Convolutional neural networks in the computer-aided diagnosis of *Helicobacter pylori* infection and non-causal comparison to physician endoscopists: a systematic review with meta-analysis

Babu P. Mohan[a], Shahab R. Khan[b], Lena L. Kassab[c], Suresh Ponnada[d], Nabeeha Mohy-Ud-Din[e], Saurabh Chandan[f], Parambir S. Dulai[g], Gursimran S. Kochhar[e]

University of Utah Health, Salt Lake City, Utah; Rush University Medical Center, Chicago, Illinois; Mayo Clinic, Rochester, Minnesota; Roanoke Medical Center, Roanoke, Virginia; Allegheny Health Network, Pittsburgh, Pennsylvania; University of Nebraska Medical Center, Omaha, Nebraska; University of California, San Diego, California, USA

Abstract

Background *Helicobacter pylori* (*H. pylori*) infection, if left untreated, can cause gastric cancer, among other serious morbidities. In recent times, a growing body of evidence has evaluated the use of a type of artificial intelligence (AI) known as "deep learning" in the computer-aided diagnosis of *H. pylori* using convolutional neural networks (CNN). We conducted this meta-analysis to evaluate the pooled rates of performance of CNN-based AI in the diagnosis of *H. pylori* infection.

Methods Multiple databases were searched (from inception to June 2020) and studies that reported on the performance of CNN in the diagnosis of *H. pylori* infection were selected. A random-effects model was used to calculate the pooled rates. In cases where multiple 2×2 contingency tables were provided for different thresholds, we assumed the data tables were independent from each other.

Results Five studies were included in our final analysis. Images used were from a combination of white-light, blue laser imaging, and linked color imaging. The pooled accuracy for detecting *H. pylori* infection with AI was 87.1% (95% confidence interval [CI] 81.8-91.1), sensitivity was 86.3% (95%CI 80.4-90.6), and specificity was 87.1% (95%CI 80.5-91.7). The corresponding performance metrics for physician endoscopists were 82.9% (95%CI 76.7-87.7), 79.6% (95%CI 68.1-87.7), and 83.8% (95%CI 72-91.3), respectively. Based on non-causal subgroup comparison methods, CNN seemed to perform equivalently to physicians.

Conclusion Based on our meta-analysis, CNN-based computer-aided diagnosis of *H. pylori* infection demonstrated an accuracy, sensitivity, and specificity of 87%.

Keywords Convolutional neural networks, *Helicobacter pylori*, meta-analysis

Ann Gastroenterol 2020; 33 (6): 1-10

[a]Gastroenterology and Hepatology, University of Utah Health, Salt Lake City, Utah (Babu P. Mohan); [b]Gastroenterology, Rush University Medical Center, Chicago, Illinois (Shahab R. Khan); [c]Internal Medicine, Mayo Clinic, Rochester, Minnesota (Lena L. Kassab); [d]Internal Medicine, Roanoke Medical Center, Roanoke, Virginia (Suresh Ponnada); [e]Gastroenterology and Hepatology, Allegheny Health Network, Pittsburgh, Pennsylvania (Nabeeha Mohy-Ud-Din, Gursimran S. Kochhar); [f]Gastroenterology and Hepatology, University of Nebraska Medical Center, Omaha, Nebraska (Saurabh Chandan); [g]Gastroenterology and Hepatology, University of California, San Diego, California (Parambir S Dulai), USA

Introduction

Helicobacter pylori (*H. pylori*) infection is a well-known risk factor for gastric cancer [1,2]. Left untreated, the disease can result in chronic gastritis, gastroduodenal ulceration, mucosal atrophy, intestinal metaplasia and mucosa-associated lymphoid tissue tumor. Treatment is therefore of paramount importance and is directed at complete eradication of *H. pylori* infection [3]. Central to this strategy is an accurate diagnostic methodology aimed at effectively ruling-in, ruling-out, and confirming the eradication of the infection.

Various endoscopic and non-endoscopic diagnostic tests are currently available. Non-endoscopic tests include a urea breath test, fecal *H. pylori* antigen test, urine anti-*H. pylori* immunoglobulin (IgG) assay, and serum *H. pylori* IgG assay. Areas targeted, using a standard endoscope, are usually chosen based on gastric mucosal redness and swelling. Current guidelines favor a repeat endoscopy to document eradication by negative biopsy results [3].

A growing body of evidence has evaluated the use of a type of artificial intelligence (AI) known as "deep learning" in the computer-aided diagnosis of health-related conditions based on medical imaging [4]. A convolutional neural network (CNN) enables machines to analyze various training images. CNN data-driven systems are trained on datasets containing large numbers of images with their corresponding labels. CNN can be seen as a system that first extracts relevant features from the input images and subsequently uses those learned features to classify a given image. The network uses convolutions of the input image in order to extract the most relevant information that helps to classify the image into different entities. Based on the accumulated data features, machines can diagnose newly acquired clinical images prospectively [5-7].

