Commentary GUT

ARTIFICIAL INTELLIGENCE IN GI ENDOSCOPY: STUMBLING BLOCKS, GOLD STANDARDS, AND THE ROLE OF ENDOSCOPY SOCIETIES

Rüdiger Schmitz[1,2,3], René Werner[2,3], Alessandro Repici[4], Raf Bisschops[5], Alexander Meining[6], Michael Zornow[7], Helmut Messmann[8], Cesare Hassan[9], Prateek Sharma[10] and Thomas Rösch[1]

[1]Department of Interdisciplinary Endoscopy, [2]Institute of Computational Neuroscience and [3]Center for Biomedical Artificial Intelligence (bAIome), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [4]Humanitas Research Hospital, Department of Gastroenterology, Rozzano, Italy, [5]Department of Gastroenterology, University Hospital Gasthuisberg, Leuven, Vlaams Brabant, Belgium, [6]Medical Department II, Gastroenterology, University of Würzburg, Würzburg, Germany, [7]Chair for Public Law and European Law, University of Göttingen, Göttingen, Germany, [8]Department of Gastroenterology, University Hospital Augsburg, Augsburg, Germany, [9]Gastroenterology Unit, Nuovo Regina Margherita Hospital, Rome, Italy, [10]Division of Gastroenterology and Hepatology, Veteran Affairs Medical Center and University of Kansas, Kansas City, Kansas, USA.

Key words: Endoscopy, Artificial Intelligence, Computer Vision, Computerised Image Analysis
Abbreviations: AI – Aritificial Intelligence. ML – Machine Learning
Word count: 2725

Corresponding Author:
Prof. Dr. Thomas Rösch, MD
Department of Interdisciplinary Endoscopy,
University Medical Center Hamburg-Eppendorf,
Martinistr. 52, 20246 Hamburg, Germany,
Fax: +49-040-7410-40004
Email: t.roesch@uke.de

Summary

Recent developments in artificial intelligence and machine learning have led to novel and promising applications in gastrointestinal endoscopy and beyond. Endoscopic AI has already become a topic of intensive research and marketing, and a recent review in this journal is a timely and thorough guide that defines terminology and outlines best practices for the development and assessment of AI systems. Our commentary highlights selected aspects of AI research and elaborates upon potential roles that the GI scientific community and professional societies may play in ensuring that the preclinical achievements translate into clinical reality. We believe that: Databases of diagnostic AI software need to be made transparent as should the conditions of the testing scenarios. Clinical expertise is required from the early phases of system development and testing. Medicolegal implications need to be clarified as AI gets increasingly involved in medical decision making. Finally, through the establishment of a continuously updated and curated, dedicated database for AI system evaluation ("rolling gold standard"), the GI societies could actively advance and frame this ever-changing field.

Artificial intelligence (AI) has been portrayed as a silver bullet for a number of challenges encountered in GI endoscopy and beyond. Intense research, commercial, and media focus have led to the publication of studies with modest patient numbers and comparatively simple technology. There is no doubt that machine learning will be a determining medical development for the years to come. However, now that the dust has begun to settle, we are at a critical juncture where the focus is shifting from pre-clinical work toward the role of machine learning in clinical practice. Current issues relate to the evaluation and testing of AI systems, especially regarding patient outcomes, and to regulatory issues surrounding implementation. Many of these aspects pertain to one overarching question: how can we ensure that preclinical results translate into trustworthy clinical reality?

For the endoscopist, whether as a reader, a reviewer or a potential user of AI, it becomes increasingly important to understand the technical aspects of the systems and their performance measurements in order to realistically assess their practical value. Therefore, with GI endoscopy machine learning at the jump-off point from proof-of-principle studies [1–7] to clinical trials [8–12] – van der Sommen et al. provide us with an accessible guide to understand, assess and critically review the current ML endoscopy literature [13].

