Article published online: 2021-11-12

Automated detection of cecal intubation with variable bowel preparation using a deep convolutional neural network

Authors
Daniel J. Low[1], Zhuoqiao Hong[2], Rishad Khan[1], Rishi Bansal[1], Nikko Gimpaya[1], Samir C. Grover[1]

Institutions
1 St. Michael's Hospital, University of Toronto
2 Massachusetts Institute of Technology

submitted 5.2.2021
accepted after revision 4.6.2021

Bibliography
Endosc Int Open 2021; 09: E1778–E1784
DOI 10.1055/a-1546-8266
ISSN 2364-3722
© 2021. The Author(s).
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Corresponding author
Samir Grover, MD, MEd, St. Michael's Hospital, 30 Bond Street, Toronto, ON M5B 1W8
Fax: +1-416 864-5882
Samir.grover@utoronto.ca

ABSTRACT

Background and study aims Colonoscopy completion reduces post-colonoscopy colorectal cancer. As a result, there have been attempts at implementing artificial intelligence to automate the detection of the appendiceal orifice (AO) for quality assurance. However, the utilization of these algorithms has not been demonstrated in suboptimal conditions, including variable bowel preparation. We present an automated computer-assisted method using a deep convolutional neural network to detect the AO irrespective of bowel preparation.

Methods A total of 13,222 images (6,663 AO and 1,322 non-AO) were extracted from 35 colonoscopy videos recorded between 2015 and 2018. The images were labelled with Boston Bowel Preparation Scale scores. A total of 11,900 images were used for training/validation and 1,322 for testing. We developed a convolutional neural network (CNN) with a DenseNet architecture pre-trained on ImageNet as a feature extractor on our data and trained a classifier uniquely tailored for identification of AO and non-AO images using binary cross entropy loss.

Results The deep convolutional neural network was able to correctly classify the AO and non-AO images with an accuracy of 94 %. The area under the receiver operating curve of this neural network was 0.98. The sensitivity, specificity, positive predictive value, and negative predictive value of the algorithm were 0.96, 0.92, 0.92 and 0.96, respectively. AO detection was >95 % regardless of BBPS scores, while non-AO detection improved from BBPS 1 score (83.95 %) to BBPS 3 score (98.28 %).

Conclusions A deep convolutional neural network was created demonstrating excellent discrimination between AO from non-AO images despite variable bowel preparation. This algorithm will require further testing to ascertain its effectiveness in real-time colonoscopy.

Introduction

Colonoscopy is the gold standard for colorectal cancer diagnosis and subsequent surveillance. The quality of colonoscopy substantially alters the efficacy of adenomatous polyp detection and colorectal cancer diagnosis. The American Society for Gastrointestinal Endoscopy (ASGE), British Society of Gastroenterology (BSE), European Society of Gastrointestinal Endoscopy (ESGE), and the Canadian Association of Gastroenterology (CAG) have baseline quality standards for colonoscopy evaluation [1–5]. These metrics include cecal intubation rate (>90 %–95 %), withdrawal time (>6 minutes), and adenoma detection rate (>15 %–25 %).

The adenoma detection rate (ADR) is a particularly well-characterized quality indicator and is inversely related to the development of post-colonoscopy colorectal cancer (PCCRC) [6]. Other quality metrics, including adequacy of bowel preparation and sufficient withdrawal time, have also been associated with higher ADR and lower rates of subsequent PCCRC [7–10]. Likewise, colonoscopy completion, defined by cecal intu-

bation, has been shown to be negatively associated with PCCRC development [11]. However, there is considerable variability in cecal intubation rates (CIR) and photodocumentation among healthcare practitioners and facilities. The CIR ranges from 58.8% to 100% with photodocumentation varying from 6% to 81% [12–17]. The variability in photodocumentation within institutions adds an additional barrier to quality improvement in colonoscopy without an objective means to assess for cecal intubation. At present, there are few automated means to record and confirm colonoscopy completion to maintain quality indicator standards.

