Volume 24, No 3 | International Journal of Radiation Research, July 2026

MRI-based radiomics models for pre-operative prediction of microsatellite instability in rectal cancer: A systematic review and meta-analysis

B. Wang, J. Gu, B. Wu*

Department of Radiology, West China Hospital, Sichuan University, Chengdu 610041, China

► ***Review article***

Corresponding author:
Bing Wu, M.D.,
E-mail: *bingwu1453@163.com*

Received: December 2025
Final revised: April 2026
Accepted: May 2026

Int. J. Radiat. Res., July 2026; 24(3): 885-893

DOI: 10.61186/ijrr.24.3.39

Keywords: *Rectal neoplasms, microsatellite instability, radiomics, magnetic resonance imaging, machine learning.*

ABSTRACT

Background: This study aimed to comprehensively assess the diagnostic capability of radiomics models derived from magnetic resonance imaging in determining microsatellite instability (MSI) status among patients with rectal cancer, while exploring variables that may affect predictive accuracy. ***Materials and Methods:*** We performed a comprehensive literature search across PubMed, EMBASE, Cochrane Library, and Web of Science databases through August 2025, targeting studies employing MRI-derived radiomics for MSI status prediction in rectal malignancies. Summary estimates for sensitivity, specificity, and the area under the summary receiver operating characteristic curve (SROC AUC) were derived through random-effects meta-analysis. Heterogeneity sources were investigated via subgroup stratification. ***Results:*** Thirteen eligible investigations encompassing 3,292 individuals with rectal cancer were incorporated. For validation cohorts, pooled estimates demonstrated a sensitivity of 0.88, specificity of 0.82, and SROC AUC of 0.91. According to Fagan nomogram interpretation, positive radiomics findings elevated MSI post-test probability from an initial 18% to 52%, whereas negative findings lowered it to 3%. Meta-regression revealed classifier selection as the predominant determinant of performance variability. Robustness was substantiated through sensitivity testing, with Deeks' funnel plot asymmetry test showing no substantial publication bias (P=0.93). ***Conclusion:*** Radiomics models based on MRI exhibit robust diagnostic performance for preoperative MSI status prediction in rectal cancer, demonstrating promise as non-invasive screening instruments to support treatment decision-making. Nevertheless, considering the methodological constraints and inter-study variability observed, rigorously designed multicenter prospective investigations are imperative to establish their clinical utility and economic feasibility.

INTRODUCTION

Rectal cancer (RC) represents a major global malignancy and continues to be among the principal contributors to cancer-associated morbidity and mortality within the gastrointestinal tract [1]. Notwithstanding substantial progress in conventional therapeutic approaches, including surgery, radiotherapy, and chemotherapy, clinical outcomes among patients exhibit marked heterogeneity. This heterogeneity is largely attributed to the diverse biological characteristics of tumors [2]. Microsatellite instability (MSI), a key molecular marker, arises from impaired DNA mismatch repair (MMR) function. High-level MSI (MSI-H) is present in approximately 14-16% of colorectal cancers [3]. Studies have demonstrated that MSI-H rectal cancer patients tend to have poor responses to conventional 5-fluorouracil-based adjuvant chemotherapy but demonstrate remarkable sensitivity to immune checkpoint inhibitor therapies [4-6]. Therefore, the National Comprehensive Cancer Network (NCCN) guidelines advocate for routine MSI testing across all RC patients [7].

MSI status detection currently employs predominantly invasive approaches, specifically immunohistochemistry (IHC) or polymerase chain reaction (PCR) testing of surgical pathology specimens [8]. The requirement for tumor tissue, however, poses significant challenges for preoperative assessment. Moreover, intratumoral heterogeneity may lead to biopsy specimens that fail to capture the complete molecular profile of the tumor [9, 10]. Therefore, the development of non-invasive, reproducible methods for predicting MSI status preoperatively holds significant clinical potential.

Radiomics is an evolving quantitative imaging field that utilizes high-throughput methods to extract large-scale imaging features and develop predictive models via machine learning (ML) algorithms [11]. Unlike conventional radiological evaluation, radiomics is capable of detecting intratumoral heterogeneity that remains invisible to visual inspection, thus providing enhanced insights into tumor biological characteristics [12]. MRI, serving as

[DOI: 10.61186/ijrr.24.3.39]

the reference imaging technique for preoperative rectal cancer evaluation, offers superior soft tissue contrast and multiparametric imaging capacity, rendering it an optimal platform for radiomic feature extraction [13].

Recent studies on MRI-based radiomics for RC have focused on predicting treatment efficacy [14], tumor stage [15], metastasis [16, 17], and patient survival [18]. Several studies have also investigated MRI-based radiomics for MSI prediction. Nevertheless, substantial heterogeneity exists in terms of sample size, modeling approaches, validation strategies, and imaging protocols, which complicates cross-study comparisons and clinical application.

This meta-analysis sought to rigorously assess how well MRI-based radiomic models predict MSI status in rectal cancer. Additionally, we aimed to determine performance-influencing factors and provide evidence-based recommendations for clinical application. To our knowledge, this is the most recent and comprehensive meta-analysis comparing machine learning approaches specifically for MRI radiomics-based MSI prediction in rectal cancer, offering new insights into classifier-specific diagnostic performance.

MATERALS AND METHODS

Methodology and registration

This systematic review and meta-analysis were performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [19]. The protocol for this review was registered a priori with the International Prospective Register of Systematic Reviews (PROSPERO) under the identifier CRD420251137444.

Search strategy

A comprehensive literature search was performed in August 2025 across four major databases: was conducted in August 2025 across four principal databases: PubMed, EMBASE, the Cochrane Library, and Web of Science. Detailed search strategies are presented in Supplementary Table S1-4. After deduplication, remaining records underwent title and abstract screening, followed by full-text retrieval and detailed evaluation of potentially eligible studies to confirm suitability for qualitative and quantitative synthesis. We also manually searched reference lists of included studies and relevant reviews [20, 21] to identify additional relevant articles.

