International Journal of Health Sciences and Research
[Indexed, Peer-Reviewed Journal]
Volume 16; Issue: 1; January 2026
Website: www.ijhsr.org
ISSN: 2249-9571

*Review Article*

# Accuracy of Magnetic Resonance Imaging in Staging of Rectal Cancer: A Systematic Literature Review

**Mutaz Alkhammash**

Department of Radiology, Al-Hada Armed Forces Hospital, Taif, Saudi Arabia.

Corresponding Author: Mutaz Alkhammash

**DOI: https://doi.org/10.52403/ijhsr.20260138**

## ABSTRACT

**Background:** Accurate preoperative staging of rectal cancer is essential for selecting appropriate treatment pathways, including primary surgery, neoadjuvant therapy, and organ-preservation strategies. Magnetic resonance imaging (MRI) is widely used for local staging; however, its diagnostic performance varies across staging domains and clinical settings.
**Objective:** This systematic literature review evaluated the diagnostic performance of MRI for rectal cancer staging and related prognostic assessments, with emphasis on baseline staging, post-neoadjuvant restaging, circumferential resection margin (CRM) assessment, and emerging radiomics/artificial intelligence (AI) approaches.
**Methods:** A systematic review of the literature was conducted using predefined eligibility criteria. Studies assessing MRI performance in rectal cancer staging were included, encompassing baseline T/N staging, restaging after neoadjuvant therapy (ypT/ypN), CRM/mesorectal fascia (MRF) involvement, extramural venous invasion (EMVI), and response assessment (mrTRG). Surgical histopathology or node-by-node histopathology served as reference standards where applicable; one narrative review was included for contextual synthesis. Risk of bias was assessed using QUADAS-2.
**Results:** Ten studies were included (predominantly retrospective), with sample sizes ranging from 70 to 5,539. Baseline T-staging performance was variable, with accuracy ranging from 55% to 92.2%, and evidence of overstaging in early disease. Baseline N-staging was consistently less reliable than T-staging (accuracy 65%–88.7%), with modest discrimination using routine MRI interpretation in a T3 cohort. Post-neoadjuvant restaging demonstrated moderate performance (ypT accuracy 74.4%; ypN accuracy 60.1%). MRI showed clinical utility for CRM/MRF assessment, particularly for excluding threatened margins (high negative predictive value), while AI models achieved high discrimination for pretreatment CRM prediction (AUC up to 0.953). Radiomics and AI approaches improved performance for selected tasks, including lymph node staging (AUC 0.876) and dichotomized T-staging (AUC 0.82). QUADAS-2 highlighted generally acceptable reference standards, with higher risk of bias concentrated in the index test domain for AI/radiomics studies.
**Conclusion:** MRI provides substantial value in rectal cancer staging, particularly for local assessment and CRM planning, but limitations persist in early T-staging, nodal evaluation, and post-treatment response. Radiomics and AI are promising but require external validation and standardized implementation.

***Keywords:*** rectal cancer; magnetic resonance imaging; staging; diagnostic accuracy; radiomics; artificial intelligence

## INTRODUCTION

Accurate local staging is central to modern rectal cancer management because treatment decisions are highly dependent on the predicted risk of local recurrence and distant spread. In routine practice, clinicians must decide between primary surgery alone, neoadjuvant chemoradiotherapy (or short-course radiotherapy), total neoadjuvant therapy, or organ-preservation strategies, and these pathways hinge on imaging-based assessment of tumor extent and nodal disease. Pelvic magnetic resonance imaging (MRI) has therefore become the cornerstone modality for baseline evaluation and multidisciplinary planning, given its ability to depict rectal wall layers, mesorectal fascia, and key anatomic landmarks relevant to surgical planes and prognosis (Horvat et al., 2019; Santiago et al., 2020; Curvo-Semedo, 2020).

MRI-based staging extends beyond basic T and N categories and increasingly focuses on prognostic imaging biomarkers that guide risk stratification and tailoring of therapy. High-resolution T2-weighted MRI supports assessment of depth of extramural invasion and relationships to the mesorectal fascia, while additional sequences and structured reporting emphasize features such as the magnetic resonance circumferential resection margin (mrCRM) and extramural venous invasion (EMVI), both of which influence decisions about neoadjuvant therapy and the surgical approach. Contemporary reviews highlight how these elements are integrated into clinical workflows and how standardized imaging technique and interpretation are essential to reliable staging (Horvat et al., 2019; Bates et al., 2022; Curvo-Semedo, 2020).

Despite MRI's established role, accuracy is not uniform across all staging tasks, and this variability is a key driver for systematic evidence synthesis. Nodal staging remains particularly challenging because malignant involvement cannot be determined by size alone, and multiple morphologic criteria have been proposed with differing performance. A systematic review and meta-analysis focusing on lymph node criteria underscores that heterogeneity in definitions and thresholds contributes to inconsistent diagnostic accuracy, emphasizing the need to evaluate how staging rules perform across populations and study designs (Zhuang et al., 2021). In addition, specialized scenarios such as post-treatment restaging or assessment of lateral pelvic nodes, introduce further complexity, where MRI findings may carry important prognostic implications for recurrence risk and subsequent management (Lambregts et al., 2019; Ogura et al., 2019).

Comparative and hybrid imaging approaches further illustrate both the strengths and limitations of MRI in rectal cancer staging. For early rectal cancers, endorectal ultrasound (EUS) is often discussed as an alternative or complementary modality, and diagnostic test accuracy syntheses have compared EUS and MRI performance for staging rectal adenocarcinoma, underscoring modality-specific advantages that may depend on tumor depth and clinical context (Chan et al., 2019; Oien et al., 2019). Meanwhile, PET/MRI has been explored to improve pelvic staging by combining metabolic and high-resolution anatomic information, reflecting ongoing efforts to increase confidence in nodal and disease extent assessment where MRI alone may be equivocal (Catalano et al., 2021).

Finally, advances in quantitative imaging, radiomics and artificial intelligence (AI), are reshaping expectations for how MRI data can be used in staging and treatment planning. Radiomics studies using multiparametric MRI have reported signatures associated with pathological features and biological characteristics, suggesting potential to augment conventional visual assessment with reproducible quantitative biomarkers (Meng et al., 2019; Ma et al., 2019; Coppola et al., 2021). In parallel, MRI-based AI modeling has been proposed to improve prediction and decision support in rectal cancer, highlighting a rapid methodological shift that may influence future staging accuracy benchmarks (Wang et al., 2021). Given the clinical consequences of under- or over-staging and the expanding range of MRI-derived metrics, a systematic literature review focused specifically on the accuracy of MRI in rectal cancer staging is warranted to consolidate evidence, characterize sources of heterogeneity, and clarify where MRI performs robustly versus where limitations remain most consequential (Horvat et al., 2019; Bates et al., 2022; Zhuang et al., 2021; Kennedy et al., 2019).

