Review | Open Access

Artificial Intelligence Surgery

Allen et al. Art Int Surg. 2026;6:682-96 DOI:10.20517/ais.2026.90

# Scoping review of digital pathology for neoadjuvant rectal cancer response

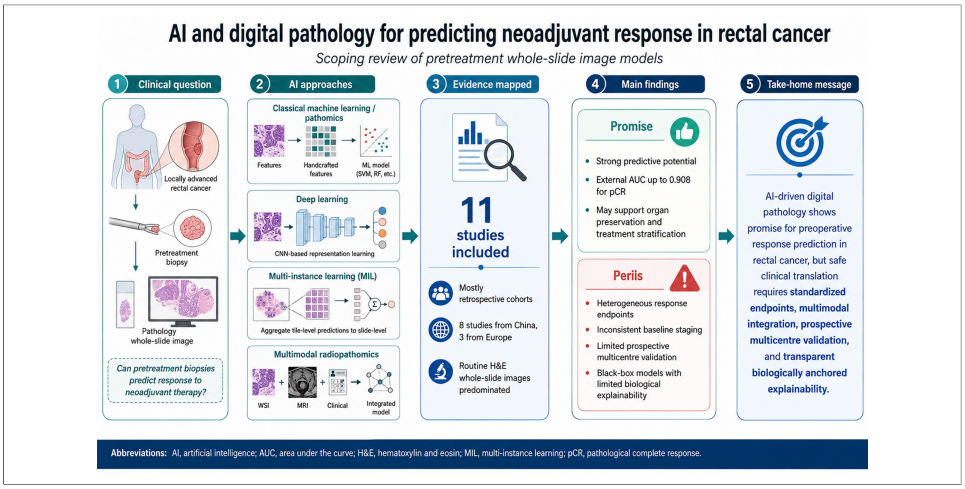
**William Allen[1], Jesse Fischer[2,3], James Atlas[4], Isaac Tranter-Entwistle[1,5], John Pearson[6], Arthur Morley-Bunker[7], Matthew Drake[8], Karolina Vernmark[9,10], Timothy Eglinton[1]**

**Keywords:**
Rectal neoplasms, artificial intelligence, digital pathology, neoadjuvant therapy, whole-slide imaging

**Citation:** Allen W, Fischer J, Atlas J, Tranter-Entwistle I, Pearson J, Morley-Bunker A, Drake M, Vernmark K, Eglinton T. Scoping review of digital pathology for neoadjuvant rectal cancer response. *Art Int Surg.* 2026;6:682-96. https://dx.doi.org/10.20517/ais.2026.90

**Received:** 27 Aug 2026 **First Decision:** 30 Sep 2026 **Revised:** 3 Nov 2026 **Accepted:** 4 Nov 2026 **Published:** 17 Nov 2026

**Academic Editor:** Andrew Gumbs
**Copy Editor:** Tong Wang
**Production Editor:** Tong Wang




## Abstract

**Aim:** Neoadjuvant therapy for locally advanced rectal cancer (LARC) yields variable patient responses. Artificial intelligence (AI) and digital pathology offer a novel approach to preoperatively predict treatment outcomes using histomorphological features, potentially guiding organ-preserving strategies. This scoping review maps the current literature on AI-driven neoadjuvant response prediction in LARC to evaluate algorithmic performance and identify methodological gaps.

**Methods:** Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) guidelines, a structured search of PubMed, Embase, and the Cochrane Library was conducted for studies published within the last 10

[1]Department of Surgery and Critical Care, Christchurch Hospital, Christchurch 8140, New Zealand.
[2]Department of Surgery, Waikato Hospital, Hamilton 3204, New Zealand.
[3]Department of Surgery, University of Auckland - Waikato Clinical Campus, Hamilton 3204, New Zealand.
[4]Department of Computer Science and Software Engineering, University of Canterbury, Christchurch 8140, New Zealand.
[5]Department of General Surgery, University of Otago Medical School, Christchurch 8140, New Zealand.
[6]Department of Pathology and Molecular Medicine, University of Otago, Christchurch 8140, New Zealand.
[7]Mackenzie Cancer Research Group, Department of Pathology and Biomedical Science, University of Otago, Christchurch 8140, New Zealand.
[8]Canterbury Health Laboratories, Christchurch 8140, New Zealand.
[9]Department of Biomedical and Clinical Sciences, Linköping University, Linköping SE-581 83, Sweden.
[10]Department of Oncology in Linköping and Department of Surgery, Orthopaedics and Cancer Care, Linköping University, Linköping SE-581 83, Sweden.

**Correspondence to:** Dr. William Allen, Department of Surgery and Critical Care, Christchurch Hospital, Christchurch 8140, New Zealand. E-mail: williamhocklyallen@gmail.com

years. Studies applying AI algorithms to pretreatment whole-slide images (WSIs) to predict neoadjuvant response in LARC were included.

**Results:** Eleven publications were included. Ten studies used retrospective cohorts only, whereas one study also included a prospective validation cohort of 100 patients. Algorithmic frameworks predominantly featured classical machine learning, using software extraction of predefined features, and deep learning for autonomous pixel analysis. Across heterogeneous endpoints and cohorts, standalone classical machine-learning models reported external validation area under the receiver operating characteristic curve (AUC) of 0.811-0.908 (prospective: 0.733), a hybrid microenvironment model reported external AUCs of 0.643-0.656 (prospective: 0.630), and standalone deep-learning models reported an external AUC of 0.723 (internal: 0.710-0.81). Critical limitations across the field include heterogeneous clinical target endpoints, inconsistent baseline staging, and a reliance on "black box" algorithms lacking true biological explainability.

**Conclusion:** AI-driven digital pathology demonstrates strong potential for predicting neoadjuvant treatment response in LARC. However, safe clinical implementation will require standardized endpoints, multimodal integration, and prospective multi-centre validation of algorithms anchored in verifiable tumour biology.

## INTRODUCTION

Colorectal cancer is a significant global health burden, ranking third in incidence and second in mortality worldwide[1]. Rectal cancer represents a distinct clinical challenge, accounting for roughly 38% of colorectal cancer cases globally[1]. The management of locally advanced rectal cancer (LARC), defined as T3-4 or lymph node-positive disease, has recently undergone a paradigm shift. Standard care increasingly incorporates total neoadjuvant therapy (TNT) to eradicate micrometastases, maximise tumour downstaging and improve overall survival[2,3]. Notably, TNT has been shown to result in a pathological complete response (pCR) or complete clinical response (cCR) in a significant subset of patients, expanding organ-preserving "watch-and-wait" strategies that originated from the observation of complete pathological response after neoadjuvant long-course chemoradiotherapy[4].

Clinical and radiological predictors of response have been investigated, but their ability to guide the selection of neoadjuvant treatment in routine practice remains limited[5-7]. This predictive deficit is particularly detrimental to poor responders who may receive negligible oncological benefit while experiencing significant adverse effects of neoadjuvant therapy, often with long-term sequelae. Furthermore, the time required for this treatment delays definitive surgical management, risking potential disease progression and poorer overall and disease-free survival.

