A Gastrointestinal Endoscopy Quality Control System Incorporated With Deep Learning Improved Endoscopist Performance in a Pretest and Post-Test Trial

Liwen Yao, MD[1,2,3], Jun Liu, MA[1,2,3], Lianlian Wu, MD[1,2,3], Lihui Zhang, MD[1,2,3], Xiao Hu, MA[4], Jinzhu Liu, MA[4], Zihua Lu, MD[1,2,3], Dexin Gong, MD[1,2,3], Ping An, MD[1,2,3], Jun Zhang, MD[1,2,3], Guiying Hu, MD[1,2,3], Di Chen, MD[1,2,3], Renquan Luo, MD[1,2,3], Shan Hu, PhD[4], Yanning Yang, MD[5] and Honggang Yu, MD[1,2,3]

INTRODUCTION: **Gastrointestinal endoscopic quality is operator-dependent. To ensure the endoscopy quality, we constructed an endoscopic audit and feedback system named Endo.Adm and evaluated its effect in a form of pretest and posttest trial.**

METHODS: **Endo.Adm system was developed using Python and Deep Convolutional Neural Ne2rk models. Sixteen endoscopists were recruited from Renmin Hospital of Wuhan University and were randomly assigned to undergo feedback of Endo.Adm or not (8 for the feedback group and 8 for the control group). The feedback group received weekly quality report cards which were automatically generated by Endo.Adm. We then compared the adenoma detection rate (ADR) and gastric precancerous conditions detection rate between baseline and postintervention phase for endoscopists in each group to evaluate the impact of Endo.Adm feedback. In total, 1,191 colonoscopies and 3,515 gastroscopies were included for analysis.**

RESULTS: **ADR was increased after Endo.Adm feedback (10.8%–20.3%, $P < 0.01$, <odds ratio (OR) 2.13, 95% confidence interval (CI) 1.317–3.447), and endoscopists' ADR without feedback remained nearly unchanged (10.8%–10.9%, $P = 0.57$, OR 1.086, 95% CI 0.814–1.447). Gastric precancerous conditions detection rate increased in the feedback group (3%–7%, $P < 0.01$, OR 1.866, 95% CI 1.399–2.489) while no improvement was observed in the control group (3.9%–3.5%, $P = 0.489$, OR 0.856, 95% CI 0.550–1.332).**

DISCUSSION: **Endo.Adm feedback contributed to multifaceted gastrointestinal endoscopic quality improvement. This system is practical to implement and may serve as a standard model for quality improvement in routine work (http://www.chictr.org.cn/, ChiCTR1900024153).**

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/A626, http://links.lww.com/CTG/A627, http://links.lww.com/CTG/A628, http://links.lww.com/CTG/A629, http://links.lww.com/CTG/A630, http://links.lww.com/CTG/A631, http://links.lww.com/CTG/A632, http://links.lww.com/CTG/A633, http://links.lww.com/CTG/A634, http://links.lww.com/CTG/A635

Clinical and Translational Gastroenterology 2021;12:e00366. https://doi.org/10.14309/ctg.0000000000000366

INTRODUCTION

Hundreds of millions endoscopic procedures are performed every year worldwide (1,2). Endoscopy is a key investigation for the diagnosis of gastrointestinal (GI) lesions and a powerful tool for its treatment. High-quality endoscopy delivers better health outcomes and better patient experience (3). However, endoscopy quality is known to vary largely among endoscopists and among units (4,5).

Audit and feedback as an intervention provide health professionals with a summary of their performance over a period (6). Studies have proven that auditing quality indicators and timely feedback can effectively improve endoscopy-related outcomes (7,8). Many societies and guidelines advocate that it is vital to have continuous monitoring of performance and feedback (9,10). Endoscopy quality auditing can be time-consuming, add labor

[1]Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China; [2]Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China; [3]Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Renmin Hospital of Wuhan University, Wuhan, China; [4]Wuhan ENDOANGEL Medical Technology Company, Wuhan, China; [5]Department of Ophthalmology, Renmin Hospital of Wuhan University, Wuhan, China. **Correspondence:** Honggang Yu. E-mail: yuhonggang1968@163.com. Yanning Yang. E-mail: ophyyn@163.com.
Received December 14, 2020; accepted April 28, 2021; published online June 15, 2021

© 2021 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology

cost, and be prone to error and bias. Existing endoscopic quality control systems involved manually report data, making it difficult to achieve timely and objective feedback (11). A fully automated, standardized electronic system to collect and calculate quality indicators for both countries and endoscopy center leaders is on highly demanded.

In recent years, there has been a tremendous advance of artificial intelligence (AI) in medical field (12). In our previous work, we have developed gastroscopy blind spot and colonoscopy withdrawal speed monitoring systems and achieved remarkable endoscopy quality improvement (13,14). Based on this work, we designed an endoscopic quality data statistics system coupled with previous Deep Convolutional Neural Ne2rk (DCNN) models. This system (Endo.Adm) can make endoscopic quality feedback by analyzing daily updated endoscopic data. Through Endo.Adm, endoscopists can be provided with their operationally relevant performance statistics—such as examination time, types of detected polyp/adenoma, and unobserved stomach sites—to make continuous quality improvement. We hypothesized that the system could improve the quality of endoscopy procedures and the lesions detection.

METHOD AND MATERIALS

Development of the Endo.Adm system

The Endo.Adm system was designed to statistics quality indicators as follow: colonoscopy withdrawal time, cecal intubation rate (CIR), adequate bowel preparation rate, polyp detection rate (PDR), adenoma detection rate (ADR), gastroscopy photo-documented stomach site, gastroscopic inspection time, and gastric precancerous condition (GPC) detection rate. First, Endo.Adm collects patient information, endoscopy images from endoscopy system, and endoscopy-associated pathology reports from a pathology database. Second, DCNN models execute the endoscopy images identification. Third, Endo.Adm calculates each quality indicator and performs data presentation (Figure 1).

