La radiologia medica (2024) 129:1099–1117
https://doi.org/10.1007/s11547-024-01840-9

ABDOMINAL RADIOLOGY

Radiomics as a tool for prognostic prediction in transarterial chemoembolization for hepatocellular carcinoma: a systematic review and meta-analysis

Kaige Deng[1] · Tong Chen[2] · Zijian Leng[1] · Fan Yang[1] · Tao Lu[3] · Jingying Cao[4] · Weixuan Pan[1] · Yongchang Zheng[1]

Received: 1 March 2024 / Accepted: 25 June 2024 / Published online: 26 July 2024
© The Author(s) 2024

Abstract

Introduction Transarterial chemoembolization (TACE) is one of the predominant locoregional therapeutic modalities for addressing hepatocellular carcinoma (HCC). However, achieving precise prognostic predictions and effective patient selection remains a challenging pursuit. The primary objective of this systematic review and meta-analysis is to evaluate the efficacy of radiomics in forecasting the prognosis associated with TACE treatment.

Methods A comprehensive exploration of pertinent original studies was undertaken, encompassing databases of PubMed, Web of Science and Embase. The studies' quality was meticulously evaluated employing the quality assessment of diagnostic accuracy studies 2 (QUADAS-2), the radiomics quality score (RQS) and the METhodological RadiomICs Score (METRICS). Pooled statistics, along with 95% confidence intervals (95% CI), were computed for sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR). Additionally, a summary receiver operating characteristic curve (sROC) was generated. To discern potential sources of heterogeneity, meta-regression and subgroup analyses were performed.

Results The systematic review incorporated 29 studies, comprising a total of 5483 patients, with 14 studies involving 2691 patients qualifying for inclusion in the meta-analysis. The assessed studies exhibited commendable quality with regard to bias risk, with mean RQS of 12.90 ± 5.13 (35.82% ± 14.25%) and mean METRICS of 62.98% ± 14.58%. The pooled sensitivity was 0.83 (95% CI: 0.78–0.87), specificity was 0.86 (95% CI: 0.79–0.92), PLR was 6.13 (95% CI: 3.79–9.90), and NLR was 0.20 (95% CI: 0.15–0.27). The area under the sROC was 0.90 (95% CI: 0.87–0.93). Significant heterogeneity within all the included studies was observed, while meta-regression and subgroup analyses revealed homogeneous and promising findings in subgroups where principal methodological variables such as modeling algorithms, imaging modalities, and imaging phases were specified.

Conclusion Radiomics models have exhibited robust predictive capabilities concerning prognosis subsequent to TACE, thereby presenting promising prospects for clinical translation.

Keywords Hepatocellular carcinoma (HCC) · Transarterial chemoembolization (TACE) · Prognosis · Radiomics · Systematic review · Meta-analysis

Abbreviations

TACE	Transarterial chemoembolization
cTACE	Conventional TACE
DEB-TACE	Drug-eluting bead TACE
HCC	Hepatocellular carcinoma
MVI	Microvascular invasion
NSCLC	Non-small cell lung cancer
RCC	Renal cell carcinoma
BCLC	Barcelona clinic liver cancer
CNLC	China liver cancer staging
RECIST	Response evaluation criteria in solid tumors
mRECIST	Modified response evaluation criteria in solid tumors
OR	Objective response
ORR	Objective response rates
OS	Overall survival
MR	Magnetic resonance
Gd-EOB-DTPA	Gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid
CT	Computed tomography
US	Ultrasound
DSA	Digital subtraction angiography

Extended author information available on the last page of the article

LI-RADS	Liver imaging reporting and data system
AP	Arterial phase
PVP	Portal venous phase
VOIs	Volumes of interest
ROIs	Regions of interest
PRISMA-DTA	Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies
PICO	Participants, intervention, control, outcome
QUADAS-2	Quality assessment of diagnostic accuracy studies 2
RQS	Radiomics quality score
METRICS	METhodological RadiomICs Score
PLR	Positive likelihood ratio
NLR	Negative likelihood ratio
sROC	Summary receiver operating characteristic curve
AUC	Area under curve
95% CI	95% confidence intervals
TP	True positive
TN	True negative
FP	False positive
FN	False negative
DL	Deep learning
ML	Machine learning

Introduction

Hepatocellular carcinoma (HCC) represents the most prevalent form of primary liver cancer, ranking sixth globally in cancer incidence and third in cancer-related mortality. In 2020, there were approximately 906,000 new cases and 830,000 deaths attributed to HCC worldwide [1]. Recognized for its highly aggressive nature, HCC is associated with a generally poor prognosis. Age-standardized 5-year survival rates exhibit regional variations, typically ranging from 5% to 30%, and falling within 10% to 19% in most countries [2]. Therapeutic advancements have contributed to an enhanced outcome for HCC patients, with observed increases of 5% to 10% in age-standardized 5-year survival rates across different regions from 1995 to 2014 [2, 3]. Current therapeutic strategies highlight optimal management through comprehensive approaches, primarily centered around surgical interventions [4–6]. Non-surgical treatments, including transarterial chemoembolization, hepatic arterial infusion chemotherapy [7], ablations [8–10], radiotherapy [11], and systemic therapies, among others, have provided opportunities to improve the prognosis, particularly for early-stage patients unsuitable for operations and those with advanced unresectable HCC.

Transarterial chemoembolization (TACE) stands as a common locoregional treatment for liver cancer. Depending on the therapeutic agents employed, it can be broadly classified into conventional TACE (cTACE), utilizing a mixture of iodized oil and chemotherapeutic drugs, and drug-eluting bead TACE (DEB–TACE) [12]. TACE is considered the preferred therapeutic modality for liver-limited diseases deemed unsuitable for surgical resection [4], especially recommended for intermediate-stage tumors according to the Barcelona clinic liver cancer (BCLC) classification [5]. Notably, TACE finds application in various scenarios, such as the management of liver cancer with localized portal vein tumor thrombosis with incomplete obstruction, bridging therapy for liver transplantation, postoperative adjuvant therapy for patients at high risk of recurrence, and conversion therapy for unresectable tumors [4–6, 13], among others. In real-world practice, TACE is commonly employed in patients with recurrent or progressed HCC and is often repeated following the initial treatment [12]. This renders the patient selection and efficacy evaluation susceptible to the influences of previous or concurrent anti-tumor treatments. Consequently, the choice and application of TACE heavily depend on the expertise and preferences of physicians, lacking well-recognized supportive informative tools.

The assessment of tumor response and long-term prognosis subsequent to TACE primarily relies on radiological criteria, such as the response evaluation criteria in solid tumors (RECIST) [14] and modified response evaluation criteria in solid tumors (mRECIST) [15, 16]. According to these criteria, the objective response rates (ORR) after TACE exhibit notable diversity across different reports, ranging broadly from 30% to 80% [17–21]. This substantial variability may be linked to factors such as distinct baseline patient characteristics, the intricacy of background therapies before TACE, and variations in TACE techniques. Consequently, predicting prognosis following TACE could enhance the process of individualized patient selection, ultimately improving clinical outcomes [13]. Researchers have explored diverse indicators for post-TACE prognosis, encompassing radiological features [22, 23], clinico-pathological factors [24], serum markers [25], and genomic alterations [26], among others. However, these predictive tools remain hindered by limitations such as insufficient predictive power, lack of validations, and the associated trauma or cost related to these tests.

First systematically defined by Lambin et al. in 2012 [27], radiomics is a technique leveraging computer programs to explore and analyze high-throughput quantitative features from medical images. Since its inception, the influence of radiomics has progressively expanded. Radiomics has primarily been applied to aid in cancer diagnosis and prognostic prediction, offering robust support for precision medicine. By transforming medical images into a high-dimensional space of features for analysis and model construction,

radiomics contributes to an enhanced understanding of medical images. In response to the unmet demand for personalized TACE, many researchers have developed radiomics models to predict prognosis after TACE. These studies have analyzed data from various imaging modalities, including magnetic resonance (MR) [21, 28–36], computed tomography (CT) [20, 37–53], as well as ultrasound (US) [54], reporting promising results. Recognizing the significance of developing reliable tools for prognostic prediction in TACE, we deem it timely and crucial to summarize the predictive performance and potential for clinical translation of radiomics models in this task. Therefore, this systematic review and meta-analysis aimed to determine the value of radiomics in predicting the prognosis of TACE treatment.

Methods

This systematic review adheres to the preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA) guidelines [55] as well as the assessing the quality of systematic reviews 2 (AMSTAR 2) guidelines [56]. The protocol was prospectively registered on PROSPERO (ID: CRD42023449278).

Literature retrieving

A computerized search was executed on PubMed, Web of Science and Embase databases to identify original studies implementing radiomics analysis of preprocedural images to predict the therapeutic outcomes of TACE. The search strategy was devised following the PICO (participants, intervention, control, outcome) principles [55] to facilitate a sensitive screening for relevant studies. The search terms encompassed four perspectives, including "hepatocellular carcinoma or liver cancer," "radiology or radiomics," "TACE," and "prognosis." The search scope included literature published up to May 16, 2023. Detailed search conditions are provided in Supplementary Table 1.

Study inclusion

After eliminating duplicates, the remaining literature underwent a comprehensive review based on the following inclusion criteria: (1) population: patients diagnosed with HCC through clinical or pathological confirmations; (2) intervention: initial TACE treatment (without prior history of TACE treatment); (3) index test: radiomics analysis/high-throughput analysis/artificial intelligence analysis based on preprocedural medical imaging; (4) outcomes: prognosis after TACE, including tumor response, overall survival, recurrence, progression, etc. Studies were then excluded based on the following criteria: (1) studies involving non-human subjects; (2) conference abstracts, reviews, case reports, commentaries, and other non-original publications; (3) studies not published in English; (4) studies that did not report standalone radiomics models without combining clinical features; (5) studies that did not provide sufficient raw data for qualitative or quantitative review.

The identified articles underwent an initial screening based on titles and abstracts. Subsequently, a full-text review was conducted for potentially eligible articles. The literature was reviewed by two authors (KGD, FY) with over four years of experience in hepatic surgery and medical imaging analysis. Any uncertainties were resolved through consensus in an author team, including a senior author with over fifteen years of experience in the management of liver cancers and meta-analysis (YCZ). When sufficient data were still unavailable after the full-text review, attempts were made to acquire the original data by contacting the corresponding authors. Endnote (version 20, Clarivate Analytics, England) was utilized for all literature management.

