EClinicalMedicine 29–30 (2020) 100622

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Real-time computer aided colonoscopy versus standard colonoscopy for improving adenoma detection rate: A meta-analysis of randomized-controlled trials

Babu P. Mohan[a,g,1], Antonio Facciorusso[b,g,1], Shahab R. Khan[c,g], Saurabh Chandan[d,g], Lena L. Kassab[e,g], Paraskevas Gkolfakis[f,g], Georgios Tziatzios[f,g], Konstantinos Triantafyllou[f,g], Douglas G. Adler[a,g,*]

[a] Gastroenterology and Hepatology, University of Utah Health, Salt Lake City, UT, USA
[b] Gastroenterology Unit, University of Foggia, Foggia, Italy
[c] Gastroenterology, Rush University Medical Center, Chicago, IL, USA
[d] Gastroenterology and Hepatology, CHI Creighton University Medical Center, Omaha, NE, USA
[e] Internal Medicine, Mayo Clinic, Rochester, MIN, USA
[f] Hepatogastroenterology Unit, 2nd Dep of Internal Medicine – Propaedeutic Medical School, National and Kapodistrian University of Athens, Athens, Greece
[g] Gastroenterology and Hepatology, University of Colorado Anshchutz Medical Campus, Aurora, CO, USA

ARTICLE INFO

Article History:
Received 14 July 2020
Revised 4 September 2020
Accepted 28 September 2020
Available online xxx

Keywords:
Convolutional neural networks
Colonoscopy
Adenoma detection rate

ABSTRACT

Background: Recent prospective randomized controlled trials have evaluated deep convolutional neural network (CNN) based computer aided detection (CADe) of lesions in real-time colonoscopy. We conducted this meta-analysis to compare the adenoma detection rate (ADR) of deep CNN based CADe assisted colonoscopy to standard colonoscopy (SC) from randomized controlled trials (RCTs).
Methods: Multiple databases were searched (from inception to May 2020) and parallel RCTs that compared deep CNN based CADe assisted colonoscopy to SC were included for this analysis. Using Mantel-Haenzel (M-H) random effects model, pooled risk ratios (RR) and mean difference (MD) were calculated. In between study heterogeneity was assessed by I^2% values. Outcomes assessed included other per patient adenoma parameters.
Findings: Six RCTs were included in our final analysis that utilized deep CNN based CADe system in real-time colonoscopy. Total numbers of patients assessed were 4962 (2480 in CADe and 2482 in SC group). CADe based colonoscopy demonstrated statistically higher pooled ADR, RR=1.5 (95% CI 1.3–1.72), $p<0.0001$, I^2=56%; and pooled PDR, RR=1.42 (95% CI 1.33–1.51), $p<0.00001$, I^2=9%; when compared to SC. Per patient adenoma detection parameters were significantly better with CADe colonoscopy when compared to SC, with increased scope withdrawal time (mean difference = 0.38, 95% CI 0.05–0.72, $p = 0.02$).
Interpretation: Based on our meta-analysis, deep CNN based CADe colonoscopy achieved significantly higher ADR metrics, albeit with increased scope withdrawal time when compared to SC.

Acronyms

CNN	convolutional neural networks
AI	artificial intelligence
ADR	adenoma detection rate
CADe	computer-aided detection
SC	standard colonoscopy
RCT	randomized controlled trial
CRC	colorectal cancer
CI	confidence interval
MAP	mean adenoma per colososcopy
RR	relative risk
MD	mean difference

1. Introduction

Adenoma detection rate (ADR) is a well-accepted quality indicator of screening colonoscopy and is defined as the proportion of patients who have one or more adenoma detected while undergoing screening colonoscopy. Higher ADR by standard colonoscopy (SC) has

* Corresponding author at: Gastroenterology and Hepatology, University of Utah Health, Salt Lake City, UT, USA.
E-mail address: Douglas.adler@hsc.utah.edu (D.G. Adler).
[1] These authors contributed equally.

https://doi.org/10.1016/j.eclinm.2020.100622

Research in context

Evidence before this study

Recently, good quality randomized controlled trials (RCTs) have assessed the impact of artificial intelligence (AI) as a computer aid in helping detect colon polyps during colonoscopy. All RCTs published to date on the use of AI in colonoscopy and reporting on the adenoma detection rate (ADR) were considered for this meta-analysis study. We searched the literature using a combination of artificial intelligence, machine learning, machine intelligence and colonoscopy. Only prospectively done RCTs were included in this analysis. Searches were run in April 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+). Two out of the six studies included for analysis were considered low on the quality scale for risk of bias due to absence of blinding. AI based colonoscopy demonstrated statistically higher pooled ADR, RR=1.5 (95% CI 1.3-1.72), $P<0.0001$; when compared to standard colonoscopy.