In this systematic review and meta-analysis, we aimed to quantitatively appraise the current reported data on the diagnostic performance of CNN-based computer-aided diagnosis of *H. pylori* infection and, if possible, compare the results to the diagnostic performance of physician endoscopists.

Materials and methods

Search strategy

A medical librarian searched the literature for the concepts of AI with endoscopy for gastrointestinal (GI) conditions. The search strategies were created using a combination of keywords and standardized index terms. Searches were run in November 2019 and an additional updated search was performed in June 2020 in ClinicalTrials.gov, Ovid EBM Reviews, Ovid Embase (1974+), Ovid Medline (1946+ including Epub ahead of print, in-process & other non-indexed citations), Scopus (1970+), and Web of Science (1975+). Results were limited to English language publications. All results were exported to Endnote X9 (Clarivate Analytics) where obvious duplicates were removed, leaving 4245 citations. The search strategy is provided in Appendix 1. The MOOSE checklist was followed and is provided as Appendix 2 [8]. Reference lists of evaluated studies were examined to identify other studies of interest.

Study selection

In this meta-analysis, we included studies that developed or validated a deep CNN learning model for the detection and/or diagnosis of *H. pylori* infection. Study selection was restricted to only those that used CNN-based machine learning models. Search terms: "Helicobacter pylori" and "convolutional neural network" were used to filter through the studies in the EndNote file. Studies were included irrespective of inpatient/outpatient setting, study sample size, optics of endoscopic imaging, follow-up time, abstract/manuscript status and geography, as long as they provided the appropriate data needed for the analysis.

Our exclusion criteria were as follows: 1) studies that used non-CNN based machine learning algorithms; and 2) studies not published in the English language. In cases of multiple publications from a single research group reporting on the same patient cohort and/or overlapping cohorts, all reported contingency tables were treated as being mutually exclusive. When necessary, authors were contacted via email for clarification of data and/or study cohort overlap.

Data abstraction and analysis

Data on study-related outcomes from the individual studies were abstracted independently onto a predefined standardized form by at least 2 authors (BPM, SRK). Disagreements were resolved by consultation with a senior author (GK). Diagnostic performance data were extracted and contingency tables were created at the reported thresholds. Contingency tables consisted of reported accuracy, sensitivity, specificity, positive predictive value and negative predictive value.

We used meta-analysis techniques to calculate the pooled estimates in each case, following a random-effects model [9]. We assessed heterogeneity between study-specific estimates using the Cochran Q statistical test for heterogeneity, 95% prediction interval, which deals with the dispersion of the effects, and the I^2 statistic [10,11], where a value <50% was considered as absence of heterogeneity. A formal publication bias assessment was not done because of the nature of the pooled results derived from the studies.

To compare the diagnostic performance of CNN to physician endoscopists, we did a subgroup analysis comparing the pooled performances of these 2 groups of datasets. All analyses were performed using Comprehensive Meta-Analysis (CMA) software, version 3 (BioStat, Englewood, NJ).

Results

Search results and study characteristics

The literature search resulted in 4245 study hits (study search and selection flowchart: Supplementary Fig. 1). All 4245 studies were screened and 106 full-length articles and/or abstracts were assessed. Five studies were included in the final analysis [12-16] (Table 1).

The following diagnostic tests were used to confirm the presence or absence of *H. pylori* infection: 1) *H. pylori* density by histology; 2) serum *H. pylori* IgG assay; 3) fecal *H. pylori* antigen test; and 4) urine *H. pylori* IgG assay. Further information is provided in Table 1.

From all the included studies, we were able to extract a total of 9 contingency table datasets for CNN and 4 for physician endoscopists' performance in detecting/diagnosing *H. pylori* infection. Values available for analysis were for accuracy, sensitivity and specificity. None of the studies reported data on positive or negative predictive value.

Meta-analysis outcomes

Diagnosis of H. pylori infection

The pooled accuracy of CNN in the computer-aided diagnosis of *H. pylori* infection was 87.1% (95% confidence interval [CI] 81.8-91.1), the pooled sensitivity was 86.3%

