

RESEARCH ARTICLE

AI based colorectal disease detection using real-time screening colonoscopy

Jiawei Jiang^{1,2,§}, Qianrong Xie^{1,§}, Zhuo Cheng^{3,§}, Jianqiang Cai^{4,§}, Tian Xia^{5,§}, Hang Yang¹, Bo Yang³, Hui Peng⁶, Xuesong Bai³, Mingque Yan³, Xue Li¹, Jun Zhou¹, Xuan Huang⁷, Liang Wang⁸, Haiyan Long⁹, Pingxi Wang¹, Yanpeng Chu¹, Fan-Wei Zeng¹, Xiuqin Zhang^{10,*}, Guangyu Wang^{11,*} and Fanxin Zeng^{1,12,*}

¹Department of Clinical Research Center, Dazhou Central Hospital, Dazhou 635000, China

²Department of Computer Science, Eidgenössische Technische Hochschule Zurich, Zurich 999034, Switzerland

³Digestive endoscopy center, Dazhou Central Hospital, Dazhou 635000, China

⁴Department of Hepatobiliary Surgery, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

⁵National Center of Biomedical Analysis, Beijing 100850, China

⁶College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

⁷Department of Ophthalmology, Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China

⁸Information Department, Dazhou Central Hospital, Dazhou 635000, China

⁹Digestive endoscopy center, Quxian People's Hospital, Dazhou 635000, China

¹⁰Institute of Molecular Medicine, Peking University, Beijing 100871, China

¹¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

¹²Department of Medicine, Sichuan University of Arts and Science, Dazhou 635000, China

*Correspondence to: Fanxin Zeng, zengfx@pku.edu.cn; Guangyu Wang, guangyu.wang24@gmail.com; Xiuqin Zhang, zhangxq@pku.edu.cn

§These authors contribute equally to this work.

Abstract

Colonoscopy is an effective tool for early screening of colorectal diseases. However, the application of colonoscopy in distinguishing different intestinal diseases still faces great challenges of efficiency and accuracy. Here we constructed and evaluated a deep convolution neural network (CNN) model based on 117 055 images from 16 004 individuals, which achieved a high accuracy of 0.933 in the validation dataset in identifying patients with polyp, colitis, colorectal cancer (CRC) from normal. The proposed approach was further validated

Received: 24 March 2021; Revised: 23 April 2021; Accepted: 17 May 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of the West China School of Medicine & West China Hospital of Sichuan University. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

on multi-center real-time colonoscopy videos and images, which achieved accurate diagnostic performance on detecting colorectal diseases with high accuracy and precision to generalize across external validation datasets. The diagnostic performance of the model was further compared to the skilled endoscopists and the novices. In addition, our model has potential in diagnosis of adenomatous polyp and hyperplastic polyp with an area under the receiver operating characteristic curve of 0.975. Our proposed CNN models have potential in assisting clinicians in making clinical decisions with efficiency during application.

Key words: artificial intelligence (AI); colorectal disease; real-time colonoscopy

Introduction

As a noninvasive examination, colonoscopy is the main method to screen colorectal diseases, including polyp, colitis, and colorectal cancer (CRC).¹⁻³ Previous studies have shown that early intestinal diseases can develop into advanced CRC without timely treatment, leading to a high mortality.⁴⁻⁶ Adenomatous polyps are the main precursor of CRC, and the detection and treatment of them could help to reduce the risk of CRC.⁷⁻⁹ However, the diagnosis of intestinal diseases is time-consuming and approximately 26% of neoplastic diminutive polyps and 3.5% CRC are missed in a single colonoscopy examination.^{10,11} Although several advanced imaging systems, such as narrow-band imaging, are developed to reduce the miss rates, the risk of missing suspicious lesion in colonoscopy examinations still exists.^{12,13} Moreover, the effective diagnosis of intestinal diseases is an interminable and time-consuming task. Therefore, the development of an AI tool for diagnosis and treatment of intestinal diseases with accuracy and efficiency is urgently needed, which are helpful to reduce the risk of CRC.

In recent years, artificial intelligence (AI) has been successfully applied throughout human life, from embryo selection for *in vitro* fertilization to mortality prediction.¹⁴⁻¹⁶ AI is also widely used in the field of medical imaging, including diagnosis of breast ultrasonography images, pathologic type classification of esophageal cancer, retinopathy classification, and COVID-19 pneumonia identification.¹⁷⁻²⁰ In gastroscopy, AI is mainly used for the detection of early cancer and the discrimination of pathological characteristics, such as gastric cancer and esophageal cancer.^{17,21} Previous studies applying AI techniques in colonoscopy have mainly focused on polyp detection, segmentation and pathological classification.²²⁻²⁴ However, with the complexity of colorectal diseases, it is challenging for these models to identify a single disease from normal cases during clinical application. Additionally, the small dataset and single-center study also limits the clinical applicability of the AI system.

By this research, we proposed an AI system to classify colorectal diseases including the polyp, colitis, CRC and normal, using real-world colonoscopy images collected from multi-center. The performance of the AI model was further tested by external validation datasets and real time diagnostic setting. In addition, we compared the performance of the AI system with novice and expert endoscopists. Finally, the performance of our

AI system in discriminating sub-polyp diseases of hyperplastic polyp and adenomatous polyp was also validated.

