

Material Supplementary

Table S1 PRISMA-DTA Checklist.

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	3-4
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A (not registered)
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4-5

Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	4-5; Supplementary
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5-6
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	5-6
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	6
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	6
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	6

Page 1 of 2

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	7
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	8 (Table 1)

Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	13 (Supplementary Table/Appendix)
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	9 (Table 2); 7-9
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	7-10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	10-11 (Table 3)
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	12-13
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	13-14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	14
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	15

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

Table S2 Quality assessment criteria – tailored questions.

Quality Assessment Criteria – Tailored Questions

Patient Selection

1. Were the inclusion and exclusion criteria for patients with gastric neoplasms clearly outlined? (Yes/No/Unclear)
 2. Was the study design identified (e.g., retrospective, prospective, multicenter)? (Yes/No/Unclear)
 3. Were patients' demographic and clinicopathological details (e.g., age, lesion size, tumor type, invasion depth) clearly described? (Yes/No/Unclear)
 4. Is there a possibility that the patient selection process (e.g., sampling bias, exclusion of non-representative cases) introduced bias? (High risk/Low risk/Unclear risk)
 5. Is there concern that the patient population or clinical setting (e.g., tertiary centers only, specific endoscopic devices) may not align with the objective of the review? (High concern/Low concern/Unclear concern)
-

Index Test (AI-Based Endoscopic Prediction Model)

1. Were the endoscopic imaging protocols (e.g., white-light, NBI, magnifying endoscopy) and AI model inputs clearly explained? (Yes/No/Unclear)
 2. Was the method for preprocessing, feature extraction, or model training described in sufficient detail? (Yes/No/Unclear)
 3. Did the study employ an internal or external validation approach (e.g., cross-validation, independent dataset)? (Yes/No/Unclear)
 4. Was the AI model development and evaluation process (e.g., architecture, parameters, training/validation/test split) described clearly enough to allow reproducibility? (Yes/No/Unclear)
 5. Could the way the AI model (index test) was conducted or interpreted have introduced bias (e.g., lack of blinding to histopathologic results)? (High risk/Low risk/Unclear risk)
 6. Is there concern that the AI method or its clinical application may not match the review objective (i.e., prediction of submucosal invasion from endoscopic images)? (High concern/Low concern/Unclear concern)
-

Reference Standard

1. Is it likely that the reference standard (e.g., histopathologic diagnosis after endoscopic resection or surgery) accurately identified submucosal invasion? (Yes/No/Unclear)
 2. Could bias have been introduced by how the reference standard was applied or interpreted (e.g., lack of blinding to AI results)? (High risk/Low risk/Unclear risk)
 3. Are there concerns that the definition of submucosal invasion used in the reference standard does not match the review question? (High concern/Low concern/Unclear concern)
-

Flow and Timing

1. Did all included patients undergo both the AI-based endoscopic assessment and the histopathologic reference standard? (Yes/No/Unclear)
 2. Was the time interval between the AI prediction and the histopathologic confirmation appropriate to prevent disease progression bias? (Yes/No/Unclear)
 3. Could the sequence or patient flow (e.g., exclusions, missing data, incomplete follow-up) have introduced bias? (High risk/Low risk/Unclear risk)
-