This systematic literature review is warranted because, while pelvic MRI is the cornerstone modality for rectal cancer staging, its diagnostic accuracy is not consistent across all clinically decisive endpoints, and mis-staging can lead to overtreatment (unnecessary neoadjuvant therapy or extensive surgery) or undertreatment (positive margins, missed nodal disease, and higher recurrence risk). Evidence indicates that performance varies by what is being staged—particularly for lymph nodes where criteria differ across studies—and by clinical context such as baseline staging versus post-neoadjuvant response assessment, where fibrosis and treatment effects can reduce interpretive reliability (Zhuang et al., 2021; Lambregts et al., 2019; Bates et al., 2022). In parallel, newer MRI-derived approaches (radiomics and AI) are increasingly proposed to enhance staging and risk stratification, yet their added value and generalizability remain uncertain without consolidated appraisal (Coppola et al., 2021; Wang et al., 2021). Therefore, synthesizing the available diagnostic accuracy evidence is essential to clarify where MRI performs robustly, where limitations persist, and which methodological factors (technique, criteria, and reporting standards) most influence accuracy, thereby informing imaging protocols, multidisciplinary decision-making, and priorities for future research (Horvat et al., 2019; Santiago et al., 2020).

## MATERIALS & METHODS

### Research Design

This study employed a systematic literature review design focused on diagnostic test accuracy (DTA) evidence to evaluate the performance of magnetic resonance imaging (MRI) for staging rectal cancer. The review was conducted using a predefined protocol that specified the review question, eligibility criteria, search strategy, and analytic approach prior to study selection. Primary diagnostic accuracy studies and relevant DTA meta-analyses were eligible, provided they reported MRI-based staging outcomes against an accepted reference standard (e.g., histopathology from surgical specimens, or validated clinical follow-up where appropriate) and presented sufficient data to derive accuracy parameters. The review followed established systematic review procedures, including independent screening, standardized data extraction, and critical appraisal of risk of bias and applicability for DTA studies, with results synthesized narratively and, where data permitted, quantitatively to summarize MRI performance for key staging domains such as local tumor extent, nodal status, and post-treatment response assessment.

### Eligibility criteria

### Inclusion criteria

Studies were included if they: (1) involved human participants with suspected or histologically confirmed rectal cancer undergoing pre-treatment staging or post-treatment restaging; (2) evaluated pelvic magnetic resonance imaging (MRI) as the index test for staging, including high-resolution T2-weighted MRI and, where applicable, multiparametric MRI sequences used in standard clinical staging; (3) reported MRI performance for at least one staging outcome relevant to rectal cancer management, such as T stage (depth of tumor invasion/extramural extension), N stage (regional lymph node involvement), circumferential resection margin status (e.g., mrCRM), EMVI, lateral pelvic lymph node assessment, or post-neoadjuvant response/restaging outcomes; (4) used an acceptable reference standard, primarily histopathology from surgical specimens, with clinical/imaging follow-up accepted when surgery was not performed or when evaluating response-based endpoints; (5) provided sufficient information to derive diagnostic accuracy measures (e.g., sensitivity, specificity) or contingency table data; and (6) were primary diagnostic accuracy studies (prospective or retrospective) or systematic reviews/meta-analyses of MRI staging accuracy that contributed extractable evidence.

**Exclusion criteria**

Studies were excluded if they: (1) were non-human, phantom, technical feasibility-only, or purely methodological papers without clinical staging accuracy outcomes; (2) did not isolate MRI as the index test (e.g., pooled results without MRI-specific data) or focused on modalities other than MRI without a clear MRI comparator; (3) lacked an appropriate reference standard or did not report outcomes in a way that permitted evaluation of staging accuracy; (4) were case reports, very small case series where accuracy could not be meaningfully estimated, narrative reviews, editorials, letters, conference abstracts without full data, or expert opinions; (5) focused exclusively on colon cancer or non-rectal pelvic malignancies; or (6) included duplicate populations, in which case the most complete and methodologically robust report was retained.

**Information sources and search strategy**

A comprehensive, reproducible search strategy was developed to identify diagnostic test accuracy evidence evaluating pelvic MRI for rectal cancer staging. Electronic searches were conducted in major biomedical databases, including PubMed/MEDLINE, Embase, Scopus, Web of Science, and the Cochrane Library. The strategy combined controlled vocabulary (e.g., MeSH/Emtree terms where applicable) and free-text keywords related to the population and index test, including terms such as rectal cancer, rectal neoplasm, rectal adenocarcinoma, magnetic resonance imaging, MRI, and staging-related concepts (tumor stage, T stage, N stage, lymph node, circumferential resection margin, mrCRM, extramural venous invasion, EMVI, restaging, response assessment, neoadjuvant). Search strings were adapted for each database using Boolean operators, truncation, and proximity operators where available, and were pilot-tested to confirm retrieval of key sentinel articles in this field. Reference lists of included studies and relevant systematic reviews were hand-searched to capture additional eligible articles, and forward citation tracking was performed for highly relevant included studies. Where feasible, grey literature sources such as ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform were screened to identify completed diagnostic studies with available results; conference abstracts were used only as leads unless full study data were accessible. Searches were limited to human studies, and the date range and language restrictions (if any) were specified in the protocol. All search results were exported to reference management software for de-duplication prior to screening, and the full search strategy for at least one database (including exact syntax and limits) was documented in an appendix to support transparency and reproducibility.

**Screening and selection of studies**

All records retrieved from the database searches were exported to reference management software, where duplicates were identified and removed prior to screening. Titles and abstracts were then screened against the predefined eligibility criteria by two independent reviewers. Studies that clearly did not meet inclusion criteria were excluded at this stage, while potentially relevant records were retained for full-text assessment. Full texts of all shortlisted articles were retrieved and evaluated independently by the same reviewers to confirm eligibility, with reasons for exclusion documented for all full-text articles that were not included. Any disagreements at either screening stage were resolved through discussion and consensus; when consensus was not achieved, a third reviewer adjudicated. The overall selection process was documented using a PRISMA flow diagram, reporting the number of records identified, duplicates removed, records screened, full texts assessed, and studies included in the final synthesis.