Results

Dataset characteristics and CNN architectures

We retrospectively obtained 117 055 colonoscopy images (including 84 605 normal, 20 537 with polyp, 7963 with colitis and 3950 with CRC) from 16 004 individuals (Fig. 1 and Supplementary Fig. S1). The characteristics of patients for each cohort are shown in Supplementary Table S1. For detection of colorectal diseases, the images in the primary cohort were randomly divided into training and validation datasets with a ratio of 8:2 (training dataset: 24 978 normal images, 10 255 polyp images, 3712 colitis images and 2170 CRC images; validation dataset: 6362 normal images, 2622 polyp images, 908 colitis images and 545 CRC images) (Fig. 1). The typical images of normal, polyp, colitis and CRC are shown in Supplementary Fig. S2. Our model (convolutional neural network 1 (CNN1)) was constructed by ResNet-50²⁵⁻²⁹ to identify colorectal diseases using transfer learning (Supplementary Fig. S3).

Performance comparison of CNN models

We first validated the CNN1 model of multi-class classification accuracy on the validation dataset. The model showed accurate performance by detecting 97.51% normal images, 91.08% polyp images, 79.41% colitis images and 79.45% CRC images with an area under the curve (AUC) of 0.983 in the validation dataset (Figs. 2A and 2B). The model achieved a precision of 0.933 (95% confidence interval (CI): 0.928, 0.938), a recall of 0.934 (95% CI: 0.933, 0.939), an F1-score of 0.933 (95% CI: 0.928, 0.938), an accuracy of 0.934 (95% CI: 0.929, 0.939) and a positive predictive value (PPV) of 0.869 (95% CI: 0.859, 0.880) (Table 1 and Supplementary Table S2). In the independent test dataset, the model showed an accuracy of 0.836, and 84% images were correctly identified, with an AUC of 0.963, a precision of 0.898 (95% CI: 0.895, 0.901), a recall of 0.836 (95% CI: 0.832, 0.840), an F1-score of 0.856 (95% CI: 0.853, 0.859) and a PPV of 0.771 (95% CI: 0.761, 0.780) (Figs. 2C and 2D, Table 1 and Supplementary Table S3). Similarly, the AI model obtained high accuracy of 0.965 in external validation dataset (Table 1). The CNN1 model achieved a diagnostic performance with AUC of 0.940, precision of 0.968 (95% CI: 0.966, 0.970), recall of 0.965 (95% CI: 0.963,

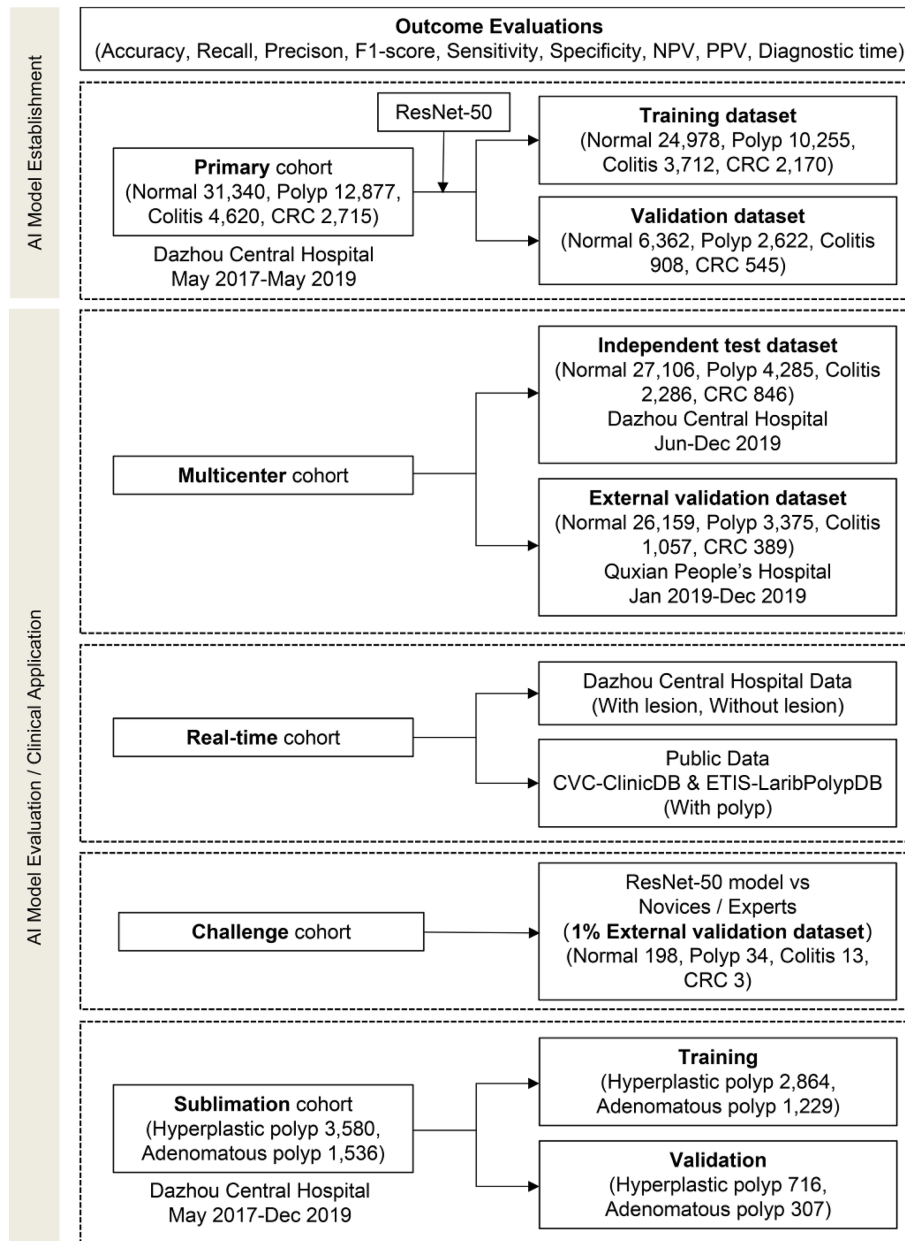


Figure 1. Illustration of the study dataset. NPV, negative predictive value; PPV, positive predictive value; AI, artificial intelligence.

