# Artificial Intelligence and Radiomics for Predicting Chemotherapy Response in Colorectal Liver Metastases: A Systematic Review and Meta-Analysis

Luis Lorenzo A. Chan, MD; Adrian Christopher A. Elio, MD

## Abstract

**Objective and Hypothesis:** The emergence of artificial intelligence (AI) and radiomics in oncology has introduced new paradigms in disease characterization, risk stratification, and treatment monitoring. In patients with colorectal liver metastases (CRLM), the ability to noninvasively predict chemotherapy response using advanced imaging biomarkers could significantly influence therapeutic decision-making. This systematic review and meta-analysis evaluates the diagnostic performance of radiomics-based AI models in predicting treatment response in CRLM.

**Methods:** A systematic review of PubMed and Google Scholar was conducted for studies from January 2015 to April 2025 using radiomics-based AI models for CRLM chemotherapy response prediction. Eligible studies reported model performance against a reference standard (Response Evaluation Criteria in Solid Tumors version 1.1, histopathology, or survival). Meta-analysis using Review Manager 5.4 generated pooled area under the curve (AUC), sensitivity, and specificity. Subgroup analyses were performed by treatment regimen, AI model type, validation strategy, and outcome definition.

**Results:** 21 studies comprising 2769 patients were included. The pooled AUC for all studies was 0.86 (95% CI 0.79-0.94), with a sensitivity of 0.83 and specificity of 0.65. Chemotherapy-only cohorts showed AUC 0.85, sensitivity 0.78, and specificity 0.66, while cohorts receiving chemotherapy plus targeted therapy showed AUC 0.88, sensitivity 0.96, and specificity 0.63.

**Conclusion:** Radiomics-based AI models achieve robust predictive accuracy in chemotherapy response stratification for CRLM, particularly when biologic agents are used. These findings support AI integration as a noninvasive, scalable adjunct to treatment planning in precision oncology.

**Keywords:** artificial intelligence, radiomics, colorectal liver metastases, chemotherapy response, precision oncology

## Introduction

Colorectal cancer (CRC) remains one of the most prevalent malignancies worldwide and a leading cause of cancer-related mortality.[1] A significant proportion of patients with CRC, which is approximately 25-50%, will develop liver metastases either at the time of diagnosis or during the course of disease progression.[1,2] These colorectal liver metastases (CRLM) represent a critical therapeutic challenge, as their presence significantly impacts prognosis, with 5-year survival rates remaining below 15% without intervention.[3] The mainstay of CRLM management includes systemic chemotherapy, targeted therapies, and surgical resection, with the latter being the only potentially curative option.[2] However, not all patients respond equally to systemic treatments, and early identification of responders versus nonresponders is essential for tailoring therapeutic

**Affiliation:** Institute of Radiology, St. Luke's Medical Center, Quezon City, Philippines.
**Disclosures:** The authors have no conflicts of interest to disclose. None of the authors received outside funding for the production of this original manuscript and no part of this article has been previously published elsewhere.
**Data Availability Statement:** All data relevant to the study are provided in the article.

https://doi.org/10.37549/AR-D-25-0146

plans, minimizing toxicity, and maximizing survival benefit.[4]

Traditional methods of evaluating treatment response rely heavily on morphological criteria such as the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1.[5] While RECIST provides standardized guidelines for assessing tumor shrinkage, it is inherently limited by its dependence on linear size measurements and its inability to capture subtle biological or functional changes within tumors.[6,7] Consequently, there has been growing interest in more sophisticated tools that can quantify intratumoral heterogeneity and predict treatment outcomes beyond size-based assessments.

Radiomics refers to the high-throughput extraction of quantitative features from medical imaging data. These features, encompassing tumor shape, intensity, texture, and wavelet transformations, reflect underlying pathophysiological processes and can provide insights into tumor behavior, aggressiveness, and response to therapy.[8-11] When paired with artificial intelligence (AI) techniques such as machine learning (ML) and deep learning (DL), radiomics can be transformed into predictive models that recognize complex patterns and make data-driven predictions. ML algorithms can be trained on large, labeled datasets to distinguish between responders and nonresponders based on imaging-derived features, while DL approaches such as convolutional neural networks can autonomously learn hierarchical representations directly from raw imaging data.[12]

The application of AI and radiomics in CRLM has gained traction in recent years, driven by advances in computational power, access to large imaging datasets, and improvements in imaging standardization. Numerous studies have explored the use of radiomics-based signatures derived from CT, MRI, or PET/CT to predict histopathologic response or RECIST-based radiologic response in patients undergoing chemotherapy. These studies vary widely in terms of patient population, treatment regimens, feature extraction methods, and validation approaches, necessitating a systematic synthesis to determine overall diagnostic performance.[13-33]

This meta-analysis aims to address this gap by pooling data from studies using radiomics-based AI models to predict chemotherapy response in CRLM. In addition to evaluating overall accuracy, we assess model performance in key subgroups, particularly those receiving chemotherapy alone versus those receiving combination therapy with biologics such as bevacizumab or cetuximab.

## Methods

### Search Strategy and Data Extraction

A systematic search of the literature was conducted using PubMed and Google Scholar for relevant articles published between January 1, 2015, and April 1, 2025. The search utilized combinations of the following keywords: "artificial intelligence," "radiomics," "colorectal liver metastases," "chemotherapy response," "machine learning," and "deep learning." Only peer-reviewed studies published in English were included. 2 reviewers independently screened titles and abstracts, followed by full-text review to identify studies meeting the eligibility criteria. Data were extracted using a standardized form that included information on study design, patient characteristics, imaging modality, AI model type, radiomics feature sets, comparator, reference standard, and diagnostic performance metrics [area under the curve (AUC), sensitivity, and specificity].

### Study Selection

Inclusion criteria required studies to involve patients with CRLM undergoing chemotherapy, utilize radiomics-based AI models as index tests, and compare results to a reference standard such as RECIST 1.1 or histopathology, with provided extractable diagnostic accuracy metrics. Both retrospective and prospective studies were eligible. Exclusion criteria included case reports, reviews, conference abstracts, nonpeer-reviewed articles, and studies with incomplete data or high methodological bias.

### Objectives

The primary objective was to evaluate the pooled diagnostic performance of radiomics-based AI models in predicting chemotherapy response in CRLM, specifically measuring AUC, sensitivity, and specificity. Secondary objectives were to compare performance by treatment regimen (chemotherapy alone vs chemotherapy with targeted therapy) and evaluate the impact of model heterogeneity (AI method), validation rigor (internal vs external), and outcome definition (RECIST vs pathology vs survival).

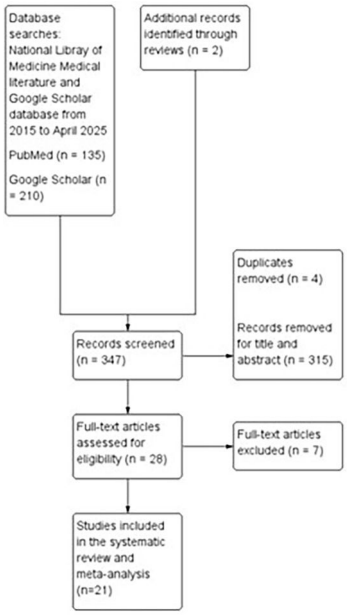
### Assessment of Publication Quality

Methodological quality was evaluated using the QUADAS-2 tool. 2 reviewers assessed risk of bias across 4 domains: patient selection, index test, reference standard, and flow/timing. All included studies were rated as low risk of bias, indicating high methodological rigor. Validation rigor was additionally classified according to TRIPOD-AI[34] guidance as (1) internal validation only, (2) external validation using independent cohorts, or (3) temporal quasi-external validation.

### Statistical Analysis

Meta-analysis was performed using Review Manager version 5.4. Pooled estimates for sensitivity, specificity, and AUC were calculated using random-effects models. Summary receiver operating characteristic curves were generated to visualize model performance. Subgroup analyses were performed for treatment regimen, AI method, validation rigor, and outcome definition subgroups. Heterogeneity was assessed using the chi-square statistic and interpreted alongside clinical variability.