Artificial intelligence (AI) encompasses computational methods designed to perform tasks that typically require human intelligence. In digital pathology, these include machine learning, in which algorithms learn predictive patterns from extracted image-derived or clinicopathological features, and deep learning, a subset of machine learning that uses multi-layered neural networks to learn complex visual features directly from whole-slide image (WSI) data, reducing reliance on manually defined pathomic features. Machine learning methods are being widely investigated and applied in medical research and offer an opportunity to improve prediction of neoadjuvant treatment response for rectal cancer. Advances in digitisation have facilitated the development of pathomics, in which quantitative histological features are extracted from digitised tissue images. AI-supported WSI analysis of digitised specimen slides allows a novel approach to diagnostic and prognostic dilemmas in pathology[8] and has been shown to predict molecular features such as microsatellite instability from histological images for gastrointestinal cancer[9]. Convolutional neural networks (CNNs) have historically been used to extract subvisual features from medical images but require intensive specialist annotation, limiting their applicability. Recent development of deep learning methods such as weakly

supervised multi-instance learning (MIL) and attention-based architectures has allowed more comprehensive utilisation of data in very large images including interpretation of complex nonlinear patterns[10] invisible to pathologists, who rely on assessment of recognised morphological features such as tumour budding and stromal content.

Although this area of research is promising, responsible clinical deployment of these technologies requires clear regulatory standards and institutional governance addressing safety, effectiveness, transparency, accountability, privacy, equity, and ongoing human oversight. Appropriate AI literacy among clinicians and information governance professionals is also necessary to support informed and ethical implementation.

This scoping review aims to map the current literature on the use of AI and digital pathology to predict preoperative neoadjuvant therapy response in LARC, including evaluation of different AI architectures and input modalities, assessment of predictive performance, and identification of methodological gaps to guide future research.

## METHODS

A scoping review methodology was chosen due to the technological novelty and rapid evolution of the field over the past decade, which have created substantial heterogeneity in the literature. Given the wide variety of reported AI architectures, a scoping review offers the required breadth to evaluate diverse approaches and direct future research. The scoping review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) framework and checklist[11]. The protocol was registered retrospectively with OSF Registries on 18 May 2026 (https://doi.org/10.17605/OSF.IO/Z9PJW).

### Eligibility criteria

Studies were included if they investigated patients predominantly with LARC (stage II-III) who received neoadjuvant chemotherapy and/or radiotherapy and applied AI algorithms to pretreatment histological images or WSIs to predict response to the neoadjuvant treatment; there was no further restriction on the neoadjuvant regimen used. Studies including a minority subset of patients with synchronous metastatic disease (Stage IV/cM1) were permitted provided the primary focus was preoperative response prediction of a locally advanced rectal primary tumour to neoadjuvant therapy. All types of machine or deep learning techniques were included. Multimodal or multiparametric investigations incorporating secondary diagnostic streams (such as radiomics or clinical data) were eligible only if the digital pathology or whole-slide imaging component was evaluated and reported separately. Studies were eligible if they applied AI methodologies to pretreatment histological images or WSIs to predict pathologically assessed response to neoadjuvant therapy, irrespective of the response definition or tumour regression grading system used. Studies published in English between 17 February 2016 and 17 February 2026 were eligible. Reviews, letters, editorials, case reports, and conference abstracts lacking full data were excluded.

### Information sources

A structured literature search was conducted on 17 February 2026 in PubMed (National Library of Medicine), Embase (Elsevier), and the Cochrane Library (Wiley). Searches were limited to English-language records published between 17 February 2016 and 17 February 2026. The exact search string and limits applied in each database are provided in Supplementary Table 1. The same free-text Boolean query was entered into all three database interfaces; database-specific controlled vocabulary and field tags were not used.

### Selection of sources of evidence

All identified citations were imported into the Rayyan web application for deduplication and screening[12]. To maintain methodological rigor, Rayyan's automated AI screening features were disabled; the platform was utilised solely for data management and blinding of reviewers. Two independent investigators (Allen W and Tranter-Entwistle I) conducted the initial title and abstract screening. Any discrepancies regarding study eligibility were resolved through consensus discussion. A third senior reviewer (Eglinton T) was nominated to perform arbitration, but this was not required as consensus was reached on all studies. Full-text articles of all potentially eligible studies were reviewed.

### Data extraction

Data were manually charted by one reviewer (Allen W) from the full-text articles and available Supplementary Materials using a standardised form in Rayyan. No duplicate extraction or independent reliability verification was performed. Extracted items comprised first author, publication year, country, study design, funding sources, cohort size and partitioning, baseline clinical stage, neoadjuvant regimen, response endpoint and definition, WSI acquisition and preprocessing, AI class and architecture, image scale and input size, segmentation method, predictive performance according to validation type, and explainability method. Unreported items were recorded as not reported.

We additionally extracted whether each publication reported confidence intervals for its main performance estimates, calibration, decision curve analysis, and sensitivity, specificity, positive predictive value, and negative predictive value at a stated prediction threshold. Validation design and any evaluation of clinical utility were recorded separately.

### Synthesis of results

Extracted data were synthesised and categorised into four overarching domains to facilitate comparative analysis across the literature: (1) image acquisition and pre-processing; (2) AI architecture and algorithm design; (3) performance metrics; and (4) model explainability. Meta-analysis was not performed due to significant heterogeneity in study designs, algorithmic frameworks, and outcome reporting. Instead, a narrative synthesis of the findings was conducted.

## RESULTS

### Overview of included studies

This scoping review included 11 publications evaluating the use of digital pathology and AI to predict treatment response in LARC[13-23]. The study selection process is summarised in Figure 1. Included studies were from China (eight studies) and Europe (three studies). Ten of the 11 included studies used retrospective cohorts only; Feng *et al.* developed and validated their model using three retrospective cohorts and subsequently evaluated it in a multicentre prospective validation cohort of 100 patients[13]. Five of the 11 studies evaluated their models in independent external cohorts[13,16-18,23], while the remaining six studies relied on internal train-test splits, internal validation, cross-validation, or single-cohort association analyses.

The patient cohorts reported across these publications were not fully independent. The 303-patient training cohort and the 480- and 150-patient retrospective validation cohorts reported by Feng *et al.*[13] corresponded to the same institutional cohorts reported by Shao *et al.*[18]. Feng *et al.* added a prospective cohort of 100 patients, whereas Shao *et al.* included an additional 48-patient retrospective validation cohort. Thus, the eight cohort entries across these two publications represented five independent cohorts. Additional patient overlap was possible in Lou *et al.*[17], which recruited from the same three institutions during an overlapping period, and between the similarly sized Yunnan Cancer Hospital datasets reported by Zhang *et al.* (2020)[21] and Zhang *et al.* (2022)[22]. However, patient-level overlap could not be confirmed from the published information. Participant numbers were therefore not pooled across publications.