0.967), F1-score: 0.965 (95% CI: 0.963, 0.967) and PPV of 0.923 (0.915, 0.930) (Figs. 2E and 2F, Table 1 and Supplementary Table S4).

Furthermore, we constructed other two CNN models based on VGG16 algorithm (CNN2), VGG19 algorithm (CNN3) in classifying normal, polyp, colitis and CRC. Confusion matrices of classification performance of CNN2 and CNN3 on validation, independent test and external validation dataset are shown in Supplementary Figs. S5 and S6. CNN2 and CNN3 showed high precision, recall, F1-score and accuracy in identifying the colorectal diseases. Compared with CNN2 and CNN3 models, CNN1 model showed a better performance in multi-class diagnosis with accuracy of 0.934 (95% CI: 0.929, 0.939) in the validation dataset (Supplementary Table S5). Taken

together, the results demonstrated that CNN1 had better performance compared with CNN2 and CNN3.

Binary classifiers were also implemented to investigate the model's detailed classification performance of distinguishing polyp, colitis, and CRC from normal (Supplementary Figs. S7–S9). The CNN1-NPo model correctly distinguished 6299 of 6362 normal images and 2447 of 2622 polyp images, with an area under the curve (AUC) of 0.992 and an accuracy of 0.966 (Supplementary Figs. S7C and S7D, Supplementary Table S6). CNN1-NCo correctly distinguished 6334 of 6362 normal images and 640 of 908 colitis images with an AUC of 0.974 (Supplementary Figs. S8C and S8D). CNN1-NCa correctly classified 6360 of 6362 normal images and 480 of 545 CRC images yielding an AUC of 0.995 (Supplementary Figs. S9C and S9D).

