SUPPLEMENTARY MATERIALS

Supplementary Table S1. Complete search strategies and final record yields for all four databases

Search	PubMed Query - May 17,2026	Items found
#19	#13 AND #17 AND #18	1,150
#18	"Magnetic Resonance Imaging"[MeSH Terms] OR "MRI Scans"[Title/Abstract] OR "MRI"[Title/Abstract] OR "CT"[Title/Abstract] OR "computed tomography"[Title/Abstract] OR "PET"[Title/Abstract] OR "positron emission tomography"[Title/Abstract] OR "multimodal imaging"[Title/Abstract] OR "multiparametric"[Title/Abstract] OR "diffusion-weighted"[Title/Abstract] OR "DWI"[Title/Abstract] OR "T2-weighted"[Title/Abstract]	46,593
#17	#14 OR #15 OR #16	30,032
#16	"Rectal Neoplasms"[MeSH Terms] OR "rectal neoplasms"[Title/Abstract] OR "rectal neoplasm"[Title/Abstract] OR "rectum neoplasms"[Title/Abstract] OR "rectum neoplasm"[Title/Abstract] OR "rectal tumors"[Title/Abstract] OR	3,141

#15	"rectal tumor"[Title/Abstract] OR "cancer of rectum"[Title/Abstract] OR "rectum cancers"[Title/Abstract] OR "rectal cancer"[Title/Abstract] OR "rectal cancers"[Title/Abstract] OR "rectum cancer"[Title/Abstract] OR "cancer of the rectum"[Title/Abstract] OR "LARC"[Title/Abstract] "Neoadjuvant Therapy"[MeSH Terms] OR "Chemoradiotherapy"[MeSH Terms] OR "neoadjuvant chemoradiotherapy"[Title/Abstract] OR "neoadjuvant chemoradiation"[Title/Abstract] OR "neoadjuvant radiotherapy"[Title/Abstract] OR "neoadjuvant therapy"[Title/Abstract] OR "nCRT"[Title/Abstract] OR "preoperative chemoradiotherapy"[Title/Abstract] OR "preoperative chemoradiation"[Title/Abstract] OR "preoperative radiotherapy"[Title/Abstract] OR "chemoradiation"[Title/Abstract]	8,108
#14	"pathological complete response"[Title/Abstract] OR "pathologic complete response"[Title/Abstract] OR "pCR"[Title/Abstract] OR "complete pathological response"[Title/Abstract] OR "complete pathologic response"[Title/Abstract] OR "ypT0N0"[Title/Abstract] OR "tumour regression grade"[Title/Abstract] OR "tumor regression	9,333

	grade"[Title/Abstract] OR "TRG"[Title/Abstract]	
#13	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12	29,766
#12	"radiomics"[Title/Abstract] OR "radiomic"[Title/Abstract] OR "texture analysis"[Title/Abstract] OR "imaging biomarker"[Title/Abstract]	28,478
#11	"resnet model"[Title/Abstract] OR "gradient boosting machine"[Title/Abstract] OR "GBM"[Title/Abstract] OR "NNet"[Title/Abstract] OR "latent dirichlet allocation"[Title/Abstract] OR "LDA"[Title/Abstract] OR "LASSO"[Title/Abstract] OR "boosting"[Title/Abstract] OR "XGBoost"[Title/Abstract] OR "radiogenomic*"[Title/Abstract]	126,763
#10	"random forest"[Title/Abstract] OR "random forest classifier"[Title/Abstract] OR "RFC"[Title/Abstract] OR "RF"[Title/Abstract] OR "CNN"[Title/Abstract] OR "convolutional neural network"[Title/Abstract]	144,632
#9	"computer neural network"[Title/Abstract] OR "computer neural networks"[Title/Abstract] OR "neural network model"[Title/Abstract] OR "neural network models"[Title/Abstract] OR "computational neural	19,831

	networks"[Title/Abstract] OR "computational neural network"[Title/Abstract] OR "Perceptrons"[Title/Abstract] OR "Perceptron"[Title/Abstract] OR "connectionist models"[Title/Abstract] OR "connectionist model"[Title/Abstract]	
#8	"Logistic Models"[MeSH Terms] OR "logistic models"[Title/Abstract] OR "logistic model"[Title/Abstract] OR "logit models"[Title/Abstract] OR "logit model"[Title/Abstract] OR "logistic regression"[Title/Abstract] OR "logistic regressions"[Title/Abstract]	661,964
#7	"Decision Trees"[MeSH Terms] OR "decision tree"[Title/Abstract] OR "decision trees"[Title/Abstract] OR "DT"[Title/Abstract]	67,319
#6	"Support Vector Machine"[MeSH Terms] OR "support vector machine"[Title/Abstract] OR "support vector machines"[Title/Abstract] OR "support vector network"[Title/Abstract] OR "support vector networks"[Title/Abstract] OR "SVM"[Title/Abstract]	47,751
#5	"Unsupervised Machine Learning"[MeSH Terms]	2,605
#4	"Supervised Machine Learning"[MeSH Terms] OR "supervised machine learning"[Title/Abstract] OR "semi supervised learning"[Title/Abstract] OR "semi-supervised	22,488

