



Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model

Michael F Byrne,¹ Nicolas Chapados,^{2,3} Florian Soudan,² Clemens Oertel,² Milagros Linares Pérez,⁴ Raymond Kelly,⁵ Nadeem Iqbal,⁶ Florent Chandelier,² Douglas K Rex⁷

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2017-314547>).

¹Division of Gastroenterology, Vancouver General Hospital, Vancouver, British Columbia, Canada

²Department of Technology, Imagia, Montreal, Quebec, Canada

³Department of Applied Mathematics, Ecole Polytechnique de Montreal, Montreal, Quebec, Canada

⁴Department of Gastroenterology, Universidad de Buenos Aires, Buenos Aires, Argentina

⁵Department of Anaesthetics, Beaumont Hospital, Dublin, Ireland

⁶Department of Gastroenterology, Saint Luke's Hospital, Kilkenny, Ireland

⁷Division of Gastroenterology and Hepatology, Indiana University Medical Center, Indianapolis, Indiana, USA

Correspondence to

Dr Michael F Byrne, Division of Gastroenterology, Vancouver General Hospital, University of British Columbia, Vancouver, BC V5Z 1M9, Canada; mike@ai4gi.com

Received 23 May 2017

Revised 20 September 2017

Accepted 11 October 2017

Published Online First

24 October 2017



To cite: Byrne MF, Chapados N, Soudan F, *et al.* *Gut* 2019;**68**:94–100.

ABSTRACT

Background In general, academic but not community endoscopists have demonstrated adequate endoscopic differentiation accuracy to make the 'resect and discard' paradigm for diminutive colorectal polyps workable. Computer analysis of video could potentially eliminate the obstacle of interobserver variability in endoscopic polyp interpretation and enable widespread acceptance of 'resect and discard'.

Study design and methods We developed an artificial intelligence (AI) model for real-time assessment of endoscopic video images of colorectal polyps. A deep convolutional neural network model was used. Only narrow band imaging video frames were used, split equally between relevant multiclass. Unaltered videos from routine exams not specifically designed or adapted for AI classification were used to train and validate the model. The model was tested on a separate series of 125 videos of consecutively encountered diminutive polyps that were proven to be adenomas or hyperplastic polyps.

Results The AI model works with a confidence mechanism and did not generate sufficient confidence to predict the histology of 19 polyps in the test set, representing 15% of the polyps. For the remaining 106 diminutive polyps, the accuracy of the model was 94% (95% CI 86% to 97%), the sensitivity for identification of adenomas was 98% (95% CI 92% to 100%), specificity was 83% (95% CI 67% to 93%), negative predictive value 97% and positive predictive value 90%.

Conclusions An AI model trained on endoscopic video can differentiate diminutive adenomas from hyperplastic polyps with high accuracy. Additional study of this programme in a live patient clinical trial setting to address resect and discard is planned.

INTRODUCTION

Endoscopists combine their knowledge of the spectrum of endoscopic appearances of precancerous lesions with meticulous mechanical exploration and cleaning of mucosal surfaces to maximise lesion detection during colonoscopy. An extension of detection is endoscopic prediction of lesion histology, including differentiation of precancerous lesions from non-neoplastic lesions, and prediction

Significance of this study

What is already known on this subject?

- Using new imaging modalities such as narrow band imaging, endoscopists have studied the potential for a 'resect and discard' strategy for management of diminutive colorectal polyps.
- Experts have good results in general but community endoscopists fall short of Preservation and Incorporation of Valuable endoscopic Innovations (PIVI) guidelines.
- Artificial intelligence (AI) is rapidly growing and shows promise in performing optical biopsies.

What are the new findings?

- We show, using a type of AI known as deep learning, that our model is accurate at 94% in differentiating diminutive adenomas from hyperplastic polyps on unaltered videos of colon polyps.
- Our model operates in quasi real-time on such videos, with a delay of just 50 ms per frame.

How might it impact on clinical practice in the foreseeable future?

If validated in planned clinical trials with patients during live procedures, this AI platform could accelerate the adoption of a 'resect and discard' strategy for diminutive colorectal polyps. If validated in planned clinical trials with patients during live procedures, this AI platform could accelerate the adoption of a 'resect and discard' strategy for diminutive colorectal polyps.

of deep submucosal invasion of cancer.^{1 2} Image analysis can guide whether lesion removal is necessary and direct an endoscopist to the best resection method.^{1–3}

Image analysis during colonoscopy has achieved increasing acceptance as a means to accurately predict the histology of diminutive lesions,^{4 5} which have minimal risk of cancer,⁶ so that these diminutive lesions could be resected and discarded without pathological assessment or left in place without resection in the case of diminutive distal

colon hyperplastic polyps.³ Discarding most diminutive lesions without pathological assessment has the potential for large cost saving with minimal risk.^{7,8}

Unfortunately, both lesion detection during colonoscopy^{9–12} and image assessment of detected lesions during colonoscopy to predict histology^{13,14} are subject to substantial operator dependence. Thus, using virtual chromoendoscopy, experts have been able to exceed the accuracy threshold for polyp differentiation recommended to permit resect and discard,⁴ but performance by community based physicians has been variable and in some cases below accepted performance thresholds.^{13,14}

