Artificial Intelligence Assisted Capsule endoscopy for Detection Lesions of Crohn's disease: A Systematic Review and Meta-analysis

Yuling Yuling Bin, Rumei Rumei Peng, Rumei Rumei Peng, Yaqian Lee, Zhijie Lee, Yang Liu

Submitted to: Journal of Medical Internet Research
on: August 29, 2024

Table of Contents

Artificial Intelligence Assisted Capsule endoscopy for Detection Lesions of Crohn's disease: A Systematic Review and Meta-analysis

Yuling Yuling Bin[1]; Rumei Rumei Peng[2]; Rumei Rumei Peng[2]; Yaqian Lee[1]; Zhijie Lee[1]; Yang Liu[1]

[1]Intensive Care Medicine Department of Hengyang Central Hospital Hengyang CN
[2]Changsha Medical College Changsha CN

Corresponding Author:
Yang Liu
Intensive Care Medicine Department of Hengyang Central Hospital
No. 12 Yancheng Road, Hengyang City
Hengyang
CN

Abstract

Background: Crohn's disease (CD), a complex member of the inflammatory bowel disease spectrum, is characterized by the diversity and skipping distribution of intestinal mucosal lesions, significantly complicating its differential diagnosis with intestinal diseases such as ulcerative colitis and intestinal tuberculosis.

Objective: With the increasing application of artificial intelligence (AI) in the medical field, its utilization in clinical diagnosis has become more widespread. However, there is a lack of systematic evaluation regarding the specific efficacy of AI in identifying CD through capsule endoscopy.

Methods: This study conducted a comprehensive search of PubMed databases, Cochrane, EMBASE, and Web of Science up to May 21, 2024, to collect relevant literature. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to rigorously assess the quality of included studies, and detailed information on study characteristics and AI algorithms was extracted. A bivariate mixed-effects model was employed to synthesize and analyze the sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Additionally, meta-regression and subgroup analyses were conducted to delve into the potential sources of heterogeneity.

Results: Ultimately, eight studies encompassing 11 distinct AI models were included in this meta-analysis. The overall area under the curve (AUC) for AI in identifying CD through capsule endoscopy (CE) was 99% (95% CI, 100%-0.00), indicating high diagnostic accuracy. Specifically, the pooled sensitivity was 94% (95% CI, 93%-96%), specificity was 97% (95% CI, 95%-98%), positive likelihood ratio (PLR) was 32.7 (95% CI, 19.9-53.6), negative likelihood ratio (NLR) was 6% (95% CI, 4%-7%), and diagnostic odds ratio (DOR) reached 576 (95% CI, 295-1127). Meta-regression analysis further revealed that AI algorithm type, study population size, and study design might be key sources of heterogeneity.

Conclusions: This study demonstrates the significant potential of AI technology in assisting endoscopists in detecting and identifying CD patients through capsule endoscopy. However, given the limitations and heterogeneity of current research, more high-quality, large-sample studies are needed to comprehensively and thoroughly evaluate the practical application value of AI in CD diagnosis, thereby promoting its widespread adoption and optimization in clinical practice.

(JMIR Preprints 29/08/2024:65914)
DOI: https://doi.org/10.2196/preprints.65914

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Artificial Intelligence Assisted Capsule endoscopy for Detection Lesions of Crohn's disease: A Systematic Review and Meta-analysis

Yuling Bin[1], Rumei Peng[2], Yaqian Lee[1], Zhijie Lee[1], Yang Liu[1]*
*Correspondence: Yang Liu
type999@163.com
[1]Intensive Care Medicine Department of Hengyang Central Hospital
[2]Changsha Medical College

Abstract

Background and Objectives Crohn's disease (CD), a complex member of the inflammatory bowel disease spectrum, is characterized by the diversity and skipping distribution of intestinal mucosal lesions, significantly complicating its differential diagnosis with intestinal diseases such as ulcerative colitis and intestinal tuberculosis. With the increasing application of artificial intelligence (AI) in the medical field, its utilization in clinical diagnosis has become more widespread. However, there is a lack of systematic evaluation regarding the specific efficacy of AI in identifying CD through capsule endoscopy.

Methods This study conducted a comprehensive search of PubMed databases, Cochrane, EMBASE, and Web of Science up to May 21, 2024, to collect relevant literature. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to rigorously assess the quality of included studies, and detailed information on study characteristics and AI algorithms was extracted. A bivariate mixed-effects model was employed to synthesize and analyze the sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Additionally, meta-regression and subgroup analyses were conducted to delve into the potential sources of heterogeneity.

Results Ultimately, eight studies encompassing 11 distinct AI models were included in this meta-analysis. The overall area under the curve (AUC) for AI in identifying CD through capsule endoscopy (CE) was 99% (95% CI, 100%-0.00), indicating high diagnostic accuracy. Specifically, the pooled sensitivity was 94% (95% CI, 93%-96%), specificity was 97% (95% CI, 95%-98%), positive likelihood ratio (PLR) was 32.7 (95% CI, 19.9-53.6), negative likelihood ratio (NLR) was 6% (95% CI, 4%-7%), and diagnostic odds ratio (DOR) reached 576 (95% CI, 295-1127). Meta-regression analysis further revealed that AI algorithm type, study population size, and study design might be key sources of heterogeneity.

Conclusion This study demonstrates the significant potential of AI technology in assisting endoscopists in detecting and identifying CD patients through capsule endoscopy. However, given the limitations and heterogeneity of current research, more high-quality, large-sample studies are needed to comprehensively and thoroughly evaluate the practical application value of AI in CD diagnosis, thereby promoting its widespread adoption and optimization in clinical practice.

Keywords Artificial Intelligence, Capsule endoscopy, Crohn's disease

Introduction

CD is a chronic inflammatory condition that can affect any part of the gastrointestinal tract, from the mouth to the anus, with the terminal ileum and proximal colon being the most commonly involved regions[1]. It is characterized by a discontinuous, patchy distribution of inflammation. The diagnosis of CD relies on identifying specific findings through endoscopic examination and histological analysis of biopsy samples. Endoscopic features typically include a cobblestone appearance of the mucosa, aphthous ulcers, and skip lesions[2]. Even lesions such as ulcers, fistulas, strictures, and multiple comorbidities may arise during the occurrence and development of CD. These characteristic findings are crucial for distinguishing CD from other inflammatory bowel diseases and for guiding appropriate treatment strategies. So distinguishing CD from intestinal tuberculosis and ulcerative colitis can be challenging.

CE is a non-invasive diagnostic technique for gastrointestinal diseases, particularly effective in small bowel exploration. It has shown superior performance compared to conventional endoscopy in terms of CD lesion detection[3, 4]. Despite the significant benefits of CE, it faces inherent challenges. These include the ability to observe the entire digestive tract, which leads to a large workload and the inability to focus on a specific area for repeated observation, posing diagnostic difficulties. The large volume of non-targeted images, which can lead to missed diagnoses, even by experienced endoscopists[5]. This problem is even more pronounced for less experienced staff, who often achieve lower detection rates and have less situational awareness. Therefore, enhancing the diagnostic capabilities of less experienced endoscopists in interpreting CE images is highly desirable.

Deep learning, a subset of AI, is primarily based on deep artificial neural networks[6]. Convolutional neural network (CNNs), main deep learning algorithm for image analysis, have demonstrated remarkable performance across a wide range of image analysis tasks[7-9]. AI is already being utilized in clinical practice, it has also been successfully applied to gastrointestinal endoscopy images, such as enhancing the detection of polyps[10], tumors[11], intestinal tuberculosis, and inflammatory bowel disease[12] in endoscopy. Thereby improving disease detection rates. AI holds the potential to automatically detect various types of lesions, shorten the reading times of CE and reduce missed diagnoses. A significant barrier to the adoption of CE for comprehensive gastrointestinal examination can be removed. This meta-analysis aims to synthesize current research on the application of AI in capsule endoscopy for the detection of CD lesions. By integrating data from multiple studies, we evaluate the effectiveness, accuracy, and clinical impact of AI-assisted capsule endoscopy, providing insights into its potential as a diagnostic tool in gastroenterology.

Method

Protocol and Registration

This systematic review adheres to the rigorous PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for comprehensive reporting. Our protocol (CRD42024545296) was duly registered with PROSPERO in May 2024, ensuring transparency and reproducibility in our research methodology. All phases of the review-including title and abstract screening, full-text screening, data extraction, assessment of adherence to reporting guidelines, and evaluation of bias and applicability-were independently performed in duplicate by two reviewers. Any disagreements were resolved through discussion with a third independent reviewer.

Study Selection for Inclusion

Studies will be selected based on inclusion and exclusion criteria that focus on the utilization of AI for the diagnosis of mucosal lesions in CD via CE. Inclusion criteria will include peer-reviewed research articles and clinical trials that detail the development or application of AI algorithms specific to CE in CD patients. Exclusion criteria will eliminate studies not involving AI, those not using CE, and studies focusing on diseases other than CD.

The search will be conducted across several databases including PubMed, Cochrane, EMBASE, and Web of science using keywords such as "artificial intelligence", "capsule endoscopy", and "Crohn's disease" (**Supplementary Material**). Initial screening will involve reviewing titles and abstracts for relevance. Subsequently, full texts of selected articles will be examined to confirm eligibility. This process will be undertaken independently by two reviewers, with discrepancies resolved by consensus or by consulting a third expert reviewer.

Data Extraction

Data will be extracted using a standardized data collection form developed to capture specific details relevant to the review objectives. Key data points will include: study author(s), publication year, sample size, AI model description (e.g., type of algorithm, training data size), capsule endoscopy findings, and diagnostic accuracy metrics (specificity, sensitivity). Extraction will be conducted by two independent reviewers to ensure accuracy, with a third reviewer available to resolve any discrepancies. Data will be recorded in a structured digital database, which facilitates data management and subsequent analysis.

Assessment of Study Characteristics

Eligible studies for inclusion in the review comprised prospective cohort studies, retrospective analyses evaluating the diagnostic accuracy of AI-assisted CE for detecting CD. To qualify, articles

needed to report estimates of overall diagnostic accuracy, sensitivity (%), and specificity (%), accompanied by 95% confidence intervals (CIs). There were no restrictions on the size of the studies.