Table 1 Study characteristics

Study, Year	Aim	Endoscopy technique	AI Model	Training strategy	Testing strategy	Accuracy	Sensitivity	Specificity	Remarks
Zheng, 2019 [16]	Evaluate for *H. pylori* infection	Standard HD-white light EGD	ResNet-50	76,146 images from 1507 patients	3755 validation images from 452 patients	84.5	81.4	90.1	For single gastric image, at optimal cut-off. Gastric biopsy, Urea breath test
						93.8	91.6	98.6	For multiple images at optimal probability score cut-off
Shichijo, 2019 [12]						nr	76	96	
Itoh, 2018 [13]	Detect *H. pylori* infection	Standard EGD	CNN (GoogLeNet)	149 images from 65 positive and 74 negative patients	30 images from 15 positive and 15 negative patients	nr	86.7	86.7	Serum *H. pylori* IgG antibodies
Nakashima, 2018 [14]	Diagnosis of *H. pylori* infection	White light imaging	CNN (GoogLeNet)	162 subjects, 648 images (90 degree rotation X4)	60 subjects	nr	66.7	60	Serum *H. pylori* IgG antibodies
		Blue laser imaging				nr	96.7	86.7	
		Linked color imaging				nr	96.7	83.3	
Shichijo, 2017 [15]	Diagnosis of *H. pylori* infection	EGD	CNN (GoogLeNet)	32,208 images from 735 positive and 1015 negative patients	11,481 images from 397 patients	83.1	81.9	83.4	Fecal antigen and urine anti-*H. pylori* IgG
						87.7	88.9	87.4	CNN based on gastric location
						82.4	79	83.2	Endoscopists
						88.9	85.2	89.3	Certified endoscopists
						84.4	81	85.1	Relatively experienced
						75.6	72.2	76.3	Beginner

AI, artificial intelligence; CNN, convolutional neural network; EGD, esophagogastroduodenoscopy; IgG, immunoglobulin G; nr, not reported; H. pylori, Helicobacter pylori

(95%CI 80.4-90.6) and specificity was 87.1% (95%CI 80.5-91.7). Forest plots are shown in Fig. 1-3.

Physician endoscopist performance outcomes

The pooled accuracy of physician endoscopists in the diagnosis of all GI lesions was 82.9% (95%CI 76.7-87.7),

the pooled sensitivity was 79.6% (95%CI 68.1-87.7) and specificity was 83.8% (95%CI 72-91.3) (Fig. 1-3). Based on a non-causal subgroup method of comparison, CNN appeared to perform comparably to physicians in terms of accuracy, sensitivity and specificity (P=0.2, P=0.2, and P=0.6, respectively) (Table 2).

Validation of meta-analysis results

Sensitivity analysis

To assess whether any one study had a dominant effect on the meta-analysis, we excluded one study at a time and analyzed its effect on the main summary estimate. In this analysis, no single study significantly affected the outcome or the heterogeneity.

Heterogeneity

A large degree of between-study heterogeneity was expected, given the broad nature of machine learning algorithms and endoscopic optics included in this study. This is reflected in our I^2 values (Table 2). Prediction interval statistics were not calculated because of the expected large degree of heterogeneity and the fact that the goal was not to provide precise point estimates.

Publication bias

Publication bias assessment largely depends on the sample size and the effect size. A publication bias assessment was deferred in this study because the final number of studies included in the analysis was less than 10.

Quality of evidence

The quality of evidence was rated for results from the meta-analysis according to the GRADE working group approach [17]. Observational studies begin with a low-quality rating and, based
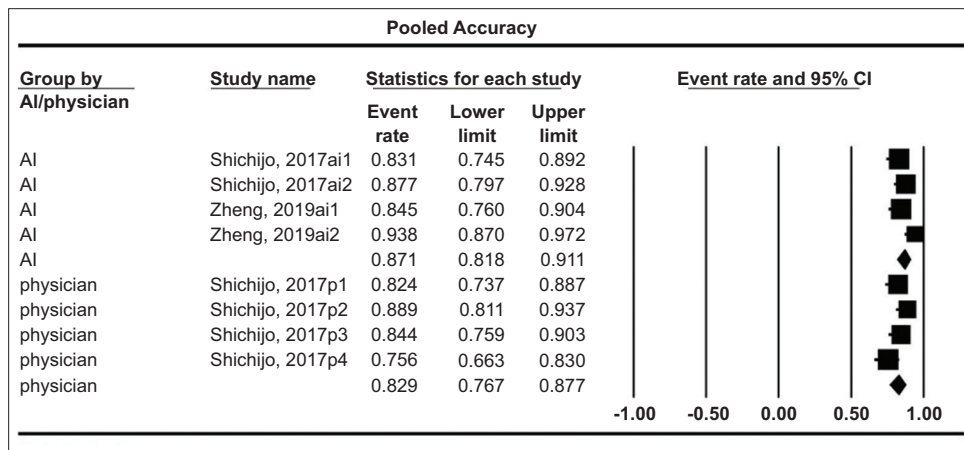