Data extraction

Predefined information was systematically extracted from the included literature according to the following categories: (1) bibliographic features, including authors, publication year, country, and study type; (2) patient characteristics, encompassing (2–1) demographic features such as sample size, age and gender; (2–2) etiological features, such as hepatitis viral infection status and liver cirrhosis; (2–3) clinical-pathological features, including pathological diagnosis and tumor staging; (3) procedural information related to TACE; (4) pre-treatment radiological examination details, covering imaging modalities, instrument parameters, and phases or sequences of the images; (5) post-treatment follow-up information, comprising follow-up endpoints, assessment criteria, follow-up intervals, and follow-up examinations; (6) methodological characteristics of radiomics analysis, involving image segmentation methods, feature extraction methods, modeling algorithms, predictive outcomes, etc.; (7) model validation methods; (8) model performance parameters, such as sensitivity, specificity, concordance index, area under the curve (AUC), number of true positive (TP), false positive (FP), true negative (TN), false negative (FN) quantities, etc.; and (9) other relevant information. Detailed data extraction strategies are presented in Supplementary Table 2.

In instances where a single study involved the construction of multiple radiomics models, the model demonstrating the highest performance was selected for data extraction. When a study presented results for both training and validation sets [57], outcomes were extracted separately from each dataset. If an article reported both pure radiomics models and combined models incorporating additional clinical features, only results from the pure radiomics models were

extracted, while the supplementary clinical features were retained for study quality evaluation.

Quality assessment

The quality of the included studies was evaluated using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) [58], the radiomics quality score (RQS) [59] criteria, and the METhodological RadiomICs Score (METRICS) [60]. QUADAS-2 was employed to assess the risk of bias across four domains: patient selection, index test, reference standard, and flow and timing within the studies. RQS is a specialized rating framework for radiomics, covering sixteen aspects such as image acquisition, feature extraction, modeling methods, model validation, and clinical applications. And the METRICS is a newly proposed assessment tool for radiomics studies, which is based on international consensus and a well-constructed framework including ranking and weighting of methodological variations (refer to Supplementary Table 3 for details).

Throughout the data extraction and quality assessment process, a systematic approach was adopted. Initially, a training phase was initiated, during which three authors (KGD, ZJL, YCZ) independently reviewed three included papers. These authors aligned their interpretations of each item in the data extraction forms and the quality rating forms, with all definitions ultimately confirmed by senior author YCZ. Subsequently, two authors (KGD, ZJL) independently conducted data extraction and quality assessment. METRICS scores were assessed using the convenient web-based tool at https://metricsscore.github.io/metrics/METRICS.html. The final electronic datasheet integrated information unanimously confirmed by both authors.

Statistical analysis

Statistical analysis and visualization were carried out using Microsoft Office Excel 2019, RevMan 5.3.5 [61], and Stata Statistical Software 17 (StataCorp., T.X., USA) [62]. 2 × 2 diagnostic contingency tables, containing numbers of TP, FP, FN, and TN, were extracted wherever possible. Meta-analysis was conducted using the Stata midas package [63]. Pooled sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) were calculated as synthetic statistics. A PLR > 10 or within 5–10, along with an NLR < 0.1 or within 0.1–0.2, indicates high or moderate informative value, respectively. Cochrane's Q-test and I^2 were utilized to assess inter-study heterogeneity, with $I^2 > 50\%$ or $P < 0.1$ indicating the presence of heterogeneity. The Galbraith method was employed to identify the impact of outliers. The summary receiver operating characteristic curve (sROC) was used to evaluate the overall predictive performance of different radiomics models. An AUC of 0.5–0.7, 0.7–0.9, and > 0.9 indicates low, moderate, and high predictive power, respectively [64]. Meta-regression was used to explore the relationship between major methodological factors and inter-study heterogeneity, followed by further subgroup analyses. Deek's funnel plot and asymmetry test were applied to evaluate publication bias. Finally, the clinical application of radiomics models was assessed. The calculated average ORR in included studies was used as prior probability, and then post-test probability was calculated using Fagan plot based on the pooled PLR and NLR and Bayesian conditional probability theorem.

Results

Literature retrieving, selection and data extraction

The literature retrieval process identified a total of 645 relevant articles, and 203 duplicated articles were subsequently removed. From the remaining 442 studies, scrutiny of titles and abstracts led to the exclusion of 396 articles that did not align with the language, study type, or PICO criteria specified for our review. After a thorough review of full texts, an additional 17 articles were excluded, resulting in the final inclusion of 29 articles (Fig. 1).

Summarized information for all 29 included studies is presented in Table 1, covering a total of 5483 patients. Details of the studies are presented in Supplementary Table 4, with the primary research features illustrated in Supplementary Fig. 1. The majority of the articles originated from China (24 articles, 82.8%). The included studies date back to 2016, with a progressive increase in the number of publications, peaking at 19 studies in 2021–2022. All included studies adopted a retrospective design. Most studies utilized the BCLC staging system to determine the inclusion of patients, while others used the China liver cancer staging (CNLC) system [32] or other criteria [21, 31, 40, 43, 48]. Mid-stage patients were the most extensively studied population, with 21 articles (in whole or in part) including BCLC-B stage patients [20, 28–30, 33, 37, 41, 28–30, 49, 53], followed by BCLC-A (13 articles) and BCLC-C stage (9 articles) patients.

In terms of imaging modalities, all studies utilized contrast-enhanced examinations. The most frequently investigated imaging modality was contrast-enhanced CT [20, 36–53], followed by contrast-enhanced MR [21, 28–35], with only one article exploring contrast-enhanced ultrasound [54]. Various phases or sequences of contrast-enhanced images were studied, with 18 articles incorporating single-phase/sequence images, while the remaining literature analyzed multi-phase/sequence images. The most commonly utilized phases for radiomics analysis were the arterial phase (AP) [21, 39, 42, 44, 46, 49] and the portal venous phase

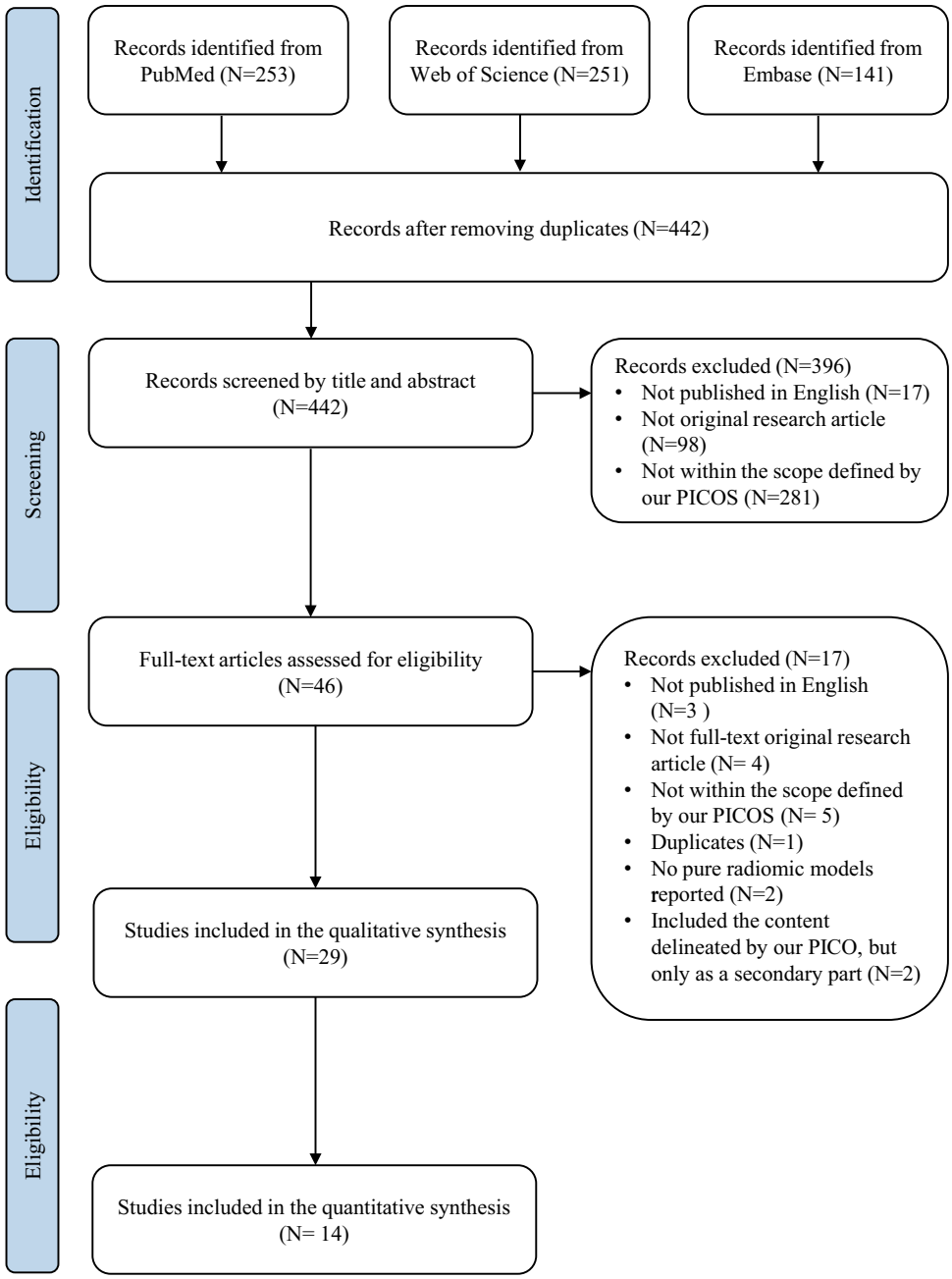

Fig. 1 PRISMA flowchart illustrating the literature selection process in this study.

(PVP) [31, 34, 36, 38, 48, 50]. Regarding MR-based radiomics, T2WI was the most frequently analyzed sequence [29, 32, 35]. Several studies compared radiomics models constructed from images of different phases/sequences, but the conclusions were inconsistent. Some indicated that multi-phase/sequence models were superior [20, 30, 42] in performance, while others demonstrated that single-phase/sequence models were comparable to, or even better than, the multi-phase/sequence models [35, 48]. In addition, comparative studies between different single-phase models suggested that the prognostic value of PVP-based models might be superior to AP-based models [31, 34, 48].