Within endoscopy, there has been a shift toward implementing artificial intelligence (AI) to improve endoscopy with enhanced diagnostics. There have been prior attempts at using artificial intelligence to detect cecal intubation without machine learning, including edge detection by geometric shapes, intensity change and saturation [18, 19]. However, artificial intelligence research in endoscopy has been accelerated by machine learning. In particular, these techniques have been best described in the computer-assisted detection of polyps (CADe), and histologic prediction of diminutive polyps (CADx) [20–25]. However, there have also been some studies implementing AI into quality indicators, including bowel preparation calculation, withdrawal time, and natural language processing in automating ADR calculation [26–33]. Certain studies have automated the detection of cecal intubation with artificial intelligence, but they have not assessed these algorithms in suboptimal conditions [33]. There have been no studies evaluating colonoscopy completion under suboptimal conditions, including variable bowel preparation. In this study, we develop a deep convolutional neural network capable of detecting the presence of the appendiceal orifice as a marker of cecal intubation and colonoscopy completion with variable bowel preparation.

Methods

This was a retrospective study using high-definition videos of colonoscopy procedures conducted at St. Michael's Hospital in Toronto, Canada, from 2015 to 2018. This study was approved by the St. Michael's Hospital Research Ethics Board (19-050).

Datasets and preprocessing

The image dataset was derived from videos of colonoscopy procedures recorded during previous interventional studies conducted at St. Michael's Hospital between 2015 and 2018 [34–36]. These videos did not have any patient identifiers and only images of bowel lumen were extracted. We screened 144 procedures from previous studies. Videos were included if the recorded colonoscopy was completed (beginning at the rectum, reaching the cecum, and withdrawn back to the rectum). Videos were excluded from the study if the recorded colonoscopy was incomplete (i. e. cecum was not reached), or if the video itself was incomplete (i. e. recording does not begin at the rectum, and/or does not show withdrawal back to the rectum after cecal intubation). A total of 35 videos were included into this study. The videos were converted into images at 10 frames per second using Adobe Photoshop CC 2019 software (San Jose, California, United States). Images were then classified as either: (1) containing the appendiceal orifice (AO); or (2) not containing the AO (non-AO). These images were subclassified into Boston Bowel Preparation Scale (BBPS) scores (a commonly used evaluative tool for assessment of quality of bowel preparation) 0, 1, 2, and 3 for a segment of bowel. Score 0 was defined as a segment of mucosa not seen because of solid stool that could not be cleared. Score 1 was defined as a portion of mucosa seen, but other areas of the segment not well seen because of staining, residual stool or opaque liquid. Score 2 was defined as a minor amount of residual staining, small fragments of stool, or opaque liquid, with the mucosa well seen. Score 3 was defined as a well visualized segment of colon without any staining, fragments of stool or opaque liquid [37]. The classification process was conducted by expert gastroenterologists (i. e. > 1000 completed procedures). The identification of the appendiceal orifice was first located in the videos to ensure correct landmarking. Images that did not provide information regarding appendiceal orifice or bowel preparation, such as images of red outs, irrigation, fluid levels, biopsy forceps, blurry images, were left unclassified. The spectrum of non-AO images was included to simulate conditions encountered in colonoscopy.

In total, 13, 522 images were collected from the videos. There were 6852 images of AO and 6670 images in non-AO group. The AO images included full-view images of the AO, partial views of the appendiceal orifice, or cecal landmarks suggestive of the appendiceal orifice (triradiate cecal folds). Within the dataset, 6,559 AO images and 6,663 non- AO images were utilized for training, validation and testing: 11,900 images for training and validation and 1,322 images for testing. We ensured that the proportion of AO to non-AO images was consistent at each of the training, validation and testing phases. We additionally rescaled all images to a size of 224 × 224 pixels. In addition, to allow for greater generalizability, we applied several data augmentation strategies to the training data. These augmentations included: resized cropping, horizontal and/or vertical flips, random rotation up to 30 degrees, and random affine transforms up to a factor of 10.