Risk of Bias/Quality Assessment Tool

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was utilized to evaluate the methodological quality of the included articles[13]. This tool encompasses four domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed for high, low, or unclear risk of bias, while the first three domains were also evaluated for high, low, or unclear concerns regarding applicability. The first three areas were also evaluated as applicability issues. Each section is classified as having high, low, or unclear risk of bias. Two investigators independently evaluated the retrieved articles for eligibility, resolving any discrepancies through mutual consensus. Review Manager version 5.3 was employed to create the summary figure for the methodological quality assessment.

Data Synthesis Methods

The synthesis of data in this review will be tailored to quantitative analyses, given the technical and clinical variability in the studies. Quantitative synthesis, where data permits, will include a meta-analysis to aggregate diagnostic performance metrics such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and AUC from the receiver operating characteristic (ROC) analysis.

Meta-Analysis Approach

We used the MIDAS module within STATA (version 18) for statistical analysis. True positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) were constructed to 2x2 contingency tables from artificial intelligence diagnostics of CD patients. The bivariate mixed effects regression model is used for the following indicators: merged sensitivity, specificity, PLR, NLR, DOR, and AUC. Firstly, the heterogeneity of the included studies is evaluated by visually examining the merged comprehensive subject operating characteristic (SROC) curves, and asymmetric shapes indicate significant heterogeneity. Use Spearman correlation analysis to test for heterogeneity caused by threshold effects; Cochran's Q test and I^2 value test were used to investigate heterogeneity caused by non-threshold effects. If $I^2 < 50\%$, it can be considered that there is low heterogeneity between research results. In this case, a fixed effects model was used for merging; If $I^2 \geq 50\%$, it can be considered that there is high heterogeneity, and a random effects model is used for merging. Explore the sources of heterogeneity through subgroup analysis and meta regression analysis. Use Deek's funnel plot to evaluate publication bias, where $P<0.1$ indicates asymmetric funnel plot. $P \leq 0.05$ is considered statistically significant. Simultaneously using Fagan's column chart to evaluate the role

and clinical value of AI assisted systems in the diagnosis of CD.

Result

Identification of Relevant Studies

A total of 155 articles were identified through a search across four electronic databases. Of these, 36 were found to be duplicates and were removed. During the initial screening, which involved a review of titles and abstracts, 106 articles were excluded. The full texts of the remaining 13 articles were thoroughly reviewed. Out of these, five studies were excluded from the final analysis because they did not align with the focus of this systematic review, which is the role of capsule endoscopy in the assessment of CD. Thus, eight studies were included in the final analysis (Figure.1).

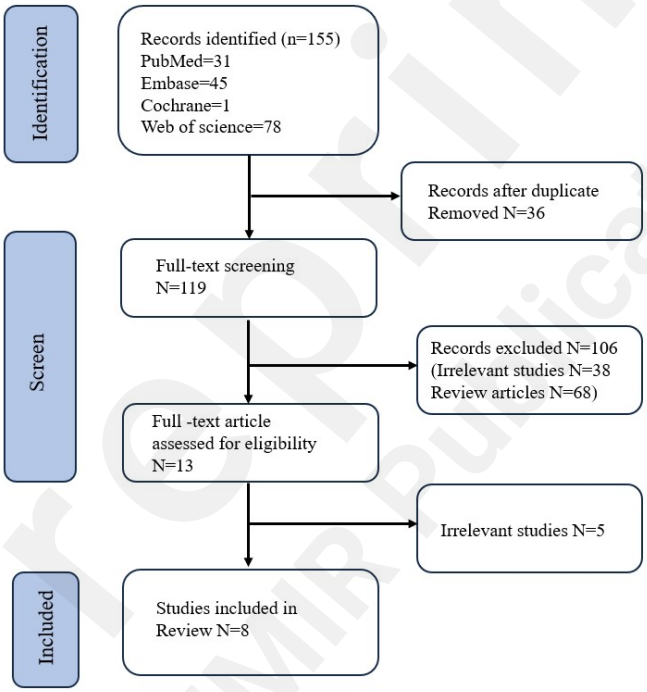


Figure 1. Flow diagram of search methodology and literature selection process.

Characteristics of the Included Studies

All studies assessed the performance of their AI algorithms using an internal validation test dataset, none reported external validation performance. Among the 8 studies diagnosing CD using CE images, a total of 444 patients (353 with CD and 91controls) were identified from 2020 to 2024, . Of these studies, 6 were retrospective and 2 were prospective. Most studies developed AI algorithms using deep learning, while three studies employed CNNs algorithms. Almost exclusively, these studies aimed to evaluate the efficacy of the algorithm in CD patients. Klang, E[14]. et al. selected CE images from 10 CD patients to train a model for detecting intestinal stenosis lesions, while Marin Santos, D[15]. et al. utilized CE images from 38 CD patients to identify various lesions in the

intestinal mucosa. An exception to this trend is the study conducted by Brodersen, J. B[16] et al., which notably did not exclude patients with ulcerative colitis and cancer. The number of participants in each study displayed a wide range, varying from a minimum of 10 to a maximum of 133 individuals. Moreover, a limitation across these retrospective studies was the absence of reporting the average age and gender of the participants. Most studies established the AI algorithm based on the deep learning. The included studies could be categorized by analysis based on design, AI algorithm, the number of enrolled patients. These characteristics were evaluated as potential sources of heterogeneity through the subgroup analysis and meta-regression. Detailed characteristics of the studies are presented in **Table 1**.

TABLE 1. Summary of studies in the literature review that applied AI techniques for CE image analysis

Study	Year of publication	Study type	Type of AI		Comparator	Total Dataset size	TP	FP	FN	TN
Klang, E[17]	2020	Retrospective study	CNN	DL Fold1	2 gastroenterology fellows	17,640 images of 49 patients	6837	195	554	10054
				DL Fold2			6933	338	458	9911
				DL Fold3			7177	410	214	9839
				DL Fold4			6999	256	392	9993
				DL Fold5			7154	348	237	9901
Vallée, R[18]	2020	Retrospective study	Neural networks	ResNet 34	3 independent experts	3498 images of 63 patients	1506	89	124	1645
				ResNext			1509	112	121	1622
				VGGNet 19			1491	109	139	1625
				VGGNet 16			1493	115	137	1619
De Maissin, A [19]	2021	Retrospective study	Neural Network	CRANN	1 gastroenterology fellow and 3 experts.	3498 images of 63 patients	1241	76	119	2048
				ResNet 34			1241	81	119	2043
				VGGNet 16			1206	89	154	2035
				VGGNet 19			1199	101	161	2023
Klang E[14]	2021	Retrospective study	Neural Network		a capsule expert	27,89 images of 10 patients	815	590	71	4772
Majtner, T[20]	2022	Prospective study	Deep learning (ResNet-50)	Small bowel	3 gastroenterologists	774 images from 38 CD patients	359	0	15	620
				Colon			174	0	6	302
				Overall			533	0	21	922
Ferreira, J. P. S.[21]	2022	Retrospective study	CNN (Xception model)		3 gastroenterologists	2449 images of 59 patients	5194	192	106	18998
Marin-Santos, D[15]	2023	Retrospective study	CNN		2 digestive tract specialists	15,972 images from 31 CD patients	1581	59	16	1538
Brodersen, J. B.[16]	2024	Prospective study	Deep learning		2 GI-Specialists	86129 images from 131 CD patients	49	6	2	74

(AXARO framework)	48	8	4	71

Quality evaluation of included literature

Regarding the QUADAS-2 tool, the quality assessment results of included studies are summarized in Figure2. Among the 8 studies included in the final analysis, 2 demonstrated a low risk of bias, 6 study at unclear risk. Bias in the included studies primarily stems from the domains of patient selection, index test, flow and timing.

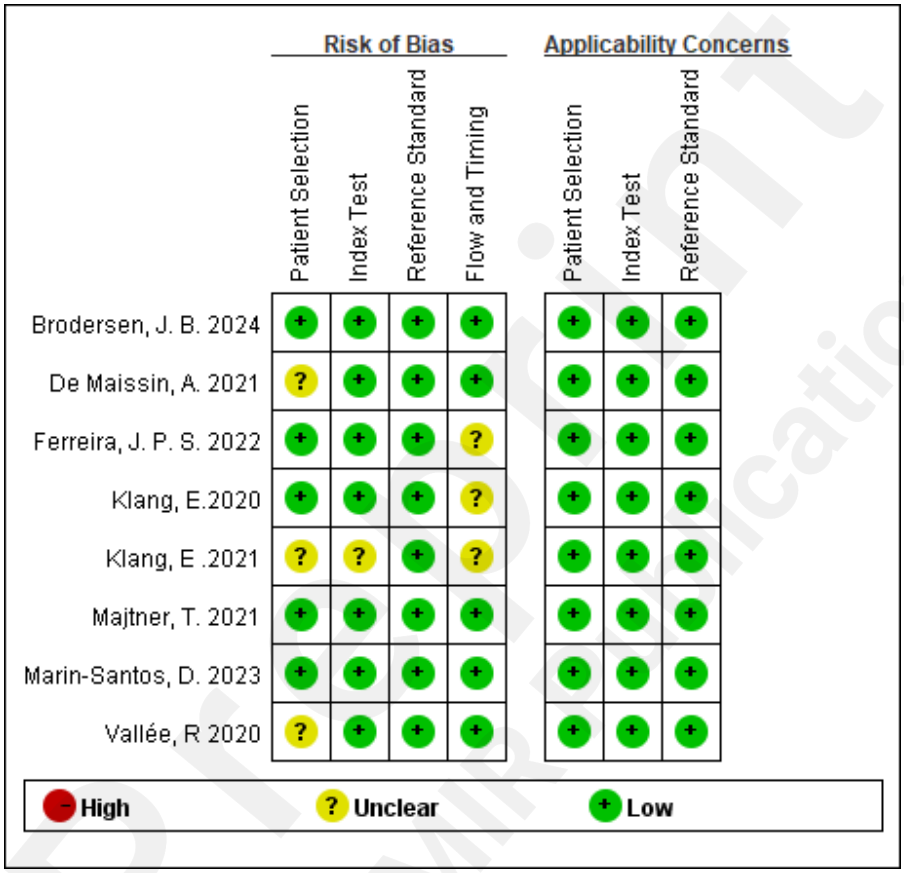


Table 2. Quality Assessment of Diagnostic Accuracy Studies-2 risk for the assessment of the methodological qualities. (+) denotes low risk of bias, (?) denotes unclear risk of bias, (-) denotes high risk of bias.