Inclusion and exclusion criteria

The inclusion criteria are as follows:

P (Population): Patients with pathologically confirmed rectal cancer.

D (Diagnostic method): Pretreatment MRI-based radiomics in MSI status prediction.

C (Comparison): Histopathological testing as the gold standard reference for model assessment.

O (Outcome): Model diagnostic performance measured by AUC, sensitivity, and specificity.

We excluded: (1) Patients with any pre-imaging treatment. (2) Studies addressing only segmentation or feature extraction methods. (3) Non-MRI imaging studies (CT, ultrasound, PET/CT). (4) Reviews, animal studies, abstracts, case reports, editorials, or letters. (5) Hybrid model studies without independent radiomics model performance data.

Study selection

All retrieved literature was managed centrally using NoteExpress software. Both automatic and manual deduplication processes were applied to eliminate duplicate records. Subsequently, two independent reviewers (Wang, B. and Gu, J.) evaluated all titles and abstracts using predefined criteria. Full-text articles of potentially eligible studies were then retrieved and assessed in detail, with exclusions and reasons recorded. Furthermore, we conducted a manual review of references and cited literature within the included articles to identify additional eligible studies. Any discrepancies were reconciled through consensus discussion or adjudication by a third investigator (Wu, B.).

Data extraction

Data extraction was performed by two independent reviewers (Wang, B. and Gu, J.) using a predefined data collection form. T The retrieved data comprised fundamental study attributes including author details, publication year, geographic location, methodological design, cohort size, and imaging protocol. Additionally, technical specifications of the prediction models were recorded, including image preprocessing strategies, segmentation techniques and software, feature selection methods, types of radiomics parameters, algorithms employed, and model validation strategies. Model diagnostic performance indicators were also extracted, encompassing AUC, sensitivity, specificity, and contingency table elements including true positives, false positives, false negatives, and true negatives.

Data were collected separately for both training and validation cohorts. Among studies reporting two or more prediction models, only the best-performing model, as indicated by the AUC value, was included. For studies evaluating hybrid models, only the pure radiomics model was considered. Comprehensive data collection was ensured by a thorough review of accompanying supplementary files. Any discrepancies in data extraction were reconciled through discussion or by consulting a third investigator (Wu, B.).

[DOI: 10.61186/ijrr.24.3.39]

Quality assessment

Two reviewers (Wang, B. and Gu, J.) independently evaluated the methodological quality of studies using the Radiomics Quality Score (RQS) checklist [22] and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [23]. The RQS checklist assesses the methodological rigor of the radiomics workflow across 16 components organized into six key domains, with scores ranging from -8 to 36. The QUADAS-2 framework assesses study quality via four primary dimensions. Risk of bias for each dimension was rated as "high," "low," or "unclear". Discrepancies during the quality assessment were managed through discussion or consensus with a third investigator (Wu, B.).

Statistical analysis

Statistical analyses were performed using STATA 14.0 software. Summary receiver operating characteristic (SROC) were derived from 2×2 diagnostic contingency tables, using AUC as the diagnostic accuracy indicator. The I^2 statistic assessed statistical heterogeneity among studies. Forest plots illustrated pooled sensitivity and specificity. Sensitivity analyses systematically removed individual studies to examine their effect on pooled outcomes. Deeks' test evaluated publication bias. Subgroup analyses and meta-regression identified heterogeneity sources. Furthermore, Fagan's nomogram quantified clinical applicability by contrasting pre-test and post-test probabilities. Random-effects modeling was applied for high heterogeneity, with $P < 0.05$ indicating significance.

RESULTS

Study selection

A total of 136 citations were retrieved from the database search. After eliminating duplicates, 87 distinct records underwent title and abstract screening, which yielded 31 articles eligible for full-text evaluation. Finally, 13 studies satisfied the inclusion criteria and were incorporated into this systematic review. Of these, 12 investigations reported adequate data to generate 2 × 2 contingency tables. We included two separate publications verified to be drawn from the same study population; data extraction was performed from both reports, though the total sample size was counted once to prevent double counting of participants. A comprehensive overview of the study selection procedure is provided in the PRISMA flowchart (figure 1).

Study characteristics

Overall, 13 investigations were incorporated into this review, comprising 3,292 rectal cancer patients [24-36]. All investigations were conducted in China, with publication dates spanning 2021 to 2025. All studies were retrospective. Most studies (10 studies) were single-center investigations, while 3 studies were multicenter. The sample sizes varied from 90 to 491 patients, with the prevalence of microsatellite instability (MSI) among the subjects ranging from 10.4% to 49.6%. Each study provided a thorough description of their model validation methodologies. Internal validation was applied in nine studies, with seven using random sampling and two relying on internal cross-validation. Three studies utilized external validation, while one study implemented both internal random sampling and external validation. With regards to image segmentation methods, the majority (12 studies) performed 3D region of interest (ROI) segmentation, whereas one study relied on 2D segmentation. PyRadiomics served as the primary segmentation software in 5 of the 13 studies. Additional technical considerations included the application of the Intraclass Correlation Coefficient (ICC) method for feature stability assessment in 9 studies, establishing specific thresholds to ensure the reliability and consistency of features. Logistic Regression (LR) was the favored modeling approach in five studies. A complete overview of study characteristics is provided in tables 1 and 2.

