International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

INPLASY202670061

doi: 10.37766/inplasy2026.7.0061

Received: 17 July 2026

Published: 17 July 2026

Diagnostic Value of Radiomic Features for Predicting Kiel 67 Expression in Hepatocellular Carcinoma: A Meta-Analysis

Jia, YS; Yu, B; Ding, LJ; Li, XM; Ru, Y; Sun, H; Song, D; Huang, C; Zhan, JJ; Yin, BY; Li, DP.

Corresponding author:
Dapeng Li

lidapeng1028@163.com

Author Affiliation:
Zhejiang Veterans Hospital.

ADMINISTRATIVE INFORMATION

Support - No support.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202670061

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 17 July 2026 and was last updated on 17 July 2026.

INTRODUCTION

Review question / Objective Review Question & Research Objective (PICOS Framework)

1. Systematic Review Question

What is the pooled diagnostic performance (sensitivity, specificity, AUC) of medical imaging-derived radiomic features for non-invasively predicting high Ki-67 expression status in patients with hepatocellular carcinoma (HCC)? And how do Ki-67 threshold cut-offs, imaging modalities, and predictive modeling algorithms affect the diagnostic efficacy of radiomic models?

2. PICOS Framework Definition

P (Population)

Patients histologically or clinically diagnosed with hepatocellular carcinoma (HCC), with histopathological Ki-67 labeling index results available to dichotomize tumours into high and low Ki-67 expression groups. Total pooled population across included studies: 1,708 HCC patients.

I (Index Test)

Radiomic predictive models built on quantitative high-throughput imaging features extracted from cross-sectional medical imaging, including MRI, ultrasound, and CT; modelling approaches include logistic regression, support vector machine (SVM), deep learning (Xception), and combined radiomics-clinical nomograms.

C (Comparator/Reference Standard)

Histopathological Ki-67 immunohistochemistry (IHC) from surgical resection or biopsy tumour specimens (gold standard). Studies used predefined Ki-67 cut-off values (10% or >10% ranging 14%–50%) to separate high vs low proliferation status.

O (Outcome)

Primary diagnostic accuracy outcomes: pooled sensitivity, pooled specificity, summary receiver operating characteristic (SROC) curve and area under the curve (AUC) for predicting high Ki-67 expression.

Secondary outcomes: subgroup diagnostic performance stratified by Ki-67 cut-off value, imaging modality, and prediction algorithm;

between-study heterogeneity, risk of bias, and publication bias assessment.
S (Study Design)
Original diagnostic accuracy clinical studies (prospective or retrospective single/multicentre cohorts) providing complete 2×2 contingency table data for radiomic Ki-67 prediction; conference abstracts, reviews, case reports, and incomplete datasets were excluded.
3. Primary Research Objective
To systematically quantify the overall diagnostic value of radiomic features for predicting high Ki-67 proliferation status in HCC via meta-analysis, and explore key sources of inter-study heterogeneity through pre-specified subgroup analyses stratified by Ki-67 cut-off thresholds, imaging modalities, and predictive modelling methods, so as to provide evidence for standardized clinical translation of radiomics as a non-invasive surrogate for Ki-67 testing.
4. Secondary Objectives
Compare diagnostic performance between MRI, ultrasound and CT radiomic signatures for Ki-67 prediction.
Evaluate performance differences between logistic regression and advanced machine/deep learning radiomic models.
Assess methodological quality and risk of bias of included diagnostic studies using QUADAS-2.
Detect publication bias and summarise major limitations hindering widespread clinical application of HCC radiomic Ki-67 prediction models.