**Figure 1.** Study flow diagram (based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses).




### Ethical Considerations

This study was reviewed and approved by the St. Luke's Medical Center Institutional Ethics Review Committee. It was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki (2013) and followed the International Council for Harmonisation–Good Clinical Practice guidelines.

As a systematic review and meta-analysis of published studies, no new patient data were collected or accessed. All extracted data were anonymized and publicly available, with data handling compliant with the Data Privacy Act of 2012.

## Results

There are 21 studies included in this systematic review and meta-analysis. These were published between 2020 and 2025. Overall, there are 2769 patients included in this study.[13-33] All studies used radiomics-based or DL-based AI models to predict treatment response.

The literature search resulted in 347 total citations, with 135 from PubMed, 210 from Google Scholar, and 2 from further reviews (Figure 1). After removing duplicates and screening through the titles and abstracts, we screened 28 unique citations. During the full-text assessment phase, 7 articles were excluded as they did not meet the inclusion criteria. 21 articles were included in the final analysis.

Studies were grouped according to treatment regimen (chemotherapy alone vs chemotherapy plus targeted therapy), AI model type (radiomics-based ML, DL, or delta-radiomics), validation design (internal vs external), and reference standard (RECIST 1.1, histopathology, or survival-based response). A detailed summary of included studies is provided in Table 1.

Considerable methodological heterogeneity was observed across studies. 14 studies relied exclusively on internal validation strategies, including cross-validation and bootstrapping, while 5 studies used a form of temporal or internal-external validation. Only 2 studies employed independent external validation cohorts, and none achieved full TRIPOD-AI level 2b or 3 validation standards. The definition of treatment response also varied. 17 studies used RECIST 1.1 as the reference standard, 3 studies used histopathologic tumor regression grade, and 1 study defined response based on an overall survival threshold.

The pooled area AUC for all studies combined was 0.86 (95% CI 0.79 to 0.94) (Figures 2, 3), indicating excellent overall diagnostic performance. Among the 21 studies, 15 provided sufficient data for pooled sensitivity and specificity analysis. Sensitivity was 0.83 (95% CI 0.81 to 0.86) (Figure 4), demonstrating the model's strength in correctly identifying responders. Specificity was 0.65 (95% CI 0.62 to 0.68) (Figure 5), suggesting moderate capability in correctly identifying nonresponders.

### Treatment Regimen Subgroup

Models applied to chemotherapy-only cohorts ($n = 14$) yielded a pooled AUC of 0.85 (95% CI 0.76-0.95) (Figure 6), with a sensitivity of 0.78 (95% CI 0.74-0.81) (Figure 7) and specificity of 0.66 (95% CI 0.62-0.70) (Figure 8). Studies including targeted agents ($n = 7$) demonstrated the highest pooled diagnostic performance, with an AUC of 0.88 (95% CI 0.77-1.00) (Figure 9), markedly higher sensitivity at 0.96 (95% CI 0.93-0.98) (Figure 10), and comparable specificity of 0.63 (95% CI 0.57-0.69) (Figure 11).

### AI Model Type Subgroup

Radiomics-based ML models ($n = 12$) produced a pooled AUC of 0.85 (95% CI 0.74-0.97), with a sensitivity of 0.74 (95% CI 0.68-0.79) and specificity of 0.74 (95% CI 0.68-0.80). DL approaches ($n = 6$) showed similar overall discrimination (AUC 0.84; 95% CI 0.71-0.97) but with higher sensitivity (0.88; 95% CI 0.85-0.90) and lower specificity (0.59; 95% CI 0.55-0.64). Delta-radiomics studies ($n = 3$) reported the strongest pooled AUC (0.91; 95% CI 0.77-1.05), although based on substantially fewer datasets.

### Validation Strategy Subgroup

Internally validated models ($n = 15$) achieved a pooled AUC of 0.83 (95% CI 0.73-0.93), with a sensitivity of 0.86 (95% CI 0.83-0.88) and specificity of 0.67 (95% CI 0.63-0.71). Temporal or internal-external validation designs ($n = 4$) yielded the highest pooled AUC (0.92; 95% CI 0.80-1.04), with a sensitivity of 0.74 (95% CI 0.67-0.80) and specificity of 0.59 (95% CI 0.52-0.65). Independent external validation ($n = 2$) showed lower pooled AUC (0.82; 95% CI 0.55-1.09) but maintained a sensitivity of 0.86 (95% CI 0.73-0.94) and higher specificity (0.77; 95% CI 0.64-0.88).

### Outcome Definition Subgroup

RECIST-based studies ($n = 17$) demonstrated a pooled AUC of 0.86 (95% CI 0.78-0.95), with a sensitivity of 0.77 (95% CI 0.74-0.81) and specificity of 0.67 (95% CI 0.63-0.71). Histopathology-based response studies ($n = 3$) produced a pooled AUC of 0.88 (95% CI 0.72-1.04) and very high sensitivity (0.98; 95% CI 0.95-0.99) but lower specificity (0.59; 95% CI 0.52-0.65). One study using a survival-based definition reported an AUC of 0.80 (95% CI 0.32-1.28).


https://doi.org/10.37549/AR-D-25-0146

**Table 1. AI Analysis of the Response to Standard Chemotherapy Alone for Patients with Colorectal Metastasis**

| STUDY | YEAR | PATIENTS | DESIGN | CHEMOTHERAPY REGIMEN | AI (ML/DL) | SIGNATURE | IMAGE MODALITY | RADIOMICS FEATURES | COMPARATOR | REFERENCE STANDARD |
|---|---|---|---|---|---|---|---|---|---|---|
| *Chemotherapy only* | | | | | | | | | | |
| Giannini | 2022 | 57 | Retrospective | FOLFOX | ML | Delta-radiomics | CT | CT scans at differential time points, 14 lesion shape-based 18 images intensity-based statistics, 75 images gray-level-based statistics | RECIST 1.1 | RECIST 1.1 |
| Wei | 2020 | 192 | Retrospective | Oxaliplatin or irinotecan | DL | Fusion radiomics | MDCT | CT scans at baseline; 583 radiomics features (traditional radiomics) | RECIST 1.1 | RECIST 1.1 |
| Nakanishi | 2021 | 42 | Retrospective | Oxaliplatin | ML | Radiomics | CT | CT scans at baseline; 1038 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Defeudis | 2022 | 92 | Retrospective | FOLFOX or FOLFIRI | ML | Radiomics | CT | CT scans at baseline; 75 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Chen | 2024 | 85 | Retrospective | XELOX, FOLFOX, FOLFIRI | ML | Radiomics | CT and MRI | MRI at baseline; 1786 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Giannini | 2020 | 95 | Retrospective | FOLFOX or FOLFIRI | ML | Radiomics | CT | CT scans at baseline; 22 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Karagkounis | 2024 | 85 | Retrospective | FOLFOX or FOLFIRI | ML | Radiomics | CT | CT scans at baseline; 21 radiomics features (19 texture features, contrast, and average gradient sum) | RECIST 1.1 and CT morphologic criteria | Histopathologic result |
| Rabe | 2022 | 29 | Retrospective | FOLFOX, FOLFIRI, FOLFOXIRI, CAPE-OX, CAPE-IRI, or capecitabine | ML | Radiomics | CT | CT scans at baseline; 175 radiomics features | RECIST 1.1 | RECIST 1.1 |