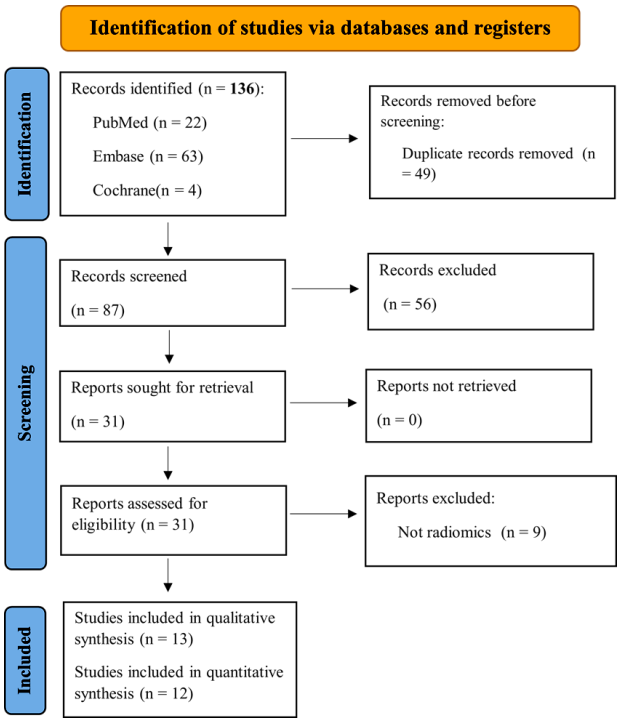

Figure 1. PRISMA flow diagram of study selection. The flowchart illustrates the systematic literature search, screening, and selection steps following the Preferred Reporting Items for PRISMA guidelines. Ultimately, 13 eligible studies were included in the meta-analysis.

Quality assessment

A summary of RQS evaluation results is presented in figure 2, with detailed ratings for each included study available in supplementary table S5. According

[DOI: 10.61186/ijrr.24.3.39]

to the RQS checklist, the studies scored between 14 and 21, yielding an overall mean score of 15.38 out of a possible 36, which corresponds to 42.7% of the maximum score achievable. The principal factors contributing to these relatively low scores were the lack of a phantom model to assess inter-scanner variability, the absence of multi-timepoint imaging data for assessing temporal consistency, and weak associations observed between radiomic features and biological endpoints. Moreover, very few studies reported on model calibration, and there was a lack of cost-effectiveness analysis. The retrospective nature of the majority of studies also contributed to the diminished quality scores observed.

Table 1. Characteristics of the included studies.

Author	Year	Country	Study design	Study setting	Sample	MSI prevalence	MRI sequence	Segmentation	Reference
Zheng [36]	2025	China	Retrospective	Single-center	120	0.250	T1CE	Semi-auto	PCR and IHC
Xing, X [35]	2024	China	Retrospective	Single-center	308	0.166	DWI, T2, T1, T1CE	Manually	IHC
Cai, Z [34]	2024	China	Retrospective	Muti-center	475	0.112	DWI, T2, T1, T1CE	Manually	IHC
Li, Z [33]	2021	China	Retrospective	Single-center	90	0.34	T2, ADC	Manually	IHC
Li [32]	2023	China	Retrospective	Muti-center	199	0.342	DWI, T2, T1, T1CE	Manually	IHC
Zhang, Y [31]	2023	China	Retrospective	Single-center	383	0.146	DWI, T2, T1, T1CE	Manually	IHC
Jin, J [29]	2023	China	Retrospective	Single-center	117	0.496	DWI, T2, T1	Manually	IHC
Xiang, S [30]	2023	China	Retrospective	Single-center	175	0.131	DWI, T2	Manually	IHC
Jing [28]	2022	China	Retrospective	Single-center	176	0.176	DWI, T2, T1CE	Manually	IHC
Zhang, W [27]	2021	China	Retrospective	Single-center	491	0.104	T2	Manually	IHC
Wang, Y [26]	2025	China	Retrospective	Single-center	291	0.168	DWI, T2	Manually	IHC
Yao, X [24]	2025	China	Retrospective	Muti-center	467	0.176	DWI, T2, T1CE	Manually	IHC
Zhang, W [25]	2021	China	Retrospective	Single-center	491	0.104	T2	Manually	IHC

MSI, Microsatellite Instability; MRI, Magnetic Resonance Imaging; DWI, Diffusion Weighted Imaging; PCR, Polymerase Chain Reaction; IHC, Immuno-histochemistry.

Table 2. Characteristics of the included studies (continued from table 1).

Author	Validation	Algorithm	Software	ROI	ICC	Imaging features	Classifier
Zheng [36]	10-fold cross-validation	ML	PyRadiomics + 3D Slicer	3D	>0.75	first-order, textural features, shape features, dynamic enhancement parameters	LASSO, RF
Xing, X [35]	5-fold cross-validation	ML	ITK-SNAP	3D	>0.8	first-order, texture features (GLCM, GLSZM, GLRLM, GLDM, NGTDM), Los features, wavelet	LASSO/GBDT, SVM
Cai, Z [34]	EV	ML	3D Slicer	3D	>0.75	first-order, textural features, shape features	LR
Li, Z [33]	random sampling	ML	AK software	3D	>0.75	first-order, texture features (GLCM, GLRLM, Haralick), shape features	LR
Li [32]	EV	ML	PyRadiomics	3D	>0.75	first-order, textural features (GLCM, GLSZM, GLRLM, GLDM), shape features	RF
Zhang, Y [31]	random sampling	ML	AK software	3D	>0.80	first-order, textural features (GLCM, GLSZM, GLRLM, GLDM, NGTDM), shape features, wavelet, LoG	LR
Jin, J [29]	random sampling	ML	3D Slicer	3D	NO	first-order, textural features (GLCM, GLSZM, GLRLM, GLDM, NGTDM), shape features, wavelet	mRMR/LASSO, LR
Xiang, S [30]	random sampling	ML	PyRadiomics	3D	>0.75	first-order, textural features (GLCM, GLRLM, NGTDM), shape features, LoG	LASSO, SVM
Jing [28]	EV	ML	Radcloud radiomics platform	3D	>0.8	first-order, shape features (GLCM, GLRLM, GLDM), higher-order statistics (logarithm, gradient, square and wavelet)	LASSO, SVM
Zhang, W [27]	random sampling	ML	PyRadiomics	3D	>0.75	first-order, texture-based (GLCM, GLSZM, GLRLM, GLDM, NGTDM), shape features	LASSO, XGBoost
Wang, Y [26]	random sampling	ML	PyRadiomics	3D	NO	first-order, textural features, shape features	mRMR/LASSO, LR
Yao, X [24]	EV and random sampling	DL	Python 3.9	2D	NO	high dimension features	mRMR/LASSO, CNN
Zhang, W [25]	random sampling	DL	ITK-SNAP	3D	NO	high dimension features	CNN