Condition being studied Condition being studied
Hepatocellular carcinoma (HCC)
Hepatocellular carcinoma is the predominant subtype of primary liver cancer, accounting for approximately 90% of all primary liver malignancies worldwide, and ranks as the fourth leading cause of cancer-related mortality globally. Its major risk factors differ geographically: chronic hepatitis B infection dominates HCC aetiology in East Asia (the population source of nearly all primary studies included in this meta-analysis), while hepatitis C, alcohol abuse and non-alcoholic steatohepatitis are more prevalent drivers in Western populations.
HCC presents insidious early clinical manifestations, resulting in delayed diagnosis for most patients. It exhibits prominent intratumoural spatial and temporal biological heterogeneity, which drives aggressive tumour behaviours including large tumour volume, microvascular invasion, poor histological differentiation and advanced TNM stage. The 5-year overall survival of HCC patients remains low, ranging only from 13% to 36%, largely due to high postoperative recurrence rates and limited effective targeted therapies for advanced disease.
Ki-67 as a core pathological biomarker for HCC
Ki-67 (Kiel 67) is a well-established immunohistochemical marker reflecting tumour cell proliferative activity in HCC. High Ki-67 labelling index correlates strongly with malignant phenotypes: larger tumour size, vascular invasion, higher Edmondson–Steiner grade, advanced tumour staging, increased postoperative recurrence and inferior overall survival. It acts as an independent prognostic biomarker to stratify HCC patients for personalised treatment planning and prognostic monitoring.
Clinical limitation of current Ki-67 testing
Standard Ki-67 detection relies on invasive biopsy or surgical histopathological specimens. Single-site tissue sampling cannot capture the full spatial heterogeneity of HCC, risking underestimation of overall tumour proliferative status; additionally, biopsy carries procedural bleeding risks and patient discomfort. This unmet clinical need motivates research into non-invasive imaging surrogates (radiomics) to predict Ki-67 expression without tissue puncture.

What are the common symptoms and signs of HCC?
How is HCC diagnosed?
What are the treatment options for HCC?

METHODS

Participant or population This systematic meta-analysis targets adult patients diagnosed with hepatocellular carcinoma (HCC), the population enrolled across all 17 eligible primary diagnostic accuracy studies (total pooled sample size = 1,708 HCC patients). Detailed participant characteristics are defined as below:
1. Core Inclusion Criteria for HCC Participants
Confirmed HCC diagnosis
Participants must have a definitive HCC diagnosis verified either via histopathological examination of surgical resection or percutaneous biopsy specimens, or by standard clinical and multi-modal imaging diagnostic guidelines for HCC.
Available Ki-67 immunohistochemistry (IHC) results
All included patients had formal pathological Ki-67 labelling index measurements from tumour tissue samples. Original studies applied explicit Ki-67 threshold cut-offs to stratify participants into two mutually exclusive subgroups:
High Ki-67 expression group (n = 962 across all studies): defined by Ki-67 index ≥10% or higher thresholds ranging 14%–50%

Low Ki-67 expression group (n = 746 across all studies): defined by Ki-67 index below the study-specific cut-off
Complete imaging data for radiomic extraction
Participants underwent pre-treatment standard liver imaging (MRI, contrast-enhanced ultrasound, or contrast-enhanced CT) to generate full tumour region-of-interest images for high-throughput radiomic feature extraction.
Adequate data for diagnostic 2×2 contingency table construction
Each study provided clear counts of true positive, false positive, true negative and false negative predictions for radiomic models versus pathological Ki-67 status, enabling pooled diagnostic meta-analysis.
2. Demographic & Study Design Features of the Population
Geographic origin
Nearly all participant cohorts were recruited from Chinese medical institutions, where chronic hepatitis B infection is the dominant aetiological driver of HCC. This creates a population bias towards East Asian HCC patients with HBV-related liver disease, with limited representation of Western HCC populations driven by hepatitis C, alcohol liver injury or non-alcoholic steatohepatitis (NASH).
Study design split
Retrospective cohorts: 12 studies, majority of the pooled population
Prospective cohorts: 5 studies, smaller participant subgroups
Model construction subgroups within the population
Two distinct participant subsets were pooled together:
Patients assessed via radiomic signature-only predictive models (10 studies)
Patients assessed via combined radiomic + clinical variable nomogram models (7 studies)
Imaging modality subgroups of participants
Participants were stratified by the imaging scan used to extract radiomic features:
MRI cohort (9 studies, largest subgroup)
Contrast-enhanced ultrasound cohort (5 studies)
Contrast-enhanced CT cohort (3 small studies, not pooled for meta-analysis due to insufficient sample volume)
3. Excluded Participant Types
The review excludes participants who met any of the following:
Patients with other primary liver malignancies (e.g., intrahepatic cholangiocarcinoma) or metastatic liver tumours (non-HCC);
Patients without complete pathological Ki-67 IHC testing or missing clear Ki-67 cut-off stratification;
Patients lacking full pre-treatment imaging scans required for radiomic feature extraction;
Duplicate patient cohorts reported across multiple publications;
Participants only described in conference abstracts, case reports, narrative reviews or non-original research without full individual patient diagnostic data.