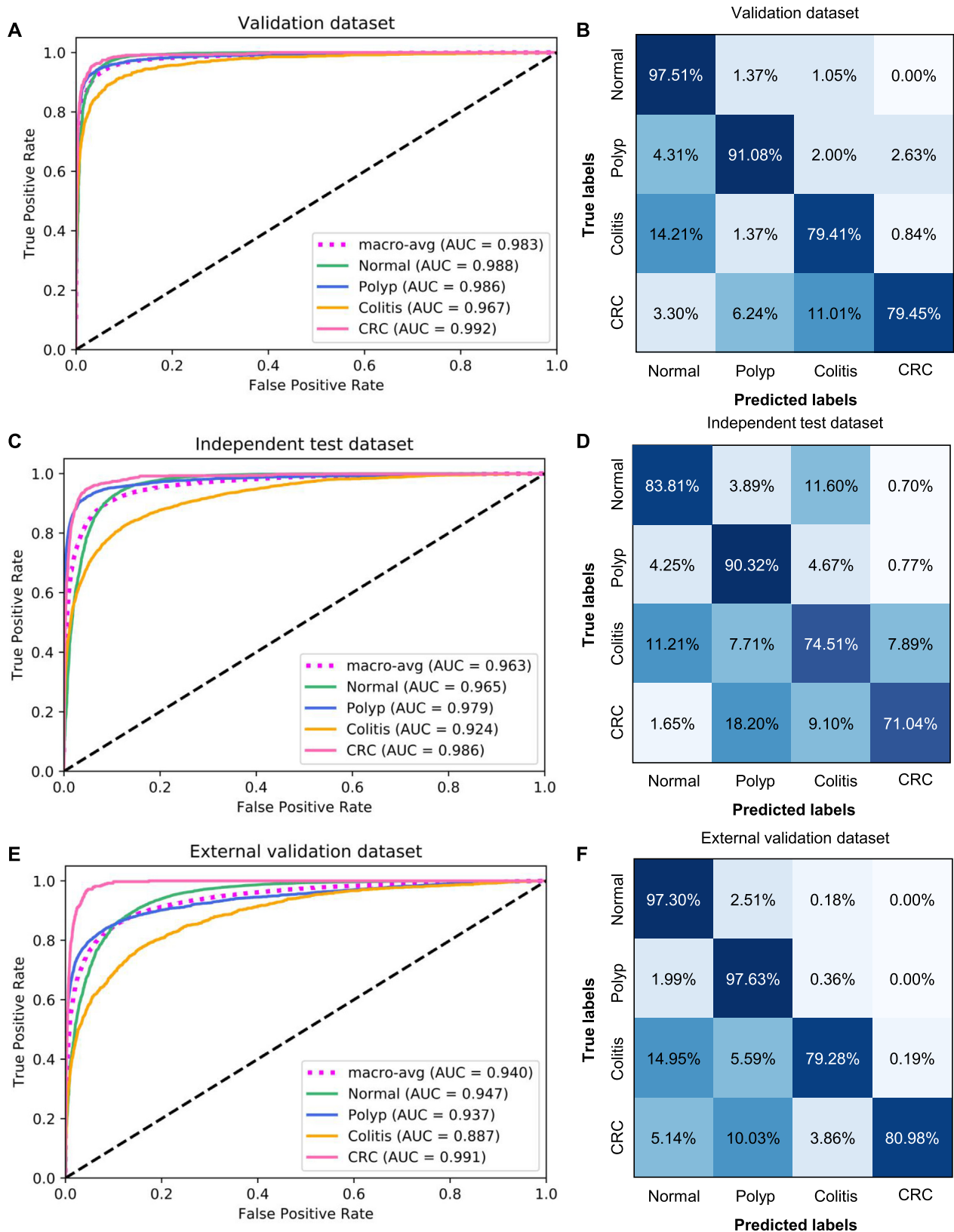


Figure 2. Diagnostic performance of AI model in multi-class classification in multi-center cohort. (A) ROC of CNN1 model in the validation dataset. (B) Confusion matrix of CNN1 model in the validation dataset. (C) ROC of CNN1 model in the independent test dataset. (D) Confusion matrix of CNN1 model in the independent test dataset. (E) ROC of CNN1 model in the external validation dataset. (F) Confusion matrix of CNN1 model in the external validation dataset.

Table 1. Performance of the CNN1 model in different validation and test dataset.

	Validation dataset	Independent test dataset	External validation dataset	P-value
Precision	0.933 (0.928, 0.938)	0.898 (0.895, 0.901)	0.968 (0.966, 0.970)	<0.001
Recall	0.934 (0.933, 0.939)	0.836 (0.832, 0.840)	0.965 (0.963, 0.967)	<0.01
F1-score	0.933 (0.928, 0.938)	0.856 (0.853, 0.859)	0.965 (0.963, 0.967)	0.001
Accuracy	0.934 (0.929, 0.939)	0.836 (0.832, 0.840)	0.965 (0.963, 0.967)	<0.01
Positive predictive value	0.869 (0.859, 0.880)	0.771 (0.761, 0.780)	0.923 (0.915, 0.930)	<0.01
Negative predictive value	0.975 (0.971, 0.979)	0.956 (0.954, 0.958)	0.973 (0.971, 0.975)	<0.001

95% confidence intervals are included in brackets.

Table 2. Performance of the CNN1-Po model in video dataset.

Data sources	Video dataset	Accuracy
Dazhou Central Hospital	No-polyp video 1	0.932 (0.922, 0.938)
	No-polyp video 2	0.902 (0.890, 0.910)
	Polyp video 1	0.957 (0.898, 1.000)
	Polyp video 2	1.000 (1.000, 1.000)
Public Data	CVC-ClinicDB	0.969 (0.955, 0.983)
	ETIS-LaribPolypDB	0.821 (0.768, 0.875)

95% confidence intervals are included in brackets.

Real-time diagnosis of the CNN model

Moreover, the performance of the AI model was tested by real time diagnostic procedure. Commonly, video is encoded from 25 to 30 frames per second (fps). In our study, all the CNN1 models could realize real-time lesion detection (0.004s/frame), which was higher than the standard frame rate (24 fps) and could meet the requirement of real-time assistance software. Furthermore, we used the colonoscopy videos from Dazhou Central Hospital and the international public dataset of CVC-ClinicDB and ETIS-LaribPolypDB to explore the real-time detection performance of CNN1-NPo in the real world. In the two full-length without polyp videos, CNN1-NPo had an accuracy of 0.932 (95% CI: 0.922–0.938) and 0.902 (95% CI: 0.890–0.910), respectively. Similarly, CNN1-NPo showed excellent diagnostic ability in two polyp videos, with an accuracy of 0.957 (95% CI: 0.898–1.000) and 1.000 (95% CI: 1.000–1.000), respectively. The accuracy of CNN1-NPo in CVC-ClinicDB and ETIS-LaribPolypDB was 0.969 (95% CI: 0.955–0.983) and 0.821 (95% CI: 0.768–0.875), respectively (Table 2).

Performance comparison between the AI model and endoscopists

We further compared the AI system with endoscopists in identifying colorectal diseases. We randomly selected 1% colonoscopy images from external validation dataset to compare the diagnostic performance of CNN1 and endoscopists. We employed endoscopists in three groups: three novices; four endoscopists in the junior group with less than 5 years of experience; and six endoscopists in the senior group with more than 5 years of clinical experience. CNN1 achieved an accuracy of 0.956, the performance of which was superior to the novices and

junior group with 0.797 and 0.833, respectively. Meanwhile, our CNN model was not inferior to experts from senior groups with an average accuracy of 0.903. Compared with experts, the novices without the assistant of CNN model had a relatively poor performance in identifying the colorectal diseases (Fig. 3). Furthermore, the accuracy of the novices and junior groups could be significantly improved with the assistant of the CNN model ($P < 0.001$) (Fig. 3). Taken together, these findings demonstrated the feasibility that CNN could assist endoscopists to detect colorectal diseases to achieve more accurate diagnosis.