We did 4 main experiments. First, we developed and tested 3 DCNN models on images. Second, we constructed a performance measurement system, Endo.Adm, for patient demographics, and endoscopic and pathological information statistics with DCNN. Third, we tested Endo.Adm accuracy in 218 colonoscopy and 96 gastroscopy cases. Fourth, we evaluated the effect of Endo.Adm on colonoscopy and gastroscopy quality in routine practice. Details of the DCNN models were described in Supplementary Digital Content 10 (see Supplementary Material, http://links.lww.com/CTG/A635) P1-4.

System design and function introduction

Endo.Adm data exchange was based on Digital Imaging and Communications in Medicine (DICOM), an international standard to transmit, store, retrieve, print, process, and display medical imaging information (15). DICOM files can be exchanged between 2 entities that are capable of receiving image and patient data in DICOM format.

Based on DICOM standards, Endo.Adm analyzes reviewable electronic medical records and images accurately and efficiently. The system, on department intranet, allows endoscopists to log on and explore their performance analysis. Endo.Adm framework was written in Python and coupled with 3 DCNN models. The statistics are completely automatic and do not involve statistical or data management staff to download or analyze data. The fully automatic feature of Endo.Adm provides solutions to integrate feedback into routine practices.

Endo.Adm consists of 3 main modules: data extraction, data staging, and data presenting.

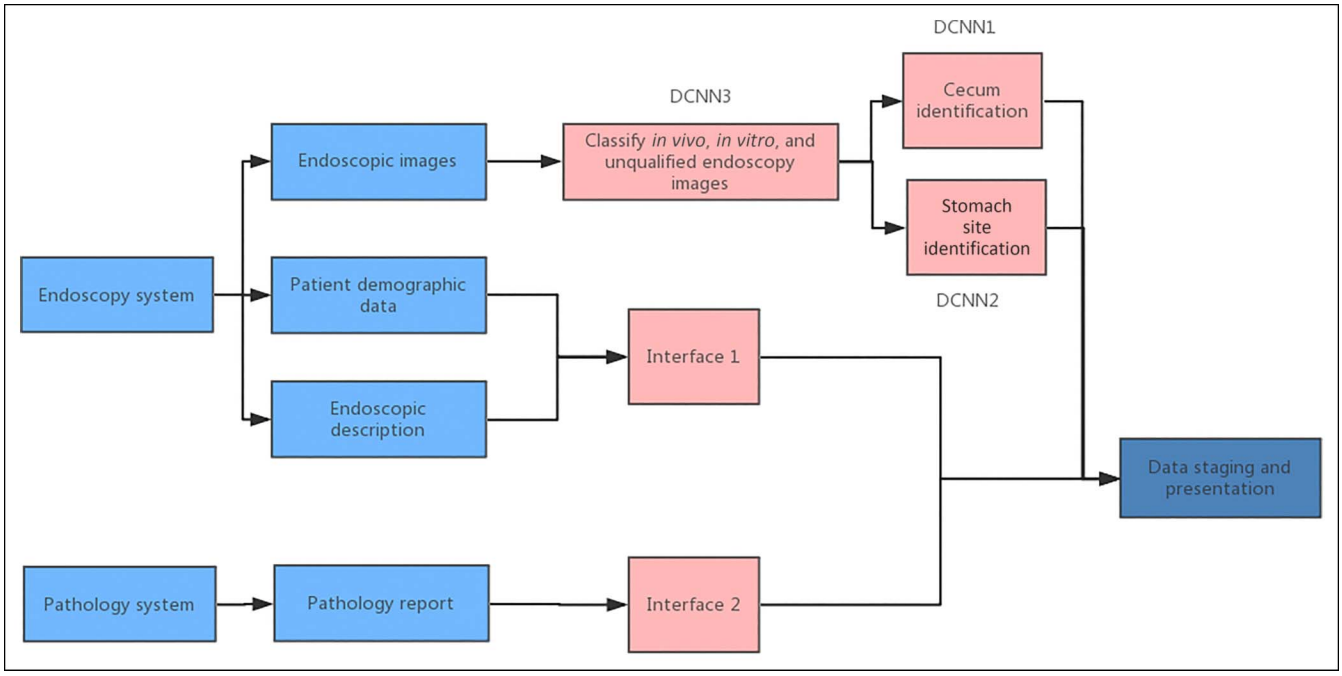

Figure 1. Technical flowchart of Endo.Adm system. Three DCNN models (DCNN1, DCNN2, and DCNN3) and 2 data interfaces were used for constructing the Endo.Adm system. DCNN, Deep Convolutional Neural Ne2rks.

Data and image extraction

Data extraction was realized by 2 customized interfaces which accessed to endoscopy and pathology information systems, respectively.

Endoscopy data. Our institution's endoscopy information system (Medcon, China, Qingdao) provides a data warehousing platform. The patient endoscopy report updates daily to a structured query language server database through Medcon system. Endo.Adm extracts demographic information (medical record number, name, age, sex, etc), endoscopic data (indications, endoscopic findings, instruments, operator, bowel preparation, biopsy, endoscopic findings, etc), and endoscopy images from the data warehouse.

Pathology data. In our hospital, the medical record numbers are sometimes absent from the pathology database. To link the endoscopic reports with corresponding pathology reports, name, age, sex, sample submission date, and the content of pathology report were applied. The pathology reports must contain specific keywords to identify it as a colonoscopy or gastroscopy sample. Colonoscopy-specific pathology keywords included colon, rectum, cecum, ileocecal valve, and colonoscopy. Gastroscopy-specific keywords contained esophagus, antrum, angular, gastric body, pylorus, fundus, duodenum, and gastroscopy. If the procedure date does not match the sample submission date, Endo.Adm matches them within 2 days of the sample's submission date. Our pathology department accepts samples submitted within 24 hours of resection, and a part of the samples may be submitted the day after the endoscopic procedure. Adenoma detection was defined if the pathology report associated with a colonoscopy report contained the following abstracted pathology fields: adenoma, adenomatous polyps, sessile serrated polyp, traditional serrated adenoma, serrated polyp, dysplasia, and adenocarcinoma. GPC detection rate was defined as the following fields: intestinal metaplasia, atrophic gastritis, and dysplasia. Regular expressions were applied to avoid false positives caused by phrases such as "no adenoma detection." Endo.Adm combines GPC detection rates and ADR with pathology extraction interfaces.