Intervention The intervention under evaluation in this meta-analysis refers to radiomic predictive models built on quantitative imaging features extracted from pre-treatment liver medical scans, which serve as a non-invasive index test to predict high Ki-67 proliferative biomarker expression in hepatocellular carcinoma (HCC). Full details of the intervention are outlined below:
1. Core Definition of the Index Test Intervention
Radiomics is the core intervention being assessed: a high-throughput quantitative imaging workflow that extracts hundreds of subvisual morphological, texture, intensity, and perfusion features from the whole tumour region of interest (ROI) on standard liver imaging. These extracted radiomic signatures are inputted into statistical or machine learning algorithms to construct binary prediction models that classify HCC tumours as either high or low Ki-67 expression status. This imaging-based radiomic prediction workflow is the non-invasive intervention compared against the invasive histopathological reference standard (Ki-67 immunohistochemistry biopsy).
2. Imaging Modalities Included in the Intervention
All radiomic models included in the review rely on one of three routine clinical liver imaging modalities as the source of feature extraction:
Magnetic resonance imaging (MRI) (9 of 17 studies): Predominantly hepatobiliary contrast-enhanced MRI with Gd-EOB-DTPA; multiparametric sequences including T1, T2, and diffusion-weighted imaging for multi-dimensional feature extraction.
Contrast-enhanced ultrasonography (CEUS) (5 of 17 studies): Sonazoid-based dynamic ultrasound perfusion imaging, capturing real-time microvascular haemodynamic radiomic features.
Contrast-enhanced computed tomography (CE-CT) (3 of 17 studies): Multiphase arterial, portal venous and delayed-phase CT scans for density and vascularity feature extraction.
3. Two Types of Radiomic Modelling Workflows (Subgroup Interventions)
Two distinct radiomic model constructions were pooled and separately analysed as sub-interventions:
Radiomics-only models (10 studies): Predictions rely solely on imaging-derived radiomic features,

without integrating clinical demographic or laboratory variables.
Combined radiomic-clinical nomogram models (7 studies): Radiomic signatures integrated with established clinical HCC risk factors (tumour size, serum AFP, liver function, tumour stage etc.) to build composite predictive nomograms.
4. Modelling Algorithms Used to Generate Radiomic Predictions
Multiple algorithmic approaches constitute variant forms of the radiomic intervention, stratified for subgroup analysis:
Logistic regression (14 studies, mainstream method): Interpretable linear statistical models for radiomic feature classification.
Machine learning nonlinear algorithms (3 studies):
Support Vector Machine (SVM, n=2): High-dimensional nonlinear classifier for radiomic data
Xception deep learning network (n=1): Deep learning combined radiomics (DLCR) model integrating imaging deep features with handcrafted radiomic metrics
5. Standard Workflow of the Radiomic Intervention
Every included study followed a consistent radiomic pipeline as the evaluated intervention:
Pre-treatment acquisition of standard liver cross-sectional imaging;
Manual or semi-automatic segmentation of the entire HCC tumour volume to generate ROI;
Standardised high-throughput extraction of quantitative radiomic features;
Feature dimension reduction to eliminate redundant, non-predictive imaging metrics;
Training and internal validation of a prediction model to dichotomise patients into high vs low Ki-67 expression;
Generation of binary diagnostic outcomes (true positive, false positive, true negative, false negative) for meta-analysis performance pooling.
6. Key Distinction from Conventional Clinical Tools
This radiomic intervention is proposed as a non-invasive alternative/complementary test to the reference standard (invasive tissue biopsy for Ki-67 IHC). Unlike traditional visual imaging assessment or serum tumour markers, the radiomic intervention quantifies invisible intratumoural spatial heterogeneity across the entire tumour, overcoming the sampling bias limitation of single-site biopsy tissue sampling.