Diagnostic Test Accuracy of Artificial Intelligence for the Crohn's disease

Among the 8 studies of patient-based analysis, the sensitivity, specificity, PLR, NLR, DOR, and area under the curve (AUC) with 95% CI of AI for the CD were 94% (95% CI, 93%-96%), 97% (95% CI, 95%-98%), 32.7 (95% CI, 19.9-53.6), 6% (95% CI, 4%-7%), 576 (95% CI, 295-1127) and 99% (95% CI, 100% - 0.00) respectively in Figure 2 .To investigate the clinical utility of AI, a Fagan nomogram was generated.

Assuming 20% diagnostic of CD the Fagan nomogram shows that the posterior probability of diagnostic of CD was 89% if the test was positive, and the posterior probability of absence of CD was 100% if the test was negative.

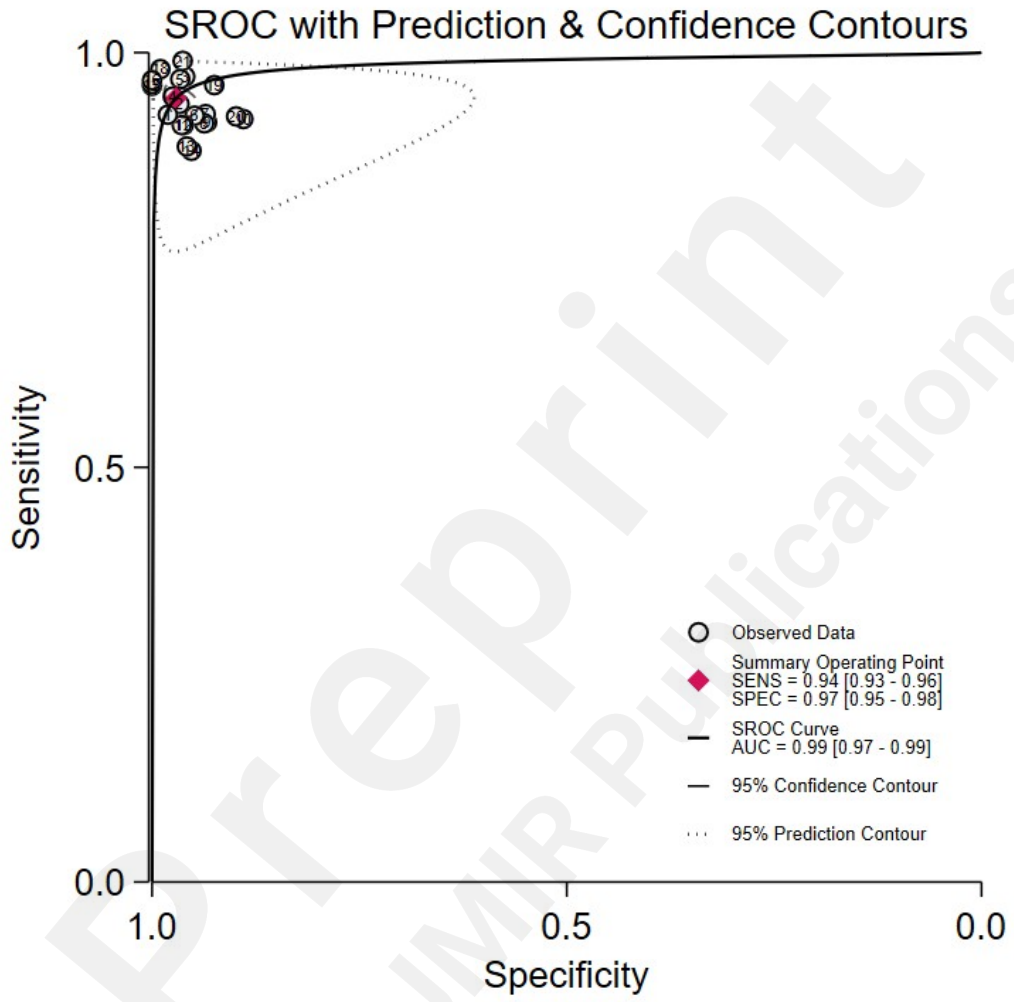
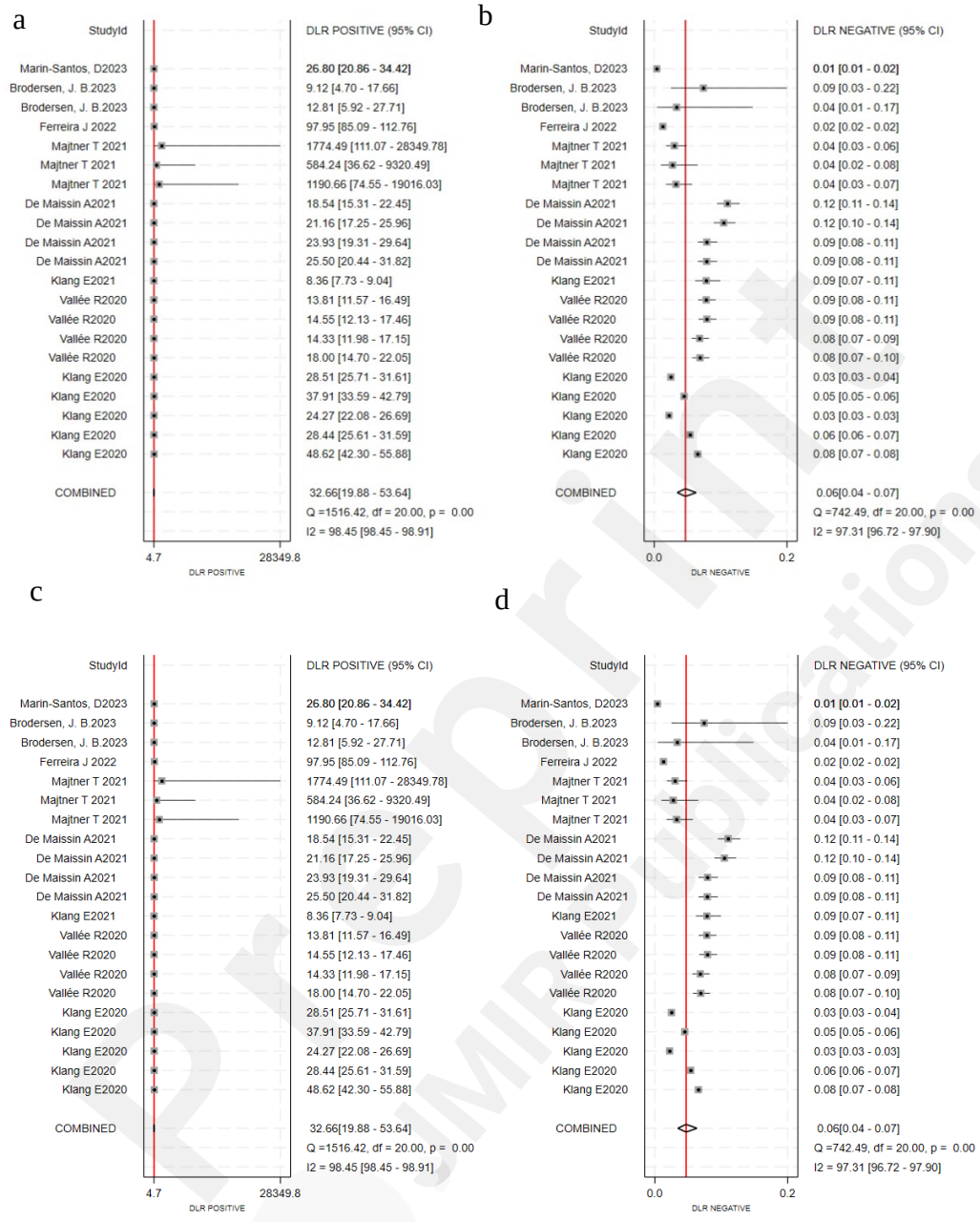


Figure 2: Summary receiver operating characteristic (SROC) curve with 95% confidence and prediction regions for Crohn's disease detection in capsule endoscopy.