Application of the AI model in polyp subgroup classification

The detection and removal of neoplastic polyp is helpful to decrease the risk of CRC.⁷ Hence, the diagnosis of the adenomatous polyp is essential for precision treatment. Here, we constructed a CNN1-HA model to perform polyp subgroup classification for the discriminating of hyperplastic and adenomatous polyp for precision treatment. A total of 5116 polyp images (including 3580 images with hyperplastic polyp and 1536 images with adenomatous polyp) from the primary cohort were collected and labeled to develop the CNN1-HA model (Supplementary Fig. S10). In the validation dataset, CNN1-HA obtained a recall of 0.942 (95% CI: 0.925, 0.959) and 0.883 (95% CI: 0.847, 0.919) to detect hyperplastic polyp and adenomatous polyp, respectively. CNN1-HA achieved 0.925 (95% CI: 0.909, 0.941) precision, 0.924 (95% CI: 0.908, 0.940) recall, 0.924 (95% CI: 0.908, 0.940) F1-score, 0.924 (95% CI: 0.908, 0.940) accuracy and an AUC of 0.975 (95% CI: 0.963, 0.987) in classified hyperplastic polyp and adenomatous polyp (Supplementary Fig. S11 and Table 3).

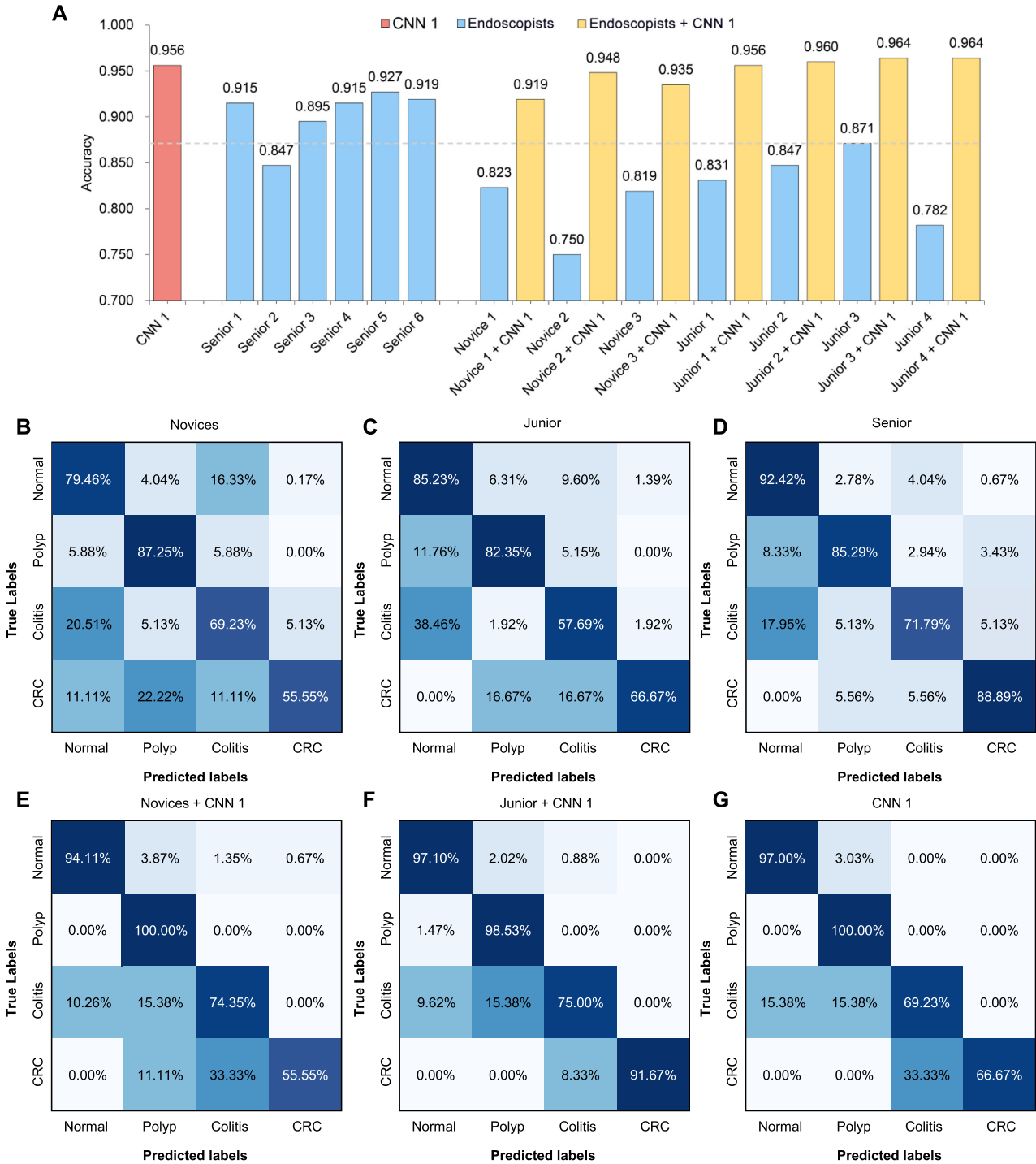


Figure 3. Comparison of diagnostic performance between AI system and endoscopists. (A) Accuracy of AI system and endoscopists. Confusion matrix of the average diagnostic performance of novices (B), junior (C), senior (D), novices with AI assistance (E), junior with AI assistance (F) and CNN 1 (G).

Discussion

In this study, a deep learning algorithm was used to construct a CNN model to diagnose the colorectal diseases. It was developed and evaluated using 117 055 colonoscopy images for endoscopic diagnosis of colorectal diseases from 16 004 individuals at multiple

institutions. The CNN model demonstrated a high precision, recall, F1-score and accuracy of diagnosing the colorectal diseases in multi-center dataset and realized real-time diagnosis. The high accuracy was superior to the novices and not inferior to the experts. Moreover, the CNN model had potential in distinguishing the diminutive hyperplastic polyp and adenomatous polyp images.

Table 3. Performance of the CNN1-HA model in polyp subgroup.