Our commentary highlights selected aspects of this review and AI as a whole and elaborates on the role of the GI endoscopy community with regards to both experiencing and framing the way ahead. In particular, we advocate a close collaboration of technology scientists and clinicians from early development phases onward to allow for the development of well-tailored AI algorithms and realistic preclinical testing. More transparency is needed with respect to the training data and the algorithm development process. In addition, in the legislative debates, the endoscopy societies need to play a critical role in defining the research priorities, minimum standards and quality metrics by having a strong voice and presence in this field. Furthermore, we propose the establishment of a "rolling gold standard" to meet requirements for continuous re-testing and benchmarking of AI systems.

How to Develop and Test Artificial Intelligence Systems – Clinical Contributions to Preclinical AI

The review in the previous issue of *Gut* starts with an excellent overview of "commonly (mis-)used terminology" of deep learning. The authors present essential methods and directly link them to specific pitfalls, with a particular focus on an important question: How shall we evaluate AI systems so that their performance translates into clinical reality?

On the technical level, this is closely related to 'generalizability'. As the authors explain, generalizability of an AI model refers to how well the model can cope with unknown data or data of a different source than its training data - e.g. images captured with SD resolution when the model was trained with HD images. Generalizability is countered by the concept of 'overfitting': As nicely summarized in Figure 1 of the review, this refers to the tendency of a ML model to capture details of the training data so tightly that it does not learn meaningful information beyond the training data set. Therefore, to examine whether the model improves or overfits, it is imperative to continuously test it on *independent* data which was *not* already utilized during the training process.

However, as the authors explain, this seemingly easy step is prone to pitfalls that lead to an overestimation of model performance: If for instance, endoscopy videos are split frame-by-frame rather than used in their entirety at the patient level, the model will be trained on one frame and validated on the next one (showing essentially the same view). This is nothing but the machine learning analogy of re-using the same exam question every year – it allows the student (or AI model) to achieve a high score by memorizing data by heart ('overfit') rather than by learning 'new' meaningful information. A similar source of performance overestimation or 'bias' is the association of certain diagnoses (e.g., presence of a visible lesion) with other imaging properties: For instance, detection of a suspicious lesion through AI could be linked to the use of NBI imaging, which the examiner always uses in these instances. Van der Sommen et al. demonstrate that problems with biases or the independence of datasets can be subtle and can occur on either the clinical or technical side. Testing of AI models is not trivial.

For the evaluation of AI models, we want to stress that one can observe considerable differences in algorithm performance based on the pre-clinical testing scenarios that are designed. For a polyp detection system, for instance, a "deliberately missed polyp scenario" [14] is a different story than high-quality close-up images from retrospectively collected examination videos, which are used in many studies. In general, testing scenarios should "[mimic] daily practice during endoscope withdrawal, where a polyp might be missed", as van der Sommen et al. claim, and realistic evaluation of AI systems requires an understanding of the intended clinical role and application which "highlights the importance of [...] continuous multidisciplinary interactions".

All these challenges from dataset design (splitting, biases) to model evaluation (testing scenarios) emphasize the value of a close and continuous interaction between engineers and endoscopists. perhaps even prior to the identification of a research question. At the first sight, model implementation might appear to be in the domain of the tech scientists. However, we want to stress that, even at this stage, clinicians can inject significant knowledge and provide input regarding the clinical question at hand and how they approach corresponding scenarios. In fact, designing machine learning models as a close mimicry of how humans approach analogous tasks can improve performance [15]. Therefore, the endoscopy community should definitely attempt to get involved in the development and testing phases on a continuous basis and at an early stage. Institutional support and dedicated funding for a closer interaction would certainly help toward the establishment of what has already been coined as "endoneering" [16].