Dense convolutional neural networks

We used DenseNet, a dense convolutional neural network architecture, that was pre-trained on approximately 1.2 million images with SIFT transforms from the ImageNet dataset as the backbone of our model [38, 39]. The DenseNet backbone connects each layer to every other layer in a feed-forward fashion. It has several advantages including stronger feature propagation, feature reuse, and fewer parameters, ultimately leading to a smaller model size. In our implementation, we adopted the DenseNet169 model architecture, but replaced the last layer with our customized classifier for appendiceal orifice detection [38, 40]. All experiments were implemented using Pytorch and scikit-learn libraries.

Training and testing

We used a batch size of 128 for both the training and validation datasets. We used an Adam optimizer with an initial learning rate of 3×10^{-4} and a scheduler to decay the learning rate of

each parameter group by 0.1 every 7 epochs. We optimized our model by using cross entropy loss as the criterion for the task, which combines a log softmax operation with a negative log likelihood loss. For validation of the algorithm, we took all the validation examples available to us and cross-verified that the minimization of the loss function was improving in each training epoch.

Outcomes and statistical analysis

The primary outcome of the study was to evaluate the operating characteristics of a deep convolutional neural network trained to detect AO vs non-AO and to assess the performance characteristics of the algorithm with variable bowel preparation scores [37]. The operating characteristics of interest include the detection rate of AO and non-AO along with the overall accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the model. The F1 score of the model was also calculated as a metric to balance precision and recall of the deep convolutional neural network.

Results

Dataset characteristics

There were a total of 13,513 images (6847 AO images and 6666 non-AO images) extracted from 35 colonoscopy videos. In terms of AO images, there were no additional findings identified in the images, including no diverticula, polyp, or vascular lesions. With respect to BBPS scores for AO images, there were 0 (0.0 %) BBPS 0 images, 2378 (34.7 %) BBPS 1 images, 3924 (57.3 %) BBPS 2 images, and 545 (8.0 %) BBPS 3 images. Within these images, 2378 (34.7 %) had inadequate BBPS scores (< 2), while 4469 (65.3 %) images had adequate BBPS scores (≥ 2). Within the non-AO images, there were 5153 images (77.3 %) that were assigned BBPS scores, and 1513 BBPS unclassifiable (22.7 %) images. There were 133 additional findings (2.6 %) within the images assigned BBPS scores, all of which were polyps. In terms of BBPS unclassifiable images, there were 1023 blurry images (67.6 %), 249 (67.6 %)images of fluid levels or irrigation (67.6 %), 34 images of instrumentation (2.2 %), and 207 images (13.7 %) of redouts. Among the images assigned BBPS scores, there were 0 (0.0 %) BBPS 0 images, 647 (9.7 %) BBPS 1 images, 3103 (46.6 %%) BBPS 2 images, and 1403 (21.1 %) BBPS 3 images. There were 647 images (12.6 %) of inadequate BBPS scores (< 2) and 4506 images (87.4 %) with adequate BBPS scores (≥ 2) (▶ **Table 1**).