Datasets	Support images	Precision	Recall	F1-score	Accuracy	AUC
Adenoma	307	0.883 (0.847, 0.919)	0.863 (0.825, 0.901)	0.873 (0.836, 0.910)	0.925 (0.896, 0.954)	0.975 (0.963, 0.987)
Hyperplastic	716	0.942 (0.925, 0.959)	0.950 (0.934, 0.966)	0.946 (0.929, 0.963)	0.924 (0.905, 0.943)	
All	1023	0.925 (0.909, 0.941)	0.924 (0.908, 0.940)	0.924 (0.908, 0.940)	0.924 (0.908, 0.940)	

95% confidence intervals are included in brackets.

CNN1-HA model represents the ResNet-50 classifying the hyperplastic and adenomatous polyp.

Abbreviations: AUC, area under the curve.

The application of deep learning in colonoscopy mainly focuses on the field of colonic polyp and ulcerative colitis.^{30–32} These models face challenges to translate into routine clinical application due to the binary classification, single center development, or small sample validation. With the complexity of colorectal diseases, multi-class classification is essential for clinical decision. In this study, our CNN model exhibited satisfactory accuracy (from 0.836 to 0.965) in discriminating normal, polyp, colitis and CRC in different centers. To the best of our knowledge, this is the first and largest study applying AI to identify multiple colorectal diseases. Furthermore, the CNN model could reduce the risk of missing inspection of colorectal diseases as a result of its high recall, leading to earlier lesion detection and precision treatment. The detection and removal of adenomatous polyp are beneficial for reducing the risk of the colorectal cancer.³³ Therefore, the diagnosis of the adenomatous polyp is essential for precision treatment. Significantly, our AI system accurately discriminated hyperplastic polyp and adenomatous polyp preoperatively with a high accuracy of 0.924.

We further analyzed the cases of false positive and false negative in using the AI model. Consistent with previous studies,^{10,34} the false detection performance was mainly caused by the poor quality of bowel preparation, the small, flat structure, intestinal air inflation and partially behind a fold. In addition, the false positive rate was mainly due to the false recognition of normal fold mucosa.

In our study, ResNet-50 CNN was used to extract the features of each input frame for the multi-class classification task. Since deep residual learning was proposed, it has been widely used in the field of image recognition for melanoma, pediatric pneumonia, and urine sediment recognition.^{25,26,35,36} By employing residual network algorithm, our model avoided gradient dissipation and overcame the problem of reduced accuracy with the increase of depth. Moreover, the diagnostic performance of our CNN model based on ResNet-50 was slightly higher than VGG16 and VGG19.

The CNN model was able to detect colorectal diseases with a speed of 0.004 seconds per frame, which met the requirement of real-time assistance software. Consistent with previous study, the diagnostic speed of our model was significantly faster than that of novices and experts.²⁴ Meanwhile, our CNN model was superior to the novices and experts with less than 5 years of experience in the diagnosis of colorectal diseases, and certainly

not inferior to experts with over 5 years of experience. Furthermore, the accuracy of the novices and experts with less than 5 years of experience with the assistant of the CNN model could be significantly improved ($P < 0.001$). These findings demonstrated the feasibility that CNN could support nonexpert endoscopists to detect colorectal diseases and achieve more accurate diagnosis. To a certain extent, our model is helpful to solve the problem of unbalanced medical resources in grass-roots hospitals and community hospitals.

Despite these remarkable results, there are some limitations in our study. Firstly, this study was validated in only one real world external validation dataset. More public data should be collected to further identify the universality of our model. A further multi-center study could be designed to optimize the performance of our model in multi-class classification and improve the generalizability. Secondly, the colonoscopy images were produced by a single type of instrument (Olympus). However, our model can also be applied to other brands of endoscopes to achieve clinical application according to the application of previous transfer learning theory in other types of images.²⁰ Thirdly, the performance of accurate detecting of CRC was lower than benign diseases, probably due to the less sufficient images of cancer compared to other intestinal diseases. As the identification of CRC is much valuable during clinical application, we would include more cancer images into our study to improve its diagnostic performance in the future. In conclusion, we demonstrated that the CNN model constructed by colonoscopy images from multiple centers could achieve high accuracy in real-time distinguishing normal, polyp, colitis, and CRC. The newly developed CNN model exhibited an excellent diagnostic performance comparable to that of expert endoscopists and superior to novices and could improve the accuracy of nonexperts to a level similar to that of an expert. The CNN model also showed potential ability in diagnosing diminutive hyperplastic polyp and adenomatous polyp to assist endoscopists to make precision clinical diagnosis and treatment decisions.

Materials and methods

Patients and images

The images in the primary cohort (normal 31 340, polyp 12 877, colitis 4620 and CRC 2715) used to establish the