Pooled Accuracy

Group by AI/physician	Study name	Event rate	Lower limit	Upper limit	Event rate and 95% CI
AI	Shichijo, 2017ai1	0.831	0.745	0.892	
AI	Shichijo, 2017ai2	0.877	0.797	0.928	
AI	Zheng, 2019ai1	0.845	0.760	0.904	
AI	Zheng, 2019ai2	0.938	0.870	0.972	
AI		0.871	0.818	0.911	
physician	Shichijo, 2017p1	0.824	0.737	0.887	
physician	Shichijo, 2017p2	0.889	0.811	0.937	
physician	Shichijo, 2017p3	0.844	0.759	0.903	
physician	Shichijo, 2017p4	0.756	0.663	0.830	
physician		0.829	0.767	0.877	

Figure 1 Forest plot comparing the accuracy of artificial intelligence (AI) vs. physicians in the detection of *Helicobacter pylori* infection
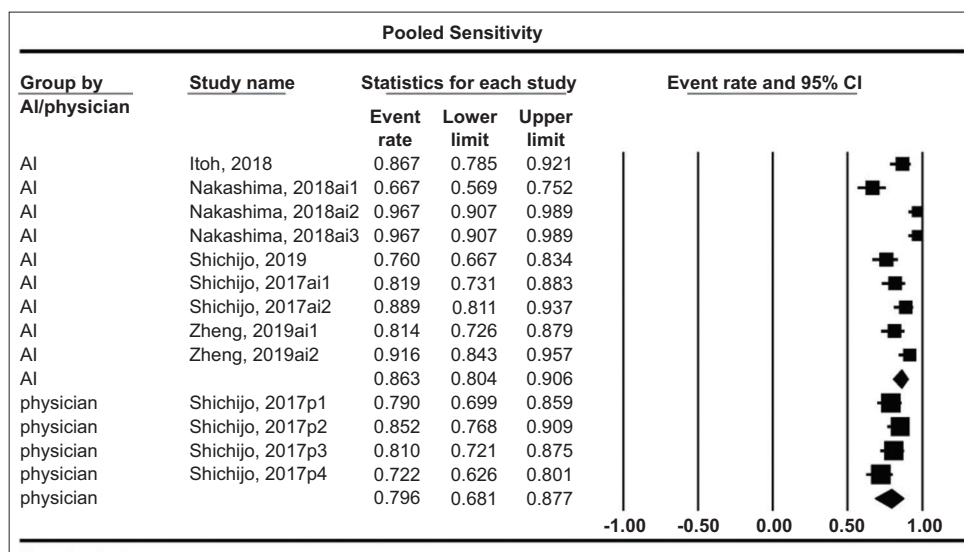*CI, confidence interval*

Pooled Sensitivity

Group by AI/physician	Study name	Event rate	Lower limit	Upper limit	Event rate and 95% CI
AI	Itoh, 2018	0.867	0.785	0.921	
AI	Nakashima, 2018ai1	0.667	0.569	0.752	
AI	Nakashima, 2018ai2	0.967	0.907	0.989	
AI	Nakashima, 2018ai3	0.967	0.907	0.989	
AI	Shichijo, 2019	0.760	0.667	0.834	
AI	Shichijo, 2017ai1	0.819	0.731	0.883	
AI	Shichijo, 2017ai2	0.889	0.811	0.937	
AI	Zheng, 2019ai1	0.814	0.726	0.879	
AI	Zheng, 2019ai2	0.916	0.843	0.957	
AI		0.863	0.804	0.906	
physician	Shichijo, 2017p1	0.790	0.699	0.859	
physician	Shichijo, 2017p2	0.852	0.768	0.909	
physician	Shichijo, 2017p3	0.810	0.721	0.875	
physician	Shichijo, 2017p4	0.722	0.626	0.801	
physician		0.796	0.681	0.877	

Figure 2 Forest plot comparing the sensitivity of artificial intelligence (AI) vs. physicians in the detection of *Helicobacter pylori* infection
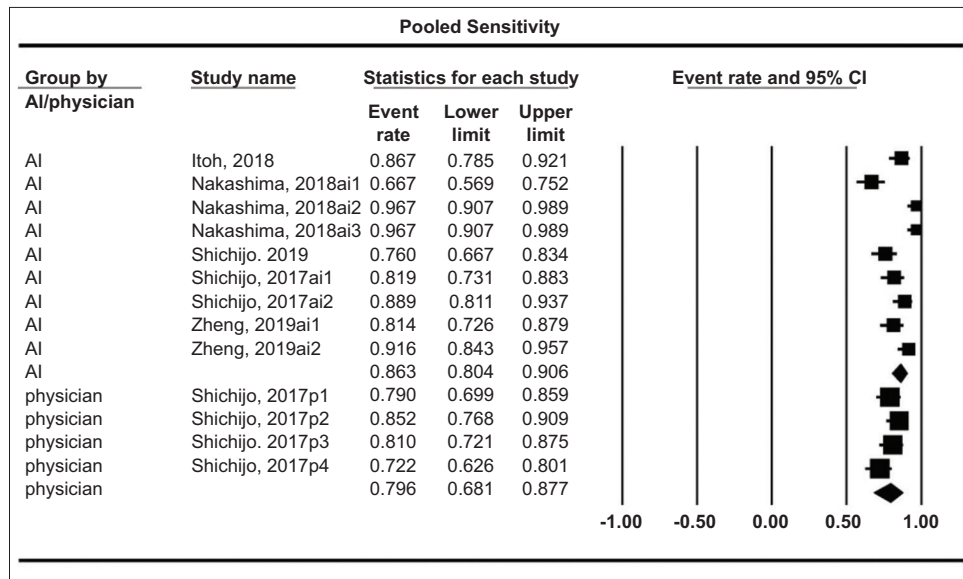*CI, confidence interval*