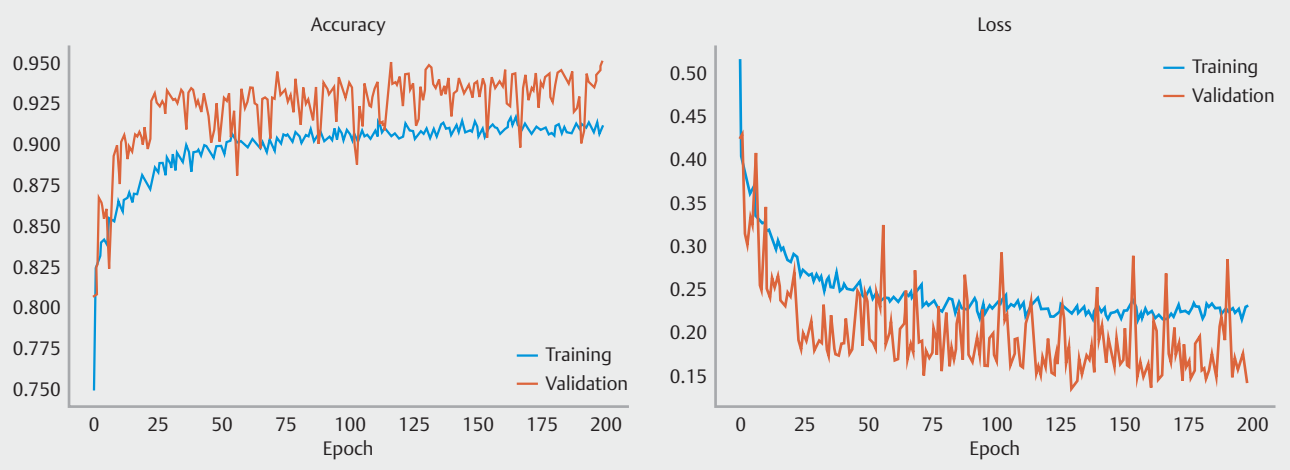
AO and non-AO detection

A total of 1,322 images were used for testing, composed of 656 AO (50.0 %) and 666 non-AO (50.0 %) images. The test set was representative of the proportion of the classes during training and validation. The proposed model is able to correctly classify the appendiceal orifice and non-appendiceal orifice with an overall accuracy of 94 % on the test dataset demonstrating excellent discrimination between the two images classes (▶ **Fig. 1**). The AUROC for this neural network is 0.98 (▶ **Fig. 2**). The operating characteristics of the model can be found in ▶ **Table 2**. The F1 score of the model was 94.3 % (▶ **Table 2**).

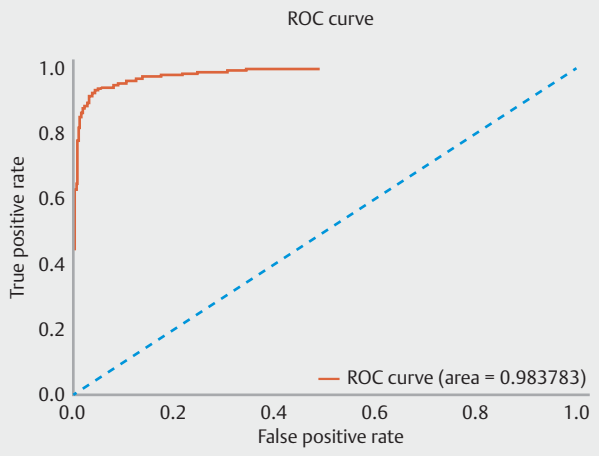
▶ **Table 1** Adequacy and subclassification of Boston Bowel Preparation Scores (BBPS) for AO and non-AO images in the dataset.

	BBPS 0	BBPS 1	BBPS 2	BBPS 3	BBPS unclassifiable	BBPS ≥ 2	BBPS ≤ 1
AO	0 (0.0 %)	2378 (34.7 %)	3924 (57.3 %)	545 (8.0 %)	0 (0.0 %)	4469 (65.3 %)	2378 (34.7 %)
Non-AO	0 (0.0 %)	647 (9.7 %)	3103 (46.6 %)	1403 (21.1 %)	1513 (22.7 %)	4506 (87.4 %)	647 (12.6 %)

AO, appendiceal orifice.



▶ **Fig. 1** Training and validation loss and accuracy curves for the deep convolutional neural network.



▶ **Fig. 2** Area under receiver operator characteristic curve for the deep convolutional neural network.

AO and non-AO Detection with variable bowel preparation

With regards to the AO test set characteristics, there were 0 (0.0%) BBPS 0 images, 355 (54.1%) BBPS 1 images, 255 (38.9%) BBPS 2 images, and 46 (7.0%) BBPS 3 images. There were 355 images (54.1%) with inadequate BBPS scores (<2) and 301 images (45.9%) with adequate BBPS scores (≥2). In the non-AO test group, there were 0 (0.0%) BBPS 0 images, 81 (12.2%) BBPS 1 images, 315 (47.3%) BBPS 2 images, 116 (17.4%) BBPS 3 images, and 154 (23.1%) BBPS unclassifiable images. There were 81 images (15.8%) of inadequate BBPS scores (<2) and 431 images (84.2%) with adequate BBPS scores (≥2) (▶ **Table 3**). The performance of the algorithm with variable bowel preparation in the test set for AO detection was 97.5%, 95.3% and 95.7% in BBPS 1, 2, and 3, respectively. When stratified for inadequate (BBPS<2) and adequate (BBPS ≥2) bowel preparation, AO detection was 97.5% and 95.4%, respectively. Likewise, non-AO detection for BBPS 1, 2, 3, and unclassifiable was 84.0%, 89.8%, 98.3%, and 96.1%, respectively. In terms of inadequate (BBPS<2) and adequate (BBPS ≥ 2) bowel preparation, non-AO detection was 84.0% and 92.1%, respectively (▶ **Table 4**). The model characteristics for BBPS1, BBPS2, and BBPS3 images, along with adequacy of bowel preparation can be found in ▶ **Table 5**.