CNN model were from the routine colonoscopy examination (Olympus EVIS LUCERA CV260 SL/CV290 SL, CV-V1 and V70; Olympus Corp; Tokyo; Japan) between May 2017 and May 2019 in the Digestive Endoscopy Center of Dazhou Central Hospital (Fig. 1 and Supplementary Fig. S1). The typical images of normal, polyp, colitis and CRC are shown in Supplementary Fig. S2. The primary cohort images were randomly divided into training and validation dataset with the ratio of 8:2. The independent test dataset included the images from Dazhou Central Hospital between June 2019 and December 2019. The external validation dataset from Quxian People's Hospital between January 2019 and December 2019 were used to prove the universality and generalizability of the model. The 1% images from external validation dataset were randomly selected to compare the performance of model with endoscopists and demonstrate the auxiliary effect of our model on novices. A total of 30 980 images were acquired using Olympus EVIS LUCERA CV260 from Quxian People's Hospital. To validate the real-time detection ability of our model, we used 2 videos without polyp in full length and 2 videos with polyps obtained from patients who underwent colonoscopy examination in August 2019 from Dazhou Central Hospital, besides the public data from CVC-ClinicDB and ETIS-LaribPolypDB (<https://polyp.grand-challenge.org/Databases/>). Two no-polyp videos were identified by endoscopists, with 4060 and 3360 still images being output at 24 frames per second, respectively. The polyp frames from the first encounter to the last were recorded and output in the videos. The CVC-ClinicDB includes 612 still images from 29 different sequences acquired from Hospital Clinic of Barcelona, Barcelona, Spain. The CRC was identified by the biopsy. Before inputting, these images were confirmed by two experts with over 15 years of clinical experience, combined with clinical examination. This study was approved by the Ethics Committee of Dazhou Central Hospital (IRB00000002-20002) and was conducted at the hospital's Clinical Research Center and Digestive Endoscopy Center. The Ethics Committee waived the need for patients to sign informed consent.

CNN models

To construct the CNN model, the images in the primary cohort were randomly divided into training (80%) and validation (20%) dataset. The images with fecal residue, numerous air bubbles and blurred images were excluded. Before inputting the images, all images were pre-treated through encoding the label of text, performing horizontal flipping, and normalizing the image size (224x224). An open source of software library TensorFlow (version 2.0) was used to train the CNN model.^{36,37} We constructed our CNN model based on the transfer learning technique pretrained with ImageNet dataset. In our study, ResNet-50 CNN was used for features learning from each input frame and realized the multi-class classification. Since the deep residual learning was proposed, it has been widely used in the field of image recognition.¹⁹ By

employing the residual network algorithm, our model avoided gradient dissipation and overcame the problem of reducing accuracy due to the increase of depth. Moreover, the diagnostic performance of our CNN model based on ResNet-50 was slightly higher than the VGG16 and VGG19.

Like other deep learning models, CNN1 based on the ResNet-50 algorithm included convolution layer, pool layer and full connection layer (Supplementary Fig. S3A). However, based on the principle of residual learning, it recasted the original mapping into $F(X) + X$, which could skip one or more layers (Supplementary Fig. S3B). CNN1 was established after 18 444 training steps. This model was run on a TITAN X GPU. After 10 epochs, the model's training steps stopped without further improvement of the cross-entropy loss and accuracy (Supplementary Fig. S4). To compare the performance of our CNN1 model, we also used the off-the-shelf VGG16 (CNN2) and VGG19 (CNN3) fine-tuned with our colonoscopy image dataset to classify the images.³⁸ The polyp subgroup differentiation model was also constructed by classifying the hyperplasia and adenomatous polyp. All the polyps were confirmed by pathology after biopsy. The names and training steps for all CNN models are summarized in Supplementary Table S7.

Quantification and statistical analysis

The accuracy and AUC were the major outcome measures of diagnostic performance. As the common indexes of model's performance, precision, recall and F1-score were also used to evaluate the model. In addition to AUC, other metrics were calculated by true positive (TP), true negative (TN), false positive (FP) and false negative (FN). If the model labeled a positive image as positive, we considered the recognition of this image was TP. Instead, FP meant negative image being identified as positive. Similarly, TN and FN indicated labeling as negative for actually negative and actually positive images, respectively.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall}$$

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

We evaluated model performance by comparing with the novices and experts. Four experts with less than 5 years of colonoscopy experience, six experts with over 5 years of colonoscopy experience, and three novices

(medical students studying endoscopic image recognition), were asked to classify the images, being blinded to classification data. We randomly selected 1% of the external validation dataset (normal 198, polyp 34, colitis 13, CRC 3) to compare the performance of the model and endoscopists.

Supplementary data

Supplementary data is available at [PCMED](#) online.

Author contributions

J.J. and H.P. conducted the algorithm. Q.X., J.C. and X.L. drafted the manuscript. Z.C., T.X., H.Y., B.Y., X.B., M.Y., X.L., Z.J., X.H., L.W., and H.L. collected and labeled the images. Q.X., Z.C., T.X., X.B., M.Y., X.L., X.H., L.W., and H.L. organized the images. P.W., Y.C., F.-W.Z., X.Z., and G.W. revised the manuscript. F.Z. designed the study, revised the manuscript and the figure. All authors approved the final draft.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (Grant No. 81902861 to F.Z. and 32000485 to X.H.), “Xinglin Scholars” Scientific Research Project Fund of Chengdu University of Traditional Chinese Medicine (Grant No. YYZX2019012 to F.Z.), the Scientific Research Fund of Technology Bureau in Dazhou (Grant No. 17YYJC0004 to F.-W.Z.), the Key Research and Development Project Fund of Science and Technology Bureau in Dazhou, Sichuan Province (Grant No. 20ZDYF0001 to F.-W.Z.). We express our deepest appreciation to J.Z, Y.C. for organizing the raw data and G.Y. for the revising the manuscript.

Conflict of interest

All authors declare no competing interests of this study.