Added value of this study

By methods of meta-analysis of good quality RCTs, this study adds robust evidence of ADR with use of AI assistance in colonoscopy.

Implications of all the available evidence

Based on the results of our study, the practice of colonoscopy in the future would most probably be done with the assistance of AI in the detection of pre-cancerous polyps and improving ADR. As ADR is technically used to define the performance of screening colonoscopies, future research is warranted in defining the role of AI strictly in screening colonoscopy.

shown to decrease CRC incidence thereby improving CRC related morbidity and mortality [1].

A growing body of evidence has evaluated the use of artificial intelligence (AI) known as computer-vision in computer-aided diagnosis (CAD) of health related conditions based on medical imaging [2–7]. Convolutional neural networks (CNN) is a type of deep machine learning algorithm that 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, a deep CNN can diagnose newly acquired clinical images prospectively [8–9]. Recent evidence has evaluated the use of CNN based algorithms in real-time colonoscopy to improve ADR by means of randomized controlled trials (RCTs) [2–7].

In this analysis, we aim to quantitatively appraise the current reported data on ADR during colonoscopy in presence of CNN based computer aided detection (CADe) from prospectively conducted parallel RCTs in real life scenario.

2. Methods

2.1. Search strategy

The literature was searched by a medical librarian for the concepts of artificial intelligence regards to endoscopy and gastrointestinal lesions. The search strategies were created using a combination of keywords and standardized index terms. Searches were run in April 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. All results were exported to Endnote X9 (Clarivate Analytics) where obvious duplicates were removed leaving 4245 citations. Search strategy is provided in Appendix-1. The PRISMA statement of adherence was followed and is provided as Appendix 2 [10]. Reference lists of evaluated studies were examined to identify other studies of interest.

2.2. Study selection

In this meta-analysis, we included parallel RCTs that evaluated ADR derived from colonoscopy procedures with real-time CNN based computer aided diagnosis (CAD) and compared it to standard colonoscopy (SC). Study selection was restricted to RCTs and CNN based machine learning models that were used during colonoscopy in the intervention group. Studies were included irrespective of inpatient/ outpatient setting; study sample-size, follow-up time, abstract/ manuscript status, and geography as long as they provided all appropriate data needed for the analysis.

Exclusion criteria were as follows: (1) studies that used non-CNN based algorithms, (2) studies that were non-clinical and reported on the mathematical development and/ or derivation of an algorithm, (3) studies not conducted as RCTs, and (4) studies that reported on training, testing and validating machine algorithms using images and/ or videos retrieved after colonoscopy. In cases of multiple publications from a single research group reporting on the same patient cohort and/or overlapping cohorts, the most comprehensive study was included. When needed, authors were contacted via email for clarification of data and/ or study-cohort overlap.

2.3. Data abstraction and analysis

Data on study-related outcomes from the individual studies were abstracted independently onto a predefined standardized form by at least three authors (AF, SRK, SC). Disagreements were resolved by consultation with another author (BPM). Risk of bias assessment was performed by evaluating the following: (1) selection bias, by means of random sequence generation and allocation concealment; (2) performance bias, by means of blinding of participants and personnel; (3) detection bias, by means of blinding of outcome assessment; (4) attrition bias, by means of incomplete outcome data; (5) reporting bias, by means of selective reporting and (6) other biases. The results of the bias assessment were reported as an overall graphical representation of the results of assessment, as well as the scoring of risks.

Primary outcome assessed was ADR. Secondary outcomes assessed were polyp detection rate (PDR), advanced adenoma detection rate (aADR), sessile serrated adenoma detection rate (SSADR), mean adenoma per colonoscopy (MAP) and other per patient parameters as available such as mean polyp per patient, mean diminutive adenoma per patient, mean flat-sessile adenoma per patient, mean large adenoma per patient, and mean right sided adenoma per patient.

We used meta-analysis techniques to calculate the pooled estimates in each case following the Mantel-Haenzel (M-H) random-effects model [11]. Summary estimates calculated were either the pooled Risk Ratio (RR) or the mean difference (MD) with corresponding 95% confidence intervals (CI), as appropriate. We assessed heterogeneity between study-specific estimates by using the I^2 statistics [12–13]. In this, values of <30%, 30% - 60%, 61% - 75%, and >75% were suggestive of low, moderate, substantial, and considerable heterogeneity, respectively. The quality of the studies was assessed using the Cochrane tool for assessing risk of bias [14]. Publication bias assessment was done qualitatively by funnel-plot assessment and quantitatively by Egger's test. Publication bias assessment was deferred if the total number of studies included in