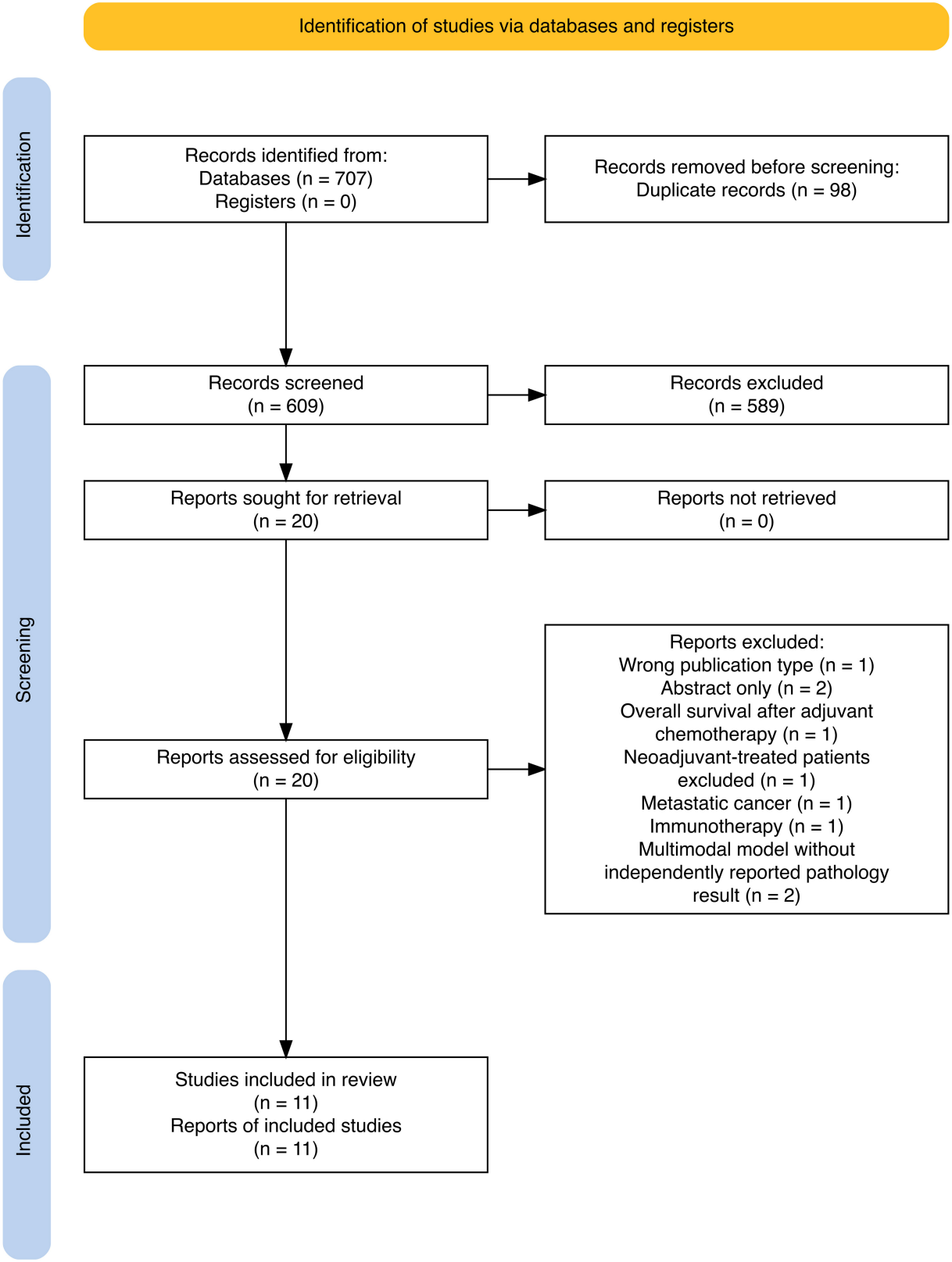



**Figure 1.** PRISMA flow diagram of study selection. Searches of PubMed, Embase, and the Cochrane Library identified 707 records (with 0 records identified from registers). After removal of 98 duplicates, 609 records underwent title and abstract screening, and 589 were excluded. All 20 reports sought for retrieval were retrieved and assessed in full text for eligibility. Nine reports were excluded because of wrong publication type (*n* = 1), abstract-only publication (*n* = 2), assessment of overall survival after adjuvant chemotherapy (*n* = 1), exclusion of patients receiving neoadjuvant therapy (*n* = 1), metastatic cancer (*n* = 1), immunotherapy (*n* = 1), or a multimodal model without separately reported pathology results (*n* = 2). Eleven publications (11 studies) were included in the review. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; *n*: number of records.

There was considerable heterogeneity in overall study population sizes and cohort partitioning [Table 1]. While larger studies (885 to 1,033 patients) commonly utilised distinct internal testing and independent external validation sets, smaller studies used single-cohort association analyses or internal model development and testing only. Lafarge *et al.* trained their algorithm on 1,057 colorectal cancer WSIs and assessed associations with treatment response in two independent rectal cancer cohorts comprising 169 patients[16].

Baseline clinical staging criteria lacked strict standardisation across the literature [Table 1]. While several studies utilised pretreatment magnetic resonance imaging (MRI) to strictly bound their cohorts to cT3-cT4 disease without distant metastasis, others grouped patients under broader clinical categories (e.g., "locally advanced" or "high-risk") without specifying the imaging modality used for staging. Notably, Jepsen *et al.* included a substantial subset (32%) of patients presenting with cM1 distant metastasis[15], and Foersch *et al.* did not report baseline clinical tumour-node-metastasis (cTNM) staging, relying instead on postoperative pathological staging (ypTNM)[14].

Treatment regimens varied across studies and included short-course radiotherapy, long-course chemoradiotherapy, and TNT. The most common regimen was long-course chemoradiotherapy, comprising pelvic radiotherapy administered concurrently with 5-fluorouracil-based chemotherapy over 5-6 weeks. Several recent studies incorporated TNT, combining induction or consolidation chemotherapy, such as capecitabine and oxaliplatin (CAPEOX) or folinic acid, fluorouracil, and oxaliplatin (FOLFOX), with long-course chemoradiotherapy. Treatment was less uniform in some smaller cohorts; for example, Jepsen *et al.* included 42 patients treated with chemoradiotherapy, seven with chemotherapy alone, and one with radiotherapy alone[15].

The clinical endpoints used to train and evaluate the AI models varied. Several studies predicted pCR, defined as ypT0N0 or tumour regression grade (TRG) 0 with no residual tumour cells. Other studies used broader regression-based classifications. Wang *et al.*[19] and Shao *et al.*[18] classified TRG 0-1 as a "good response" and TRG 2-3 as a "poor response". Jepsen *et al.*[15] used the Mandard system, defining TRG 1-2 as a good response and TRG 3-5 as a poor response. Foersch *et al.*[14] used the Dworak system, grouping grades 0-2 as "no response" and grades 3-4 as "response". Wu *et al.* developed a multiclass classifier for three AJCC TRG categories: TRG 0 (pCR), TRG 1-2 (moderate to minimal response), and TRG 3 (poor response)[20]. Some studies also evaluated disease-free survival, relapse-free survival, or overall survival.