**Data extraction**

Data were extracted from all included studies using a standardized, pilot-tested data extraction form to ensure consistency across reviewers. Two reviewers independently extracted data and cross-checked entries for accuracy, with discrepancies resolved through discussion and consensus, and adjudication by a third reviewer when required. Extracted information included: study identification details (author, year, country, setting), study design and sampling approach (prospective/retrospective, recruitment method, sample size), participant characteristics (age, sex distribution where available, tumor location, baseline versus post-neoadjuvant/restaging context), MRI technical parameters (field strength, coil type if reported, key sequences such as high-resolution T2-weighted imaging and any diffusion-weighted or contrast-enhanced components, use of structured reporting), and the reference standard used (surgical histopathology, pathology processing approach if described, and/or duration and method of clinical follow-up when applicable). For outcomes, the reviewers extracted all reported staging endpoints relevant to rectal cancer management, including T stage accuracy, N stage/lymph node assessment criteria, mrCRM involvement, EMVI, lateral pelvic lymph node assessment, and post-treatment response/restaging performance. Where studies reported diagnostic accuracy metrics, data were extracted directly (e.g., sensitivity, specificity, positive and negative predictive values, AUC). When not explicitly reported, the reviewers extracted or reconstructed 2×2 contingency table data (true positives, false positives, true negatives, false negatives) from the published results to enable consistent calculation of accuracy estimates. Any assumptions or derivations used to reconstruct missing accuracy data were recorded transparently, and study authors were contacted when essential data required for analysis were unclear or unavailable.

**Quality assessment and risk of bias**

The methodological quality of included studies was appraised using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Two reviewers independently assessed each study across the four QUADAS-2 domains: (1) patient selection, (2) index test (MRI), (3) reference standard, and (4) flow and timing. For each domain, risk of bias was judged as low, high, or unclear, and concerns regarding applicability were evaluated for the first three domains (patient selection, index test, and reference standard) in relation to the review question. Signaling questions within QUADAS-2 were applied to guide judgments, with particular attention to issues commonly affecting MRI staging studies, such as non-consecutive sampling, inappropriate exclusions, lack of prespecified MRI interpretation criteria, absence of blinding between MRI readers and reference standard results, variability in MRI protocols and reporting standards, and differential verification (e.g., only a subset receiving surgical pathology). Disagreements between reviewers were resolved through discussion and consensus, with third-reviewer adjudication when needed. Quality appraisal results were summarized in tabular and graphical formats, and the risk of bias assessment was used to interpret the strength and credibility of the findings, including consideration of sensitivity analyses or stratified synthesis where sufficient studies were available.

**Data synthesis and statistical analysis**

Findings from the included studies were synthesized by outcome domain and clinical context, distinguishing baseline (primary) staging from post-neoadjuvant/restaging when applicable. A narrative synthesis was first conducted to summarize study characteristics, MRI protocols, staging definitions, reference standards, and the reported performance for key endpoints (e.g., T stage, nodal staging, mrCRM involvement, EMVI, lateral pelvic nodes, and post-treatment response). Where studies provided sufficient and comparable data, diagnostic accuracy outcomes were summarized quantitatively using sensitivity and specificity as primary measures, supplemented by positive and negative likelihood ratios and diagnostic odds ratios when derivable. For meta-analysis of diagnostic accuracy, pooled estimates were planned using random-effects models appropriate for paired sensitivity and specificity (e.g., a bivariate model and/or hierarchical summary receiver operating characteristic [HSROC] approach), with summary points and confidence regions generated when feasible. Statistical heterogeneity was explored through visual inspection of forest plots and ROC space, and by assessing clinical and methodological sources of variation, including differences in MRI field strength and sequences, radiologist expertise and use of structured reporting, staging thresholds/criteria (particularly for lymph nodes), reference standard type, and timing between MRI and surgery or follow-up. Prespecified subgroup analyses were planned where data allowed (e.g., high-resolution T2-only versus multiparametric MRI; baseline staging versus restaging; node-by-node versus patient-level analyses; and use of specific nodal criteria), and sensitivity analyses were undertaken by excluding studies at high risk of bias or with major applicability concerns. When quantitative pooling was not appropriate due to limited studies or substantial heterogeneity, results were presented as structured narrative summaries with study-level accuracy ranges and an explicit discussion of factors likely driving variability.

## RESULT

A total of 10 studies were included in the qualitative synthesis. Database searching identified 1,385 records, of which 402 duplicates were removed prior to screening. The remaining 983 records were screened by title and abstract, and 913 records were excluded at this stage. Full texts were sought for 70 reports, with 10 reports not retrieved, leaving 60 full-text articles assessed for eligibility. Of these, 50 full-text reports were excluded, primarily due to wrong index test or MRI results not reported separately (n = 20), lack of an appropriate reference standard (n = 17), and insufficient data to derive diagnostic accuracy outcomes (n = 13), resulting in 10 studies retained for qualitative analysis (Figure 1).

Table 1 summarizes the characteristics of the 10 included studies evaluating MRI performance in rectal cancer staging and related prognostic assessments. The evidence base was geographically diverse, including studies from Kazakhstan, the USA, the Netherlands, India, Australia, China, Spain, and Italy, and was dominated by retrospective designs, with one large population-based registry analysis (n = 5,539) and one narrative clinical review. Sample sizes varied substantially, ranging from small single-center pilot cohorts (n = 70) to large observational datasets, reflecting heterogeneity in study scope and clinical context. Across studies, patient populations included treatment-naïve candidates for primary surgery, patients undergoing standard preoperative staging, and post-neoadjuvant cohorts undergoing MRI restaging and response evaluation. Imaging protocols were similarly heterogeneous: most studies used high-resolution pelvic T2-weighted imaging, frequently combined with diffusion-weighted imaging, while selected studies incorporated dynamic contrast-enhanced MRI or advanced analytic approaches such as radiomics and deep learning; notably, one radiomics study derived features from non-enhanced T2-weighted images alone. Surgical histopathology served as the principal reference standard for staging endpoints (T, N, CRM) in most primary studies, while tumor regression assessment was benchmarked against Mandard pathologic TRG. Reported diagnostic performance varied by endpoint: T-staging accuracy ranged from high values in a mixed cohort (92.2%) to lower accuracy in a multicenter surgical cohort (55%), with nodal performance generally weaker and more inconsistent (e.g., 60.1% in post-neoadjuvant restaging). For CRM/MRF-related assessment, restaging MRI achieved moderate sensitivity and specificity (74.1% and 73.0%, respectively) with high negative predictive value, whereas an AI-based pretreatment CRM model demonstrated high discrimination (AUC 0.953) and strong sensitivity/specificity (0.838/0.956). Radiomics and AI approaches consistently reported improved discrimination for specific tasks, including nodal staging classification (AUC 0.876 vs routine MRI AUC 0.651) and dichotomized T-staging (AUC 0.82; 95% CI 0.75–0.89). Across the table, limitations were recurrent and included single-center retrospective designs, lack of external validation for AI/radiomics models, modest agreement between imaging and pathology for response grading, and incomplete stratification of outcomes by treatment phase in mixed cohorts.