**Table 1. continued**

| STUDY | YEAR | PATIENTS | DESIGN | CHEMOTHERAPY REGIMEN | AI (ML/DL) | SIGNATURE | IMAGE MODALITY | RADIOMICS FEATURES | COMPARATOR | REFERENCE STANDARD |
|---|---|---|---|---|---|---|---|---|---|---|
| Su | 2022 | 100 | Retrospective | XELOX, FOLFOX, FOLFIRI | ML | Radiomics | MRI | MRI at baseline; 1688 radiomics features; used 10 single-classifier radiomics models and IVIM parameter images | RECIST 1.1 | RECIST 1.1 |
| Davis | 2024 | 95 | Prospective | 5-FU with oxaliplatin and/or irinotecan | DL | Radiomics | CT | CT scans at baseline; used 2 DIM (ABMIL and MIL) | RECIST 1.1 | RECIST 1.1 |
| Gennaro | 2025 | 84 | Retrospective | Systemic chemotherapy | DL | Delta radiomics | CT | CT scans at baseline; 62 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Piliposyan | 2025 | 355 | Retrospective | Systemic chemotherapy | DL | Radiomics + CNN | CT | CT scans at baseline; 107 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Qi | 2023 | 116 | Retrospective | Irinotecan-based chemotherapy | ML | Radiomics | CT | CT scans at baseline; first-order, shape, and texture features from multiscale voxel sampling (1 × 1 × 1 to 5 × 5 × 5 $mm^3$) | RECIST 1.1 | RECIST 1.1 |
| Ma | 2021 | 102 | Retrospective | FOLFOX | ML | Delta-radiomics | MRI | MRI at baseline; 396 radiomics features | RECIST 1.1 | RECIST 1.1 |
| *Chemotherapy with targeted therapy* | | | | | | | | | | |
| Yoon | 2024 | 17 | Prospective | FOLFIRI, folinic acid fluorouracil, or irinotecan with bevacizumab or cetuximab | ML | Radiomics | CT and MRI | MRI baseline; 107 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Zhu | 2021 | 180 | Mixed | FOLFOX, FOLFIRI, XELOX with bevacizumab or anti-EGFR antibody | DL | Radiomics | MRI | Used 3 DL models | RECIST 1.1 and histopathology | RECIST 1.1 and histopathology |
| Dercle | 2020 | 129 | Retrospective | FOLFIRI + cetuximab | ML | Radiomics | CT | CT scans at differential time | Overall survival > 17.7 months | Overall survival > 17.7 months |



https://doi.org/10.37549/AR-D-25-0146

**Table 1. continued**

| STUDY | YEAR | PATIENTS | DESIGN | CHEMOTHERAPY REGIMEN | AI (ML/DL) | SIGNATURE | IMAGE MODALITY | RADIOMICS FEATURES | COMPARATOR | REFERENCE STANDARD |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | points, 1749 radiomics features, 8 rim specific features | | |
| Maaref | 2020 | 202 | Retrospective | FOLFOX or FOLFIRI with bevacizumab | ML+ DL (DCNN) | Radiomics + DCNN | CT | CT scans at baseline; 14 textural features | Histopathology | Histopathology |
| Miyamoto | 2024 | 150 | Retrospective | FOLFOX or irinotecan with bevacizumab or anti-EGFR antibody | ML | Radiomics | CT | CT scans at baseline; 107 radiomics features | RECIST 1.1 | RECIST 1.1 |
| Zhou | 2023 | 307 | Retrospective | FOLFOX with bevacizumab | DL | Radiomics | PET/CT | PET/CT baseline; multimodal fusion | RECIST 1.1 | RECIST 1.1 |
| Qu | 2023 | 76 | Retrospective | FOLFOX/XELOX/ FOLFIRI with bevacizumab | ML | Radiomics | CT | CT scan baseline; used 5 models | RECIST 1.1 | RECIST 1.1 |

*Data are grouped according to chemotherapy-only and chemotherapy-with-targeted-therapy subgroups.*

*AI, artificial intelligence; CNN, convolutional neural network; DL, deep learning; ML, machine learning; RECIST, Response Evaluation Criteria in Solid Tumors.*

At the study level, several methodological patterns recurred across higher-performing models. Approaches that incorporated temporal or multiphasic information such as delta-radiomics[16] and dynamic radiomics,[18] longitudinal MRI or multiparametric imaging,[13,20,23,24] and DL-based feature extraction[21,24-28] frequently achieved AUC values above 0.80-0.90. Models that incorporated peritumoral or microenvironmental features,[17,19] or fused imaging with clinical and biologic data,[28-32] also tended to yield stronger performance.

## Discussion

The findings of this systematic review and meta-analysis provide evidence that AI and radiomics-based models can play a significant role in predicting chemotherapy response in patients with CRLM. The pooled AUC of 0.86 for all studies supports the clinical utility of these tools in stratifying patients based on their likelihood of responding to systemic therapy.

Our findings are reinforced by multiple innovative radiomics strategies across included studies. Delta-radiomics[16] and dynamic feature analysis[18] offered superior early response prediction, aligning with the need for early identification of nonresponders to avoid unnecessary treatment and toxicity.[35] Su[20] and Yoon[23] demonstrated the power of serial and longitudinal MRI, highlighting the importance of capturing temporal changes in tumor characteristics.[36] Ensemble learning[20] and attention-based DL[25] showed performance advantages over traditional radiomics, supporting the impression that more sophisticated AI techniques can better extract and learn from complex imaging data.[36] Integration of tumor microenvironment,[17] peritumoral features,[21] and biologic markers[28] further improved prediction, underscoring the multifactorial nature of treatment response and the benefit of incorporating diverse data sources.

The particularly high sensitivity observed in the chemotherapy with targeted therapy subgroup is notable. This group included studies such as Zhou et al[28], who developed a multimodal model integrating PET/CT, clinical, and histopathological features (DERBY+), which achieved an AUC of 0.83. Similarly, Dercle et al demonstrated an AUC of 0.80 using a CT-based radiomics response signature in patients treated with anti-EGFR therapy.[29] Yoon et al further validated the potential of MRI radiomics, achieving an AUC of 0.857 in a longitudinal DL model predicting early response to biologic therapy.[23] Other key contributors to this subgroup include Zhu et al,[24] Maaref et al,[30] Miyamoto et al,[22] and Qu et al,[18] all of which underscore the value of integrating temporal and multiparametric imaging in biologic therapy response prediction. This finding may reflect the enhanced biological impact of agents like bevacizumab and cetuximab, which influence tumor vasculature and molecular expression profiles. These

**Figure 2.** Pooled under the curve (AUC) of radiomics-based artificial intelligence (AI) models predicting chemotherapy response in colorectal liver metastases. The pooled random-effects estimate (diamond) demonstrates excellent overall discrimination (AUC = 0.86, 95% CI 0.79-0.94). Values to the right of zero favor AI-based prediction over conventional assessment methods.