	learning"[Title/Abstract] OR "inductive machine learning"[Title/Abstract] OR "active machine learning"[Title/Abstract] OR "learning from labeled data"[Title/Abstract]	
#3	"Deep Learning"[MeSH Terms] OR "deep learning"[Title/Abstract] OR "hierarchical learning"[Title/Abstract]	128,633
#2	"Machine Learning"[MeSH Terms] OR "machine learning"[Title/Abstract] OR "transfer learning"[Title/Abstract]	274,883
#1	"Artificial Intelligence"[MeSH Terms] OR "artificial intelligence"[Title/Abstract] OR "computational intelligence"[Title/Abstract] OR "machine intelligence"[Title/Abstract] OR "computer reasoning"[Title/Abstract] OR "AI"[Title/Abstract] OR "computer vision systems"[Title/Abstract] OR "computer vision system"[Title/Abstract] OR "knowledge acquisition"[Title/Abstract] OR "knowledge representation"[Title/Abstract] OR "knowledge representations"[Title/Abstract]	435,398

Search	Embase Query - May 17,2026	Items found
#18	#16 AND #17	980
#17	Magnetic Resonance Imaging':ab,ti OR 'MRI Scans':ab,ti OR 'MRI':ab,ti OR 'CT':ab,ti OR 'computed tomography':ab,ti OR 'PET':ab,ti OR 'positron emission tomography':ab,ti OR 'multimodal imaging':ab,ti OR 'multiparametric':ab,ti OR 'diffusion-weighted':ab,ti OR 'DWI':ab,ti OR 'T2-weighted':ab,ti	657,080
#16	#11 AND #15	15,071
#15	#12 OR #13 OR #14	375,594
#14	rectum cancer'/exp OR 'cancer rectum':ab,ti OR 'rectal cancer':ab,ti OR 'rectal malignancy':ab,ti OR 'rectum malignancy':ab,ti OR 'LARC':ab,ti	146,421
#13	neoadjuvant chemoradiotherapy'/exp OR 'neoadjuvant chemoradiotherapy':ab,ti OR 'neoadjuvant chemoradiation':ab,ti OR 'neoadjuvant radiotherapy':ab,ti OR 'neoadjuvant therapy':ab,ti OR 'nCRT':ab,ti OR 'preoperative chemoradiotherapy':ab,ti OR 'preoperative chemoradiation':ab,ti OR 'preoperative radiotherapy':ab,ti OR 'chemoradiation':ab,ti	64,130
#12	pathological complete response'/exp OR 'pathological complete	351,861

	response':ab,ti OR 'pathologic complete response':ab,ti OR 'pCR':ab,ti OR 'complete pathological response':ab,ti OR 'complete pathologic response':ab,ti OR 'ypT0N0':ab,ti OR 'tumour regression grade':ab,ti OR 'tumor regression grade':ab,ti OR 'TRG':ab,ti	
#11	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10	1,111,714
#10	support vector machine'/exp OR 'support vector machine':ab,ti OR 'support vector':ab,ti OR 'support vector classification':ab,ti OR 'support vector classifier':ab,ti OR 'support vector classifiers':ab,ti OR 'support vector machines':ab,ti OR 'support vector network':ab,ti OR 'support vector regression':ab,ti OR 'support vectors':ab,ti OR 'svc':ab,ti OR 'svm':ab,ti OR 'svr':ab,ti	63,875
#9	resnet model':ab,ti OR 'gradient boosting machine':ab,ti OR 'nnet':ab,ti OR 'latent dirichlet allocation':ab,ti OR 'texture':ab,ti OR 'lasso':ab,ti OR 'radiogenomic*' OR 'radiomics'/exp OR 'radiomics':ab,ti	58,964
#8	random forest'/exp OR 'random forest':ab,ti OR 'random decision forest':ab,ti OR 'random forest algorithm':ab,ti OR 'random forests':ab,ti OR 'RF':ab,ti	5,182

