Colonoscopy is commonly used to screen for colorectal cancer (CRC). We develop a deep learning model called CRCNet for optical diagnosis of CRC by training on 464,105 images from 12,179 patients and test its performance on 2263 patients from three independent datasets. At the patient-level, CRCNet achieves an area under the precision-recall curve (AUPRC) of 0.882 (95% CI: 0.828–0.931), 0.874 (0.820–0.926) and 0.867 (0.795–0.923). CRCNet exceeds average endoscopists performance on recall rate across two test sets (91.3% versus 83.8%; two-sided t-test, p < 0.001 and 96.5% versus 90.3%; p = 0.006) and precision for one test set (93.7% versus 83.8%; p = 0.02), while obtains comparable recall rate on one test set and precision on the other two. At the image-level, CRCNet achieves an AUPRC of 0.990 (0.987–0.993), 0.991 (0.987–0.995), and 0.997 (0.995–0.999). Our study warrants further investigation of CRCNet by prospective clinical trials.
Cancer-related cancer (CRC) ranks the second leading cause of cancer-related death and the third most common cancer types worldwide. Colonoscopy is the most frequently used tool to screen for CRC, offering direct biopsy of intestinal tumor masses for pathological diagnosis. The advantage of colonoscopic screening is its ability to detect precancerous lesions and CRC at early stage, during which surgical removal is often curative. Colonoscopic screening has contributed to reduced mortality of CRC in observational studies, where ~80% of CRC can be prevented by polypectomy. Comparative studies have shown that prompt treatment of severe atypical hyperplasia or CRC at early stage prolonged overall survival. In addition, incomplete biopsy often leads to misdiagnosis of early CRC as mild or moderate dysplasia, subsequently leading to inappropriate treatment. Therefore, accurate differentiation between malignant and benign lesions under colonoscopy is medically important to select optimal treatment regimen, avoid inappropriate endoscopic resection, and improve cost-effectiveness. A key aspect of endoscopists is the differential diagnosis of colonoscopic lesions, according to NCI Intestinal Polyp Classification. CRCNet is a deep learning model that can improve the efficiency of colonoscopy, and can be used to detect precancerous lesions and CRC at early stage prolonging overall survival and improving cost-effectiveness.

### Results

#### Baseline characteristics of training and test datasets

We obtained 464,105 images from 12,179 patients between 2011 and 2019 at TCH as training set. We subsequently assembled three test sets that consisted of 20,783 images from 363 patients between April 2019 and May 2019 at TCH, 15,441 images from 430 patients between January 2018 and February 2019 at Tianjin First Central Hospital (TFCH), and 48,391 images from 1470 patients between January 2018 and December 2018 at Tianjin General Hospital (TGH), respectively. All CRC patients \(n = 3176\) and 56.1\% (5050/9003) of control patients in training set and all patients in three test sets have pathological examination results. The training set consisted of 3176 CRC patients and 9003 controls. Male sex accounted for 62.5\% (1985/3176) in CRC patient group versus 54.5\% (4909/9003) in the control group. Ages were 60 (53–67) in the CRC patient group versus 57 (49–64) for the control group. For CRC patients, tumors from different sites were included: ascending colon (19.9\%, n = 631), transverse colon (4.5\%, n = 143), descending colon (5.6\%, n = 178), sigmoid colon (17.7\%, n = 561), and rectum (52.4\%, n = 1663). There were 11.4\% of CRC patients at stage I, 44.3\% at stage II, 7.3\% at stage III, and 2\% at stage IV; whereas 35.7\% patients had no TNM stage information because they did not receive surgical resection but only colonoscopy-guided biopsy (Table 1). Besides normal mucosa, multiple benign diseases were encompassed in the control group, including adenoma, hyperplastic polyps, sessile serrated adenomas/polyps, inflammatory bowel disease, and chronic mucosal inflammation (Supplementary Table 1). There were 146 CRC patients and 217 controls in TCH test set (n = 363), 90 CRC patients and 340 controls in TFCH test set (n = 430), and 71 CRC patients and 1399 controls in TGH test set (n = 1470). The additional detailed number of patients with regards to sex, age, tumor site, TNM stage, and a flowchart depicting these processes were provided in Table 1 and Fig. 1, respectively.