Pooled Sensitivity					
Group by **AI/physician**	**Study name**	**Statistics for each study**			**Event rate and 95% CI**
		Event rate	Lower limit	Upper limit	
AI	Itoh, 2018	0.867	0.785	0.921	
AI	Nakashima, 2018ai1	0.667	0.569	0.752	
AI	Nakashima, 2018ai2	0.967	0.907	0.989	
AI	Nakashima, 2018ai3	0.967	0.907	0.989	
AI	Shichijo. 2019	0.760	0.667	0.834	
AI	Shichijo, 2017ai1	0.819	0.731	0.883	
AI	Shichijo, 2017ai2	0.889	0.811	0.937	
AI	Zheng, 2019ai1	0.814	0.726	0.879	
AI	Zheng, 2019ai2	0.916	0.843	0.957	
AI		0.863	0.804	0.906	
physician	Shichijo, 2017p1	0.790	0.699	0.859	
physician	Shichijo, 2017p2	0.852	0.768	0.909	
physician	Shichijo. 2017p3	0.810	0.721	0.875	
physician	Shichijo, 2017p4	0.722	0.626	0.801	
physician		0.796	0.681	0.877	

-1.00 -0.50 0.00 0.50 1.00

Figure 3 Forest plot comparing the specificity of artificial intelligence (AI) vs. physicians in the detection of *Helicobacter pylori* infection
CI, confidence interval

Table 2 Summary of results

Pooled results (95% confidence intervals)

Variable	CNN	Physician endoscopists
Accuracy	87.1% (81.8-91.1) I^2=47 4 datasets P=0.2	82.9% (76.7-87.7) I^2=52 4 datasets
Sensitivity	86.3% (80.4-90.6) I^2=83 9 datasets P=0.2	79.6% (68.1-87.7) I^2=43 4 datasets
Specificity	87.1% (80.5-91.7) I^2=87 9 datasets P=0.6	83.8% (72-91.3) I^2=51 4 datasets

CNN, convolutional neural network

on the risk of bias, indirectness, heterogeneity and publication bias, this meta-analysis would be considered as low-quality evidence.

Discussion

To the best of our knowledge, this is the first systematic review and meta-analysis to assess CNN-based computer aided diagnosis of *H. pylori* infection. Based on our analysis, CNN-based deep machine learning demonstrated a pooled accuracy of 87.1%, a sensitivity of 86.3% and a specificity of 87.1% in the computer-aided diagnosis of *H. pylori* infection based on endoscopic images.

The respective pooled accuracy, sensitivity and specificity of CNN in the detection of *H. pylori* infection were each approximately 87%. The corresponding pooled performance values for physician endoscopists were 83%, 80% and 84%,

respectively. Based on a non-causal subgroup comparison, the pooled accuracy, sensitivity and specificity values seemed comparable between AI and physicians. Although the estimates seem to support the claim that deep learning algorithms can match physician-level diagnostic performance, several methodological limitations need to be kept in mind.

The included studies evaluated the performance of CNN under experimental conditions and not in a real-life clinical scenario. Prospective studies are lacking. Only high-quality images were used to train CNN. Procedural limitations such as less insufflation of air, post-biopsy bleeding, halation, blur, defocus or mucus can all affect the accuracy of a computer-aided diagnosis in a real clinical setting. Not all studies reported comparison outcomes with physician endoscopists. There was variability in the choice of thresholds used to report sensitivity and specificity. There was a lack in uniformity in the validation of the algorithm's training process before it was used for testing.

How does our study compare to currently published data? Although there are no other current meta-analyses evaluating the use of CNN in the diagnosis of *H. pylori* infection, a recently published review by Liu *et al* [4] evaluated the performance of deep machine learning and compared it to healthcare professionals in detecting diseases from medical imaging. They found the diagnostic performance of deep learning models to be equivalent to that of healthcare professionals and the pooled results were comparable to this study.