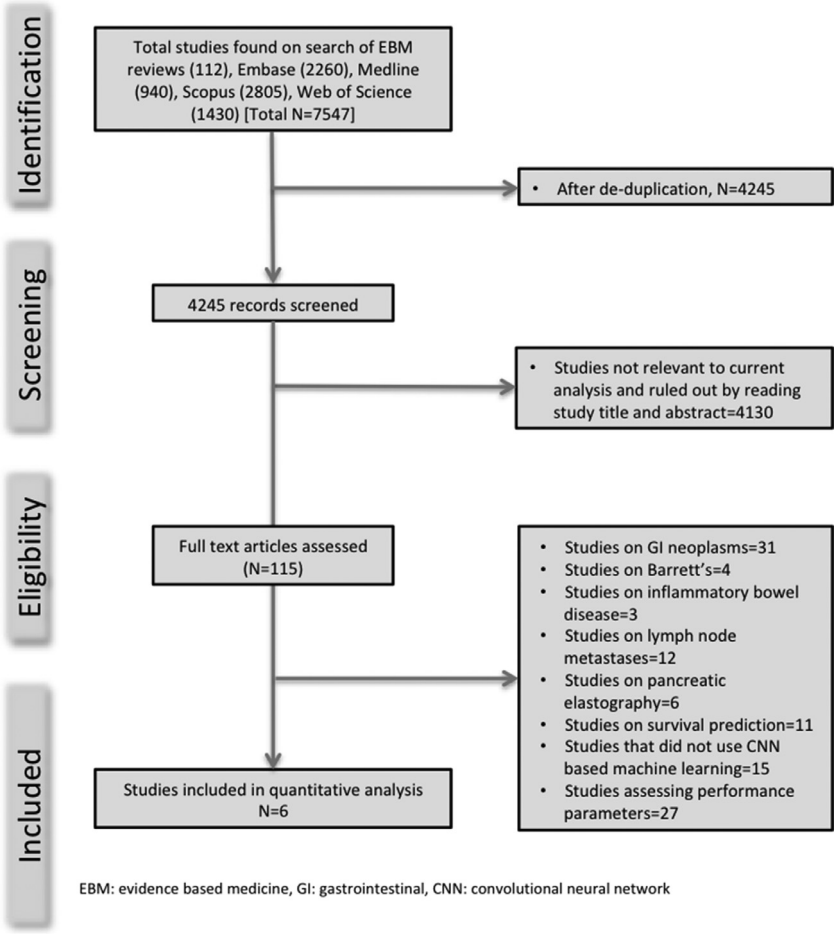

Fig. 1. Literature search flowchart.

the analysis were less than ten. All analyses were performed using RevMan version 5.3 and Comprehensive Meta-Analysis (CMA) software, version 3 (BioStat, Englewood, NJ).

2.4. Role of funding source

No funding was received for this study and there was no role of any funding source.

3. Results

3.1. Search results and study characteristics

From an initial search of 7547 studies, 4245 studies were screened after removing the duplicates. 115 full text articles were reviewed, and six prospective studies were included in the final analysis (Study selection flowchart: Fig. 1) [2–7]. All studies used deep CNN based machine learning algorithm with capability of detecting lesions in real time. Five studies [3–7] were performed in China and one study [2] was performed in Italy. Study and population characteristics are summarized in Table 1. The total number of patients included in the analysis was 4962, with 2480 in CADe arm and 2482 in SC group. Baseline age range (50–52 vs 51), male gender (50% vs 51%) and screening/ surveillance indication (13% vs 14%) were comparable between the CADe and SC arms.

The risk of bias assessment of the studies are summarized in Supplementary Figs. 15 and 16. Based on the assessment two studies scored low on the scale due to lack of blinding.

3.2. Meta-analysis outcomes

The pooled ADR with use of CADe endoscopy was significantly greater when compared to standard colonoscopy (SC); RR=1.5 (95% CI 1.3–1.72), $p<0.0001$, I^2=56% (Forest plot: Fig. 2). The pooled proportion of ADR with CADe was 32.8% (95% CI 24.2–42.7) and the pooled proportion of ADR with SC was 21.1% (95% CI 14.5–29.7). A subgroup analysis based on ADR from studies published from China (east) and ADR of high-quality studies did not affect the pooled rates and/ or the level of statistical significance (Forest plots: Supplementary Figs. 2, 3). Additionally, the pooled RR of PDR was significantly greater with CADe when compared to SC (1.42, 95% CI 1.33–1.51, $p<0.0001$, I^2=9%; Forest plot: Supplementary Figure-1) and the mean difference of scope withdrawal time was statistically increased with CADe (0.38 min, 95% CI 0.05–0.72, p = 0.02, I^2=97%).