Discussion

We present a deep convolutional neural network with an accuracy of 94% and an area under the receiver operating curve of 0.98 in discriminating images of the AO from those that do not depict the AO, with variable bowel preparation. The algorithm had overall excellent operating characteristics in sensitivity, specificity, positive predictive value and negative predictive.

▶ **Table 2** Detection characteristics of deep convolutional neural network in AO and Non-AO images.

	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score
Model operating characteristics	96.5%	92.0%	92.3%	96.4%	94.3%
	True positive	True negative	False positive	False negative	
Model test set	633	613	53	23	

AO, appendiceal orifice.

▶ **Table 3** Adequacy and subclassification of Boston Bowel Preparation Scores (BBPS) for AO and non-AO images in test set.

	BBPS 0	BBPS 1	BBPS 2	BBPS 3	BBPS unclassifiable	BBPS ≥ 2	BBPS ≤ 1
AO	0 (0.0%)	355 (54.1%)	255 (38.9%)	46 (7.0%)	0 (0.0%)	355 (45.9%)	301 (54.1%)
Non-AO	0 (0.0%)	81 (12.2%)	315 (47.3%)	116 (17.4%)	154 (23.1%)	431 (84.2%)	81 (15.8%)

AO, appendiceal orifice.

▶ **Table 4** Performance of deep convolutional neural network with varying Boston Bowel Preparation Scores (BBPS) in AO and non-AO images.

	BBPS 0	BBPS 1	BBPS 2	BBPS 3	BBPS unclassifiable	BBPS ≥ 2	BBPS ≤ 1
AO	N/A	97.5%	95.3%	95.7%	N/A	95.4%	97.5%
Non-AO	N/A	84.0%	89.8%	98.3%	96.1%	92.1%	84.0%

AO, appendiceal orifice.

▶ **Table 5** Detection characteristics of deep convolutional neural network in AO and Non-AO images stratified by Boston Bowel Preparation Scale (BBPS) scores.

	BBPS 1	BBPS 2	BBPS 3	Unclassifiable	BBPS ≥ 2	BBPS ≤ 1
False negative	9	12	2	–	14	9
False positive	13	32	2	6	34	13
True positive	346	243	44	–	287	346
True negative	68	283	114	148	397	68
Sensitivity	97.5 %	95.3 %	95.7 %	–	95.4 %	97.5 %
Specificity	84.0 %	89.8 %	98.3 %	96.1 %	92.1 %	84.0 %
PPV	96.4 %	88.4 %	95.7 %	–	89.4 %	96.4 %
NPV	88.3 %	95.9 %	98.3 %	100.0 %	96.6 %	88.3 %

AO, appendiceal orifice; PPV, positive predictive value; NPV, negative predictive value.

When assessing bowel preparation, AO detection was > 95 % irrespective of BBPS score and adequacy of bowel preparation. However, non-AO detection progressively improved (from 84.0 % to 98.3 %) with BBPS score and was superior with adequate (92.1 %) compared to inadequate (84.0 %) bowel preparation.