a
StudyId
DLR POSITIVE (95% CI)
Marin-Santos, D2023 26.80 [20.86 - 34.42]
Brodersen, J. B.2023 9.12 [4.70 - 17.66]
Brodersen, J. B.2023 12.81 [5.92 - 27.71]
Ferreira J 2022 97.95 [85.09 - 112.76]
Majtner T 2021 1774.49 [111.07 - 28349.78]
Majtner T 2021 584.24 [36.62 - 9320.49]
Majtner T 2021 1190.66 [74.55 - 19016.03]
De Maissin A2021 18.54 [15.31 - 22.45]
De Maissin A2021 21.16 [17.25 - 25.96]
De Maissin A2021 23.93 [19.31 - 29.64]
De Maissin A2021 25.50 [20.44 - 31.82]
Klang E2021 8.36 [7.73 - 9.04]
Vallée R2020 13.81 [11.57 - 16.49]
Vallée R2020 14.55 [12.13 - 17.46]
Vallée R2020 14.33 [11.98 - 17.15]
Vallée R2020 18.00 [14.70 - 22.05]
Klang E2020 28.51 [25.71 - 31.61]
Klang E2020 37.91 [33.59 - 42.79]
Klang E2020 24.27 [22.08 - 26.69]
Klang E2020 28.44 [25.61 - 31.59]
Klang E2020 48.62 [42.30 - 55.88]
COMBINED 32.66[19.88 - 53.64]
Q =1516.42, df = 20.00, p = 0.00
I2 = 98.45 [98.45 - 98.91]
4.7
28349.8
DLR POSITIVE
b
StudyId
DLR NEGATIVE (95% CI)
Marin-Santos, D2023 0.01 [0.01 - 0.02]
Brodersen, J. B.2023 0.09 [0.03 - 0.22]
Brodersen, J. B.2023 0.04 [0.01 - 0.17]
Ferreira J 2022 0.02 [0.02 - 0.02]
Majtner T 2021 0.04 [0.03 - 0.06]
Majtner T 2021 0.04 [0.02 - 0.08]
Majtner T 2021 0.04 [0.03 - 0.07]
De Maissin A2021 0.12 [0.11 - 0.14]
De Maissin A2021 0.12 [0.10 - 0.14]
De Maissin A2021 0.09 [0.08 - 0.11]
De Maissin A2021 0.09 [0.08 - 0.11]
Klang E2021 0.09 [0.07 - 0.11]
Vallée R2020 0.09 [0.08 - 0.11]
Vallée R2020 0.09 [0.08 - 0.11]
Vallée R2020 0.08 [0.07 - 0.09]
Vallée R2020 0.08 [0.07 - 0.10]
Klang E2020 0.03 [0.03 - 0.04]
Klang E2020 0.05 [0.05 - 0.06]
Klang E2020 0.03 [0.03 - 0.03]
Klang E2020 0.06 [0.06 - 0.07]
Klang E2020 0.08 [0.07 - 0.08]
COMBINED 0.06[0.04 - 0.07]
Q =742.49, df = 20.00, p = 0.00
I2 = 97.31 [96.72 - 97.90]
0.0
0.2
DLR NEGATIVE
c
StudyId
DLR POSITIVE (95% CI)
Marin-Santos, D2023 26.80 [20.86 - 34.42]
Brodersen, J. B.2023 9.12 [4.70 - 17.66]
Brodersen, J. B.2023 12.81 [5.92 - 27.71]
Ferreira J 2022 97.95 [85.09 - 112.76]
Majtner T 2021 1774.49 [111.07 - 28349.78]
Majtner T 2021 584.24 [36.62 - 9320.49]
Majtner T 2021 1190.66 [74.55 - 19016.03]
De Maissin A2021 18.54 [15.31 - 22.45]
De Maissin A2021 21.16 [17.25 - 25.96]
De Maissin A2021 23.93 [19.31 - 29.64]
De Maissin A2021 25.50 [20.44 - 31.82]
Klang E2021 8.36 [7.73 - 9.04]
Vallée R2020 13.81 [11.57 - 16.49]
Vallée R2020 14.55 [12.13 - 17.46]
Vallée R2020 14.33 [11.98 - 17.15]
Vallée R2020 18.00 [14.70 - 22.05]
Klang E2020 28.51 [25.71 - 31.61]
Klang E2020 37.91 [33.59 - 42.79]
Klang E2020 24.27 [22.08 - 26.69]
Klang E2020 28.44 [25.61 - 31.59]
Klang E2020 48.62 [42.30 - 55.88]
COMBINED 32.66[19.88 - 53.64]
Q =1516.42, df = 20.00, p = 0.00
I2 = 98.45 [98.45 - 98.91]
4.7
28349.8
DLR POSITIVE
d
StudyId
DLR NEGATIVE (95% CI)
Marin-Santos, D2023 0.01 [0.01 - 0.02]
Brodersen, J. B.2023 0.09 [0.03 - 0.22]
Brodersen, J. B.2023 0.04 [0.01 - 0.17]
Ferreira J 2022 0.02 [0.02 - 0.02]
Majtner T 2021 0.04 [0.03 - 0.06]
Majtner T 2021 0.04 [0.02 - 0.08]
Majtner T 2021 0.04 [0.03 - 0.07]
De Maissin A2021 0.12 [0.11 - 0.14]
De Maissin A2021 0.12 [0.10 - 0.14]
De Maissin A2021 0.09 [0.08 - 0.11]
De Maissin A2021 0.09 [0.08 - 0.11]
Klang E2021 0.09 [0.07 - 0.11]
Vallée R2020 0.09 [0.08 - 0.11]
Vallée R2020 0.09 [0.08 - 0.11]
Vallée R2020 0.08 [0.07 - 0.09]
Vallée R2020 0.08 [0.07 - 0.10]
Klang E2020 0.03 [0.03 - 0.04]
Klang E2020 0.05 [0.05 - 0.06]
Klang E2020 0.03 [0.03 - 0.03]
Klang E2020 0.06 [0.06 - 0.07]
Klang E2020 0.08 [0.07 - 0.08]
COMBINED 0.06[0.04 - 0.07]
Q =742.49, df = 20.00, p = 0.00
I2 = 97.31 [96.72 - 97.90]
0.0
0.2
DLR NEGATIVE

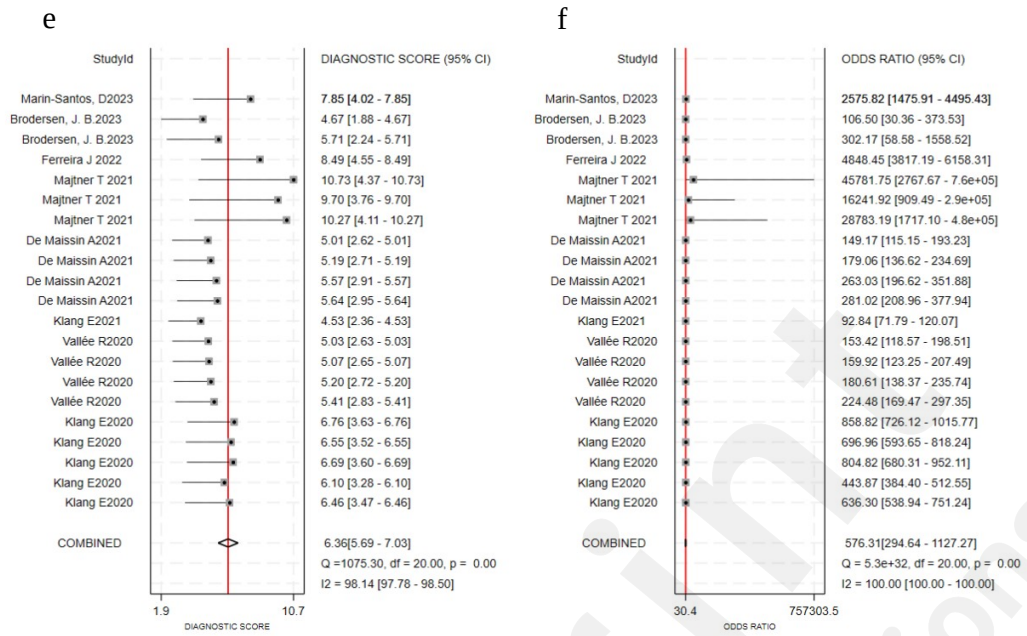


Figure 3. Forest plots reveal sensitivity, specificity, positive likelihood ratio (PLR) , negative likelihood ratio (NLR), diagnostic score and diagnostic odds ratio(DOR) estimates of Artificial Intelligence for Crohn's disease in capsule endoscopy images. sensitivity (a), specificity (b), PLR(c), NLR (d), diagnostic score (e), and DOR(f).

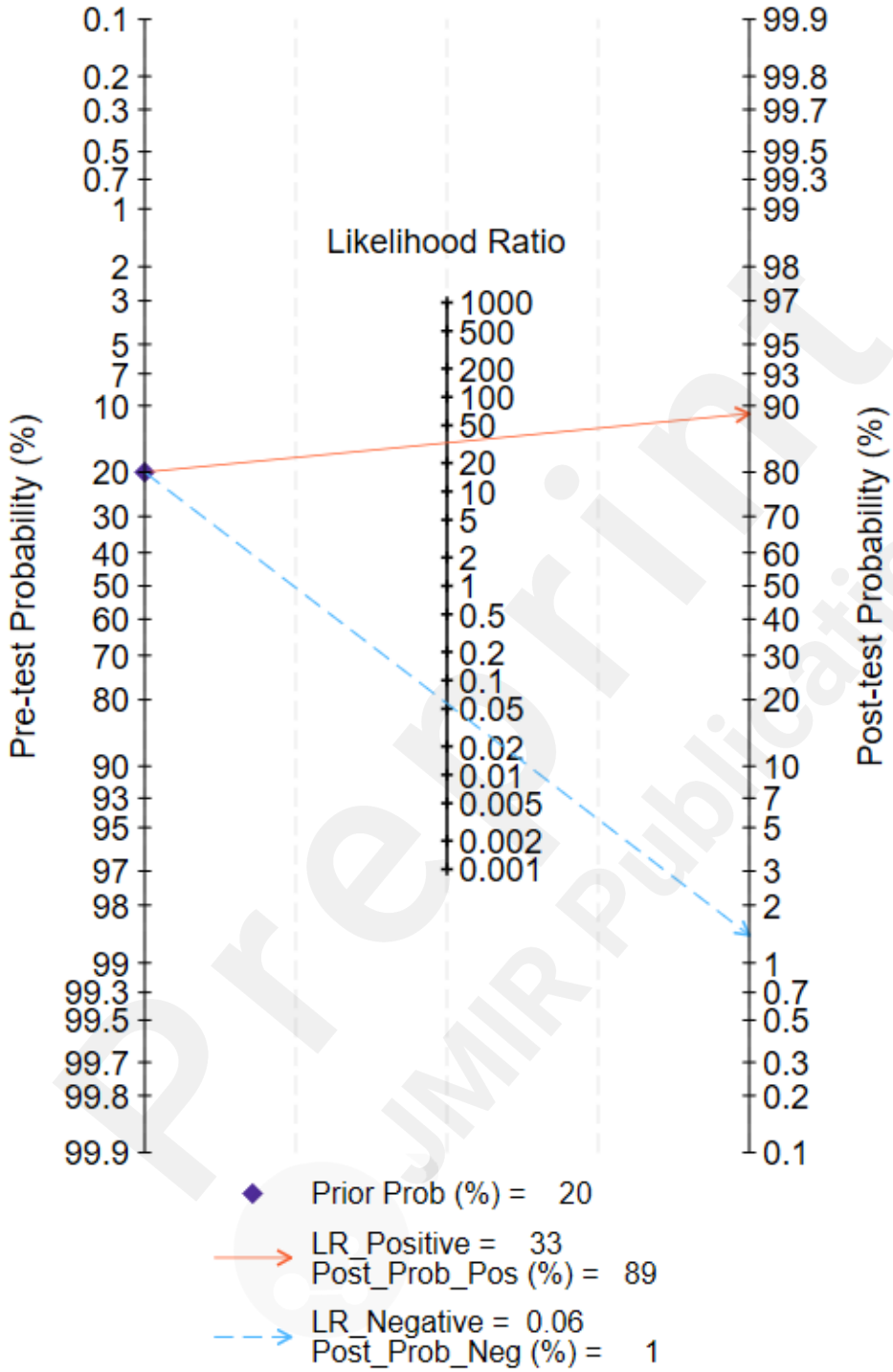


Figure 4. Fagan nomogram for the diagnosis of Crohn's disease in capsule endoscopy images.