The pooled RR of advanced ADR (1, 95% CI 0.74–1.36, p = 0.93, I^2=0%; Forest plot: Supplementary Figure-4) and sessile serrated ADR (1.29, 95% CI 0.89–1.89, p = 0.18, I^2=0%; Forest plot: Supplementary Figure-5) were comparable between CADe and SC; however the mean adenoma detected per colonoscopy was significantly better with CADe colonoscopy (mean difference = 0.19, 95% CI 1.16 – 0.21, $p<0.001$, I^2=90%; Forest plot: Supplementary Figure 6). The individual pooled proportions of the analyzed outcomes are summarized in Table 2. The pooled proportion of false positives on CADe colonoscopy was 10.3% (95% CI 6.1–16.8), I^2=93%, with comparable cecal intubation time (mean difference = 0.04, 95% CI −0.29 – 0.38, p = 0.8, I^2=60%) between CADe and SC.

In terms of per patient analysis, the mean difference of the polyp per patient, the mean rate of detection of diminutive adenoma, flat-

Table 1
Study and population characteristics.

Details	Gong, 2020		Repici, 2020		Liu, 2020		Su, 2019		Wang, 2019		Wang, 2020	
	AI	SC	AI	SC	AI	SC	AI	SC	AI	SC	AI	SC
Study details	RCT, June 2019 to Sept 2019, Single center, China.	RCT, Sep to Nov 2019, Multicenter, Italy	RCT, Oct 2018 to Mar 2019, Multicenter, China.	RCT, Oct 2018 to May 2019, Single center, China	RCT, Sep 2017 to Feb 2018, Single center, China	Double-blind RCT, Sept 2018 to Jan 2019, Single center, China.						
Study aim	Detection of colorectal adenomas, Time insertion and withdrawal, Avoid blind spots caused by endoscope slipping, Monitor real-time withdrawal speed during colonoscopy	Efficacy of CADe system for the detection of colorectal neoplasia	Colonoscopic polyp and Adenoma detection rates (ADR)	Polyp detection, withdrawal time, withdrawal stability, bowel preparation	Colonoscopic polyp and Adenoma detection rates (ADR)	Double-blind study with sham control to rigorously assess the effectiveness of CADe system in improving ADR						
Deep CNN details	ENDOANGEL system - deep CNN trained and tested using VGG-16, DenseNet-169, ResNet-50 & Inception-v3. VGG-16 was finally used to develop the system. TensorFlow deep learning framework was used.	GI-Genius, Medtronic - deep CNN architecture details not available	Convolutional three−dimensional (3D) neural network. The convolutional 3D network is designed for spatiotemporal data.	5 deep CNN models to automatically time the withdrawal phase, supervise withdrawal stability, evaluate bowel preparation, and detect polyps in real time. Models developed baed on Alex-Net, ZFNet, YOLO V2	deep CNN was based on SegNet architecture	EndoScreener - based on SegNet architecture						
Total patients	704	685	1026	623	1058	962						
	355	349	341	344	508	518	308	315	522	536	484	478
Age (SD)	50·0 (37·0−58·0)	49·0 (36·0−57·0)	61.5 (9.7)	61.1 (10.6)	51.02 (12.26)	50.13 (12.68)	50.54 (10.28)	51.63 (9.04)	51.07 (13.15)	49.94 (13.79)	49 (39−60)	49 (40.3−56)
Female (%)	168 (47)	191 (55)	169 (49.6)	179 (52)	244 (48.03)	231 (44.59)	149 (48.38)	167 (53.02)	259 (49.62)	287 (53.54)	243 (50)	224 (47)
Male (%)	187 (53)	158 (45)	172 (50.4)	165 (49.6)	264 (51.97)	287 (55.41)	159 (51.62)	148 (46.98)	263 (50.38)	249 (46.46)	241 (50)	254 (53)
Colonoscopy indication												
FIT+ (%)	–	–	102 (29.9)	105 (30.5)	–	–	–	–	–	–	–	–
Primary CRC screening (%)	60 (17)	63 (18)	77 (22.6)	76 (22.1)	30 (5.91)	36 (6.95)	115 (37.34)	101 (32.06)	40 (7.66)	44 (8.21)	82 (17)	76 (16)
Surveillance (%)	14 (4)	22 (6)	86 (25.2)	78 (22.7)	–	–	69 (22.4)	78 (24.76)	–	–	–	–
GI symptoms (%)	281 (79)	264 (76)	76 (22.3)	85 (24.7)	478 (94.09)	482 (93.05)	193 (62.66)	214 (67.94)	482 (92.34)	492 (91.79)	402 (83)	402 (84)
Adequate BBPS (>2 in all segments) (%)	334 (94.08)	327 (93.69)	339 (99.4)	342 (99.4)	442 (87.01)	447 (86.29)	–	–	449 (86.02)	457 (85.26)	413 (85%)	413 (86%)
Cecal intubation (insertion) time; min (SD)	–	–	9 (5−11)	8.1 (2−10)	5.68 (4.09)	5.96 (4.06)	6.38 (2.25)	6.27 (2.17)	5.63 (4.03)	5.71 (3.9)	5.58 (3.96)	5.58 (3.7)
Withdrawal time; min (SD)	6.38 (2.48)	4.76 (2.54)	6.95 (1.68)	7.25 (2.48)	6.82 (1.78)	6.74 (1.62)	7.03 (1.01)	5.68 (1.26)	6.89 (1.79)	6.39 (1.21)	7.46 (2.02)	6.99 (1.57)
Total procedure time; min (SD)	–	–	–	–	12.41 (4.25)	12.7 (4.16)	–	–	12.52 (4.38)	12.2 (4.08)	–	–
ADR, n/N (%)	58/355 (16)	27/349 (8)	187/341 (54.8)	139/344 (40.4)	198/508 (39.2)	119/518 (23.9)	89/308 (28.9)	52/315 (16.5)	152/522 (29.12)	108/536 (20.3)	164/484 (34)	134/478 (28)
PDR, n/N (%)	166/355 (47)	118/349 (34)	279/341 (82)	214/344 (62)	221/508 (43.7)	144/518 (27.8)	118/308 (38.3)	80/315 (25.4)	235/522 (45)	156/536 (29.1)	252/484 (52)	177/478 (37)
False positive on CADe	–	–	–	–	36	–	62	–	39	–	48	–
Total adenomas (n)	61	27	177	136	250	142	113	56	262	160	281	181
Location of adenomas (n): right/ left	26/ 35	12/ 15	123/ 109	97/ 72	131/119	81/ 61	48/ 65	18/ 38	122/140	76/ 84	132/ 149	85/ 96
Size of the Adenomas (n): <10 mm (APS); >10 mm (APS)	50 (14%); 10 (3%)	26 (7.45%); 1 (<1%)	151 (44.3%); 36 (10.6%)	111 (32.3%); 28 (9.1%)	229 (0.451); 21 (0.041)	132 (0.255); 10 (0.019)	–	–	246 (0.437); 16 (0.031)	152 (0.268); 8 (0.014)	271 (96); 10 (4)	174 (96); 7 (4)
Advanced adenomas (n)	–	–	35 (10.3)	33 (7.3)	14	16	–	–	17	16	11 (2)	13 (4)