Comparator The comparator in this diagnostic test meta-analysis is the reference standard test: immunohistochemical (IHC) staining of Ki-67 antigen on surgically resected or percutaneous biopsy tumour tissue specimens from patients with hepatocellular carcinoma (HCC). This invasive histopathological assay serves as the gold standard against which the radiomic index test (intervention) is compared.
Full definition of the comparator
Testing procedure
Tumour tissue samples obtained via liver biopsy or surgical resection undergo formalin fixation, paraffin embedding, sectioning, and standard Ki-67 IHC staining by clinical pathology laboratories. Pathologists quantify the Ki-67 labelling index, defined as the percentage of tumour cell nuclei with positive Ki-67 staining under light microscopy.
Binary classification standard
Each primary study applied a predefined numerical cutoff value of the Ki-67 index to dichotomise patients into two mutually exclusive comparator groups:
High Ki-67 expression group: Ki-67 index ≥ study-specific threshold (either exactly 10%, or thresholds ranging from 14% to 50% for the >10% subgroup)
Low Ki-67 expression group: Ki-67 index below the defined cutoff
Rationale for this comparator
Ki-67 IHC is the universally accepted clinical reference for assessing tumour proliferative activity in HCC. High Ki-67 expression is an established independent pathological marker of aggressive tumour biology, vascular invasion, postoperative recurrence and poor overall survival. All radiomic prediction models are validated against this tissue-based pathological result to calculate diagnostic accuracy metrics (true positives, false positives, true negatives, false negatives).
Key limitations of the comparator (as contextual comparison)
While acting as the gold standard, this invasive comparator has critical clinical drawbacks that motivate the radiomic index test:
Invasive sampling carries risks of bleeding, infection and patient discomfort;
HCC exhibits strong intratumoural spatial heterogeneity; single-site biopsy only reflects local tumour proliferation status and may misrepresent global Ki-67 levels, leading to sampling error;
Biopsy cannot be repeatedly performed for dynamic longitudinal monitoring of tumour proliferative behaviour.
Other indirect comparative benchmarks (secondary comparison)
Beyond the primary histopathological reference standard, the review indirectly compares multiple variant forms of the radiomic intervention against one another via subgroup analyses:
MRI radiomics vs ultrasound radiomics vs CT radiomics
Logistic regression predictive models vs machine learning/deep learning models

Radiomics-only prediction models vs combined radiomic-clinical nomogram models
Models using Ki-67 cutoff = 10% vs models using Ki-67 cutoff >10%
These subgroup comparisons quantify how different radiomic workflows perform relative to each other when measured against the identical Ki-67 IHC reference comparator.

Study designs to be included Only peer-reviewed diagnostic cohort studies (prospective & retrospective) were included. Eligible studies must supply complete 2×2 diagnostic tables for radiomic Ki-67 prediction in HCC. Excluded: abstracts, case reports, reviews, animal experiments, duplicate datasets and articles without full extractable data.Can you provide more details about the diagnostic accuracy cohort studies?What are the specific requirements for the radiomic Ki-67 prediction in HCC?Are there any limitations or potential biases in the included study designs?Only original diagnostic accuracy cohort studies.