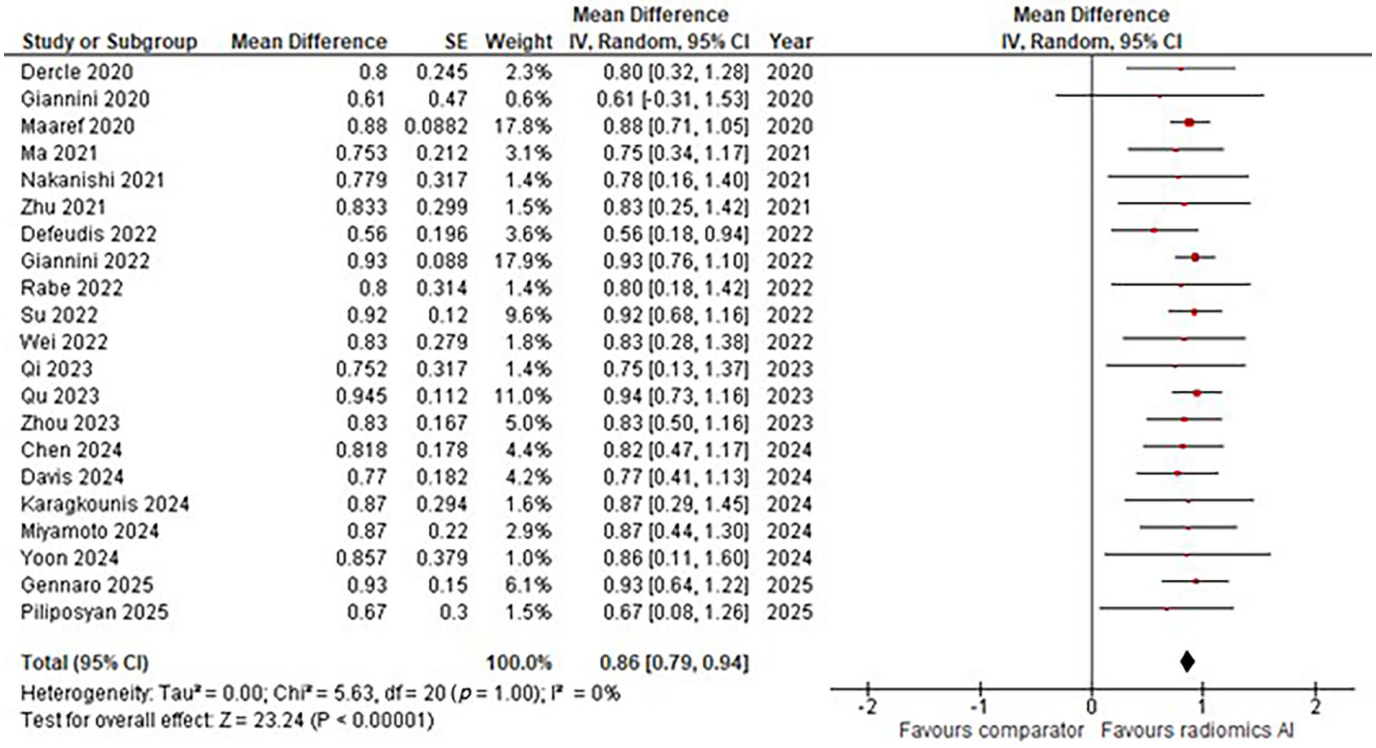
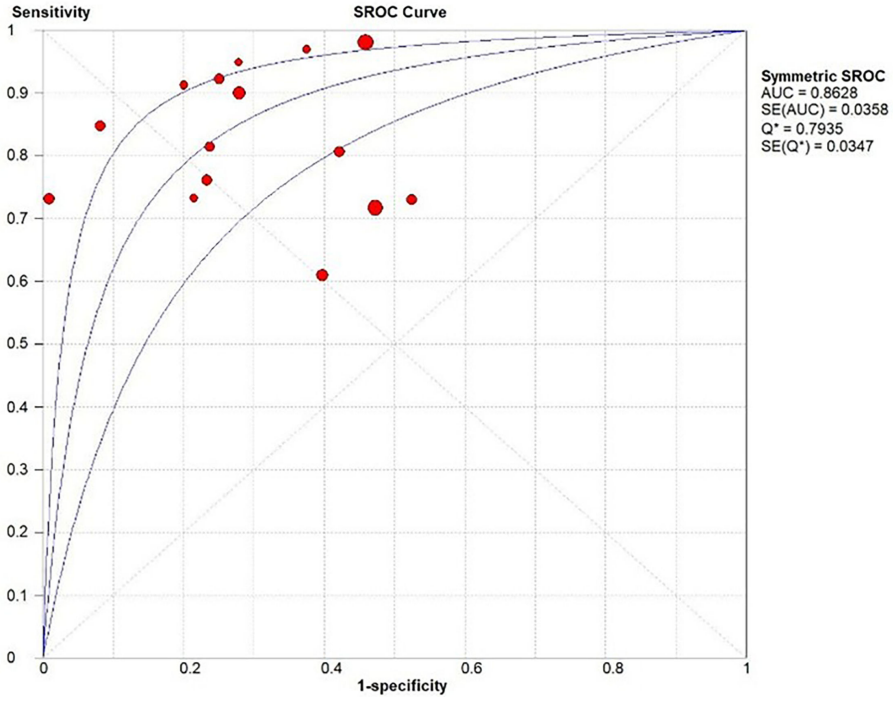



**Figure 3.** Summary receiver operating characteristic (SROC) curve demonstrating the diagnostic performance of radiomics-based artificial intelligence models versus standard comparators across all included studies. The pooled area under the curve (AUC) was 0.87, indicating good overall discriminative ability.




subtle changes are not always readily apparent through size-based RECIST criteria but may be detected through quantitative imaging biomarkers. AI models, especially those utilizing DL, are well-suited to capture these nuances due to their ability to learn hierarchical imaging features.[37]

In contrast, the chemotherapy-only subgroup showed slightly lower sensitivity and similar specificity. However, several high-performing studies still emerged within this group. For instance, Chen et al achieved an AUC of 0.818 by incorporating hepatobiliary phase MRI radiomics,[13] and Karagkounis et al reported an AUC of 0.87 with 100% sensitivity using post-treatment CT features from tumoral and peritumoral regions.[17] These findings illustrate that even without adjunct biologic therapy, advanced radiomics frameworks can deliver high diagnostic accuracy. Additional models such as those by Su et al,[20] Rabe et al,[19] and Giannini et al[15,16] further support the feasibility of using imaging-based AI approaches,


https://doi.org/10.37549/AR-D-25-0146

**Figure 4.** Pooled sensitivity of radiomics-based artificial intelligence models. Sensitivity was 0.83 (95% CI 0.81-0.86). Heterogeneity ($I^2$ = 86.8%) suggests performance differences across model types and patient cohorts.

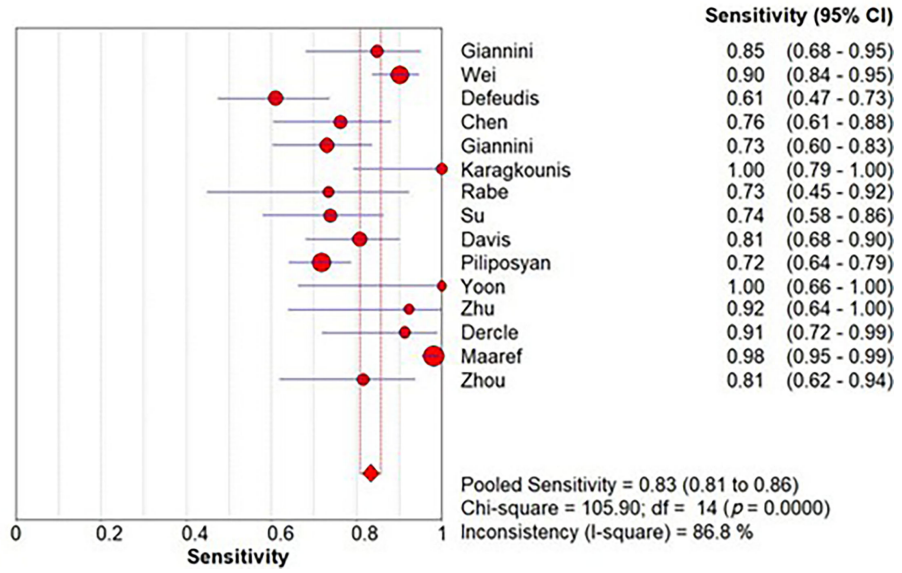
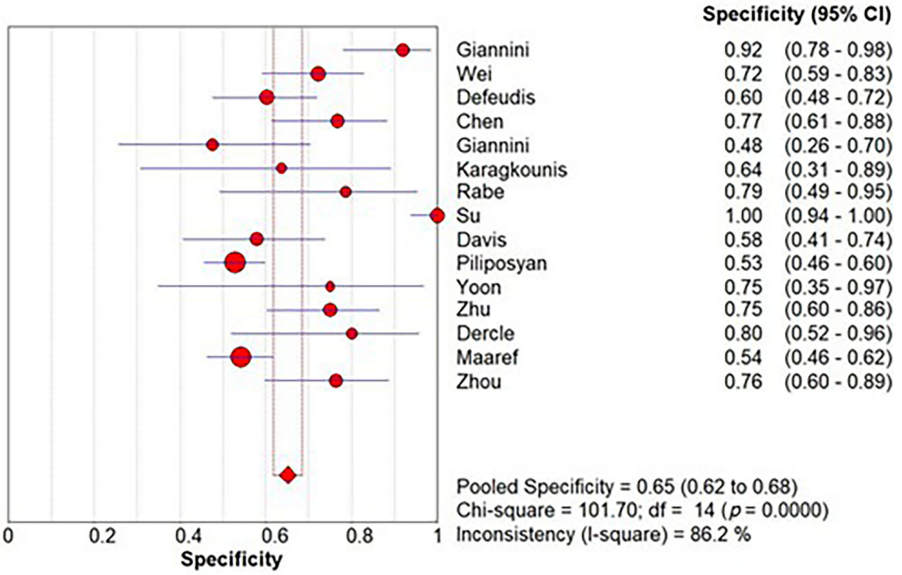



**Figure 5.** Pooled specificity of radiomics-based artificial intelligence models. The overall pooled specificity was 0.65 (95% CI 0.62-0.68), with substantial heterogeneity ($I^2$ = 86.2%).




particularly those leveraging longitudinal analysis, lesion-level heterogeneity, and texture-specific features, for predicting chemotherapy response with strong clinical relevance.