(*continued on next page*)

Table 1 (*Continued*)

Details	Gong, 2020		Repici, 2020		Liu, 2020		Su, 2019		Wang, 2019		Wang, 2020	
	AI	SC	AI	SC	AI	SC	AI	SC	AI	SC	AI	SC
Sessile Serrated adenomas (n)	–	–	24 (7)	18 (5.2)	18	13	–	–	17	14	18 (4)	14 (5)
Morphology of adenomas (n) Flat/ Sessile Pedunculated	–	–	–	–	210/ 40	106/ 36	98/ 15	43/ 13	223/ 39	127/ 33	257/ 19	160/ 20
Total polyps (n)	178	124	126	91	486	248	177	96	498	269	501	308
Location of polyp (n) right/ left	84/174	60/108	123/109	97/72	218/268	96/152	75/102	34/62	212/286	95/174	207/294	127/181
Size of the polyp (n): <10 mm (mean polyp per patient); ≥10 mm (mean polyp per patient)	167 (48%); 11 (3%)	121 (35%); 3 (1%)	–	–	464 (0.913); 22 (0.043)	232 (0.448); 16 (0.031)	–	–	482 (0.856); 16 (0.031)	259 (0.457); 10 (0.018)	490 (98); 11 (2)	297 (96); 11 (4)
Cancer	–	–	10 (2.9)	3 (0.9)	0	0	–	–	0	0	0	0

CADe: computer assisted detection, SC: standard colonoscopy, RCT: randomized controlled trial, CNN: convoluted neural networks, SD: standard deviation, ADR: adenoma detection rate, PDR: polyp detection rate, FIT: fecal immunochemical testing, GI: gastrointestinal, CRC: colorectal cancer.

sessile adenoma, large adenoma (≥10 mm), small adenoma (<10 mm) and right-sided adenomas were significantly greater with CAD aided colonoscopy. The pooled rates are summarized in Table 2 and forest plots are provided in supplementary materials: Supplementary Figs. 7–14.

3.3. Validation of meta-analysis results

3.3.1. 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.

3.3.2. Heterogeneity

The I^2% values are summarized in Table 2. Moderate to no heterogeneity was observed in the analysis reflecting the real-life applicability and reproducibility of the results on this study.

3.3.3. Publication bias

A publication bias assessment was deferred in this study due to the fact that the total number of studies included was less than ten.