Data and images staging

There are 7 steps to the data processing protocol.

Step 1: Data on demographics and endoscopic records were extracted from the Endoscopy Information System into the Staging Database.

Step 2: DICOM images from each procedure were extracted and converted into .joint Picture Group (JPG) format. Medical record number, examination item, and image generation time contained in DICOM images were used for naming .JPG images.

Step 3: .JPG images were applied to DCNN3 for screening out *in vitro* and unqualified images.

Step 4: The eligible filtered images were applied to DCNN1 (colonoscopy) or DCNN2 (gastroscopy) according to their examination item. After DCNN processes, corresponding results were recorded into the staging database.

Step 5: Pathology results were stored in the pathological system database. Endo.Adm accessed the Structured Query Languag pathology database through interface and extracted pathology report for each endoscopy procedure.

Step 6: Each procedure report is presented as raw data along with the associated pathology results and DCNN processed results.

Step 7: Quality indicators were calculated according to the raw data. Specific calculation methods used are shown in Table 1.

Table 1. Endo.Adm function and test results

Function modules	Statistics method	Data extract method	Test method	Accuracy in test set
Colonoscopy				
Withdrawal time analysis	Time of last *in vivo*—cecum image	DCNN1 + DCNN3	Cecum images + endoscopist judgment	91.3%
Cecal intubation rate analysis	Cases identified by DCNN1/(total procedures–exclusions) × 100%	DCNN1 + DCNN3 + data extraction	Cases images + endoscopist judgment	96.3%
Adequate bowel preparation rate analysis	Patients with BBPS[2] scores of 2 or 3 for all colon segments/total procedures × 100%	Data extraction	Manual checking	100%
Polyp detection rate analysis	Polyps detected/total procedures × 100%	Data extraction	Manual checking	100%
Gastroscopy				
Photodocumented stomach site analysis	Sum of photodocumented stomach sites.	DCNN2 + DCNN3	Stomach site images + endoscopist judgment	91%
Inspection time analysis	Time of last–first *in vivo* images	DCNN3	*In vivo* image + endoscopist judgment	99%
Pathology				
Pathology report linking analysis	Cases with correct pathology results linking/ total procedures × 100%	Data extraction	Manual checking	97.8%

BBPS, Boston Bowel Preparation Scale; DCNN, Deep Convolutional Neural Ne2rks.

Data presentation

The data presentation was in 3 main functional interfaces: colonoscopic quality analysis, gastroscopic quality analysis, and quality report generation.

Colonoscopic quality analysis interface. CIR, withdrawal time, PDR, ADR, adequate bowel preparation rate, detected colorectal cancer, and colorectal cancer detection rate were shown. Endoscopists are capable to check their incomplete colonoscopies and make improvements (see Supplementary Figure 1, Supplementary Digital Content 1, http://links.lww.com/CTG/A626).

Gastroscopic quality analysis interface. Esophagogastroduodenoscopy inspection time, photodocumented stomach site, detected early gastric cancer (EGC) and gastric cancer (GC), detection rate of EGC and GC, and GPC detection rate were shown (see Supplementary Figure 2, Supplementary Digital Content 2, http://links.lww.com/CTG/A627). Endo.Adm also provides link for endoscopists to view the results of photodocumented stomach sites, showing detected and missed for each site. Missed photodocumentation rates are also shown, as illustrated in Supplementary Digital Content 3 (see Supplementary Figure 3, http://links.lww.com/CTG/A628).

Quality report generation interface. Endo.Adm provides endoscopists with automatically generated quality report for any period in the doctor performance module. The quality report shows ADR, PDR, withdrawal time, CIR for colonoscopy and GPC detection rate, inspection time, and photodocumented part count for gastroscopy in a table. The quality report also shows ADR and GPC detection rate change tendencies in the form of line chart for a 6-month period (see Supplementary Figure 4, Supplementary Digital Content 4, http://links.lww.com/CTG/A629).

Testing the system in endoscopy cases

Different from images, photodocumentation for each patient was stored as a document listing images in chronological order. We tested the accuracy of Endo.Adm in case test set by selecting 218 colonoscopies and 96 gastroscopies from March 4, 2019, through October 3, 2019, independent from the training image test sets, for manual validation. The case selection period was isolated from our training sets. The same validation set was selected (96 gastroscopies and 218 colonoscopies) for the pathology linking function testing. Testing method and accuracy were listed in Table 1.

Prospectively validation of Endo.Adm effectiveness in clinical settings

Setting and period. The testing was performed in an endoscopy center at a university hospital in the center of China from June 2019 to September 2019. The study was conducted in 3 phases: (i) baseline phase (phase 1, April 20, 2019, to May 31, 2019), (ii) informing and randomization phase, and (iii) postintervention phase (phase 2, July 1, 2019, to August 20, 2019).

Study endoscopists and procedures. The institutional review board at Renmin Hospital of Wuhan University approved this study. Since this project examined endoscopy quality in routine clinical practice, the review board only required an informed consent for endoscopists as reported by previous studies (16,17). All endoscopists who routinely perform endoscopy in our endoscopy center agreed to be included before randomization. Endoscopists who neither presented for both parts (phase 1 and phase 2) of the study nor performed endoscopy (<10 procedures/phase) were excluded. Endoscopists with less than 1 year of endoscopy experience were also excluded because of potential unstable performance.