ROI, Region of Interest; ICC, Intraclass Correlation Coefficient; EV, External Validation; ML, Machine Learning; DL, Deep Learning; GLCM, Gray Level Co-occurrence Matrix; GLDM, Gray Level Dependence Matrix; GLRLM, Gray Level Run Length Matrix; GLSZM, Gray Level Size Zone Matrix; LoG, Laplacian of Gaussian; NGTDM, Neighborhood Gray Tone Difference Matrix; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; mRMR, maximal relevance and minimal redundancy; RF, Random Forest; SVM, Support Vector Machine; KNN, K-Nearest Neighbor; GBDT, Gradient Boosting Decision Tree; CNN, Convolutional Neural Network.

The QUADAS-2 evaluation demonstrated a generally low risk of bias across included investigations, as shown in figure 3. Some studies received an "unclear" bias risk rating owing to insufficient reporting of methodological procedures, including image acquisition protocols and preprocessing techniques. The other three evaluated domains were classified as presenting "low" bias risk. Regarding applicability concerns, one investigation received a "high" risk rating in patient selection, as its reported MSI prevalence of 49.6% was significantly higher than the prevalence generally observed in the wider population. This discrepancy raised questions regarding the representativeness of its study population.

[DOI: 10.61186/ijrr.24.3.39]

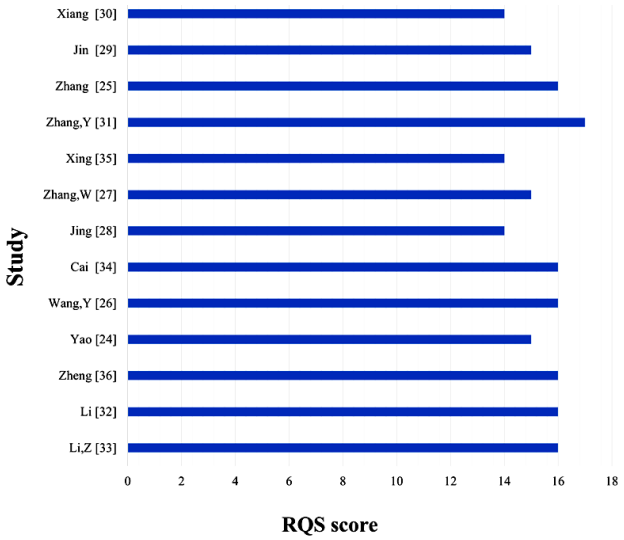

Figure 2. Overview of quality assessment of included investigations using RQS score.

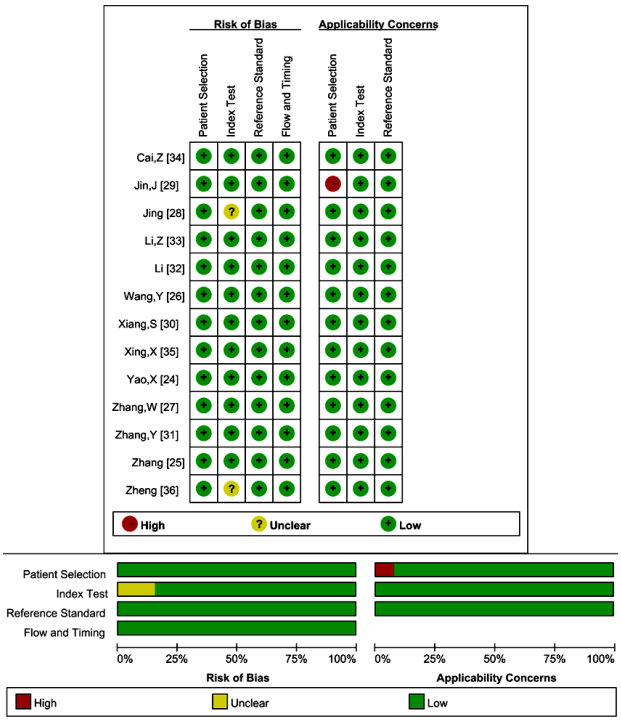

Figure 3. Summary of risk of bias and applicability concerns according to QUADAS-2.

Meta-analysis

Training set

This review included nine radiomics models within training sets. Figure 4 summarizes the meta-analysis results for these models. The pooled sensitivity was found to be 0.85 (95% CI: 0.72-0.92, I^2 = 84.83%), with a pooled specificity of 0.83 (95% CI: 0.75-0.88, I^2 = 90.05%), and a SROC AUC of 0.90 (95% CI: 0.87-0.92). Deeks' funnel test revealed no significant publication bias (P = 0.30) (Supplementary figure S1). Sensitivity analyses, employing a "one-study removal" approach, demonstrated no substantial variations in overall results following the sequential exclusion of each included study (Supplementary figure S2). The prevalence of MSI within our review cohort was 18% (582/3292). With a baseline pre-test probability of 18% and positive likelihood ratio (PLR) of 5, a positive test result raised the post-test probability to 51%. Conversely, a negative result reduced the post-test probability to 4%, corresponding to a negative likelihood ratio (NLR) of 0.19.