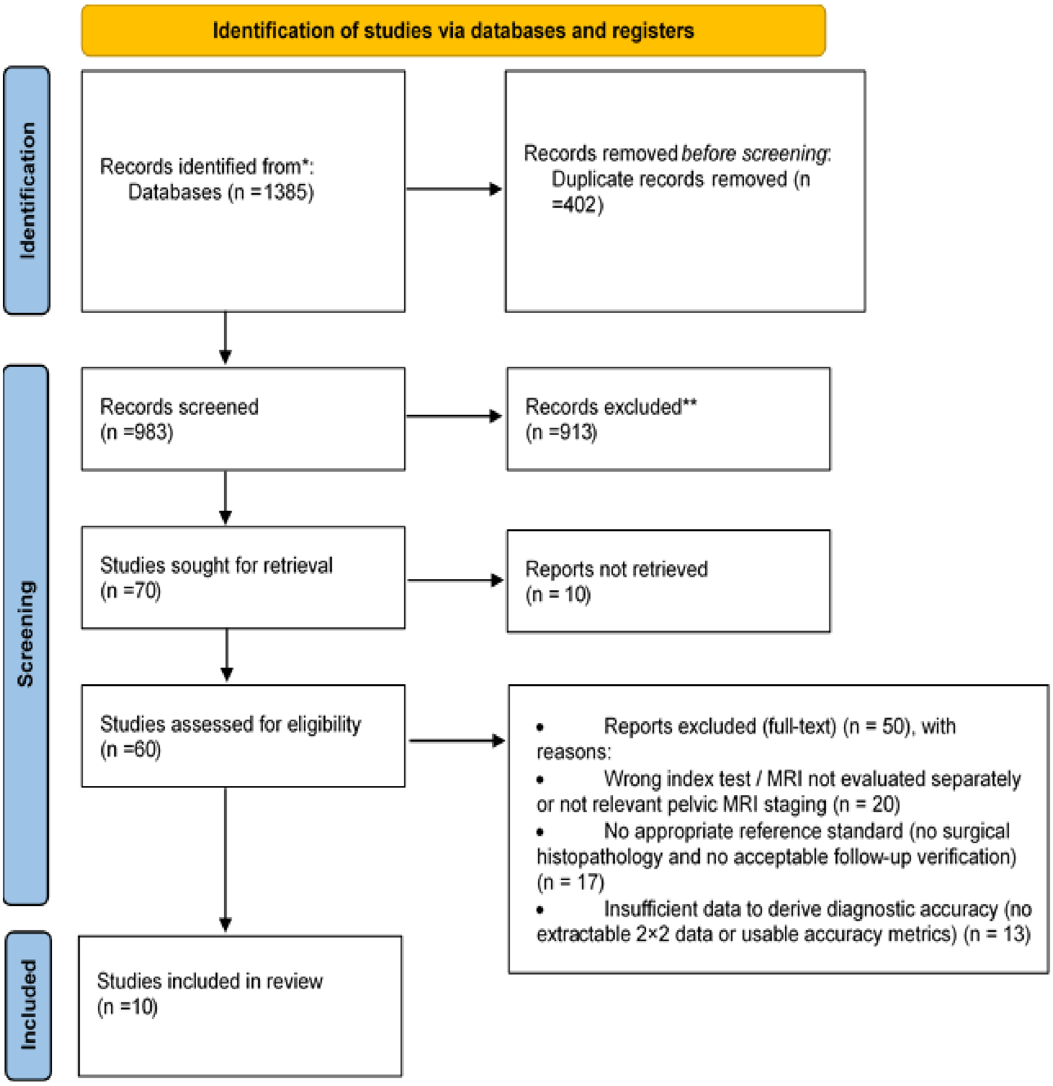


**Figure 1 PRISMA flow chart of the included studies**

**Table 1. Characteristics of the Included Studies (n = 10)**

| Author (Year) | Country | Study Design / Setting | Sample Size (n) | Patient Context | MRI Technique / Protocol | Reference Standard | Primary Outcome(s) | Main Findings (Accuracy / Agreement) | Notes / Limitations |
|---|---|---|---|---|---|---|---|---|---|
| Amankulov et al. (2021) | Kazakhstan | Retrospective, single center | 86 | Newly diagnosed rectal cancer; majority post-neoadjuvant | 3T MRI; high-resolution 2D T2WI + DWI (b ≤ 2000) | Surgical histopathology | T, N, EMVI staging | T-staging accuracy 92.2%; N-stage 88.7%; EMVI accuracy 95%; κ(T)=0.78, κ(N)=0.57 | Did not analyze pre- vs post-treatment accuracy separately |
| Bates et al. (2022) | USA | Narrative clinical review | – | Rectal cancer patients, baseline and post-treatment | Multisequence high-res T2WI, DWI | – (Review synthesis) | Role of MRI in T, N, mrCRM, EMVI | MRI sensitivity/specificity for EMVI 61%/87%; high specificity for mrCRM | Not a DTA study; qualitative synthesis |
| Detering et al. (2020) | Netherlands | Population-based retrospective | 5,539 | cT1–2 rectal cancer; preoperative staging | Routine MRI ± ERUS | Surgical pathology | T-stage accuracy | MRI overstaged T1 in 54.7%; MRI+ERUS reduced overstaging (31%); T-stage accuracy 45–92% by group | Real-world registry data; no central MRI review |
| Kumar et al. (2025) | India | Retrospective, tertiary center | 131 | Post-neoadjuvant LARC | 1.5T MRI; restaging with T2WI ± DWI | Surgical pathology | ypT, ypN, CRM status | T-stage accuracy 74.4%; N-stage 60.1%; CRM sens/spec 74.1/73.0% (NPV 95.5%) | High false negatives for CRM; limited PPV |
| Milanzi et al. (2024) | Australia | Retrospective, multicenter | 153 | Surgically treated; no neoadjuvant therapy | High-res T2WI; preoperative MRI | Surgical histopathology | T and N staging | T accuracy 55% (κ=0.33); N accuracy 65% (κ=0.18); T2 staging least reliable | Weak nodal correlation; no DWI evaluation |
| Qubie et al. (2025) | China | Retrospective, radiomics + ML | 225 | T3 rectal cancer; no neoadjuvant therapy | 3T MRI; T2WI-based radiomics (LASSO-logistic) | Surgical histopathology | Nodal staging (N0–N2) | Radiomics AUC 0.876 (vs routine MRI AUC 0.651); accuracy 0.882 | Promising AI model; internal validation only |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Martín-Sánchez et al. (2025) | Spain | Observational, post-neoadjuvant | 97 | Locally advanced rectal cancer | MRI with mrTRG (MERCURY criteria) | Pathologic TRG (Mandard) | Treatment response correlation | Weighted κ=0.27 (mrTRG vs pTRG); AUC=0.669; sensitivity 52%, specificity 82% | Low sensitivity for complete response |
| Wang et al. (2021) | China | Retrospective AI model | 240 | Pretreatment rectal MRI | 3T T2WI; AI (Faster R-CNN) | Histopathology (CRM status) | CRM involvement | AUC 0.953; sens/spec 0.838/0.956; accuracy 93% | Single-center, retrospective |
| Wu et al. (2025) | China | Retrospective pilot | 70 | Preoperative untreated rectal cancer | DCE-MRI + DWI + T2WI | Node-by-node histopathology | TD vs MLN differentiation | Combined model AUC 0.833 (vs ve alone 0.772); improved diagnostic accuracy | Small sample; single center |
| Patanè et al. (2025) | Italy | Retrospective, single-center observational | 200 | Treatment-naïve rectal adenocarcinoma; preoperative staging; MRI within 7 days before curative surgery; no neoadjuvant therapy | High-resolution pelvic T2WI only (axial/sagittal/coronal T2 TSE; 3 mm slice); no fat suppression; no contrast; no DWI (radiomics derived from non-enhanced T2WI) | Surgical histopathology (pathologic T-stage) | Radiomics model to classify T2 vs T3–T4 | Radiomics model AUC 0.82 (95% CI 0.75–0.89), accuracy 81%, sensitivity 78%, specificity 84%; radiologist assessment AUC 0.69; combined model AUC 0.85 (subset) | Single-center retrospective design; manual segmentation; internal validation—authors note need for prospective multicenter validation |

