

RESEARCH ARTICLE

Open Access



# Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis

Sergei Bedrikovetski<sup>1,2\*</sup> , Nagendra N. Dudi-Venkata<sup>1,2</sup>, Hidde M. Kroon<sup>1,2</sup>, Warren Seow<sup>1</sup>, Ryash Vather<sup>2</sup>, Gustavo Carneiro<sup>3</sup>, James W. Moore<sup>1,2</sup> and Tarik Sammour<sup>1,2</sup>

## Abstract

**Background:** Artificial intelligence (AI) is increasingly being used in medical imaging analysis. We aimed to evaluate the diagnostic accuracy of AI models used for detection of lymph node metastasis on pre-operative staging imaging for colorectal cancer.

**Methods:** A systematic review was conducted according to PRISMA guidelines using a literature search of PubMed (MEDLINE), EMBASE, IEEE Xplore and the Cochrane Library for studies published from January 2010 to October 2020. Studies reporting on the accuracy of radiomics models and/or deep learning for the detection of lymph node metastasis in colorectal cancer by CT/MRI were included. Conference abstracts and studies reporting accuracy of image segmentation rather than nodal classification were excluded. The quality of the studies was assessed using a modified questionnaire of the QUADAS-2 criteria. Characteristics and diagnostic measures from each study were extracted. Pooling of area under the receiver operating characteristic curve (AUROC) was calculated in a meta-analysis.

**Results:** Seventeen eligible studies were identified for inclusion in the systematic review, of which 12 used radiomics models and five used deep learning models. High risk of bias was found in two studies and there was significant heterogeneity among radiomics papers (73.0%). In rectal cancer, there was a per-patient AUROC of 0.808 (0.739–0.876) and 0.917 (0.882–0.952) for radiomics and deep learning models, respectively. Both models performed better than the radiologists who had an AUROC of 0.688 (0.603 to 0.772). Similarly in colorectal cancer, radiomics models with a per-patient AUROC of 0.727 (0.633–0.821) outperformed the radiologist who had an AUROC of 0.676 (0.627–0.725).

**Conclusion:** AI models have the potential to predict lymph node metastasis more accurately in rectal and colorectal cancer, however, radiomics studies are heterogeneous and deep learning studies are scarce.

**Trial registration:** PROSPERO CRD42020218004.

**Keywords:** Colorectal cancer, Artificial intelligence, Radiomics, Deep learning, Machine learning, meta-analysis, Lymph node metastasis

\* Correspondence: [sergei.bedrikovetski@gmail.com](mailto:sergei.bedrikovetski@gmail.com)

<sup>1</sup>Discipline of Surgery, Faculty of Health and Medical Sciences, School of Medicine, University of Adelaide, Adelaide, South Australia, Australia

<sup>2</sup>Department of Surgery, Colorectal Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Colorectal cancer (CRC) is the second most common malignancy and the third leading cause of cancer-related mortality in the world, accounting for 862,000 deaths annually [1]. CRC nodal metastases play a pivotal role in disease-free survival and in determining appropriate adjuvant and neoadjuvant treatment [2]. As a result of the application of preoperative staging MRI in patients with rectal cancer, neoadjuvant chemoradiation has become the standard of care in locally advanced tumours, resulting in improved local control and resectability. Owing to the lower accuracy of lymph node staging in colon cancer at diagnosis, neoadjuvant treatment is not as commonly recommended [3, 4]. However, this may change following the results of the recent Fluoropyrimidine, Oxaliplatin and Targeted Receptor Pre-Operative Therapy (FOXTROT) trial showing the safety and efficacy of neoadjuvant chemotherapy in patients with locally advanced colon cancer [5]. Therefore, improved accuracy in clinical nodal staging at diagnosis may become critical in surgical planning and targeting effective neoadjuvant treatment for these patients [6, 7].

Clinical staging of CRC is typically performed by radiologists assessing contrast enhanced computer tomography (CT) images in patients with colorectal cancer, and in addition, magnetic resonance imaging (MRI) in patients with rectal cancer. The staging accuracy of CT and MRI is affected by multiple factors, such as equipment performance, standardised imaging protocols, the reporting radiologist's experience, and patient-specific factors. Overall, published series have reported a 70% accuracy of diagnosing lymph node metastasis on CT, and 69% on MRI using standard criteria [8, 9].

Current staging paradigms with its limited diagnostic and staging accuracy may be able to overcome by using Artificial intelligence (AI) models. AI-enabled radiomics involves the extraction of a large number of investigator defined features from medical images using advanced computational algorithms [10]. While radiomics models have been used to predict lymph node metastasis in CRC with partial success, previous studies by Ding et al. and Wang et al. demonstrate that deep learning algorithms have the potential to identify more subtle patterns that may elude conventional radiological and statistical methods [11–13]. Deep learning is a technique that involves the use of convolutional neural networks to self-educate an algorithm based on useful representations of images, thus bypassing the step of extracting manually designed features [14]. In recent years, radiomics nomograms and deep learning models have started to make a meaningful contribution to radiological diagnoses [15].

The aim of this systematic review and meta-analysis is to evaluate the accuracy of AI models in diagnosing

lymph node metastasis on CT and/or MRI in colorectal cancer patients.

## Methods

### Search strategy

This systematic review and meta-analysis was performed according to the recommendations of the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines and was registered with the International Prospective Register of Systematic Reviews with an analysis plan prior to conducting the research. A systematic search of the Cochrane Library, PubMed (MEDLINE), EMBASE and IEEE Xplore databases was performed for studies published between January 1st 2010 and October 1st 2020. The following search terms were used: artificial intelligence, deep learning, convolutional neural network, machine learning, automatic detection, radiomics, radiomic, CT/MRI, lymph node, lymph node metastasis, colon, rectal, colorectal (Additional file 1: Table S1). Reference lists of articles retrieved were also searched manually to identify additional eligible studies.