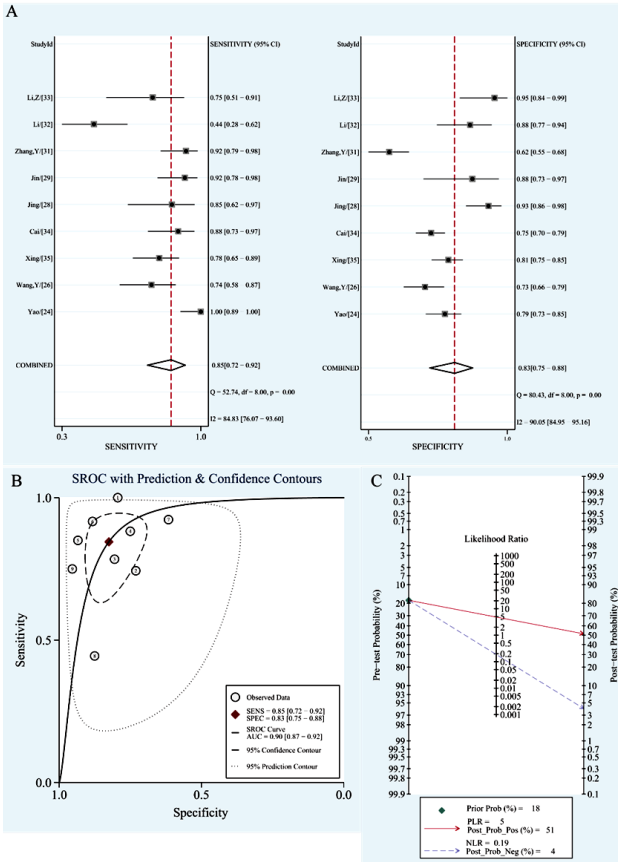

Figure 4. Diagnostic performance of radiomics models for MSI status prediction within training datasets. **(A)** Paired forest plots illustrating pooled sensitivity and specificity. **(B)** SROC curve. **(C)** Fagan nomogram. AUC: area under the curve; PLR: positive likelihood ratio; NLR: negative likelihood ratio; Prob: probability; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristic.

Validation set

This review included 12 radiomics model validation sets. The pooled sensitivity, specificity, and SROC AUC were calculated as 0.88 (95% CI: 0.75-0.95, I^2 = 80.05%), 0.82 (95% CI: 0.74-0.88, I^2 = 85.59%), and 0.91 (95% CI: 0.88-0.93), respectively. Deeks' test showed no publication bias (P = 0.93) (Supplementary Figure S3). Sensitivity analysis indicated no relevant changes in overall metrics when each included study was sequentially removed (Supplementary Figure S4). Given a pre-test probability of 18%, the post-test probability for a positive finding increased to 52%, while the post-test probability for a negative finding decreased to 3% (figure 5).

[DOI: 10.61186/ijrr.24.3.39]

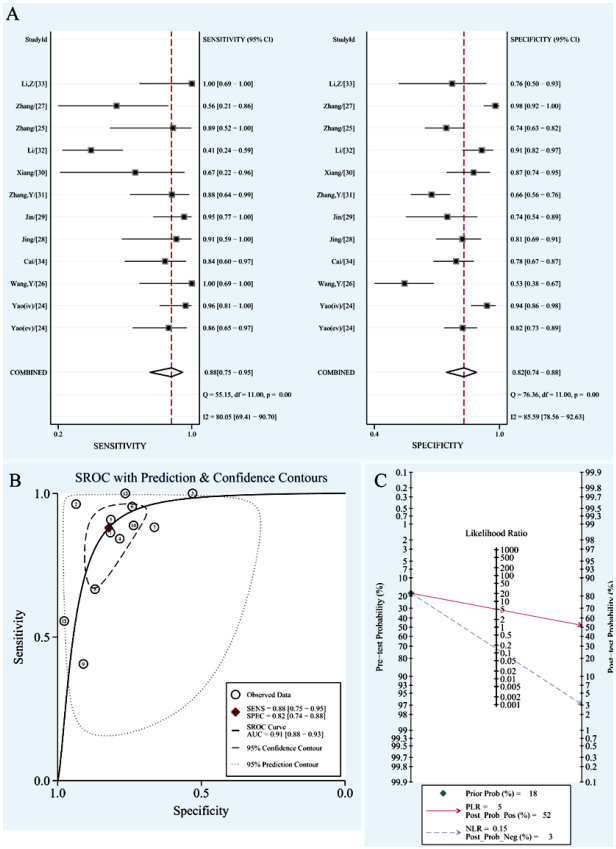

Figure 5. Diagnostic performance of radiomics models for MSI status prediction within validation datasets. **(A)** Paired forest plots illustrating pooled sensitivity and specificity. **(B)** SROC curve. **(C)** Fagan nomogram.

Subgroup analysis and meta-regression

Subgroup analysis outcomes are presented in Table 3. Findings indicate that none of the factors potentially influencing diagnostic accuracy reached statistical significance across in training and validation sets. Specifically, subgroups defined by study setting (single-center vs. multi-center), external validation status, and traditional ML versus DL algorithms exhibited comparable sensitivity and specificity, with no significant performance differences (all P > 0.05). Meta-regression analysis exploring heterogeneity sources (see Table 4) identified classifier type as the primary contributor to between-study variability. The combined test for this factor approached the threshold for statistical significance (P=0.07) and explained up to 63% of heterogeneity (I^2 =63%). Diagnostic performance comparisons among subgroups revealed the highest diagnostic odds ratio (DOR) in the CNN subgroup (DOR=19.965, 95% CI: 8.175 – 48.759), followed by the XGBoost subgroup (DOR=15.893, 95% CI: 5.472 – 46.156). However, pairwise comparisons between classifier subgroups demonstrated no statistically significant differences (overall model P=0.236).

Table 3. Summary of subgroup analysis.