#### Performance of CRCNet on three independent test sets

We trained CRCNet iteratively and evaluated the performance of the best model on three test sets (see “Methods”). Malignancy score for each patient was calculated according to Eq. (1). We found that CRCNet achieved high performance in identifying CRC patients: for the TCH test set, area under the PR curve (AUPRC) was 0.882 (95% CI: 0.828–0.931), accuracy was 87.3\% (83.5–90.6\%), recall rate was 90.4\% (0.844–0.947), specificity was 85.3\% (79.8–89.7\%), and F1 was 85.2\% for TFCH test set, AUPRC was 0.874 (0.820–0.926), accuracy was 91.6\% (88.6–94.1\%), recall rate was 78.9\% (69.0–86.8\%), specificity was 95.0\% (92.1–97.1\%), and F1 was 79.8\% for TGH test set, AUPRC was 0.867 (0.795–0.923), accuracy was 98.0\% (97.2–98.7\%), recall rate was 74.6\% (62.9–84.2\%), specificity was 99.2\% (98.6–99.6\%), and F1 was 78.5\% (Table 2 and Supplementary Fig. 1). The area under the receiver-operating characteristic (AUROC) curve was 0.930 (0.903–0.956) for TCH, 0.961 (0.943–0.979) for TFCH, and 0.989 (0.980–0.997) for TGH test sets (Supplementary Fig. 2). The other classification metrics such as precision, negative predictive value and kappa coefficient were provided in Table 2.

Notably, the predicted malignancy scores of CRC patients at early stage (I and II) and advanced stage (III and IV) are comparable (mean 0.679 versus 0.689; two-sided t-test, \(p = 0.843\);
The AUPRC values of CRCNet were provided as Supplementary Data 1. The activation mapping (Grad-CAM) images of these 36 cases were provided in Table 3. Detailed classification metrics for each endoscopist were provided in Supplementary Table 2. The interrater agreement rate for this group of five experienced endoscopists, CRCNet obtained an F1 metric of 0.852, 0.857, and 0.928 versus 0.768, 0.878, and 0.873 on these three subsets from test sets. The other classification metrics such as accuracy, recall rate, specificity, precision, negative predictive value, and kappa coefficient were provided in Table 3. Detailed classification metrics for each endoscopist were provided in Supplementary Table 3. The interrater agreement rate for this group of five experienced endoscopists was 56.7% (206/363, Fleiss Κappa 0.58; two-sided z-test, \( p < 0.001 \)) in TCH test set, 76.6% (222/290, Fleiss Κappa 0.75; two-sided z-test, \( p < 0.001 \)) in TFCH test set, and 88.2% (239/271, Fleiss Κappa 0.87; two-sided z-test, \( p < 0.001 \)) in TGH test set. The average precision and recall rate of this group of five endoscopists were situated below PR curves in TCH and TFCH test sets (Fig. 2a, c), while it was marginally situated above PR curve in TGH test set (Fig. 2b). At the average precision level of this group of five endoscopists, CRCNet obtained higher recall rate in TCH (91.3% versus 83.8%; two-sided binomial test, \( p < 0.001 \)) and TGH (96.5% versus 90.3%; two-sided binomial test, \( p = 0.02 \)) test set, and comparable precision rate in TFCH (82.9% versus 87.6%; two-sided binomial test, \( p = 0.29 \)) test set. Whereas at the average recall rate level of this group of five endoscopists, CRCNet obtained higher precision in TGH (93.7% versus 83.8%; two-sided binomial test, \( p = 0.02 \)) test set, and comparable precision in TCH (81.3% versus 77.9%; two-sided binomial test, \( p = 0.32 \)) and TFCH (81.1% versus 83.4%; two-sided binomial test, \( p = 0.52 \)) test sets.