Eligibility criteria Eligibility Criteria
1. Supplementary Inclusion Criteria (beyond PICOS framework)
These additional inclusion rules supplement the PICOS criteria and must be fully satisfied for studies to enter quantitative meta-analysis:
Published full-text original clinical research with formal peer review; only complete articles are eligible, with no restriction on publication language.
The study must provide complete raw diagnostic data to construct a standard 2×2 contingency table (true positives, false positives, true negatives, false negatives). If raw counts were missing from the manuscript, corresponding authors were contacted via email to request original datasets; studies with persistently unavailable core diagnostic data were excluded.
Imaging data used for radiomic feature extraction must be pre-treatment scans of primary hepatocellular carcinoma lesions, without prior locoregional or systemic anti-tumour therapy that could alter tumour imaging phenotypes.
Ki-67 immunohistochemistry testing must be performed on surgical resection or percutaneous biopsy tumour tissue from the index HCC lesion, rather than metastatic liver lesions or non-tumour liver parenchyma.
Radiomic workflows must include full tumour volume segmentation (not single-slice partial sampling) to capture intratumour spatial heterogeneity.
2. Supplementary Exclusion Criteria (beyond PICOS framework)
Studies will be excluded if they meet any of the following extra conditions not covered in PICOS:
Non-full-text publications: conference abstracts, meeting posters, brief communications, letters to editors, editorial comments, and conference proceedings without complete original data.
Non-clinical research: animal experiments, cell line in vitro studies, phantom imaging tests, computational simulation papers without human HCC patient cohorts.
Duplicate publications: multiple articles reporting overlapping or identical patient cohorts; only the largest, most complete dataset was retained for meta-analysis to avoid double-counting participants.
Unavailable full-text manuscripts after repeated retrieval attempts, including papers without open access and no response from authors to full-text requests.
Studies only reporting correlation coefficients between radiomic features and continuous Ki-67 labelling index, without binary classification into high/low Ki-67 groups and corresponding 2×2 diagnostic table.
Studies combining HCC with other liver malignancies (intrahepatic cholangiocarcinoma, metastatic liver tumours) without separate subgroup diagnostic data exclusively for HCC patients.
Models relying solely on subjective visual imaging features (no quantitative high-throughput radiomic extraction pipeline).
3. Additional methodological filtering rules
Studies without clear reporting of Ki-67 cutoff thresholds to dichotomise proliferation status were excluded, as subgroup meta-analysis by Ki-67 cut-off could not be performed.
Diagnostic studies with fewer than 20 total HCC patients were not excluded outright but were retained for pooling and marked for sensitivity analysis to assess the impact of small-sample cohorts on overall pooled estimates.
Studies lacking a clear description of imaging acquisition parameters, tumour segmentation protocols, or feature reduction methods were still included in the meta-analysis but flagged as having unclear risk of bias under QUADAS-2 assessment.

Information sources All information sources used to retrieve eligible primary diagnostic accuracy studies are categorised below, fully aligned with PRISMA 2020 reporting standards, including electronic bibliographic databases, manual secondary searches, and direct author contact for missing raw data. The final search cutoff date was 20 August 2025.

1. Electronic Bibliographic Databases (Primary Search Sources)
Four core biomedical databases were searched from database inception up to 20 August 2025, with no language restrictions applied. Search syntax was customised for each database using subject headings (MeSH, Emtree) and free-text keywords across title, abstract and topic fields.
PubMed (MEDLINE)
Global clinical and preclinical medical database with full MeSH indexing for radiology, hepatology and tumour biomarkers.
Web of Science Core Collection
Citation-indexed database covering peer-reviewed radiology, oncology and medical imaging journals, enabling cross-reference tracking.
Cochrane Library
Repository of systematic reviews, diagnostic test accuracy meta-analyses and clinical trial records for methodological reference.
Embase
Comprehensive European and Asian biomedical database with dedicated Emtree subject terms for radiomics, ultrasound/MRI and liver pathology research.
2. Supplementary Manual Search Sources (Grey & Secondary Literature Screening)
After database retrieval, additional manual searches were performed to capture missed eligible publications:
Reference list backward screening
Full-text reference sections of all initially retrieved relevant articles, published systematic reviews and meta-analyses focused on HCC radiomics or Ki-67 prediction were manually scanned to identify overlooked primary studies.
No dedicated grey literature databases / conference abstract repositories
Conference abstracts were not targeted for full inclusion per pre-specified exclusion criteria; only complete peer-reviewed full-text original articles were retained.
3. Direct Author Correspondence (Data Recovery Source)
When potentially eligible full-text articles lacked complete 2×2 contingency table diagnostic data (true positive, false positive, true negative, false negative counts) required for meta-analysis, the corresponding authors were contacted via institutional email to request raw patient-level diagnostic results. Studies with permanently unavailable data after author follow-up were excluded from quantitative synthesis.
4. Protocol Registration Repository
The review protocol was retrospectively registered in the PROSPERO International Prospective Register of Systematic Reviews, serving as a formal public record of pre-defined search and analysis methods (registration number withheld in manuscript draft as xxxxx).
5. Excluded Information Sources
No additional platforms (Google Scholar, clinical trial registries, preprint servers, regional Chinese medical databases) were searched for this review to maintain consistent, internationally standardised literature retrieval aligned with mainstream diagnostic meta-analysis practice.