4. Discussion

We report a statistically significant increase in ADR (RR=1.5, $p<0.0001$) and PDR (RR=1.42, $p<0.00001$) with the aid of CNN based AI during colonoscopy as compared to standard colonoscopy, with an increased scope withdrawal time (MD=0.38, p = 0.02). This meta-analysis seems to confirm the hypothesis that AI-based CADe systems in real-time colonoscopy can improve ADR. Although ADR is the primary endpoint of this study, the inherent imperfections of ADR needs to be acknowledged and therefore we report the mean adenoma per colonoscopy that was also statistically significant with CADe (MD=0.19, $p<0.01$).

Different types of deep CNN algorithms were used in the analyzed studies; however, the underlying mathematical concepts are comparable.[9] Using transfer learning, large neural networks can be trained faster with minimal image data in addition to avoiding overfitting. A recent meta-analysis of eighteen studies established the accuracy parameters of CNN based CADe systems in lesion detection during colonoscopy [15]. Prospective real-time studies are being published at a rapid rate evaluating the role of deep CNN based CADe system in real-time colonoscopy.

Our pooled results in terms of per patient ADR data including mean polyp per patient (MD=0.64, $p<0.0001$), mean diminutive adenoma per patient (RR=1.68, p = 0.09), mean flat sessile adenoma per patient (RR=1.75, p = 0.07), mean large adenoma per patient (RR=1.56, p = 0.009) and mean small adenoma per patient (RR=1.39, p = 0.0008) are encouraging and significantly greater with CADe assisted colonoscopy. However, the pooled rates of relative risk for advanced ADR and sessile serrated ADR were comparable between AI assisted and standard colonoscopy. In other words, the use of AI in real-time colonoscopy did not seem to significantly improve the detection of these high-risk lesions with high chances of malignant transformation. Possible explanations include the limited aADR and SSADR training data for the learning of the CADe algorithm, or physician endoscopists being extra careful in identifying these lesions. Another possible reason is that the trials were underpowered to detect these lesions as compared to conventional adenomas.

The efficient detection of a precancerous lesion on colonoscopy depends on various factors including adequate bowel preparation and endoscopist's experience. Physician fatigue and examination time are factors that can potentially lead to a missed lesion. A deep CNN based CADe aid can help circumvent this problem. Other advantages are the monitoring of withdrawal time, which is a quality parameter of

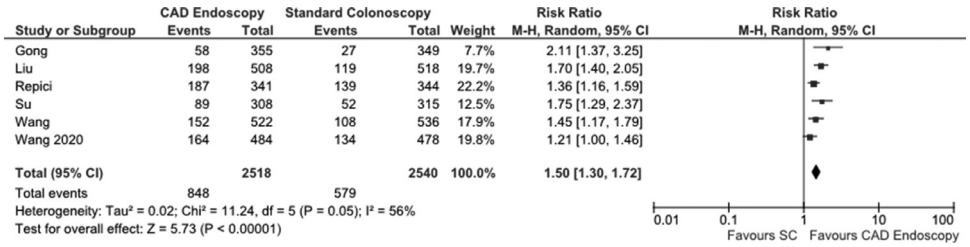

Fig. 2. Forest plot, ADR.

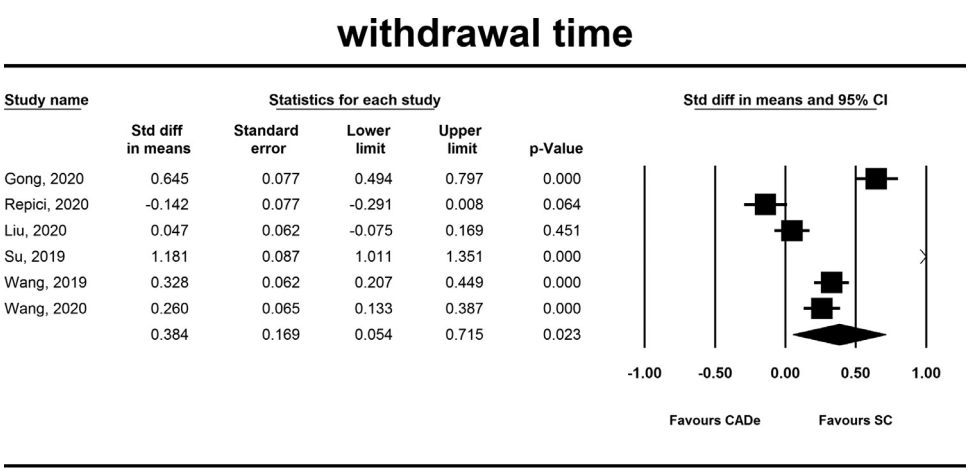

Fig. 3. Forest plot, mean difference in withdrawal time.

Table 2
Summary of results.