The improving operating characteristics from non-AO BBPS 1 to BBPS 3 can be attributed to a number of factors. For example, the clear visualization of the lack of cecal landmarks is more difficult with worsening bowel preparation because of increased background noise. In a review of the false-positive images (non-AO interpreted as AO) in BBPS 1 and BBPS 2 classes, the majority of the images had some feature that could be misinterpreted as part of the triradiate cecal folds. These features coupled with increasing noise from worsening bowel preparation led to the misclassification of these images. This was compounded by the fact that there were limited non-AO BBPS 1 images (9.7 %) for training. This is dissimilar to AO BBPS 3 images, in which the performance was excellent (96.7 %) despite a limited data set (8.0 %), as the presence of cecal landmarks can be clearly identified. Of note, there were no BBPS 0 images in the dataset. Although we do not expect our algorithm to have difficulty with this classification given that cecal landmarks would be completely obscured, it would be an important addition to simulate real-life colonoscopy conditions. To improve the false-positive rate and lower BBPS score classifications of non-AO images, a larger number of images with more variations are required for training, validation, and testing.

Although the distribution of BBPS scores was not equal, this did not bias our algorithm as it was trained for detection of the AO and non-AO, and not bowel preparation. Likewise, the fluctuation in the BBPS proportions in the test set compared to the overall dataset is attributed to random allocation that was conducted for AO and non-AO images, but not for bowel preparation. Despite the excellent accuracy and operating characteristics in AO detection across all bowel preparation classes, our system was only trained and tested on 35 videos with a relatively limited number of images in our dataset. The model, particularly lower BBPS non-AO images, can be improved with a larger balanced data set for training and testing to enhance variability and to improve generalizability. As the model was trained and validated with static images, this algorithm's application to recorded videos and real-time colonoscopy have not yet been determined and require further research.

In our review of the literature, existing applications of AI in gastroenterology have focused primarily on developing computer-assisted devices for detection and pathology prediction of polyps [21–25]. There is growing interest in the implementation of AI in assessing quality indicators in colonoscopy. In particular, algorithms have been used to assess for bowel preparation and withdrawal time [32, 33]. However, this is among the first machine learning algorithms created to assess for cecal intubation in the presence of variable bowel preparation. The algorithm adds to the pre-existing literature in synthesizing differing quality metric parameters and simulates greater real-world conditions in colonoscopy. Moreover, the robustness of the algorithm is demonstrated under variable and suboptimal conditions. Given that colonoscopy quality indicators occur within a spectrum, the ability for algorithms to perform under variable circumstances is particularly relevant. Although the validation of artificial intelligence algorithms in controlled environments is important, their impact may be greater under subpar circumstances. For example, greater benefit may be obtained from the detection of polyps in inadequate bowel preparation, or in lower-quality colonoscopy systems with worse spatial resolution. As such, machine learning algorithms should be evaluated in both optimal and suboptimal conditions to broaden the applicability of their use cases and to derive maximal benefit in imperfect circumstances.

The applications of AI pertaining to colonoscopy completion are significant. Although all major gastroenterology societies have thresholds for colonoscopy completion, there is considerable variability in cecal intubation and photodocumentation among hospitals and practitioners. Of concern, the rates of cecal intubation and photodocumentation among certain providers pale in comparison to standards set forth by multiple gastroenterology societies [12–17]. The maintenance of this quality metric is significant, as lower rates of colonoscopy comple-

tion are associated with higher rates of PCCRC [11]. Despite this, there are no formal auditing practices to ensure the maintenance of quality indicators in endoscopy as it is both cost and time-intensive. One common quality improvement initiative has demonstrated that providing intermittent feedback to clinicians regarding their cecal intubation rates through report cards can improve cecal intubation rates [41, 42]. Likewise, other studies have shown a possible association between time of day and worsening endoscopy quality with reductions in ADR and cecal intubation rates as a workday progresses, possibly related to practitioner fatigue [43–45]. As a result, implementing a computer-assisted device for detection of colonoscopy completion may provide a method of quality indicator feedback by facilitating automated documentation and objective detection of cecal intubation.

Conclusions

In summary, we successfully created an algorithm using a deep convolutional neural network with excellent accuracy for detection of the AO under variable bowel preparation. Moving forward, this algorithm requires a larger dataset for training, and implementation in real-time colonoscopy to elucidate its applications more clearly. Within the domain of quality indicators in colonoscopy, the synthesis of other AI quality metric algorithms in suboptimal conditions is necessary for future testing to derive greater benefit in improving and maintaining colonoscopy quality.