Exploring Heterogeneity with Meta-Regression and Subgroup Analysis

For the diagnosis of CD using CE images, the design among the included studies (P < 0.001), the total number of included patients (P < 0.001) and the AI algorithm (P < 0.001) were identified as sources of heterogeneity in model of meta-regression.

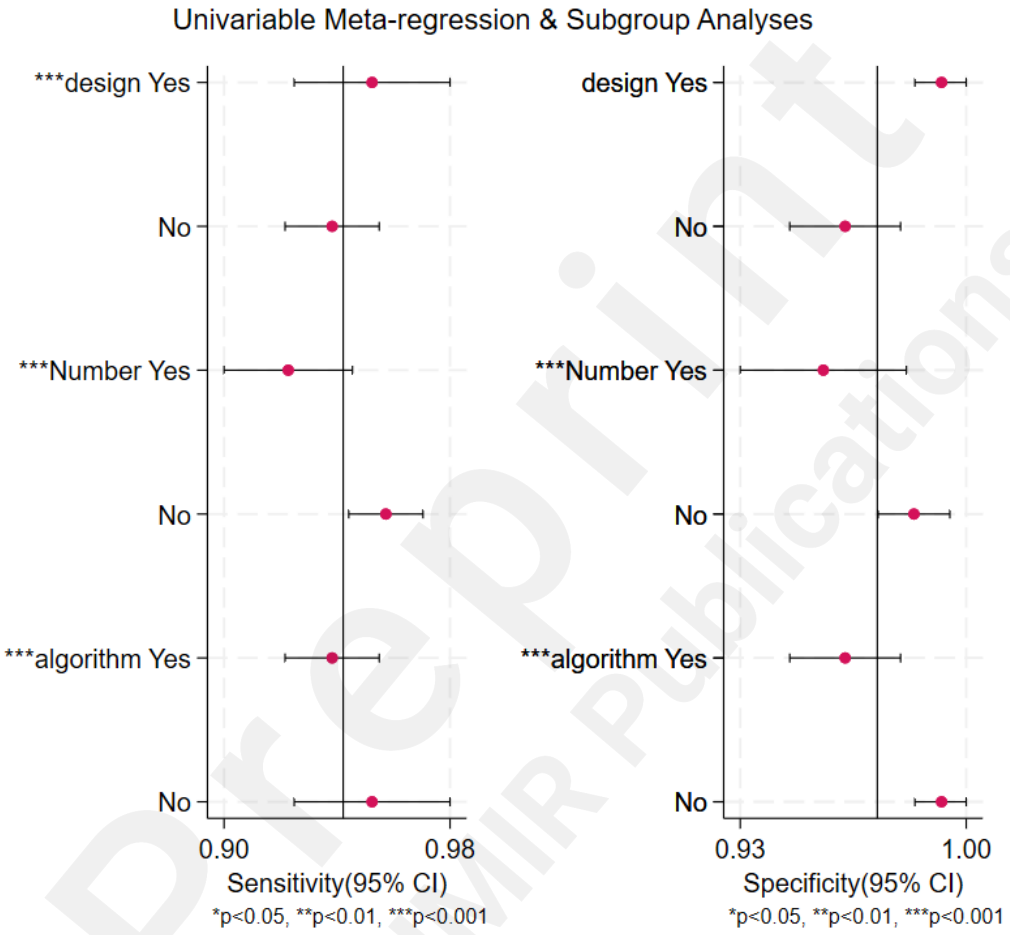


Figure 5. Meta-regression analysis identifying sources of heterogeneity in the diagnostic test accuracy meta-analysis.

Publication Bia

We performed a publication bias analysis for the included studies. No significant bias was observed in diagnostic Crohn's disease efficacy of AI via capsule endoscopic images.

Figure shows the Deek' funnel plot indicated no evidence of publication bias (P = 0.56).

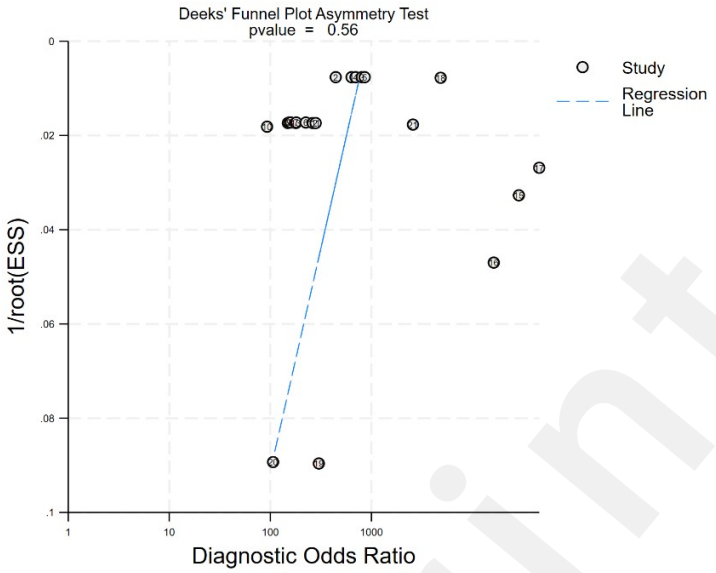


Figure 6. Deek' funnel plot for the studies of assessing publication bias (*P*= 0.56), indicating no significant publication bias. Each dot represents an individual study.

Discussion

CD is an autoimmune disorder that affects the entire digestive tract and can progress rapidly. CE offers significant advantages for diagnosing CD, as it minimizes the need for multiple invasive procedures and reduces patient discomfort. However, CE produces a large volume of images, some of which may be blurry or irrelevant. Moreover, CD gastrointestinal mucosal lesions include various manifestations such as edema, ulcers, fistulas, stenosis, etc., all of which contribute to the complexity and challenge of making an accurate diagnosis. Recent studies have demonstrated the substantial potential of AI in medical diagnostics[7, 22, 23]. Traditionally, CE diagnosis is operator-dependent and subjective, but AI-assisted endoscopy can provide a valuable second opinion, potentially reducing operator dependency. This approach, characterized as computer-aided diagnosis, significantly improves the diagnostic accuracy of capsule endoscopy. Endoscopist combined with AI analysis will enhance the probability of uncovering crucial gastrointestinal mucosal lesions under endoscopy, thereby advancing the accuracy and effectiveness of diagnostic assessments. Both patient-based and image-based analyses commonly presented good

AI performance for diagnose in CD. To enhance diagnostic capabilities, AI must be trained and rigorously validated using vast amounts of data, allowing it to learn the intricacies of these diverse manifestations and ultimately improve its diagnostic accuracy. Therefore, more prospective studies focusing on the application of AI in clinical practice, particularly in diagnostic equipment, is essential.

This article provides a comprehensive review of the diverse applications of AI in diagnosing CD patients through the analysis of CE images. A significant finding of this study is the robustness of the AI algorithm's diagnostic performance, while studies involving a large patient population and high methodological quality demonstrated higher diagnostic performance, the difference was not substantial. Despite employing a comprehensive and thorough search strategy, several limitations must be acknowledged regarding both the evidence and the review itself. We were able to identify only 8 eligible studies. any of these studies presented an unclear risk of bias due to their design. Additionally, most studies either had small sample sizes or included only patients with CD, which may have compromised the precision of the effect estimates. However, the studies included in the meta-analysis lack external data validation, which compromises the reliability of their findings. Most studies utilize deep learning as the AI algorithm, with only 3 studies employing CNNs. Additional, the heterogeneity observed in this review is notably high and is attributed to several factors, including the type of AI algorithm, the number of participants, and the research methodology. These factors have been further validated through subgroup analysis.

Conclusion

AI demonstrates high overall diagnostic accuracy for CD patients in capsule endoscopy. However, heterogeneity and small sample sizes compromise the quality and validity of these findings. Moreover, the currently lack sufficient data to accurately identify intestinal mucosal lesions such as ulcers, stenosis, and bleeding. Further research, including rigorous patients selection and randomized clinical trials,

is needed to fully assess the effectiveness of AI in capsule endoscopy. CNNs, a novel recognition algorithm based on artificial neural networks and deep learning theory, hold the potential to transform its treatment in the diagnosis of CD. This systematic review offers a comprehensive overview of these models, serving as a foundation for future research in this field. Such studies could help promote more individualized and flexible screening options for suspected CD patients, to minimize invasive examination.

Author Contributions

Yuling Bin conceptualized and designed the study, conducted the majority of data collection, analyzed the results, and drafted the initial manuscript. Rumei Peng contributed to refining the study design by providing valuable insights and assisted in data collection and analysis. Yaqian Lee collaborated closely with Yuling Bin in drafting and editing the manuscript. Zhijie Lee contributed to interpreting complex findings, offering alternative perspectives and explanations, and assisted in preparing the figures and tables to ensure they accurately represented the study's outcomes. Yang Liu actively engaged in finalizing the manuscript in response to reviewers' comments and feedback.

Conflict of interest statement

The authors affirm their commitment to ensuring the impartiality and objectivity of this work, rigorously guarding against any potential conflicts of interest that may compromise the integrity of our findings.

Reference

[1]. Torres, J., S. Mehandru, J. F. Colombel and L. Peyrin-Biroulet (2017). "Crohn's disease." Lancet **389**(10080): 1741-1755.

[2]. Feuerstein, J. D. and A. S. Cheifetz (2017). "Crohn Disease: Epidemiology, Diagnosis, and Management." Mayo Clin Proc **92**(7): 1088-1103.