Training Set		No	Sensitivity	P	Specificity	P
Setting	Muti-center	3	0.85 (0.68-1.00)	0.66	0.81 (0.70 -0.93)	0.18
	Single-center	6	0.84 (0.73-0.96)		0.83 (0.75-0.91)	
Algorithm	ML	8	0.81 (0.71-0.90)	0.15	0.83 (0.76-0.90)	0.71
	DL	1	1.00 (1.00-1.00)		0.80 (0.59-1.00)	
Validation Set		No	Sensitivity	P	Specificity	P
Setting	Muti-center	4	0.81 (0.62-1.00)	0.22	0.87 (0.79 -0.96)	0.50
	Single-center	8	0.91 (0.82-1.00)		0.79 (0.69-0.88)	
Algorithm	ML	9	0.85 (0.72-0.97)	0.29	0.81 (0.72-0.90)	0.15
	DL	3	0.93 (0.83-1.00)		0.84 (0.71-0.97)	
External validation	No	8	0.77 (0.56-0.98)	0.31	0.84 (0.73-0.96)	0.10
	Yes	4	0.92 (0.85-1.00)		0.81 (0.72-0.91)	

ML, Machine Learning; DL, Deep Learning

Table 4. Summary of meta-regression analysis

Parameter	Sensitivity	Specificity	p	I^2
Setting	0.95 [0.70-0.99]	0.71 [0.45-0.88]	0.46	0%
External validation	0.97 [0.80-1.00]	0.79 [0.53-0.92]	0.27	25%
Classifier	0.92 [0.82-0.96]	0.82 [0.71-0.90]	0.07	63%

DISCUSSION

MRI-based radiomic models in this meta-analysis exhibited high diagnostic performance, comparing favorably with earlier evidence from more heterogeneous colorectal cancer cohorts. The pooled validation AUC of 0.91 exceeds the 0.83 reported by Wang *et al.* in their systematic review encompassing multiple imaging modalities across colorectal cancer [20], and is consistent with the MRI-specific pooled AUC of 0.90 reported by Capello Ingold *et al.* in their colorectal cancer meta-analysis [21]. The slightly superior performance observed in the present rectal cancer–specific cohort likely reflects the standardized role of pelvic MRI in rectal cancer preoperative workup, which provides superior soft-tissue visualization and multiparametric advantages over CT [13]. MSI-H rectal cancer characteristically demonstrates distinct pathological features (such as mucinous differentiation, indistinct tumor margins, and marked peritumoral lymphocytic infiltration) that lend themselves to objective quantification through MRI-based textural and morphological radiomics signatures [34, 37]. These features offer compelling biological justification for utilizing MRI radiomics exclusively for MSI prediction in rectal cancer. Applying Fagan's nomogram with a baseline MSI-H probability of 18%, aligned with documented

[DOI: 10.61186/ijrr.24.3.39]

population-level prevalence in rectal cancer [3], a positive radiomics finding elevated the post-test probability to 52%, whereas a negative finding lowered it to merely 3%. This negative predictive capacity holds clinical significance in markedly decreasing the necessity for invasive biopsy procedures in low-risk individuals, while channeling molecular diagnostic resources toward higher-likelihood cases.

Substantial between-study heterogeneity was observed across included studies. Meta-regression identified classifier type as the primary contributor to between-study variability, explaining up to 63% of observed heterogeneity (P=0.07). Despite CNN-based models yielding the greatest estimated diagnostic odds ratio (DOR=19.965), pairwise comparisons between classifier subcategories did not achieve significance, indicating that no single algorithm carries a universally superior diagnostic profile in this domain. This observation aligns with a recent meta-analysis in head and neck cancer that documented pooled AUCs of 92% for deep learning and 91% for handcrafted radiomics approaches with no statistically significant difference [38]. This suggests that model complexity doesn't guarantee better diagnostics. Future studies should select classifiers based on dataset characteristics (sample size, class balance, dimensionality, interpretability) rather than assuming complex models are superior [39, 40].

Beyond classifier type, a major source of heterogeneity is insufficient methodological standardization across included studies. Studies employed heterogeneous MRI sequences, feature extraction software platforms, and preprocessing pipelines, limiting cross-study feature comparison. The Image Biomarker Standardization Initiative (IBSI) framework provides consensus-based, mathematically rigorous definitions for radiomic feature computation and preprocessing, and its adoption has been shown to substantially reduce inter-software feature variability in multicenter settings [41]. However, IBSI compliance was inconsistent across the included studies, contributing to the low mean RQS score of 42.7%. Despite a 2024 IBSI update formalizing filter specifications, adoption remains incomplete in recent publications [42]. Future work requires strict IBSI-compliant pipelines and complete preprocessing documentation to achieve inter-center reproducibility needed for regulatory validation and clinical use.

Although deep learning models exhibited a numerically higher diagnostic odds ratio than traditional machine learning methods, the absence of statistical significance warrants a cautious interpretation. Theoretically, deep learning algorithms, particularly CNNs, can autonomously extract hierarchical high-dimensional features from raw images, bypassing manual feature engineering and potentially uncovering novel imaging biomarkers [43, 44]. However, the efficacy of deep learning heavily relies on large, well-balanced datasets. Within the relatively small cohorts, the risk of overfitting substantially diminishes the practical advantages of deep learning over well-regularized traditional machine learning models [26, 45]. Furthermore, the inherent "black-box" nature of deep learning limits its interpretability, posing a significant barrier to clinical adoption, as transparent decision-making is a prerequisite for clinical and regulatory acceptance [46]. Therefore, future investigations must incorporate explainable AI techniques to enhance model transparency and facilitate clinical translation.

Several limitations in this study warrant consideration. First, all included studies were conducted in China, which may limit the generalizability of our findings to other populations and introduce selection bias. Second, the retrospective nature of studies introduces potential selection and information biases. Third, variability in imaging protocols and preprocessing techniques may have contributed to residual heterogeneity, despite the use of random-effects modeling. Additionally, most models relied solely on internal validation, which limits the assessment of external generalizability.

Future research should prioritize large-scale, prospective, multi-center, and international studies employing standardized imaging and modeling protocols to validate the robustness of these models across different populations and imaging systems. Comparative studies between algorithms and integrative approaches combining radiomics with clinical, genomic, and pathological data are necessary to develop more comprehensive predictive frameworks. Furthermore, advancing automation in feature extraction, model calibration, and cost-effectiveness evaluation will be crucial for translating MRI-based radiomics into precision oncology and personalized treatment planning.