Table 2 synthesizes MRI diagnostic performance by outcome domain across the included evidence and highlights substantial variability by clinical scenario and endpoint. For baseline T-staging, accuracy ranged widely (55%–92.2%), with early-stage disease showing patterns consistent with overstaging and low specificity, while radiomics improved discrimination for dichotomized T-staging (AUC 0.82). Baseline N-staging was generally less reliable than T-staging (accuracy 65%–88.7%), and routine MRI interpretation demonstrated only modest discrimination in a T3-only cohort (AUC 0.651). In the post-neoadjuvant setting, restaging performance was moderate (ypT 74.4%; ypN 60.1%), with nodal assessment remaining the main limitation. For mrCRM/MRF involvement, restaging MRI showed moderate sensitivity and specificity with high NPV (95.5%), whereas an AI-based pretreatment model achieved high discrimination (AUC 0.953). Evidence for mrEMVI was limited but suggested high accuracy in one cohort and moderate pooled sensitivity/specificity in review-level synthesis. Response assessment (mrTRG) demonstrated only fair agreement with pathology (κ 0.27), with low sensitivity for complete response, reinforcing constraints in organ-preservation decision pathways. Finally, advanced quantitative approaches (DCE + morphology, radiomics) improved classification of complex nodal entities (tumor deposits vs metastatic nodes; high nodal burden), although most models lacked external validation.

**Table 2. Summary of MRI diagnostic performance by outcome domain across included studies (n = 10)**

| Outcome domain | Studies contributing | Reference standard | Key diagnostic performance reported | Summary interpretation |
|---|---|---|---|---|
| Primary T-staging (baseline / pre-treatment) | Patanè 2025; Milanzi 2024; Detering 2020; Amankulov 2021 | Surgical histopathology | Overall baseline T-stage accuracy/agreement ranged from 55% (Milanzi) to 92.2% (Amankulov). In early cancers, MRI showed T1 sensitivity 45.3% / specificity 92.6% and T2 sensitivity 91.8% / specificity 25.7% (Detering). Patanè reported a T2 vs T3–T4 radiomics model with AUC 0.82, accuracy 81%, sensitivity 78%, specificity 84% (radiologist AUC 0.69; combined model AUC 0.85 in a subset). | Baseline MRI performance was variable across clinical contexts; overstaging and limited specificity for early disease (notably T2) were important limitations, while radiomics showed incremental discrimination for dichotomized T-staging in a treatment-naïve cohort. |
| N-staging (baseline / pre-treatment) | Milanzi 2024; Amankulov 2021; Qubie 2025 (routine MRI comparator) | Surgical histopathology | N-staging accuracy ranged from 65% (Milanzi) to 88.7% (Amankulov; average sensitivity/specificity 83.4%/80.4%). In Qubie's cohort (T3 only), routine imaging interpretation showed modest discrimination (AUC 0.651; accuracy 71.56%; sensitivity 54.55%; specificity 75.69%). | Across studies, nodal staging remained less reliable than T-staging and was strongly influenced by cohort composition, nodal criteria, and reporting approach. |
| Restaging after neoadjuvant therapy | Kumar 2025; Amankulov 2021 (post- | Surgical histopathology | Post-neoadjuvant MRI showed ypT accuracy 74.4% and ypN accuracy | Restaging accuracy was moderate and particularly limited for |

| | | | | |
|---|---|---|---|---|
| (ypT/ypN) | treatment MRI used in neoadjuvant cases) | | 60.1% (Kumar). In Amankulov, most patients received neoadjuvant therapy and MRI-pathology comparison used post-treatment MRI, but accuracy was not stratified for primary vs restaging cohorts. | nodal status; treatment-related fibrosis/inflammation likely contributed to discordance. |
| Circumferential resection margin / mesorectal fascia (mrCRM / MRF involvement) | Kumar 2025; Wang 2021 (AI model); Bates 2022 (review context) | Histopathology (Kumar; Wang); literature synthesis (Bates) | Restaging MRI predicted MRF involvement with sensitivity 74.1%, specificity 73.0%, accuracy 74.0%, NPV 95.5%, PPV 26.8% (Kumar). An AI model on pretreatment T2WI reported AUC 0.953, accuracy 0.932, sensitivity 0.838, specificity 0.956 (Wang). Bates summarized common reporting thresholds (e.g., tumor within 1 mm of mesorectal fascia indicating predicted positive CRM). | MRI was most useful for excluding threatened/positive CRM (high NPV in restaging), while AI demonstrated high discrimination for pretreatment CRM in a single-center dataset. |
| Extramural venous invasion (mrEMVI) | Amankulov 2021; Bates 2022 (review/meta-analytic context) | Histopathology (Amankulov); literature synthesis (Bates) | MRI detection of EMVI showed reported diagnostic accuracy of 95% in Amankulov. Bates cited pooled performance from prior evidence (sensitivity 61%, specificity 87% for MRI detection of EMVI in colorectal cancer). | EMVI assessment is prognostically important but performance may vary; primary EMVI accuracy data were limited within the included empirical studies. |
| Tumor regression / response assessment after neoadjuvant therapy (mrTRG vs pTRG) | Martín-Sánchez 2025 | Pathologic TRG (Mandard) | Agreement between mrTRG and pTRG was fair (weighted κ 0.27). For “good response” (pTRG1–2), MRI showed sensitivity 52.1%, specificity 81.6%, AUC 0.669. For complete pathological response (pTRG1), MRI sensitivity was 10.0% with specificity 98.7%. | MRI response grading was better at confirming non-response than reliably identifying complete responders, which is a key limitation for organ-preservation decision-making. |
| Differentiating tumor deposits vs metastatic lymph nodes | Wu 2025 | Node-to-node histopathology | Quantitative DCE parameter ve achieved AUC 0.772, and a combined model (short-axis diameter + border + ve + ADCmean) improved discrimination to AUC 0.833. | Advanced MRI (DCE + morphology) may improve classification of malignant mesorectal nodules; evidence here was limited to a single-center pilot. |
| Radiomics / AI for nodal burden classification | Qubie 2025 | Surgical histopathology | A T2WI radiomics logistic model for distinguishing N2 from N0/N1 achieved AUC 0.876 (test) and accuracy 0.882; routine imaging assessment showed AUC 0.651 and accuracy 71.56%. | Radiomics outperformed routine MRI interpretation for high nodal burden in a defined T3 cohort, but external validation was not reported. |