[3]. Kopylov, U., D. E. Yung, T. Engel, S. Vijayan, O. Har-Noy, L. Katz, et al. (2017). "Diagnostic yield of capsule endoscopy versus magnetic resonance enterography and small bowel contrast ultrasound in the evaluation of small bowel Crohn's disease: Systematic review and meta-analysis." Dig Liver Dis **49**(8): 854-863.

[4]. Tamilarasan, A. G., Y. Tran, S. Paramsothy and R. Leong (2022). "The diagnostic yield of pan-enteric capsule endoscopy in inflammatory bowel disease: A systematic review and meta-analysis." J Gastroenterol Hepatol **37**(12): 2207-2216.

[5]. Brodersen, J. B., J. Kjeldsen, M. A. Juel, T. Knudsen, S. R. Rafaelsen and M. D. Jensen (2024). "Changes in endoscopic activity and classification of lesions with panenteric capsule endoscopy in patients treated for Crohn's disease - a prospective blinded comparison with

ileocolonoscopy, faecal calprotectin and C-reactive protein." J Crohns Colitis.

[6]. Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al. (2017). "A survey on deep learning in medical image analysis." Med Image Anal **42**: 60-88.

[7]. Khan, S., K. H. Barve and M. S. Kumar (2020). "Recent Advancements in Pathogenesis, Diagnostics and Treatment of Alzheimer's Disease." Curr Neuropharmacol **18**(11): 1106-1125.

[8]. Xie, F., K. Zhang, F. Li, G. Ma, Y. Ni, W. Zhang, et al. (2022). "Diagnostic accuracy of convolutional neural network-based endoscopic image analysis in diagnosing gastric cancer and predicting its invasion depth: a systematic review and meta-analysis." Gastrointest Endosc **95**(4): 599-609.e597.

[9]. Jahagirdar, V., J. Bapaye, S. Chandan, S. Ponnada, G. S. Kochhar, U. Navaneethan, et al. (2023). "Diagnostic accuracy of convolutional neural network-based machine learning algorithms in endoscopic severity prediction of ulcerative colitis: a systematic review and meta-analysis." Gastrointest Endosc **98**(2): 145-154.e148.

[10]. Aziz, M., R. Fatima, C. Dong, W. Lee-Smith and A. Nawras (2020). "The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis." J Gastroenterol Hepatol **35**(10): 1676-1683.

[11]. Keshtkar, K., A. R. Safarpour, R. Heshmat, R. Sotoudehmanesh and A. Keshtkar (2023). "A Systematic Review and Meta-analysis of Convolutional Neural Network in the Diagnosis of Colorectal Polyps and Cancer." Turk J Gastroenterol **34**(10): 985-997.

[12]. Carter, D., A. Albshesh, C. Shimon, B. Segal, A. Yershov, U. Kopylov, et al. (2023). "Automatized Detection of Crohn's Disease in Intestinal Ultrasound Using Convolutional Neural Network." Inflamm Bowel Dis **29**(12): 1901-1906.

[13]. Whiting, P. F., A. W. Rutjes, M. E. Westwood, S. Mallett, J. J. Deeks, J. B. Reitsma, et al. (2011). "QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies." Ann Intern Med **155**(8): 529-536.

[14]. Klang, E., A. Grinman, S. Soffer, R. Margalit Yehuda, O. Barzilay, M. M. Amitai, et al. (2021). "Automated Detection of Crohn's Disease Intestinal Strictures on Capsule Endoscopy Images Using Deep Neural Networks." J Crohns Colitis **15**(5): 749-756.

[15]. Marin-Santos, D., J. A. Contreras-Fernandez, I. Perez-Borrero, H. Pallares-Manrique and M. E. Gegundez-Arias (2023). "Automatic detection of crohn disease in wireless capsule endoscopic images using a deep convolutional neural network." Applied Intelligence **53**(10): 12632-12646.

[16]. Brodersen, J. B., M. D. Jensen, R. Leenhardt, J. Kjeldsen, A. Histace, T. Knudsen, et al. (2024). "Artificial Intelligence-assisted Analysis of Pan-enteric Capsule Endoscopy in Patients with Suspected Crohn’s Disease: A Study on Diagnostic Performance." Journal of Crohn's and Colitis **18**(1): 75-81.

[17]. Klang, E., Y. Barash, R. Y. Margalit, S. Soffer, O. Shimon, A. Albshesh, et al. (2020). "Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy." Gastrointest Endosc **91**(3): 606-613.e602.

[18]. Vallée, R., A. De Maissin, A. Coutrot, H. Mouchère, A. Bourreille and N. Normand (2020). CrohnIPI: An endoscopic image database for the evaluation of automatic Crohn's disease lesions recognition algorithms. SPIE Medical Imaging Conference - Biomedical Applications in Molecular, Structural, and Functional Imaging, Houston, TX.

[19]. De Maissin, A., R. Valleé, M. Flamant, M. Fondain-Bossiere, C. L. Berre, A. Coutrot, et al. (2021). "Multi-expert annotation of Crohn's disease images of the small bowel for automatic detection

using a convolutional recurrent attention neural network." Endoscopy International Open **9**(7): E1136-E1144.

[20]. Majtner, T., J. B. Brodersen, J. Herp, J. Kjeldsen, M. L. Halling and M. D. Jensen (2021). "A deep learning framework for autonomous detection and classification of Crohn's disease lesions in the small bowel and colon with capsule endoscopy." Endosc Int Open **9**(9): E1361-e1370.

[21]. Ferreira, J. P. S., Mjdqec De Mascarenhas Saraiva, J. P. L. Afonso, T. F. C. Ribeiro, H. M. C. Cardoso, A. P. Ribeiro Andrade, et al. (2022). "Identification of Ulcers and Erosions by the Novel Pillcam™ Crohn's Capsule Using a Convolutional Neural Network: A Multicentre Pilot Study." J Crohns Colitis **16**(1): 169-172.

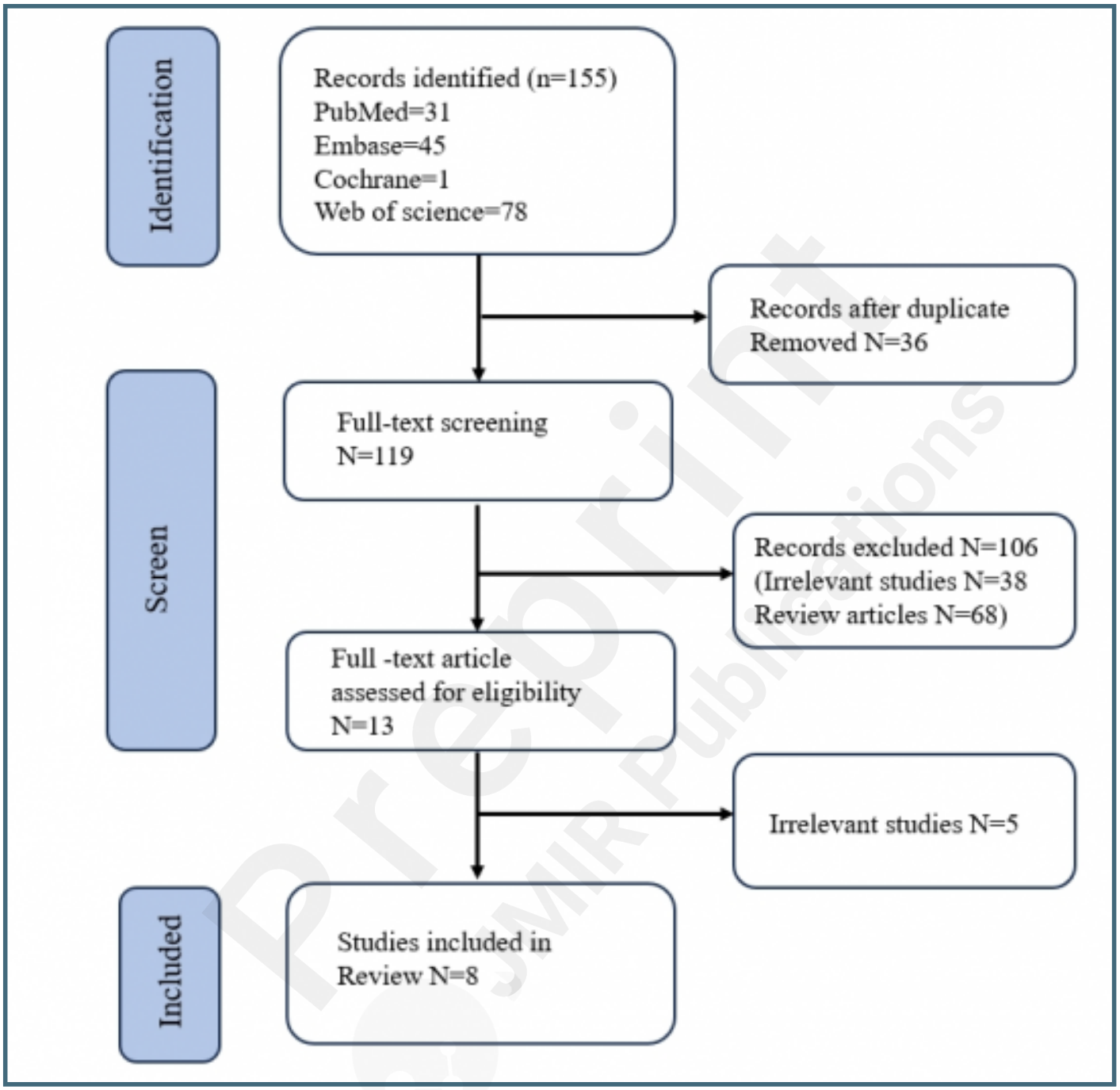
[22]. Mitsala, A., C. Tsalikidis, M. Pitiakoudis, C. Simopoulos and A. K. Tsaroucha (2021). "Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era." Curr Oncol **28**(3): 1581-1607.

[23]. Huang, X., H. Wang, C. She, J. Feng, X. Liu, X. Hu, et al. (2022). "Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy." Front Endocrinol (Lausanne) **13**: 946915.