Accordingly, different initiatives were developed to investigate cost-effective approaches to derive qualitative histological information from endoscopic images, also referred to as optical biopsy. Originating from such effort, a sound body of evidence suggests that a simple narrow band imaging (NBI)-based classification system, the NBI International Colorectal Endoscopic (NICE) classification, could enable differentiating hyperplastic from adenomatous polyps (including diminutive polyps). The NICE classification scheme was designed to enable trained endoscopists to recognise visual cues such as colour, presence of vessels and surface patterns and to be readily applicable in routine practice without optical magnification by endoscopists without extensive experience in endoscopic imaging, chromoendoscopy or pit-pattern diagnosis.¹ However, NICE is not perfect and does not, for example, address the issue of sessile serrated polyps (SSPs), which is clearly problematic in the efforts to deliver ‘true’ optical biopsy. Attempts have been made to address this problem, such as the workgroup on serrated polyp and polyposis (WASP) scheme from the Dutch Workgroup on serrated polyp and polyposis (WASP).¹⁵ This is based on the NICE classification, but serrated lesions have now been added in the WASP classification. However, WASP also has its limitations, and very reliable optical biopsy continues to prove elusive in general usage.

The National Institute of Health and Care Excellence in the UK has very recently published evidence-based recommendations in an online document stating that virtual chromoendoscopy using NBI, Fuji Intelligent Chromo Endoscopy or Pentax i-SCAN is recommended to assess polyps of 5 mm or less during colonoscopy, instead of histopathology, to determine whether they are adenomatous or hyperplastic, only if high-definition enabled virtual chromoendoscopy equipment is used, the endoscopist has been trained to use virtual chromoendoscopy and accredited to use the technique under a national accreditation scheme, the endoscopy service includes systems to audit endoscopists and provide ongoing feedback on their performance and, importantly, the assessment is made with high confidence (<https://www.nice.org.uk/guidance/dg28>).

A potential solution to mitigate both the variability in endoscopic detection and histology prediction is to apply computerised image analysis to deliver computer decision support solutions. Recent studies have successfully used automatic image analysis techniques to accurately predict histology based on images captured with endocytoscopy¹⁶ and magnification endoscopy¹⁷ or to improve lesion detection.¹⁸ Studies using traditional machine learning^{16,17} have the limitations inherent to hand-crafted feature extraction, guided by the desire to ‘visually capture what is seen’ and are inherently limited by such. Considerable hand-engineering of imaging features is required for presentation to a support vector classifier. In addition, previous work has focused on high magnification endoscopy,^{16,17} which is not commonly used in clinical practice.

The European Society of Gastrointestinal Endoscopy (ESGE) published a technology review in 2016 in relation to advanced

endoscopic imaging.¹⁹ In this comprehensive review, the topic of decision support tools, and computer-aided diagnosis (CAD) was covered, with questions from this paper around the role of CAD assistance in training for optical diagnosis or whether such systems would initially be a ‘second reader’ to support the endoscopist’s diagnosis. The ESGE committee went on to further state that ‘the *stand alone* use of such systems to completely replace clinical judgment for decision making would require a much higher diagnostic performance and additional safeguards’ but that ‘availability of CAD combined with advanced endoscopic imaging is likely to emerge in clinical practice in the next few years’.

More recently, a field of artificial intelligence known as deep learning has opened the door to more detailed image analysis and real-time application by automatically extracting relevant imaging features, departing from human perceptual biases. Deep learning^{20,21} is an umbrella term for a wide range of machine learning models and methods, typically based on artificial neural networks,^{22,23} which aim at learning multilevel representations of data useful for making predictions or classifications. In particular, the development of deep convolutional neural networks (DCNN) has transformed the field of computer vision.^{23,24} In contrast to even recently published work in the gastroenterology literature, the DCNN approach in our study works in almost real-time with raw, unprocessed frames from the video sequence captured from the endoscope. In this study, we used a DCNN to train a deep learning-based AI model to differentiate conventional adenomatous from hyperplastic polyps. We tested the model on unaltered videos of 125 consecutively identified diminutive polyps with proven histology.

METHODS

We developed a deep learning-based AI model for real-time assessment of endoscopic video images of colorectal polyps. We used stored videos of unaltered endoscopic polyps provided by DKR to train the model. The videos used were available from a previous study and were deidentified, and hence the institutional review board waived review of this current study. We trained the model on videos containing NBI segments only of colorectal polyps captured with 190 series Olympus (Olympus Corp, Center Valley, Pennsylvania, USA) colonoscopes. All polyps were first detected in the normal ‘far focus’ mode. Then the colonoscope was moved close to the polyp and the near focus mode activated. There was no effort to video the polyp in the near focus mode for a set time interval. Once clear views had been subjectively obtained in the near focus mode, the polyp was resected and retrieved. The training video sequences comprised polyps of all size ranges including many polyps >10 mm in size, were previously de-identified and sorted by their respective pathology but were not of consecutive polyps. Videos of normal mucosa containing no polyp were also used to train the model. All recordings were made using high-definition Olympus video recorders.