	No of studies analyzed; n/N in CADe group & n/N in SC group	Pooled rate (95% CI) Pooled proportions (95% CI)	I^2% heterogeneity	*P*-value
ADR	6 studies;CADe: 848/2518 SC: 579/2540	RR=1.5 (1.3–1.72) CADe: 32.8% (24.2–42.7) SC: 21.1% (14.5–29.7)	56%	$p<0.00001$
ADR (East: studies published in China)	5 studies;CADe: 661/2177 SC: 440/2196	RR=1.55 (1.3–1.85) CADe: 29% (22.5–36.4) SC: 18.3% (13.1–24.9)	60%	$p = 0.04$
ADR (higher quality studies)	5 studies;CADe: 650/2010 SC: 460/2022	RR=1.45 (1.25–1.68) CADe: 31.5% (21.4–43.8) SC: 20.7% (12.8–31.7)	51%	$p = 0.09$
Advanced ADR (aADR)	4 studies;CADe: 77/1855 SC: 78/1876	RR=1 (0.74–1.36) CADe: 3.9% (1.8–8.4) SC: 4% (2–7.9)	0%	$p = 0.93$
PDR	6 studies;CADe: 1271/2518 SC: 889/2540	RR=1.42 (1.33–1.51) CADe: 52% (41–62.8) SC: 35.3% (26.1–45.8)	9%	$p<0.00001$
Sessile serrated ADR	3 studies;CADe: 59/1347 SC: 46/1358	RR=1.29 (0.89–1.89) CADe: 4.5% (2.7–7.2) SC: 3.5% (2.2–5.4)	0%	$p = 0.18$
Mean Adenoma per colonoscopy	6 studies;2518 in CADe & 2540 in SC	MD=0.19 (0.16–0.21)	90%	$p<0.001$
Withdrawal time	6 studies	MD=0.38 (0.054–0.715)	97%	$p = 0.02$
Cecal intubation time	5 studies;2168 in CADe & 2191 in SC	MD=0.04 (−0.29–0.38)	60%	$p = 0.8$
False positives on CADe	4 studies	Pooled rate= 10.3% (6.1–16.8)	93%	-na-
Per patient analysis				
Mean polyp per patient	5 studies;2177 in CADe & 2196 in SC	MD=0.64 (0.45–0.83)	90%	$P<0.00001$
Mean diminutive adenoma per patient	5 studies;CADe: 802/2163 SC: 481/2191	RR=1.68 (1.46–1.92) CADe: 37.1% (33–41.4) SC: 22.2% (18.5–26.3)	50%	$p = 0.09$
Mean flat-sessile adenoma per patient	5 studies;CADe: 980/2163 SC: 566/2191	RR=1.75 (1.54–1.98) CADe: 45.2% (39.1–51.6) SC: 25.8% (21.5–30.6)	54%	$p = 0.07$
Mean large adenoma per patient	4 studies;CADe: 83/1855 SC: 53/1876	RR=1.56 (1.12–2.19) CADe: 4.2% (2–8.7) SC: 2.5% (0.9–6.7)	0%	$p = 0.009$
Mean small adenoma per patient	4 studies;CADe: 220/1855 SC: 159/1876	RR=1.39 (1.15–1.69) CADe: 11.9% (10.5–13.4) SC: 8.5% (7–10.2)	0%	$p = 0.0008$
Mean right sided adenoma per patient	6 studies;CADe: 332/2163 SC: 243/2191	RR=1.36 (1.18–1.58) CADe: 14.8% (8.1–25.5) SC: 10.2% (5.1–19.1)	0%	$p<0.0001$

CADe: computer aided detection, SC: standard colonoscopy, ADR: adenoma detection rate, PDR: polyp detection rate, RR: risk ratio, MD: mean difference.

paramount importance, and the assisted view of blurred images captured during rapid movement of colonoscope. Still, the output results of a machine algorithm are only as good as its input training. Therefore, great impetus must to given to the learning curve of the machine learning software with new training images and/ or videos [8,16–18].

Although the results of ADR improvement seemed to be modest with AI, a previous large network meta-analysis comparing different techniques to improve ADR found that low-cost optimization of existing resources, such as water-aided colonoscopy or addition of a second observer, represents the most cost-effective strategy in this setting and performs even better than newer expensive scopes [19]. Based on our findings, a combination approach with newer technologies based on deep CNNs could potentially enhance the overall ADR, although further trials are needed to confirm these assumptions.

The strengths of this review reside on the careful selection of RCTs reporting on deep CNN based on real-time colonoscopy procedure. With six total studies, this is the largest meta-analysis and therefore adds important data to the current literature on this topic. A recent meta-analysis focused on diagnostic performance of AI systems but not on direct comparison with HD colonoscopy for ADR [15]. Few other meta-analyses have recently been published highlighting similar findings as this study [20–22]. However, this study differs in the reporting of expanded pooled rates of colonoscopy parameters including the scope withdrawal time.