The strength of this review resides in the careful selection of studies reporting on machine-based learning, limiting them solely to CNN-based algorithms and avoiding other redundant studies. There are limitations to this study, most of which are inherent to any meta-analysis. The included studies were not entirely representative of the general population and community practice, with most studies being performed in an experimental environment. Our analysis had studies that were retrospective in nature, contributing to selection bias.

Our analysis has the limitation of non-causal comparison and heterogeneity. We were unable to statistically ascertain a cause for the observed heterogeneity. However, we hypothesize that the observed heterogeneity was primarily due to the following variables: threshold cutoff used, different training algorithms and training methodologies employed, and the variability in endoscopic optics (standard white light, blue laser imaging, linked color imaging). There exists a considerable degree of uncertainty and incomprehensibility due to the small dataset, as only 5 studies were available for analysis. Nevertheless, this study is the best available pooled evaluation of the diagnostic performance of CNN in computer-aided assessment of *H. pylori* infection thus far.

In conclusion, based on our meta-analysis, deep machine learning by means of convolutional neural network based algorithms demonstrates a pooled accuracy, sensitivity and specificity of 85% in the computer-aided diagnosis of *H. pylori* infection based on endoscopic imaging. CNN seems to demonstrate better accuracy, and equivalent sensitivity and specificity, compared to physician endoscopists. Deep learning in gastroenterology is in its infancy and is witnessing a rapid, steep growth in terms of learning as well as technological development. Future studies are needed to streamline the machine-learning process and define its role in the computer-aided diagnosis of *H. pylori* infections in real-life clinical scenarios.

Summary Box

What is already known:

- There is no other meta-analysis evaluating the use of artificial intelligence (AI) based on convolutional neural networks (CNN) in the endoscopic image-based diagnosis of *H. pylori*
- A handful of studies have reported equivalent diagnostic performance between AI and physician endoscopists

What the new findings are:

- Diagnostic accuracy, sensitivity and specificity are >85% with AI based on CNN in the diagnosis of *H. pylori*
- Diagnostic accuracy of CNN seemed comparable to physician endoscopists based on non-causal subgroup comparison

References

1. Sugano K. Effect of *Helicobacter pylori* eradication on the incidence of gastric cancer: a systematic review and meta-analysis. *Gastric Cancer* 2019;**22**:435-445.
2. Uemura N, Okamoto S, Yamamoto S, et al. *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* 2001;**345**:784-789.
3. Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG Clinical Guideline: treatment of *Helicobacter pylori* infection. *Am J Gastroenterol* 2017;**112**:212-239.
4. Liu X, Faes L, Kale AU, 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 Digit Health* 2019;**1**:e271-e297.
5. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;**68**:94-100.
6. Krizhevsky A, Sutskever I, Hinton GE. Advances in neural information processing systems. Neural Information Processing Systems Foundation 2012;1269.
7. Szegedy C, Vanhoucke V, Ioffe S, et al. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
8. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA* 2000;**283**:2008-2012.
9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177-188.
10. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-560.
11. Mohan BP, Adler DG. Heterogeneity in systematic review and meta-analysis: how to read between the numbers. *Gastrointest Endosc* 2019;**89**:902-903.
12. Shichijo S, Endo Y, Aoyama K, et al. Application of convolutional neural networks for evaluating *Helicobacter pylori* infection status on the basis of endoscopic images. *Scand J Gastroenterol* 2019;**54**:158-163.
13. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc Int Open* 2018;**6**:E139-E144.
14. Nakashima H, Kawahira H, Kawachi H, et al. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018;**31**:462-468.
15. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images. *EBioMedicine* 2017;**25**:106-111.
16. Zheng W, Zhang X, Kim JJ, et al. High accuracy of convolutional neural network for evaluation of *Helicobacter pylori* infection based on endoscopic images: preliminary experience. *Clin Transl Gastroenterol* 2019;**10**:e00109.
17. Puhan MA, Schünemann HJ, Murad MH, et al; GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;**349**:g5630.

Supplementary material

Number of results before and after de-duplication

Database	Number of initial hits	After de-duplication
EBM Reviews	112	38
Embase	2260	1508
Medline	940	874
Scopus	2805	1512
Web of Science	1430	313
Totals	7547	4245

EBM Reviews

(((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)).ab,hw,ti.) AND ((endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab,hw,ti.) AND (("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)).ab,hw,ti.)

Embase (1974+)

(digestive system cancer/ or exp esophagus cancer/ or exp intestine cancer/ or exp stomach cancer/ or digestive system tumor/ or exp esophagus tumor/ or exp gastrointestinal tumor/ or exp intestine tumor/ or exp stomach tumor/ or ((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)).ab,kw,ti.)