The differential performance between chemotherapy-only and biologic-inclusive cohorts suggests that radiomics may be particularly sensitive to biologically targeted treatment effects. This supports the notion that AI-derived imaging biomarkers can serve as noninvasive surrogates for molecular or histologic assessment.[38] The high AUCs seen in biologic-treated cohorts[28-30] also validate AI's role in predicting response to agents with complex mechanisms beyond cytotoxicity.[39]

Radiomics-based ML and DL models demonstrated comparable overall discrimination, although DL models showed higher pooled sensitivity and lower specificity, indicating a tendency to favor responder classification at the expense of misclassifying nonresponders. Delta-radiomics models yielded the highest pooled AUC despite fewer included studies, maintaining a more balanced sensitivity-specificity profile. These findings suggest that models leveraging temporal or treatment-induced imaging changes may better reflect true biologic response than those relying solely on baseline tumor features. From a clinical standpoint, highly sensitive models may reduce the risk of prematurely halting effective therapy, whereas models with more balanced accuracy may better prevent continuation of ineffective regimens.

However, the observed performance must be interpreted in the context of nontrivial methodological limitations. A large majority of studies relied on internal validation only, with relatively few employing temporal or internal-external designs and very few using independent external cohorts. The attenuation in AUC observed across validation strata illustrates the risk of optimism bias when models are not tested on independent datasets. This pattern mirrors the broader AI and radiomics literature and emphasizes the need for robust, multicenter external validation and adherence to reporting standards such as TRIPOD-AI.

Outcome definitions also varied, with most studies using RECIST 1.1, some using histopathologic tumor regression, and one using survival-based endpoints. RECIST-based models dominated the pooled estimates but also carried the intrinsic limitations of size-based criteria, which may not fully reflect biologic response, especially in the setting of targeted therapies. Histopathology-based models showed numerically higher AUCs, suggesting that biologically grounded endpoints may yield more meaningful predictive signatures; however, they were fewer in number and often limited to resectable cohorts. Survival-based

**Figure 6.** Forest plot showing the pooled area under the curve (reported as mean difference) for radiomics-based artificial intelligence models in the chemotherapy-only subgroup. The pooled mean difference was 0.85 (95% CI 0.76-0.95; $I^2$ = 0%).

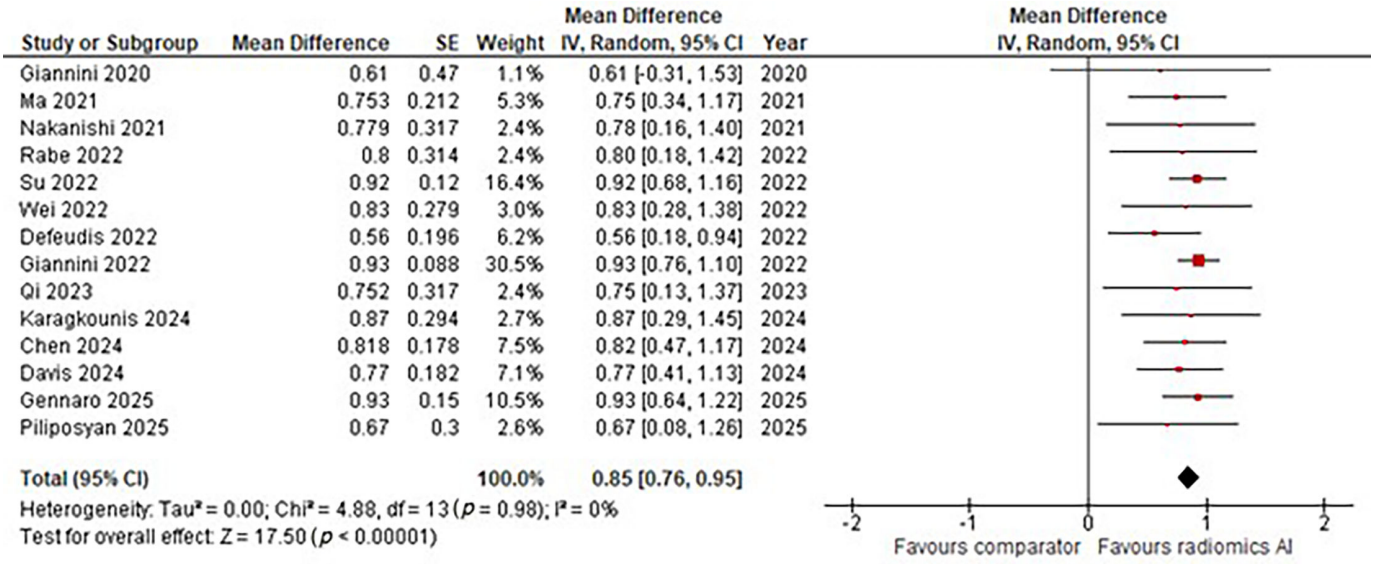
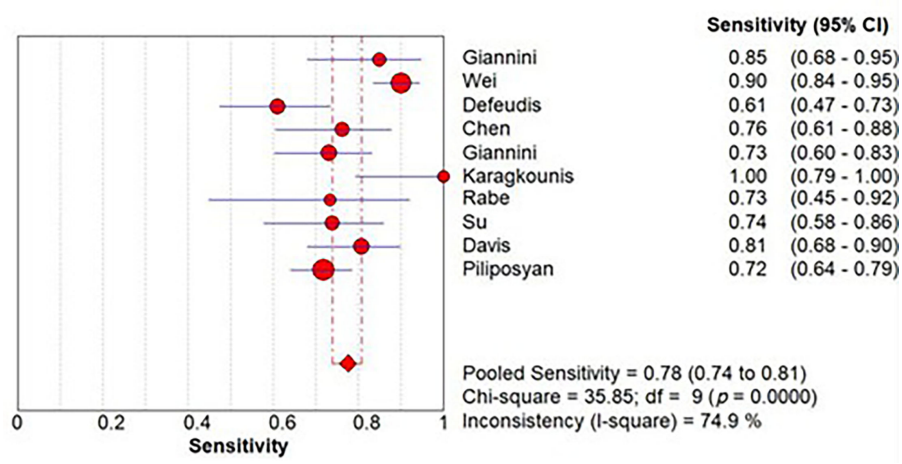



**Figure 7.** Forest plot showing the pooled sensitivity of radiomics-based artificial intelligence models in the chemotherapy-only subgroup. The pooled sensitivity was 0.78 (95% CI 0.74-0.81), with moderate heterogeneity ($I^2$ = 74.9%).




models are conceptually attractive but more challenging to standardize and were underrepresented in the current literature. Collectively, this heterogeneity in outcome definitions complicates direct comparison between studies. A shift toward histopathologic endpoints, survival metrics, or composite response criteria incorporating functional imaging (eg, DWI, PET-CT) would enable more clinically meaningful prediction models.