	artificial neural network'/exp OR 'artificial neural network':ab,ti OR 'algorithmic neural network':ab,ti OR 'ann approach':ab,ti OR 'ann method':ab,ti OR 'ann methodology':ab,ti OR 'ann methods':ab,ti OR 'ann model':ab,ti OR 'ann modeling':ab,ti OR 'ann modelling':ab,ti OR 'ann models':ab,ti OR 'ann output':ab,ti OR 'ann technique':ab,ti OR 'ann techniques':ab,ti OR 'ann training':ab,ti OR 'anns':ab,ti OR 'artificial neural networks':ab,ti OR 'artificial nns':ab,ti OR 'computational neural network':ab,ti OR 'computer neural network':ab,ti OR 'computerized neural network':ab,ti OR 'connectionist model':ab,ti OR 'connectionist network':ab,ti OR 'connectionist neural network':ab,ti OR 'connectionist system':ab,ti OR 'mathematical neural network':ab,ti OR 'neural network artificial':ab,ti OR 'neural network	
#7	algorithm':ab,ti OR 'neural network model':ab,ti OR 'CNN':ab,ti	22,999
	decision tree'/exp OR 'decision trees':ab,ti OR 'decision tree':ab,ti	
#6	OR 'DT':ab,ti	76,591
	statistical model'/exp OR 'likelihood functions':ab,ti OR 'linear model':ab,ti OR 'linear models':ab,ti OR 'logistic models':ab,ti OR 'statistic model':ab,ti OR 'statistical models':ab,ti OR 'statistics	
#5	model':ab,ti OR 'logistic regression':ab,ti	23,503
#4	deep learning'/exp OR 'hierarchical learning':ab,ti OR 'deep	684,047

	learning':ab,ti	
#3	machine learning'/exp OR 'machine learning':ab,ti OR 'learning machine':ab,ti OR 'learning machines':ab,ti OR 'transfer learning':ab,ti	38,149
#2	artificial intelligence'/exp OR 'artificial intelligence':ab,ti OR 'machine intelligence':ab,ti OR 'computational intelligence':ab,ti	329,566
#1	radiomics'/exp OR 'radiomics':ab,ti OR 'radiomic':ab,ti	81,337

Search	**Web of Science Query - May 17,2026**	**Items found**
#4	#3 AND #2 AND #1	520
#3	TS=(Magnetic Resonance Imaging OR MRI Scans OR MRI OR CT OR computed tomography OR PET OR positron emission tomography OR multimodal imaging OR multiparametric OR diffusion-weighted OR DWI OR T2-weighted)	570,758
#2	TS=(Rectal Neoplasms OR Rectal Neoplasm OR Rectum Neoplasms OR Rectum Neoplasm OR Rectal Tumors OR Rectal Tumor OR Cancer of Rectum OR Rectum Cancers OR Rectal Cancer OR Rectal Cancers OR Rectum Cancer OR Cancer of the Rectum OR LARC OR	345,469

#1	neoadjuvant chemoradiotherapy OR neoadjuvant chemoradiation OR neoadjuvant radiotherapy OR neoadjuvant therapy OR nCRT OR preoperative chemoradiotherapy OR preoperative chemoradiation OR preoperative radiotherapy OR chemoradiation OR pathological complete response OR pathologic complete response OR pCR OR complete pathological response OR complete pathologic response OR ypT0N0 OR tumour regression grade OR tumor regression grade OR TRG) TS=(artificial intelligence OR computational intelligence OR machine intelligence OR Machine Learning OR transfer learning OR deep learning OR hierarchical learning OR supervised machine learning OR semi supervised learning OR inductive machine learning OR Unsupervised Machine Learning OR support vector machine OR support vector network OR SVM OR decision tree OR logistic models OR logistic regression OR computer neural network OR neural network model OR connectionist models OR random forest OR random forest classifier OR RFC OR RF OR CNN OR convolutional neural network OR resnet model OR gradient boosting machine OR GBM OR latent dirichlet allocation OR texture OR LASSO OR boosting OR XGBoost OR radiomics OR radiomic OR imaging biomarker OR predictive model OR	1,732,726

computer-aided OR computer assisted OR algorithm)

Search	The Cochrane CENTRAL Query - May 17,2026	Items found
#25	#13 AND #23 AND #24	197
#24	(Magnetic Resonance Imaging):ti,ab,kw OR (MRI Scans):ti,ab,kw OR (MRI):ti,ab,kw OR (CT):ti,ab,kw OR (computed tomography):ti,ab,kw OR (PET):ti,ab,kw OR (positron emission tomography):ti,ab,kw OR (multimodal imaging):ti,ab,kw OR (multiparametric):ti,ab,kw OR (diffusion-weighted):ti,ab,kw OR (DWI):ti,ab,kw OR (T2-weighted):ti,ab,kw	38,784
#23	#16 OR #19 OR #22	26,464
#22	#20 OR #21	8,053
#21	(Rectal Neoplasms):ti,ab,kw OR (Rectal Neoplasm):ti,ab,kw OR (Rectum Neoplasms):ti,ab,kw OR (Rectum Neoplasm):ti,ab,kw OR (Rectal Tumors):ti,ab,kw OR (Rectal Tumor):ti,ab,kw OR (Cancer of	8,026