The NICE classification¹ was used as the foundation for training the deep learning programme in association with the endoscopic video images. A DCNN model was used in this study. A convolutional neural network (CNN) is a type of artificial neural network used in deep learning and has been applied by several groups to analysis of visual imagery. CNNs incorporate very little preprocessing in comparison with other image classification algorithms, and these networks (such as used in our study) learn the filters that were previously hand-engineered in more traditional algorithms. This independence from

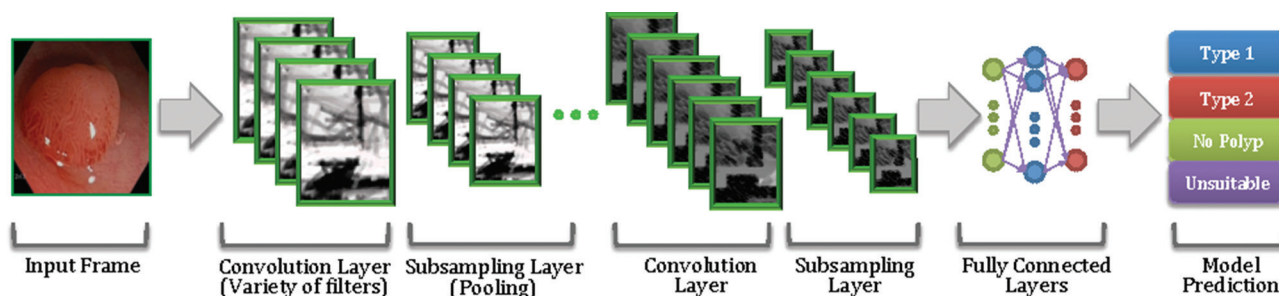


Figure 1 Schematic of the deep convolutional neural network model used.

prior knowledge and human effort in feature design represents a significant advantage of neural network models over other types of machine learning.^{25 26} The DCNN model used is based on the inception network architecture²⁴ (figure 1). Following standard procedure with DCNNs, model training was carried out with stochastic gradient descent from randomly initialised weights to minimise a frame-level cross-entropy loss function. To construct each mini batch of 128 frames during training, frames were randomly selected from the training set such that they were approximately balanced across classes and source video. For each frame, we applied a data augmentation procedure to create a richer diversity of frames by a random resizing and cropping of the frame, followed by a random flipping along either axis. Training stopped when the loss started increasing on an independent validation set.

Each frame was reviewed according to the multiclass model under consideration, by medical students and GI fellows (figure 2).

Frames (NBI only) used to train the model were split equally between relevant multiclass. The processing time of our DCNN model required 50 ms per frame on a PC with an NVIDIA graphics processing unit.

The DCNN model allowed essentially real-time analysis of endoscopic polyp videos and calculates a probability that a polyp is a conventional adenoma or a serrated class lesion. The

probability of a hyperplastic or adenomatous polyp (NICE types 1 and 2) is displayed immediately on each endoscopic video image.

To give a sense of how the system operates in real-time, the model builds a credibility score by analysing how the NICE class predictions fluctuate across successive frames (figure 3). The idea is to mimic the human perceptual system that promotes longitudinal coherence over short-lived information in order to provide clinically relevant information to the endoscopist. Accordingly, the credibility is also updated in real-time, in a form of exponential smoothing: $\text{credibility}(t) = \alpha * \text{credibility}(t-1) + (1-\alpha) * \text{update}(t)$, where $\text{update}(t)$ is an indication of whether the model's predictions have changed between frames at $t-1$ and t , and α is a parameter between 0 and 1 that is found by searching over a validation set. If the credibility is below 50%, the model is considered to have insufficient confidence to make a prediction. Videos with such a low credibility score were excluded from all accuracy calculations in a manner equivalent to a low confidence interpretation by an endoscopist. The process for determining confidence is quantitative and reproducible.

The training, validation, and final testing sets of endoscopic videos of polyps had no overlap. All frames were in NBI only and were a mixture of normal focus and near focus. For the training set, we used 223 polyp videos (29% NICE type 1, 53% NICE

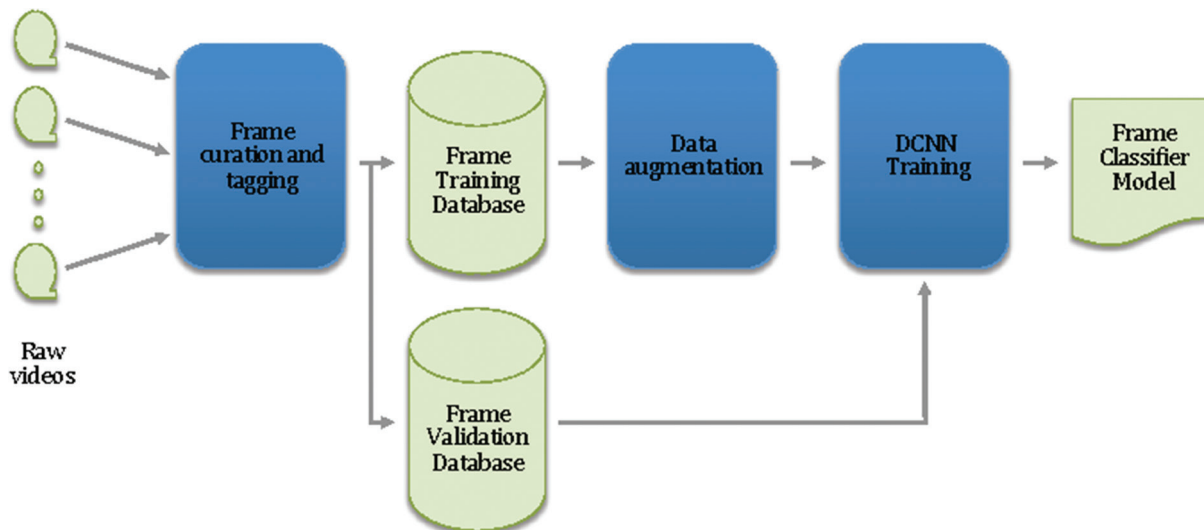


Figure 2 Schematic of the data preparation and training procedure of the deep convolutional neural network (DCNN) frame classifier. Raw videos are curated and tagged on a frame-by-frame basis. Then videos are split into disjoint databases: the larger serving as the training set and the smaller serving as a validation set. The purpose of the latter is to carry out 'early stopping' during the training procedure. Data augmentation is performed on the training frames only. After training, the resulting frame classification model can be used for prediction on new videos.

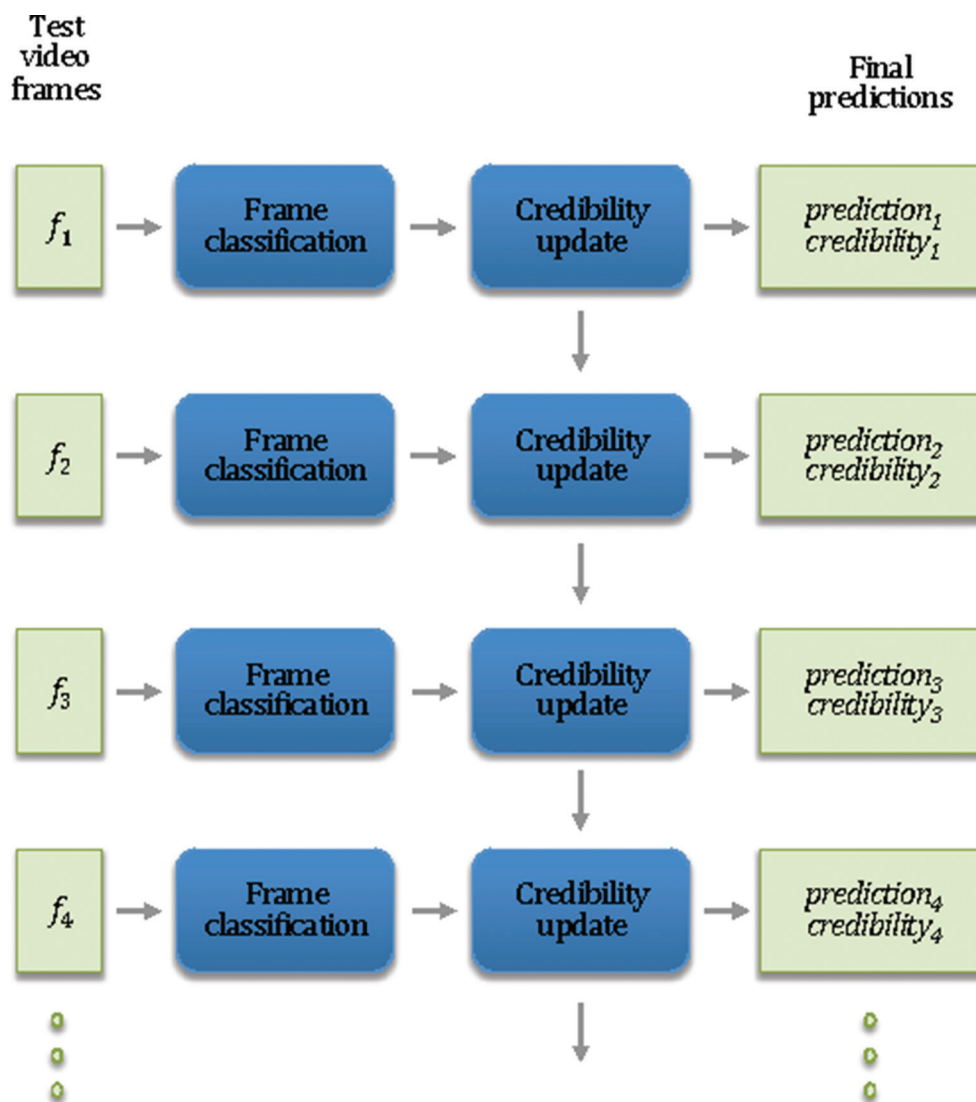


Figure 3 Illustration of the real-time prediction on a new video. Individual frames from the video are presented to the classification model (resulting from the training procedure), whose output is then processed by the credibility update mechanism. The result is a class probability for each frame (where the class may be one of 'NICE Type 1', 'NICE Type 2', 'No Polyp', 'Unsuitable'), as well as a credibility score between 0% and 100%. NICE, narrow band imaging International Colorectal Endoscopic.