While the pooled sensitivity of 0.86 across all studies, and 0.88 within the biologics subgroup, is encouraging, clinical applicability depends not only on the ability to detect responders but also to accurately identify nonresponders. The observed specificity of approximately 0.65 suggests that a considerable proportion of patients may be misclassified as likely responders despite potentially limited benefit. This could lead to unnecessary exposure to chemotherapy-related toxicities or delay in transitioning to alternative therapies. One contributing factor may be the difficulty in distinguishing stable disease from partial response using imaging alone. These models, though highly sensitive to subtle image-derived changes, may lack access to molecular or serum biomarkers that could refine response prediction. Therefore, future models must strike a balance between sensitivity and specificity to optimize treatment planning. Integrating genomic data or circulating tumor markers such as CEA, KRAS/NRAS status, or ctDNA levels may improve the discriminatory power of AI models and reduce false-positive rates.[40,41]

Heterogeneity in imaging protocols, segmentation approaches, feature extraction pipelines, and model architectures further complicates generalizability. Most studies were single-center, used institution-specific imaging parameters, and applied heterogeneous feature engineering and selection strategies. This variability likely contributed to the statistical heterogeneity observed in pooled estimates and limits the immediate transferability of any single model into routine practice. Standardization efforts, such as harmonized acquisition protocols, robust feature harmonization pipelines, and transparent reporting of preprocessing steps, will be critical for translating radiomics models beyond the research setting.

For clinical integration, several practical barriers must be addressed. These include the computational requirements


https://doi.org/10.37549/AR-D-25-0146

**Figure 8.** Forest plot showing the pooled specificity of radiomics-based artificial intelligence models in the chemotherapy-only subgroup. The pooled specificity was 0.66 (95% CI 0.62-0.70), with high heterogeneity ($I^2$ = 89.6%).

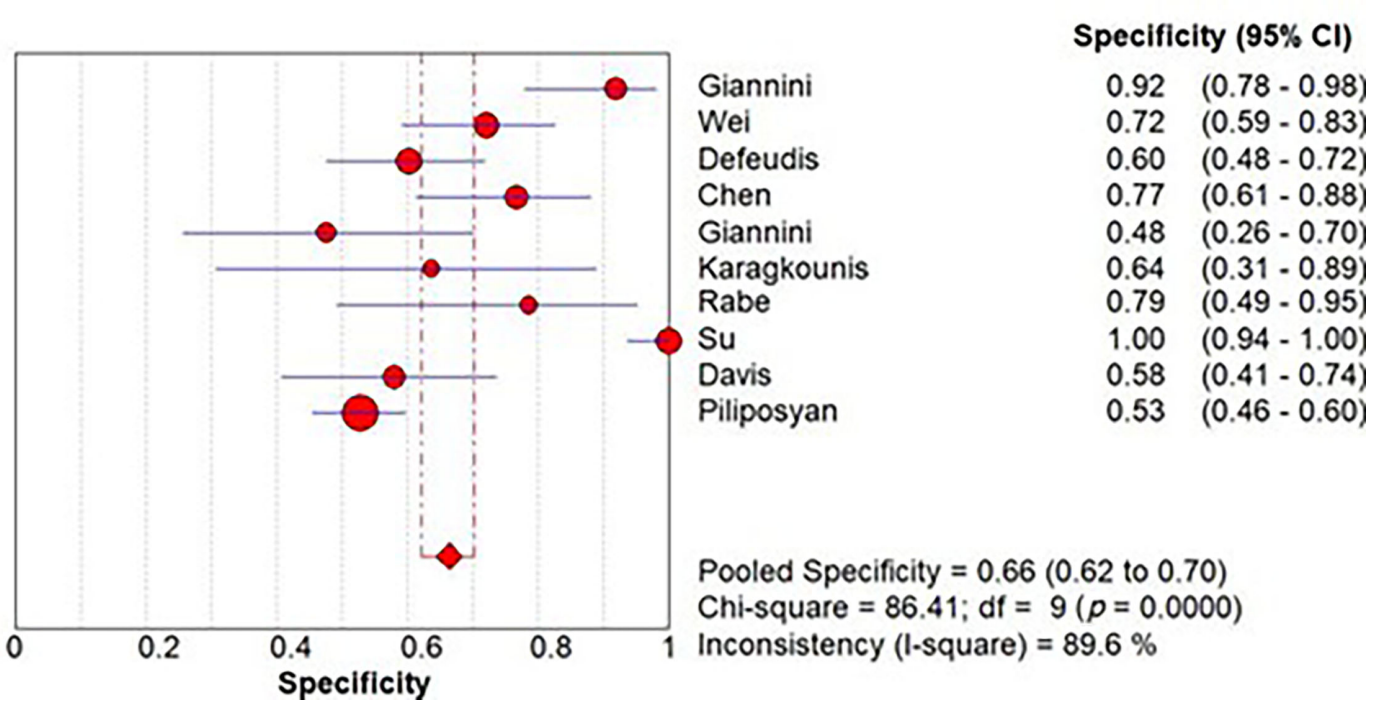
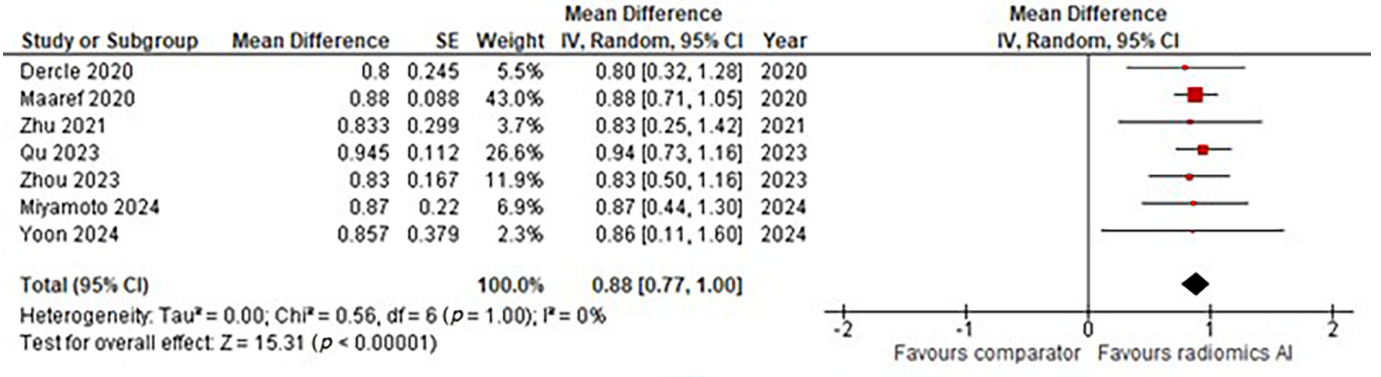



**Figure 9.** Forest plot showing the pooled area under the curve (expressed as mean difference) for radiomics-based artificial intelligence models in the chemotherapy plus targeted therapy subgroup. The pooled mean difference was 0.88 (95% CI 0.77-1.00).




of DL models, lack of interpretability ("black-box" nature), and the need for radiomics software validation.[42] Explainable AI techniques, such as SHAP values, Grad-CAM, or attention heatmaps, may help bridge the gap between algorithm output and clinical trust.[43] Additionally, institutions with limited access to advanced imaging modalities or GPU infrastructure may face challenges in model deployment. Cost-effectiveness studies and integration into clinical decision support systems are needed to evaluate real-world feasibility.

Despite these limitations, the present analysis confirms that AI and radiomics have strong potential to augment traditional imaging criteria in CRLM, particularly for early response assessment and treatment stratification. The consistent signal of high AUCs across diverse models and settings suggests that quantitative imaging biomarkers can capture relevant aspects of tumor biology and therapeutic effect that are not readily visible to the human eye. For institutions seeking to adopt such tools, the current evidence supports prioritizing models that incorporate longitudinal or multiphasic imaging, leverage DL or hybrid architectures, and have undergone at least some form of external validation.