AND (digestive tract endoscopy/ or exp chromoendoscopy/ or exp endoscopic retrograde cholangiopancreatography/ or exp esophagogastroduodenoscopy/ or exp esophagoscopy/ or exp gastrointestinal endoscopy/ or digestive endoscope/ or exp anoscope/ or exp balloon enteroscope/ or exp capsule endoscope/ or exp colonoscope/ or exp digestive endomicroscope/ or exp duodenoscope/ or exp esophagoscope/ or exp gastroduodenoscope/ or exp gastroscope/ or exp proctoscope/ or exp sigmoidoscope/ or

(endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab,kw,ti.) AND (exp artificial intelligence/ or exp machine learning/ or ("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)).ab,kw,ti.) NOT (exp animal/ not exp human/, exp child/ not exp adult/, "case report".kw,pt,ti.) Limit to English

Ovid MEDLINE(R) 1946 to Present and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) Daily

(exp Gastrointestinal Neoplasms/ or ((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach

or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)). ab,kf,ti.) AND (exp Endoscopy, Digestive System/ or exp Endoscopes, Gastrointestinal/ or (endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab,kf,ti.) AND (exp Artificial Intelligence/ or ("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)). ab,kf,ti.) NOT (exp Animals/ not Humans/, exp CHILD/ not exp ADULT/, "case report".kf,pt,ti.) Limit to English

Scopus

(TITLE-ABS-KEY ((digestive OR gastr* OR gi OR alimentary OR esophag* OR oesophag* OR stomach OR intestin* OR bowel* OR colon* OR colorectal OR rectal OR rectum OR sigmoid OR duoden* OR ileum OR ileal OR jejun* OR anal OR anus) W/3 (polyp* OR mass* OR lesion* OR tumor* OR tumour* OR carcin* OR adeno* OR neoplas* OR cancer* OR malignan* OR sarcoma* OR lymphoma* OR leiomyosarcoma*))) AND (TITLE-ABS-KEY (endoscop* OR enteroscop* OR gastroscop* OR colonoscop* OR duodenoscop* OR rectoscop* OR sigmoidoscop* OR ileocolonoscop* OR chromoendoscop* OR esophagogastroduodenoscop* OR esophagoscop* OR oesophagogastroduodenoscop* OR proctoscop* OR ercp OR anoscop* OR endomicroscop* OR oesophagoscop* OR gastroduodenoscop* OR sigmoidoscop* OR diagnos* OR patholog*)) AND (TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "machine intelligen*" OR computer-aided OR "computational intelligen*" OR "deep learning" OR "deep unified network*" OR "data mining" OR datamining OR "supervised learning" OR "semi-supervised learning" OR "unsupervised learning" OR "automated pattern recognition" OR "Bayesian learning" OR "computer heuristics" OR "hidden Markov model*" OR "k-nearest neighbor*" OR "kernel method*" OR "learning algorithm*" OR "natural language processing" OR "support vector" OR "vector machine" OR gaussian OR bootstrap OR "regression tree*" OR "linear discriminant analysis" OR "naive Bayes" OR "learning vector" OR "random forest*" OR "chi-square automatic interaction detection" OR "iterative dichotom*" OR fuzzy OR "neural network*" OR perceptron* OR (computer AND W/1 AND heuristic*))) AND (LIMIT-TO (LANGUAGE , "English"))

Web of Science

TS=((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) NEAR/3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)) AND TS=(endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*) AND TS=("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer NEAR/1 heuristic*)) Limit to English

Appendix 2 MOOSE checklist

Item No	Recommendation	Reported on Page No
Reporting of background should include		
1	Problem definition	5
2	Hypothesis statement	-
3	Description of study outcome(s)	5
4	Type of exposure or intervention used	5
5	Type of study designs used	5
6	Study population	5
Reporting of search strategy should include		
7	Qualifications of searchers (e.g., librarians and investigators)	7, Appendix 1
8	Search strategy, including time period included in the synthesis and key words	7, Appendix 1
9	Effort to include all available studies, including contact with authors	7
10	Databases and registries searched	7, Appendix 1
11	Search software used, name and version, including special features used (e.g., explosion)	Appendix 1
12	Use of hand searching (e.g., reference lists of obtained articles)	-na-
13	List of citations located and those excluded, including justification	Appendix 1
14	Method of addressing articles published in languages other than English	-na-
15	Method of handling abstracts and unpublished studies	7
16	Description of any contact with authors	7
Reporting of methods should include		
17	Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	7-8
18	Rationale for the selection and coding of data (e.g., sound clinical principles or convenience)	7-8
19	Documentation of how data were classified and coded (e.g., multiple raters, blinding and inter-rater reliability)	-na-
20	Assessment of confounding (e.g., comparability of cases and controls in studies where appropriate)	-na-
21	Assessment of study quality, including blinding of quality assessors, stratification or regression on possible predictors of study results	7-8
22	Assessment of heterogeneity	8
23	Description of statistical methods (e.g., complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated	8
24	Provision of appropriate tables and graphics	Table 1 supplementary materials
Reporting of results should include		
25	Graphic summarizing individual study estimates and overall estimate	Supplementary materials
26	Table giving descriptive information for each study included	Table 1
27	Results of sensitivity testing (e.g., subgroup analysis)	11
28	Indication of statistical uncertainty of findings	12