### Study selection

Articles were included if they met the following criteria: (1) included patients with histopathological diagnosis of CRC; (2) developed or used a radiomics or deep learning algorithm to assess CT or MRI pre-operative lymph node metastasis detection and (3) published in English language. Exclusion criteria were (1) case reports, review articles, editorials, letters, comments, and conference abstracts; (2) studies focusing on segmentation or feature extraction methods only and (3) animal studies. After removing duplicates, titles and abstracts were reviewed for eligibility by two independent reviewers (SB and NNDV) using Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia, available at [www.covidence.org](http://www.covidence.org)). Any disagreements were resolved by consensus arbitrated by a third author (TS).

### Data extraction

Data from selected full-text articles were reviewed for reporting on the type of radiomics or deep learning model, study characteristics and outcome measures. The extracted data included the first author, year of publication, country, study type, number of patients, sample size for diagnostic accuracy, age, imaging modality, type of malignancy, AI model, and referenced standard. Data related to the accuracy of the radiologists' assessment derived from studies using clinical nodal staging or clinical nomograms solely based on N-staging was also collected. To obtain diagnostic accuracy data of AI models and radiologists' assessment, two-by-two contingency tables, sensitivity, specificity, accuracy, and area under the

receiver operating characteristic curve (AUROC) were extracted or reconstructed. The primary endpoint was AUROC, secondary endpoints included sensitivity, specificity, and accuracy.

#### Quality assessment and publication Bias

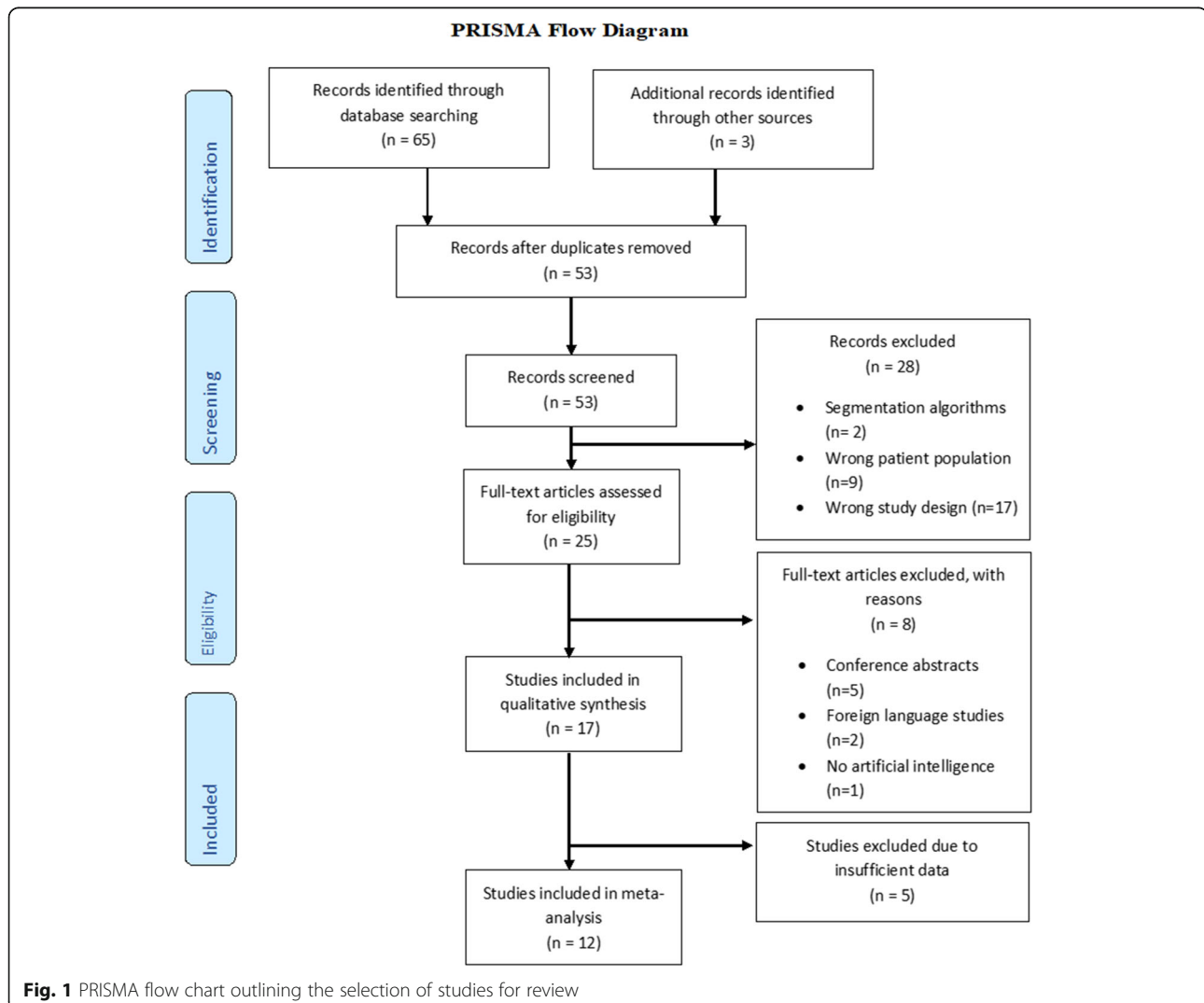
The modified version as proposed by Sollini et al. of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to assess the methodological quality of the included studies [16]. Minimum criteria for fulfilling each QUADAS-2 item were discussed by two reviewers (SB and NNDV) and disagreements were resolved by consensus. Publication bias was assessed using the Egger regression test and is presented as a funnel plot of diagnostic AUROC.

#### Statistical analysis

Meta-analysis was performed using testing set results of studies that presented absolute numbers for AUROC

and 95% confidence intervals, contingency tables or provided sufficient information to derive the numbers manually. If results were not reported in an independent test set, cross validation or full test sample results are presented in this review. When results of different AI algorithms were reported in one article, the proposed algorithm with the highest diagnostic performance was analysed.