Future research should focus on prospective, multicenter trials that incorporate harmonized imaging protocols, standardized radiomics workflows, and predefined validation plans. Studies that link radiomics signatures with molecular and genomic data, as well as with circulating biomarkers, may enable more comprehensive and biologically

**Figure 10.** Forest plot showing the pooled sensitivity of radiomics-based artificial intelligence models in the chemotherapy plus targeted therapy subgroup. The pooled sensitivity was 0.96 (95% CI 0.93-0.98), with moderate heterogeneity ($I^2$ = 69.9%).

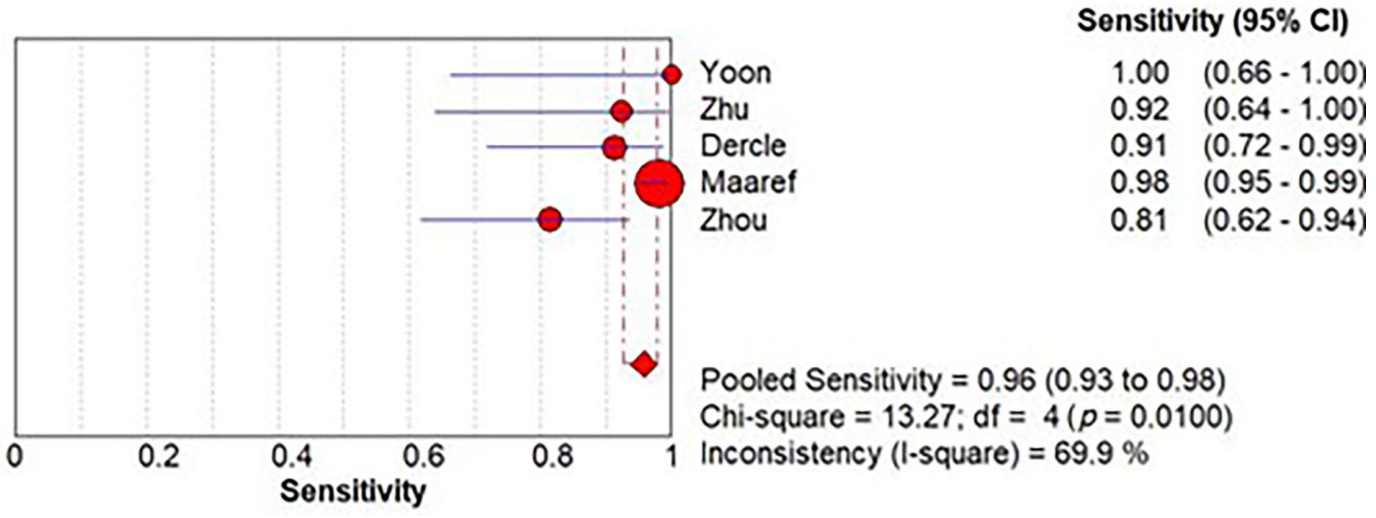
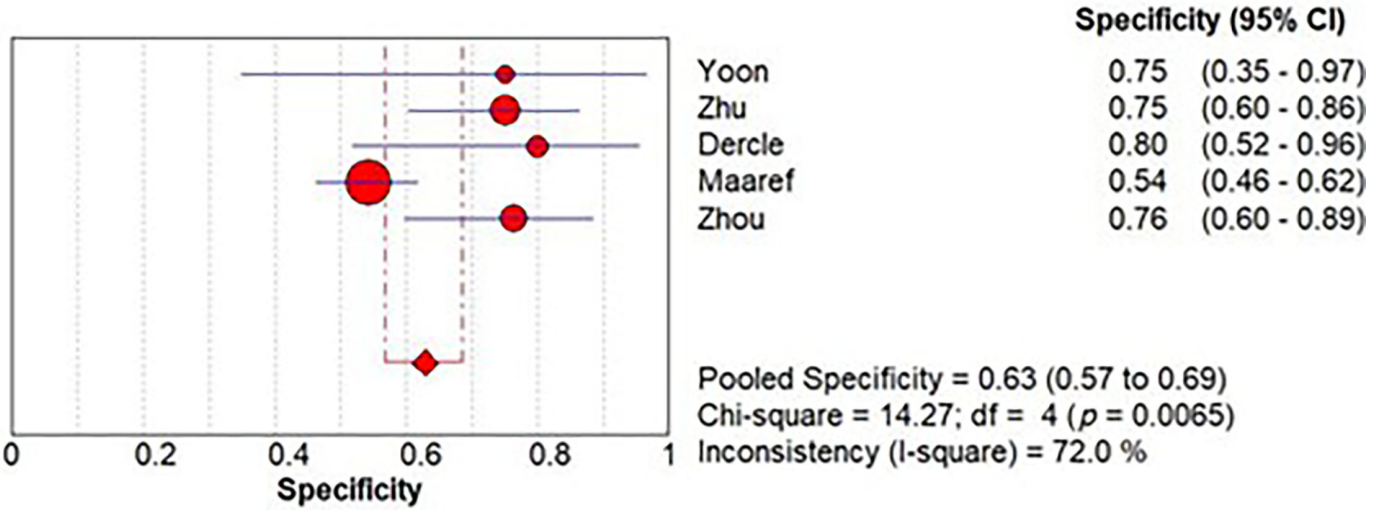



**Figure 11.** Forest plot showing the pooled specificity of radiomics-based artificial intelligence models in the chemotherapy plus targeted therapy subgroup. The pooled specificity was 0.63 (95% CI 0.57-0.69) with substantial heterogeneity ($I^2$ = 72.0%).




grounded predictive models. In parallel, development and testing of explainable AI frameworks and formal cost-effectiveness analyses will be necessary to clarify how these tools can be safely and efficiently integrated into real-world oncologic decision-making.

## Conclusion

This meta-analysis affirms the promising diagnostic value of radiomics-based AI models in predicting chemotherapy response in CRLM. These models perform particularly well in patients receiving chemotherapy in combination with targeted therapy, reflecting the value of integrating biologically rich treatment data into AI frameworks. With continued refinement and validation, these technologies have the potential to revolutionize how oncologic imaging is interpreted and applied in treatment decision-making.

## References

1) Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases - a population-based study on incidence, management and survival. *BMC Cancer*. 2018;18(1):78. doi:10.1186/s12885-017-3925-x

2) Martin J, Petrillo A, Smyth EC. Colorectal liver metastases: current management and future perspectives. *World J Clin Oncol*. 2020;11(10):761-808. doi: 10.5306/wjco.v11.i10.761

3) Valderrama-Treviño AI, Barrera-Mera B, Ceballos-Villalva JC, Montalvo-Javé EE. Hepatic metastasis from colorectal cancer. *Euroasian J Hepatogastroenterol*. 2017;7(2):166-175. doi:10.5005/jp-journals-10018-1241


https://doi.org/10.37549/AR-D-25-0146

4) Mitchell D, Puckett Y, Nguyen QN. Literature review of current management of colorectal liver metastasis. *Cureus*. 2019;11(1):e3940. doi:10.7759/cureus.3940

5) Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026

6) van Klaveren RJ, Aerts JGJV, de Bruin H, et al. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. *Lung Cancer*. 2004;43(1):63-69. doi:10.1016/s0169-5002(03)00292-7

7) Ko CC, Yeh LR, Kuo YT, Chen JH. Imaging biomarkers for evaluating tumor response: RECIST and beyond. *Biomark Res*. 2021;9(1):52. doi:10.1186/s40364-021-00306-8

8) Park JE, Kim HS. Radiomics as a quantitative imaging biomarker: practical considerations and the current standpoint in neuro-oncologic studies. *Nucl Med Mol Imaging*. 2018;52(2):99-108. doi:10.1007/s13139-017-0512-7

9) Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30(9):1234-1248. doi:10.1016/j.mri.2012.06.010

10) O'Connor JP. Rethinking the role of clinical imaging. *Elife*. 2017;6:e30563. doi:10.7554/eLife.30563