The majority of studies employed 3D volumes of interest (VOIs) for analysis, with only two articles utilizing 2D regions of interest (ROIs) based on the maximum tumor section [46, 47]. Most studies only delineated the tumor regions, while a few explored radiomics features in the peritumor areas [36, 37, 45]. Song et al. suggested that the prognostic value of radiomics models based on tumor regions plus peritumor extensions was not as good as the models considering solely the tumor regions [34]. 28 articles reported the algorithms used for feature selection and model construction, with the majority employing machine learning (ML) algorithms. Five articles adopted deep learning

Table 1 Baseline characteristics of the included studies

Year, author, nation	Imaging modality	Number of patients	Tumor staging	TACE procedure	Region of interest	Modeling algorithm	Number, phase of features	Validation datasets	Predicted endpoints
2020, Liu, China [54]	CEUS	130	BCLC-B	cTACE	3D	CNN	NA, NA	internal validation	OR
2021, Chen, China [37]	CECT	595	BCLC-B	NA	3D	LASSO	18, AP + PS	internal and external validation	OR
2021, Jin, China [38]	CECT	256	BCLC-A/B	cTACE	3D	LDA	14, PVP	external validation	EVIT
2021, Peng, China [39]	CECT	310	BCLC-B	cTACE	3D	CNN, RF	14, AP	internal validation	OR
2022, Bai, China [20]	CECT	111	BCLC-A/B	cTACE	3D	SVM	22, AP + VP	internal validation	OR
2022, Li, China [40]	CECT	48	< = 2 tumors	NA	3D	LASSO	5, early AP + mid AP + late AP + courage period	resampling validation	OR
2022, Wang, China [41]	CECT	243	BCLC-A/B	cTACE or DEB-TACE	3D	LASSO, XGBoost	6, AP + PVP	internal validation	Unsuitable for TACE
2023, Sun, China [42]	CECT	399	BCLC-B	cTACE	3D	LASSO	19, AP	internal validation	OR
2020, Sun, China [28]	CEMR	84	BCLC-B	cTACE	3D	LASSO	NA, T2WI + DWIb0_DWIb500 + ADC	internal validation	PD
2021, Kong, China [29]	CEMR	99	BCLC-B/C	cTACE or DEB-TACE	3D	LASSO	6, T2WI	internal validation	OR
2021, Zhao, China [30]	CEMR	122	BCLC-A/B/C	cTACE	3D	LASSO	6, AP + PVP + DP	internal validation	OR
2022, Cannella, Italy [31]	CEMR	51	Unifocal tumor	NA	3D	hybrid descriptive-inferential method	3, PVP	fivefold cross-validation	OR
2022, Tian, China [32]	CEMR	71	CNLC-Ib/IIa	cTACE	3D	3D-CNN, SVM-RFE	20, T2WI	fivefold cross-validation	PD

Table 1 (continued)

Year, author, nation	Imaging modality	Number of patients	Tumor staging	TACE procedure	Region of interest	Modeling algorithm	Number, phase of features	Validation datasets	Predicted endpoints
2023, Chen, China [33]	CEMR	172	BCLC-A/B/C	cTACE or DEB-TACE	3D	DNN	NA, T2WI + AP + PVP + DP	internal and external validation	OR
2016, Hinrichs, Germany [43]	CECT	19	tumor > = 2 cm	cTACE	3D	NA	1, AP + PVP	NA	OS
2018, Kim, Korea [44]	CECT	88	BCLC-A/B/C	cTACE	3D	LASSO	12, AP	NA	OS
2020, Meng, China [45]	CECT	162	BCLC-A/B	cTACE	3D	LASSO	6, AP + PVP	internal validation	OS
2020, Peng, China [46]	CECT	798	BCLC-B	NA	2D	CNN	NA, AP	internal and external validation	OR
2021, Guo, China [47]	CECT	94	BCLC-A/B/C	cTACE	2D	LASSO	4, PS	internal validation	OR, OS
2021, Ivanics, Canada [48]	CECT	88	within Milan / Toronto criteria	NA	3D	LASSO, SVM	27, PVP	fivefold cross-validation	Progression dropout in LT list or after LT
2021, Niu, China [49]	CECT	218	BCLC-A/B	cTACE	3D	LASSO	8, AP	external validation	OS
2021, Tipaldi, Italy [50]	CECT	50	BCLC-A/B/C	DEB-TACE	3D	LR	4, PVP	Bootstrap validation	OS
2022, Dai, China [51]	CECT	102	BCLC-B/C	cTACE or DEB-TACE	3D	LASSO	9, PS	internal validation	OS
2022, Liu, China [52]	CECT	70	NA	NA	3D	LASSO	3, NA	tenfold cross-validation	OS

Table 1 (continued)

Year, author, nation	Imaging modality	Number of patients	Tumor staging	TACE procedure	Region of interest	Modeling algorithm	Number, phase of features	Validation datasets	Predicted endpoints
2022, Wang, China [36]	CECT	543	NA	NA	3D	LASSO	25, PVP	internal and external validation	TTP, OS
2023, Fan, China [53]	CECT	92	BCLC-A/B/C	cTACE	3D	adaLASSO	17, AP + PVP	internal validation	Nonviable lesion
2020, Song, China [34]	CEMR	184	BCLC-0/A/B	cTACE	3D	LASSO	14, PVP	internal validation	RFS
2021, Kuang, China [21]	CEMR	153	tumor < = 5 cm, < = 2 lesions	cTACE	3D	mRMR, LASSO	11, AP	external validation	OR
2022, Liu, China [35]	CEMR	140	BCLC-A/B/C	cTACE	3D	mRMR, LASSO	6, T2WI	external validation	OR

NA, not applicable/not available; CECT, contrast-enhanced computed tomography; CEMR, contrast-enhanced magnetic resonance; Gd-EOB-DTPA MR, Gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid-enhanced magnetic resonance; CEUS, contrast-enhanced ultrasound; HCC, hepatocellular carcinoma; BCLC, Barcelona Clinic Liver Cancer; CNLC, China Liver Cancer Staging; TACE, transarterial chemoembolization; cTACE, conventional TACE; DEB-TACE, TACE with drug-eluting beads; mRECIST, modified response evaluation criteria in solid tumors; OR, objective response; OS, overall survival; TTP, time to progression; EVIT, extrahepatic spread or vascular invasion after initial TACE monotherapy; AP, arterial phase; PS, plain scan; PVP, portal vein phase; DP, delayed phase; DL, deep learning; ML, machine learning; CNN, convolutional neural network; LASSO, least absolute shrinkage and selection operator; XGBoost, extreme gradient boosting; LR, logistic regression; mRMR, maximum correlation–minimum redundancy; SVM, support vector machines; RF, random forest; DNN, Deep-Learning Neural Network; LDA, linear discriminant analysis

(DL) algorithms in the modeling process [32, 33, 39, 46, 54]. A total of 44 independent datasets (training sets or validation sets) reported both sample sizes of the cohorts and the number of features in the predictive models. The median sample-ize-to-feature-number ratio was 9.30 (p 25–p 75: 5.16–12.43), with only 19 datasets having a value over 10.

Quality assessment

A quality assessment was conducted for the included studies. When assessing the risk of bias using QUADAS-2, prevalent biases were identified, originating from unclear case selection procedures (whether all patients were included consecutively), the absence of specified assessment criteria for diagnostic models (cut-off values), and the lack of validation of established models in independent datasets (Fig. 2a, Supplementary Fig. 2A). The latter two factors were also the primary sources of concerns about the studies' applicability.

The summary of study quality according to RQS and METRICS is depicted in Fig. 2b, c, with detailed results presented in Supplementary Fig. 2B and Supplementary Table 5. The overall RQS across all 29 included studies averaged 12.90 ± 5.13 (35.82% ± 14.25%). METRICS scores showed positive correlation with RQS scores (Fig. 2c), averaged 62.98% ± 14.58%. Generally, most studies demonstrated satisfactory quality in terms of reporting imaging strategies, multiple segmentations, integrating non-radiomics clinical variables, assessing the model's discriminative powers and calibration, model validation, and disclosing radiomics features within the model. Using the METRICS assessment, there were 2, 14, 10 and 3 studies meeting the criteria of "excellent", "good", "moderate" and "low", respectively.

Study data synthesis (*Meta*-analysis)

Pooled predictive performance of radiomics for TACE response

A total of 23 independent datasets from 14 studies provided sufficient information to extract a complete 2 × 2