	Rectum):ti,ab,kw OR (Rectum Cancers):ti,ab,kw OR (Rectal Cancer):ti,ab,kw OR (Rectal Cancers):ti,ab,kw OR (Rectum Cancer):ti,ab,kw OR (Cancer of the Rectum):ti,ab,kw OR (LARC):ti,ab,kw	
#20	MeSH descriptor: [Rectal Neoplasms] explode all trees	1,913
#19	#17 OR #18	7,939
#18	(neoadjuvant chemoradiotherapy):ti,ab,kw OR (neoadjuvant chemoradiation):ti,ab,kw OR (neoadjuvant radiotherapy):ti,ab,kw OR (neoadjuvant therapy):ti,ab,kw OR (nCRT):ti,ab,kw OR (preoperative chemoradiotherapy):ti,ab,kw OR (preoperative chemoradiation):ti,ab,kw OR (preoperative radiotherapy):ti,ab,kw OR (chemoradiation):ti,ab,kw	7,825
#17	MeSH descriptor: [Neoadjuvant Therapy] explode all trees OR MeSH descriptor: [Chemoradiotherapy] explode all trees	2,025
#16	#14 OR #15	20,382
#15	(pathological complete response):ti,ab,kw OR (pathologic complete response):ti,ab,kw OR (pCR):ti,ab,kw OR (complete pathological response):ti,ab,kw OR (complete pathologic response):ti,ab,kw OR (ypT0N0):ti,ab,kw OR (tumour regression grade):ti,ab,kw OR (tumor regression grade):ti,ab,kw OR (TRG):ti,ab,kw	17,954

#14	MeSH descriptor: [Pathological Complete Response] explode all trees	9,249
#13	#3 OR #6 OR #9 OR #12	6254
#12	#10 OR #11	232
#11	(Semi-supervised Learning):ti,ab,kw OR (Semi supervised Learning):ti,ab,kw OR (Inductive Machine Learning):ti,ab,kw OR (Active Machine Learning):ti,ab,kw OR (Learning from Labeled Data):ti,ab,kw	210
#10	MeSH descriptor: [Supervised Machine Learning] explode all trees	25
#9	#7 OR #8	1,289
#8	(Deep Learning):ti,ab,kw OR (Hierarchical Learning):ti,ab,kw	1,289
#7	MeSH descriptor: [Deep Learning] explode all trees	65
#6	#4 OR #5	3,378
#5	(Machine Learning):ti,ab,kw OR (Transfer Learning):ti,ab,kw	3,307
#4	MeSH descriptor: [Machine Learning] explode all trees	252
#3	#1 OR #2	2,510
#2	(Computational Intelligence):ti,ab,kw OR (Artificial Intelligence):ti,ab,kw OR (Computational Intelligence):ti,ab,kw OR (Machine Intelligence):ti,ab,kw OR (Computer Reasoning):ti,ab,kw	1,327

	OR (AI):ti,ab,kw	
#1	MeSH descriptor: [Artificial Intelligence] explode all trees	1,457

AI: Artificial intelligence; CNN: Convolutional neural network; CT: Computed tomography; DWI: Diffusion-weighted imaging; GBM: Gradient boosting machine; LASSO: Least absolute shrinkage and selection operator; LDA: Latent Dirichlet allocation; MeSH: Medical subject headings; MRI: Magnetic resonance imaging; nCRT: Neoadjuvant chemoradiotherapy; PET: Positron emission tomography; pCR: Pathological complete response; RF: Random forest; SVM: Support vector machine; TRG: Tumour regression grade.

Supplementary Table S2. Studies included in and excluded from the meta-analysis

S2a. Studies included in the meta-analysis (n = 13)

Study	**Country**	**Total cohort (N)**	**Validation cohort (n)**	**Validation type**
Bibault *et al* 2018	France	95	95	Internal
Ju *et al* 2023	South Korea	156	40	External
Wu *et al* 2025	China	768	367	External

Liu *et al* 2017	China	222	70	Temporal
Li *et al* 2021	China	80	80	Internal
Peng *et al* 2023	China	83	83	Internal
Boldrini *et al* 2022	Italy/China	59	59	External
Jin *et al* 2021	China	141	141	External
Cao *et al* 2026	China	181	181	External
Shi *et al* 2023	China	77	77	External
Nie *et al* 2016	China	48	48	Internal
Horvat *et al* 2022	USA/Brazil	50	50	External
Petkovska *et al* 2020	USA	102	102	Internal
Total		**2,062**	**1,393**	

S2b. Studies excluded from the meta-analysis (n = 22)

Ref.	Study	Country	n	Reason for exclusion
1.	Lee *et al* 2024	South Korea	104	Outcome includes near-pCR (pTRG 3–4), not pure ypT0N0
2.	Defeudis *et al* 2022	Italy	28	Outcome includes near-pCR (TRG 1–2 responders)
3.	Zhou *et al* 2023	China	155	Outcome = good response (TRG 0–1) includes near-pCR
4.	Kaval *et al* 2024	Turkey	28	Outcome = CR includes cCR; pure pCR not reported separately
5.	Crimi *et al* 2024	Italy	30	Very small pCR events (n = 4 in external validation)
6.	Yi *et al* 2019	China	41	No external validation; ROI on single 2D slice only
7.	Pang *et al* 2021	China	34	Low sensitivity (50%) on external validation; small pCR events
8.	Wang *et al* 2022	USA	37	CT modality only; small external validation
9.	Huang *et al* 2023	China	110	Mixed neoadjuvant regimens (some chemotherapy alone without radiotherapy)