CONCLUSION

This meta-analysis reveals that MRI-based radiomics models demonstrate high diagnostic performance in predicting MSI status in rectal cancer, emphasizing their potential as a non-invasive preoperative evaluation tool. These models could aid in optimizing treatment decisions and advancing precision medicine. However, due to the methodological limitations and heterogeneity across current studies, multi-center, prospective research are essential to validate their clinical value and cost-effectiveness.

***Acknowledgments*:** The authors would like to thank all researchers whose published works contributed to this systematic review and meta-analysis. No external funding was received for this study.

[DOI: 10.61186/ijrr.24.3.39]

Conflict of interest: The authors declare no conflicts of interest relevant to the content of this article.

Funding: This research received no specific grant from any funding agency.

Ethical considerations: This study is a systematic review and meta-analysis of previously published studies. As it does not involve direct patient contact, primary data collection, or intervention, formal ethical approval was not required. All included studies were conducted in accordance with the ethical standards of their respective institutions.

Authors' contributions: Bo Wang: Conceptualization, data extraction, statistical analysis, writing – original draft. Jiamin Gu: Data extraction, quality assessment, writing – review and editing. Bing Wu: Supervision, conflict resolution, writing – review and editing. All authors have read and approved the final version of the manuscript.

AI usage disclosure: No artificial intelligence tools were used in the preparation, drafting, or editing of this manuscript. All literature searches, data extraction, quality assessments, and analyses were performed by the authors.

REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, *et al.* (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians,* ***74****(3): 229-63.*
2. Glynne-Jones R, Wyrwicz L, Tiret E, Brown G, Rödel C, Cervantes A, *et al.* (2017) Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology,* ***28*** *(suppl_4): iv22-iv40.*
3. Gelsomino F, Barbolini M, Spallanzani A, Pugliese G, Cascinu S (2016) The evolving role of microsatellite instability in colorectal cancer: A review. *Cancer Treat Rev,* ***51****: 19-26.*
4. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, *et al.* (2015) PD-1 blockade in tumors with mismatch-repair deficiency. *The New England Journal of Medicine,* ***372****(26): 2509-20.*
5. André T, Shiu K-K, Kim TW, Jensen BV, Jensen LH, Punt C, *et al.* (2020) Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *The New England journal of medicine,* ***383****(23): 2207-18.*
6. Trojan J, Stintzing S, Haase O, Koch C, Ziegler P, Demes M, *et al.* (2021) Complete pathological response after neoadjuvant short-course immunotherapy with ipilimumab and nivolumab in locally advanced MSI-H/dMMR rectal cancer. *Oncologist,* ***26****(12): e2110-e4.*
7. Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen Y-J, Ciombor KK, *et al.* (2018) Rectal cancer, version 2.2018, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network : JNCCN,* ***16****(7): 874-901.*
8. Gilson P, Merlin J-L, Harlé A (2021) Detection of microsatellite instability: State of the art and future applications in circulating tumour DNA (ctDNA). *Cancers (Basel),* ***13****(7): 1491.*
9. Russano M, Napolitano A, Ribelli G, Iuliani M, Simonetti S, Citarella F, *et al.* (2020) Liquid biopsy and tumor heterogeneity in metastatic solid tumors: the potentiality of blood samples. *J Exp Clin Cancer Res,* ***39****(1): 95.*
10. Styk J, Pös Z, Pös O, Radvanszky J, Turnova EH, Buglyó G, *et al.* (2023) Microsatellite instability assessment is instrumental for predictive, preventive and personalised medicine: status quo and outlook. *EPMA J,* ***14****(1): 143-65.*
11. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, *et al.* (2020) Introduction to radiomics. *J Nucl Med,* ***61****(4): 488-95.*
12. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, *et al.* (2019) The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. *Theranostics,* ***9****(5): 1303-22.*
13. Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo -Semedo L, *et al.* (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *European radiology,* ***28****(4): 1465-75.*
14. Shaish H, Aukerman A, Vanguri R, Spinelli A, Armenta P, Jambawalikar S, *et al.* (2020) Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. *European radiology,* ***30****(11): 6263-73.*
15. Liu Z, Zhang J, Wang H, Chen X, Song J, Xu D, *et al.* (2024) MRI-based radiomics feature combined with tumor markers to predict TN staging of rectal cancer. *J Robot Surg,* ***18****(1): 229.*
16. Ao W, Wu S, Wang N, Mao G, Wang J, Hu J, *et al.* (2025) Novel deep learning algorithm based MRI radiomics for predicting lymph node metastases in rectal cancer. *Scientific Reports,* ***15****(1): 12089.*
17. Liu J, Sun L, Lu X, Geng Y, Zhang Z (2023) Development and validation of a stacking nomogram for predicting regional lymph node metastasis status in rectal cancer via deep learning and hand-crafted radiomics. *International Journal of Radiation Research,* ***21*** *(2): 267-74.*
18. Liu J, Liu K, Cao F, Hu P, Bi F, Liu S, *et al.* (2025) MRI-based radiomic nomogram for predicting disease-free survival in patients with locally advanced rectal cancer. *Abdominal Radiology (New York),* ***50****(6): 2388-400.*
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al.* (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* ***372****: n71.*
20. Wang Q, Xu J, Wang A, Chen Y, Wang T, Chen D, *et al.* (2023) Systematic review of machine learning-based radiomics approach for predicting microsatellite instability status in colorectal cancer. *La radiologia medica,* ***128****(2): 136-48.*
21. Capello Ingold G, Martins Da Fonseca J, Kolenda Zloić S, Verdan Moreira S, Kago Marole K, Finnegan E, *et al.* (2025) Preoperative radiomics models using CT and MRI for microsatellite instability in colorectal cancer: a systematic review and meta-analysis. *Abdominal Radiology,* ***50****(11): 5148–61.*
22. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, *et al.* (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol,* ***14****(12): 749 -62.*
23. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med,* ***155****(8): 529-36.*
24. Li Z, Dai H, Liu Y, Pan F, Yang Y, Zhang M (2021) Radiomics analysis of multi-sequence MR images for predicting microsatellite instability status preoperatively in rectal cancer. *Frontiers in Oncology,* ***11****: 697497.*
25. Zhang W, Huang Z, Zhao J, He D, Li M, Yin H, *et al.* (2021) Development and validation of magnetic resonance imaging-based radiomics models for preoperative prediction of microsatellite instability in rectal cancer. *Annals of Translational Medicine,* ***9****(2): 134.*
26. Zhang W, Yin H, Huang Z, Zhao J, Zheng H, He D, *et al.* (2021) Development and validation of MRI-based deep learning models for prediction of microsatellite instability in rectal cancer. *Cancer Medicine,* ***10****(12): 4164-73.*
27. Jing G, Chen Y, Ma X, Li Z, Lu H, Xia Y, *et al.* (2022) Predicting mismatch‐repair status in rectal cancer using multiparametric MRI-based radiomics models: A preliminary study. *BioMed Research International,* ***2022****(1): 6623574.*
28. Jin J, Wang H, Chen Y (2023) Predictive value of multimodal magnetic resonance imaging based radiomics model for microsatellite instability of rectal cancer. *Chinese Journal of Digestive Surgery,* ***22****(6): 779-87.*
29. Li Z, Zhang J, Zhong Q, Feng Z, Shi Y, Xu L, *et al.* (2023) Development and external validation of a multiparametric MRI-based radiomics model for preoperative prediction of microsatellite instability status in rectal cancer: a retrospective multicenter study. *European Radiology,* ***33****(3): 1835-43.*
30. Xiang S, Zheng LB, Zhu L, Gao Y, Wang DS, Liu SL, *et al.* (2023) Radiomics-based prediction of microsatellite instability in stage II and III rectal cancer patients based on T2WI MRI and diffusion-