Item No	Recommendation	Reported on Page No
Reporting of discussion should include		
29	Quantitative assessment of bias (e.g., publication bias)	12
30	Justification for exclusion (e.g., exclusion of non-English language citations)	-na-
31	Assessment of quality of included studies	-na-
Reporting of conclusions should include		
32	Consideration of alternative explanations for observed results	13-15
33	Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review)	13-15
34	Guidelines for future research	15
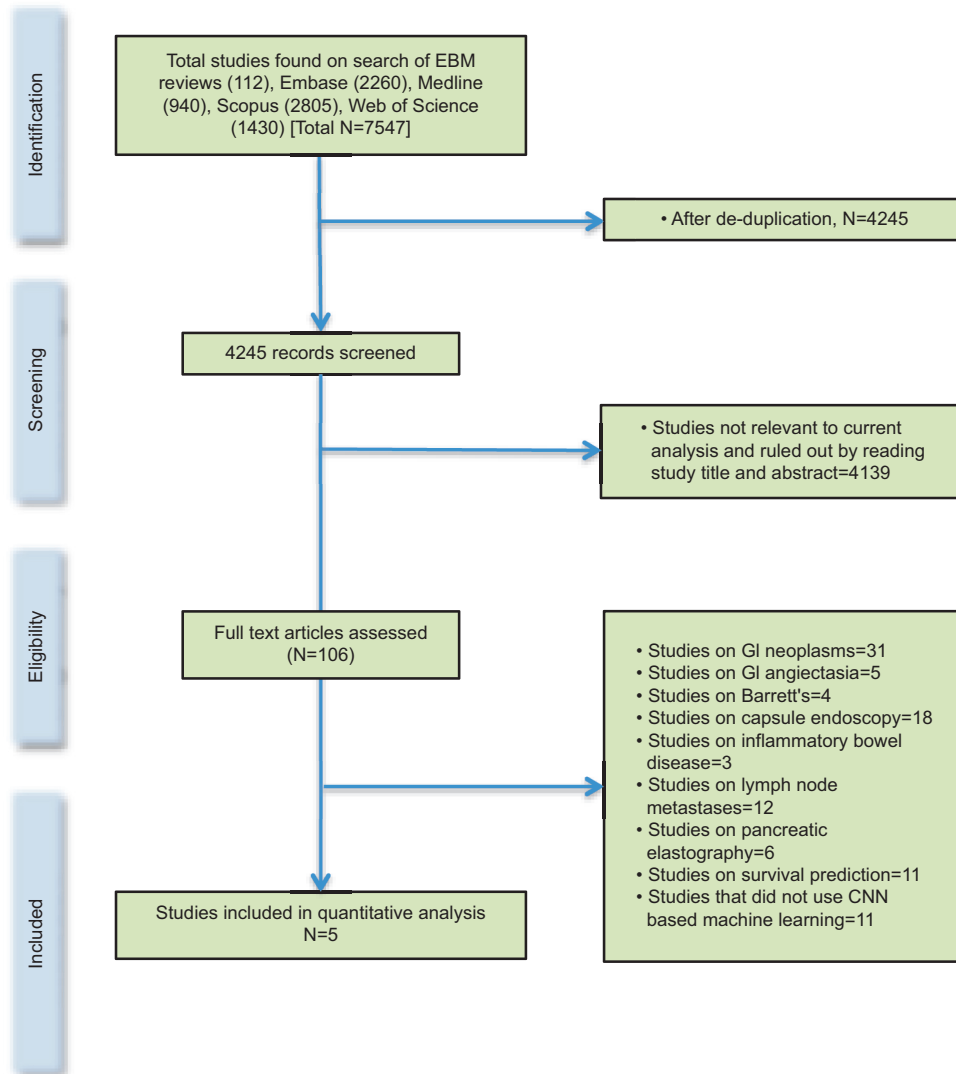
Supplementary Figure 1 Literature search flow chart

EBM, evidence-based medicine; GI, gastrointestinal; H. pylori, Helicobacter pylori; CNN, convolutional neural network