10.	Bulens *et al* 2020	Belgium/ Netherlands	55	Outcome = ypT0-1N0 (includes near-pCR, not pure pCR)
11.	Antunes *et al* 2020	USA	44	Single-slice ROI
12.	Jayaprakasam *et al* 2022	USA	236	No external validation (cross-validation only)
13.	Miranda *et al* 2023	Brazil	55	No external validation; train/test split only
14.	Pang *et al* 2025	China	207	Prospective validation but moderate AUC (0.795); not pooled
15.	Wen *et al* 2023	China	42	No external validation; small validation set (n = 42, 9 pCR)
16.	Horvat *et al* 2018	USA	114	No external validation (CV only); perfect sensitivity suggests overfitting
17.	Losurdo *et al* 2021	Italy	40	Multimodal (miRNA + imaging); outcome includes near-pCR
18.	Cusumano *et al* 2021	Italy/Netherlands	195	No external validation; modest AUC (0.72)
19.	Lu *et al* 2024	China	52	Very low specificity (30%) on external

				validation
20.	Shin *et al* 2022	South Korea	306	Temporal validation only; no external validation
21.	Bellini *et al* 2022	Italy	40	No validation; severe overfitting (100% accuracy on derivation set)
22.	Chee *et al* 2017	South Korea	95	No classification model; only statistical group comparisons

pCR: Pathological complete response; cCR: Clinical complete response; pTRG: Pathological tumour regression grade; TRG: Tumour regression grade; ROI: Region of interest; CV: Cross-validation; miRNA: MicroRNA.

Supplementary Table S3. Detailed assessment of potential patient overlap across the 13 included studies

Study	Institution(s)	Recruitment period	Sample size	pCR rate	Overlap risk
Bibault *et al* 2018	France	2008–2015	95	23.20%	None identified
Ju *et al* 2023	South Korea	2015–2019	156	17.30%	None identified
Wu *et al* 2025	China	2018–2023	768	30.10%	None identified
Liu *et al* 2017	China	2010–2015	222	17.10%	None identified

Li *et al* 2021	China	2015–2019	80	18.80%	None identified
Peng *et al* 2023	China	2017–2022	83	26.50%	None identified
Boldrini *et al* 2022	Italy/China	2012–2020	59	16.90%	None identified
Jin *et al* 2021	China	2016–2019	141	30.50%	None identified
Cao *et al* 2026	China	2020–2024	181	17.10%	None identified
Shi *et al* 2023	China	2018–2022	77	22.10%	None identified
Nie *et al* 2016	China	2012–2015	48	22.90%	None identified
Horvat *et al* 2022	USA/Brazil	2015–2020	50	16.00%	None identified
Petkovska *et al* 2020	USA	2013–2018	102	18.60%	None identified

No definitive evidence of patient overlap was identified for any pair of studies; however, in the absence of patient-level identifiers, residual overlap cannot be fully excluded. pCR: Pathological complete response.

Supplementary Table S4. Radiomics quality score (RQS) criteria

Criteria	Description	Points
1. Image protocol	Well-documented image protocols (e.g.,	=+1 (if protocols are

quality	contrast, slice thickness, energy, etc.) and/or usage of public image protocols allowing reproducibility/replicability	well-documented) +1 (if public protocol is used)
2. Multiple segmentations	Possible actions: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities	1
3. Phantom study on all scanners	Detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability	1
4. Imaging at multiple time points	Collect images of individuals at additional time points. Analyse feature robustness to temporal variabilities (e.g., organ movement, organ expansion/shrinkage)	1
5. Feature reduction or adjustment for	Decreases the risk of overfitting. Overfitting is inevitable if the number of	−3 (if neither measure is implemented)

multiple testing	features exceeds the number of samples. Consider feature robustness when selecting features	+3 (if either measure is implemented)
6. Multivariable analysis with non-radiomics features	Inclusion of non-radiomics features (e.g., EGFR mutation) is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non-radiomics features	1
7. Detect and discuss biological correlates	Demonstration of phenotypic differences (possibly associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology	1
8. Cut-off analyses	Determine risk groups by either the median, a previously published cut-off, or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	1
9. Discrimination statistics	Report discrimination statistics (e.g., C-statistic, ROC curve, AUC) and their statistical significance (e.g., p-values,	=+1 (if a discrimination statistic and its statistical significance are reported)

	confidence intervals). One can also apply resampling methods (e.g., bootstrapping, cross-validation)	+1 (if a resampling method technique is also applied)
10. Calibration statistics	Report calibration statistics (e.g., Calibration-in-the-large/slope, calibration plots) and their statistical significance (e.g., p-values, confidence intervals). One can also apply resampling methods (e.g., bootstrapping, cross-validation)	=+1 (if a calibration statistic and its statistical significance are reported) +1 (if a resampling method technique is also applied)
11. Prospective study registered in a trial database	Provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	=+7 (for prospective validation of a radiomics signature in an appropriate trial) −5 (if validation is missing)
12. Validation	The validation is performed without retraining and without adaptation of the cut-off value, providing crucial information with regard to credible clinical performance	+2 (if validation is based on a dataset from the same institute) +3 (if validation is based on a dataset from another institute)