Three software packages, MedCalc for Windows, version 16.4.3 (MedCalc Software, Ostend, Belgium), RevMan, version 5.3.21 and Meta-DiSc version 1.4, were utilised for statistical analysis. Missing data were computed using confusion matrix calculator or manually derived using formulas in Additional file 1: Table S2. Pooling sensitivity, specificity and AUROC data was conducted using the Mantel-Haenszel method (fixed-effects model) and the DerSimonian Laird method (random-effects model) [17, 18]. To assess heterogeneity between studies, the inconsistency index ( $I^2$ ) was used [19].



Heterogeneity was quantified as low, moderate, and high, with upper limits of 25, 50 and 75% for  $I^2$ , respectively. Forrest plots were drawn to show AUROC estimates in each study in relation to the summary pooled estimate. A funnel plot was constructed to visually assess publication bias.

## Results

### Study selection

A total of 68 studies were identified and 53 remained after removing duplicates. Review of titles and abstracts left 25 studies for full-text review. Finally, 17 studies were included in the systematic review, 12 of which could be used in the meta-analysis and five studies were excluded due to insufficient information (Fig. 1) [11, 12, 20–34].

### Study characteristics

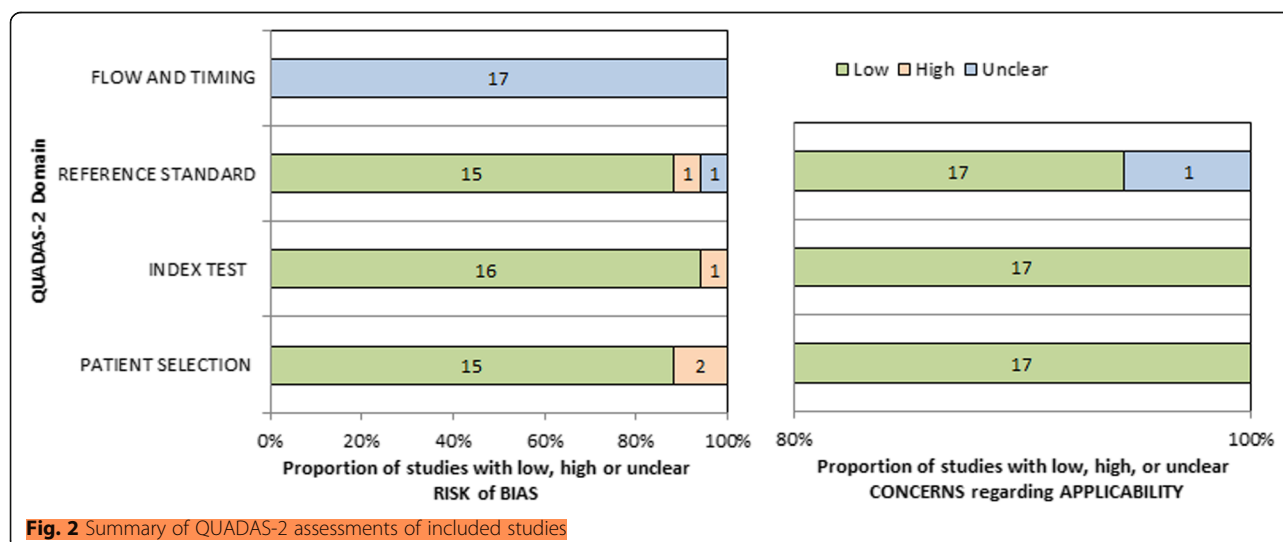
Twelve studies used radiomics models and five used deep learning models (Additional file 1: Table S3). All included studies were published between 2011 and 2020. Study design was retrospective in 11 and prospective in six studies. Fourteen studies were single-center and three were multi-center. Patients were predominantly male with a median age of 60 years (54–64). Eight studies used MRI and nine used CT to train their algorithm. The type of malignancy was colorectal in three studies, colon only in two, and rectal only in 12. Eleven studies used per-patient diagnostic output (the patient is node positive or negative) and 6 studies used per-nodal diagnostic output of lymph node metastasis (each individual node analysed separately). Fifteen studies used the post-operative pathology report as reference standard, and one study used a radiology report as the reference standard. The reference standard for the one remaining study was not reported.

### Quality assessment and publication Bias

The methodologic quality of included studies is summarized in Fig. 2. As per the QUADAS-2 tool, risk of bias in patient selection was low in 15 (88%) studies and high in two (12%) studies. Risk of bias in the index test was high in one study (6%) and low in 16 (94%). Risk of bias in the reference standard test was low in 15 (88%), high in one study (6%) and unclear in one study (6%). Flow and timing had all 17 studies with unclear risk of bias. Overall applicability concerns were low (Additional file 1: Table S4). Funnel plot assessment (Additional file 1: Figure S1) showed no significant publication bias (Egger's intercept 1.11, 95%CI – 1.22 to 3.42,  $p = 0.313$ ).

### Diagnostic accuracy

For the 12 studies that could be included in the quantitative analysis, 10 used radiomics and two used deep learning. For each outcome, summary estimates of sensitivity, specificity and AUROC were produced with 95% confidence intervals on a per-patient and per-nodal basis (Table ). Pooled colorectal and rectal, per-patient and per-nodal detailed diagnostic measures reported by individual studies are shown in Table . The data for radiomics models in rectal cancer showed high heterogeneity with the exception of per-node AUROC and sensitivity. On a per-patient basis, radiomics in rectal cancer pooled AUROC was 0.808 (95%CI 0.739–0.876; Fig. ) and pooled sensitivity and specificity were 0.776 (95%CI 0.685–0.851) and 0.676 (95%CI 0.608–0.739), respectively. On a per-nodal basis radiomics in rectal cancer pooled AUROC was 0.846 (95%CI 0.803–0.890) and pooled sensitivity and specificity were 0.896 (95%CI 0.834–0.941) and 0.743 (95%CI 0.665–0.811), respectively. On a per-patient basis radiomics in CRC pooled AUROC was 0.727 (95%CI 0.633–0.821). The radiologist per-patient assessment in rectal cancer pooled AUROC