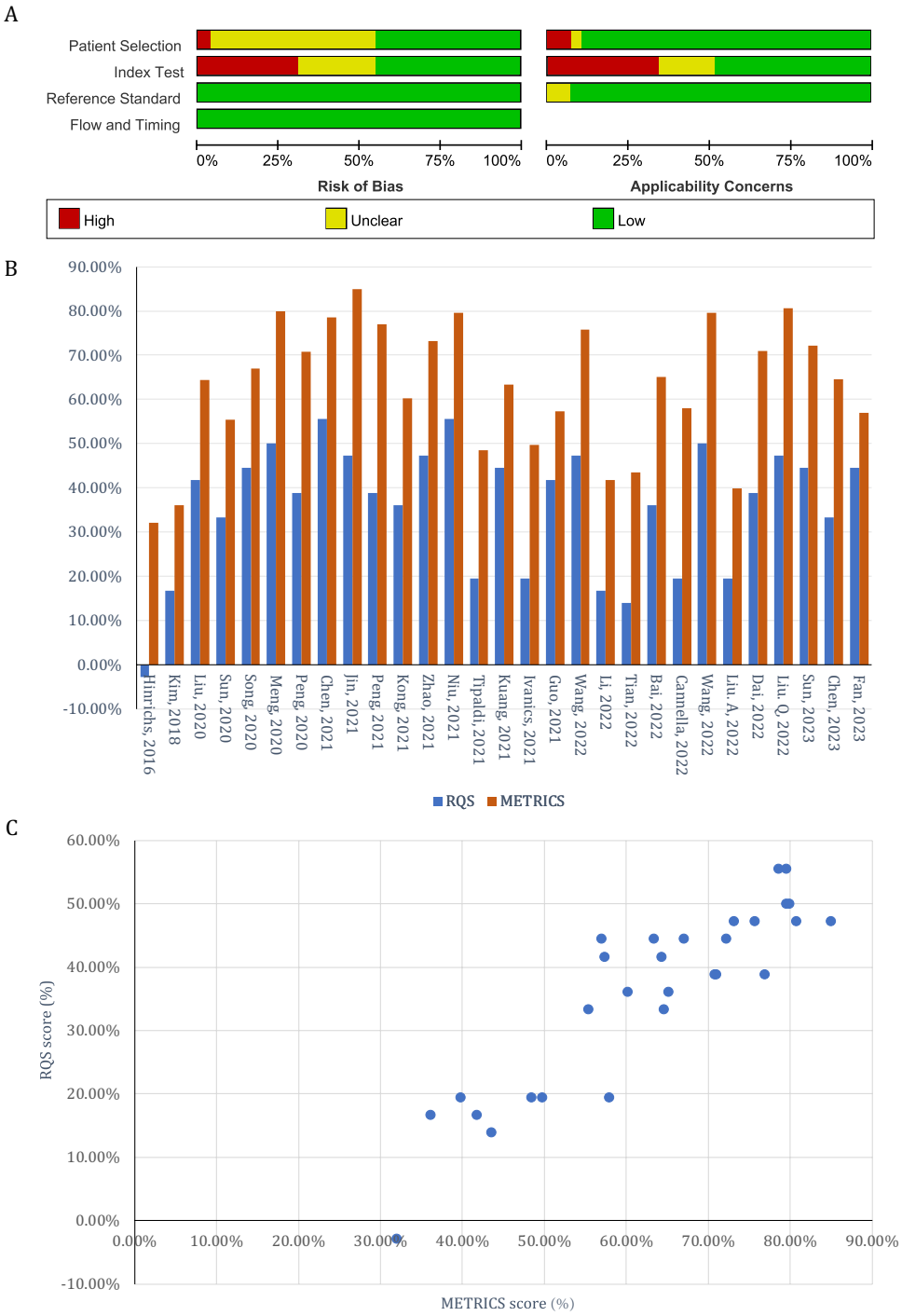

Fig. 2 Quality assessment of included articles. **A** The bias risk assessment of included studies utilizing the QUADAS-2 scale. **B** RQS and METRICS scores of included articles. **C** Relationship between METRICS and RQS for each included study, illustrated by scatter plot

contingency table and thus entered the meta-analysis. The included datasets comprised 12 training datasets (with or without resampling validation), including 1628 patients, and 11 independent validation datasets (internal or external validation), comprising 815 patients. In all of these datasets, the predictive endpoint of radiomics models was the tumor response after TACE, defined by objective response (OR) according to the RECIST [14] or mRECIST [15] criteria in the majority of the studies. All datasets included in the meta-analysis and their key methodological information are outlined in Table 2. For the total of 23 datasets, a synthetic analysis of predictive performance was conducted on 2443 subjects (Fig. 3a–e, Supplementary Table 6). The pooled sensitivity was 0.83 (95% CI: 0.78–0.87) (Fig. 3a), specificity was 0.86 (95% CI: 0.79–0.92) (Fig. 3b), and the pooled PLR and NLR were 6.13 (95% CI: 3.79–9.90) (Fig. 3c)

and 0.20 (95% CI: 0.15–0.27) (Fig. 3d), respectively. The AUC of the sROC was 0.90 (95% CI: 0.87–0.93) (Fig. 3e). Heterogeneity tests revealed I^2 exceeding 70% for sensitivity, specificity, PLR, and NLR, with Q-test P-values below 0.01, indicating significant heterogeneity (Supplementary Table 6). Further Galbraith plot suggested a limited and symmetrical impact of outliers on the result of the meta-analysis (Fig. 4a).

Origin of heterogeneity and subgroup analysis

Meta-regression analysis was performed based on the key methodological parameters, which incorporated six variables: imaging modalities (CT/MR/US), training or validation datasets, modeling algorithms (ML/DL), imaging phases (single-phase/multi-phase), single imaging phase (AP/PVP), and the inclusion of peritumoral features (Yes/No). In relation to pooled sensitivity, noteworthy inter-subgroup variances were identified between training sets and validation sets, MR studies and other studies, ML models and DL models, single-phase models and multi-phase models, as well as AP models and PVP models (Fig. 4c, Supplementary Table 7). Concerning specificity, significant disparities were discerned between MR studies and other studies, and between ML models and DL models (Fig. 4d, Supplementary Table 7).

Subsequent subgroup analyses were conducted for variables deemed significant in the meta-regression. Due to the limited studies reporting single-phase AP models and PVP models, subgroup analyses were executed based on the remaining factors, and pooled sensitivity, specificity, and heterogeneity within each subgroup was calculated (Table 3). Inter-study heterogeneity was significantly lower among the independent validation datasets compared to the training datasets, while comparable model performance was reported (I^2: 51.32% vs. 82.11%, Supplementary Fig. 3). In the 11 validation datasets, further subgrouping was based on imaging modalities, imaging phases and modeling algorithms. Inter-study homogeneity with $I^2 < 50\%$ was observed in the MR subgroup, ML subgroup, and the multi-phase model subgroup (Supplementary Fig. 4–6). Furthermore, the five studies combining CT and ML also showed satisfactory homogeneity (I^2: 45.07, Supplementary Fig. 7).

Table 2 Major results and methodological features of the studies in meta-analysis

Author year	TP	FP	FN	TN	Dataset	Modality	Algorithm	Phase	AP/ PVP	Peritumor
Bai 2022 (v)	8	0	3	21	validation	CT	ML	Multi		No
Cannella 2022 (t)	21	9	8	13	training	MR	ML	Single	PVP	No
Chen 2021 (t)	164	31	20	140	training	CT	ML	Multi		Yes
Chen 2021 (v)	52	13	9	44	validation	CT	ML	Multi		Yes
Chen 2023 (t)	51	10	5	49	training	MR	DL	Multi		No
Chen 2023 (v)	32	8	5	32	validation	MR	DL	Multi		No
Jin 2021 (t)	27	24	7	78	training	CT	ML	Single	PVP	No
Jin 2021 (v)	19	27	7	67	validation	CT	ML	Single	PVP	No
Kong 2021 (t)	25	5	12	27	training	MR	ML	Single		No
Kong 2021 (v)	13	3	3	11	validation	MR	ML	Single		No
Li 2022 (t)	20	2	4	22	training	CT	ML	Multi		No
Liu 2020 (t)	30	1	1	57	training	US	DL	Multi		No
Liu 2020 (v)	13	2	2	24	validation	US	DL	Multi		No
Peng 2021 (t)	80	0	3	56	training	CT	DL	Single	AP	No
Peng 2021 (v)	97	0	7	67	validation	CT	DL	Single	AP	No
Sun 2020 (v)	5	3	2	7	validation	MR	ML	Multi		No
Sun 2023 (t)	96	44	26	133	training	CT	ML	Single	AP	No
Sun 2023 (v)	36	20	7	37	validation	CT	ML	Single	AP	No
Tian 2022 (t)	12	1	8	50	training	MR	DL	Single		No
Wang 2022 (t)	38	14	20	99	training	CT	ML	Multi		No
Wang 2022 (v)	14	8	9	41	validation	CT	ML	Multi		No
Zhao 2021 (t)	32	9	12	32	training	MR	ML	Multi		No
Zhao 2021 (v)	17	9	2	9	validation	MR	ML	Multi		No

(t), training dataset; (v), validation dataset; TP, true positive; FP, false positive; FN, false negative; TN, true negative; CT, contrast-enhanced computed tomography; MR, magnetic resonance; AP, arterial phase; PVP, portal vein phase; DL, deep learning; ML, machine learning

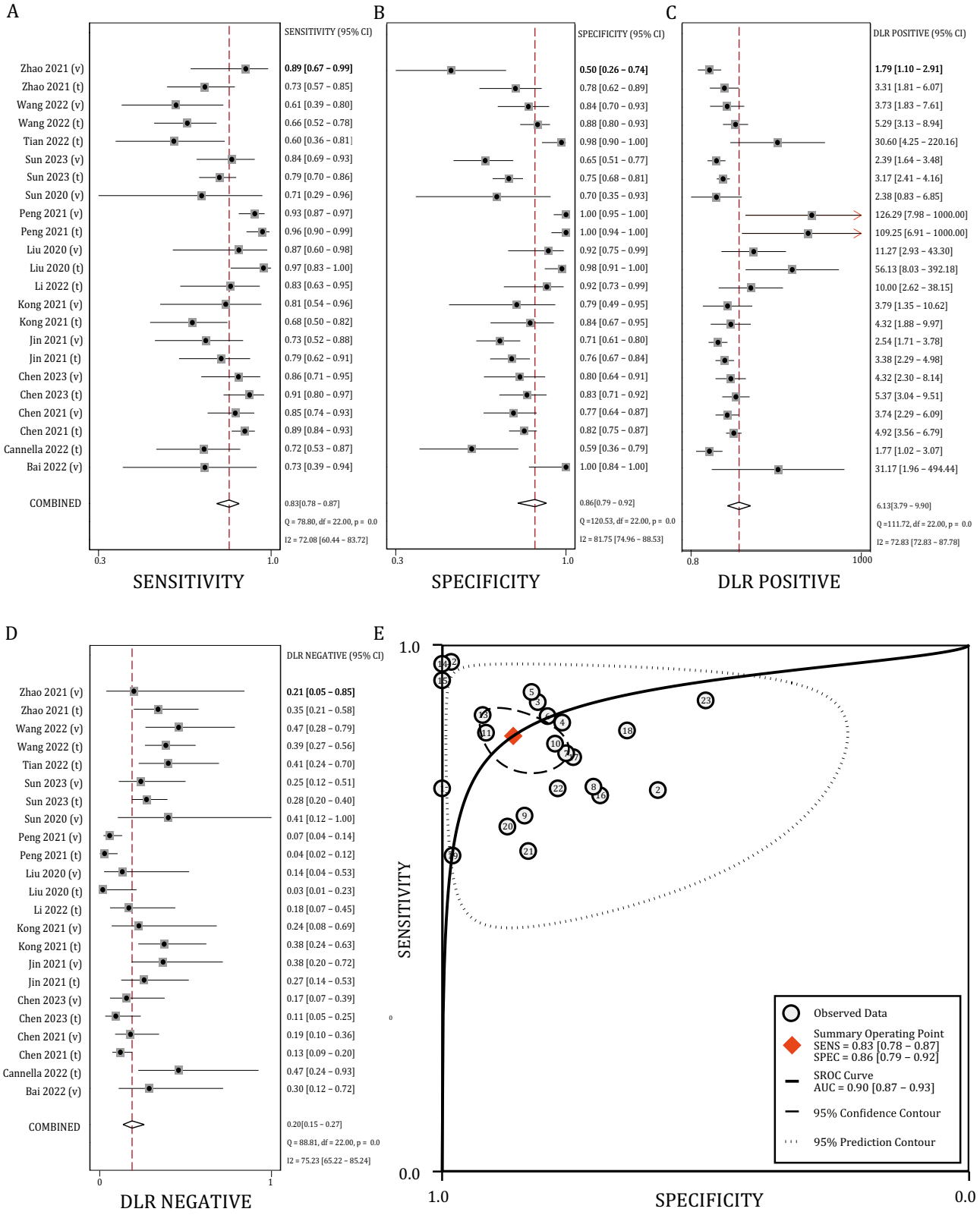

Fig. 3 Summarized and pooled performance of radiomics models for predicting TACE response. **A** Pooled sensitivity. **B** Pooled specificity. **C** Pooled positive likelihood ratio. **D** Pooled negative likelihood ratio. **E** The summary receiver operating characteristic curve