## Image acquisition and pre-processing

The foundation of all predictive algorithms was the digitisation of standard histopathological slides into WSIs. Most models generated predictions using routine haematoxylin and eosin (H&E) staining. Foersch *et al.* used a multistain immunohistochemistry (IHC) model[14], while Jepsen *et al.* aligned H&E and IHC slides to quantify immune cell populations within the tumour microenvironment[15].

A significant bottleneck in digital pathology is the necessity of isolating the tumour region of interest to prevent the AI from analysing irrelevant tissue or background artefact. While many studies relied on manual segmentation by pathologists, several recent pipelines automated segmentation [Table 2]. Approaches included automated binarisation algorithms [Zhang *et al.* (2022)[22]] or processing unannotated tissue blocks directly (Wang *et al.*[19]). These advancements demonstrate a trajectory toward fully automated workflows that reduce pathologist supervision.

**Table 1. Clinical characteristics, study design, and neoadjuvant treatment regimens of the included literature**

| Authors | Country & study design | Total patients, cohort split & validation | Clinical stage | Neoadjuvant regimen | Outcome attempted to predict | Reported discrimination and response associations |
|---|---|---|---|---|---|---|
| Feng *et al.* (2022)[13] | China; Retrospective with prospective validation | *N* = 1,033 (Train: 303, Ext Val 1: 480, Ext Val 2: 150, Prospective Val: 100) | cT1-T4b, cN0-N2; via MRI | Long-course CRT (50 Gy pelvic RT + concurrent 5-FU) | pCR *vs.* non-pCR. The pathological report served as the gold standard | Pathomics nucleus model: retrospective external validation AUCs of 0.811 (0.774-0.848) and 0.832 (0.764-0.900); prospective validation AUC of 0.733 (0.620-0.845). Pathomics microenvironment model: retrospective external validation AUCs of 0.656 (0.599-0.712) and 0.643 (0.542-0.743); prospective validation AUC of 0.630 (0.507-0.754) |
| Foersch *et al.* (2023)[14] | Germany; Retrospective | *N* = 117 (Neoadjuvant predictor cohort; 11-fold cross-validation) | Did not report baseline cTNM. Included 7 patients with metastatic (M+) disease | Long-course CRT (pelvic radiochemotherapy; details unspecified) | Dworak Classification; Response (Grades 3-4) *vs.* No Response (Grades 0-2) | Overall response-classification accuracy of 73.5% using 11-fold cross-validation; CI not reported. |
| Zhang *et al.* (2024)[23] | China; Retrospective | *N* = 294 (Train: 244, Ext Test: 50) | Pretreatment MRI - Clinical stage II or III (cT3/4 or cN1/2) disease | Mixed (Short-course RT, Long-course CRT, or RAPIDO-style TNT) | TRG scale; pCR *vs.* non-pCR | XGBoost AUC of 1.000 (0.999-1.000) in the internal training dataset and 0.908 (0.815-1.000) in external validation |
| Lou *et al.* (2022)[17] | China; Retrospective | *N* = 885 (Train: 666, Internal Test: 117, Ext Val: 102) | Locally advanced (cT2-T4, cN0-N2) | Long-course CRT (50 Gy pelvic RT + concurrent FOLFOX/FOLFIRI) | pCR *vs.* non-pCR based on their postoperative pathological diagnosis | Internal testing AUC of 0.710 (0.595-0.808) and external validation AUC of 0.723 (0.591-0.844) |
| Wang *et al.* (2023)[19] | China; Retrospective | *N* = 146 overall; 500-tile analysis: 69 patients (training: 55; internal validation: 14) | LARC (cT2-T4, cN0-N2) | TNT (Induction FOLFOX/CAPEOX + long-course pelvic CRT) | TRG scale; Good Response (TRG 0-1) and a Poor Response (TRG 2-3) | Tile-level classifier achieved a primary AUC of 0.779 and an internal validation AUC of 0.733; CIs not reported |
| Lafarge *et al.* (2024)[16] | UK & Austria; Retrospective | *N* = 169 (External test set, trained on 1,057 pan-colorectal WSIs) | States "high-risk" or "advanced" rectal cancer | Long-course CRT (45-50.4 Gy pelvic RT + concurrent Capecitabine) | pCR *vs.* non-pCR not further classified | Adjusted ORs for pCR were 2.69 (1.01-7.17) for imCMS1 and 0.25 (0.07-0.88) for imCMS4 |
| Wu *et al.* (2023)[20] | China; Retrospective | *N* = 94; 133 WSIs (training: 113; testing: 20). Patient counts by split and patient-level separation were not clearly reported | LARC (cT2-cT4) | Mixed (Preoperative chemotherapy ± concurrent pelvic RT) | AJCC TRG system; TRG0 (pCR), TRG1/2 (moderate/minimal response), and TRG3 (poor response) | Slide-level macro-AUC of 0.81 on the held-out internal test set; CI not reported |
| Jepsen *et al.* (2024)[15] | Denmark; Retrospective | *N* = 50 (Single-cohort association analysis; pre-trained on an independent set of 35 WSIs) | cT3-T4; MMR-proficient via MRI. Included 32% (*n* = 16) with cM1 distant metastasis | Mixed (Long-course CRT, short-course RT, or chemo/RT alone) | Mandard TRG system; Complete Response (Mandard TRG 1) *vs.* other responders, and Good Response (Mandard TRG 1-2) *vs.* poor response (Mandard TRG 3-5) | Per 10 percentage point increase in $CD8^+/CD3^+$ ratio, ORs were 2.15 (1.19-4.41) for good response and 2.19 (1.12-4.87) for complete response. Per 100 cells/$mm^2$ increase in $CD8^+$ density, corresponding ORs were 1.20 (1.01-1.46) and 1.33 (1.10-1.70) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Shao *et al.* (2020)[18] | China; Retrospective | *N* = 981 (Train: 303, Ext Val 1: 480, Ext Val 2: 150, Ext Val 3: 48) | cT1-T4, cN0-N2; via MRI | Long-course CRT (50 Gy GTV/45 Gy CTV in 25 fractions + concurrent 5-FU) | AJCC/CAP TRG system; TRG 0 *vs.* TRG 1-3 and good response (TRG 0-1) *vs.* non-GR | Pathomics signature: primary AUCs of 0.929 for TRG 0 *vs.* TRG 1-3 and 0.897 for good response; pooled external validation AUCs of 0.895 and 0.770, respectively. AUC CIs not reported. Pooled external four-class TRG accuracy was 62.0% (58.28%-65.72%) |
| Zhang *et al.* (2022)[22] | China; Retrospective | 150 WSIs (training: 90; internal validation: 30; internal testing: 30). Number of unique patients not explicitly reported | LARC without further specification | Long-course CRT (Standard pelvic nCRT; details unspecified) | AJCC TRG system; Good Response (TRG 0-1) and a non-Good Response (TRG 2-3) | Internal test AUC of 0.7589 and accuracy of 71.67% for differentiating good from non-good response; CIs not reported |
| Zhang *et al.* (2020)[21] | China; Retrospective | *N* = 151 (Primary: 120; Internal validation: 31) | LARC (T2-T4 and N0-N2) | Mixed: Long-course CRT or Induction TNT | TRG scale; non-poor response (TRG 0-1) and poor response (TRG 2-3) | Pathology signature achieved a primary AUC of 0.930 (0.883-0.966) and an internal validation AUC of 0.877 (0.719-0.970) |