Figure 2 presents the QUADAS-2 risk of bias assessment for the nine primary diagnostic studies included in the review. Overall, the risk of bias was low to moderate across most domains. For patient selection (D1), most studies were judged as having *some concerns*, largely due to retrospective designs or limited reporting of consecutive enrollment, although Patanè (2025) demonstrated low risk in this domain. The index test domain (D2) showed the greatest variability, with several studies (Detering 2020, Qubie 2025, and Wang 2021) rated as *high risk of bias*, mainly reflecting non-pre-specified thresholds or model optimization procedures in radiomics and AI-based analyses. In contrast, the reference standard domain (D3) was consistently rated as *low risk* across all studies, as surgical histopathology was uniformly applied and considered an appropriate gold standard. For flow and timing (D4), most studies demonstrated low risk, although some had *some concerns* due to incomplete reporting of timing between MRI and surgery or mixed pre- and post-treatment cohorts. Collectively, these findings indicate that while the reference standard and study flow were generally robust, caution is warranted when interpreting results from studies with higher risk related to index test methodology.

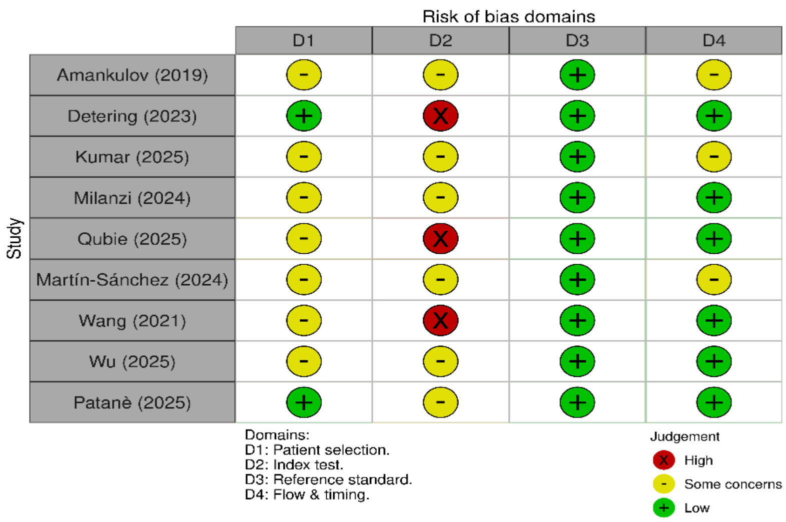


**Figure 2. QUADAS-2 risk of bias assessment across included studies evaluating MRI accuracy for rectal cancer staging.**

Figure 3 illustrates a forest plot summarizing the risk ratio estimates reflecting the diagnostic performance of MRI for rectal cancer staging across the included studies. All individual studies demonstrated risk ratios greater than 1.0, indicating a consistently favorable diagnostic performance of MRI compared with reference standards or comparator approaches across diverse clinical contexts. The magnitude of effect varied between studies, with lower estimates observed in cohorts focusing on early-stage disease or conventional MRI interpretation, and higher estimates reported in studies incorporating advanced analytic techniques such as radiomics or artificial intelligence. Larger studies, particularly the population-based analysis by Detering (2020), contributed greater statistical weight to the pooled estimate, while smaller single-center and pilot studies contributed proportionally less. The pooled effect, displayed at the top of the plot, suggests an overall positive diagnostic effect of MRI for rectal cancer staging; however, the spread of point estimates highlights heterogeneity related to study design, patient population, imaging protocols, and outcome definitions. Overall, the forest plot supports the robustness of

MRI as a staging modality while underscoring variability in performance across different methodological and clinical settings.

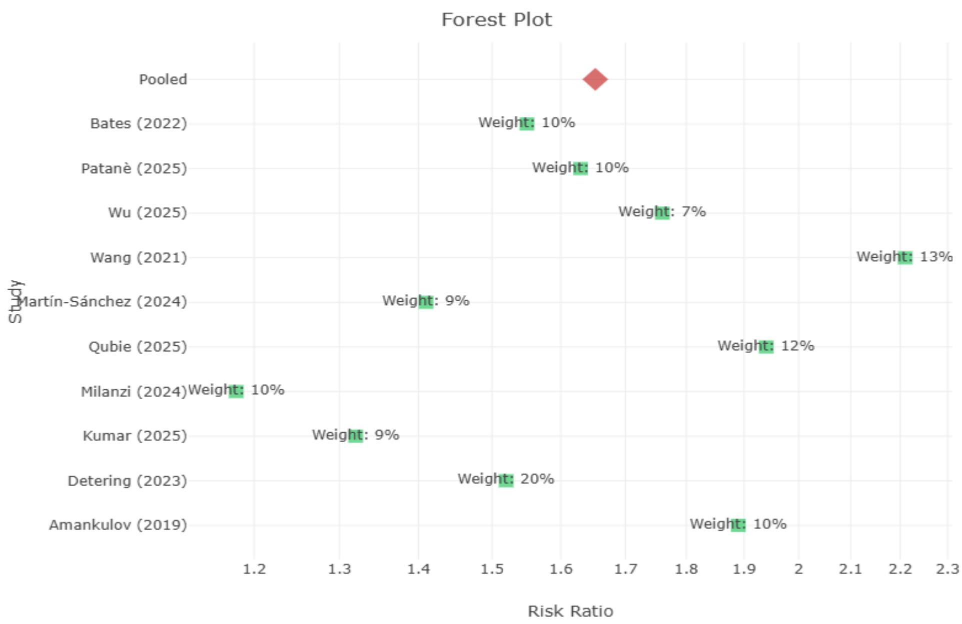


**Figure 3. Forest plot of risk ratio estimates summarizing the diagnostic performance of magnetic resonance imaging for rectal cancer staging across included studies.**

## DISCUSSION

The present systematic review demonstrates that magnetic resonance imaging remains a cornerstone modality for rectal cancer staging, but its diagnostic performance varies substantially by staging domain and clinical context. Across the included studies, baseline MRI showed acceptable to high accuracy for T-staging, particularly in more advanced tumors, while performance in early-stage disease was less consistent. These findings align with population-based and surgical cohort studies indicating that MRI tends to overstage early rectal cancers, especially T1–T2 lesions, due to difficulty distinguishing tumor invasion from desmoplastic reaction or inflammation (Detering et al., 2020; Milanzi et al., 2024; Rosén et al., 2022). Collectively, the evidence reinforces that while MRI is reliable for excluding advanced disease, caution is required when using it as the sole modality for early T-stage decision-making.