Instruments used in this study included gastroscopes and colonoscopes from 2 vendors (Olympus Optical Company Tokyo, Japan, and Fujifilm Company, Kanagawa, Japan). The models of the scopes contained 590, 600 from Fujifilm and 260, 290 from Olympus. Procedures performed by enrolled endoscopists from phase 1 and phase 2 were included. Phase 2 procedures were enrolled the day after endoscopists received their quality report. Colonoscopy procedures were excluded from our study if the patients had polyposis syndromes, lumen obstruction, a history of colorectal surgery, and inflammatory bowel disease. For gastroscopy, the exclusion criteria were a history of gastric surgery and obstruction.

Randomization and procedure. In June 2019, 12 eligible endoscopists performed both colonoscopy and gastroscopy, and 5 eligible endoscopists performed gastroscopy-only were randomly assigned to control and feedback groups (in approximate 1:1 ratio). Randomization was computer-generated and stratified according to their frequency of procedures. Both feedback and control group endoscopists were informed of the standard quality indicators requirements and the corresponding references during informed consent. In addition to the quality requirements, endoscopists randomized to the feedback group received customized quality reports feedback from Endo.Adm weekly (see Supplementary Figure 9, Supplementary Digital Content 9, http://links.lww.com/CTG/A634). Feedback endoscopists could also access to Endo.Adm for more detailed quality statistics.

The pathology result of each procedure was automatically linked by Endo.Adm and manually rechecked by researchers.

Definitions and endpoints. For colonoscopy procedures, ADR was defined as the proportion of colonoscopy in whom at least 1 adenoma was identified (including traditional serrated adenoma, sessile serrated adenomas, or carcinoma) (18). Advanced adenomas were defined as adenomas that were either ≥10 mm in size, or adenomas with histopathology of tubulovillous, villous, adenocarcinoma, or high-grade dysplasia. All patients were followed up to November 25, 2019, in both groups. Polyps not removed or retrieved were categorized as non-neoplastic and not taken into account when calculating ADR. Withdrawal was considered to start after the cecum was photodocumented. For colonoscopy, primary endpoint was ADR and predefined secondary endpoints were advanced ADR, withdrawal time, PDR, and CIR.

For gastroscopy, primary endpoint was defined as the GPC detection rate, and predefined secondary endpoints were inspection time and photodocumentation completeness. Diagnoses of ADR and GPC detection rate were based on the pathology report's description. Gastroscopic inspection time was considered as the time from the first *in vivo* images to the last *in vivo* images. In 2015, the European Society of GI Endoscopy (ESGE) systematically investigated available evidence and proposed that the entire stomach should be fully mapped during gastroscopy (19). In our current study, photodocumentation completeness was defined as the number of stomach sites being

photodocumented during a gastroscopy (20). GPCs contained dysplasia, gastric atrophy, and intestinal metaplasia (21). EGC was defined as gastric adenocarcinomas confined to the mucosa and submucosa of the stomach with or without regional lymph node metastases (22).

Power estimates and statistical analyses. The primary endpoint was to investigate the effect of Endo.Adm quality improvement program on ADR and GPC detection rate. The baseline ADR was 10.8%, and we estimated statistical power with a goal of achieving at least 80% power at the 5% significance level to detect an increase in ADR from 10.8% to 20%. Cluster randomized designed was applied for sample size calculation, with 6 clusters in each group. Group sample size was 262 in each phase of the study. For GPC detection rate, 592 patients in each phase were required to demonstrate an increase from 4% to 8% with a 5% significance level and 80% power with 8 clusters in each group (PASS 15, Tennessee).

Baseline characteristics, withdrawal time, and gastroscopy inspection time between study groups were compared using the χ^2 test for categorical variables and the Mann-Whitney *U* test for continuous variables.

In addition to routine descriptive summaries, analysis of ADR, PDR, advanced ADR, CIR, GPC detection rate, and other outcomes was used to create a generalized estimating equation model. Details of the model are illustrated in the Supplementary Digital Content 10 (see Supplementary Material, http://links.lww.com/CTG/A635) P4. A 2-sided *P* value of 0.05 was considered to be statistically significant. All analyses were performed using SPSS 20 (IBM, Chicago, IL). All authors had access to the study data and reviewed and approved the final manuscript.

RESULTS

The performance of Endo.Adm in endoscopy cases

Among the 96 gastroscopy examinations, Endo.Adm identified stomach sites with an average accuracy of 91% and a separate accuracy for each site ranging from 75% to 100% in the 96 gastroscopic cases (see Supplementary Figure 5, Supplementary Digital Content 5, http://links.lww.com/CTG/A630). No significant difference was found in stomach sites count between Endo.Adm and endoscopist-labeled results ($P = 0.47$). For esophagogastroduodenoscopy procedure timing, Endo.Adm correctly predicted the start time in 99% (95/96) cases and end time in 100% (96/96) cases. Only 1 *in vitro* image (an image close to the patient face) was misidentified *in vivo* by Endo.Adm. Among 218 colonoscopy cases, Endo.Adm had a 91.3% accuracy on withdrawal time calculation and 96.3% accuracy on cecal intubation prediction. Different performance in withdrawal time calculation and intubation prediction was due to the