Competing interests

Rishad Khan has received research grants from AbbVie and Ferring Pharmaceuticals and research funding from Pendopharm.
Samir C. Grover has received research grants and personal fees from AbbVie and Ferring Pharmaceuticals, personal fees from Takeda, education grants from Janssen, and has equity in Volo Healthcare.
All other authors have no relevant disclosures.

References

[1] Ponich T, Enns R, Romagnuolo J et al. Canadian credentialing guidelines for esophagogastroduodenoscopy. Can J Gastroenterol 2008; 22: 349–354

[2] Cohen J, Pike IM. Defining and measuring quality in endoscopy. Gastrointest Endosc 2015; 81: 1–2

[3] Rembacken B, Hassan C, Riemann JF et al. Quality in screening colonoscopy: Position statement of the European Society of Gastrointestinal Endoscopy (ESGE). Endoscopy 2012; 44: 957–968

[4] Rees CJ, Thomas-Gibson S, Rutter MD et al. UK key performance indicators and quality assurance standards for colonoscopy. Gut 2016; 65: 1923–1929

[5] Kaminski MF, Thomas-Gibson S, Bugajski M et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy 2017; 49: 378–397

[6] Kaminski MF, Regula J, Kraszewska E et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010; 362: 1795–1803

[7] Lund M, Trads M, Njor SH et al. Quality indicators for screening colonoscopy and colonoscopist performance and the subsequent risk of interval colorectal cancer: A systematic review. JBI Database Syst Rev Implement Reports 2019; 17: 2265–2300

[8] Hilsden RJ, Dube C, Heitman SJ et al. The association of colonoscopy quality indicators with the detection of screen-relevant lesions, adverse events, and postcolonoscopy cancers in an asymptomatic Canadian colorectal cancer screening population. Gastrointest Endosc 2015; 82: 887–894

[9] Lebwohl B, Kastrinos F, Glick M et al. The impact of suboptimal bowel preparation on adenoma miss ratesa and the factors associated with early repeat colonoscopy. Gastrointest Endosc 2011; 73: 1207–1214

[10] Shaukat A, Rector TS, Church TR et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. Gastroenterology 2015; 149: 952–957

[11] Baxter NN, Warren JL, Barrett MJ et al. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol 2012; 30: 2664–2669

[12] Zorzi M, Senore C, Da Re F et al. Quality of colonoscopy in an organised colorectal cancer screening programme with immunochemical faecal occult blood test: The EQuIPE study (Evaluating Quality Indicators of the Performance of Endoscopy). Gut 2015; 64: 1389–1396

[13] Gonçalves AR, Ferreira C, Marques A et al. Assessment of quality in screening colonoscopy for colorectal cancer. Clin Exp Gastroenterol 2011; 4: 277–281

[14] Singh H, Kaita L, Taylor G et al. Practice and documentation of performance of colonoscopy in a central Canadian health region. Can J Gastroenterol Hepatol 2014; 28: 185–190

[15] Lee TJW, Rutter MD, Blanks RG et al. Colonoscopy quality measures: Experience from the NHS Bowel Cancer Screening Programme. Gut 2012; 61: 1050–1057

[16] De Jonge V, Sint Nicolaas J, Cahen DL et al. Quality evaluation of colonoscopy reporting and colonoscopy performance in daily clinical practice. Gastrointest Endosc 2012; 75: 98–106

[17] Lund M, Erichsen R, Valori R et al. Data quality and colonoscopy performance indicators in the prevalent round of a FIT-based colorectal cancer screening program. Scand J Gastroenterol 2019; 54: 471–477

[18] Wang Y, Tavanapong W, Wong J et al. Edge cross-section features for detection of appendiceal orifice appearance in colonoscopy videos. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2008; 2008: 3000–3003

[19] Cao Y, Liu D, Tavanapong W et al. Automatic classification of images with appendiceal orifice in colonoscopy videos. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2006; 1: 2349–2352