Main outcome(s) Primary outcome: Overall diagnostic accuracy of radiomics to predict high Ki-67 in HCC, with pre-treatment imaging as index test and post-biopsy/resection Ki-67 IHC as reference standard. Effect measures: pooled sensitivity, specificity, SROC-AUC with 95% CIs. Global pooled results: sensitivity 0.87 (0.81–0.91), specificity 0.79 (0.71–0.93), AUC 0.90 (0.87–0.93).
Secondary subgroup outcomes stratified by Ki-67 cutoff, imaging modality and modelling method, reporting identical accuracy metrics. Heterogeneity metrics (I^2, Q-test, Spearman correlation for threshold effect), QUADAS-2 risk-of-bias ratings and Deeks' funnel plot for publication bias were supplementary analytical outcomes. All tests adopted two-sided α=0.05.

Quality assessment / Risk of bias analysis
QUADAS-2 tool was adopted to evaluate risk of bias for all included diagnostic accuracy studies. Four core evaluation domains were assessed independently: patient selection, index test, reference standard, flow and timing. Each domain was rated as low, high or unclear risk of bias based on published signalling questions. Two researchers finished assessments separately; disagreements were resolved via discussion or a third reviewer's judgement. No composite numerical quality score was generated as QUADAS-2 discourages this practice. Visual outputs including methodological quality graph and summary plot were produced in Review Manager 5.3 to display bias distribution across all 17 studies. Beyond QUADAS-2 evaluation, Deeks' funnel plot asymmetry test was conducted in Stata's midas module to detect publication bias, with p80%).

Strategy of data synthesis
All statistical analyses were performed with Stata 16.0 (midas module) and Review Manager 5.3. A bivariate random-effects model was selected as the core synthesis method, given substantial inter-study heterogeneity (I^2 > 80.81 for sensitivity, I^2 > 86.86 for specificity). This model jointly pooled sensitivity and specificity while accounting for between-study variance, generating pooled point estimates with 95% CIs, and plotting the summary receiver operating characteristic (SROC) curve to

calculate overall AUC as the primary discriminative metric.

Heterogeneity was quantified via Q-test and I^2 statistic; $I^2 > 50\%$ denoted substantial heterogeneity. Spearman correlation between logit(sensitivity) and logit(1-specificity) was calculated to test threshold effect ($r = -0.062$, $p = 0.814$, no significant threshold-driven heterogeneity).

Predefined subgroup analyses were conducted to explore heterogeneity sources: stratification by Ki-67 cutoff (10% vs >10%), imaging modality (MRI, ultrasound; CT only narratively summarised due to n=3), and modelling algorithm (logistic regression pooled; machine learning described separately for limited samples). Each eligible subgroup adopted the identical bivariate random-effects framework.

Multiple sensitivity analyses were implemented to test result robustness: leave-one-out iteration, separation of prospective/retrospective cohorts, comparison of radiomics-only versus clinical-combined models, exclusion of small-sample studies and high QUADAS-2 risk-of-bias papers. Pooled metrics were recalculated for each filtered dataset and cross-compared with global results.

Publication bias was assessed via Deeks' funnel plot asymmetry test ($p = 0.42$, no significant bias). Risk of bias across included studies was evaluated using QUADAS-2, with visual summary graphs produced in RevMan. Two independent researchers extracted 2×2 contingency table data; missing raw data were requested from corresponding authors, with incomplete studies excluded. All two-sided statistical tests adopted $\alpha = 0.05$. Narrative synthesis supplemented quantitative pooling for subgroups with insufficient study numbers.