Table 2. Colonoscopy baseline characteristics

	Control group			Feedback group		
	Phase 1 (N = 342)	Phase 2 (N = 367)	*P* value	Phase 1 (N = 251)	Phase 2 (N = 231)	*P* value
Endoscopist variables						
No. of endoscopists	6			5		
Age (SD)	35 (3)	39.2 (5.9)				0.01[a]
No. of yr since training, (SD)	5.7 (2.9)	8.2 (4.8)				0.4[a]
Patient variables						
Age, mean (SD)	49 (14.3)	47 (14)	0.6	47.5 (13.6)	48 (14.1)	0.4
Male, n (%)	199 (58.2)	209 (56.9)	0.7	140 (55.8)	126 (54.5)	0.9
Indications for colonosopy, n (%)			0.1			0.4
Screening	113 (33)	146 (39.8)		97 (38.7)	98 (42.4)	
Surveillance	14 (4.1)	18 (4.9)		10 (3.98)	13 (5.63)	
Diagnosis	215 (62.9)	203 (55.3)		144 (57.4)	120 (52)	
Recruitment, n (%)			0.4			0.2
Outpatient	260 (76)	267 (72.8)		193 (76.9)	165 (71.4)	
Inpatient	82 (24)	98 (26.7)		58 (23.1)	66 (28.6)	
Sedation during endoscopy, n (%)			0.4			1
Yes	136 (39.8)	134 (36.5)		104 (41.4)	96 (41.6)	
No	206 (60.2)	233 (63.5)		147 (58.6)	135 (58.4)	
Bowel preparation, n (%)			0.3			0.7
Inadequate (sum <6.0 or anyone <2.0), n(%)	48 (14)	61 (16.6)		30 (12.0)	24 (10.4)	
Adequate (sum ≥6.0 and everyone ≥2.0), n (%)	294 (86)	306 (83.4)		221 (88.0)	207 (89.6)	

[a]Represented the comparison between control and feedback groups.

misidentification of cecum images in 11 cecal intubated cases, although Endo.Adm correctly predicted intubation. Therefore, withdrawal time was incorrectly calculated. For pathology results linking, Endo.Adm correctly matched 330 in 345 cases and achieved an accuracy of 95.7%.

Outcome of practical testing

We enrolled and randomized 12 endoscopists performing both gastroscopy and colonoscopy and 5 endoscopists performing gastroscopy in an approximate 1:1 ratio, separately. One endoscopist performing both gastroscopy and colonoscopy in the feedback group was excluded because of the teaching task while examinations were performed by his trainees. Thus, all analyses comparing phase 1 and phase 2 include 16 endoscopists. Data analyses are based on 1,191 colonoscopies (593 in phase 1 and 598 in phase 2) and 3,515 gastroscopies performed by endoscopists throughout the trial phases (1,878 in phase 1 and 1,637 in phase 2). As each endoscopist in our department performs colonoscopy and gastroscopy routinely, colonoscopic and gastroscopic quality feedback was conducted simultaneously, which is similar to the practical environment. The baseline characteristics of colonoscopy and gastroscopy are displayed in Tables 2 and 4, respectively.

Colonoscopy

As shown in the Table 3, the mean ADR of endoscopists in the feedback group improved from 10.8% to 20.3% ($P < 0.01$, odds ratio [OR] 2.13, 95% confidence interval [CI] 1.317–3.447) while the ADR remained unchanged in the control group (10.8%–10.9%, $P = 0.57$, OR 1.086, 95% CI 0.814–1.447). Advanced ADR also improved significantly in the feedback group (4.4%–8.7%, $P = 0.04$, OR 0.96, 95% CI 0.939–0.982). PDR in feedback group endoscopists increased from 40.6% to 53.3% ($P < 0.01$; OR 1.761, 95% CI 1.030–5.237) while no increase was observed in the control group (Table 3). The colonoscopy withdrawal time among cases with no polyps significantly increased in the feedback group (4.9–5.9 minutes, $P < 0.01$). However, the CIR did not improve significantly after Endo.Adm audit and feedback (94.2%–96.6%, $P = 0.077$, OR 0.59, 95% CI 0.329–1.059) (see Supplementary, Supplementary Digital Content 10, http://links.lww.com/CTG/A635).

Table 3. Polyp and adenoma characteristics

	Control group			Feedback group		
Polyp subtype[a], n (%)	Phase 1 (N = 342)	Phase 2 (N = 367)	*P* value	Phase 1 (N = 251)	Phase 2 (N = 231)	*P* value
Polyps	120 (35.1)	127 (34.6)	0.94	102 (40.6)	123 (53.3)	<0.01
Adenomas	37 (10.8)	40 (10.9)	0.57	27 (10.8)	47 (20.3)	<0.01
Advanced	9 (2.6)	15 (4.1)	0.30	11 (4.4)	20 (8.7)	0.04
Nonadvanced	28 (8.2)	27 (7.4)	0.78	16 (6.4)	32 (13.9)	0.04
Hyperplastic and inflammatory	87 (25.4)	87 (23.7)	0.6	76 (30.3)	61 (26.4)	0.36
Polyp shape						
Polypoid	114 (33.3)	121 (33)	0.93	101 (40.2)	104 (45)	0.31
Nonpolypoid (flat)	18 (5.3)	13 (3.5)	0.28	4 (1.6)	12 (5.2)	0.04
Polyp location						
Right	45 (13.2)	55 (15)	<0.01	54 (21.5)	52 (22.5)	0.83
Left	97 (28.4)	91 (24.8)	0.31	73 (29.1)	91 (39.4)	0.02
Polyp size						
Size 1–5 mm	104 (30.4)	103 (28.1)	0.51	86 (34.3)	113 (48.9)	<0.01
Size 6–9 mm	26 (7.6)	33 (9)	0.59	8 (3.2)	8 (3.5)	1
Size 10+ mm	5 (1.5)	11 (3)	0.21	7 (2.8)	7 (3)	1
Adenoma shape						
Polypoid	31 (9.1)	37 (10.1)	0.7	25 (10)	43 (18.6)	<0.01
Nonpolypoid (flat)	6 (1.8)	3 (0.8)	0.33	2 (0.8)	4 (1.7)	0.43
Adenoma location						
Right	18 (5.3)	24 (6.5)	0.53	12 (4.8)	21 (9.1)	0.072
Left	27 (7.9)	27 (7.4)	0.89	19 (7.6)	35 (15.2)	<0.01
Adenoma size						
Size 1–5 mm	31 (9.1)	29 (7.9)	0.59	20 (8)	38 (16.4)	<0.01
Size 6–9 mm	10 (2.9)	14 (3.8)	0.54	4 (1.6)	6 (2.6)	0.53
Size 10+ mm	4 (1.2)	6 (1.6)	0.75	5 (2)	7 (3)	0.56

[a]Shown is the number (and percent) of patients with at least 1 polyp or adenoma of the given subtype.