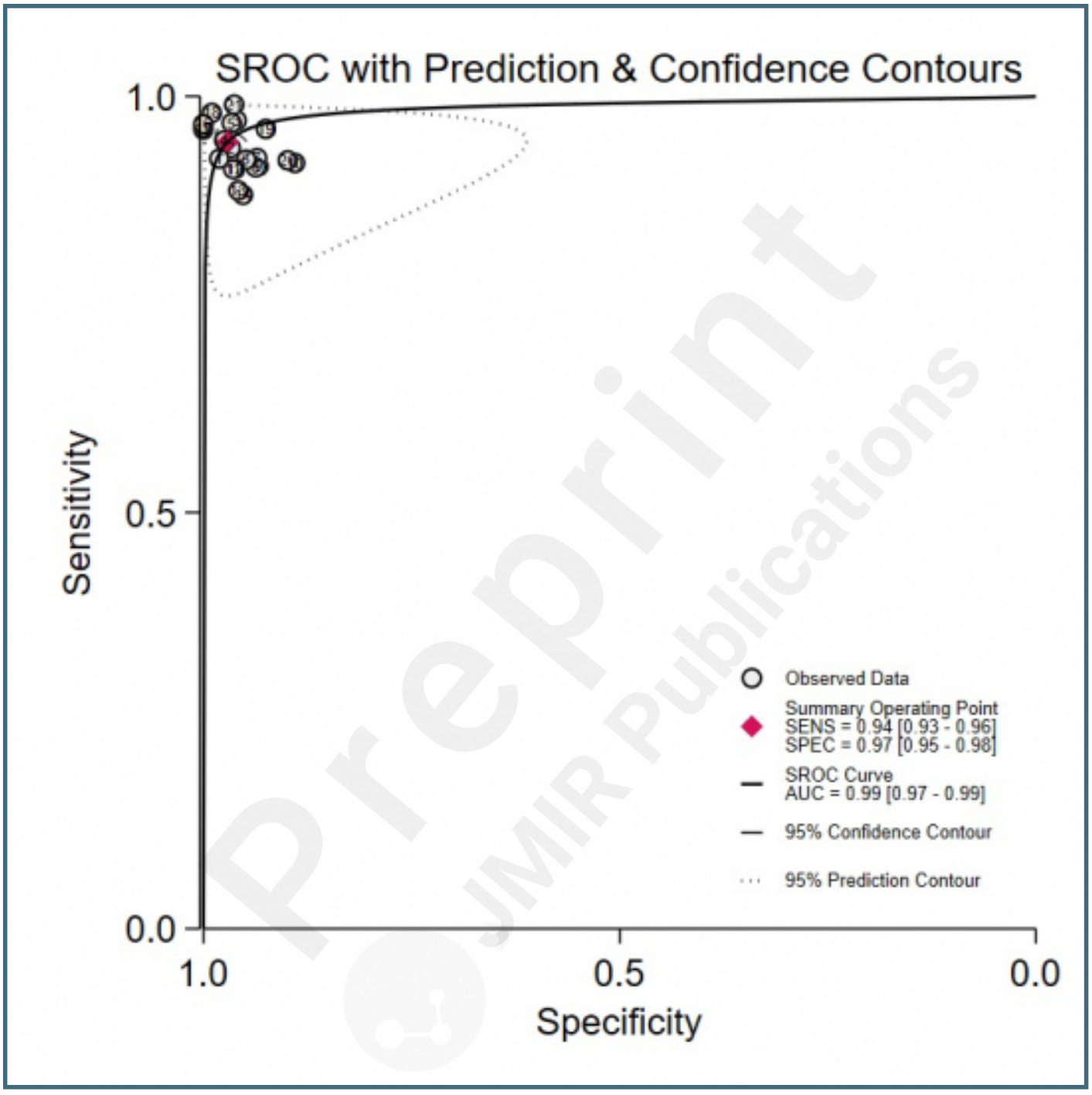
Supplementary Files

Figures

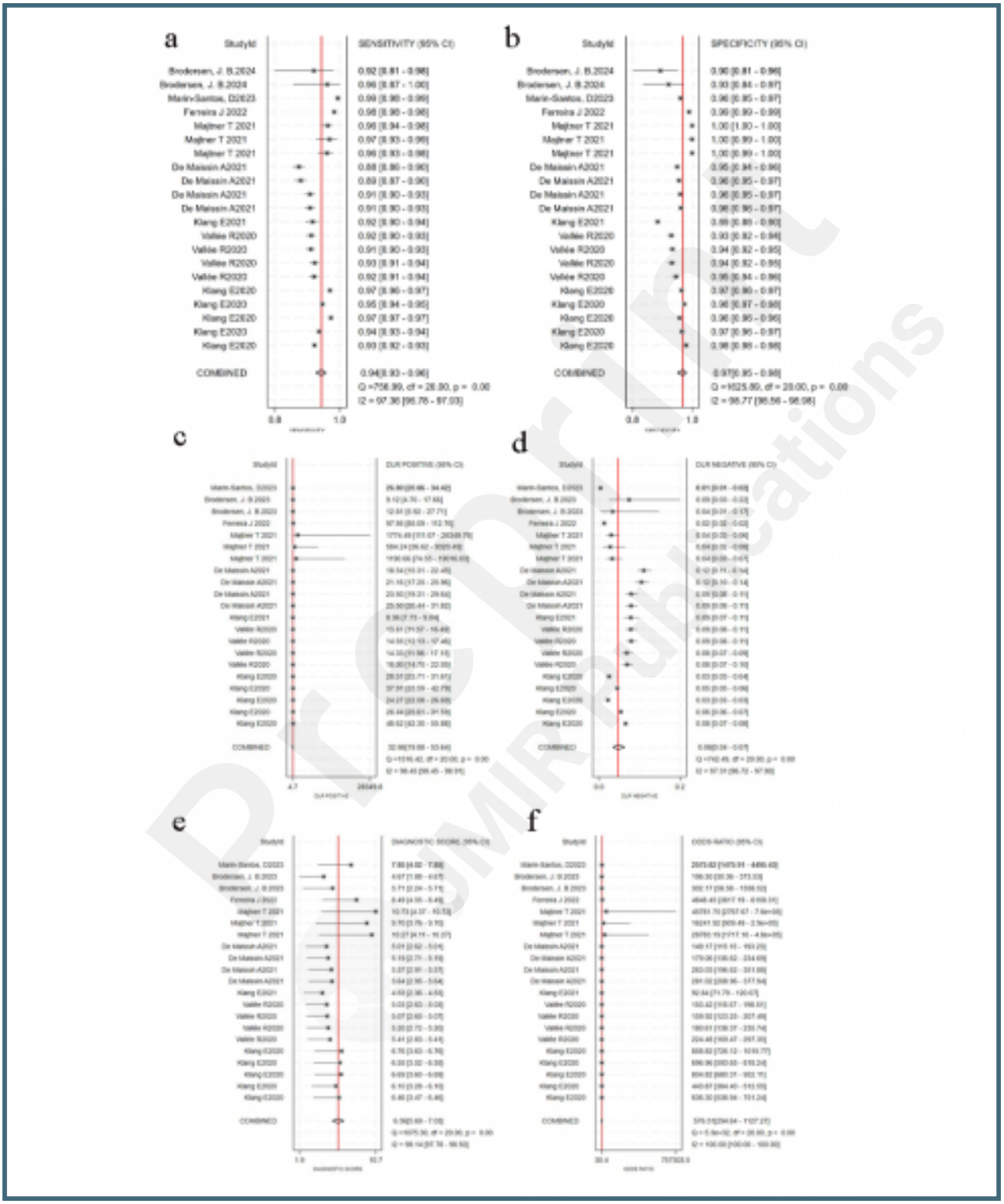
Flow diagram of search methodology and literature selection process.



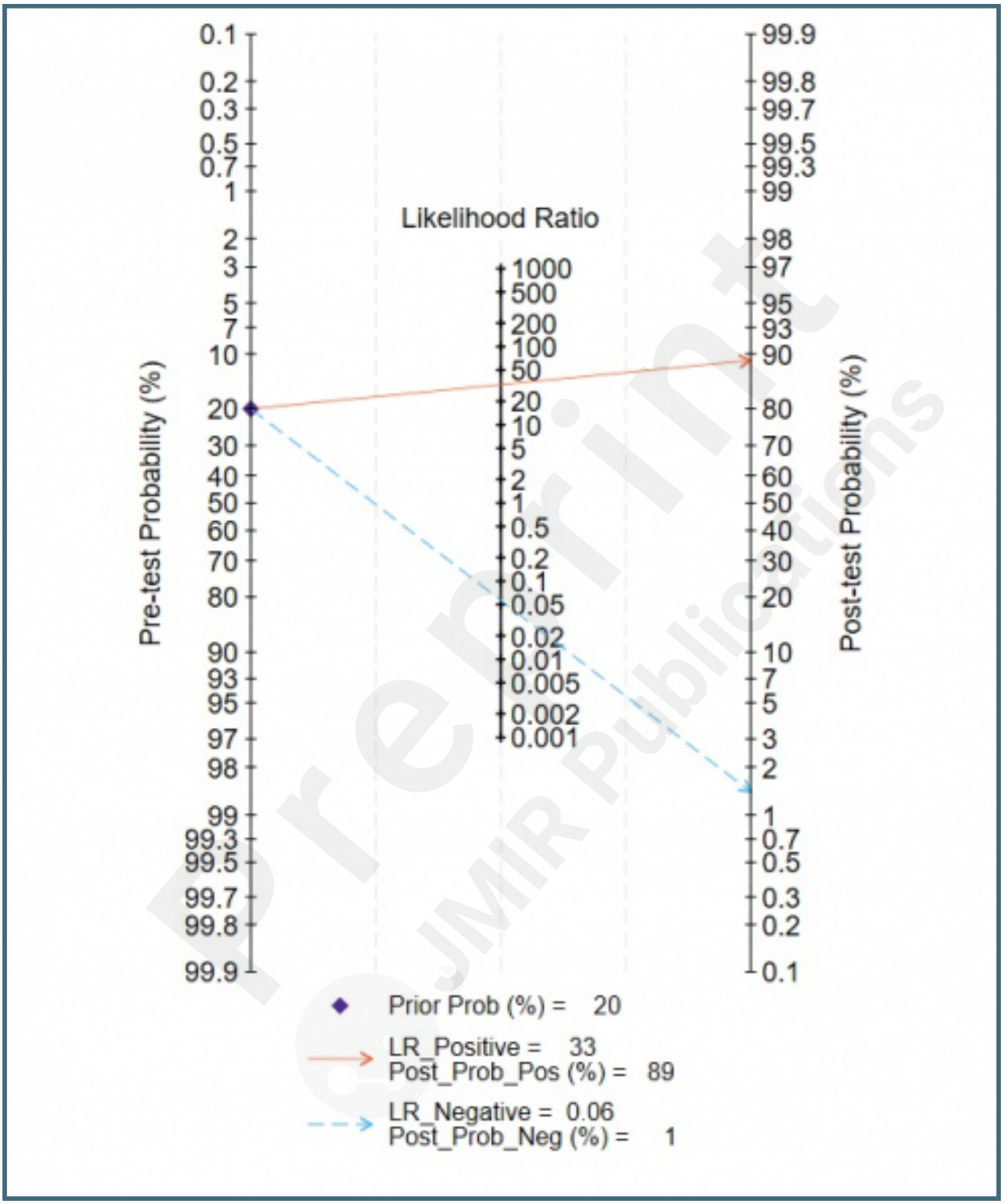
Summary receiver operating characteristic (SROC) curve with 95% confidence and prediction regions for Crohn's disease detection in capsule endoscopy.