Subgroup analysis Predefined subgroup analyses were performed to explore sources of substantial inter-study heterogeneity (I^2>80%) using consistent bivariate random-effects models, limited to strata with ≥5 studies for quantitative pooling. First, studies were split by Ki-67 cutoff: 8 papers using 10% threshold yielded pooled AUC=0.89; 9 papers with cutoffs >10% (14%–50%) had AUC=0.90, with nearly identical sensitivity (0.87 for both), slightly higher specificity in the higher-cutoff group. Second, imaging modality subgroups: 9 MRI studies produced pooled AUC=0.88, while 5 ultrasound cohorts reached AUC=0.92 with overlapping confidence intervals, precluding definitive superiority. Three CT studies were too few for pooling and only summarised narratively. Third, modelling algorithm subgroup: 14 logistic regression studies had pooled AUC=0.89; 3 machine learning papers (2 SVM, 1 Xception) lacked sufficient sample size for meta-analysis and were tabulated separately. All subgroup SROC curves and sensitivity/specificity forest plots were generated as supplementary figures. Subgroup differences were mild without statistically significant gaps, indicating radiomic performance remains stable across clinical and methodological variations.Predefined subgroup analyses using bivariate random-effects models were conducted to explore high inter-study heterogeneity (I^2>80%). First, stratified by Ki-67 cutoff: 8 studies with 10% threshold yielded AUC=0.89; 9 studies using cutoffs over 10% reached AUC=0.90, with identical pooled sensitivity (0.87) and slightly better specificity for higher thresholds. Second, imaging modality subgroups: nine MRI cohorts had pooled AUC=0.88, while five ultrasound studies produced AUC=0.92; overlapping confidence intervals ruled out definitive superiority of ultrasound. Three CT studies had too few samples for pooled meta-analysis and were only described narratively. Third, stratified by modelling methods: 14 logistic regression studies obtained AUC=0.89; three machine learning papers (2 SVM, 1 Xception) were not pooled due to limited sample size and summarised separately. All subgroups generated independent SROC curves and sensitivity/specificity forest plots as supplementary materials. Across all stratifications, diagnostic metrics had minor differences without statistically meaningful gaps, proving radiomic predictive performance remained stable regardless of pathological cutoffs, imaging tools or modelling algorithms.

Sensitivity analysis Multiple pre-planned sensitivity analyses were performed to test the robustness of pooled diagnostic metrics and explore sources of high inter-study heterogeneity ($I^2 > 80\%$ for sensitivity and specificity). All analyses reused the bivariate random-effects model consistent with the primary meta-analysis.

First, leave-one-out analysis was conducted by sequentially removing each single study and recalculating pooled sensitivity, specificity and AUC. No single study drastically altered overall point estimates, confirming the main findings were not driven by outlier cohorts. Second, sensitivity analysis stratified by study design separated 5 prospective and 12 retrospective studies. Retrospective cohorts yielded slightly inflated performance, while prospective data produced marginally lower AUC, indicating retrospective selection bias modestly overestimated predictive power.

Third, analyses split models into radiomics-only (10 studies) and radiomics-clinical nomograms (7 studies). Integrated nomograms showed marginally better specificity, but global AUC shifts were

minimal, proving radiomic signatures remained the core predictive driver regardless of clinical covariate integration. Fourth, we excluded small-sample studies (fewer than 50 patients). After removal, pooled AUC changed negligibly, ruling out small overfitted cohorts as the primary source of heterogeneity.

Fifth, sensitivity filtering based on QUADAS-2 risk of bias: only studies rated low risk across all four domains were re-analysed. Pooled performance slightly declined, showing moderate-quality biased studies marginally overoptimised diagnostic accuracy. Sixth, we eliminated CT and machine learning small subgroups separately; overall AUC remained stable, demonstrating these minority subsets did not distort main pooled results.

An auxiliary sensitivity check reconfirmed threshold effect via Spearman correlation ($r = -0.062$, $p = 0.814$), consistent across all filtered datasets, proving differing Ki-67 cutoffs were not the dominant heterogeneity source. All sensitivity test outcomes were tabulated as supplementary material with overlapping confidence intervals observed across nearly all subsets, verifying the primary conclusion (overall AUC = 0.90, robust radiomic predictive capacity) was stable and credible after multiple filtering adjustments.

Country(ies) involved China.

Keywords Hepatocellular carcinoma, Ki-67 antigen, radiomics, diagnostic value, meta-analysis.

Contributions of each author

Author 1 - Yushan Jia.
Author 2 - Bo Yu.
Author 3 - Lijie Ding.
Author 4 - Xinmiao Li.
Author 5 - Yun Ru.
Author 6 - Hui Sun.
Author 7 - Da Song.
Author 8 - Chen Huang.
Author 9 - Jinjie Zhan.
Author 10 - Boyang Yin.
Author 11 - Dapeng Li.