		+4 (if validation is based on two datasets from two distinct institutes) +4 (if the study validates a previously published signature) +5 (if validation is based on three or more datasets from distinct institutes)
13. Comparison to 'gold standard'	Assess the extent to which the model agrees with/is superior to the current 'gold standard' method (e.g., TNM-staging for survival prediction). This comparison shows the added value of radiomics	2
14. Potential clinical utility	Report on the current and potential application of the model in a clinical setting (e.g., decision curve analysis)	2
15. Cost-effectiveness analysis	Report on the cost-effectiveness of the clinical application (e.g., QALYs generated)	1

		=+1 (if scans are open source) +1 (if region of interest segmentations are open source) +1 (if code is open source) +1 (if radiomics features are calculated on a set of representative ROIs and the calculated features and representative ROIs are open source)
16. Open science and data	Make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	
	Total points (36=100%)	

Supplementary Table S5. Radiomics quality score (RQS) assessment of the 13 included studies

Ref.	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D11**	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	
Bibault *et*	1	0	0	0	3	1	0	0	2	0	0	2	0	0	0	0	9 (25%)	

al 2018																	
Ju *et al*																	
2023	1	0	0	0	3	0	0	0	2	0	0	3	0	0	0	0	9 (25%)
Wu *et al*																	11
2025	1	0	0	0	3	1	0	0	2	0	0	4	0	0	0	0	(31%)
Liu *et al*																	11
2017	1	0	0	1	3	1	0	0	2	1	0	2	0	0	0	0	(31%)
Li *et al*																	
2021	1	0	0	1	3	0	0	0	2	0	0	2	0	0	0	0	9 (25%)
Peng *et al*																	10
2023	1	0	0	1	3	1	0	0	2	0	0	2	0	0	0	0	(28%)
Boldrini *et*																	10
al 2022	1	0	0	0	3	1	0	0	2	0	0	3	0	0	0	0	(28%)
Jin *et al*																	13
2021	1	0	0	1	3	1	0	0	2	1	0	3	0	0	0	1	(36%)
Cao *et al*																	15
2026	1	0	0	1	3	1	0	0	2	1	0	5	0	0	0	1	(42%)
Shi *et al*																	
2023	1	0	0	0	3	0	0	0	2	0	0	3	0	0	0	0	9 (25%)
Nie *et al*	1	0	0	0	3	0	0	0	2	0	0	2	0	0	0	0	8 (22%)

2016																	
Horvat *et al* 2022	1	0	0	0	3	0	0	0	2	0	0	3	0	0	0	0	9 (25%)
Petkovska *et al* 2020	1	0	0	0	3	1	0	0	2	0	0	2	0	0	0	0	9 (25%)

D1: Image protocol quality; D2: Multiple segmentations; D3: Phantom study; D4: Imaging at multiple time points; D5: Feature reduction; D6: Multivariable analysis with non-radiomics features; D7: Biological correlates; D8: Cut-off analyses; D9: Discrimination statistics; D10: Calibration statistics; D11: Prospective study; D12: Validation; D13: Comparison to gold standard; D14: Potential clinical utility; D15: Cost-effectiveness analysis; D16: Open science and data; D17: Total points.

Supplementary Table S6: GRADE-DTA certainty assessment

Domain	Assessment	Rating	Rationale
Risk of Bias	Serious	⬇ -1	8/13 studies (61.5%) rated as high risk of bias using PROBAST, primarily due to limitations in the Analysis domain (small sample size, lack of external validation, no calibration reporting). QUADAS-AI identified similar concerns in the Index Test domain.

Indirectness	Not serious	0	Population (LARC patients), index test (AI-based imaging models), and reference standard (histopathological ypT0N0) directly matched the review question. Predominance of MRI studies (84.6%) limits generalizability to CT/PET modalities, but not enough to warrant downgrading.
Inconsistency	Not serious	0	Between-study heterogeneity was low (I^2 = 11.3% by the Zhou and Dendukuri approach). The bivariate model explained most of the between-study variance. No downgrading for inconsistency.
Imprecision	Not serious	0	Confidence intervals around pooled estimates were relatively narrow (sensitivity 70.6–86.4%; specificity 78.6–88.5%; HSROC AUC 0.889).
Publication bias	Not serious	0	Egger-type regression test was non-significant (P = 0.65), though with limited power (n = 13); publication bias cannot be definitively ruled out
Overall certainty	MODERATE ⬇️		Downgraded by one level (risk of bias only); the current estimate is likely to be close to the true effect.

