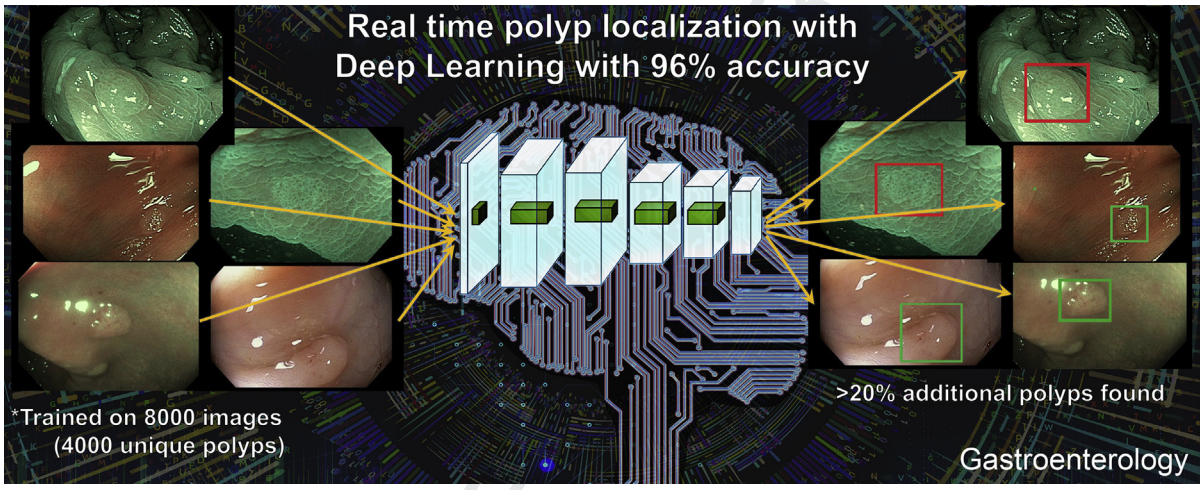
Gastroenterology 2018;■:1–10

Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy

Gregor Urban,[1,2] Priyam Tripathi,[4] Talal Alkayali,[4,5] Mohit Mittal,[4] Farid Jalali,[4,5] William Karnes,[4,5] and Pierre Baldi[1,2,3]

[1]Department of Computer Science, University of California, Irvine, California; [2]Institute for Genomics and Bioinformatics, University of California, Irvine, California; [3]Center for Machine Learning and Intelligent Systems, University of California, Irvine, California; [4]Department of Medicine, University of California, Irvine, California; and [5]H.H. Chao Comprehensive Digestive Disease Center, University of California, Irvine, California



BACKGROUND & AIMS: The benefit of colonoscopy for colorectal cancer prevention depends on the adenoma detection rate (ADR). The ADR should reflect the adenoma prevalence rate, which is estimated to be higher than 50% in the screening-age population. However, the ADR by colonoscopists varies from 7% to 53%. It is estimated that every 1% increase in ADR lowers the risk of interval colorectal cancers by 3%–6%. New strategies are needed to increase the ADR during colonoscopy. We tested the ability of computer-assisted image analysis using convolutional neural networks (CNNs; a deep learning model for image analysis) to improve polyp detection, a surrogate of ADR. **METHODS:** We designed and trained deep CNNs to detect polyps using a diverse and representative set of 8,641 hand-labeled images from screening colonoscopies collected from more than 2000 patients. We tested the models on 20 colonoscopy videos with a total duration of 5 hours. Expert colonoscopists were asked to identify all polyps in 9 de-identified colonoscopy videos, which were selected from archived video studies, with or without benefit of the CNN overlay. Their findings were compared with those of the CNN using CNN-assisted expert review as the reference. **RESULTS:** When tested on manually labeled images, the CNN identified polyps with an area under the receiver operating characteristic curve of 0.991 and an accuracy of 96.4%. In the analysis of colonoscopy videos in which 28 polyps were removed, 4 expert reviewers identified 8 additional polyps without CNN assistance that had not been removed and identified an additional 17 polyps with CNN assistance (45 in total). All polyps removed and identified by expert review were detected by the CNN. The CNN had a false-positive rate of 7%. **CONCLUSION:** In a set of 8,641 colonoscopy images containing 4,088 unique polyps, the CNN identified polyps with a cross-validation accuracy of 96.4% and an area under the receiver operating characteristic curve of 0.991. The CNN system detected and localized polyps well within real-time constraints using an ordinary desktop machine with a contemporary graphics processing unit. This system could increase the ADR and decrease interval colorectal cancers but requires validation in large multicenter trials.

Keywords: Machine Learning; Convolutional Neural Networks; Colorectal Cancer Prevention; Adenoma Detection Rate Improving Technology.

Abbreviations used in this paper: **ADR, adenoma detection rate; AUC, area under the curve; CNN, convolutional neural network; CRC, colorectal cancer; FPR, false-positive rate; NBI, narrow-band imaging; NPI, not pre-initialized; PI, pre-initialized; WLE, white light endoscopy.**

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2018.06.037

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text •

NEW FINDINGS

Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text •

LIMITATIONS

Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text •

IMPACT

Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text • Placeholder text •

CLINICAL AT

Colorectal cancer (CRC) is the second leading cause of cancer-related death in the United States.[1] CRC arises from precancerous polyps[2] with a mean dwell time of at least 10 years.[3] The National Polyp Study showed that 70%–90% of CRCs are preventable with regular colonoscopies and removal of polyps.[4] Seven percent of 9% of CRCs occur despite being up to date with colonoscopy.[5] It is estimated that 85% of these "interval cancers" are due to missed polyps or incompletely removed polyps during colonoscopy.[6]

The prevalence of precancerous polyps in the screening population older than 50 years is estimated to be at least 50%.[7] Adenomas are the most prevalent precancerous polyp. The adenoma detection rate (ADR; percentage of screening colonoscopies with $\geq$1 adenoma found) is a measure of a colonoscopist's ability to find adenomas. Ideally, the ADR should reflect adenoma prevalence. Unfortunately, the ADR varies widely (7%–53%) among colonoscopists performing screening colonoscopies.[8] In tandem colonoscopies, 22%–28% of polyps and 20%–24% of adenomas were missed[7] and CRC had a diagnostic miss rate of 5%.[9] The ADR is dependent on a colonoscopist's level of training, time spent, and technique used during withdrawal, preparation quality, and other colonoscopist- and procedure-dependent factors.[10] A large Kaiser Permanente study showed that for each 1% increase in the ADR, the interval cancer rate was decreased by 3%.[8] A subsequent study with nearly 1 million person-years of follow-ups in Poland showed a 6% decrease in interval cancer rates for each 1% increase in the ADR.[11] This study also showed an 82% decrease in interval cancer rates among colonoscopists that improved their ADRs to the top quintile. Not surprisingly, the ADR currently is a key quality measure reportable in the United States to the Centers for Medicare and Medicaid and is tied to reimbursement under the Medicare Access and CHIP Reauthorization Act of 2015 and the Merit Based Incentive Payments System beginning in the 2017.[12]

Several novel technologies have been developed to improve the ADR, including enhanced optics (resolution, zoom and wide angle, chromoendoscopy, digital autofluorescence, extra lenses for side and forward views) and attachments and modifications to aid view behind and between folds, including cap-assisted techniques and a balloon-assisted device.[13] Extra-wide angle colonoscopes and multi-camera systems initially showed promise to increase the ADR compared with standard forward-facing camera systems.[13] However, a recent meta-analysis and large randomized study showed no difference in the ADR compared with standard forward-viewing colonoscopy.[14] A review of 5 studies on the effect of high-definition colonoscopes on the ADR showed conflicting evidence,[13] with 1 study concluding that the ADR is improved only for endoscopists with a low ADR (<20%).[15] Similarly, most studies on digital chromoendoscopy, specifically narrow-band imaging (NBI), have found that NBI does not improve the ADR compared with white light imaging.[13] Evidence suggests positive effects of autofluorescence, but it is associated with added expense and poor image resolution.[13]

Computer-assisted image analysis has the potential to further aid adenoma detection but has remained underdeveloped. A notable benefit of such a system is that no alteration of the colonoscope or procedure is necessary.

Deep learning has been successfully applied to many areas of science and technology,[16] such as computer vision,[17–21] speech recognition,[22] natural language processing,[23] games,[24,25] particle physics,[26,27] organic chemistry,[28] and biology,[29–34] to name just a few areas and examples. A convolutional neural network (CNN) is a type of deep learning model that is highly effective at performing image analysis.

Ideally, a polyp-detection assistance module should have a sensitivity of 1 (or close to it) to avoid false-negative results, but this comes at the cost of an increased false-positive rate (FPR) when the area under the curve (AUC; Performance Evaluation and Metrics section in the Supplement) is not close to 1. A large FPR, even with perfect sensitivity, diminishes the benefits of an assistance system when user desensitization comes into play. A polyp-detection module also must process images at a minimum of 30 frames per second to be applicable during colonoscopy. Therefore, surmounting the constraints of accuracy and processing speed were our primary goals.

Methods

Convolutional Neural Networks

We trained different CNN architectures in this study, including models with weights initialized by training on the ImageNet data corpus,[35] before refining the weights in our dataset. All trained CNNs consisted of the same fundamental building blocks, including (1) convolutional layers, (2) fully