Table 4. Gastroscopy baseline characteristics

	Control group			Feedback group		
	Phase 1 (N = 925)	Phase 2 (N = 913)	*P* value	Phase 1 (N = 953)	Phase 2 (N = 724)	*P* value
Endoscopist variables						
No. of endoscopists	8			8		
Age (SD)	34.2 (2.9)			36.8 (5.6)		0.3[a]
No. of yr since training, (SD)	4.87 (2.9)			5.88 (4.8)		0.4[a]
Patient variables						
Age, mean (SD)	46.5 (15.1)	45.7 (13.9)	0.07	47.1 (15.3)	46.4 (14.6)	0.1
Male, n (%)	451 (48.8)	445 (48.7)	1	468 (49.1)	366 (50.6)	0.9
Indications, n (%)			0.5			0.7
Epigastric pain	16 (1.7)	16 (1.8)		19 (2)	16 (2.2)	
Reflux	25 (2.7)	29 (3.2)		31 (3.2)	27 (3.7)	
Other abdominal pain	88 (9.5)	109 (11.9)		94 (9.9)	71 (9.8)	
Health examination	352 (38)	342 (37.5)		368 (38.6)	257 (35.5)	
Others	444 (48)	417 (45.7)		441 (46.3)	353 (48.8)	
Recruitment, n (%)			0.02			0.2
Inpatient	179 (19.4)	217 (23.8)		178 (18.7)	152 (21)	
Outpatient	746 (80.6)	696 (76.2)		775 (81.3)	572 (79)	
Sedation during endoscopy, n (%)			<0.01			<0.01
Yes	634 (68.5)	466 (51)		699 (73.4)	341 (47.1)	

[a]Represented the comparison between control and feedback groups.

Gastroscopy

The overall GPC detection rate was 3% and 3.9% for feedback and control groups, respectively, in phase 1. In phase 2, the feedback group's GPC detection rate increased from 3% to 7% ($P < 0.01$, OR 1.866, 95% CI 1.399–2.489), whereas the decrease of GPC detection rate in the control group was not improved (3.9%–3.5%, $P = 0.489$, OR 0.856, 95% CI 0.550–1.332) (Table 5). Photodocumentation to support the extent of examination has been endorsed by expert consensus and guidelines. The endoscopist should perform a complete examination that includes visualization of the esophagus, stomach (including retroflexion), and proximal duodenum and document it in the procedure report (23). In our current trial, photodocumentation completeness significantly improved in both control and feedback group endoscopists. Photodocumentation completeness generated by feedback group endoscopists significantly increased from 14.2 to 17.6 ($P < 0.01$) while in the control group was from 14.1 to 15.5 ($P < 0.01$). However, no significant improvement of inspection time was observed in either control or feedback groups ($P = 0.112$ and $P = 0.097$) (see Supplementary Table 2, Supplementary Digital Content 10, http://links.lww.com/CTG/A635).

Table 5. Gastric precancerous conditions detected on gastroscopy (as confirmed by histology) stratified by control and feedback endoscopists

	Control group			Feedback group		
	Phase 1 (N = 925)	Phase 2 (N = 913)	*P* value	Phase 1 (N = 953)	Phase 2 (N = 724)	*P* value
Advanced gastric cancer, n (%)	7 (0.8)	10 (1.1)	0.48	9 (0.9)	3 (0.4)	0.25
Early gastric cancer, n (%)	1 (0.1)	0 (0)	1	0 (0)	2 (0.3)	0.19
Gastric precancerous condition						
Dysplasia, n (%)	9 (1)	4 (0.4)	0.27	4 (0.4)	7 (1)	0.22
Gastric atrophy, n (%)	6 (0.6)	3 (0.3)	0.51	5 (0.6)	10 (1.4)	0.07
Intestinal metaplasia, n (%)	28 (3)	28 (3)	1	27 (3)	43 (5.9)	<0.01
No. of patients with gastric precancerous conditions, n (%)[a]	36 (3.9)	32 (3.5)	0.49	29 (3)	51 (7)	<0.01

[a]Some patients had >1 high-risk gastric lesions. The final analysis was a patient-based analysis where 1 positive outcome was registered.

DISCUSSION

In the current study, we constructed a GI endoscopic quality control system coupled with DCNN models. The performance of Endo.Adm was tested through endoscopy images and cases, with the end result proving the system to be reliable. We also evaluated its effect in a practical settings; Endo.Adm audit and feedback resulted in a comprehensive quality improvement.

Deep learning has played an important role in endoscopic quality control. In our previous work, we constructed real-time colonoscopy quality control system with timing withdrawal time, evaluating bowel preparation, and monitoring withdrawal speed based on DCNN models (14,24). Different from the aforementioned modalities, the aim of Endo.Adm is to provide more timely, extensive clinical performance summaries acquired directly after endoscopy or pathology reports are presented. Therefore, the center leader or administration officer can access updated data to achieve a real-time audit.

Endo.Adm accesses endoscopy report system data to calculate both colonoscopy and gastroscopy quality indicators automatically. Quality indicators generated by the first iteration of Endo.Adm are based on ESGE and the American Society of GI Endoscopy quality improvement initiatives that were available at the time of Endo.Adm development (9,10,25). Endo.Adm's framework was designed with good extensibility which is compatible with many DCNN models. Work of incorporating colonoscopy withdrawal speed, gastroscopy blind spot rate, and bowel preparation score assessed by AI is under way to allow a more comprehensive quality assessment.

Our clinical verification of Endo.Adm effectiveness was designed in the form of a pretest and posttest trial. The substantial increase in our feedback group ADR (from 10.8% to 20.3%), and GPC detection rate (from 3% to 7%), compared with the unchanged control group, suggested that the intervention may have positive effect on endoscopy quality. Through interviewing enrolled endoscopists after the study, we learned that the quality improvement was mainly due to the following reasons: continuous attention to quality issues, improvement of quality issues, and more focused on searching lesions. Our results support that quality of both colonoscopy and gastroscopy can be improved through Endo.Adm audit and feedback.