How to Report on Artificial Intelligence Systems – a Demand for Transparency

Artificial intelligence, as of today, *inherently* is a 'black box'. Millions of internal parameters make it hard or impossible to really understand the inner workings of a system and to explain which features lead to which decision. And despite being a topic of intense research, [17,18], explainability remains a relevant limitation, and will probably continue to be in the future. This has implications on the transparency that should be required by the GI community. In publications, presentations and live demos, we often get the impression that most systems are heavily trained on small polyps, so that large and flat lesions are highlighted by only one or two small frames somewhere within the polyp. It could be argued that a weaker performance on large polyps may not be too problematic. However, a similar effect in the detection of more subtle lesions cannot be excluded without knowledge of the training set, as long as the inner workings of the system cannot be analyzed in depth. Therefore, we regard it essential to know the basis of which data an AI system was developed to assess its scope of application to the end user. To date, this is only known from one system, which used data from a prospective chromoendoscopy study [19]. For the inherent opacity of current AI, the assessment of the scope of application is dependent on knowledge of the training data, and more data transparency will be needed in the future.

Also, even if this is a sensitive issue and by no means new, commercial interests are strongly in favor of publishing positive results or may lead to "[production of] black box systems without supporting scientific evidence" to protect intellectual property. Generally, conflicts of interest are as important as in other fields and should be declared with the same scrutiny and we should push for high-class evidence and transparency. As a side note, AI is not immune to generating favorable results by selecting a favorable performance metric, as van der Sommen et al. explain. In the AI world, there are huge differences between the various valid metrics utilized for machine learning models and this topic in itself has become the subject of heated discussions [20]. Moreover, even seemingly basic measures such as sensitivity and false positives rates can be implemented in many different ways [21] – this should lead to all of us demanding further transparency also on why a given metric is chosen.

Coming back to the inherent opacity of the AI systems, this also implies medicolegal challenges regarding, e.g., model certification or the liability of the manufacturer. For instance, its black box nature may make it harder to check the validity of an AI system prior to deployment or prove its defectiveness in the court of law [22]. In the EU, this has led to debates on possible legislation

regarding the transparency of the algorithms [23] or obligations to detail or even preserve the training data [22]. Adaptive algorithms, i.e. AI systems that continue to learn after deployment, pose an additional challenge and will require specific regulation and further transparency on performance objectives and change protocols (as proposed by the FDA under the "total product lifecycle (TPLC) regulatory approach" [24]).

Yet another topic of the current regulatory debate is 'risk classification' of AI systems. For the clinician, it seems obvious that a device may help with detecting a few (or at best many) additional small polyps. However, interference with tissue diagnosis and 'objective' clinical management is an entirely different field. To avoid a scenario in which well-intended but overly strict risk classifications hamper the introduction and improvement of these systems, it is necessary that regulatory debates acknowledge the differences between specific applications [25]. Therefore, to ensure that future regulatory frameworks balance safety, transparency and permissiveness for innovation, the endoscopy societies should inject their knowledge of foreseeable techniques and application scenarios into the political debates.

Thus, to assess the validity and the scope of application of the systems, the community should at least require that the development process, the training data and the testing scenarios are described in detail and with particular focus on clinical aspects. Concerning basic and pre-clinical studies, publication of model source codes may be encouraged to ensure transparency and reproducibility – in line with corresponding trends in computer science. Research on AI explainability should be welcomed by the community. These aspects are paralleled by regulatory debates on data and algorithm transparency, and the societies should play a major role in this process.

How to Benchmark and (Re-)Certify Artificial Intelligence Systems

With regards to the clinical adoption of AI systems, van der Sommen et al. mention that, aside from the clinical trials which each individual system will have to pass, there may be an additional need to compare different systems, which society-collected "benchmarking datasets" might address. We expect the need for independent benchmarking to reach even beyond this scenario. Therefore, we propose, support and expand upon the idea of society-curated benchmark datasets [26,27].