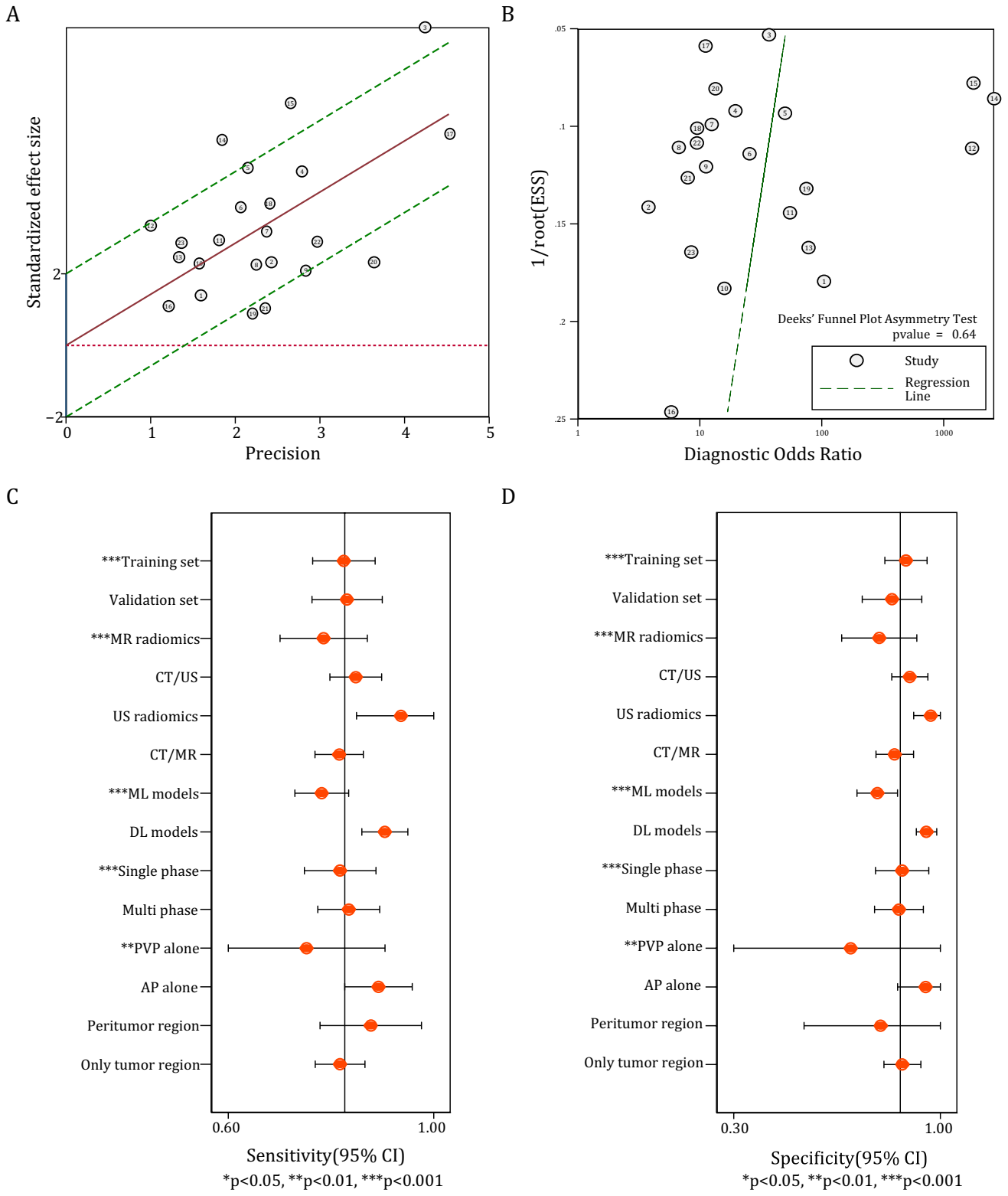

Fig. 4 Heterogeneity among included studies. **A** Impact of outliers on the meta-analysis, illustrated by Galbraith plot. **B** No significant publication bias was indicated by Deek's funnel plot and asymmetry test. **C–D** Impact of methodological factors on (**C**) pooled sensitivity and (**D**) specificity of radiomics models, according to meta-regression

Table 3 The pooled results from subgroup analysis

Subgroups			No. of datasets	Sensitivity (95%CI)	I^2% (P value)	Specificity (95%CI)	I^2% (P value)
Datasets	Training		12	0.83 (0.74–0.89)	82.11 (0.00)	0.88 (0.79–0.94)	83.43 (0.00)
	Validation		11	0.83 (0.77–0.88)	51.32 (0.02)	0.84 (0.71–0.92)	80.54 (0.00)
Validation Datasets	Modality	CT	6	0.83 (0.72–0.89)	73.45 (0.00)	0.90 (0.64–0.98)	88.65 (0.00)
		MR	4	0.85 (0.75–0.91)	0.00 (0.67) *	0.71 (0.57–0.82)	49.19 (0.12) *
	Algorithm	DL	3	NA (limited data)			
		ML	8	0.79 (0.71–0.85)	22.88 (0.25) *	0.75 (0.66–0.82)	57.08 (0.02)
	Phase	Single	4	0.85 (0.75–0.92)	65.38 (0.00)	0.87 (0.52–0.98)	90.91 (0.00)
		Multi	7	0.81 (0.72–0.88)	38.59 (0.13) *	0.81 (0.71–0.88)	65.32 (0.01)
	MR + ML		3	NA (limited data)			
	CT + ML		5	0.77 (0.67–0.85)	45.07 (0.12) *	0.77 (0.68–0.85)	63.27 (0.03)

*refer to synthesized results with $I^2 < 50\%$ and $P > 0.1$

Evaluation of publication *bias* and clinical interpretation

The funnel plot disclosed no significant evidence of publication bias among the studies included in the analysis ($P = 0.64$, see Fig. 4b). Of the 20 datasets extracted from 11 studies that reported ORR following TACE, the weighted average ORR was determined to be 51.24% (Supplementary Table 8). Using this average ORR as the pre-test probability, the posterior probabilities of achieving an objective response were calculated to be 87% and 17% when the radiomics models yielded positive and negative predictions, respectively (Fig. 5a). The likelihood ratio matrix illustrated that the majority of studies failed to attain optimal PLR and NLR, with only three records falling within the upper left quadrant, indicative of high predictive informativeness (Fig. 5b).

Discussion

This systematic review encompassed a total of 29 primary studies, involving 5483 patients, with 14 articles integrated into the meta-analysis, constituting 2691 patients. The studies manifested noteworthy regional disparities, as 25 studies

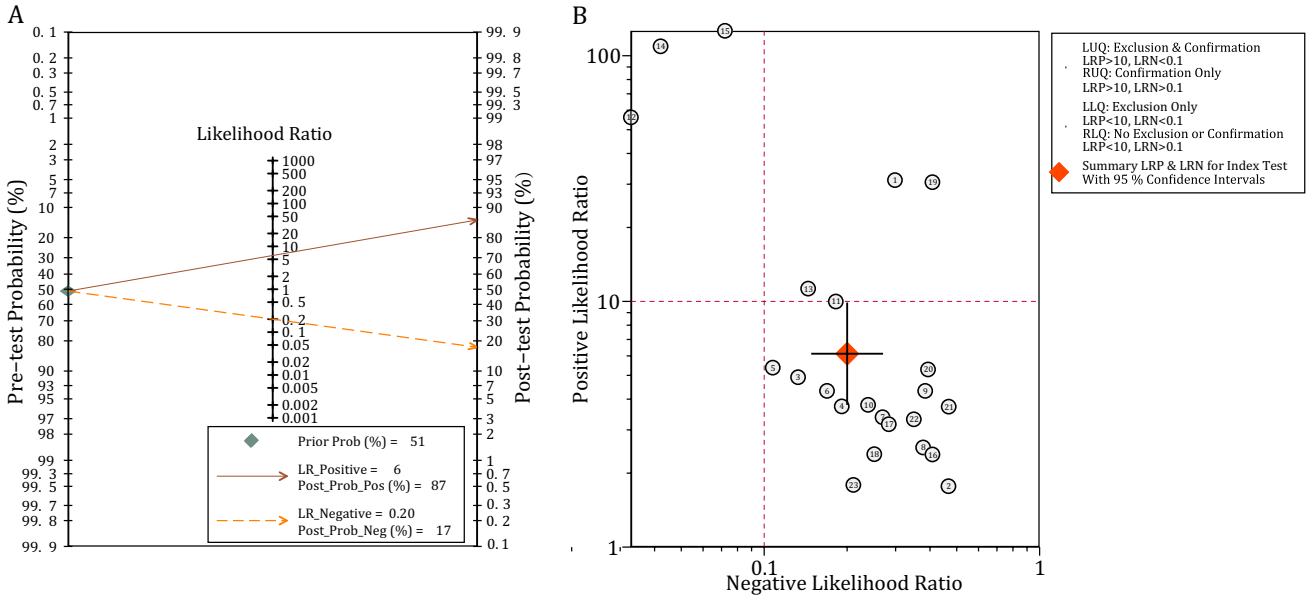

Fig. 5 Clinical utility of the radiomics models. **A** Fagan plot illustrating the posterior probabilities of the radiomics models. **B** Likelihood ratio matrix illustrating the informative value of each individual radiomics model

(86.2%) emanated from East Asia, of which 24 hailed from China (82.8%), including a total of 5187 Chinese patients (2640 Chinese patients in the meta-analysis). This observation aligns with the high incidence of HCC in East Asia [1] and the prevalent utilization of TACE within this region [6]. Additionally, this regional concentration determined a high proportion of viral-hepatitis-related HCC cases in this review (Supplementary Table 4). We further performed an exploratory bibliometric analysis by searching the Web of Science core database. The findings revealed that 42.3% of publications within the purview of "radiomics" originated from China. Similarly, Chinese researchers contributed 30.2% and 31.4% of the studies on "HCC" and "TACE", respectively, establishing China as the foremost source of studies in these domains. From the perspective of practical implication, radiomics tools may confer cost advantages and greater convenience compared to molecular-pathological tests, such as next-generation sequencing, especially in middle-income countries like China. Nevertheless, this does not diminish the applicational potential these studies in the context of global HCC management and precision medicine.