Patients with chronic atrophic gastritis or intestinal metaplasia should be considered at a higher risk of gastric adenocarcinoma (26). It is important to accurately identify patients with precancerous conditions. When comparing phase 2 with phase 1, the GPC detection rate improved from 3% to 7% (Table 5). The increasing number of detected GPCs lesions in the feedback group implied that Endo.Adm had a positive impact on GPC detection and contribute to the detection of GC at early stages. Moreover, it is also worth noted that the detected EGC increased from 0 to 2 in the feedback group which also illustrated the effectiveness of Endo.Adm.

After more than 10 years of exploration and standardization, audit and feedback has shown its effect on colonoscopy quality improvement (4,27,28). Imperiali et al. (29) validated the effectiveness of quality data audit and feedback on CIR and PDR. Kahi et al. (30) and Keswani et al. (31) applied quality report card for colonoscopy audit and feedback. Their interventions have significantly improved ADR. Abdul-Baki et al. (32) publicly reported endoscopists' quality data, and their initiative was associated with a significant improvement in the ADR. However, for the time being, endoscopy audit and feedback is mainly conducted using manual methods, which are not only laborious but also cumbersome because of interrogation of pathology databases and photodocumentation analysis. Manual performance measurement requires dedicated data statisticians or management staff, which is extremely time-consuming and costly. Thomas et al. (11) constructed the National Endoscopy Database for providing endoscopic quality audit and feedback. To calculate withdrawal time, CIR, and ADR statistics, endoscopists must manually record the cecum images and upload pathology results.

The advantage of AI statistics is that the statistical results are more objective, efficient, and automated. The use of a stopwatch or a foot pedal to record withdrawal time can theoretically achieve 100% statistical accuracy, but subjective interference is unavoidable. Moreover, for endoscopy centers with heavy workload, it is very difficult to maintain strict and accurate manual recording of withdrawal time over a long period. Endo.Adm applied the DCNN models to identify cecum and landmarks in the stomach to automatically complete the evaluation of the withdrawal time, CIR, and gastroscopic photodocumentation completeness. In addition, Endo.Adm also coupled an interface to link the pathological report with corresponding endoscopy procedure to calculate ADR and GPC detection rate. In conclusion, Endo.Adm achieved fully automated statistical analysis, eliminating the cost of manual statistics and errors caused by subjective factors.

There are limitations to Endo.Adm. Although it can provide sufficient quality control information, the recently reported deep learning-based bowel preparation and withdrawal speed evaluation have not been incorporated. However, work is under way to incorporate more quality indicators in the next version. Secondly, the statistics of photodocumentation completeness was based on still images instead of full-length videos. There may be some parts that have been observed, but there are no photodocumentation left. However, photodocumentation of all normal anatomical landmarks during gastroscopy has been proposed as key performance indicators by ESGE, and it might be an indirect quality indicator for careful inspection of the digestive lumen (10). Thirdly, although the CIR improved from 94.2% to 96.6%, there was no significant difference between phase 1 and phase 2 in the feedback group. The CIR not only is a quality indicator but also reveals the endoscopic skills of a physician. Reasons for failing to reach the cecum were diverse, include excessive loop formation, inadequate bowel preparation, and failure to traverse angulated, fixed, or strictured sigmoids (33). The Endo.Adm feedback can help endoscopists to pay more attention on their endoscopy quality but cannot compensate for technical defects. On the other hand, in our current study, high CIR was observed in feedback group phase 1, and thus, the CIR was not significantly improved.

As a software product, Endo.Adm is easy to be generalized among different hospitals. The data exchange is based on DICOM standards which are quite mature and have been used widely. Endo.Adm accesses an endoscopy information system and pathology database through customized interface modules according to the data structures. The interfaces are designed as an independent program, which minimizes the effect of changes in electronic records and ensures Endo.Adm's high flexibility. Therefore, this system could be installed in many different endoscopy centers and applied at a large scale.

In summary, we present a quality improvement system for GI endoscopy coupled with DCNN models. We verified the effect of this system in our routine practice, and our results indicate that multifaceted improvements in GI endoscopic quality can be

achieved with Endo.Adm system. In the future, the coverage and depth of Endo.Adm data will further increase. Endo.Adm will be generalized among different sites and establish benchmark for quality control indicators. The next version of Endo.Adm will provide (i) performance measurement for endoscopic ultrasound and endoscopic retrograde pancreatography; (ii) more comprehensive deep learning-based quality indicators; and (iii) building mobile quality data display software. Implementing Endo.Adm will facilitate a significant shift in endoscopic quality assurance. This concept may be adopted by the broader health care system, promoting progress in assessing performance, and ultimately improving the prognosis of endoscopy-related outcomes.

CONFLICTS OF INTEREST

Guarantor of the article: Honggang Yu, MD.

Specific author contributions: Liwen Yao and Jun Liu, and Yanning Yang and Honggang Yu contributed equally to this work. H.G.Y. and Y.Y.N.: conceived and designed the study; J.Z.L., S.H., and X.H.: trained and tested the models; G.Y.H., J.L., L.W.Y., L.L.W., and R.Q.L.: collected and reviewed images; Z.H.L., D.X.G., L.H.Z., D.H., and L.W.Y.: collected, collated, and analyzed the data; L.W.Y.: wrote the manuscript; J.Z. and P.A.: performed extensive editing of the manuscript; all authors reviewed and approved the final manuscript for submission. All authors were involved in data acquisition, general design of the trial, interpretation of the data, and critical revision of the manuscript.

Financial support: This work was partly supported by the grant from Project of Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision (grant no. 2018BCC337); Hubei Province Major Science and Technology Innovation Project (grant no. 2018-916-000-008); and the National Natural Science Foundation of China (grant no. 81770899). The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Potential competing interests: The authors declared no conflict of interest. The coauthor lists do not include endoscopists who participated in the clinical trial.