GRADE: Grading of Recommendations Assessment, Development and Evaluation; DTA: Diagnostic test accuracy; LARC: Locally advanced rectal cancer; I^2: I-squared heterogeneity statistic.

Supplementary Table S7: Combined QUADAS-AI assessment

Study	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	D10	**D11**
Bibault *et al* 2018	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Ju *et al* 2023	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Wu *et al* 2025	Low	Low	Low	Low	Yes	Yes	Yes	Yes	Yes	Low	Low
Liu *et al* 2017	High	High	Low	Low	No	Uncle ar	Unclear	Unclear	No	High	High
Li *et al* 2021	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Peng *et al* 2023	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Boldrini *et al* 2022	Low	Low	Low	Low	Yes	Yes	Yes	Yes	No	Low	Low
Jin *et al* 2021	Low	Low	Low	Low	Yes	Yes	Yes	Yes	Yes	Low	Low
Cao *et al* 2026	Low	Low	Low	Low	Yes	Yes	Yes	Yes	Yes	Low	Low

Shi *et al* 2023	Low	Low	Low	Low	Yes	Yes	Yes	Unclear	No	Low	Low
Nie *et al* 2016	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Horvat *et al* 2022	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High
Petkovska *et al* 2020	High	High	Low	Low	No	Yes	Unclear	Unclear	No	High	High

D1: Patient selection; D2: Index test (AI model); D3: Reference standard (histopathological ypT0N0); D4: Flow and timing; D5: Algorithm description; D6: Training/validation split; D7: Data leakage avoidance; D8: Missing data handling; D9: Calibration reporting; D10: Overall risk of bias; D11: Applicability concern.

Supplementary Table S8. Study-level 2×2 data used in the primary meta-analysis

Study	**Validation cohort used**	**n**	**TP**	**FP**	**FN**	**TN**	**Threshold**	**AUC source**	**Rationale for cohort/model selection**
Bibault *et al* 2018	Internal (5-fold	95	15	12	7	61	Youden	Reported	Single model; CV performance used

	CV)								
Ju *et al* 2023	External	40	5	4	1	30	Youden	Reported	RF external selected over CNN
Wu *et al* 2025	External (Test 1 + Test 2)	367	91	47	13	216	Youden	Reported (Table 2 of source)	Combined external cohorts; training excluded
Liu *et al* 2017	Temporal	70	11	3	1	55	Youden	Reported	Validation cohort used
Li *et al* 2021	Internal (LOO-CV)	80	12	1	2	64	Youden	Reported	Best of four classifiers
Peng *et al* 2023	Internal	83	19	7	3	54	Youden	Reported	Space-time model selected

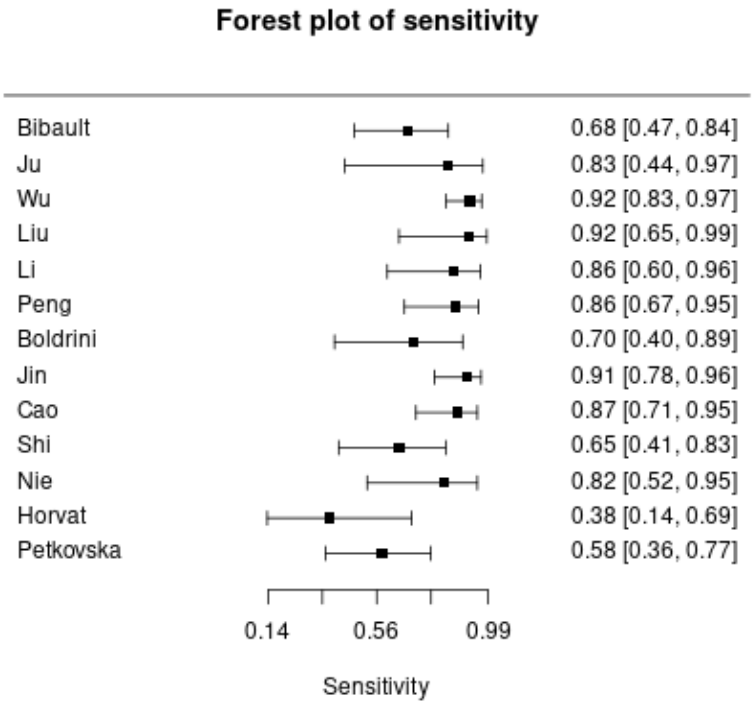
Boldrin i *et al* 2022	External	59	7	18	3	31	Published	Reported	External validation of Dinapoli model
Jin *et al* 2021	External	141	39	8	4	90	Youden	Reported	External cohort used
Cao *et al* 2026	External	181	27	19	4	131	Youden	Reported	External cohort used
Shi *et al* 2023	External	77	11	18	6	42	Youden	Reported	Prospective external validation
Nie *et al* 2016	Internal	48	9	8	2	29	Youden	Reported	Internal validation
Horvat *et al* 2022	External	50	3	5	5	37	Youden	Reported	External validation
Petko	Intern	102	11	19	8	64	Youden	Reported	Combined radiomics +

vska *et al* 2020	al	MRI model

[1]AUC: Area under the curve; TP: True positive; FP: False positive; FN: False negative; TN: True negative; RF: Random forest; CNN: Convolutional neural network; LOO-CV: Leave-one-out cross-validation