**Performance of CRCNet versus five endoscopists.** All CRC patients and a random subset of 200 controls from TCH and TGH test sets, and all patients from TCH test set were used to evaluate the performance of CRCNet versus a group of five skilled endoscopists. The total number of images read and interpreted by each endoscopist were 38,788. CRCNet achieved an AUPRC of 0.882 (95% CI 0.828–0.931) for TCH test set, 0.920 (95% CI 0.874–0.955) for TFCH test set, and 0.969 (95% CI 0.937–0.992) for TGH test set (Fig. 2). As compared with endoscopists, CRCNet achieved an F1 metric of 0.852, 0.857, and 0.928 versus 0.768, 0.878, and 0.873 on these three subsets from test sets. The other classification metrics such as accuracy, recall rate, specificity, precision, negative predictive value, and kappa coefficient were provided in Table 3. Detailed classification metrics for each endoscopist were provided in Supplementary Table 2. The interrater agreement rate for this group of five experienced endoscopists was 56.7% (206/363, Fleiss Κappa 0.58; two-sided z-test, \( p < 0.001 \)) in TCH test set, 76.6% (222/290, Fleiss Κappa 0.75; two-sided z-test, \( p < 0.001 \)) in TFCH test set, and 88.2% (239/271, Fleiss Κappa 0.87; two-sided z-test, \( p < 0.001 \)) in TGH test set. The average precision and recall rate of this group of five endoscopists were situated below PR curves in TCH and TFCH test sets (Fig. 2a, c), while it was marginally situated above PR curve in TGH test set (Fig. 2b). At the average precision level of this group of five endoscopists, CRCNet obtained higher recall rate in TCH (91.3% versus 83.8%; two-sided binomial test, \( p < 0.001 \)) and TGH (96.5% versus 90.3%; two-sided binomial test, \( p = 0.006 \)) test sets, and comparable recall rate in TFCH (82.9% versus 87.6%; two-sided binomial test, \( p = 0.29 \)) test set. Whereas at the average recall rate level of this group of five endoscopists, CRCNet obtained higher precision in TGH (93.7% versus 83.8%; two-sided binomial test, \( p = 0.02 \)) test set, and comparable precision in TCH (81.3% versus 77.9%; two-sided binomial test, \( p = 0.32 \)) and TFCH (81.1% versus 83.4%; two-sided binomial test, \( p = 0.52 \)) test sets.

**Performance of CRCNet at image level.** We also measured the classification performance of CRCNet at the image level on images of CRC patients from three test sets (see "Methods"). CRCNet achieved high performance in detecting consensus malignant images (Fig. 3): for the TCH test set, AUPRC was 0.990 (95% CI 0.987–0.993) and F1 was 95.6%; for TFCH test set,
Colonoscopic imaging data collected from Tianjin Cancer Hospital between August 2011 and March 2019, 12,179 individuals, 464,105 images. Non-CRC disease: 5050 individuals. Doctors directed control: 3953 individuals. CRC: 3176 individuals. Control: 9003 individuals, 436,034 images.

Training set: 12,179 individuals, 464,105 images.

CRCNet classification model.

Training and model development:

Aim: CRC vs. Non-CRC classification.

Fig. 1 Flowchart depicting the development and evaluation of CRCNet. a Model development consisted data curation and CRCNet training. b Evaluation of CRCNet on three test sets. c Comparison between CRCNet and five endoscopists on a subset of randomly selected cases. All CRC patients and 5050 control patients in the training set and all patients in three test sets have surgical specimen or biopsy for pathological evaluation.

Table 2 Classification metrics of CRCNet at the patient level.

| Performance metrics                  | The performance of CRCNet across three test sets |
|--------------------------------------|--------------------------------------------------|
|                                      | Tianjin Cancer Hospital (n = 363) | Tianjin First Central Hospital (n = 430) | Tianjin General Hospital (n = 1470) |
| Accuracy (95% CI)                    | 0.873 (0.835–0.906)                | 0.916 (0.886–0.941)                | 0.980 (0.972–0.987)                |
| Recall rate (95% CI)                 | 0.904 (0.864–0.947)                | 0.789 (0.690–0.868)                | 0.746 (0.629–0.842)                |
| Specificity (95% CI)                 | 0.853 (0.798–0.897)                | 0.950 (0.921–0.971)                | 0.992 (0.986–0.996)                |
| Precision (95% CI)                   | 0.805 (0.736–0.863)                | 0.807 (0.709–0.883)                | 0.828 (0.713–0.911)                |
| Negative predicted value (95% CI)    | 0.930 (0.885–0.961)                | 0.944 (0.915–0.966)                | 0.987 (0.980–0.992)                |
| Kappa$^a$                            | 0.742                              | 0.745                              | 0.775                              |
| $F_1$$^b$                            | 0.852                              | 0.798                              | 0.785                              |

$^a$Measures the agreement between predicted classification with pathological report.