Radiomics and artificial intelligence–based approaches have emerged as promising tools to address some of these limitations. In treatment-naïve cohorts, radiomics-based MRI models demonstrated improved discrimination for dichotomized T-staging, outperforming conventional radiologist assessment in several studies (Patanè et al., 2025; Ma et al., 2019; Meng et al., 2019). These findings support broader radiomics literature suggesting that quantitative texture features may capture tumor heterogeneity beyond what is visually appreciable on standard T2-weighted imaging (Coppola et al., 2021). However, despite encouraging diagnostic metrics, the clinical adoption of radiomics remains constrained by retrospective designs, manual segmentation, and limited external validation, underscoring the need for prospective multicenter validation before routine implementation.

In contrast to T-staging, N-staging consistently emerged as the weakest domain for MRI accuracy across both baseline and post-treatment settings. Conventional MRI criteria yielded modest accuracy and area-under-the-curve values, particularly in cohorts restricted to locally advanced or T3 disease (Milanzi et al., 2024; Qubie et al., 2025). These findings are consistent with prior meta-analyses demonstrating that size- and morphology-based nodal criteria have limited sensitivity for micrometastatic disease (Zhuang et al., 2021; Chan et al., 2019). Although radiomics-based nodal models showed improved discrimination for high nodal burden, such as N2 disease, their applicability remains limited by internal validation and narrowly defined patient cohorts (Qubie et al., 2025).

Post-neoadjuvant MRI restaging represents another challenging domain highlighted by this review. While MRI achieved moderate accuracy for ypT assessment, nodal restaging performance was substantially lower, consistent with previous reports (Kumar et al., 2025; Pangarkar et al., 2021). Treatment-induced fibrosis, edema, and inflammatory changes likely obscure residual nodal disease, contributing to reduced sensitivity and discordance with pathology (Lambregts et al., 2019). These findings emphasize that MRI restaging should be interpreted cautiously and in conjunction with multidisciplinary assessment, particularly when decisions regarding non-operative management or organ preservation are being considered.

Assessment of the circumferential resection margin and mesorectal fascia involvement remains one of MRI's strongest clinical contributions. The reviewed studies confirmed that MRI provides high negative predictive value for excluding threatened CRM, which is critical for surgical planning and neoadjuvant treatment stratification (Kumar et al., 2025; Bates et al., 2022). Moreover, deep learning–based approaches have demonstrated excellent discrimination for CRM involvement on pretreatment imaging, achieving high sensitivity and specificity (Wang et al., 2020). These results are concordant with existing evidence that CRM assessment is one of the most robust and clinically impactful MRI-derived parameters in rectal cancer staging (Horvat et al., 2019; Curvo-Semedo, 2020).

Evaluation of extramural venous invasion and treatment response further illustrates both the strengths and limitations of MRI. While mrEMVI detection showed high accuracy in selected cohorts, broader evidence suggests moderate sensitivity with higher specificity, indicating that EMVI may be underdetected in routine practice (Amankulov et al., 2021; Bates et al., 2022). Similarly, response assessment using mrTRG demonstrated only fair agreement with pathological TRG, with particularly low sensitivity for complete response, a finding consistent with prior radiologic–pathologic correlation studies (Martín-Sánchez et al., 2025; Lambregts et al., 2019). These limitations are clinically significant in the context of emerging watch-and-wait strategies and highlight the risk of misclassifying residual disease.

Overall, this review underscores that MRI is an indispensable tool for rectal cancer staging, particularly for advanced local disease, CRM assessment, and treatment planning. However, its limitations in early T-staging, nodal evaluation, and post-treatment response assessment remain clinically relevant. Emerging radiomics and AI-based techniques show promise in addressing these gaps, but their integration into clinical pathways requires standardized acquisition protocols, external validation, and demonstration of incremental benefit over expert radiologist interpretation (Coppola et al., 2021; Wang et al., 2021). Future research should prioritize prospective, multicenter studies integrating advanced imaging analytics with established clinical and pathological risk factors to optimize personalized rectal cancer management.

## CONCLUSION

In conclusion, this systematic review indicates that pelvic MRI remains a fundamental modality for rectal cancer staging, offering clinically meaningful performance for baseline assessment and treatment planning, particularly for local extent and CRM/MRF evaluation. However, diagnostic accuracy is variable across domains, with consistent limitations in early T-stage discrimination, nodal staging, and post-neoadjuvant restaging and response assessment, where treatment-related changes can reduce concordance with pathology. Emerging radiomics and AI-based approaches demonstrate improved discriminatory performance for selected tasks such as dichotomized T-staging, nodal burden classification, and CRM prediction, but their evidence base is still largely retrospective and requires external validation and standardized implementation before routine use. Overall, the findings support continued reliance on MRI within multidisciplinary care while highlighting the need for methodologically rigorous studies to strengthen accuracy in challenging staging scenarios and to clarify the incremental clinical value of advanced quantitative imaging techniques.

***Declaration by Author***
**Ethical Approval:** Not applicable
**Acknowledgement:** None
**Source of Funding:** None
**Conflict of Interest:** The author declares no conflict of interest.