Limitations of this study are primarily related to the fact that the majority of the studies come from one geographical location in addition to a lack in uniformity of the CADe algorithms used across the centers. This limits the generalized global applicability of our results and possibly reflect the performance of Chinese centers in general. Studies did not evaluate ADR strictly for screening indications alone and a lack of stratification of outcomes based on the indication for colonoscopy prevented us from performing a sensitivity analysis restricting to ADR with CADe in screening colonoscopy. Furthermore, bias pertaining to performance and outcomes detection was unavoidable to the unblinded nature of included trials. With time and increasing use of a global endoscopy related image database, algorithms could potentially be trained uniformly across centers. Although the technology is rapidly advancing in AI, we do not anticipate CNN based deep learning to get obsolete before further real-life prospective studies are reported.

In conclusion, based on our meta-analysis, deep CNN based CADe system significantly increases ADR during real-time colonoscopy, albeit with increased withdrawal time.

Declaration of Competing Interest

All other authors declare they have nothing to disclose and have no conflicts of interest.

Unnikrishnan M Pattath, BTECH, MBA; Artificial intelligence solutions, Bangalore, India, for help with technical details on deep CNN based algorithms.

Data sharing statement

This is a meta-analysis of already published studies. The data used can be found in the original studies and in the data-table provided in this study and/ or its supplementary files.

Funding statement

No funding was received for this study.

Supplementary materials legend

Supplementary Figure 1: Forest plot, PDR
Supplementary Figure 2: Forest plot, ADR (East)
Supplementary Figure 3: Forest plot, ADR (higher quality studies)
Supplementary Figure 4: Forest plot, aADR
Supplementary Figure 5: Forest plot, SSADR
Supplementary Figure 6: Forest plot, mean adenoma per colonoscopy
Supplementary Figure 7: Forest plot, mean polyp per patient
Supplementary Figure 8: Forest plot, mean diminutive adenoma per patient
Supplementary Figure 9: Forest plot, mean flat-sessile adenoma per patient
Supplementary Figure 10: Forest plot, mean large adenoma per patient
Supplementary Figure 11: Forest plot, mean small adenoma per patient
Supplementary Figure 12: Forest plot, mean right-sided adenoma per patient
Supplementary Figure 13: Forest plot, mean insertion time
Supplementary Figure 14: Forest plot, false positive on CADe
Supplementary Figure 15: Risk of bias graph
Supplementary Figure 16: Risk of bias summary
Appendix-1: Literature search strategy
Appendix-2: PRISMA checklist

Acknowledgements

Dana Gerberi, MLIS, Librarian, Mayo Clinic Libraries, for help with the systematic literature search.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2020.100622.

References

[1] Corley DA, Levin TR, Doubeni CA. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:2541.
[2] Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology.
[3] Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–9.
[4] Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol 2020;5:352–61.
[5] Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26:13–9.
[6] Su JR, Li Z, Shao XJ, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020;91 415-424.e4.
[7] Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5:343–51.
[8] Ebigbo A, Palm C, Probst A, et al. A technical review of artificial intelligence as applied to gastrointestinal endoscopy: clarifying the terminology. Endosc Int Open 2019;7:E1616–23.
[9] Gao J, Jiang Q, Zhou B, et al. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: an overview. Math Biosci Eng 2019;16:6536.
[10] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Ann. Intern. Med. 2009;151:264–9.
[11] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
[12] Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ: Br Med J 2003;327:557.
[13] Mohan BP, Adler DG. Heterogeneity in systematic review and meta-analysis: how to read between the numbers. Gastrointest Endosc 2019;89:902–3.
[14] Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj 2011;343:d5928.
[15] Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. Gastrointestinal Endosc
[16] Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. Lancet Gastroenterol Hepatol 2019;4:71–80.

[17] Hoerter N, Gross SA, Liang PS. Artificial intelligence and polyp detection. Curr Treat Options Gastroenterol 2020;18:120–36.
[18] Hsieh Y-H, Leung FW. An overview of deep learning algorithms and water exchange in colonoscopy in improving adenoma detection. Expert Rev Gastroenterol Hepatol 2019;13:1153–60.
[19] Facciorusso A, Triantafyllou K, Murad MH, et al. Compared abilities of endoscopic techniques to increase colon adenoma detection rates: a network meta-analysis. Clin Gastroenterol Hepatol 2019;17 2439-2454.e25.
[20] Aziz M, Fatima R, Dong C, et al. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: a systematic review with meta-analysis. J Gastroenterol Hepatol 2020.
[21] Barua I, Vinsard D, Jodal H, Løberg M, Kalager M, Ø Holme, Misawa M, Bretthauer M, Mori Y. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy 2020 (AAM).
[22] Ishita B, Daniela V, Henriette J, Magnus L, Mette K, Øyvind H, Masashi M, Michael B, Yuichi M. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy.