Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies

Silvia Pecere, Giulio Antonelli, Mario Dinis Ribeiro, Yuichi Mori, Cesare Hassan, Lorenzo Fuccio, Raf Bisschops, Guido Costamagna, Eun Hyo Jin, Dongheon Lee, Masashi Misawa, Helmut Messmann, Federico Iacopini, Lucio Petruzziello, Alessandro Repici, Yutaka Saito, Prateek Sharma, Masayoshi Yamada, Cristiano Spada, Leonardo Frazzoni

Angaben zur Veröffentlichung / Publication details:

Pecere, Silvia, Giulio Antonelli, Mario Dinis Ribeiro, Yuichi Mori, Cesare Hassan, Lorenzo Fuccio, Raf Bisschops, et al. 2022. "Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies." *United European Gastroenterology Journal* 10 (8): 817–26. https://doi.org/10.1002/ueg2.12285.

Received: 15 March 2022 | Accepted: 18 July 2022

DOI: 10.1002/ueg2.12285

REVIEW ARTICLE

UNITED EUROPEAN GASTROENTEROLOGY ueg journal WILEY

Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies

Silvia Pecere[1,2] | Giulio Antonelli[3,4] | Mario Dinis-Ribeiro[5] | Yuichi Mori[6,7] | Cesare Hassan[8,9] | Lorenzo Fuccio[10] | Raf Bisschops[11] | Guido Costamagna[1,2] | Eun Hyo Jin[12] | Dongheon Lee[13] | Masashi Misawa[7] | Helmut Messmann[14] | Federico Iacopini[4] | Lucio Petruzziello[1,2] | Alessandro Repici[8] | Yutaka Saito[15] | Prateek Sharma[16] | Masayoshi Yamada[15] | Cristiano Spada[2,17] | Leonardo Frazzoni[10]

[1]Digestive Endoscopy Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

[2]Centre for Endoscopic Research Therapeutics and Training (CERTT), Università Cattolica del Sacro Cuore, Rome, Italy

[3]Department of Anatomical, Histological, Forensic Medicine and Orthopedics Sciences, "Sapienza" University of Rome, Rome, Italy

[4]Gastroenterology and Digestive Endoscopy Unit, Ospedale dei Castelli Hospital, Rome, Italy

[5]CIDES/CINTESIS, Faculty of Medicine, University of Porto, Porto, Portugal

[6]Clinical Effectiveness Research Group, University of Oslo, Oslo, Norway

[7]Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan

[8]Department of Biomedical Sciences, Humanitas University, Milan, Italy

[9]Department of Gastroenterology, IRCCS Humanitas Research Hospital, Milan, Italy

[10]Department of Medical and Surgical Sciences (DIMEC), University of Bologna, S. Orsola-Malpighi Hospital, Bologna, Italy

[11]Department of Gastroenterology and Hepatology, University Hospitals Leuven, TARGID, KU Leuven, Belgium

[12]Department of Internal Medicine, Healthcare Research Institute, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Korea

[13]Department of Biomedical Engineering, College of Medicine, Chungnam National University and Hospital, Daejeon, South Korea

[14]III Medizinische Klinik, Universitatsklinikum Augsburg, Augsburg, Germany

[15]Endoscopy Division, National Cancer Center Hospital, Tokyo, Japan

[16]Department of Gastroenterology and Hepatology, University of Kansas Medical Center, Kansas City, Kansas, USA

[17]Fondazione Poliambulanza, Brescia, Italy

Correspondence
Silvia Pecere, Digestive Endoscopy Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy.
Email: silvia.pecere@gmail.com

Funding information
Open access funding provided by BIBLIOSAN.

[Correction added on 25 August 2022, after first online publication: Surname spelling of author 'Eun Hyo Jin' and affiliation of author 'Dongheon Lee' have been corrected.]

Abstract

Widespread adoption of optical diagnosis of colorectal neoplasia is prevented by suboptimal endoscopist performance and lack of standardized training and competence evaluation. We aimed to assess diagnostic accuracy of endoscopists in optical diagnosis of colorectal neoplasia in the framework of artificial intelligence (AI) validation studies.

Literature searches of databases (PubMed/MEDLINE, EMBASE, Scopus) up to April 2022 were performed to identify articles evaluating accuracy of individual endoscopists in performing optical diagnosis of colorectal neoplasia within studies

validating AI against a histologically verified ground-truth. The main outcomes were endoscopists' pooled sensitivity, specificity, positive and negative predictive value (PPV/NPV), positive and negative likelihood ratio (LR) and area under the curve (AUC for sROC) for predicting adenomas versus non-adenomas.

Six studies with 67 endoscopists and 2085 (IQR: 115–243,5) patients were evaluated. Pooled sensitivity and specificity for adenomatous histology was respectively 84.5% (95% CI 80.3%–88%) and 83% (95% CI 79.6%–85.9%), corresponding to a PPV, NPV, LR+, LR− of 89.5% (95% CI 87.1%–91.5%), 75.7% (95% CI 70.1%–80.7%), 5 (95% CI 3.9%–6.2%) and 0.19 (95% CI 0.14%–0.25%). The AUC was 0.82 (CI 0.76–0.90). Expert endoscopists showed a higher sensitivity than non-experts (90.5%, [95% CI 87.6%–92.7%] vs. 75.5%, [95% CI 66.5%–82.7%], $p < 0.001$), and Eastern endoscopists showed a higher sensitivity than Western (85%, [95% CI 80.5%–88.6%] vs. 75.8%, [95% CI 70.2%–80.6%]). Quality was graded high for 3 studies and low for 3 studies. We show that human accuracy for diagnosis of colorectal neoplasia in the setting of AI studies is suboptimal. Educational interventions could benefit by AI validation settings which seem a feasible framework for competence assessment.

KEYWORDS
artificial intelligence, colonoscopy, endoscopist performance, human factor, polyp characterization, polyp detection

INTRODUCTION

A substantial proportion of the cost of population-based Colorectal Cancer (CRC) screening program is due to removal and subsequent pathology assessment of diminutive colorectal polyps that represent more than 80% of all the detectable lesions.[1–4] Optical diagnosis has been shown to be able to in vivo predict histology of these diminutive lesions in expert centers, opening the way to cost-saving strategies, namely the Leave-in-Situ for ≤5 mm rectosigmoid hyperplastic lesions, and Resect and Discard for all the others.[5,6]

Disappointingly, implementation of these cost-saving strategies has been hampered by suboptimal results in community-based controlled trials, questioning on the actual accuracy of endoscopists in the optical diagnosis of diminutive lesions.[7] However, a direct assessment of the accuracy of individual endoscopists in optical diagnosis is limited to few studies, leaving uncertainty on the actual need of educational interventions as well as on the best approach.[8]

Artificial Intelligence (AI) has been claimed to predict histology of diminutive polyps in real-time endoscopy. For this reason, several AI-algorithms have been tested in standalone performance studies against a ground-truth generally represented by pathologically verified polyps selected by expert endoscopists, resulting in an overall accuracy of over 90%.[9] In order to better define it, AI performance has been generally benchmarked against multiple endoscopists with different degrees of competence which were administered the same sets of images/videos analyzed by AI.[10,11] This framework provides a unique modality of assessing the performance of human endoscopists when dealing with optical diagnosis, and could be used as testing ground for the application of PIVI criteria.[6]

Aim of our study was to evaluate the accuracy of human endoscopists in performing optical diagnosis of colorectal polyps extracting their performances from studies on standalone performance of AI systems, as well as on possible associated factors. Such analysis could set the grounds for new modalities of training and competence evaluation in colorectal lesion evaluation.

METHODS

The methods of our analysis and inclusion criteria were based on Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) recommendations.[12]

The PRISMA Checklist is available in Supporting Information S1.

Study registration

This study was registered on the PROSPERO international database (University of York Centre for Reviews, www.crd.york.ac.uk/prospero/). Number: 279321.

Inclusion and exclusion criteria

Only original full articles published in English have been considered for the study. Abstract, letter or review articles were excluded. All studies reporting the use of AI for characterization of colorectal adenoma compared to human characterization with histological confirmation as ground truth have been included.

Search strategy and data extraction

We performed a comprehensive literature search of two scientific databases (PubMed/Medline and Scopus) up to April 2022 to identify full articles evaluating the diagnostic accuracy of AI-assisted colonoscopy for characterization of colorectal adenoma compared to "human control" of expert and non-expert endoscopists. Electronic searches were supplemented by manual searches of references of included studies. Complete search strategy and search strings used are available in Supporting Information S2. Two authors (SP and GA) independently evaluated all titles and abstracts of the identified articles to exclude papers not strictly related to the aim of the study or meeting inclusion criteria. Remaining abstract and full text were further screened for eligibility. Finally, any disagreement was discussed and solved with senior authors. Data extraction from eligible study was performed using the following scheme (a blank example of our data extraction table is available in Supporting Information S3):

- the total number of images/cases and the number of total positive images/cases (predicted as adenoma/non adenoma by endoscopists and confirmed by histology)
- the numbers of images/cases classified as true positive (images/cases showing colorectal lesion predicted-as-adenoma by AI), true negative (images/cases showing non-neoplastic mucosa without AI detection or lesions predicted as non-neoplastic), false positive (FP, images/cases showing non-neoplastic mucosa or lesions detected/predicted-as-neoplastic by AI) or false negative (images/cases showing a neoplastic lesion missed by AI or predicted as non-neoplastic)

In addition, country of provenience, type of study, number of patients, characteristics of polyps detected were also considered. Corresponding authors were contacted for data extraction in case of missing information from published studies.

Study outcomes

Primary endpoint of the study was the pooled diagnostic endoscopists' accuracy for the characterization of colorectal adenoma in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), likelihood ratio (LR+) and negative likelihood ratio (LR). The accuracy of endoscopists was defined as "hierarchical summary receiver-operating characteristic (SROC) curve (area under the curve; AUC)."

Secondary outcomes were the diagnostic performance according to study design, endoscopist's level of expertise and country of provenience.

Quality of studies

The degree of bias was assessed using a modified version of the QUADAS (quality for assessment of diagnostic studies score)[10,13] score that was already used in previous publications. We include specific bias domains for diagnostic studies in AI. We divided in two main domains and respective subdomains, namely Training set bias (subdomains: Selection bias, Spectrum Bias and Operator bias) and Validation set bias (subdomains: Overfitting bias and Operator bias). For Overfitting bias we considered at low risk of bias papers explicitly describing the use of overfitting mitigation techniques as data augmentation, dropout, batch normalization, regularization, early stopping, and transfer learning from large datasets.

Statistical analysis

We computed summary estimates of sensitivity, specificity, LR+ and LR− of GI endoscopists on a "per-endoscopist" basis, through the bivariate mixed-effects regression model proposed by Reitsma et al[14]; 95% Confidence Intervals (CIs) for the diagnostic accuracy parameters were computed through the bivariate model, as well. Positive predictive value (PPV) and negative predictive value (NPV) were obtained for the pooled prevalence of lesions. Forest plots for sensitivity and specificity, and summary receiving operating characteristic (SROC) curve were drawn. Positive and negative likelihood ratios were applied to the pooled prevalence of the various types of UGI premalignant and malignant lesions (i.e. pre-test probability), to derive the post-test probability in case of a positive or negative test result; a Fagan's plot was derived, accordingly.