Study Highlights

WHAT IS KNOWN

- ✓ Audit and feedback was effective in colonoscopy quality improvement.
- ✓ No full-automatic performance measurement system has been constructed yet, especially using deep learning.

WHAT IS NEW HERE

- ✓ We constructed a performance measurement system on GI endoscopy with deep learning.
- ✓ The system was effective in improving GI endoscopy quality in a clinical trial.

TRANSLATIONAL IMPACT

- ✓ Deep learning-based quality statistics system has potential to improve the daily endoscopy quality.

REFERENCES

1. Tan G, Rao SS. Part I: How to ergonomically design a modern endoscopic suite. Tech Gastrointest Endosc 2019;21(3):133–9.
2. Wang LW, Lin H, Xin L, et al. Establishing a model to measure and predict the quality of gastrointestinal endoscopy. World J Gastroenterol 2019; 25(8):1024–30.
3. Rutter MD, Rees CJ. Quality in gastrointestinal endoscopy. Endoscopy 2014;46(6):526–8.
4. Rutter MD, Senore C, Bisschops R, et al. The European Society of Gastrointestinal Endoscopy quality improvement initiative: Developing performance measures. Endoscopy 2016;48(1):81–9.
5. Burr NE, Derbyshire E, Taylor J, et al. Variation in post-colonoscopy colorectal cancer across colonoscopy providers in English National Health Service: Population based cohort study. BMJ 2019;367:l6090.
6. Tinmouth J, Patel J, Hilsden RJ, et al. Audit and feedback interventions to improve endoscopist performance: Principles and effectiveness: Best practice & research. Clin Gastroenterol 2016;30(3):473–85.
7. Benson ME, Reichelderfer M, Said A, et al. Variation in colonoscopic technique and adenoma detection rates at an academic gastroenterology unit. Dig Dis Sci 2010;55(1):166–71.
8. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. Gastroenterology 2017;153(1):98–105.
9. 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(04):378–97.
10. Bisschops R, Areia M, Coron E, et al. Performance measures for upper gastrointestinal endoscopy: A European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. Endoscopy 2016;48(9):843–64.
11. Lee TJ, Siau K, Esmaily S, et al. Development of a national automated endoscopy database: The United Kingdom National Endoscopy Database (NED). United Eur Gastroenterol J 2019;7(6):798–806.
12. Torkamani A, Andersen KG, Steinhubl SR, et al. High-definition medicine. Cell 2017;170(5):828–43.
13. Wu L, Zhou W, Wan X, et al. A deep neural ne2rk improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019; 51(6):522–31.
14. 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(4):352–61.
15. Mildenberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. Eur Radiol 2002;12(4):920–7.
16. Wallace MB, Crook JE, Thomas CS, et al. Effect of an endoscopic quality improvement program on adenoma detection rates: A multicenter cluster-randomized controlled trial in a clinical practice setting (EQUIP-3). Gastrointest Endosc 2017;85(3):538–45.e4.
17. Coe SG, Crook JE, Diehl NN, et al. An endoscopic quality improvement program improves detection of colorectal adenomas. Am J Gastroenterol 2013;108(2):219–26.
18. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. Am J Gastroenterol 2006;101(4):873.
19. Bretthauer M, Aabakken L, Dekker E, et al. Requirements and standards facilitating quality improvement for reporting systems in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) position statement. Endoscopy 2016;48(3):291–4.
20. Yao K. The endoscopic diagnosis of early gastric cancer. Ann Gastroenterol 2013;26(1):11.
21. Wani S, Hall M, Keswani RN, et al. Variation in aptitude of trainees in endoscopic ultrasonography, based on cumulative sum analysis. Clin Gastroenterol Hepatol 2015;13(7):1318–25.
22. Carter KJ, Schaffer HA, Ritchie WP Jr. Early gastric cancer. Ann Surg 1984;199(5):604.
23. Park WG, Cohen J. Quality measurement and improvement in upper endoscopy. Tech Gastrointest Endosc 2012;14(1):13–20.
24. Zhou J, Wu L, Wan X, et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). Gastrointest Endosc 2019; 91(2):428–35.e2.
25. Faigel DO, Pike IM, Baron TH, et al. Quality indicators for gastrointestinal endoscopic procedures: An introduction. Gastrointest Endosc 2006;63(4 Suppl):S3–S9.
26. Pimentel-Nunes P, Libânio D, Marcos-Pinto R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European

Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa De Endoscopia Digestiva (SPED) guideline update 2019. Endoscopy 2019;51(4):365–88.
27. Keswani RN, Yadlapati R, Gleason KM, et al. Physician report cards and implementing standards of practice are both significantly associated with improved screening colonoscopy quality. Am J Gastroenterol 2015;110(8):1134–9.
28. Tinmouth J, Patel J, Hilsden RJ, et al. Audit and feedback interventions to improve endoscopist performance: Principles and effectiveness. Best Pract Res Clin Gastroenterol 2016;30(3):473–85.
29. Imperiali G, Minoli G, Meucci GM, et al. Effectiveness of a continuous quality improvement program on colonoscopy practice. Endoscopy 2007; 39(4):314–8.
30. Kahi CJ, Ballard D, Shah AS, et al. Impact of a quarterly report card on colonoscopy quality measures. Gastrointest Endosc 2013;77(6):925–31.
31. Keswani RN, Yadlapati R, Gleason KM, et al. Physician report cards and implementing standards of practice are both significantly associated with improved screening colonoscopy quality. Am J Gastroenterol 2015; 110(8):1134–9.
32. Abdul-Baki H, Schoen RE, Dean K, et al. Public reporting of colonoscopy quality is associated with an increase in endoscopist adenoma detection rate. Gastrointest Endosc 2015;82(4):676–82.
33. Matyja M, Pasternak A, Szura M, et al. Cecal intubation rates in different eras of endoscopic technological development. Videosurgery Other Miniinvasive Tech 2018;13(1):67.

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.