Forest plots reveal sensitivity, specificity, positive likelihood ratio (PLR) , negative likelihood ratio (NLR), diagnostic score and diagnostic odds ratio(DOR) estimates of Artificial Intelligence for Crohn's disease in capsule endoscopy images. sensitivity (a), specificity (b), PLR(c), NLR (d), diagnostic score (e), and DOR(f).



Fagan nomogram for the diagnosis of Crohn's disease in capsule endoscopy images.



Meta-regression analysis identifying sources of heterogeneity in the diagnostic test accuracy meta-analysis.

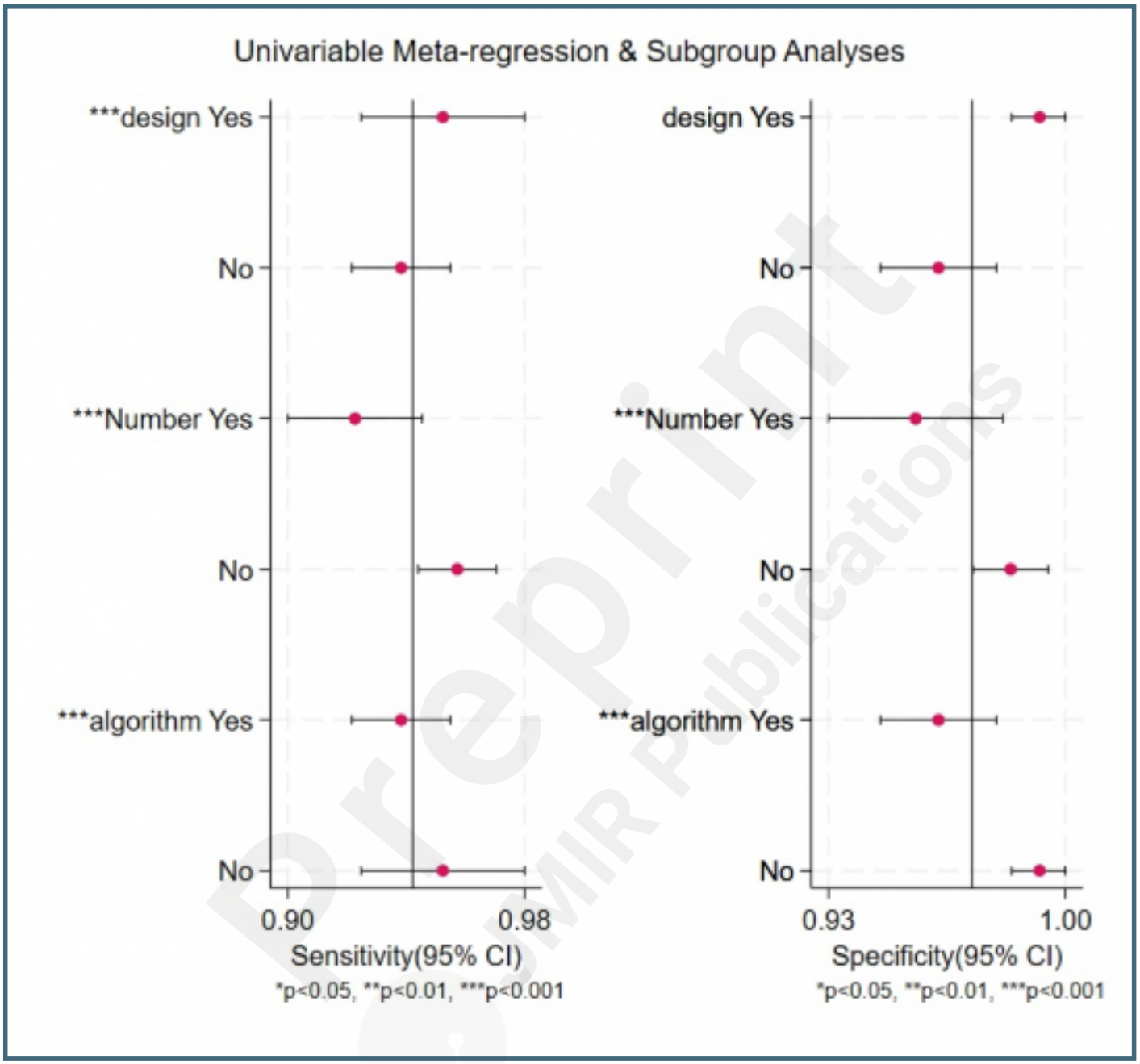
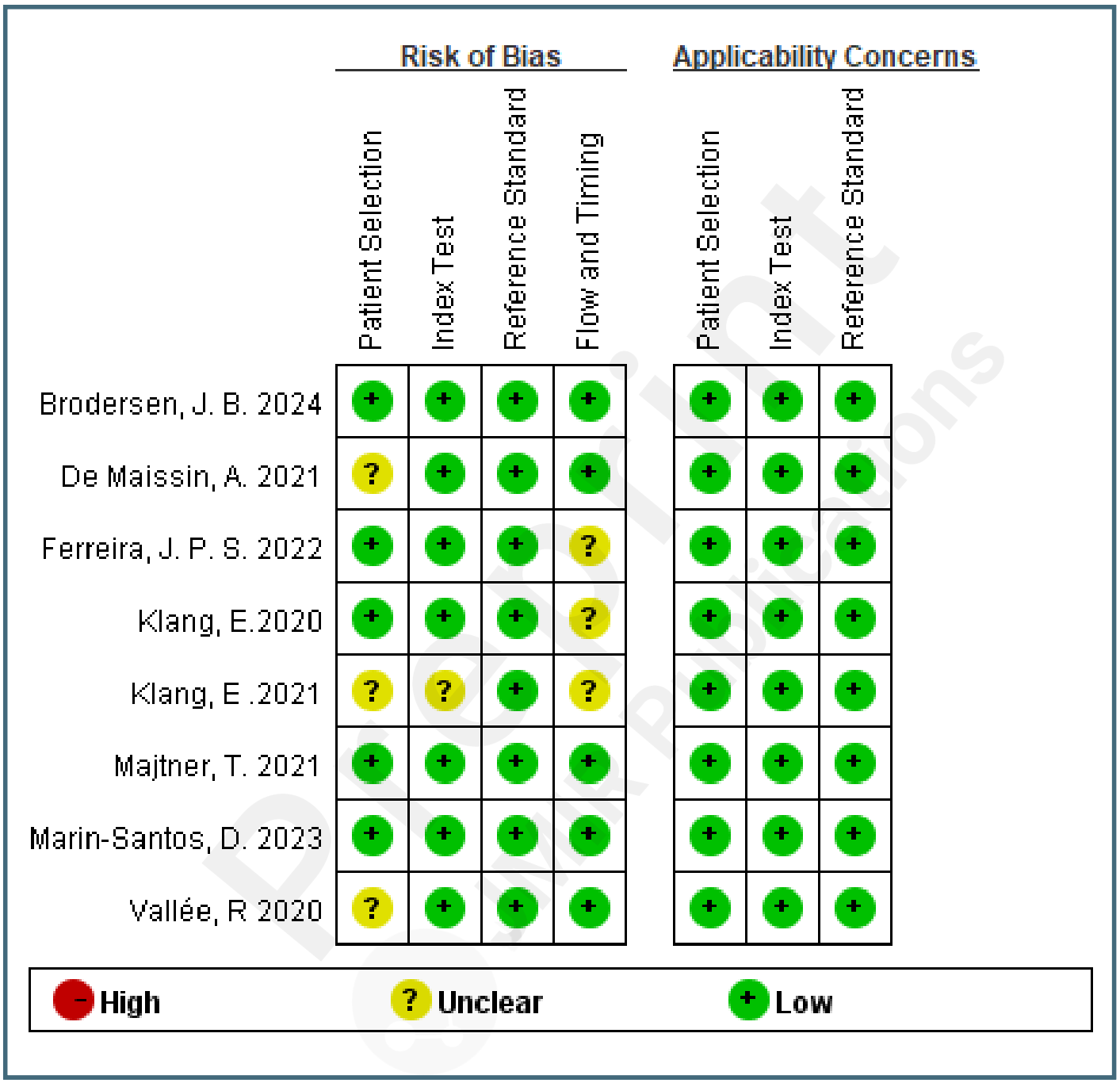


Table 2. Quality Assessment of Diagnostic Accuracy Studies-2 risk for the assessment of the methodological qualities. (+) denotes low risk of bias, (?) denotes unclear risk of bias, (-) denotes high risk of bias.



Multimedia Appendixes

Table 1. Summary of studies in the literature review that applied AI techniques for CE image analysis.
URL: http://asset.jmir.pub/assets/9604cbd33bf7f3311beb8ed988b40a4b.docx

Searching strategy to find the relevant articles.
URL: http://asset.jmir.pub/assets/fd38d53a290d375a1e4dd49a6eac41b1.docx

Primsa checklist.
URL: http://asset.jmir.pub/assets/9e2e0518a5070fc5c9618f8ec48ea34b.docx