Heterogeneity was assessed through visual inspection of forest plots and SROC curve, and quantified by the between-study standard deviation (SD) for logit-transformed sensitivity and specificity.[15] We assessed heterogeneity through sensitivity analyses based on subgroup meta-analyses and bivariate meta-regression models. Variables which might have influenced the diagnostic accuracy of GI endoscopists were defined a priori at two levels: (i) the "endoscopist" level, that is, the experience of the endoscopists participating to the included studies as dichotomized into expert and non-expert according to study definitions; (ii) the "study" level, that is, study size, mono versus multicenter studies, country, number of images provided, percentage of lesions in the right colon, percentage of adenomas, and quality of studies. All the analyses were performed with the package *mada*[16] for R.[17]

RESULTS

Search data

The search strategy yielded a total of 1267 studies. Once duplicates were removed, a total of 987 studies were screened by analysis of title and abstract and 959 studies were removed because not related to the study topic or not meeting inclusion criteria. Then, 28 studies were entirely evaluated for eligibility and among them 20 were excluded, all for the absence of histological confirmation and two of them were excluded since accuracy data of performance for each endoscopy was not available in the full text. Finally, six

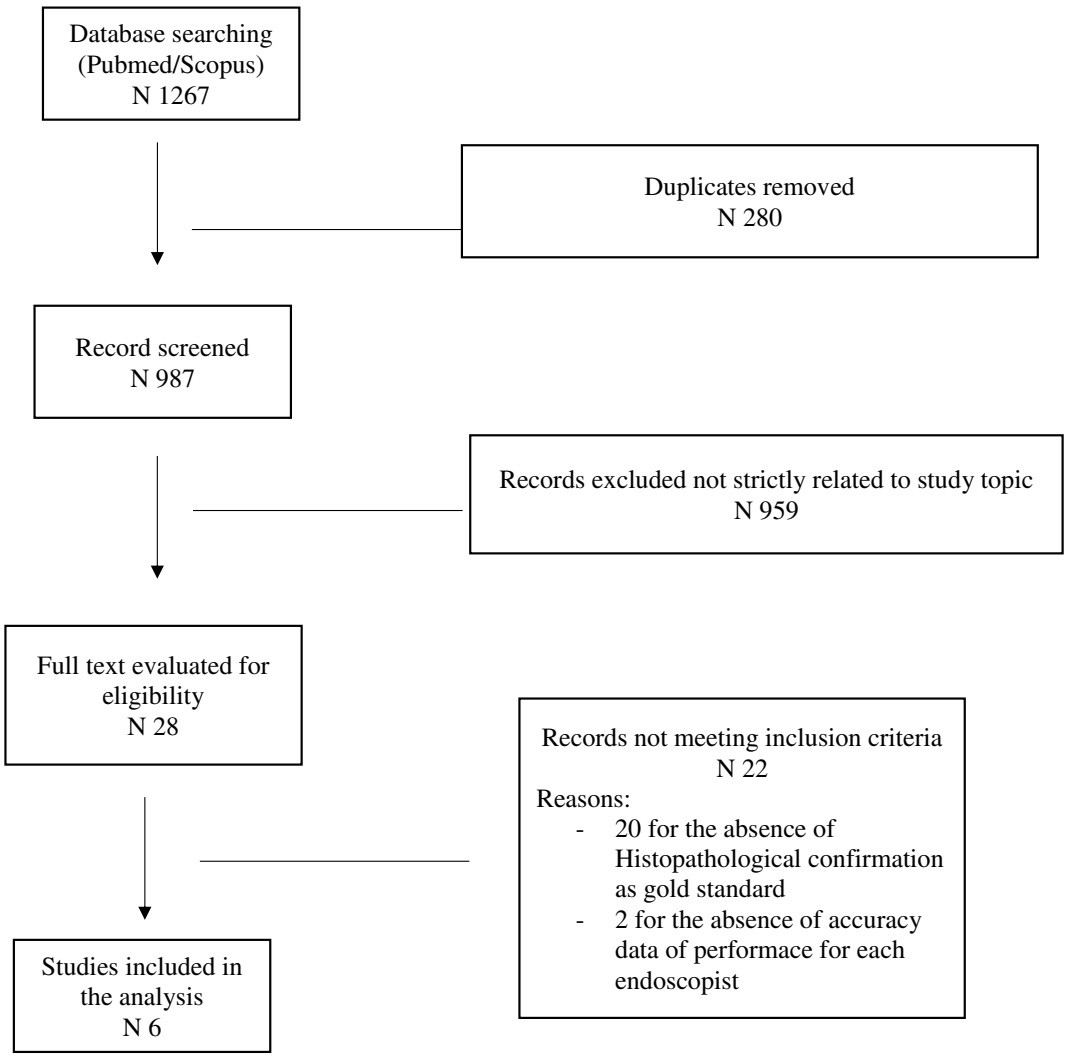

FIGURE 1 Flow-chart of included studies

articles[18–23] were included in the statistical analysis (Figure 1—Study Flowchart).

Study details

Among 6 studies included (Table 1), 4 had a single center design,[18–20,22] while 2 had a multicenter design[21,23](6). Regarding geographical area, 4 studies were performed in Eastern centers[18,20–22] and two in Western centers.[19,23] Olympus endoscopes with narrow-band imaging (NBI) filter were used by five out of six studies and two of them added endocitoscopy (EC). All studies were based on characterization of adenoma/non-adenoma lesions and had histo-pathological evaluation as standard reference. Median number of included patients was 208.5 (IQR: 115–243.5). All studies reported the number of total images used for adenoma/non adenoma character-ization, accounting for a total of 1368 (median: 209; IQR: 108.5–296). Regarding polyp details, all studies considered colorectal polyps <10 mm (for one study the data was missing). Polyp morphology was protruded type (Paris type Is, Isp or Ip) and slightly elevated type (Paris type IIa) in five studies, while an Eastern study included also slightly depressed type polyps (Paris type IIc). Complete characteristics of polyps are reported in Table 2.

Endoscopists characteristics

Overall, 67 endoscopists from 6 studies were included in the analysis. Of these, 5/67 (7.46%) endoscopists came from Western centers and 62/67 (92.55%) from Eastern centers. Expert endoscopists were 34/67 (50.75%) while 33/67 (49.25%) were considered non-experts and were all located in Eastern countries.

Primary outcome

The pooled prevalence of colorectal adenoma among all images shown to endoscopists was 8576/13705 (63.2%; 95% CI 62.1%–

TABLE 1 Details of included studies

First Author, Year	Design	Country	Patients (*n*)	Consecutive Y/N	Images (*n*)	Endoscopists (*n*)	Expert (*n*)	Non expert (*n*)	AI type	Imaging type	Setting
Chen, 2018	U	E	193	Y	284	6	2	4	CAOB	HDWL/ magnifying NBI	Experimental images only
Renner, 2018	U	W	250	Y	100	2	2	0	DNN-CAD	HDWL/NBI	Experimental images only
Mori et al., 2018	U	E	320	Y	450	4	2	2	CAD system	EC-NBI	Real time images and videos
Kudo, 2020	M	E	89	N	100	30	10	20	EndoBRAIN	WL/EC-NBI	Experimental images only
Jin, 2020	U	E	224	N	300	22	15	7	CNN system	NBI/near-focus	Experimental images only
Weigt, 2021	M	W	80	Y	134	3	3	0	CAD-EYE	WL/LCI/BLI	Experimental images and videos

Abbreviations: BLI, Blue Light Imaging; CAD, Computer Aided Detection; CAOB, computer-assisted optical biopsy; CNN, convolutional neural network; DNN, deep neural network; E, Eastern; EC, endocitoscopy; HDWL, high definition white light; LCI, Linked Color Imaging; M, multicentric; NBI, narrow band imaging; U, unicentric; W, Western.

TABLE 2 Polyps characteristics

Polyps characteristics			
Study	**Size (mm)**	**Shape (Paris class)**	**Location % (right/left colon)**
Chen, 2018[18]	<5	Is – Isp – IIa	34.8/65.2
Renner, 2018[19]	<5	Is – Ip – IIa	51/49
Mori, 2018[22]	<5	Is – Ip – IIa – IIc	40.4/59.6
Kudo, 2020[21]	<10	Is – Isp – IIa	38/62
Jin, 2020[20]	<5	Is – Isp – IIa	54.3/45.7
Weigt, 2021[23]	-	Is – IIa	44/35.5 (20.9 missing)

64.4%). Overall, 67 endoscopists from 6 studies had a pooled sensitivity and specificity of 84.5% (95% CI 80.3%–88%) and 83% (95% CI 79.6%–85.9%), respectively for adenomatous histology. In addition, PPV and NPV were 89.5% (95% CI 87.1%–91.5%) and 75.7% (95% CI 70.1%–80.7%), respectively, corresponding to positive and negative likelihood ratio (LR+/LR) of 5 (95% CI 3.9%–6.2%) and 0.19 (95% CI 0.14%–0.25%), with AUC of 0.82 (95% CI 0.76–0.90). Relative SROC curve is available in Figure 2.

Secondary outcomes

Experienced endoscopists had a significantly higher sensitivity (90.5%, [95% CI 87.6%–92.7%] vs. 75.5%, [95% CI 66.5%–82.7%], $p < 0.001$) and specificity than non-expert endoscopists (84.8% [95% CI 82.3%–87.8%] vs. 81.4% [95% CI 75.1%–86.4%], $p < 0.84$), corresponding to a NPV of 84% (95% CI 79.4%–87.6%) versus 66.1% (95% CI 64.9%–80.5%). The forest plots for sensitivity and specificity can be found in Figure 3.

Eastern endoscopists showed higher sensitivity than Western endoscopists (85%, [95% CI 80.5%–88.6%] vs. 75.8%, [95% CI

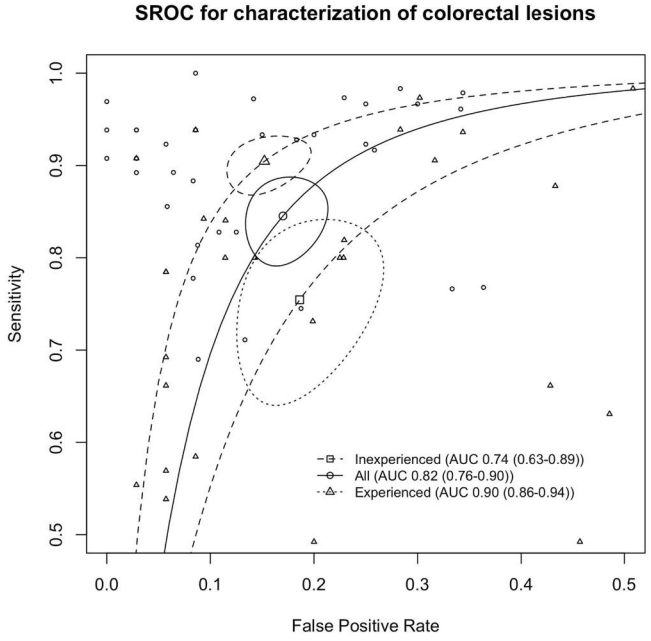

FIGURE 2 Summary receiver-operating characteristic curve

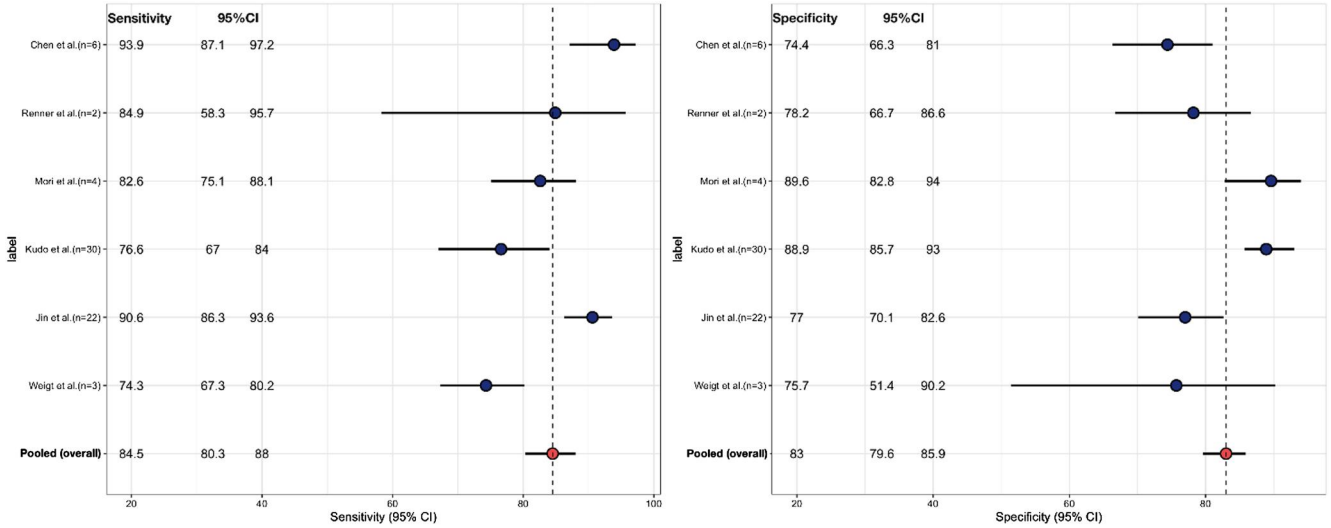

FIGURE 3 Forest plots for sensitivity and specificity by study

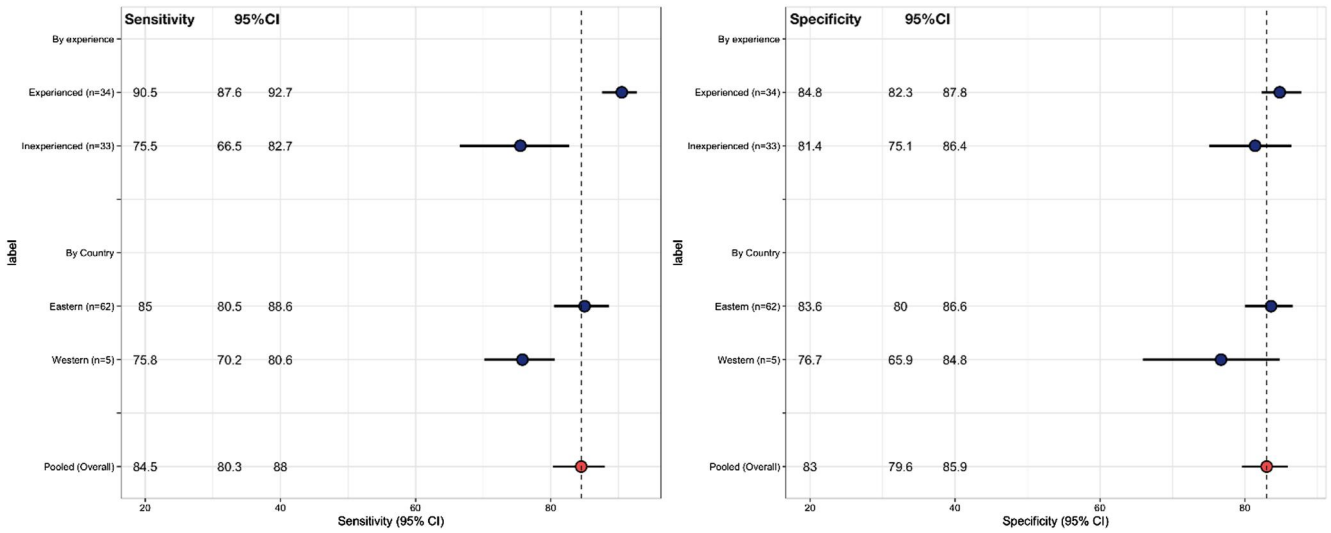

FIGURE 4 Forest plots for sensitivity and specificity by experience and country

70.2%–80.6%]) and higher specificity (83.6%, [95% CI 80%–86.6%] vs. 76.7%, [95% CI 65.9%–84.8%]). The forest plots is available in Figure 4. Moreover, sensitivity is significantly higher for endoscopists of single center design studies (90.2% [95% CI 86.9%–92.8%]) than multicenter studies (76.2% [95% CI 76.2%–88.8%], $p < 0.001$), while on the contrary, endoscopists of multicenter studies seem to have a better rate of specificity (88.8% [95% CI 84.5%–92.1%]) than single center studies (78.5% [95% CI 73.7%–82.7%], $p = 0.001$). Details in Table 3.

Additional analysis

Meta-regression analysis for other studies variables showed a positive relation between the number of images and sensitivity ($p = 0.002$) and a negative relation with specificity ($p = 0.032$). Also the rate of right colon lesions had a significant impact on sensitivity ($p < 0.001$). Details available in Table 4.

Quality of studies

Study quality assessment according to the modified QUADAS score is available in Table 5. In detail, 3 out of 6 studies[20–22] were considered of High quality, and 3 studies[18,19,23] was considered of Low quality. There was a tendency to spectrum bias in the included studies, as often the images were only selected among high quality images or best framing of the polyp. Meta regression analysis including study quality is available in Table 3.

DISCUSSION

By exploiting the artificial setting represented by AI validation studies, we measured a suboptimal performance of human endoscopists in the optical diagnosis of diminutive to small polyps that appears to be not compatible with the implementation of clinical

TABLE 3 Subgroup meta-analyses for summary diagnostic accuracy measures of endoscopists for adenoma characterization at colonoscopy, according to study variables

Study variable (*n* of endoscopists)	Sensitivity (95% CI)	*p*-value for sensitivity	Specificity (95% CI)	*p*-value for specificity
Endoscopists' experience				
Experienced (*n* = 34)	90.5 (87.6–92.7)	<0.001	84.8 (82.3–87.8)	0.084
Inexperienced (*n* = 33)	75.5 (66.5–82.7)		81.4 (75.1–86.4)	
Country				
Eastern (*n* = 62)	85 (80.5–88.6)	0.436	83.6 (80–86.6)	0.28
Western (*n* = 5)	75.8 (70.2–80.6)		76.7 (65.9–84.8)	
Study design				
Monocenter (*n* = 34)	90.2 (86.9–92.8)	<0.001	78.5 (73.7–82.7)	0.001
Multicenter (*n* = 33)	76.2 (67.7–83.1)		88.8 (84.5–92.1)	
Study quality				
High (*n* = 56)	83.6 (78.6–87.6)	0.359	84.6 (80.9–87.8)	0.051
Low (*n* = 11)	89 (80.8–93.9)		75.6 (70.1–80.4)	

TABLE 4 Meta-regression analysis for continuous moderators

Study variable	Coefficient for sensitivity (95% CI)	*p*-value for impact on sensitivity	Coefficient for 1-specificity (95% CI)	*p*-value for impact on specificity
Number of images	0.004 (0.001–0.006)	0.002	0.002 (0.001–0.004)	0.032
Percentage of right colon lesions	−0.081 (−0.126–−0.037)	<0.001	0.086 (−1.084–1.256)	0.886
Relative frequency of adenomas	−5.310 (−11.429–0.809)	0.089	−1.305 (−6.053–3.443)	0.590

TABLE 5 Quality assessment

	Reference standard/Training set			Index test/Validation set		
Study	Selection bias	Spectrum bias	Operator bias	Overfitting bias	Operator bias	Overall quality
Chen, 2018	↑	↑	↓	↑	↓	Low
Renner, 2018	↑	↑	↑	↓	↑	Low
Mori et al., 2018	↓	↓	↓	↓	↑	High
Kudo, 2020	↓	↑	↓	↓	↓	High
Jin, 2020	↓	↓	↓	↓	↓	High
Weigt, 2021	↓	↑	↓	↓	↑	Low

Note: ↓ low risk of bias ↑ high risk of bias.

strategies based on a human-alone evaluation. Even though sensitivity and specificity overall would be slightly over the 80% threshold recently proposed by ESGE for the resect-and-discard strategy (24), the 76% NPV for adenomatous histology is disappointingly far from the 90% cut-off universally recognized as the minimum cut-off to implement the leave-in-situ strategy.[6,24,25] This result is not fully unexpected since previous literature[26,27] has shown how, especially in the community setting, endoscopists fail to reach required thresholds. However, we show for the first time how the development framework of CADx systems may be an optimal platform to assess endoscopist competence.

The main clinical relevance of our study is the intimate association between the level of competence as defined by the degree of experience and the accuracy of individual endoscopists. Remarkably, the much higher sensitivity of experts versus non-experts—90.5% versus 75.5%—indicates a much higher risk of false-negative cases for adenomatous histology that is adenomas misinterpreted as hyperplastic polyps by non-expert endoscopists. This

is by far the worst error that can come from an inaccurate in vivo prediction as potentially condemning a high-risk patient with multiple adenomas (i.e., ≥3 low-risk adenomas) who needs an intensive post-polypectomy surveillance to a low-risk category without the necessary endoscopic surveillance.[28] In addition, we showed that an Eastern location of the endoscopists is also associated with a higher accuracy in optical diagnosis, irrespectively of the level of experience. This shows that the training approach that is much more meticulous and image-based in Eastern as compared to Western school is critical in the development of adequate skills in polyp characterization. Thus, a dedicated image-based training is needed for Western endoscopists, and in this regard the artificial setting adopted in our pooled studies could be at least a good benchmarking when testing the outcome of such educational interventions.

Our study indirectly supports the validity of community-based studies on optical diagnosis of diminutive polyps showing a suboptimal performance not matching the required standards.[27] Indeed, recent meta-analysis on training modalities for optical diagnosis has shown pre-training accuracy levels as low as 68.1%, as well as an unsatisfactory post-training performance ranging from 77.1% to 81.6%.[8] Such results are likely to be the direct consequence of the low sensitivity we measured rather than related with the clinical setting where such accuracy was tested, that is, distraction related with real-life endoscopy, blurred or out-of-focus images, and difficult polyp position.