We argue that, in principle, every AI system may require continued re-testing with the introduction of a new generation of scopes. This is because the generalizability of an AI system to new hardware conditions is difficult to guarantee. Also, minor updates of the AI system itself, introduction of additional features or continuous training of adaptive systems might require repeated re-testing. It is clearly impractical to achieve this with clinical trials only. Therefore, even though randomized clinical trials will remain the gold standard, there will be a need for an additional means for AI testing *before*, *aside* and *after* clinical trials. We have to develop cost-effective means to conduct such testing and ensure they are trustworthy. When benchmarking against a dataset, the dataset itself needs to be continuously updated to new hardware and new applications. This is for the same reason that requires re-testing of the AI system itself, namely insecure generalizability to changing conditions.

We, therefore, suggest to establish a 'rolling gold standard' for endoscopic AI benchmarking – a closed, hidden and endoscopy society-curated dataset that is regularly updated to new endoscope generations and other changing conditions (also cf. [27]). Further technical requirements need to be established by the GI community. In order to ensure that the benchmarking dataset remains hidden, for instance, tests need to be conducted on dedicated society hardware, and strictly only on random subsets of the entire benchmark dataset. Moreover, with up-coming applications for outcome prediction in mind, clinical follow-up might be sought from the beginning. To ensure that model uncertainty can be acknowledged and in line with further trends in machine learning, annotations by multiple individual experts should be maintained and not merged into a consensus annotation. Moreover, such efforts might also be used as an opportunity to support recent claims [27,28] and steps [29,30] toward the establishment of additional, public datasets for AI development, following the example of the well-established ImageNet dataset [31].

Furthermore and most importantly, it needs to be debated *whether and how* all this can be realized given the current market conditions in GI endoscopy. Also, we need solutions on *how* a 'rolling gold standard' can be implemented in the certification processes to handle (re-)certification and re-testing for software and hardware updates. For instance, one could envision a framework in which the independent (re-)evaluation of commercial AI systems is offered as a service to the industry, in exchange for a fee that enables the societies to fund the curation and updating of the 'rolling gold standard'. Current regulatory debates on how to involve stakeholders in the development of ex-post tests, [22], as well as attempts toward a "total product lifecycle regulatory approach", [24], might create a window of opportunity for the endoscopy societies to put forward and establish a rolling gold standard for AI benchmarking.

What to expect

The ultimate question remains: where is AI headed in clinical GI endoscopy**?** Looking into other areas in and also outside of medicine, the implementation of AI into daily routine seems inevitable. This creates enormous potential for improvements in our daily work life. As usual with innovations, market forces and scientific evidence both compete and cooperate in many different and often unforeseen ways. However, it is anticipated that AI will become part of the next generation of instruments of virtually all manufacturers and competing independent and software solutions may find their way to the endoscopy practice as well.

Endoscopists will probably use AI in routine endoscopy for lesion detection as well as eventually tissue diagnosis and perhaps even prediction of prognosis: The latter aspects have a potential to become a game changer for optical diagnosis. Furthermore, AI can improve endoscopy through many other applications. These range from automated report generation and procedure documentation, e.g. based on the recognition of anatomical landmarks [32] and endoscopic equipment [33], over procedural [8] to preparatory [34] quality monitoring. Also, complementary to automated report generation, AI-based methods have already been used for the extraction of standardized information [35] or the detection of genetic polyposis syndrome from previous medical records [36]. AI may push the rapid development of capsule-based imaging or telemedicine applications and interactions. Various AI applications with regards to patient management, logistical support and virtual training can be envisioned, in parallel to developments in other areas of service provision. Therefore, it is very welcome that the societies have entered debates on how to focus research and prioritize applications [26]. Additionally, we should continuously push for high-class evidence from outcome trials and monitor AI implementation over time, even though this might also lead to disillusions in certain cases [37].