[DOI: 10.61186/ijrr.24.3.39]

weighted imaging. *Chinese Journal of Digestive Surgery,* ***61****(9): 782-7.*
31. Zhang Y, Liu J, Wu C, Peng J, Wei Y, Cui S (2023) Preoperative prediction of microsatellite instability in rectal cancer using five machine learning algorithms based on multiparametric MRI radiomics. *Diagnostics,* ***13****(2): 269.*
32. Cai Z, Xu Z, Chen Y, Zhang R, Guo B, Chen H, *et al.* (2024) Multiparametric MRI subregion radiomics for preoperative assessment of high-risk subregions in microsatellite instability of rectal cancer patients: a multicenter study. *International Journal of Surgery (London, England),* ***110****(7): 4310-9.*
33. Xing X, Li D, Peng J, Shu Z, Zhang Y, Song Q (2024) A combinatorial MRI sequence-based radiomics model for preoperative prediction of microsatellite instability status in rectal cancer. *Scientific Reports,* ***14****(1): 11760.*
34. Wang Y, Xie B, Wang K, Zou W, Liu A, Xue Z, *et al.* (2025) Multi-parametric MRI habitat radiomics based on interpretable machine learning for preoperative assessment of microsatellite instability in rectal cancer. *Academic Radiology,* ***32****(7): 3975-88.*
35. Yao X, Deng S, Han X, Huang D, Cao Z, Ning X, *et al.* (2025) Deep learning algorithm□based MRI radiomics and pathomics for predicting microsatellite instability status in rectal cancer: a multicenter study. *Academic Radiology,* ***32****(4): 1934-45.*
36. Zheng C-Y, Zhang J-M, Lin Q-S, Lian T, Shi L-P, Chen J-Y, *et al.* (2025) Noninvasive prediction of microsatellite instability in stage II/III rectal cancer using dynamic contrast-enhanced magnetic resonance imaging radiomics. *World Journal of Gastrointestinal Oncology,* ***17****(8): 108362.*
37. Ma W, Hou C, Yang M, Wei Y, Mao J, Guan L, *et al.* (2025) Different MRI-based radiomics machine learning models to predict CD3+ tumor-infiltrating lymphocytes in rectal cancer. *Frontiers in Oncology,* ***15****: 1509207.*
38. Valizadeh P, Jannatdoust P, Pahlevan-Fallahy MT, Hassankhani A, Amoukhteh M, Bagherieh S, *et al.* (2025) Diagnostic accuracy of radiomics and artificial intelligence models in diagnosing lymph node metastasis in head and neck cancers: a systematic review and meta-analysis. *Neuroradiology,* ***67****(2): 449-67.*
39. Charilaou P, Battat R (2022) Machine learning models and over-fitting considerations. *World Journal of Gastroenterology,* ***28****(5): 605-7.*
40. Zabor EC, Reddy CA, Tendulkar RD, Patil S (2022) Logistic regression in clinical studies. *Int J Radiat Oncol, Biol, Phys,* ***112****(2): 271-7.*
41. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, *et al.* (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology,* ***295****(2): 328-38.*
42. Whybra P, Zwanenburg A, Andrearczyk V, Schaer R, Apte AP, Ayotte A, *et al.* (2024) The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. *Radiology,* ***310****(2): e231319.*
43. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, *et al.* (2021) Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med,* ***4****(1): 65.*
44. Li M, Jiang Y, Zhang Y, Zhu H (2023) Medical image analysis using deep learning algorithms. *Front Public Health,* ***11****: 1273253.*
45. Viswam D, Abraham Francis A, Dawale S, Kanakaraj R, Sharma Y, Karki L, *et al.* (2025) Artificial intelligence in radiology and diagnostic imaging. *Bioinformation,* ***21****(7): 1891-4.*
46. Cheng CH, Shi S-S (2025) Artificial intelligence in cancer: applications, challenges, and future perspectives. *Mol Cancer,* ***24*** *(1): 274.*

[DOI: 10.61186/ijrr.24.3.39]