Where reported, 95% confidence intervals are presented in parentheses after performance estimates and ORs. ORs describe response associations rather than classifier discrimination. No included publication evaluated whether AI-guided treatment decisions improved patient outcomes. AI: Artificial intelligence; AJCC: American Joint Committee on Cancer; AUC: area under the receiver operating characteristic curve; CAP: College of American Pathologists; CAPEOX: capecitabine and oxaliplatin; $CD8^+$: cluster of differentiation 8-positive; CI: confidence interval; CRT: chemoradiotherapy; cT/cN/cM: clinical tumour/node/metastasis staging; cTNM: clinical tumour-node-metastasis; CTV: clinical target volume; Ext: external; FOLFIRI: folinic acid, fluorouracil, and irinotecan; FOLFOX: folinic acid, fluorouracil, and oxaliplatin; GR: good response; GTV: gross tumour volume; imCMS: image-based consensus molecular subtype; LARC: locally advanced rectal cancer; MMR: mismatch repair; MRI: magnetic resonance imaging; N: number of patients; nCRT: neoadjuvant chemoradiotherapy; OR: odds ratio; pCR: pathological complete response; RAPIDO: Rectal cancer And Preoperative Induction therapy followed by Dedicated Operation; RT: radiotherapy; Test: testing dataset; TNT: total neoadjuvant therapy; Train: training dataset; TRG: tumour regression grade; Val: validation; WSI: whole-slide image; XGBoost: extreme gradient boosting; 5-FU: 5-fluorouracil.

## AI architecture and algorithm design

The computational approaches generally fell into two distinct methodological classes: classical machine learning and advanced deep learning [Table 2]. Classical pipelines (such as those utilising XGBoost or Support Vector Machines) required an intermediate processing step using software to extract hundreds of nuclear and morphometric features before the algorithm could predict a response[13,18,21,23]. In Feng *et al.*, the pathomics nucleus model used classical machine learning [CellProfiler feature extraction with a support vector machine (SVM) classifier], whereas the pathomics microenvironment model used a hybrid approach combining visual geometry group (VGG)-19 convolutional neural network feature extraction with an SVM classifier[13].

Conversely, deep learning frameworks (specifically CNNs and MIL networks) were designed to autonomously learn predictive patterns directly from raw image pixels[13-17,19,20,22]. A major computational challenge in digital pathology is the gigapixel size of WSIs. To prevent the AI from losing the global spatial context of the tissue when making predictions on small image patches, MIL and Graph Neural Networks (GNNs) were deployed[17,19,22]. These architectures represent a potential practical advantage, as they allow algorithms to process slides of entire biopsies and identify complex spatial relationships within the tumour microenvironment without requiring pixel-by-pixel human annotation.

## Model explainability

A significant barrier to adoption of AI into clinical workflow is the "black box" nature of deep learning, which results in output predictions without interpretable clinical rationale. Most studies in this review integrated explainability protocols [Table 2], though the depth of true biological validation varied. Many

**Table 2. Computational methodology, AI architectures, and image processing techniques**

| Study | Algorithm class | Specific architecture | Magnification scale | Image input/patch size | Tissue segmentation method | Explainability method |
|---|---|---|---|---|---|---|
| Feng *et al.* (2022)[13] | Classical Machine Learning & Deep Learning | CellProfiler + SVM (nucleus model); VGG-19 feature extraction + SVM (microenvironment model) | Not explicitly specified | Not explicitly specified | Manual segmentation of tumour regions using a standard pipeline | Heat maps with Pearson coefficients and RFE |
| Foersch *et al.* (2023)[14] | Deep Learning | ResNet-18 with gating-based attention module | 20× | 520 × 520 pixels | Automated TMA core segmentation using QuPath | Guided Grad-CAM |
| Zhang *et al.* (2024)[23] | Classical Machine Learning | XGBoost | 400× | 1,000 × 1,000 pixel "snapshot" | Manual review to identify focal areas and select a single representative snapshot | SHAP values |
| Lou *et al.* (2022)[17] | Deep Learning | Multi-Instance Learning (ResNet-18, K-means, MI-FCM) | 20× | 299 × 299 pixels | Hand-delineated by pathologists using ImageScope software | Grad-CAM, t-SNE clustering, and Raster Fairy method |
| Wang *et al.* (2023)[19] | Deep Learning | CNN combined with GNN | Not explicitly specified | 224 × 224 pixels | Automated (WSIs cut directly into smaller blocks without manual ROI drawing) | Visual assessment of histomorphological features |
| Lafarge *et al.* (2024)[16] | Deep Learning | ResNet with multi-instance collective assumption | Scanned at 20×; model tiles extracted at 5× | 318 × 318 pixels | Manual annotation by a pathologist, explicitly excluding folds and debris | Slide-level classification maps and transcriptomic deconvolution (MCP-counter, xCell) |
| Wu *et al.* (2023)[20] | Deep Learning | ResNet-101 | 40× and 10× | 512 × 512 pixels (resized to 256 × 256 for input) | Manual annotation using QuPath, splitting into overlapping tiles with SPCN | CAM |
| Jepsen *et al.* (2024)[15] | Deep Learning | DeepLabv3+ (segmentation) and U-Net (cell counting) | 20× | Not explicitly specified | Manual outline on H&E by a GI pathologist, digitally aligned with IHC using the Tissue Align module | Direct biological quantification of aligned IHC cell populations |
| Shao *et al.* (2020)[18] | Classical Machine Learning | XGBoost | Delineated at 10×, rechecked at 20× | 512 × 512 pixels | Manual delineation by multiple pathologists independently rechecked to ensure boundary validity | XGBoost average gain (Gain) and Spearman test |
| Zhang *et al.* (2022)[22] | Deep Learning | Multi-Instance Learning (ResNet-18) with bilinear attention multi-scale feature fusion | 20× and 5× | Pyramid multi-scale sampling | Automated Otsu binarization algorithm | Gated attention weight matrices |
| Zhang *et al.* (2020)[21] | Classical Machine Learning | SVM and Logistic Regression | Downsampled from 400× to 200× (20×) | 512 × 512 pixels | Rough manual annotation cropped into bounding rectangles, downsampled, and manually selected for dense tumour cells | LASSO |