$^b$Harmonic average of the precision and recall rate.
AUPRC was 0.991 (95% CI 0.987–0.995) and F1 was 96.3%; and for TGH, AUPRC was 0.997 (95% CI 0.995–0.999) and F1 was 97.3%. The other classification metrics such as accuracy, recall rate, specificity, precision, negative predictive value, and kappa coefficient were provided in Supplementary Table 3.

Visual explanation of decision made by CRCNet. We used Grad-CAM algorithm\(^{21}\) to identify image regions contributed the most to the prediction made by CRCNet. Representative examples of malignant colonoscopic images with accompanying saliency heatmaps highlighting features most influenced CRCNet prediction was shown in Fig. 4 and Supplementary Data 2. The Grad-CAM heatmaps of flat and sessile serrated polyps were provided as Supplementary Data 3 and 4. In addition, we asked five endoscopists to inspect 255 randomly selected colonoscopic images and their accompanying saliency heatmaps. The full list of these 255 images were provided in Supplementary Data 5. The percentage of these 255 saliency heatmaps for which these five endoscopists agreed that the heatmaps captured the regions of malignant lesions was 94.3% (95% CI: 91.4–97.1%). The accuracy assessed by each endoscopist was shown in Supplementary Table 4.

Discussion

Results from this study showed that CRCNet model achieved high precision and recall rates in identifying CRC patients as compared with a group of five skilled endoscopists. CRCNet achieved consistent and robust performance across three test sets with significant improvement in precision on two test sets and recall rate on one test set. At the image level, CRCNet performed satisfactorily in distinguishing between malignancy and benignity of colonic lesions. In addition, endoscopists found the evidence-based visual explanation derived from CRCNet useful for routine clinical practice.

Improvement in medical imaging technique and endoscopic classification system such as WASP have facilitated the optical diagnosis of benign lesions, however, optical diagnosis of malignant lesions is also an important clinical application and remains challenging. For example, invasive tumor often needs surgical treatment, whereas precancerous disease such as adenoma
requires endoscopic submucosal dissection. Therefore, it is important to devise new ways to aid endoscopists in differentiating benign lesions from malignant ones on white-light images, which is the most commonly used imaging modality in routine clinical practice. In addition, white-light imaging mode can avoid false positive caused by poor intestinal preparation under NBI to some extent.

In recent studies, deep learning has been widely applied in endoscopy. Previous studies have developed deep learning models on images from colonoscopy for detecting polyps or adenomas. Our study was dedicated to optical diagnosis of CRC instead of lesion detection as conducted by aforementioned studies. Optical diagnosis of CRC has previously been investigated by Mori et al. by training on image features extracted from endocytoscopy, which can provide morphological images of the nuclei and gland duct lumens that beyond the capability of colonoscopy. Takemura et al. developed HuPAS software for predicting histology of colorectal tumors from 1519 cut-out images from magnifying NBI images. To the best of our knowledge, we developed CRCNet for optical diagnosis of CRC using by far the largest number of colonoscopic images from standard white-light imaging modality. CRC is initially identified by endoscopists during colonoscopic examination based on tumor surface features. It is difficult to accurately recognize all malignant lesions and there is a low interendoscopists agreement rate, which was also reflected by the varying agreement rates among endoscopists in test sets in this study (range from 56.7 to 88.2%). This is likely due to increased difficulty in identifying noncancerous patients in TCH cohort because patients came to TCH are often suspected to be cancerous. This was reflected by higher proportion of CRC patients in test set from TCH (40.2%, n = 146/363) as compared with those from TFCH (20.9%, n = 90/430) and TGH (4.8%, n = 71/1470). On the contrary, CRCNet can make consistent interpretation once deployed. CRCNet has the potential to reduce the reliance on expertise of colonoscopists for optical diagnosis of CRC and improve diagnostic consistency. One advantage of our study was the inclusion of diverse noncancerous diseases (i.e., polyps, adenomas, chronic

Fig. 3 Performance of CRCNet in identifying malignant colonoscopic images. Precision-recall curves of CRCNet on a TCH test set, b TFCH test set, and c TGH test sets. Area under the precision-recall curve and associated 95% confidence intervals are included.