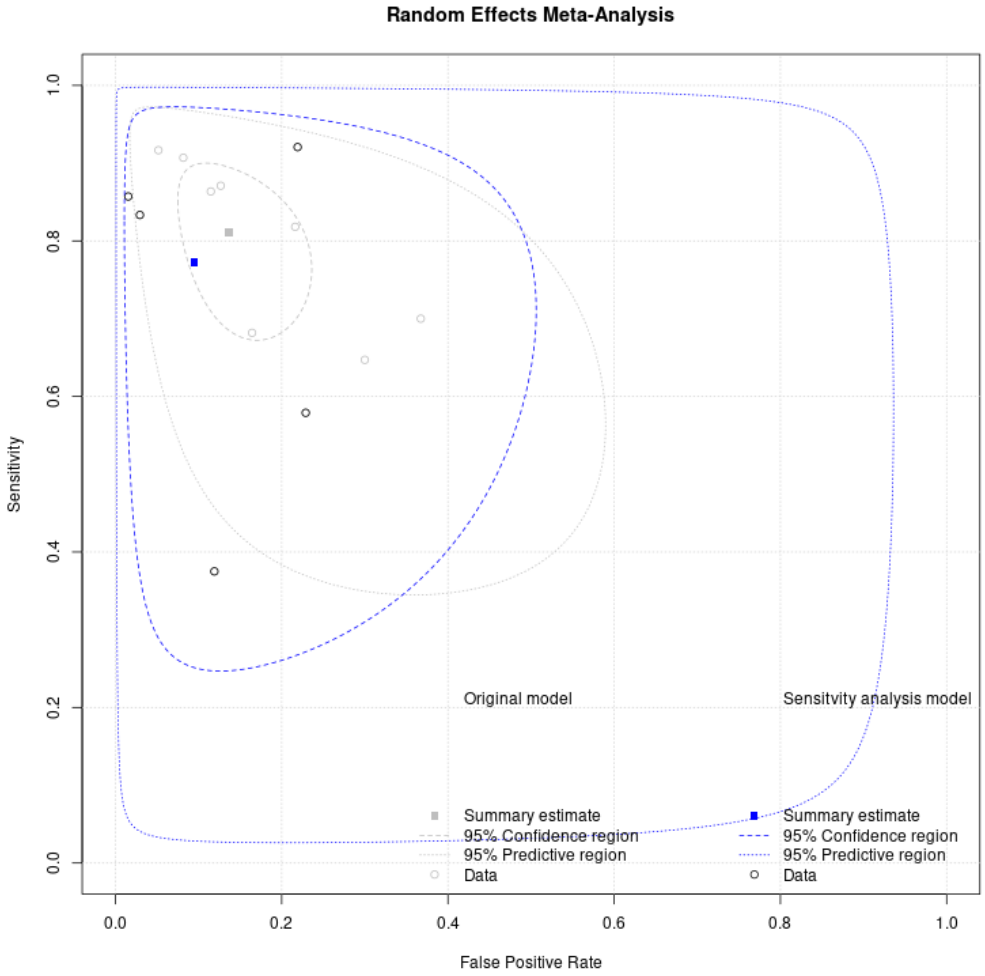
[1] The Ju et al 2023 2×2 counts were corrected to match the sensitivity (83.3%) and specificity (88.2%) reported in the source publication. The Wu et al 2025 2×2 counts were derived from the reported sensitivity (0.921, external test 1; 0.805, external test 2) and specificity (0.781, external test 1; 0.872, external test 2), combined across both external cohorts (n = 367), corresponding to a combined sensitivity of 87.5% and specificity of 82.1%. All other 2×2 values were taken from the primary publications or their supplementary materials.

Supplementary Figure 1: Forest plot with study-specific AUCs



Forest plot displaying individual study sensitivity values with 95% confidence intervals, alongside the pooled sensitivity. Studies are ordered by year of publication.

Supplementary Figure 2: Sensitivity analysis



Sensitivity analysis showing the HSROC curve for the primary model (solid line) and the sensitivity-analysis model restricted to low risk-of-bias studies (dashed line). The two curves overlap substantially, indicating that the primary estimate was not materially inflated by high-risk studies.