AI: Artificial intelligence; CAM: class activation mapping; CNN: convolutional neural network; GI: gastrointestinal; GNN: graph neural network; Grad-CAM: gradient-weighted class activation mapping; H&E: haematoxylin and eosin; IHC: immunohistochemistry; LASSO: least absolute shrinkage and selection operator; MCP-counter: microenvironment cell populations-counter; MI-FCM: multi-instance fuzzy c-means; ResNet: residual network; RFE: recursive feature elimination; ROI: region of interest; SHAP: SHapley additive exPlanations; SPCN: structure-preserving colour normalization; SVM: support vector machine; TMA: tissue microarray; t-SNE: t-distributed stochastic neighbour embedding; VGG: visual geometry group; WSI: whole-slide image; XGBoost: extreme gradient boosting.

deep learning architectures utilised gradient-based spatial heatmaps [such as Grad-class activation mapping (CAM)][14,17,20] to visually highlight the tissue regions driving the prediction. While these heatmaps illustrated that model attention was concentrated on histologically relevant regions (such as tumour nuclei and

tumour-infiltrating lymphocytes rather than background stroma), they did not explicitly define the underlying cellular biology.

To bridge the gap between mathematical outputs and true tumour biology, three studies utilised external molecular or immunohistochemical validation[14-16]. For example, Lafarge *et al.* utilised transcriptomic deconvolution to demonstrate that their AI-derived visual classifications were associated with independently estimated immune and fibroblast cell populations[16]. Similarly, Jepsen *et al.* built biological explainability directly into their image processing pipeline[15]. By digitally aligning standard H&E slides with IHC stains, they showed that their algorithm quantified cluster of differentiation 8-positive ($CD8^+$) cytotoxic T-cell infiltration associated with treatment response.

## Model performance

Evaluating and directly comparing the predictive efficacy of AI models across the literature is inherently challenging due to significant heterogeneity in study designs, the utilisation of internal *vs.* external validation sets, and the variance in target clinical endpoints. Performance metrics are therefore presented according to each study's target outcome and validation design rather than being used to rank the different model architectures.

Performance estimates for models predicting pCR are summarised in Table 1. Zhang *et al.* (2024) reported an area under the receiver operating characteristic curve (AUC) of 1.000 in the internal training dataset[23]. In the independent external cohort, the detailed Results and published receiver operating characteristic curves of the original publication report an XGBoost AUC of 0.908 (95%CI: 0.815-1.000), although the source abstract reports 0.950 for the same cohort. We therefore used 0.908 throughout this review. Feng *et al.* evaluated their multimodal RAdioPathomics Integrated preDiction System (RAPIDS) model, incorporating pathological and MRI-derived features, in retrospective external cohorts and a prospective multicentre cohort[13]. For the standalone pathomics models, the classical machine-learning nucleus model reported retrospective external validation AUCs of 0.811 and 0.832 and a prospective validation AUC of 0.733, while the hybrid microenvironment model reported retrospective external validation AUCs of 0.656 and 0.643 and a prospective validation AUC of 0.630.

Lou *et al.* evaluated a MIL model, which reported an AUC of 0.710 (sensitivity: 72.6%, Specificity: 46.9%) in internal testing and 0.723 (sensitivity: 72.5%, Specificity: 62.7%) in an external validation cohort[17]. Two studies reported odds ratios (OR) as efficacy measures rather than AUC. Jepsen *et al.* demonstrated that a 10 percentage point increase in the ratio of $CD8^+/CD3^+$ lymphocytes calculated by their algorithm increased the odds of achieving a pCR by 2.19 times[15]. Additionally, they found a 100 cell/$mm^2$ increase in raw $CD8^+$ density alone increased the odds of a complete response by 33% (OR 1.33). Similarly, Lafarge *et al.* reported that biopsies classified by their ensemble model as the "imCMS1" subtype were significantly more likely to achieve pCR (OR 2.69), whereas those classified as "imCMS4" had lower odds of pCR (OR 0.25)[16].

When models were tasked with differentiating broader binary outcomes based on TRG, predictive performance remained discriminatory but varied based on the classification thresholds. Zhang *et al.* (2020) achieved an AUC of 0.930 in the primary dataset and 0.877 in an internal validation dataset when discriminating a poor response (TRG 2-3) from a non-poor response (TRG 0-1) using a logistic regression-based pathology signature[21]. Using a similar TRG threshold, Wang *et al.* evaluated a GNN tile-level classifier, achieving an AUC of 0.779 in the primary set and 0.733 in the internal validation set[19].

Zhang *et al.* (2022) utilised a multi-scale MIL model to separate "Good" from "non-Good" responses, achieving an AUC of 0.7589 and an overall accuracy of 71.67% in an internal test set[22]. Foersch *et al.*[14] evaluated response *vs.* no response utilising dichotomisation of the Dworak classification system, achieving

an overall accuracy of 73.5%, with a true predictive value of 78.1% for identifying non-responders. Finally, Shao *et al.* reported pooled external validation AUCs of 0.895 for TRG 0 *vs.* TRG 1-3 and 0.770 for good response (TRG 0-1 *vs.* TRG 2-3) using their pathomics signature[18].

Wu *et al.* developed a multiclass algorithm capable of predicting three distinct TRG categories[20]. Evaluated at 40× magnification, the model achieved a slide-level macro-AUC of 0.81 on the held-out internal test set.

Confidence intervals for response predictive performance were reported in five publications (5/11), although reporting was incomplete for some models and endpoints; two additional publications reported confidence intervals for response associations. Calibration of treatment-response predictions was reported in one publication (1/11), Zhang *et al.* (2020), which also reported decision curve analysis (1/11)[21]. Zhang *et al.* (2024) reported calibration of a separate disease-free survival nomogram rather than the treatment-response classifier[23]. For standalone pathology models, only two publications (2/11) reported sensitivity, specificity, positive predictive value, and negative predictive value[13,17], while only one publication (1/11), Lou *et al.*, reported them alongside an explicit numerical probability threshold (> 0.7)[17]; Shao *et al.* reported these metrics only for their multimodal radiopathomics model[18]. These operating characteristics are summarised in Supplementary Table 2. Regarding clinical utility, no publication (0/11) evaluated whether AI-guided treatment decisions improved patient outcomes.