Fig. 4 Exemplified class activation maps. a Raw colonoscopic images. b Gradient-weighted class activation map. c Guided gradient-weighted class activation map. d Haematoxylin–eosin staining images with scale bars. The length of scale bar was shown above the bar. e Tumor location and TNM stage.
mucosal inflammation, and inflammatory bowel disease) in control group and employment of pathological examination as gold standard. Therefore, CRCNet could learn noncancerous features that often complicate CRC diagnosis, presumably increasing model performance and avoiding verification bias. Meanwhile, the performance of endoscopists in identifying CRC depends on their working experience and often varies considerably given that human-derived diagnostic systems are often established by expert consensus and subjected to human variation20, whereas CRCNet can provide objective optical diagnosis.

Another advantage of CRCNet is that it can report results immediately and consistently on the graphics processing unit thus facilitating real-time CRC detection and decreasing the workload, inconsistency, and misdiagnosis. In addition, CRCNet can overcome inherent limitations of endoscopists such as perceptual bias and visual fatigue17. CRCNet showed high degree of fidelity and uniformity in differentiating malignant lesions from benign lesions at image level, suggesting that it could serve as a second-read tool during colonoscopic examination. The proportion of trained patients with CRC is closely associated with pathological stage. The comparable prediction malignancy scores of CRC patients at early stage (I and II) and advanced stage (III and IV) (mean 0.679 versus 0.689; two-sided t-test, p = 0.843) suggested that CRCNet performs equally good in detecting CRC patients at early stage as those at advanced stage. For false positives with high prediction malignancy scores and also interpreted as CRC patients by a group of five endoscopists, it is likely due to incomplete biopsy during colonoscopy examination. This highlighted that CRCNet might complement endoscopists in optical diagnosis. Additional follow-up evaluation for these patients are required. Besides, the visual explanation derived from CRCNet can further provide evidence-based classification to assist endoscopists in interpreting the images. This can justify the predictions made by CRCNet, especially when there are unreliable predictions. However, the accuracy of CRCNet to pinpoint small tumors would decrease as Grad-CAM algorithm only performs well when the class being detected covers a big portion of the image.

Integration of CRCNet into colonoscopy interpretation system can help endoscopists speed up interpretation process. A second read from CRCNet could augment the capability of endoscopists to manage patients at high risk of CRC. Marginal improvement in CRCNet to discriminate between malignant and benign lesions could reduce unnecessary biopsy. Using the white-light imaging modality and not having to revert to NBI would be more clinically relevant and potentially more translational as it relies on images seen in routine colonoscopy. CRCNet can further expand the recognition of early colorectal lesions by white light, especially for community hospitals without NBI equipment. Although CRCNet was developed with white-light imaging data, it could to some extent correctly classify sessile polyps as benign, which is considered to be better recognized on NBI imaging mode. However, the aforementioned clinical diagnostic validity conferred by CRCNet requires further investigation in clinical trials.

There are several limitations of this study. Firstly, CRCNet was trained only on static images from a single medical center, it may not be able to capture all the real-world data variation in relation to different devices and skills of endoscopists. In daily clinical practice, optical diagnosis of colonoscopic lesions are evaluated in a real-time scenario and image quality would be affected by operational fluctuation. This will deteriorate the performance of CRCNet. On the other hand, it might be easier to characterize colonoscopic lesions in real-time colonoscopy as we are able to visualize lesions from multiple perspectives, thus potentially increasing the performance of optical diagnosis13. To overcome this limitation, we employed extensive data augmentation during model training, the impact of lacking training data from other centers could be mitigated and image variations were also increased. This was demonstrated by comparable, robust, and generalizable performance of CRCNet on external test sets. Results from Urban et al. demonstrated that deep learning model trained on static images could achieve high accuracy in real-time colonoscopy settings23. Secondly, the current model can only distinguish between malignancy and benignity. In this study, we did not include other rare malignant diseases observed by colonoscopy such as sarcoma, melanoma, gastrointestinal stromal tumor, lymphoma, and carcinosarcoma because of their limited number of cases. Future work will be focused on expanding our model to further classify whether the detected benign intracavity lesion is adenoma, hyperplastic polyp, or sessile serrated polyps/adenomas. Thirdly, the enrolled patients in training set and test sets were mainly northern Chinese, the performance of CRCNet on other ethnic groups remains to be tested. Fourthly, the improvement of clinical outcomes brought by CRCNet remains to be determined as this is a retrospective study. In addition, we did not consider other types of data, for example, age, sex, family history, and tumor site, beyond colonoscopic images.