Among the included articles, mid-term HCC patients, predominantly BCLC-B stage patients, constituted the predominant cohorts, aligning with the prevailing indications for TACE [5, 13]. The majority of the studies delved into the predictive value of cross-sectional imaging, encompassing CT and MR, both acknowledged as diagnostic modalities for HCC [65]. This alignment with the traditional clinical and radiological practice might set the basis of convenient application of these radiomic models. Since 2018, a discernible upswing in the number of studies in this field has been noted, and the studies showed considerable consistency in research patterns, indicating a trend of maturation and standardization of radiomics research [59, 60].

To our knowledge, this represents the inaugural systematic review and meta-analysis delving into the predictive efficacy of radiomics in determining the prognosis of ACE. Previous investigations endeavored to harness pre-treatment conventional imaging features for predicting TACE outcomes. Parameters such as tumor size, location, enhancement patterns, enhancement heterogeneity, and intraprocedural digital subtraction angiography (DSA) characteristics were explored [66, 67]. Additionally, established liver imaging assessment tools, like the liver imaging reporting and data system (LI-RADS), were also employed for this purpose [68]. Though these traditional imaging features had demonstrated some correlations with TACE response, they were not specifically tailored tools for this purpose, resulting in imprecise prediction and limited clinical value. Consequently, radiomics methodologies emerge as a promising avenue in this context. Radiomics has demonstrated substantial potential across various oncological domains since its inception. Previous systematic reviews and meta-analyses have consolidated evidence of its efficacy in diverse areas, such as predicting microvascular invasion (MVI) in HCC [69, 70], post-radiotherapy survival in non-small cell lung cancer (NSCLC) [71], and differential diagnosis and prognosis prediction for renal cell carcinoma (RCC) [72]. Diverging from preceding meta-analyses, our study placed particular emphasis on the generalizability of radiomics models. We independently analyzed outcomes from both the datasets utilized for model training and the separate datasets designated for validation. The combined results from validation sets did not disclose insufficient model performance compared to training sets. This finding lays the groundwork for future endeavors in validating and applying these radiomics models. Of note, utilization of public datasets for model testing is a practical way to enhance the generalizability of radiomic models. In terms of the specific tasks explored in this meta-analysis, the HCC-TACE-SEG dataset from The Cancer Imaging Archive (TCIA) might provide a proper external validation cohort [73], which is worth exploiting in future studies.

In this systematic review, we utilized QUADAS-2 [58], RQS [59] and METRICS [60] to appraise the quality of radiomics studies. In general, the included studies showed moderate to good quality. In 2020, Ursprung et al. summarized the role of radiomics in RCC, reporting an average RQS of 3.41 ± 4.43 for 57 articles [72]. In contrast, this systematic review, comprising 29 studies, reported a significantly higher average RQS of 12.90 ± 5.13, reflecting the advancements in this field in recent years. Further analysis of the 16 scoring categories in RQS revealed the most pronounced disparity in the "validation" category, while the lowest scores were noted in three domains: "detect and discuss biological correlates," "prospective study registered in a trial database," and "cost-effectiveness analysis." This discrepancy may be partly attributed to our specific inclusion criteria, where prospective studies could be constrained by the intricacies of treatment indications [12], and the joint analysis with other biomarkers might be impeded by the absence of reliable predictive biomarkers for TACE efficacy. However, it is noteworthy that integrating cost-effectiveness analysis could play a pivotal role in the subsequent validation and application of radiomics models, particularly in evaluating methodological factors such as the workforce and time required for image segmentation, additional resources for multi-phase image segmentation, and net benefits compared to traditional imaging examinations. In addition, we evaluated METRICS scores of the studies, which is a novel assessment tool that is more detailed and balanced compared to RQS. For most included studies, further improvement could be made in handling clinical confounding factors, appropriate use of machine learning algorithms, appropriate choice of imaging phases, independent validation of the results, as well as the transparency and reproducibility of scientific data. Hence,

we advocate for future radiomics research to incorporate these considerations in their analysis and reports.

The meta-analysis involving all the included radiomics models indicated pooled sensitivity and specificity surpassing 0.80. As these models were developed in highly bespoke manners within retrospective cohorts, lacking consistent cut-off values across different models, we employed a summary receiver operating characteristic to assess the general prognostic performance, which unveiled an AUC of 0.90, signifying robust discriminative capacity for post-TACE prognosis. As indicative parameters for clinical application, the pooled PLR and NLR were 6.13 and 0.20, respectively, with only three studies meeting the criteria of PLR > 10 and NLR < 0.1. Overall, the current radiomics models demonstrated moderate clinical informative value. Furthermore, with prior probability taken into consideration, the calculated post-test probabilities of objective response were 87% and 17% when positive and negative results were given by the radiomics models, respectively. Accordingly, radiomics tools could forecast totally different outcomes after TACE, and thus harbor significant potential in informing therapeutic decisions of TACE.

Significant heterogeneity surfaced among the incorporated studies, a revelation consistent with prior meta-analyses in the radiomics field [69–72], aligned with our expectations. Given the retrospective design of all studies, heterogeneity was inevitable [70]. Although similar and standard research workflows were adopted by most of the included studies, numerous methodological variables in the whole analyzing process, from imaging data to the final predictive model, could contribute to substantial disparities among studies, which is an inherent limitation of highly customized models. Nevertheless, the combined statistical results of the current models are still of considerable value. In this study, we utilized the Galbraith plot to evaluate the impact of heterogeneity on the meta-analysis results, uncovering few extreme outliers and a generally symmetric impact. Furthermore, Deek's funnel plot and asymmetry tests were utilized, suggesting no apparent publication bias.

Through meta-regression, we scrutinized the origins of heterogeneity, we observed that several crucial methodological factors including imaging modalities, modeling algorithms, and validation or training datasets could potentially be major contributors to inter-study heterogeneity. In subgroup analyses, results from testing sets among studies that adopted the same imaging method (MR), imaging phases (multi-phase imaging data), or the same type of modeling (ML-based modeling) showed better homogeneity. And the synthesized results in subgroups did not show serious deviation from the combined results of all studies. Taken together, we believe that radiomics tools have great potential for predicting treatment outcome of TACE, from both the research and clinical application perspectives.

We analyzed the performance of different models only in the validation datasets (Table 3), CT-based radiomics models demonstrated comparable pooled sensitivity and superior specificity compared to MR models. Considering the lower time cost and expense of CT compared to MR, CT-based radiomics models may hold an advantage in practical clinical application. Interestingly, models using single-phase/sequence imaging data did not show insufficient performance compared to multi-phase/sequence models in combined sensitivity and specificity. A few studies directly compared the predictive value between uniparametric and multi-parametric models [20, 30, 35, 42, 48], and there was no consistent evidence supporting joint analysis of multiple imaging phases or sequences. As pointed out by the METRICS guideline [60], multi-parametric analysis may unnecessarily increase the data dimensionality and risk of overfitting. Therefore, future studies should try to use simple radiographic data or prove the added value of multi-parametric models. Notably, while many studies considered multiple phases, very few conducted voxel-to-voxel image matching across different phases [43, 49], with no study analyzing the evolution of feature values across different time points. Future research employing image alignment and feature matching to analyze the time evolution of features might expand the role of delta-radiomics [59] and fully leverage the potential of multi-phase imaging.

The choice of modeling algorithms influences the model's performance in a decisive manner [74, 75]. In our subgroup analysis, ML-based models demonstrated inferior sensitivity and specificity compared to DL models. In addition, two studies directly compared DL and ML models and indicated the superior predictive power of DL models [33, 54]. However, it is noteworthy that the transparency of details in DL models was limited in the included studies, potentially restricting the reproducibility of their results. Generally, deep learning demonstrated significant potential for radiomics modeling, albeit its current limited use in studies [32, 33, 39, 46, 54]. There is presently no consensus on the necessity of including peritumoral features. Our meta-regression and the direct comparison by Song et al. [34] did not reveal significant benefit of considering peritumoral regions. Nevertheless, in other tasks, such as predicting MVI [76] and postoperative recurrence [77] in HCC, analyzing both tumor and peritumoral regions proved to be more effective. Hence, whether incorporating peritumoral features improves model performance might be task-specific.

Finally, while our meta-analysis suggested potential advantages of some methods, there have been no universally recognized optimal choices for these methodological options, and further validations are imperative. The ongoing exploration and standardization of radiomics research methods will undoubtedly lead to more ideal predictive models,

guiding decision-making and management for HCC patients undergoing TACE treatment.

Limitations and future perspectives

This study presents several limitations. Firstly, all included studies are retrospective in design. Despite conducting meta-regression and subgroup analyses, the sources of heterogeneity could not be fully elucidated. Therefore, caution is advised when directly applying the results of these studies due to potential concerns about generalizability. This limitation underscores the necessity for further methodological exploration and additional radiomics research to address these gaps. Secondly, not all included datasets met the ideal sample size criterion of ten times the number of model features, emphasizing the need for future studies to establish larger clinical cohorts or reduce the number of model features.