## DISCUSSION

This scoping review evaluated AI and digital pathology applied to the preoperative prediction of neoadjuvant therapy response for LARC. Across the included studies, reported discrimination varied by model class, clinical endpoint, and validation design. Standalone classical machine-learning models reported external validation AUCs of 0.811-0.908 for predicting pathological response[13,18,23], a hybrid VGG-19 feature-extraction and SVM microenvironment model reported external validation AUCs of 0.643-0.656[13], and Lou *et al.* reported an external validation AUC of 0.723 for a standalone end-to-end deep-learning model using MIL[17]. For descriptive context, previously published conventional clinical prediction models have reported validation AUCs of 0.60-0.70[24-26], and a recent meta-analysis of MRI radiomics models reported a pooled AUC of 0.91[27]. However, these figures are presented for context only and cannot be directly compared, as the underlying studies differed in their patient populations, clinical endpoints, validation designs, and analytical methods, and no formal statistical comparison was performed. While our findings highlight the potential of using H&E slide digital pathology to help inform neoadjuvant therapy decisions in rectal cancer, this review also identified critical methodological gaps that must be addressed before translating these algorithms into routine clinical practice, including the development of standardized clinical endpoints, assessment of calibration, and biological explainability.

The reviewed studies demonstrated significant methodological heterogeneity with regard to pre-processing and model architecture. Image pre-processing ranged from manual delineation of tumour regions by pathologists to fully automated segmentation pipelines. Following pre-processing, two distinct architectural approaches emerged: classical machine learning and deep learning. Classical models required the explicit extraction of sub-visual nuclear and cellular morphological features prior to analysis. Conversely, advanced deep learning frameworks (including CNNs, MIL, and GNNs) were designed to autonomously process raw pixel data and capture whole-slide spatial context. Notably, none of the reviewed cohorts deployed modern self-supervised vision transformers or digital pathology foundation models, representing a key developmental gap in the current literature. These transformer-based architectures excel at modelling long-range, multi-scale spatial dependencies across massive gigapixel WSIs and provide a promising opportunity for advancement of the field.

Reported performance varied across the different architectural approaches. Among classical machine-learning pathomics models, external validation AUCs were 0.908 in Zhang *et al.* (2024)[23], 0.895 in Shao *et al.*[18] (for TRG 0 *vs.* TRG 1-3), and 0.811-0.832 for the nuclear pathomics model in Feng *et al.*[13], while Feng *et al.*'s hybrid VGG-19/SVM microenvironment model reported external validation AUCs of 0.643-0.656[13]. Among standalone deep-learning models, Lou *et al.* reported an internal test AUC of 0.710 and an external validation AUC of 0.723[17], while Wang *et al.*[19] reported an AUC of 0.779 in the primary set and 0.733 in an internal validation set[19]. These between-study values should not be interpreted as evidence of architectural superiority because the studies differed substantially in their populations, sample sizes, endpoints, preprocessing methods, and validation designs. No included study compared classical machine learning and deep learning for neoadjuvant response prediction using a common dataset. This review therefore cannot determine whether one architecture performs better than another or establish mechanistic explanations for the observed differences.

Integration of AI-driven digital pathology into routine surgical oncology workflow requires careful consideration and robust, explainable methodology. This review identified substantial heterogeneity in study design, inconsistently and poorly defined clinical eligibility, with some studies using pretreatment MRI staging to define inclusion whereas others reported only ypTNM. The absence of accurate pretreatment clinical staging data introduces clear limitations when evaluating pretreatment prediction models. Clinical endpoints also varied substantially. Although pCR and cCR are established response measures, they are not interchangeable: pCR is determined following surgical resection, whereas cCR is directly relevant to consideration of organ-preserving management. Several studies used alternative response categories, such as "good response" defined as TRG 0-1, which limits direct comparison and clinical applicability. Future studies should prospectively predefine clinically meaningful endpoints and include outcomes relevant to watch-and-wait strategies, such as sustained cCR, local regrowth, organ-preservation rate, and salvage surgery, rather than relying predominantly on pCR as a surrogate endpoint[4]. The "black box" nature of deep learning also remains a barrier to clinical trust, and incorporation of explainability is a key requirement going forward. Although spatial heatmaps identified tissue regions influencing model predictions, most studies did not link these regions to independently validated cellular biology. Finally, the evidence base was predominantly retrospective. External validation strengthened several studies, but prospective multicentre validation remains necessary.

Prediction errors have different clinical consequences. A false-positive prediction of complete response could support inappropriate watch-and-wait management despite residual disease, making specificity and positive predictive value particularly important. Conversely, a false-negative prediction could contribute to radical resection in a patient who might preserve the rectum, highlighting the importance of sensitivity. AUC alone cannot establish safety at a chosen threshold, and model predictions should complement clinical response assessment rather than independently determine organ preservation.

Generalisability may also be affected by dataset shift across institutions, including differences in tissue fixation, staining, slide preparation, scanner hardware, image resolution, and image-acquisition protocols. Pretreatment biopsy samples may also provide incomplete representation of spatial and intratumoural heterogeneity. Future studies should therefore evaluate model robustness across institutions and scanner platforms and report calibration alongside discrimination. Local recalibration may be required before implementation in a new clinical setting.

AI-based digital pathology may also identify morphological, molecular, and spatial features associated with immune responsiveness, including immune-cell infiltration, spatial immune architecture, and potentially deficient mismatch repair/microsatellite instability-high (dMMR/MSI-H) status. This may become clinically

relevant to immune-checkpoint blockade in selected patients with dMMR/MSI-H LARC. However, prediction of immunotherapy response was not evaluated in the studies included in this review and remains an area for future research[28].

Heterogeneity in clinical endpoints and study design presented inherent limitations to this scoping review and precluded meta-analysis. Additionally, because the intersection of AI and digital pathology lacks universally standardised clinical and computational terminology, establishing highly specific search criteria proved challenging. Consequently, this introduces the possibility that some relevant studies or alternative algorithmic architectures may have been inadvertently excluded from our final analysis.

In regard to limitations of this scoping review, the protocol was registered retrospectively after completion of the literature search and therefore could not prospectively constrain the review methods. In addition, using a common free-text query without database-specific controlled vocabulary or field tags may have reduced search sensitivity and resulted in relevant studies being missed. Data charting was also conducted by one reviewer without duplicate extraction or independent verification, increasing the possibility of extraction or transcription errors.

Evaluating biopsies in isolation may ultimately limit their clinical reliability. For instance, Shao *et al.* reported lower discrimination when their pathomics model predicted broader response categories[18]; integrating multiparametric MRI (mp-MRI) data improved discrimination for good response, with a multimodal radiopathomics validation AUC of 0.933. Feng *et al.* also evaluated a multimodal pipeline integrating WSI features with radiological imaging, including prospective multicentre validation[13]. However, three retrospective cohorts were shared with Shao *et al.*[18], so these publications cannot be considered fully independent evidence of the value of multimodal integration. Nevertheless, multimodal integration may improve the predictive performance of pathomics models by capturing both macroscopic and microstructural features associated with tumour response.

## CONCLUSION

The integration of AI and digital pathology offers a promising preoperative tool for predicting neoadjuvant treatment response in LARC. Although current pathomics and multimodal architectures have demonstrated promising predictive performance, clinical implementation remains limited. In surgical oncology, critical decisions regarding treatment de-escalation or radical resection cannot be justified safely by uninterpretable algorithmic outputs. To support shared clinical decision-making, AI predictions must move beyond mathematical probabilities and be anchored in verified tumour biology. Ultimately, the successful translation of these models into routine clinical practice requires rigorous, prospective multicentre trials and the integration of transparent, biologically sound explainability frameworks directly into the algorithmic design.