type 2 and 18% of normal mucosa with no polyp), comprising 60 089 frames. For the validation set, we used 40 videos (NICE type 1, NICE type 2 and two videos of normal mucosa). The final test set included 125 consecutively identified diminutive polyps, comprising 51 hyperplastic polyps and 74 adenomas. Overfitting can be a concern in such studies. To address this, we ensured that all test images were from a completely separate dataset, never seen by the model in the training or validation phases, such that the reported results can represent the expected out-of-sample accuracy.

After training the model with the high-definition videos, the validation dataset was used to modify/fine-tune the 'hyper-parameters' for the AI system architecture (number of layers in the neural network, size of such layers, and so on). We then tested the model's accuracy on a consecutive sample of diminutive (≤ 5 mm) colorectal polyps that were video recorded in NBI and resected by DKR for histological analysis. DKR recorded the test videos for prospective use in another trial.²⁷ All videos were fully deidentified.

The video recordings for the test set were typically 10–20 s in length (median 16 s). All included normal and near focus imaging and at least one short frozen segment when a photograph was taken (the Olympus 190 video image freezes briefly when a photograph is taken). Each polyp was resected using cold methods (snare or forceps) and submitted to pathology separately.

Conventional adenomas were lesions that were dysplastic (high or low grade) and further characterised as tubular, tubulovillous or villous. All NICE type 1 lesions in the study were hyperplastic polyps by pathology (sessile serrated polyps (SSPs) were excluded). In the test set, polyps that were reported as normal tissue at pathology were not recut. We used the pathology report provided for routine patient care. Pathologists at Indiana University can access endoscopy reports at their discretion, and these reports include photographs of lesions in many cases. However, to our knowledge, none of the pathologists routinely access the reports and none of the pathologists are trained in endoscopic prediction of polyp pathology. Furthermore, none

Table 1 Assignment of narrow band imaging International Colorectal Endoscopic classification (NICE) type 1 vs NICE type 2 compared with the pathology determined histology

		Predicted by the model	
		NICE type 1	NICE type 2
Pathology	Hyperplastic	33	7
	Adenoma	1	65

of the reports contained verbal descriptions of the endoscopist's prediction of histology. The pathologists at Indiana University use widely accepted terminology to describe colorectal polyps, including the descriptors tubular, tubulovillous, and villous as well as low or high grade dysplasia for conventional adenomas. Serrated class lesions are characterised as traditional serrated adenomas, hyperplastic polyps or SSPs, without or with cytological dysplasia according to criteria recommended by a National Institute of Health consensus panel.²⁸

RESULTS

In the testing dataset, a total of 158 consecutive diminutive polyps were identified, video-recorded, resected, and submitted for pathological examination by DKR. Thirty polyps were excluded from the study because the pathological report was SSP (n=3), normal tissue or lymphoid aggregate (n=25) or faecal material (n=2); one video was excluded because it was corrupted and two had frames with multiple polyps.

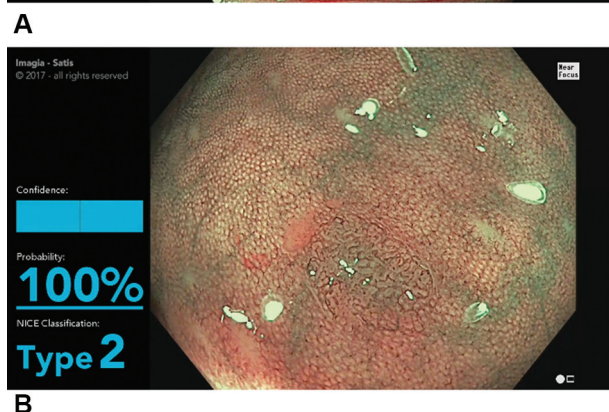
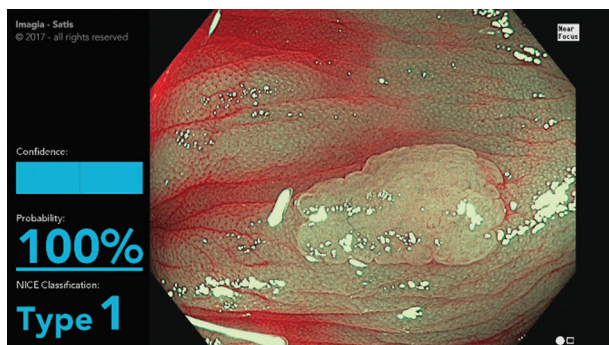


Figure 4 (A) Screen shot of the model during the evaluation of a NICE type 1 lesion (hyperplastic polyp). The display shows the type determined by the model (type 1) and the probability (100%). (B) Screen shot of the model in the evaluation of a NICE type 2 lesion (conventional adenoma). The display shows the type 2 determined by the model and the probability (100%) (see video). NICE, narrow band imaging International Colorectal Endoscopic.