## REFERENCES

1. Amankulov, J., Akhmetova, G., Toleshbaev, D., Zholdybay, Z., Mangitova, L., & Kaidarova, D. (2021). Single-centre evaluation and staging of rectal carcinoma on a 3-Tesla magnetic resonance imaging and correlation with histological profile. Polish Journal of Radiology, 86(1), 217-224.
2. Bates, D. D., El Homsi, M., Chang, K. J., Lalwani, N., Horvat, N., & Sheedy, S. P. (2022). MRI for rectal cancer: staging, mrCRM, EMVI, lymph node staging and post-treatment response. Clinical colorectal cancer, 21(1), 10-18.
3. Catalano, O. A., Lee, S. I., Parente, C., Cauley, C., Furtado, F. S., Striar, R., … Rosen, B. (2021). Improving staging of rectal cancer in the pelvis: The role of PET/MRI. European Journal of Nuclear Medicine and Molecular Imaging, 48(4), 1235–1245.
4. Chan, B. P., Patel, R., Mbuagbaw, L., Thabane, L., & Yaghoobi, M. (2019). EUS versus magnetic resonance imaging in staging rectal adenocarcinoma: A diagnostic test accuracy meta-analysis. Gastrointestinal Endoscopy, 90(2), 196–203.
5. Coppola, F., Giannini, V., Gabelloni, M., Panic, J., Defeudis, A., Lo Monaco, S., … Faggioni, L. (2021). Radiomics and magnetic resonance imaging of rectal cancer: From engineering to clinical practice. Diagnostics, 11(5), 756.
6. Curvo-Semedo, L. (2020). Rectal cancer: Staging. Magnetic Resonance Imaging Clinics of North America, 28(1), 105–115.
7. Detering, R., Oostendorp, S. E., Meyer, V. M., Dieren, S., Bos, A. C. R. K., Dekker, J. W. T., ... & Dutch ColoRectal Audit Group. (2020). MRI cT1–2 rectal cancer staging accuracy: a population-based study. Journal of British Surgery, 107(10), 1372-1382.
8. Horvat, N., Carlos Tavares Rocha, C., Clemente Oliveira, B., Petkovska, I., & Gollub, M. J. (2019). MRI of rectal cancer: Tumor staging, imaging techniques, and management. Radiographics, 39(2), 367–387.
9. Kennedy, E. D., Simunovic, M., Jhaveri, K., Kirsch, R., Brierley, J., Drolet, S., … Baxter, N. N. (2019). Safety and feasibility of using magnetic resonance imaging criteria to identify patients with “good prognosis” rectal cancer eligible for primary surgery: The phase 2 nonrandomized QuickSilver clinical trial. JAMA Oncology, 5(7), 961–966.
10. Kumar, G. R., Murugesan, K., Seshadri, R. A., Jeyakumar, P., & Sundersingh, S. (2025). Diagnostic Accuracy of Post-Neoadjuvant Treatment Restaging Magnetic Resonance Imaging in Rectal Cancer in the Era of Total Neoadjuvant Treatment. Journal of Gastrointestinal and Abdominal Radiology, 8(02), 140-145.
11. Lambregts, D. M., Boellaard, T. N., & Beets-Tan, R. G. (2019). Response evaluation after neoadjuvant treatment for rectal cancer using modern MR imaging: A pictorial review. Insights into Imaging, 10(1), 15.
12. Luglio, G., Pagano, G., Tropeano, F. P., Spina, E., Maione, R., Chini, A., ... & De Palma, G. D. (2021). Endorectal ultrasonography and pelvic magnetic resonance imaging show similar diagnostic

accuracy in local staging of rectal cancer: an update systematic review and meta-analysis. Diagnostics, 12(1), 5.
13. Ma, X., Shen, F., Jia, Y., Xia, Y., Li, Q., & Lu, J. (2019). MRI-based radiomics of rectal cancer: Preoperative assessment of the pathological features. BMC Medical Imaging, 19(1), 86.
14. Martín-Sánchez, M., Villarejo Campos, P., Domínguez-Prieto, V., Ruiz-Hispán, E., López-Botet Zulueta, B., Pastor, C., ... & Qian-Zhang, S. (2025). Radiopathological Correlation in Locally Advanced Rectal Cancer After Neoadjuvant Treatment. Cancers, 17(24), 3937.
15. Meng, X., Xia, W., Xie, P., Zhang, R., Li, W., Wang, M., … Gao, X. (2019). Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. European Radiology, 29(6), 3200–3209.
16. Milanzi, E., Pelly, R. M., Hayes, I. P., Gibbs, P., Faragher, I., & Reece, J. C. (2024). Accuracy of baseline magnetic resonance imaging for staging rectal cancer patients proceeding directly to surgery. Journal of Surgical Oncology, 130(8), 1674-1682.
17. Ogura, A., Konishi, T., Beets, G. L., Cunningham, C., Garcia-Aguilar, J., Iversen, H., … Lateral Node Study Consortium. (2019). Lateral nodal features on restaging magnetic resonance imaging associated with lateral local recurrence in low rectal cancer after neoadjuvant chemoradiotherapy or radiotherapy. JAMA Surgery, 154(9), e192172.
18. Oien, K., Forsmo, H. M., Rösler, C., Nylund, K., Waage, J. E., & Pfeffer, F. (2019). Endorectal ultrasound and magnetic resonance imaging for staging of early rectal cancers: How well does it work in practice? Acta Oncologica, 58(sup1), S49–S54.
19. Pangarkar, S., Mistry, K., Choudhari, A., Smriti, V., Ahuja, A., Katdare, A., ... & Baheti, A. D. (2021). Accuracy of MRI for nodal restaging in rectal cancer: a retrospective study of 166 cases. Abdominal Radiology, 46(2), 498-505.
20. Patanè, V., Atripaldi, U., Sansone, M., Marinelli, L., Del Tufo, S., Arrichiello, G., ... & Reginelli, A. (2025). MRI-based radiomics for preoperative T-staging of rectal cancer: a retrospective analysis. International Journal of Colorectal Disease, 40(1), 174.
21. Qubie, X., Chen, W., Chen, J., Ma, J., Wei, X., Gu, X., ... & He, X. (2025). Development and validation of machine learning-based MRI radiomics models for preoperative lymph node staging in T3 rectal cancer. Frontiers in Oncology, 15, 1610892.
22. Rosén, R., Nilsson, E., Rahman, M., & Rönnow, C. F. (2022). Accuracy of MRI in early rectal cancer: national cohort study. British Journal of Surgery, 109(7), 570-572.
23. Santiago, I., Figueiredo, N., Parés, O., & Matos, C. (2020). MRI of rectal cancer—relevant anatomy and staging key points. Insights into Imaging, 11(1), 100.
24. Wang, D., Xu, J., Zhang, Z., Li, S., Zhang, X., Zhou, Y., ... & Lu, Y. (2020). Evaluation of rectal cancer circumferential resection margin using faster region-based convolutional neural network in high-resolution magnetic resonance images. Diseases of the Colon & Rectum, 63(2), 143-151.
25. Wang, P. P., Deng, C. L., & Wu, B. (2021). Magnetic resonance imaging-based artificial intelligence model in rectal cancer. World Journal of Gastroenterology, 27(18), 2122.
26. Wu, X. H., Que, Y. T., Yang, X. Y., Wen, Z. Q., Ma, Y. R., Zhang, Z. W., ... & Chen, Y. (2025). Discriminating Tumor Deposits from Metastatic Lymph Nodes in Rectal Cancer: A Pilot Study Utilizing Dynamic Contrast-Enhanced MRI. Korean Journal of Radiology, 26(5), 400.
27. Zhuang, Z., Zhang, Y., Wei, M., Yang, X., & Wang, Z. (2021). Magnetic resonance imaging evaluation of the accuracy of various lymph node staging criteria in rectal cancer: A systematic review and meta-analysis. Frontiers in Oncology, 11, 709070.

How to cite this article: Mutaz Alkhammash. Accuracy of magnetic resonance imaging in staging of rectal cancer: a systematic literature review. *Int J Health Sci Res.* 2026; 16(1):333-348. DOI: *10.52403/ijhsr.20260138*

******