## DECLARATIONS

### Authors' contributions

Conception and design of the review, literature search, data extraction, data analysis and interpretation, and drafting of the manuscript: Allen W

Study selection and critical review of the manuscript: Tranter-Entwistle I.

Critical review, editorial input, and expert advice on the interpretation and presentation of the findings: Fischer J, Atlas J, Pearson J, Morley-Bunker A, Drake M, Vernmark K, Eglinton T

All authors reviewed and approved the final manuscript.

### Availability of data and materials

Not applicable.

### AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tools Google Gemini (version Gemini 3 Flash, released

2025-12-17, and, version Gemini 3.5 Flash, released 2026-05-19) and Google NotebookLM (web application) were used to assist with language editing and improving readability. During revision, the OpenAI Codex desktop application (version GPT-5.6 Sol, released 2026-07-09) was used to assist with language editing. ChatGPT Images (OpenAI, version 2.5, released 2026-09-08) was used to generate the graphical abstract from an author-supplied prompt. The graphical abstract is a schematic illustration and contains no patient data or original study images. The tools did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

### Financial support and sponsorship

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Funding information for the individual papers in this scoping review is provided in Supplementary Table 3.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2026.

### Supplementary Materials

Supplementary Materials

## REFERENCES

1. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer; 2024. Available from: https://gco.iarc.who.int/today/. [Last accessed on 9 Nov 2026].
2. Bahadoer RR, Dijkstra EA, van Etten B, et al. Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. *Lancet Oncol.* 2021;22:29-42. DOI PubMed
3. Conroy T, Bosset JF, Etienne PL, et al. Neoadjuvant chemotherapy with FOLFIRINOX and preoperative chemoradiotherapy for patients with locally advanced rectal cancer (UNICANCER-PRODIGE 23): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol.* 2021;22:702-15. DOI PubMed
4. Verheij FS, Omer DM, Williams H, et al. Long-term results of organ preservation in patients with rectal adenocarcinoma treated with total neoadjuvant therapy: the randomized phase II OPRA trial. *J Clin Oncol.* 2024;42:500-6. DOI PubMed PMC
5. Peng H, Wang C, Xiao W, et al. Analysis of clinical characteristics to predict pathologic complete response for patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *J Cancer.* 2018;9:2687-92. DOI PubMed PMC
6. Patel UB, Taylor F, Blomqvist L, et al. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol.* 2011;29:3753-60. DOI PubMed
7. Miranda J, Causa Andrieu P, Nincevic J, et al. Advances in MRI-based assessment of rectal cancer post-neoadjuvant therapy: a comprehensive review. *J Clin Med.* 2023;13:172. DOI PubMed PMC
8. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol.* 2019;20:e253-61. DOI PubMed PMC
9. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* 2019;25:1054-6. DOI PubMed PMC
10. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer.* 2021;124:686-96. DOI PubMed PMC
11. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* 2018;169:467-73. DOI PubMed

12. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. *Syst Rev.* 2016;5:210. DOI PubMed PMC
13. Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health.* 2022;4:e8-17. DOI PubMed
14. Foersch S, Glasner C, Woerl AC, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med.* 2023;29:430-9. DOI PubMed
15. Jepsen DNM, Høeg H, Bzorek M, et al. Digitally assessed lymphocyte infiltration in rectal cancer biopsies is associated with pathological response to neoadjuvant therapy. *Hum Pathol.* 2024;144:61-70. DOI PubMed
16. Lafarge MW, Domingo E, Sirinukunwattana K, et al. Image-based consensus molecular subtyping in rectal cancer biopsies and response to neoadjuvant chemoradiotherapy. *NPJ Precis Oncol.* 2024;8:89. DOI PubMed PMC
17. Lou X, Zhou N, Feng L, et al. Deep learning model for predicting the pathological complete response to neoadjuvant chemoradiotherapy of locally advanced rectal cancer. *Front Oncol.* 2022;12:807264. DOI PubMed PMC
18. Shao L, Liu Z, Feng L, et al. Multiparametric MRI and whole slide image-based pretreatment prediction of pathological response to neoadjuvant chemoradiotherapy in rectal cancer: a multicenter radiopathomic study. *Ann Surg Oncol.* 2020;27:4296-306. DOI PubMed PMC
19. Wang A, Ding R, Zhang J, Zhang B, Huang X, Zhou H. Machine learning of histomorphological features predict response to neoadjuvant therapy in locally advanced rectal cancer. *J Gastrointest Surg.* 2023;27:162-5. DOI PubMed
20. Wu Y, Liu X, Liu F, et al. A multi-class classification algorithm based on hematoxylin-eosin staining for neoadjuvant therapy in rectal cancer: a retrospective study. *PeerJ.* 2023;11:e15408. DOI PubMed PMC
21. Zhang F, Yao S, Li Z, et al. Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. *Clin Transl Med.* 2020;10:e110. DOI PubMed PMC
22. Zhang D, Duan Y, Guo J, et al. Using multi-scale convolutional neural network based on multi-instance learning to predict the efficacy of neoadjuvant chemoradiotherapy for rectal cancer. *IEEE J Transl Eng Health Med.* 2022;10:4300108. DOI PubMed PMC
23. Zhang Y, Huang Y, Xu M, et al. Pathomics-based machine learning models for predicting pathological complete response and prognosis in locally advanced rectal cancer patients post-neoadjuvant chemoradiotherapy: insights from two independent institutional studies. *BMC Cancer.* 2024;24:1580. DOI PubMed PMC
24. Fischer J, Eglinton TW, Richards SJ, Frizelle FA. Predicting pathological response to chemoradiotherapy for rectal cancer: a systematic review. *Expert Rev Anticancer Ther.* 2021;21:489-500. DOI PubMed
25. Mbanu P, Saunders MP, Mistry H, et al. Clinical and radiomics prediction of complete response in rectal cancer pre-chemoradiotherapy. *Phys Imaging Radiat Oncol.* 2022;23:48-53. DOI PubMed PMC
26. Tanaka MD, Geubels BM, Grotenhuis BA, et al. Validated pretreatment prediction models for response to neoadjuvant therapy in patients with rectal cancer: a systematic review and critical appraisal. *Cancers.* 2023;15:3945. DOI PubMed PMC
27. Liao Z, Luo D, Tang X, Huang F, Zhang X. MRI-based radiomics for predicting pathological complete response after neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review and meta-analysis. *Front Oncol.* 2025;15:1550838. DOI PubMed PMC
28. Olawade DB, David-Olawade AC, Adereni T, Egbon E, Teke J, Boussios S. Integrating AI into cancer immunotherapy - a narrative review of current applications and future directions. *Diseases.* 2025;13:24. DOI PubMed PMC

**Disclaimer/Publisher's Note:** All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.