**Table 1** Results for deep learning radiomics models and radiologist in accuracy to detect lymph node metastasis

First author	TP	FP	TN	FN	PPV, %	NPV, %	Sensitivity, %	Specificity, %	Accuracy, %	AUROC	95%CI	Standard Error <sup>c</sup>
<b>Deep learning</b>												
<b>Per-patient</b>												
Ding [12]	–	–	–	–	–	–	–	–	–	0.920	0.876–0.964	0.0224
Wang [11]	40	4	58	5	90.9	92.1	88.9	93.5	91.6	0.912 <sup>c</sup>	0.842–0.958	0.0296
Glaser [25] <sup>a</sup>	–	–	–	–	–	–	–	–	–	0.860	–	–
<b>Per-node</b>												
Lu [28]	–	–	–	–	–	–	–	–	–	0.912	–	–
Li [29]	–	–	–	–	–	–	–	–	94.4	–	–	–
<b>Radiomics</b>												
<b>Per-patient</b>												
Eresen [20]	29 <sup>c</sup>	6 <sup>c</sup>	33 <sup>c</sup>	10 <sup>c</sup>	82.8 <sup>c</sup>	76.7 <sup>c</sup>	74.36	84.62	79.49	0.825	0.778–0.872	0.0240
Li [21]	69 <sup>c</sup>	44 <sup>c</sup>	128 <sup>c</sup>	67 <sup>c</sup>	61.06	65.64	50.74	74.42	63.96	0.650	0.583–0.713	0.0331
Yang [22]	13 <sup>c</sup>	5 <sup>c</sup>	21 <sup>c</sup>	2 <sup>c</sup>	73.2 <sup>c</sup>	90.5 <sup>c</sup>	85.0	82.0	83.0	0.780	0.630–0.920	0.0740
Nakanishi [23]	–	–	–	–	–	–	–	–	–	0.900	0.800–0.990	0.0485
Zhou [24]	24 <sup>c</sup>	27 <sup>c</sup>	74 <sup>c</sup>	5 <sup>c</sup>	47.1	93.7	82.8	73.3	75.4	0.818	0.731–0.905	0.0444
Meng [26]	46 <sup>c</sup>	36 <sup>c</sup>	47 <sup>c</sup>	17 <sup>c</sup>	56.1 <sup>c</sup>	73.4 <sup>c</sup>	73.0	56.6	63.7	0.697	0.612–0.781	0.0431
Chen [30]	–	–	–	–	–	–	–	–	–	0.857	0.726–0.989	0.0671
Huang [31]	–	–	–	–	–	–	–	–	–	0.788	0.779–0.797	0.0046
<b>Per-node</b>												
Zhu [27]	18 <sup>c</sup>	21 <sup>c</sup>	32 <sup>c</sup>	1 <sup>c</sup>	46.2	97.0	94.7	60.4	69.4 <sup>c</sup>	0.812	0.703–0.895	0.0490
Cai [32]	–	–	–	–	–	–	89	82	88	–	–	–
Tse [34]	–	–	–	–	–	–	–	–	91.0	–	–	–
Cui [33]	111 <sup>c</sup>	17 <sup>c</sup>	78 <sup>c</sup>	14 <sup>c</sup>	86.7 <sup>c</sup>	85.0 <sup>c</sup>	89	82	88	0.855 <sup>c</sup>	0.801–0.898 <sup>c</sup>	0.0247
<b>Radiologist</b>												
<b>Per-patient</b>												
Li [21] <sup>d</sup>	94 <sup>c</sup>	63 <sup>c</sup>	109 <sup>c</sup>	42 <sup>c</sup>	59.9	72.2	69.1	63.4	65.9	0.708	0.645–0.765	0.0306
Eresen [20]	33 <sup>c</sup>	23 <sup>c</sup>	16 <sup>c</sup>	6 <sup>c</sup>	58.9 <sup>c</sup>	72.7 <sup>c</sup>	84.6	41.0	62.8	0.772	0.718–0.825	0.0273
Yang [22]	41	41	43	14	50.0	75.4	74.6	51.2	60.4	0.629 <sup>c</sup>	0.543–0.709 <sup>c</sup>	0.0423
Nakanishi [23] <sup>b</sup>	71	0	147	29	100.0	83.5	71.0	100.0	88.3 <sup>c</sup>	0.855 <sup>c</sup>	0.805–0.896 <sup>c</sup>	0.0232
Zhou [24] <sup>b</sup>	49	89	215	38	35.5	85.0	56.3	70.7	67.5	0.635	0.585–0.683	0.0250
Meng [26] <sup>b</sup>	88 <sup>c</sup>	96 <sup>c</sup>	124 <sup>c</sup>	37 <sup>c</sup>	47.8 <sup>c</sup>	77.0 <sup>c</sup>	55.9	70.4	61.3	0.632 <sup>c</sup>	0.578–0.683 <sup>c</sup>	0.0268
Chen [30]	–	–	–	–	–	–	–	–	–	0.671	0.511–0.831	0.0816
Huang [31]	58 <sup>c</sup>	30 <sup>c</sup>	69 <sup>c</sup>	43 <sup>c</sup>	65.9 <sup>c</sup>	61.6 <sup>c</sup>	57.4 <sup>c</sup>	69.7 <sup>c</sup>	63.5 <sup>c</sup>	0.636 <sup>c</sup>	0.565–0.702 <sup>c</sup>	0.0349
<b>Per-node</b>												
Cui [33]	39 <sup>c</sup>	101 <sup>c</sup>	52 <sup>c</sup>	36 <sup>c</sup>	27.7 <sup>c</sup>	59.1 <sup>c</sup>	52 <sup>c</sup>	34 <sup>c</sup>	39.9 <sup>c</sup>	0.430 <sup>c</sup>	0.365–0.497 <sup>c</sup>	0.0337