Conclusion

Radiomics models demonstrated good accuracy and promising clinical utility in predicting the prognosis of initial TACEtreatment. Methodological factors such as imaging modalities, phases of imaging, and modeling algorithms are significant sources of literature heterogeneity.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11547-024-01840-9.

Author contributions KGD, TC and YCZ were involved in the conceptualization. FY, KGD and YCZ contributed to the literature screening and selection. ZJL, KGD and YCZ extracted the data and assessed the quality. KGD and TC contributed to the statistical analysis. KGD assisted in the original draft of the manuscript. TC, WXP and TL checked and edited the manuscript. YCZ contributed to the supervision.

Funding This study was supported by grants from the Beijing Municipal Natural Science Foundation Project (7222130) and National High-Level Hospital Clinical Research Funding (2022-PUMCH-A-236).

Data availability statement All articles included in this review are available from PubMed, Web of Science, and Embase. All data generated during the analysis in this study are included in the article and supplementary materials.

Declarations

Conflicts of interest The authors have no conflict of interests related to this publication.

Statement of Ethics Institutional Review Board approval was not required for this systematic review and meta-analysis.

Research involving human participants and/or animals Not applicable for this systematic review and meta-analysis.

Informed consent Not applicable for this systematic review and meta-analysis.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/.

References

1. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 71(3):209–249. https://doi.org/10.3322/caac.21660
2. Allemani C, Matsuda T, Di Carlo V et al (2018) Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 391(10125):1023–1075. https://doi.org/10.1016/s0140-6736(17)33326-3
3. Raffetti E, Portolani N, Molfino S et al (2021) Is survival for hepatocellular carcinoma increasing? A population-based study on survival of hepatocellular carcinoma patients in the 1990s and 2000s. Clin Res Hepatol Gastroenterol 45(1):101433. https://doi.org/10.1016/j.clinre.2020.04.004
4. Benson AB, D'Angelica MI, Abbott DE et al (2021) Hepatobiliary cancers, version 2.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 19(5):541–565. https://doi.org/10.6004/jnccn.2021.0022
5. Vogel A, Cervantes A, Chau I et al (2018) Hepatocellular carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 29(Suppl 4):238–255. https://doi.org/10.1093/annonc/mdy308
6. Zhou J, Sun H, Wang Z et al (2023) Guidelines for the diagnosis and treatment of primary liver cancer (2022 Edition). Liver Cancer 12(5):405–444. https://doi.org/10.1159/000530495
7. Llovet JM, De Baere T, Kulik L et al (2021) Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol 18(5):293–313. https://doi.org/10.1038/s41575-020-00395-0
8. Nault JC, Sutter O, Nahon P, Ganne-Carrié N, Séror O (2018) Percutaneous treatment of hepatocellular carcinoma: state of the art and innovations. J Hepatol 68(4):783–797. https://doi.org/10.1016/j.jhep.2017.10.004
9. Sun H, Xu W, Yang H (2022) Anatomical thermal ablation: a valuable treatment method for hepatocellular carcinoma smaller than 5 cm. Hepatol Res 52(10):893. https://doi.org/10.1111/hepr.13805
10. Sadigh G, Abi-Jaoudeh N (2022) Surgery versus ablation for small hepatocellular carcinoma treatment and role of personalized medicine. J Am Coll Radiol 19(11):1224–1225. https://doi.org/10.1016/j.jacr.2022.09.023
11. Sonbol MB, Riaz IB, Naqvi SAA et al (2020) Systemic therapy and sequencing options in advanced hepatocellular carcinoma:

a systematic review and network meta-analysis. JAMA Oncol 6(12):e204930. https://doi.org/10.1001/jamaoncol.2020.4930
12. Cho Y, Choi JW, Kwon H et al (2023) Transarterial chemoembolization for hepatocellular carcinoma: 2023 expert consensus-based practical recommendations of the Korean Liver Cancer Association. J Liver Cancer 23(2):241–261. https://doi.org/10.17998/jlc.2023.05.22
13. Sieghart W, Hucke F, Peck-Radosavljevic M (2015) Transarterial chemoembolization: modalities, indication, and patient selection. J Hepatol 62(5):1187–1195. https://doi.org/10.1016/j.jhep.2015.02.010
14. Eisenhauer EA, Therasse P, Bogaerts J et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45(2):228–247. https://doi.org/10.1016/j.ejca.2008.10.026
15. Lencioni R, Llovet JM (2010) Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 30(1):52–60. https://doi.org/10.1055/s-0030-1247132
16. Gillmore R, Stuart S, Kirkwood A et al (2011) EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. J Hepatol 55(6):1309–1316. https://doi.org/10.1016/j.jhep.2011.03.007
17. Wei J, Hu Y, Yu J, Yin C, Chen G, Jin L (2022) Predictive value of hepatic venous pressure gradient and efficacy and significance of early PTVE for gastrointestinal bleeding after TACE for liver cancer. J Cancer Res Ther 18(7):1967–1972. https://doi.org/10.4103/jcrt.jcrt_331_22
18. Saito N, Nishiofuku H, Sato T et al (2023) Predictive factors of complete response to transarterial chemoembolization in intermediate stage hepatocellular carcinoma beyond Up-To-7 criteria. Cancers 15(9):2609
19. Lu H, Zheng C, Xiong B, Xia X (2023) TACE versus TACE + entecavir versus TACE + tenofovir in the treatment of HBV associated hepatocellular carcinoma. BMC Cancer 23(1):235. https://doi.org/10.1186/s12885-023-10694-9
20. Bai H, Meng S, Xiong C et al (2022) Preoperative CECT-based radiomic signature for predicting the response of transarterial chemoembolization (TACE) therapy in hepatocellular carcinoma. Cardiovasc Interv Radiol 45(10):1524–1533. https://doi.org/10.1007/s00270-022-03221-z
21. Kuang Y, Li R, Jia P et al (2021) MRI-Based Radiomics: Nomograms predicting the short-term response after transcatheter arterial chemoembolization (TACE) in hepatocellular carcinoma patients with diameter less than 5 cm. Abdom Radiol (NY) 46(8):3772–3789. https://doi.org/10.1007/s00261-021-02992-2
22. Lee S-W, Yen C-L, Cheng Y-C, Shun Yang S, Lee T-Y (2022) The radiological prognostic factors of transcatheter arterial chemoembolization to hepatocellular carcinoma. Medicine 101(41):e30875. https://doi.org/10.1097/md.0000000000030875
23. Byun J, Kim SY, Kim JH et al (2021) Prediction of transarterial chemoembolization refractoriness in patients with hepatocellular carcinoma using imaging features of gadoxetic acid-enhanced magnetic resonance imaging. Acta Radiol 62(12):1548–1558. https://doi.org/10.1177/0284185120971844
24. He K, Yang Z, Liu X et al (2023) Identification of potential predictors of prognosis and sorafenib-associated survival benefits in patients with hepatocellular carcinoma after transcatheter arterial chemoembolization. Curr Oncol 30(1):476–491
25. Minici R, Siciliano MA, Ammendola M et al (2023) Prognostic role of neutrophil-to-lymphocyte ratio (NLR), lymphocyte-to-monocyte ratio (LMR), platelet-to-lymphocyte ratio (PLR) and lymphocyte-to-C reactive protein ratio (LCR) in patients with hepatocellular carcinoma (HCC) undergoing chemoembolizations (TACE) of the liver: the unexplored corner linking tumor microenvironment, biomarkers and interventional radiology. Cancers 15(1):257
26. Zeng J, Han L, Wang T et al (2023) The allelic expression of RNA editing gene ADARB1 in hepatocellular carcinoma treated with transarterial chemoembolization. Pharmgenomics Pers Med 16:229–238. https://doi.org/10.2147/pgpm.S402115
27. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48(4):441–446. https://doi.org/10.1016/j.ejca.2011.11.036
28. Sun Y, Bai H, Xia W et al (2020) Predicting the outcome of transcatheter arterial embolization therapy for unresectable hepatocellular carcinoma based on radiomics of preoperative multiparameter MRI. J Magn Reson Imaging 52(4):1083–1090. https://doi.org/10.1002/jmri.27143
29. Kong C, Zhao Z, Chen W et al (2021) Prediction of tumor response via a pretreatment MRI radiomics-based nomogram in HCC treated with TACE. Eur Radiol 31(10):7500–7511. https://doi.org/10.1007/s00330-021-07910-0
30. Zhao Y, Wang N, Wu J et al (2021) Radiomics analysis based on contrast-enhanced MRI for prediction of therapeutic response to transarterial chemoembolization in hepatocellular carcinoma. Front Oncol 11:582788. https://doi.org/10.3389/fonc.2021.582788
31. Cannella R, Camma C, Matteini F et al (2022) Radiomics analysis on gadoxetate disodium-enhanced MRI predicts response to transarterial embolization in patients with HCC. Diagnostics 12(6):1308. https://doi.org/10.3390/diagnostics12061308
32. Tian Y, Komolafe TE, Chen T, Zhou B, Yang X (2022) Prediction of TACE treatment response in a preoperative MRI via analysis of integrating deep learning and radiomics features. Journal of Medical and Biological Engineering 42(2):169–178. https://doi.org/10.1007/s40846-022-00692-w
33. Chen M, Kong C, Qiao E, et al. Multi-algorithms analysis for pre-treatment prediction of response to transarterial chemoembolization in hepatocellular carcinoma on multiphase MRI. *Insights Imaging*. 2023;14(1):38. https://doi.org/10.1186/s13244-023-01380-2
34. Song W et al (2020) MRI-based radiomics: associations with the recurrence-free survival of patients with hepatocellular carcinoma treated with conventional transcatheter arterial chemoembolization. J Magn Reson Imaging 52:461–473. https://doi.org/10.1002/jmri.26977
35. Liu QP, Yang KL, Xu X, Liu XS, Qu JR, Zhang YD (2022) Radiomics analysis of pretreatment MRI in predicting tumor response and outcome in hepatocellular carcinoma with transarterial chemoembolization: a two-center collaborative study. Abdom Radiol (NY) 47(2):651–663. https://doi.org/10.1007/s00261-021-03375-3
36. Wang H, Liu Y, Xu N et al (2022) Development and validation of a deep learning model for survival prognosis of transcatheter arterial chemoembolization in patients with intermediate-stage hepatocellular carcinoma. Eur J Radiol 156:110527. https://doi.org/10.1016/j.ejrad.2022.110527
37. Chen M, Cao J, Hu J et al (2021) Clinical-radiomic analysis for pretreatment prediction of objective response to first transarterial chemoembolization in hepatocellular carcinoma. Liver Cancer 10(1):38–51. https://doi.org/10.1159/000512028
38. Jin Z, Chen L, Zhong B et al (2021) Machine-learning analysis of contrast-enhanced computed tomography radiomics predicts patients with hepatocellular carcinoma who are unsuitable for initial transarterial chemoembolization monotherapy: a multicenter study. Transl Oncol 14(4):101034. https://doi.org/10.1016/j.tranon.2021.101034
39. Peng J, Huang J, Huang G, Zhang J (2021) Predicting the initial treatment response to transarterial chemoembolization in