Machine learning in GI endoscopy builds on the collaboration of clinical endoscopy, basic computer science, regulatory agencies and the GI societies. To avoid major pitfalls, a mutual understanding of these disciplines is vital. We hope that this commentary and the original review will ignite and influence a number of debates on how the endoscopy community may pave the path ahead: For instance, we need to assess the validity and the scope of systems despite their inherent 'black box nature'. Therefore, the community needs to debate how to systematically and transparently report on training and testing scenarios. The societies should assume a role in the political debate to ensure a balanced, permissive and application-specific regulatory approach. With regards to an anticipated demand to test, re-test and compare machine learning systems before, aside and after clinical trials, we envision that regularly updated benchmarking datasets as a 'rolling gold standard' could provide an ideal framework.

References

1 Groof J de, van der Sommen F, van der Putten J, *et al.* The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterology Journal* 2019;**7**:538–47. doi:10.1177/2050640619837443

2 Byrne MF, Chapados N, Soudan F, *et al.* Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*

2019;**68**:94–100. doi:10.1136/gutjnl-2017-314547
3 Misawa M, Kudo S, Mori Y, *et al.* Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy. *Gastroenterology* 2016;**150**:1531-1532.e3. doi:10.1053/j.gastro.2016.04.004
4 Ebigbo A, Mendel R, Probst A, *et al.* Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* 2019;:gutjnl-2019-319460. doi:10.1136/gutjnl-2019-319460
5 Struyvenberg MR, de Groof AJ, van der Putten J, *et al.* A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus. *Gastrointestinal Endoscopy* Published Online First: June 2020. doi:10.1016/j.gie.2020.05.050
6 de Groof AJ, Struyvenberg MR, van der Putten J, *et al.* Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020;**158**:915-929.e4. doi:10.1053/j.gastro.2019.11.030
7 Ebigbo A, Mendel R, Probst A, *et al.* Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut* 2018;:gutjnl-2018-317573. doi:10.1136/gutjnl-2018-317573
8 Wu L, Zhang J, Zhou W, *et al.* Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019;**68**:2161–9. doi:10.1136/gutjnl-2018-317366
9 Wang P, Berzin TM, Glissen Brown JR, *et al.* Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019;**68**:1813–9. doi:10.1136/gutjnl-2018-317500
10 Luo H, Xu G, Li C, *et al.* Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *The Lancet Oncology* Published Online First: October 2019. doi:10.1016/S1470-2045(19)30637-0
11 Mori Y, Kudo S, Misawa M, *et al.* Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Annals of Internal Medicine* 2018;**169**:357. doi:10.7326/M18-0249
12 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. doi:10.1053/j.gastro.2020.04.062
13 van der Sommen F, de Groof J, Struyvenberg M, *et al.* Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. *Gut* 2020;:gutjnl-2019-320466. doi:10.1136/gutjnl-2019-320466
14 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. doi:10.1053/j.gastro.2018.06.037
15 Schmitz R, Madesta F, Nielsen M, *et al.* Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. *arXiv:190910726 [cs, eess, q-bio]* Published Online First: 24 September 2019.http://arxiv.org/abs/1909.10726
16 Meining A. Endoneering: A new perspective for basic research in gastrointestinal endoscopy. *United European Gastroenterology Journal* 2020;**8**:241–5. doi:10.1177/2050640620913433
17 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: : IEEE 2017. 618–26. doi:10.1109/ICCV.2017.74
18 Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for Deep Networks. *arXiv:170301365 [cs]* Published Online First: 12 June 2017.http://arxiv.org/abs/1703.01365 (accessed 24 Nov 2020).
19 Repici A, Wallace MB, East JE, *et al.* Efficacy of Per-oral Methylene Blue Formulation for Screening Colonoscopy. *Gastroenterology* 2019;**156**:2198-2207.e1. doi:10.1053/j.gastro.2019.02.001
20 Maier-Hein L, Eisenmann M, Reinke A, *et al.* Why rankings of biomedical image analysis competitions should be interpreted with care. *Nature Communications* 2018;**9**. doi:10.1038/s41467-018-07619-7
21 Holzwanger EA, Bilal M, Glissen Brown JR, *et al.* Benchmarking False Positive Definitions for Computer Aided Polyp Detection in Colonoscopy. *Endoscopy* Published Online First: 2 November 2020. doi:10.1055/a-1302-2942
22 European Commission. White Paper On Artificial Intelligence - A European Approach to Excellence and Trust. *COM(2020) 65 final (19 Feb 2020)*https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed 5 Jul 2020).
23 European Commission. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. *COM(2020) 64 final (19 Feb 2020)*https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0064&from=en (accessed 20 Oct 2020).
24 U.S. Food & Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. 2019, p. 3-4. https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf (accessed 8 Oct 2020).
25 UEG Position Paper on AI. https://ueg.eu/files/1468/cec6f62cfb44b1be110b7bf70c8362d8.pdf (accessed 18 Nov 2020).
26 Berzin TM, Parasa S, Wallace MB, *et al.* Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE Task Force. *Gastrointestinal Endoscopy* Published Online First: June 2020. doi:10.1016/j.gie.2020.06.035
27 Parasa S, Wallace M, Bagci U, *et al.* Proceedings from the First Global Artificial Intelligence in Gastroenterology and Endoscopy Summit. *Gastrointestinal Endoscopy* 2020;**92**:938-945.e1. doi:10.1016/j.gie.2020.04.044
28 van der Sommen F. Gastroenterology needs its own ImageNet. *Journal of Medical Artificial Intelligence* 2019;**2**:23–23. doi:10.21037/jmai.2019.11.03
29 Borgli H, Thambawita V, Smedsrud PH, *et al.* HyperKvasir, a comprehensive multi-class image and video dataset