[20] Mori Y, Kudo S, East JE et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). American Society for Gastrointestinal Endoscopy 2020: doi:10.1016/j.gie.2020.03.3759

[21] Urban G, Tripathi P, Alkayali T et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018; 155: 1069–1078.e8

[22] Chen PJ, Lin MC, Lai MJ et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018; 154: 568–575

[23] Kudo S, Misawa M, Mori Y et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol 2019; 18: 1874–1881.e2

[24] Zachariah R, Samarasena J, Luba D et al. Prediction of polyp pathology using convolutional neural networks achieves “resect and discard” thresholds. Am J Gastroenterol 2020; 115: 138–144

[25] Repici A, Badalamenti M, Maselli R et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020; 159: 512–520.e7

[26] Nayor J, Borges LF, Goryachev S et al. Natural language processing accurately calculates adenoma and sessile serrated polyp detection rates. Dig Dis Sci 2018; 63: 1794–1800

[27] Lee JK, Jensen CD, Levin TR et al. Accurate identification of colonoscopy quality and polyp findings using natural language processing. J Clin Gastroenterol 2019; 53: E25–E30

[28] Imler TD, Morea J, Kahi C et al. Multi-center colonoscopy quality measurement utilizing natural language processing. Am J Gastroenterol 2015; 110: 543–552

[29] Raju GS, Lum PJ, Slack RS et al. Natural language processing as an alternative to manual reporting of colonoscopy quality metrics. Gastrointest Endosc 2015; 82: 512–519

[30] Imler TD, Morea J, Kahi C et al. Natural language processing accurately categorizes findings from colonoscopy and pathology reports. Clin Gastroenterol Hepatol 2013; 11: 689–694

[31] Mehrotra A, Dellon ES, Schoen RE et al. Applying a natural language processing tool to electronic health records to assess performance on colonoscopy quality measures. Gastrointest Endosc 2012; 75: 1233–1239.e14

[32] Zhou J, Wu L, Wan X et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). Gastrointest Endosc 2020; 91: 428–435.e2

[33] 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: 352–361

[34] Grover SC, Garg A, Scaffidi MA et al. Impact of a simulation training curriculum on technical and nontechnical skills in colonoscopy: a randomized trial. Gastrointest Endosc 2015; 82: 1072–1079

[35] Grover SC, Scaffidi MA, Khan R et al. Progressive learning in endoscopy simulation training improves clinical performance: a blinded randomized trial. Gastrointest Endosc 2017; 86: 881–889

[36] Walsh CM, Scaffidi MA, Khan R et al. Non-technical skills curriculum incorporating simulation-based training improves performance in colonoscopy among novice endoscopists: Randomized controlled trial. Dig Endosc 2020; 32: 940–948

[37] Lai EJ, Calderwood AH, Doros G et al. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009; 69: 620–625

[38] Huang G, Liu Z, Van Der Maaten L et al. Densely connected convolutional networks. Proc – 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2017: 2261–2269

[39] Deng J, Dong W, Socher R et al. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition 2009: 248–255

[40] Tan C, Sun F, Kong T et al. A survey on deep transfer learning. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2018; 11141: 270–279

[41] Uche-Anya EN, Brown JJ, Asumeng C et al. Impact of a citywide benchmarking intervention on colonoscopy quality performance. Dig Dis Sci 2020; 65: 2534–2541

[42] Kahi CJ, Ballard D, Shah AS et al. Impact of a quarterly report card on colonoscopy quality measures. Gastrointest Endosc 2013; 77: 925–931

[43] Wells CD, Heigh RI, Sharma VK et al. Comparison of morning versus afternoon cecal intubation rates. BMC Gastroenterol 2007; 7: 1–5

[44] Harewood GC, Chrysostomou K, Himy N et al. Impact of operator fatigue on endoscopy performance: implications for procedure scheduling. Dig Dis Sci 2009; 54: 1656–1661

[45] Almadi MA, Sewitch M, Barkun AN et al. Adenoma detection rates decline with increasing procedural hours in an endoscopist's workload. Can J Gastroenterol Hepatol 2015; 29: 304–308