The strength of our study is to show that an artificial setting, that is, exposure of multiple endoscopists against images of histologically-verified lesions, may be suitable to assess the skills of individual endoscopists in polyp characterization, as much as in benchmarking them against experts or standalone performance of Artificial Intelligence algorithms. In this regard, a recent meta-analysis[9] comprising 7680 images of colorectal polyps from 18 studies showed an accuracy (AUC) of AI of 96% (95% CI 0.95–0.98), corresponding to a sensitivity of 92.3% (95% CI 88.8%–94.9%) and a specificity of 89.8% (95% CI 85.3%–93.0%). When compared with our pooled estimate of 84% for endoscopist-based sensitivity, this would suggest a relevant role for AI in assisting human endoscopists for polyp characterization. Secondly, all studies included diminutive to small polyps that is exactly what is required for the proposed cost-saving resect-and-discard and leave-in-situ strategies.

The main limitations of our study is the per-polyp rather than per-patient analysis due to the image-oriented rather than patient-oriented collection of cases. However, a high per-polyp accuracy should be the base rather than the consequence of a successful clinically-oriented strategy rather than vice versa, and the cost-saving strategies proposed by the PIVI document are indeed polyp- rather than patient-based. Secondly, the actual number of western endoscopists was low, prompting the need for additional data. Third, our quality assessment found a possible spectrum-related bias: indeed, endoscopists were shown images or video frames of lesions specifically selected for the purpose of characterization, which may be an ideal setting. Further, the prevalence of the disease (i.e. adenomas) may not represent clinical practice. Nevertheless, sensitivity and specificity are independent from the prevalence of the disease, therefore such estimates have external validity. Fourth, we cannot fully rule out an underperformance of benchmarking endoscopists leading to investigator bias. However, it must also be noted that benchmarking human endoscopists are ofthen not involved in the study conduction in the first place. Indeed, whether or not benchmarking is undergone by endoscopists from different centers and not involved in the data acquisition and annotation is a major quality indicator of a preclinical AI paper.[10,29] Fifth, current CADx systems do not account for sessile serrated lesions as adenomatous, leading to a possible partial reduction of their accuracy. However, it must be noted that the primary aims of these systems have been first of all to implement cost saving strategies for diminutive polyps. This limits the impact of serrated lesions as their prevalence in the RS tract is negligible and all serrated lesions >5 mm of the whole colon are to be in any case resected and sent to pathology. Last, we provided diagnostic accuracy for adenomatous lesions irrespective of colonic site, therefore the inference on leave-in-situ strategy may be biased. However, although we could not separately assess diagnostic accuracy for rectosigmoid lesions, we performed a metaregression analysis showing that sensitivity tended to reduce when a higher proportion of lesions in the proximal colon was shown to endoscopists. This is in line with current recommendations suggesting to limit the leave-in-situ strategy to rectosigmoid lesions.[30]

In conclusion, we show a disappointingly low accuracy in optical diagnosis of diminutive to small polyps when extracting them from the artificial setting of AI standalone performance studies. The exploitation of the AI development framework for endoscopist competence assessment is feasible and effective.

AUTHOR CONTRIBUTIONS

Silvia Pecere, Giulio Antonelli, Cesare Hassan: conception and design; Silvia Pecere, Giulio Antonelli, Yuichi Mori, Cesare Hassan: data extraction and interpretation; Lorenzo Fuccio, Leonardo Frazzoni: statistical analysis; Silvia Pecere, Giulio Antonelli, Cesare Hassan, Leonardo Frazzoni: drafting of the article; Raf Bisschops, Mario Dinis-Ribeiro, Helmut Messmann, Yuichi Mori, Federico Iacopini, Lucio Petruzziello, Eun Hyo Jin, Yutaka Saito, Masayoshi Yamada, Alessandro Repici, Prateek Sharma, Guido Costamagna, Cristiano Spada: data provision and/or critical revision of the article for important intellectual content. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENT

No funding was obtained for this study.

Open access funding provided by BIBLIOSAN.

CONFLICT OF INTEREST

The authors declare no COI relevant to this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Silvia Pecere https://orcid.org/0000-0002-4401-7344
Masayoshi Yamada https://orcid.org/0000-0003-3979-5560