<sup>a</sup> Values extracted from training set<sup>b</sup> Values extracted from total cohort<sup>c</sup> Manually derived/reconstructed values using formulas from Additional file 1: Table S2<sup>d</sup> Values extracted from clinical models

FN, false negative; FP, false positive; TN, true negative and TP, true positive; PPV, positive predictive value; NPV, negative predictive value; AUROC, area under the receiver operating characteristic; CI, confidence interval

was 0.688 (95%CI 0.603 to 0.772), sensitivity was 0.678 (95%CI 0.628–0.726) and specificity was 0.701 (95%CI 0.667–0.733). Further, the radiologists per-patient assessment in CRC pooled AUROC was 0.676 (95%CI 0.627–0.725), sensitivity was 0.641 (95%CI 0.577–0.702) and specificity was 0.657 (95%CI 0.597–0.713). The deep learning data demonstrated low heterogeneity ( $I^2 = 0.00\%$ ,  $p = 0.829$ ), and on a per-patient basis, deep learning models outperformed radiomics and radiologist

assessment in rectal cancer with an AUROC of 0.917 (95%CI 0.882–0.952). Deep learning sensitivity and specificity were reported in a single study as 0.889 and 0.935, respectively (Table 1).

## Discussion

To our knowledge, this is the first systematic review and meta-analysis of deep learning and radiomics performance in the assessment of lymph node metastasis in



**Table 2** Pooled results of per-patient and per-node diagnosis from deep learning, radiomics and radiologists

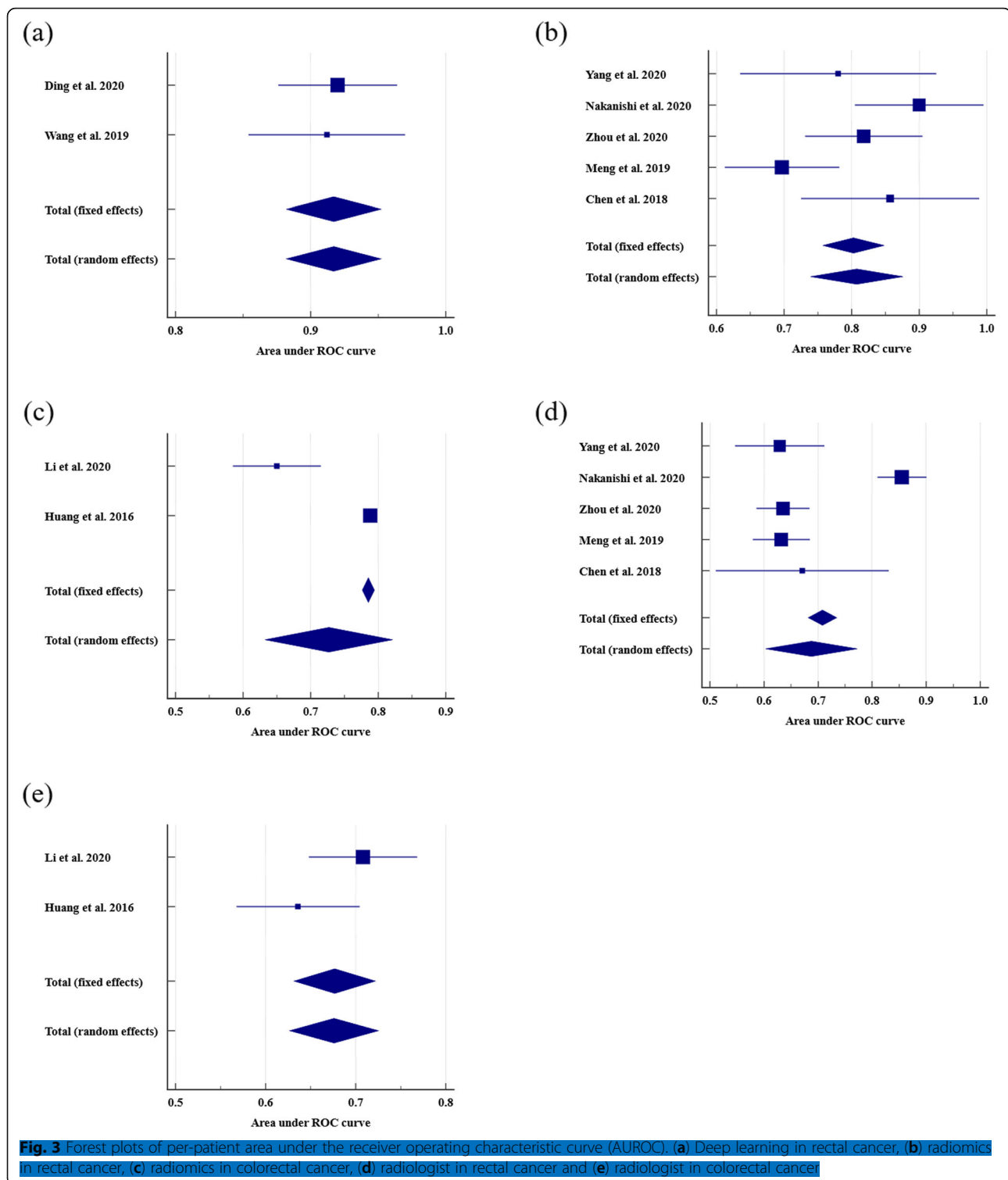
Variable	Studies analysed	Type of malignancy	No. of studies	Pooled results (95% CI)	Heterogeneity ( $I^2$ , %)	Heterogeneity P Value
<b>Deep learning</b>						
AUROC per-patient	[11, 12]	Rectal	2	0.917 (0.882–0.952)	0.00	0.829
<b>Radiomics</b>						
Sensitivity per-patient	[22, 24, 26]	Rectal	3	0.776 (0.685–0.851)	0.00	0.368
Sensitivity per-node	[27, 33]	Rectal	2	0.896 (0.834–0.941)	0.00	0.393
Specificity per-patient	[22, 24, 26]	Rectal	3	0.676 (0.608–0.739)	75.4	0.017
Specificity per-node	[27, 33]	Rectal	2	0.743 (0.665–0.811)	87.8	0.004
AUROC per patient	[21, 31]	Colorectal	2	0.727 (0.633–0.821)	94.1	< 0.0001
AUROC per patient	[22–24, 26, 30]	Rectal	5	0.808 (0.739–0.876)	63.3	0.028
AUROC per node	[27, 33]	Rectal	2	0.846 (0.803–0.890)	0.00	0.433
<b>Radiologist</b>						
Sensitivity per-patient	[21, 31]	Colorectal	2	0.641 (0.577–0.702)	70.9	0.064
Specificity per-patient	[21, 31]	Colorectal	2	0.657 (0.597–0.713)	11.1	0.289
Sensitivity per-patient	[22–24, 26]	Rectal	4	0.678 (0.628–0.726)	57.5	0.070
Specificity per-patient	[22–24, 26]	Rectal	4	0.701 (0.667–0.733)	97.8	< 0.0001
AUROC per-patient	[21, 31]	Colorectal	2	0.676 (0.627–0.725)	58.4	0.121
AUROC per-patient	[22–24, 26, 30]	Rectal	5	0.688 (0.603 to 0.772)	93.4	< 0.0001