intermediate-stage hepatocellular carcinoma by the integration of radiomics and deep learning. Front Oncol 11:730282. https://doi.org/10.3389/fonc.2021.730282
40. Li Q, Luo G, Li J (2022) Evaluation of therapeutic effects of computed tomography imaging classification algorithm-based transcatheter arterial chemoembolization on primary hepatocellular carcinoma. Comput Intell Neurosci. https://doi.org/10.1155/2022/5639820
41. Wang DD, Zhang JF, Zhang LH et al (2022) Clinical-radiomics predictors to identify the suitability of transarterial chemoembolization treatment in intermediate-stage hepatocellular carcinoma: a multicenter study. Hepatobiliary Pancreat Dis Int. https://doi.org/10.1016/j.hbpd.2022.11.005
42. Sun Z, Shi Z, Xin Y et al (2023) Contrast-enhanced CT imaging features combined with clinical factors to predict the efficacy and prognosis for transarterial chemoembolization of hepatocellular carcinoma. Acad Radiol. https://doi.org/10.1016/j.acra.2022.12.031
43. Hinrichs JB, Shin HO, Kaercher D et al (2016) Parametric response mapping of contrast-enhanced biphasic CT for evaluating tumour viability of hepatocellular carcinoma after TACE. Eur Radiol 26(10):3447–3455. https://doi.org/10.1007/s00330-015-4203-4
44. Kim J, Choi SJ, Lee SH, Lee HY, Park H (2018) Predicting survival using pretreatment CT for patients with hepatocellular carcinoma treated with transarterial chemoembolization: comparison of models using radiomics. AJR Am J Roentgenol 211(5):1026–1034. https://doi.org/10.2214/ajr.18.19507
45. Meng XP, Wang YC, Ju S et al (2020) Radiomics analysis on multiphase contrast-enhanced CT: a survival prediction tool in patients with hepatocellular carcinoma undergoing transarterial chemoembolization. Front Oncol 10:1196. https://doi.org/10.3389/fonc.2020.01196
46. Peng J, Kang S, Ning Z et al (2020) Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. Eur Radiol 30(1):413–424. https://doi.org/10.1007/s00330-019-06318-1
47. Guo Z, Zhong N, Xu X et al (2021) Prediction of hepatocellular carcinoma response to transcatheter arterial chemoembolization: a real-world study based on non-contrast computed tomography radiomics and general image features. J Hepatocell Carcinoma 8:773–782. https://doi.org/10.2147/jhc.S316117
48. Ivanics T, Salinas-Miranda E, Abreu P et al (2021) A pre-TACE radiomics model to predict HCC progression and recurrence in liver transplantation: a pilot study on a novel biomarker. Transplantation 105(11):2435–2444. https://doi.org/10.1097/tp.0000000000003605
49. Niu XK, He XF (2021) Development of a computed tomography-based radiomics nomogram for prediction of transarterial chemoembolization refractoriness in hepatocellular carcinoma. World J Gastroenterol 27(2):189–207. https://doi.org/10.3748/wjg.v27.i2.189
50. Tipaldi MA, Ronconi E, Lucertini E et al (2021) Hepatocellular carcinoma drug-eluting bead transarterial chemoembolization (DEB-TACE): outcome analysis using a model based on pretreatment CT texture features. Diagnostics (Basel). https://doi.org/10.3390/diagnostics11060956
51. Dai Y, Jiang H, Feng ST et al (2022) Noninvasive imaging evaluation based on computed tomography of the efficacy of initial transarterial chemoembolization to predict outcome in patients with hepatocellular carcinoma. J Hepatocell Carcinoma 9:273–288. https://doi.org/10.2147/jhc.S351077
52. Liu A, Liu B, Duan X et al (2022) Development of a novel combined nomogram model integrating Rad-score, age and ECOG to predict the survival of patients with hepatocellular carcinoma treated by transcatheter arterial chemoembolization. J Gastrointest Oncol 13(4):1889–1897. https://doi.org/10.21037/jgo-22-548
53. Fan XL, Wang YH, Chen YH et al (2023) Computed tomography texture analysis combined with preoperative clinical factors serve as a predictor of early efficacy of transcatheter arterial chemoembolization in hepatocellular carcinoma early access. Abdom Radiol. 48(6):2008–2018. https://doi.org/10.1007/s00261-023-03868-3
54. Liu D, Liu F, Xie X et al (2020) Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. Eur Radiol 30(4):2365–2376. https://doi.org/10.1007/s00330-019-06553-6
55. Salameh JP, Bossuyt PM, McGrath TA et al (2020) Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ 370:m2632. https://doi.org/10.1136/bmj.m2632
56. Shea BJ, Reeves BC, Wells G et al (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 358:j4008. https://doi.org/10.1136/bmj.j4008
57. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 13:1. https://doi.org/10.1186/s12916-014-0241-z
58. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155(8):529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
59. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14(12):749–762. https://doi.org/10.1038/nrclinonc.2017.141
60. Kocak B, Akinci D'Antonoli T, Mercaldo N et al (2024) METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. Insights Imaging 15(1):8. https://doi.org/10.1186/s13244-023-01572-w
61. The Cochrane Collaboration, Review Manager (RevMan). Version 5.3.5. 2020.
62. StataCorp. Stata Statistical Software: Release 17. Collection Station, TX: StataCorp. LLC. 2021.
63. Dwamena BA (2007) Meta-analytical integration of diagnostic accuracy studies in Stata. Ben Dwamena
64. Moses LE, Shapiro D, Littenberg B (1993) Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 12(14):1293–1316. https://doi.org/10.1002/sim.4780121403
65. Marrero JA, Kulik LM, Sirlin CB et al (2018) Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the study of liver diseases. Hepatology 68(2):723–750. https://doi.org/10.1002/hep.29913
66. Vesselle G, Quirier-Leleu C, Velasco S et al (2016) Predictive factors for complete response of chemoembolization with drug-eluting beads (DEB-TACE) for hepatocellular carcinoma. Eur Radiol 26(6):1640–1648. https://doi.org/10.1007/s00330-015-3982-y
67. Reis SP, Sutphin PD, Singal AG et al (2017) Tumor enhancement and heterogeneity are associated with treatment response to drug-eluting bead chemoembolization for hepatocellular carcinoma. J Comput Assist Tomogr 41(2):289–293. https://doi.org/10.1097/rct.0000000000000509
68. Tachibana Y, Takaji R, Maruno M et al (2022) LI-RADS classification and outcomes of hepatocellular carcinoma treated with transcatheter arterial chemoembolization plus radiofrequency ablation. Cancer Diagn Progn 2(4):471–481. https://doi.org/10.21873/cdp.10130

69. Huang J, Tian W, Zhang L et al (2020) Preoperative prediction power of imaging methods for microvascular invasion in hepatocellular carcinoma: a systemic review and meta-analysis. Front Oncol. https://doi.org/10.3389/fonc.2020.00887
70. Hong SB, Choi SH, Kim SY et al (2021) MRI features for predicting microvascular invasion of hepatocellular carcinoma: a systematic review and meta-analysis. Liver Cancer 10(2):94–106. https://doi.org/10.1159/000513704
71. Kothari G, Korte J, Lehrer EJ et al (2021) A systematic review and meta-analysis of the prognostic value of radiomics based models in non-small cell lung cancer treated with curative radiotherapy. Radiother Oncol 155:188–203. https://doi.org/10.1016/j.radonc.2020.10.023
72. Ursprung S, Beer L, Bruining A et al (2020) Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis. Eur Radiol 30(6):3558–3566. https://doi.org/10.1007/s00330-020-06666-3
73. Moawad AW, Morshid A, Khalaf AM et al (2023) Multimodality annotated hepatocellular carcinoma data set including pre- and post-TACE with imaging segmentation. Sci Data 10(1):33. https://doi.org/10.1038/s41597-023-01928-3
74. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H (2015) Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087. https://doi.org/10.1038/srep13087
75. Avanzo M, Wei L, Stancanello J et al (2020) Machine and deep learning methods for radiomics. Med Phys 47(5):e185–e202. https://doi.org/10.1002/mp.13678
76. Xia TY, Zhou ZH, Meng XP et al (2023) Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. Radiology 307(4):e222729. https://doi.org/10.1148/radiol.222729
77. Li N, Wan X, Zhang H, Zhang Z, Guo Y, Hong D (2022) Tumor and peritumor radiomics analysis based on contrast-enhanced CT for predicting early and late recurrence of hepatocellular carcinoma after liver resection. BMC Cancer 22(1):664. https://doi.org/10.1186/s12885-022-09743-6

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

Authors and Affiliations

Kaige Deng[1] · Tong Chen[2] · Zijian Leng[1] · Fan Yang[1] · Tao Lu[3] · Jingying Cao[4] · Weixuan Pan[1] · Yongchang Zheng[1]

✉ Yongchang Zheng
zhengyongchang@pumch.cn

Kaige Deng
dkg_9917@163.com

Tong Chen
ct_9917@163.com

Zijian Leng
lengzj_10@163.com

Fan Yang
pumc_YangFan@student.pumc.edu.cn

Tao Lu
lt_bjdx@163.com

Jingying Cao
caojy02202@163.com

Weixuan Pan
pan2001weixuan@126.com

1 Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

2 Department of Medical Oncology, Cancer Hospital, Chinese Academy of Medical Sciences, Beijing 100021, China

3 Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

4 Zunyi Medical University, Zunyi, Guizhou 563000, China