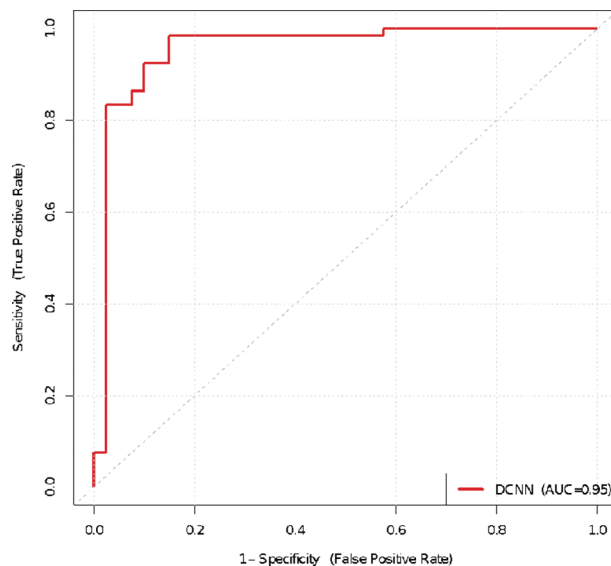


Figure 5 Receiver operator characteristic curve for the model differentiation of adenomatous versus hyperplastic polyps. AUC, area under the curve; DCNN, deep convolutional neural network.

Accordingly, 125 polyp videos were evaluated using the AI model. The final pathology of the 125 lesions was 51 hyperplastic polyps and 74 adenomas. Of these, the model did not build enough confidence to predict the histology in 19 polyps, leaving 106 in which the model made a high confidence prediction. Table 1 shows the predictions of the model for these 106 diminutive polyps compared with the histologies of the polyps. Figure 4 shows screenshots of the model as it appears in real-time in the evaluation of NICE type 1 and 2 lesions (unaltered) and for which the model reached high confidence. The video shows the model as it appears during colonoscopy (online supplementary video 1). For the 106 polyps, the accuracy of the model was 94% (95% CI 86% to 97%), the sensitivity for identification of adenomas was 98% (95% CI 92% to 100%), specificity was 83% (95% CI 67% to 93%), negative predictive value was 97% and positive predictive value was 90%.

DISCUSSION

Colonoscopy plays a pivotal role in diagnosis and prevention of colorectal cancer (CRC), which is overall the third leading cause of cancer death in the USA.²⁹ Unfortunately, colonoscopy is technically a highly operator dependent procedure, including detection of adenomas^{9 10} and serrated lesions^{11 12} and polyp resection.³⁰ This operator dependence leads to substantial variation between endoscopists in their effectiveness in preventing CRC with colonoscopy,^{31 32} which is the fundamental goal of most colonoscopies. Increasingly, clinicians are advised to make quality measurements,³³ and clinical trials address educational and technical adjuncts that could improve detection.³⁴

Although not all aspects of colonoscopy performance are currently amenable to reduction in performance variability by use of software, detection of polyps and prediction of histology are two aspects of performance that could potentially be enhanced by imaging analytics. In this study, we showed that an AI model trained in polyp differentiation could accurately identify whether consecutive diminutive polyps were conventional adenomas with an overall accuracy of 94%. For conventional adenomas verified by pathology, the sensitivity, specificity, positive predictive value and negative predictive value were 98%,

83%, 97%, and 90%, respectively. Acknowledging that our study is on unaltered videos rather than live patients, nonetheless, the AI model performed as well as experts typically perform using the NICE criteria and better than many community endoscopists have performed.⁴ Furthermore, the computer analysis of histological prediction is available in almost real-time (delay of 50 ms per frame). If the model's accuracy is verified in prospective clinical trials, it could revolutionise the management of diminutive colorectal polyps by essentially enabling the 'resect and discard' and 'leave distal colon hyperplastic polyps in place' paradigms to be accurately executed by both academic and community colonoscopists.

In this study, we apply deep learning to the real-time challenge for polyp differentiation into NICE types 1 and 2, using non-magnification colonoscopy, and most importantly where computer decision support is provided in real-time on unaltered endoscopic video streams. Previous studies of computer decision support for colorectal polyps have used magnifying colonoscopes¹⁷ or endocytoscopy,¹⁶ both of which are rarely available in the USA or Europe, and while acknowledging the great work of these investigators in this field, our DCNN approach is very different. Our model works with unprocessed frames and can operate in quasi real-time, with a frame processing time of 50 ms on consumer-grade hardware. Our model also works regardless of the polyp location in the frame (the operator does not need to precisely locate the polyp in the middle of the frame). The DCNN is trained end-to-end, meaning that the complete image preprocessing and classification task is solved within the same learning procedure, resulting in a much more robust model than previous work^{16,17} that consisted of hand-specified preprocessing followed by a trainable classifier. In the broader computer vision community, the end-to-end training of DCNNs has been, since 2013, systematically overtaking hand-engineered features and support vector classification. Ongoing work will determine if such an AI-based clinical decision support system could aid in the widespread adoption of a 'resect and discard' strategy.