AUROC, area under the receiver operating characteristic; CI, confidence interval

rectal and CRC patients. The results demonstrate a very high AUROC of 0.917 (95%CI, 0.882–0.952) when a deep learning model is used as a diagnostic tool compared with a radiomics model (AUROC 0.808, 95%CI 0.739–0.876). The diagnostic performance of both deep learning and radiomics models surpassed that of the radiologist assessment with an AUROC of 0.688 (95%CI, 0.603 to 0.772).

A number of research studies have already suggested AI has the potential to transform the healthcare sector particularly in areas where image recognition can be applied [35–37]. In terms of colorectal diseases, AI has been applied to colonic polyps, adenomas, colorectal cancer, ulcerative colitis and intestinal motility disorders [38–41]. Owing to the rapid development of AI technology, AI is bound to continually play an important role in the field of colorectal diagnosis and treatment [42]. Furthermore, the increase in computing power paired with the availability of large imaging databases offer the opportunity to develop more accurate AI algorithms.(10) At present, applications of deep learning to medical imaging are in vogue. However, deep learning models have several drawbacks, including variability in the images, large sample size, poor generalization and extensive computing resources. These models tend to rely on superficial data patterns and often fail when external factors such as different imaging acquisition parameters and types of scanners cause a distribution shift [43].

In this review, most studies used radiomics ( $n = 12$ ), rather than deep learning methodology ( $n = 5$ ) largely owing to deep learning technology being more recent, but also because it requires specific expertise. This limits the ability to draw definitive comparisons between the two AI models as one is somewhat over-represented in the data. Additionally, most studies were retrospective in design, making them prone to confounding and selection bias. Several studies focused on the technical aspects of the algorithm and did not address key limitations such as input variation, absence of clinical information (age, tumour site, patient history) and potential data overfitting often caused by noise in the data, overcomplicated models, and small sample sizes. Another issue, particularly common in deep learning studies, is the failure to report contingency tables or sufficient detail to enable reconstruction. We had to exclude five (29%) studies from the meta-analysis due to incomplete data. Most studies were conducted at a single-center and used internal verification or resampling methods (cross validation). Internal validation, however, tends to overestimate the AUROC due to the model's lack of generalizability, limiting the integration of AI models into the clinical setting [44]. Therefore, external validation prediction models using images from different hospitals are required to create reliable estimates on the level of performance at other sites [45]. The number of studies diagnosing lymph node metastasis on a per-



nodal basis in this meta-analysis is small. This is understandable, given that lymph node metastasis is staged on a per-patient basis in the clinical setting. Interestingly, five studies on rectal cancer extracted radiomics features

from CT despite MRI being the gold standard imaging modality for lymph node detection in clinical practice.

This meta-analysis has some limitations that merit consideration. Firstly, a relatively small number of deep

learning studies were available for inclusion. This, along with the heterogeneity seen in radiomics studies, means that the summary estimates of AUROCs have to be interpreted with caution. Secondly, because of incomplete reporting of results by several studies, estimates of diagnostic performance were calculated using limited data. Thirdly, given the majority of the included studies originate from China, there is a potential for geographical bias. Lastly, the wide range of scanner types, imaging protocols, and criteria for lymph node metastasis used may have affected accuracy of results. Results for radiomics and the radiologist assessment were highly heterogeneous, which may be attributed to the different imaging modalities and small sample sizes. In the future, diagnostic AI models will have to be rigorously evaluated on their clinical benefit in comparison to current standard of care, as not all are suitable for clinical practice. Therefore, studies comparing AI with the clinicians' performance are most valuable and are more likely to ensure safe and effective implementation of AI technology into daily practice [46, 47].

## Conclusion

AI models have the potential to predict lymph node metastasis more accurately on a per-patient basis in colorectal cancer than the radiologists' assessment, however, radiomics studies are heterogeneous and deep learning studies are scarce. With further development and refinement, AI models capable of accurately predicting nodal stage may represent a significant advance in pre-operative staging of colorectal cancer to better inform clinician and patient.

## Abbreviations

CRC: Colorectal cancer; MRI: Magnetic resonance imaging; CT: Computer tomography; AI: Artificial intelligence; QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies tool 2; AUROC: Area under the receiver operating characteristic curve

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-021-08773-w>.

**Additional file 1: Table S1.** Search Strategy. **Table S2.** Diagnostic accuracy measures. **Table S3.** Selected characteristics of included studies. **Table S3.** Quality assessment of studies included in systematic review, according to the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) Tool adapted with signalling questions by Sollini et al. **Figure S1.** Publication bias presentation using funnel plot of included studies.

## Acknowledgments

Presented to the Annual Scientific Congress of the Royal Australasian College of Surgeons, Melbourne, Australia, May 2021 [48].

## Authors' contributions

SB designed, developed, and refined the study protocol with contributions from NNDV, HMK, WS, RV, GC, JWM, and TS. SB and NNDV developed the search strategy and designed the literature search. SB and NNDV screened

titles and abstracts and undertook the data extraction. SB, NNDV, HMK, WS, RV, GC, JWM, and TS interpreted the data for the work; SB drafted the manuscript. All authors were involved in critically revising the draft. All authors approved the final version to be published.

## Funding

This work was supported by the Colorectal Surgical Society of Australia and New Zealand (CSSANZ) Foundation Grant and Australian Research Council Discovery Project Grant 180103232. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

All data generated or analysed during this study are included in this published article and supplementary material.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests

## Author details

<sup>1</sup>Discipline of Surgery, Faculty of Health and Medical Sciences, School of Medicine, University of Adelaide, Adelaide, South Australia, Australia.

<sup>2</sup>Department of Surgery, Colorectal Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia. <sup>3</sup>Australian Institute for Machine Learning, School of Computer Science, University of Adelaide, Adelaide, South Australia, Australia.

Received: 14 July 2021 Accepted: 8 September 2021

Published online: 26 September 2021

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. <https://doi.org/10.3322/caac.21492>.
- Baxter NN, Virnig DJ, Rothenberger DA, Morris AM, Jessurun J, Virnig BA. Lymph node evaluation in colorectal cancer patients: a population-based study. *J Natl Cancer Inst*. 2005;97(3):219–25. <https://doi.org/10.1093/jnci/dji020>.
- Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: Colon Cancer, version 2.2018. *J Natl Compr Cancer Netw*. 2018;16(4):359–69. <https://doi.org/10.6004/jnccn.2018.0021>.
- Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. Rectal Cancer, version 2.2018, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. 2018;16(7):874–901. <https://doi.org/10.6004/jnccn.2018.0061>.
- Seymour MT, Morton D, Investigators obotIFT: FOxTROT: an international randomised controlled trial in 1052 patients (pts) evaluating neoadjuvant chemotherapy (NAC) for colon cancer. 2019; 37(15\_suppl):3504–3504.
- Dighe S, Swift I, Brown G. CT staging of colon cancer. *Clin Radiol*. 2008; 63(12):1372–9. <https://doi.org/10.1016/j.crad.2008.04.021>.
- Sammour T, Malakorn S, Thampy R, Kaur H, Bednarski BK, Messick CA, et al. Selective central vascular ligation (D3 lymphadenectomy) in patients undergoing minimally invasive complete mesocolic excision for colon cancer: optimizing the risk-benefit equation. *Color Dis*. 2020;22(1):53–61. <https://doi.org/10.1111/codi.14794>.
- Iannicelli E, Di Renzo S, Ferri M, Pillozzi E, Di Girolamo M, Saporì A, et al. Accuracy of high-resolution MRI with lumen distention in rectal cancer staging and circumferential margin involvement prediction. *Korean J Radiol*. 2014;15(1):37–44. <https://doi.org/10.3348/kjr.2014.15.1.37>.
- Fernandez LM, Parlade AJ, Wasser EJ, Dasilva G, de Azevedo RU, Ortega CD, et al. How reliable is CT scan in staging right Colon Cancer? *Dis Colon Rectum*. 2019;62(8):960–4. <https://doi.org/10.1097/DCR.0000000000001387>.



10. Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol*. 2019;25(6): 485–95. <https://doi.org/10.5152/dir.2019.19321>.
11. Wang H, Wang H, Song L, Guo Q: Automatic Diagnosis of Rectal Cancer Based on CT Images by Deep Learning Method. In: *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI): 19–21 Oct. 2019 2019*; 2019: 1–5.
12. Ding L, Liu G, Zhang X, Liu S, Li S, Zhang Z, et al. A deep learning nomogram kit for predicting metastatic lymph nodes in rectal cancer. *Cancer Med*. 2020;9(23):8809–20. <https://doi.org/10.1002/cam4.3490>.
13. Bedrikovetski S, Dudi-Venkata NN, Maicas G, Kroon HM, Seow W, Carneiro G, et al. Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: a systematic review and meta-analysis. *Artif Intell Med*. 2021;113:102022. <https://doi.org/10.1016/j.artmed.2021.102022>.
14. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29(2):102–27. <https://doi.org/10.1016/j.zemedi.2018.11.002>.
15. Benjamins S, Dhunoo P, Mesko B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3(1):118. <https://doi.org/10.1038/s41746-020-00324-0>.
16. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging*. 2019;46(13):2656–72. <https://doi.org/10.1007/s00259-019-04372-x>.
17. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719–48.
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–60. <https://doi.org/10.1136/bmj.327.7414.557>.
20. Eresen A, Li Y, Yang J, Shanguan J, Velichko Y, Yaghami V, et al. Preoperative assessment of lymph node metastasis in Colon Cancer patients using machine learning: A pilot study. *Cancer Imaging*. 2020;20(1). <https://doi.org/10.1186/s40644-020-00308-z>.
21. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, et al. A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *J Transl Med*. 2020;18(1):46. <https://doi.org/10.1186/s12967-020-02215-0>.
22. Yang YS, Feng F, Qiu YJ, Zheng GH, Ge YQ, Wang YT. High-resolution MRI-based radiomics analysis to predict lymph node metastasis and tumor deposits respectively in rectal cancer. *Abdominal Rad*. 2020.
23. Nakanishi R, Akiyoshi T, Toda S, Murakami Y, Taguchi S, Oba K, et al. Radiomics approach outperforms diameter criteria for predicting pathological lateral lymph node metastasis after Neoadjuvant (chemo) radiotherapy in advanced low rectal Cancer. *Ann Surg Oncol*. 2020;27(11): 4273–83. <https://doi.org/10.1245/s10434-020-08974-w>.
24. Zhou X, Yi Y, Liu Z, Zhou Z, Lai B, Sun K, et al. Radiomics-based preoperative prediction of lymph node status following Neoadjuvant therapy in locally advanced rectal Cancer. *Front Oncol*. 2020;10. <https://doi.org/10.3389/fonc.2020.00604>.
25. Glaser S, Maicas G, Bedrikovetski S, Sammour T, Carneiro G: Semi-Supervised Multi-Domain Multi-Task Training for Metastatic Colon Lymph Node Diagnosis from Abdominal CT. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI): 3–7 April 2020 2020*; 2020: 1478–1481.
26. Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *Eur Radiol*. 2019;29(6):3200–9. <https://doi.org/10.1007/s00330-018-5763-x>.
27. Zhu H, Zhang X, Li X, Shi Y, Zhu H, Sun Y. Prediction of pathological nodal stage of locally advanced rectal cancer by collective features of multiple lymph nodes in magnetic resonance images before and after neoadjuvant chemoradiotherapy. *Chin J Cancer Res*. 2019;31(6):984–92. <https://doi.org/10.21147/j.issn.1000-9604.2019.06.14>.
28. Lu Y, Yu Q, Gao Y, Zhou Y, Liu G, Dong Q, et al. Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. *Cancer Res*. 2018;78(17):5135–43. <https://doi.org/10.1158/0008-5472.CAN-18-0494>.
29. Li J, Wang P, Li Y, Zhou Y, Liu X, Luan K: Transfer Learning of Pre- Trained Inception-V3 Model for Colorectal Cancer Lymph Node Metastasis Classification. In: *2018 IEEE International Conference on Mechatronics and Automation (ICMA): 5–8 Aug. 2018 2018*; 2018: 1650–1654.
30. Chen LD, Liang JY, Wu H, Wang Z, Li SR, Li W, et al. Multiparametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics. *Life Sci*. 2018;208:55–63. <https://doi.org/10.1016/j.lfs.2018.07.007>.
31. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34(18):2157–64. <https://doi.org/10.1200/JCO.2015.65.9128>.
32. Cai H, Cui C, Tian H, Zhang M, Li L. A novel approach to segment and classify regional lymph nodes on computed tomography images. *Comput Math Methods Med*. 2012;2012:1–9. <https://doi.org/10.1155/2012/145926>.
33. Cui C, Cai H, Liu L, Li L, Tian H, Li L. Quantitative analysis and prediction of regional lymph node status in rectal cancer based on computed tomography imaging. *Eur Radiol*. 2011;21(11):2318–25. <https://doi.org/10.1007/s00330-011-2182-7>.
34. Tse DM, Joshi N, Anderson EM, Brady M, Gleeson FV. A computer-aided algorithm to quantitatively predict lymph node status on MRI in rectal cancer. *Br J Radiol*. 2012;85(1017):1272–8. <https://doi.org/10.1259/bjr/13374146>.
35. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw Open*. 2019;2(10): e1913436. <https://doi.org/10.1001/jamanetworkopen.2019.13436>.
36. Rauschecker AM, Rudie JD, Xie L, Wang J, Duong MT, Botzakis EJ, et al. Artificial intelligence system approaching Neuroradiologist-level differential diagnosis accuracy at brain MRI. *Radiology*. 2020;295(3):626–37. <https://doi.org/10.1148/radiol.2020190283>.
37. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology*. 2020;296(2):E65–71. <https://doi.org/10.1148/radiol.2020200905>.
38. Luo Y, Zhang Y, Liu M, Lai Y, Liu P, Wang Z, et al. Artificial intelligence-assisted colonoscopy for detection of Colon polyps: a prospective, Randomized Cohort Study. *J Gastrointest Surg*. 2020;25(8):2011–8. <https://doi.org/10.1007/s11605-020-04802-4>.
39. Kudo SE, Ichimasa K, Villard B, Mori Y, Misawa M, Saito S, et al. Artificial intelligence system to determine risk of T1 colorectal Cancer metastasis to lymph node. *Gastroenterology*. 2021;160(4):1075–1084 e1072. <https://doi.org/10.1053/j.gastro.2020.09.027>.
40. Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: emerging technologies and future directions. *World J Gastroenterol*. 2021;27(17):1920–35. <https://doi.org/10.3748/wjg.v27.i17.1920>.
41. Segui S, Drozdal M, Pascual G, Radeva P, Malagelada C, Azpiroz F, et al. Generic feature learning for wireless capsule endoscopy analysis. *Comput Biol Med*. 2016;79:163–72. <https://doi.org/10.1016/j.combiomed.2016.10.011>.
42. Wang Y, He X, Nie H, Zhou J, Cao P, Ou C. Application of artificial intelligence to the diagnosis and therapy of colorectal cancer. *Am J Cancer Res*. 2020;10(11):3575–98.
43. Perone CS, Cohen-Adad J. Promises and limitations of deep learning for medical image segmentation. *J Med Artif Intel*. 2019;2. <https://doi.org/10.21037/jmai.2019.01.01>.
44. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol*. 2019;20(3):405–10. <https://doi.org/10.3348/kjr.2019.0025>.
45. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. <https://doi.org/10.1371/journal.pmed.1001381>.
46. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health*. 2019;1(6):e271–97. [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2).

47. Ding L, Liu GW, Zhao BC, Zhou YP, Li S, Zhang ZD, et al. Artificial intelligence system of faster region-based convolutional neural network surpassing senior radiologists in evaluation of metastatic lymph nodes of rectal cancer. *Chin Med J*. 2019;132(4):379–87. <https://doi.org/10.1097/CM9.000000000000095>.
48. Abstract Journal Colorectal Surgery. *ANZ J Surg* 2021, 91(S1):29–51.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