References

- Muller B, Voigtsberger P, Scheibner K, et al. Results of colonoscopy in surgically relevant diseases of the large intestine. (Article in German). *Zentralbl Chir* 1986;**111**:1091–9.
- US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for colorectal cancer: US preventive services task force recommendation statement. *JAMA* 2016;**315**:2564–75. doi:10.1001/jama.2016.5989.
- Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US multi-society task force on colorectal cancer and the American Cancer Society. *CA Cancer J Clin* 2006;**56**:143–59. doi:10.3322/canjclin.56.3.143.
- Nowacki TM, Brückner M, Eveslage M, et al. The risk of colorectal cancer in patients with ulcerative colitis. *Dig Dis Sci* 2015;**60**:492–501. doi:10.1007/s10620-014-3373-2.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;**61**:759–67. doi:10.1016/0092-8674(90)90186-i.
- Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424. doi:10.3322/caac.21492.
- Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 1993;**329**:1977–81. doi:10.1056/NEJM199312303292701.
- Citarda F, Tomaselli G, Capocaccia R, et al. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut* 2001;**48**:812–5. doi:10.1136/gut.48.6.812.
- Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. *Cancer* 1975;**36**:2251–70. doi:10.1002/cncr.2820360944.
- van Rijn JC, Reitsma JB, Stoker J, et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006;**101**:343–50. doi:10.1111/j.1572-0241.2006.00390.x.
- Than M, Witherspoon J, Shami J, et al. Diagnostic miss rate for colorectal cancer: An audit. *Ann Gastroenterol* 2015;**28**:94–8.
- Pasha SF, Leighton JA, Das A, et al. Narrow band imaging (NBI) and white light endoscopy (WLE) have a comparable yield for detection of colon polyps in patients undergoing screening or surveillance colonoscopy: A meta-analysis. *Gastrointest Endosc* 2009;**69**:PAB363. doi:10.1016/j.gie.2009.03.1079.
- Machida H, Sano Y, Hamamoto Y, et al. Narrow-band imaging in the diagnosis of colorectal mucosal lesions: a pilot study. *Endoscopy* 2004;**36**:1094–8. doi:10.1055/s-2004-826040.
- Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;**1**:18. doi:10.1038/s41746-018-0029-1.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;**25**:44–56. doi:10.1038/s41591-018-0300-7.
- Manna C, Nanni L, Lumini A, et al. Artificial intelligence techniques for embryo and oocyte classification. *Reprod Biomed Online* 2013;**26**:42–9. doi:10.1016/j.rbmo.2012.09.015.
- Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019;**89**: 25–32. doi:10.1016/j.gie.2018.07.037.
- Zhang K, Liu X, Shen J, et al. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 2020;**181**:1423–33. doi:10.1016/j.cell.2020.04.045.
- Qi X, Zhang L, Chen Y, et al. Automated diagnosis of breast ultrasonography images using deep neural networks. *Med Image Anal* 2019;**52**:185–98. doi:10.1016/j.media.2018.12.006.
- Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018;**172**:1122–31. doi:10.1016/j.cell.2018.02.010.
- Li L, Chen Y, Shen Z, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* 2020;**23**:126–32. doi:10.1007/s10120-019-00992-2.
- Guo X, Zhang N, Guo J, et al. Automated polyp segmentation for colonoscopy images: A method based on convolutional neural networks and ensemble learning. *Med Phys* 2019;**46**: 5666–76. doi:10.1002/mp.13865.
- 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**:1069–78. doi:10.1053/j.gastro.2018.06.037.
24. Chen PJ, Lin MC, Lai MJ, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018;**154**:568–75. doi:10.1053/j.gastro.2017.10.010.
 25. Yu L, Chen H, Dou Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging* 2017;**36**:994–1004. doi:10.1109/TMI.2016.2642839.
 26. Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput Methods Programs Biomed* 2019;**187**:104964. doi:10.1016/j.cmpb.2019.06.023.
 27. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Netw* 2020;**128**:47–60. doi:10.1016/j.neunet.2020.05.003.
 28. Jimenez-Sanchez A, Kazi A, Albarqouni S, et al. Precise proximal femur fracture classification for interactive training and surgical planning. *Int J Comput Assist Radiol Surg* 2020;**15**:847–57. doi:10.1007/s11548-020-02150-x.
 29. Lu L, Daigle BJ, Jr. Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma. *PeerJ* 2020;**8**:e8668. doi:10.7717/peerj.8668.
 30. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019;**89**:408–15. doi:10.1016/j.gie.2018.09.024.
 31. Mori Y, Kudo SE, Misawa M, et al. Real-Time Use of Artificial intelligence in identification of diminutive polyps during colonoscopy: A prospective study. *Ann Intern Med* 2018;**169**: 357–66. doi:10.7326/M18-0249.
 32. Wang P, Xiao X, Brown JRG, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018;**2**:741–8. doi:10.1038/s41551-018-0301-3.
 33. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988;**319**:525–32. doi:10.1056/NEJM198809013190901.
 34. Heresbach D, Barrioz T, Lapalus MG, et al. Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy* 2008;**40**:284–90. doi:10.1055/s-2007-995618.
 35. Zhang X, Jiang L, Yang D, et al. Urine sediment recognition method based on multi-view deep residual learning in microscopic image. *J Med Syst* 2019;**43**:325. doi:10.1007/s10916-019-1457-4.
 36. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–8.
 37. Abadi M, Agarwal A, Barham P, et al. *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*. <https://www.mendeley.com/catalogue/tensorflow-largescale-machine-learning-heterogeneous-distributed-systems>.
 38. Simonyan K, Zisserman AJCS. *Very deep convolutional networks for large-scale image recognition*. 2014. arXiv:1409.1556v4.