REFERENCES

1. Greuter MJE, de Klerk CM, Meijer GA, Dekker E, Coupé VM. Screening for colorectal cancer with fecal immunochemical testing with and without postpolypectomy surveillance colonoscopy: a cost-effectiveness analysis. Ann Intern Med. 2017;167:544–54. https://doi.org/10.7326/M16-2891
2. Gupta N, Bansal A, Rao D, Early DS, Jonnalagadda S, Wani SB, et al. Prevalence of advanced histological features in diminutive and small colon polyps. Gastrointest Endosc. 2012;75(5):1022–30. https://doi.org/10.1016/j.gie.2012.01.020
3. Laish I, Sergeev I, Stein A, Naftali T, Konikoff FM. Risk of metachronous advanced lesions after resection of diminutive and small, non-advanced adenomas. Clin Res Hepatol Gastroenterol. 2019;43(2):201–7. https://doi.org/10.1016/j.clinre.2018.03.001
4. Schachschal G, Sehner S, Choschzick M, Aust D, Brandl L, Vieth M, et al. Impact of reassessment of colonic hyperplastic polyps by expert GI pathologists. Int J Colorectal Dis. 2016;31(3):675–83. https://doi.org/10.1007/s00384-016-2523-8
5. Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. Clin Gastroenterol Hepatol. 2010;8(10):865–9.e3. https://doi.org/10.1016/j.cgh.2010.05.018
6. Rex DK, Kahi C, O'Brien M, Levin T, Pohl H, Rastogi A, et al. The American Society for Gastrointestinal Endoscopy PIVI (preservation and incorporation of valuable endoscopic innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc. 2011;73(3):419–22. https://doi.org/10.1016/j.gie.2011.01.023
7. Vu HT, Sayuk GS, Hollander TG, Clebanoff J, Edmundowicz SA, Gyawali CP, et al. Resect and discard approach to colon polyps: real-world applicability among academic and community gastroenterologists. Dig Dis Sci. 2015;60(2):502–8. https://doi.org/10.1007/s10620-014-3376-z
8. Smith SCL, Siau K, Cannatelli R, Shivaji UN, Ghosh S, Saltzman JR, et al. Training methods in optical diagnosis and characterization of colorectal polyps: a systematic review and meta-analysis. Endosc Int Open. 2021;9(05):E716–26. https://doi.org/10.1055/a-1381-7181
9. 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. Gastrointest Endosc. 2020;92(1):11–22.e6. https://doi.org/10.1016/j.gie.2020.02.033
10. Arribas J, Antonelli G, Frazzoni L, Fuccio L, Ebigbo A, van der Sommen F, et al. Standalone performance of artificial intelligence for upper GI neoplasia: a meta-analysis. Gut. 2020;70(8):1458–68. https://doi.org/10.1136/gutjnl-2020-321922
11. Frazzoni L, Arribas J, Antonelli G, Libanio D, Ebigbo A, van der Sommen F, et al. Endoscopist diagnostic accuracy in detecting upper-GI neoplasia in the framework of artificial intelligence studies. Endoscopy. 2021;54(04):403–11. https://doi.org/10.1055/a-1500-3730
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62(10):1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005
13. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
14. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–90. https://doi.org/10.1016/j.jclinepi.2005.02.022
15. Naaktgeboren CA, Ochodo EA, Van Enst WA, de Groot JAH, Hooft L, Leeflang MMG, et al. Assessing variability in results in systematic reviews of diagnostic studies. BMC Med Res Methodol. 2016;16(1):6. https://doi.org/10.1186/s12874-016-0108-4
16. Doebler P, Holling H. Meta-analysis of diagnostic accuracy with mada. p. 21.
17. R: a language and environment for statistical computing. Im Internet. Accessed 7 May 2020. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing
18. Chen P-J, Lin M-C, Lai M-J, Lu HHS, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology. 2018;154(3):568–75. https://doi.org/10.1053/j.gastro.2017.10.010
19. Renner J, Phlipsen H, Haller B, Navarro-Avila F, Saint-Hill-Febles Y, Mateus D, et al. Optical classification of neoplastic colorectal polyps – a computer-assisted approach (the COACH study). Scand J Gastroenterol. 2018;53(9):1100–6. https://doi.org/10.1080/00365521.2018.1501092
20. Jin EH, Lee D, Bae JH, Kang HY, Kwak MS, Seo JY, et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. Gastroenterology. 2020;158(8):2169–79.e8. https://doi.org/10.1053/j.gastro.2020.02.036
21. Kudo S-E, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2019;18(8):1874–81.e2. https://doi.org/10.1016/j.cgh.2019.09.009
22. Mori Y, Kudo S-E, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med. 2018;169(6):357–66. https://doi.org/10.7326/M18-0249
23. Weigt J, Repici A, Antonelli G, Afifi A, Kliegis L, Correale L, et al. Performance of a new integrated computer-assisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. Endoscopy. 2021;54(2):180–4. https://doi.org/10.1055/a-1372-0419
24. Kaminski MF, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees C, Dekker E, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. Endoscopy. 2017;49(04):378–97. https://doi.org/10.1055/s-0043-103411
25. Dekker E, Houwen BBSL, Puig I, Bustamante-Balen M, Coron E, Dobru DE, et al. Curriculum for optical diagnosis training in europe: European Society of Gastrointestinal Endoscopy (ESGE) position statement. Endoscopy. 2020;52(10):899–923. https://doi.org/10.1055/a-1231-5123
26. Ladabaum U, Fioritto A, Mitani A, Desai M, Kim JP, Rex DK, et al. Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions. Gastroenterology. 2013;144(1):81–91. Epub 2012 Oct 3. PMID: 23041328; PMCID: PMC5518757. 28. https://doi.org/10.1053/j.gastro.2012.09.054
27. Rees CJ, Rajasekhar PT, Wilson A, Close H, Rutter MD, Saunders BP, et al. Narrow band imaging optical diagnosis of small colorectal

polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. Gut. 2017;66(5):887–95. https://doi.org/10.1136/gutjnl-2015-310584
28. Hassan C, Antonelli G, Dumonceau J-M, Regula J, Bretthauer M, Chaussade S, et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) guideline – update 2020. Endoscopy. 2020;52(8):a-1185-3109. https://doi.org/10.1055/a-1185-3109
29. van der Sommen F, de Groof J, Struyvenberg M, van der Putten J, Boers T, Fockens K, et al. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. Gut. 2020;69(11):2035–45. Epub 2020 May 11. PMID: 32393540; PMCID: PMC7569393. https://doi.org/10.1136/gutjnl-2019-320466
30. Houwen BBSL, Hassan C, Coupé VMH, Greuter MJE, Hazewinkel Y, Vleugels JLA, et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) position statement. Endoscopy. 2022;54(1):88–99. Epub 2021 Dec 6. PMID: 34872120. https://doi.org/10.1055/a-1689-5130

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pecere S, Antonelli G, Dinis-Ribeiro M, Mori Y, Hassan C, Fuccio L, et al. Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies. United European Gastroenterol J. 2022;10(8):817–26. https://doi.org/10.1002/ueg2.12285