann J, Koch V, Dos Santos DP, Grünewald LD, et al. Computed tomography radiomics to differentiate intrahepatic cholangiocarcinoma and hepatocellular carcinoma. Clin Oncol 2023;35:e312–8. https://doi.org/10.1016/j.clon.2023.01.018.
[36] Ponnoprat D, Inkeaw P, Chaijaruwanich J, Traisathit P, Sripan P, Inmutto N, et al. Classification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma based on multi-phase CT scans. Med Biol Eng Comput 2020;58:2497–515. https://doi.org/10.1007/s11517-020-02229-2.
[37] Guo L, Li X, Zhang C, Xu Y, Han L, Zhang L. Radiomics based on dynamic contrast-enhanced magnetic resonance imaging in preoperative differentiation of combined hepatocellular-cholangiocarcinoma from hepatocellular carcinoma: a Multi-center study. Journal of Hepatocellular Carcinoma 2023;10:795–806. https://doi.org/10.2147/JHC.S406648.
[38] Zhen S, Cheng M, Tao Y, Wang Y, Juengpanich S, Jiang Z, et al. Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data. Front Oncol 2020;10:680. https://doi.org/10.3389/fonc.2020.00680.
[39] Xia T, Zhou Z, Meng X, Zha J, Yu Q, Wang W, et al. Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. Radiology 2023;307:e222729. https://doi.org/10.1148/radiol.222729.
[40] Calderaro J, Seraphin TP, Luedde T, Simon TG. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol 2022;76:1348–61. https://doi.org/10.1016/j.jhep.2022.01.014.