Although CRCNet could achieve satisfactory performance, it does not necessarily mean improved clinical outcomes. Class imbalance is a common challenge in medical area. It is difficult to obtain positive samples in that these samples were under-represented in real-world scenario and imaging database. The ratio of positive samples to negative samples was 6.4% in training set. We addressed this issue by using focal loss for training. The focal loss focuses training on hard-classified examples while prevents the well-classified negative examples from overwhelmingly affecting the model26.

Future prospective studies should be conducted to compare clinical efficacy of colonoscopy with or without assistance of CRCNet. Given the unbalanced medical resource in many countries including China, such a model may benefit community hospitals in rural areas. In the future, we intend to associate pathological findings such as the depth of tumor invasion, vascular thrombus, perineural invasion, lymph node, and distant metastasis with features of extracted colonoscopic images, in order to predict the preoperative CRC staging and aid selection of alternative treatments.

In summary, CRCNet can achieve high performance in differentiating CRC from benign diseases such as adenomas and polyps. CRCNet achieved consistent and robust performance in identifying CRC patients across three test sets. Its performance was comparable with a group of five skilled endoscopists. Prospective randomized clinical trials are required to test the performance of this model in real-world clinical settings.

Methods

Study design and participants. We did retrospective, multicohort, diagnostic study using colonoscopic images from three tertiary hospitals in China. We retrieved colonoscopic images from Medical Imaging Database as training set at TCH between August 2011 and March 2019. We used images from patients who received colonoscopic examination between April 2019 and May 2019 at TCH as internal test sets, January 2018 and February 2019 at TCH as first external test set, and January 2018 and December 2018 at TCH as the second external test set. All images and pathological examination reports were deidentified before they were transferred to investigators. All CRC patients and 56.1% (5050/9003) of controls in the training set and all patients in three test sets underwent surgical resection or endoscopy-guided biopsy; therefore, they had pathological examination as gold standard to diagnose CRC. The training set and test sets were obtained from real-world colonoscopic imaging cohort. The control group included patients diagnosed with normal mucosa, adenoma, polyps, familial adenomatous polyposis syndrome, inflammatory bowel disease, and chronic mucosal inflammation. The CRC patient group consisted of patients diagnosed with solitary tumor. We excluded other rare malignant diseases such as sarcoma, melanoma, gastrointestinal stromal tumor, lymphoma, and carcinosarcoma. All patients included in this study were 18 years of age or older. Surgically resected CRC tumors were staged according to the 7th
We used a random subset of images, which was not reported in the paper and its supplementary information.

The authors declare that the data supporting the findings of this study are available in the Nature Research Reporting Summary linked to this article.

The data availability section of the paper states: The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Restrictions apply to the availability of the training and test sets, which were used with permission for the current study, and so are not publicly available. Databases used include Colonoscopic Imaging Databases and ImageNet (http://www.image-net.org/).

The code used to train and evaluate the model is available on GitHub (https://github.com/xioliangchun/AIplus/tree/master/CRCNet).