11) van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339

12) Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5

13) Chen Y, Lu T, Zhang Y, et al. Baseline hepatobiliary MRI for predicting chemotherapeutic response and prognosis in initially unresectable colorectal cancer liver metastases. *Abdom Radiol*. 2024;49(8):2585-2594. doi:10.1007/s00261-024-04492-5

14) Defeudis A, Cefaloni L, Giannetto G, et al. Comparison of radiomics approaches to predict resistance to 1st line chemotherapy in liver metastatic colorectal cancer. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021;2021:3305-3308. doi:10.1109/EMBC46164.2021.9630316

15) Giannini V, Defeudis A, Rosati S, et al. An innovative radiomics approach to predict response to chemotherapy of liver metastases based on CT images. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in Conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society*. Montreal, QC, Canada. 2020. doi:10.1109/EMBC44109.2020.9176627

16) Giannini V, Pusceddu L, Defeudis A, et al. Delta-radiomics predicts response to first-line oxaliplatin-based chemotherapy in colorectal cancer patients with liver metastases. *Cancers*. 2022;14(1):241. doi:10.3390/cancers14010241

17) Karagkounis G, Horvat N, Danilova S, et al. Computed tomography-based radiomics with machine learning outperforms radiologist assessment in estimating colorectal liver metastases pathologic response after chemotherapy. *Ann Surg Oncol*. 2024;31(13):9196-9204. doi:10.1245/s10434-024-15373-y

18) Qu H, Zhai H, Zhang S, et al. Dynamic radiomics for predicting the efficacy of antiangiogenic therapy in colorectal liver metastases. *Front Oncol*. 2023;13:992096. doi:10.3389/fonc.2023.992096

19) Rabe E, Cioni D, Baglietto L, et al. Can the computed tomography texture analysis of colorectal liver metastases predict the response to first-line cytotoxic chemotherapy?. *World J Hepatol*. 2022;14(1):244-259. doi:10.4254/wjh.v14.i1.244

20) Su X, Zhang H, Wang Y. A predictive model for early therapeutic efficacy of colorectal liver metastases using multimodal MRI data. *J Xray Sci Technol*. 2023;31(2):357-372. doi:10.3233/XST-221317

21) Wei J, Cheng J, Gu D, et al. Deep learning-based radiomics predicts response to chemotherapy in colorectal liver metastases. *Med Phys*. 2021;48(1):513-522. doi:10.1002/mp.14563

22) Miyamoto Y, Nakaura T, Ohuchi M, et al. Radiomics-based machine learning approach to predict chemotherapy responses in colorectal liver metastases. *J Anus Rectum Colon*. 2025;9(1):117-126. doi:10.23922/jarc.2024-077

23) Yoon S, Kim YJ, Jeon JS, Ahn SJ, Choi SJ. Radiomics and machine learning analysis of liver magnetic resonance imaging for prediction and early detection of tumor response in colorectal liver metastases. *Korean J Clin Oncol*. 2024;20(1):27-35. doi:10.14216/kjco.24005

24) Zhu H-B, Xu D, Ye M, et al. Deep learning-assisted magnetic resonance imaging prediction of tumor response to chemotherapy in patients with colorectal liver metastases. *Int J Cancer*. 2021;148(7):1717-1730. doi:10.1002/ijc.33427

25) Davis JMK, Niazi MKK, Ricker AB, et al. Predicting response to neoadjuvant chemotherapy for colorectal liver metastasis using deep learning on prechemotherapy cross-sectional imaging. *J Surg Oncol*. 2024;130(1):93-101. doi:10.1002/jso.27673

26) Gennaro N, Soliman M, Borhani AA, et al. Delta radiomics and tumor size: a new predictive radiomics model for chemotherapy response in liver metastases from breast and colorectal cancer. *Tomography*. 2025;11(3):20. doi:10.3390/tomography11030020

27) Piliposyan M. Treatment response prediction for colorectal liver metastases using pre-treatment computed tomography scans (master's thesis, queen's university). QSpace. 2025. https://qspace.library.queensu.ca/items/420de053-27ff-412b-aa7f-71575cb83665

28) Zhou S, Sun D, Mao W, et al. Deep radiomics-based fusion model for prediction of bevacizumab treatment response and outcome in patients with colorectal cancer liver metastases: a multicentre cohort study. *EClinicalMedicine*. 2023;65:102271. doi:10.1016/j.eclinm.2023.102271

29) Dercle L, Lu L, Schwartz LH, et al. Radiomics response signature for identification of metastatic colorectal cancer sensitive to therapies targeting EGFR pathway. *J Natl Cancer Inst*. 2020;112(9):902-912. doi:10.1093/jnci/djaa017

30) Maaref A, Romero FP, Montagnon E, et al. Predicting the response to FOLFOX-based chemotherapy regimen from untreated liver metastases on baseline CT: a deep neural network approach. *J Digit Imaging*. 2020;33(4):937-945. doi:10.1007/s10278-020-00332-2

31) Ma YQ, Wen Y, Liang H, Zhong JG, Pang PP. Magnetic resonance imaging-radiomics evaluation of response to chemotherapy for synchronous liver metastasis of colorectal cancer. *World J Gastroenterol*. 2021;27(38):6465-6475. doi:10.3748/wjg.v27.i38.6465

32) Nakanishi R, Oki E, Hasuda H, et al. Radiomics texture analysis for the identification of colorectal liver metastases sensitive to first-line oxaliplatin-based chemotherapy. *Ann Surg Oncol*. 2021;28(6):2975-2985. doi:10.1245/s10434-020-09581-5

33) Qi W, Yang J, Zheng L, et al. CT-based radiomics for the identification of colorectal cancer liver metastases sensitive to first-line irinotecan-based chemotherapy. *Medical Physics*. 2023;50(5):2705-2714. doi:10.1002/mp.16325

34) Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378

35) Granata V, Fusco R, Catalano O, et al. Early assessment of colorectal cancer patients with liver metastases treated with antiangiogenic drugs: the role of intravoxel incoherent motion in diffusion-weighted imaging. *PLoS One*. 2015;10(11):e0142876. doi:10.1371/journal.pone.0142876

36) Wang Z-X, Yao Y-C, Mai Z-J, et al. Temporal change in treatment patterns of metastatic colorectal cancer and its association with patient survival: a retrospective cohort study based on an intelligent big-data platform. *Engineering*. 2021;7(4):526-533. doi:10.1016/j.eng.2020.10.017

37) Zhao T, Wang S, Ouyang C, et al. Artificial intelligence for geoscience: progress, challenges, and perspectives. *Innovation*. 2024;5(5):100691. doi:10.1016/j.xinn.2024.100691

38) Khalifa M, Albadawy M. AI in diagnostic imaging: revolutionising accuracy and efficiency. *Comput Methods Programs Biomed Update*. 2024;5:100146. doi:10.1016/j.cmpbup.2024.100146

39) Russo V, Lallo E, Munnia A, et al. Artificial intelligence predictive models of response to cytotoxic chemotherapy alone or combined to targeted therapy for metastatic colorectal cancer patients: a systematic review and meta-analysis. *Cancers*. 2022;14(16):4012. doi:10.3390/cancers14164012

40) Franczak C, Filhine-Tressarieu P, Broséus J, et al. Clinical interest of circulating tumor DNA in oncology. *Arch Med Res*. 2018;49(5):297-305. doi:10.1016/j.arcmed.2018.10.007

41) Duffy MJ, Crown J. Use of circulating tumour DNA (ctDNA) for measurement of therapy predictive biomarkers in patients with cancer. *J Pers Med*. 2022;12(1):99. doi:10.3390/jpm12010099

42) Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. *Inf Fusion*. 2022;77:29-52. doi:10.1016/j.inffus.2021.07.016

43) Hassan SU, Abdulkadir SJ, Zahid MSM, Al-Selwi SM. Local interpretable model-agnostic explanation approach for medical imaging analysis: a systematic literature review. *Comput Biol Med*. 2025;185:109569. doi:10.1016/j.compbiomed.2024.109569