Limitations of this study are several. These include collection of the videos by a single operator who is also a recognised expert colonoscopist and the use of video recordings rather than real-time assessments of polyps. However, we expect that the ability to manoeuvre and stabilise the instrument to allow stable imaging of colorectal polyps in focus will be achievable by colonoscopists with a wide range of skills. Even though of course this is not 'clinical' real-time in that we have not yet used this model in an actual patient setting, as mentioned in detail above, the testing dataset is raw, untouched, colon polyp screening footage, and our AI model performs in almost real-time (50 ms delay).

In addition, 19 of the 125 videos (15%) of consecutive diminutive polyps in the test set were excluded by the AI model, because it did not develop at least 50% confidence in the diagnosis. This low confidence determination by the model is analogous to a low confidence interpretation by an endoscopist.^{3,4} There was no particular trend for a certain type of histology or morphology in polyps where there was not enough confidence generated by the model. In a resect and discard paradigm, a polyp that the model could not generate >50% confidence in a diagnosis would be resected and sent to pathology. However, the videos used to train and test the model in this study were not originally recorded for the purpose of this study. In this retrospective video dataset, in some cases, images were blurred, or only very partial views were obtained, and the confidence in prediction was low. When the model is used in a true live patient scenario, an endoscopist would be able to move the colonoscope tip and change the image in an attempt to allow the model to

build up its confidence. This is essentially no different than what happens in day-to-day practice right now where endoscopists spend additional (few seconds) time looking more closely at a 'possible' polyp, washing the lens and so on. Thus, in actual clinical practice, the fraction of polyps with low confidence ratings may be lower than observed in this study. Such a clinical study is currently being planned.

The NICE classification system has been criticised for not incorporating sessile serrated adenomas (SSAs), and the WASP schema has been suggested as an improvement as it incorporates SSAs.¹⁵ We chose traditional adenomas and hyperplastic polyps, and the NICE classification, for this study as a proof of concept. Our AI algorithm is pathology agnostic. It is important to point out that we do not specify 'a priori' any imaging features that may distinguish between type 1 and type 2 polyps, for example. We never code within the system any features that can help distinguish between types 1 and 2. Our model discriminates this from raw pixels and nothing more. A 'binary' decision could easily be a 'three way or a four way or more' decision, determining if a polyp is a traditional adenoma, a benign hyperplastic polyp or a SSA, a lymphoid aggregate or normal tissue, for example. The only difference in such a model would be the composition of the training dataset. We are already collecting datasets to work on this clinical question of SSAs, but in our current work, we chose two polyp classes with our novel AI approach. Furthermore, there is currently great difficulty in studying AI or any other endoscopist method to identify SSP/SSA because the pathology gold standard is subject to marked interobserver variation in differentiation from hyperplastic polyps. Despite this limitation, the AI programme described here could still be used to support a resect and discard paradigm for diminutive polyp management because experts in this field are now endorsing a strategy of resect and discard for diminutive adenomas anywhere in the colon, identify and leave in place NICE type 1 lesions in the rectosigmoid (which are hyperplastic in >98% of cases), and resect and submit to pathology for NICE type 1 lesions proximal to the sigmoid (to allow the opportunity to identify SSAs by pathology in these lesions).

Incorporation of AI into widespread community clinical use will be challenging. Any new technology will have to have minimal impact on the workflow of the endoscopist and also not be distracting with its onscreen presence. The 'form factor' for incorporation of AI into clinical endoscopy will be crucial to its adoption and safe use. There will also be significant regulatory and reimbursement hurdles to overcome before artificial intelligence in endoscopy becomes a reality in clinical practice. Gaining the confidence of the physician community will be key to allow said physicians to gain the confidence of our patients that a computer can help the doctor to make a diagnosis or the even bigger challenge of placing trust in the exclusive decision making of a computer.

For the future, a similar deep learning approach also holds substantial potential to facilitate detection by highlighting areas of possible adenomatous or serrated mucosa for close inspection by the endoscopist. In addition, the strategy followed here for training the programme to differentiate hyperplastic polyps from adenomatous could potentially be used to improve diagnostic assessment of a variety of endoscopic images, addressing clinical problems such as identification of dysplasia in Barrett's oesophagus and detection of intestinal metaplasia and dysplasia in the gastric mucosa. Furthermore, alternative endoscopic images such as confocal laser and endocytoscopy can potentially be used to train this platform to provide automatic interpretation of clinically acquired images.

In summary, we have demonstrated that an AI model can achieve high accuracy in sorting diminutive colorectal polyps into conventional adenoma versus hyperplastic polyps when used on unaltered colon polyp video sequences. We are planning clinical trials to evaluate the potential of this imaging analytics AI technology in day-to-day practice.

Contributors Study conception and design: MFB. Drafting the manuscript: MFB and DKR. Data analysis: all authors. Development of the artificial intelligence model: NC, FS and FC. Video recording: DKR. Critical revision of the manuscript: all authors.