Received: 20 January 2020; Accepted: 19 May 2020; Published online: 11 June 2020

References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424 (2018).
2. Regula, J. et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. N. Engl. J. Med. 355, 1863–1872 (2006).
3. Lieberman, D. A. et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. N. Engl. J. Med. 343, 162–168 (2000).
4. Citarda, F., Tomasselli, G., Capocaccia, R., Barcherini, S. & Crespi, M. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. Gut 48, 812–815 (2001).
5. Zauber, A. G. et al. Colonoscopic Polypectomy and long-term prevention of colorectal cancer deaths. Obstet. Gynecol. Surv. 61, 353–356 (2012).
6. Aitken, W. S. et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 375, 1624–1633 (2010).
7. Singh, H. et al. The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer. Gastroenterology 139, 1128–1137 (2010).
8. Schoen, R. E. et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. N. Engl. J. Med. 366, 2345–2357 (2012).

9. Neerinck, M. et al. Colonic work-up after incomplete colonoscopy: significant new findings during follow-up. Endoscopy 42, 730–735 (2010).

10. Bick, B. L., Vemulapalli, K. C. & Rex, D. K. Regional center for complex colonoscopy: yield of neoplasia in patients with prior incomplete colonoscopy. Gastrointest. Endosc. 83, 1239–1244 (2016).

11. Hermans, C., van der Zee, D. & Gijssen, I. Double-balloon endoscopy after incomplete colonoscopy and its comparison with computed tomography colonography. Clin. Endosc. 51, 66–71 (2018).

12. National Institute for Health and Care Excellence. Virtual chromoendoscopy to assess colorectal polyps during colonoscopy. https://www.nice.org.uk/guidance/dg28 (2017).

13. IJspeert, J. E. G. et al. Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. Gut 65, 963–970 (2016).

14. Siegel, R. et al. Cancer treatment and survivorship statistics, 2012. CA Cancer J. Clin. 62, 220–241 (2012).

15. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90 (2017).

16. Li, X. et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. Lancet Oncol. 20, 193–201 (2019).

17. Chen, P. J. et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 154, 568–575 (2018).

18. Misawa, M. et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. Gastroenterology 154, 2027–2029 (2018).

19. Luo, H. et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol. 2045, 1–10 (2019).

20. Artifical intelligence and computer-aided diagnosis in colorectal cancer: current evidence and future directions. Lancet Gastroenterol. Hepatol. 4, 71–80 (2019).

21. Selvaraju, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. IEEE Int. Conf. Comput. Vis. 40, 618–626 (2017).

22. Wang, P. et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat. Biomed. Eng. 2, 741–748 (2018).

23. Mori, Y. et al. Impact of an automated system for endoscopic diagnosis of small colorectal lesions: an international web-based study. Endoscopy 48, 1110–1118 (2016).

24. Takemura, Y. et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). Gastrointest. Endosc. 75, 179–185 (2012).

25. Urban, G. et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 155, 1069–1078 (2018).

26. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 318–327 (2017).

27. Huang, G., Liu, Z., Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. IEEE Conf. Comput. Vis. Pattern Recognit. 39, 2261–2269 (2017).

28. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252 (2015).

29. Grau, J., Grosse, I. & Keilwagen, J. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics 31, 2595–2597 (2015).

Acknowledgements

The work was supported by the National Natural Science Foundation of China (no. 31801117 to X.L. and no. 81903055 to X.T.) and IRT_14B40 from the Program for Changjiang Scholars and Innovative Research Team in University in China (K.C.). W.Z. is supported by a fellowship from the National Foundation for Cancer Research and a Hanes and Wills Family endowed professorship in cancer at the Wake Forest Baptist Comprehensive Cancer Center. W.Z. is supported by the Cancer Center Support Grant from the National Cancer Institute to the Comprehensive Cancer Center at Wake Forest Baptist Medical Center (P30 CA012197).

Author contributions

X.L. and K.C. designed and supervised the study; X.L., D.Z., F.T., and X.T. performed data analysis and wrote the manuscript; D.Z., X.T., B.W., G.Z., and W.Z. read and interpreted colonoscopic images; L.S., X.G., and B.S. curated pathological data; X.H., F.Z., Z.C., M.Y., Y.Y., Q.Z., Z.L., J.L., and J.W. reviewed and separated images; and X.L., F.T., and W.Z. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16777-6.

Correspondence and requests for materials should be addressed to K.C. or X.L.

Peer review information Nature Communications thanks Faisal Mahmood, Ulas Bagci and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.