for gastrointestinal endoscopy. *Scientific Data* 2020;**7**. doi:10.1038/s41597-020-00622-y
30 Misawa M, Kudo S, Mori Y, *et al.* Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointestinal Endoscopy* Published Online First: July 2020. doi:10.1016/j.gie.2020.07.060
31 Deng J, Dong W, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009. 248–55. doi:10.1109/CVPR.2009.5206848
32 Su J-R, Li Z, Shao X-J, *et al.* Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointestinal Endoscopy* Published Online First: August 2019. doi:10.1016/j.gie.2019.08.026
33 Samarasena J, Yu AR, Torralba EJ, *et al.* Artificial Intelligence Can Accurately Detect Tools Used During Colonoscopy: Another Step Forward Toward Autonomous Report Writing: Presidential Poster Award: 1075. *Official journal of the American College of Gastroenterology | ACG* 2018;**113**.
34 Zhou J, Wu L, Wan X, *et al.* A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointestinal Endoscopy* Published Online First: November 2019. doi:10.1016/j.gie.2019.11.026
35 Laique SN, Hayat U, Sarvepalli S, *et al.* Application of optical character recognition with natural language processing for large-scale quality metric data extraction in colonoscopy reports. *Gastrointestinal Endoscopy* doi:10.1016/j.gie.2020.08.038
36 Parthasarathy G, Lopez R, McMichael J, *et al.* A natural language–based tool for diagnosis of serrated polyposis syndrome. *Gastrointestinal Endoscopy* Published Online First: May 2020. doi:10.1016/j.gie.2020.04.077
37 Lehman CD, Wellman RD, Buist DSM, *et al.* Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Internal Medicine* 2015;**175**:1828. doi:10.1001/jamainternmed.2015.5231

Contributors

All authors have together written the paper and approved the final version.

Competing interests

RW: Research support Siemens. AR: Medtronic and Cosmo consultancy, research support Cosmo, Fujifilm, Pentax. RB: Research support and consultancy Pentax, Fujifilm and Medtronic. CH: Medtronic and Fujifilm consultancy. PS: Consultancy Olympus and Boston Scientific, research support Cosmo, CDx Laboratories, Erbe, Fujifilm, Medtronic, and US Endoscopy. TR: Research support Olympus and Fujifilm. All other authors disclosed no financial relationships.

Patient consent for publication

Not required.