Funding This work was primarily supported by 'ai4gi', a joint venture between Satis Operations Inc and Imagia Cybernetics.

Competing interests MFB: CEO and shareholder, Satis Operations Inc, 'ai4gi' joint venture; research support: Boston Scientific. NC: Imagia shareholder, 'ai4gi' joint venture. FS: Imagia shareholder, 'ai4gi' joint venture. CO: Imagia shareholder, 'ai4gi' joint venture. FC: Imagia shareholder, 'ai4gi' joint venture. DKR: consultant: Olympus Corp and Boston Scientific; research support: Boston Scientific, Endochoice and EndoAid.

Ethics approval Indiana University.

Provenance and peer review Not commissioned; internally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2019. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Hewett DG, Kaltenbach T, Sano Y, *et al.* Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology* 2012;143:599–607.
- Hayashi N, Tanaka S, Hewett DG, *et al.* Endoscopic prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging international colorectal endoscopic (NICE) classification. *Gastrointest Endosc* 2013;78:625–32.
- Rex DK, Kahi C, O'Brien M, *et al.* The American society for gastrointestinal endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2011;73:419–22.
- Abu Dayyeh BK, Thosani N, Konda V, *et al.* ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015;81:502.e1–6.
- Kamiński MF, Hassan C, Bisschops R, *et al.* Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2014;46:435–57.
- Ponugoti PL, Cummings OW, Rex DK. Risk of cancer in small and diminutive colorectal polyps. *Dig Liver Dis* 2017;49:34–7.
- Kessler WR, Imperiale TF, Klein RW, *et al.* A quantitative assessment of the risks and cost savings of forgoing histologic examination of diminutive polyps. *Endoscopy* 2011;43:683–91.
- Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. *Clin Gastroenterol Hepatol* 2010;8:865–9.
- Barclay RL, Vicari JJ, Doughty AS, *et al.* Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *N Engl J Med* 2006;355:2533–41.
- Chen SC, Rex DK. Endoscopist can be more powerful than age and male gender in predicting adenoma detection at colonoscopy. *Am J Gastroenterol* 2007;102:856–61.
- Hetzel JT, Huang CS, Coukos JA, *et al.* Variation in the detection of serrated polyps in an average risk colorectal cancer screening cohort. *Am J Gastroenterol* 2010;105:2656–64.
- Kahi CJ, Li X, Eckert GJ, *et al.* High colonoscopic prevalence of proximal colon serrated polyps in average-risk men and women. *Gastrointest Endosc* 2012;75:515–20.
- Ladabaum U, Fioritto A, Mitani A, *et al.* Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions. *Gastroenterology* 2013;144:81–91.
- Rees CJ, Rajasekhar PT, Wilson A, *et al.* Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. *Gut* 2017;66:887–95.
- Ilispeert JE, Bastiaansen BA, van Leerdam ME, *et al.* Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut* 2016;65:963–70.
- Mori Y, Kudo SE, Wakamura K, *et al.* Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc* 2015;81:621–9.
- Kominami Y, Yoshida S, Tanaka S, *et al.* Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016;83:643–9.
- Tajbakhsh N, Gurudu SR, Liang J. A comprehensive computer-aided polyp detection system for colonoscopy videos. *Inf Process Med Imaging* 2015;24:327–38.
- East JE, Vleugels JL, Roelandt P, *et al.* Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy* 2016;48:1029–45.
- LeCun Y, Bengio Y, Hinton G, *et al.* Deep learning. *Nature* 2015;521:436–44.
- Murphy KP. *Machine learning: a probabilistic perspective*: MIT press, 2012.
- LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proc IEEE Inst Electr Electron Eng* 1998;86:2278–324.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012:1097–105.
- Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision. IEEE Conference on Computer Vision and Pattern Recognition 2016;2818–26.
- Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
- Goodfellow I, Bengio Y, Courville A. *Deep learning*: Mit press, 2016.
- Rex DK, Ponugoti P, Kahi C. The "valley sign" in small and diminutive adenomas: prevalence, interobserver agreement, and validation as an adenoma marker. *Gastrointest Endosc* 2017;85:614–21.
- Rex DK, Ahnen DJ, Baron JA, *et al.* Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol* 2012;107:1315–29.
- Siegel RL, Miller KD, Jemal A, *et al.* Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7–30.
- Pohl H, Srivastava A, Bensen SP, *et al.* Incomplete polyp resection during colonoscopy: results of the Complete Adenoma Resection (CARE) study. *Gastroenterology* 2013;144:74–80.
- Kamiński MF, Regula J, Kraszewska E, *et al.* Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;362:1795–803.
- Corley DA, Jensen CD, Marks AR, *et al.* Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014;370:1298–306.
- Rex DK, Schoenfeld PS, Cohen J, *et al.* Quality indicators for colonoscopy. *Am J Gastroenterol* 2015;110:72–90.
- Coe SG, Crook JE, Diehl NN, *et al.* An endoscopic quality improvement program improves detection of colorectal adenomas. *Am J Gastroenterol* 2013;108:219–26.