

ARTICLE

CB-HRNet: A Class-Balanced High-Resolution Network for the evaluation of endoscopic activity in patients with ulcerative colitis

Ge Wang¹ | Shujiao Zhang² | Jie Li¹ | Kai Zhao¹ | Qiang Ding¹ | Dean Tian¹ | Ruixuan Li² | Fuhao Zou² | Qin Yu¹ 

¹Department of Gastroenterology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

²School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China

Correspondence

Fuhao Zou, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China.

Email: fuhao_zou@hust.edu.cn

Qin Yu, Department of Gastroenterology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 Jiefang Avenue, Wuhan, Hubei 430030, China.

Email: yuqin@tjh.tjmu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81770528, 81974063 and 81800493; Hubei Pathophysiology Society Foundation, Grant/Award Number: 2021HBAP011

Abstract

Endoscopic evaluation is the key to the management of ulcerative colitis (UC). However, there is interobserver variability in interpreting endoscopic images among gastroenterologists. Furthermore, it is time-consuming. Convolutional neural networks (CNNs) can help overcome these obstacles and has yielded preliminary positive results. We aimed to develop a new CNN-based algorithm to improve the performance for evaluation tasks of endoscopic images in patients with UC. A total of 12,163 endoscopic images from 308 patients with UC were collected from January 2014 to December 2021. The training set and test set images were randomly divided into 37,515 and 3191 after excluding possible interference and data augmentation. Mayo Endoscopic Subscores (MES) were predicted by different CNN-based models with different loss functions. Their performances were evaluated by several metrics. After comparing the results of different CNN-based models with different loss functions, High-Resolution Network with Class-Balanced Loss achieved the best performances in all MES classification subtasks. It was especially great at determining endoscopic remission in UC, which achieved a high accuracy of 95.07% and good performances in other evaluation metrics with sensitivity 92.87%, specificity 95.41%, kappa coefficient 0.8836, positive predictive value 93.44%, negative predictive value 95.00% and area value under the receiver operating characteristic curve 0.9834, respectively. In conclusion, we proposed a new CNN-based algorithm, Class-Balanced High-Resolution Network (CB-HRNet), to evaluate endoscopic activity of UC with excellent performance. Besides, we made an open-source dataset and it can be a new benchmark in the task of MES classification.

Ge Wang and Shujiao Zhang contributed equally to this work.

Fuhao Zou and Qin Yu contributed equally to this work and share last authorship.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

There is interobserver variability in interpreting endoscopic images of patients with ulcerative colitis (UC) among gastroenterologists. Artificial intelligence (AI) is presenting as a promising tool to solve this problem.

WHAT QUESTION DID THIS STUDY ADDRESS?

Recent studies have reported the possible roles of AI in inflammatory bowel disease (IBD). However, accuracy, sensitivity, and specificity are the major concerns for the application of AI in medical fields. There were no comprehensive reports comparing the performance of different convolutional neural network (CNN) algorithms in evaluating endoscopic inflammation of UC based on a large dataset.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study for the first time compared the performance of different CNN architectures on a large dataset of endoscopic images from patients with UC via comprehensive tasks. We further developed a new CNN-based algorithm, named Class-Balanced High-Resolution Network (CB-HRNet) to assist in evaluating endoscopic activities of UC, which was especially great at determining endoscopic remission. We also described the AI-assisted image recognition in detail and made an open-source dataset to be a new benchmark in the task of Mayo Endoscopic Subscores classification.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

CB-HRNet will have great potential to help to improve the accuracy and consistency of endoscopic evaluation in all IBD centers and hospitals, saving the labor and reading time as well.

INTRODUCTION

Inflammatory bowel disease (IBD) is considered a chronic nonspecific intestinal inflammatory disease, mainly including ulcerative colitis (UC) and Crohn's disease (CD). The definite diagnosis of IBD depends on a comprehensive evaluation combined with clinical symptoms, laboratory findings, and endoscopic and pathologic features. Representative endoscopic findings of UC include geographic superficial ulcers, erosions, and mucosal ground glass changes, usually covered with fibrinous exudate. The lesions are continuously diffused from the rectum.¹ Endoscopic activity is considerably important for the evaluation of mucosal inflammation and disease activity, further guiding treatment.²

Some score systems have been used for measuring the endoscopic disease activity, among which the Mayo Endoscopic Subscore (MES) classification has been reliable and widely used in the clinic as well as research.³ Under the Mayo score system, UC can be divided into endoscopic remission state and active state, with MES less than or equal to 1, and greater than 1, respectively.⁴ However, some endoscopic characteristics of UC, such as

superficial ulcers and fibrinous exudate, erythema and erosions, and erosions and ulcers, are sometimes difficult for endoscopists to distinguish, resulting in the classification of different Mayo scores. Therefore, there is interobserver variability in interpreting endoscopic images among different gastroenterologists,⁵ some of whom are even not trained well about the recognition and evaluation of endoscopic activity in patients with UC. Actually, human reviewers were significantly more likely to disagree on intermediate MES 1 and MES 2 compared with MES 0 and MES 3, which was 38.7% and 21.8%, respectively.⁶ Another study reported 73% interobserver disagreement for "normal" appearances and 63% disagreement for moderate appearances of video sigmoidoscopies from patients with UC.⁵ Although standardized training by experts in IBD or introduction of a supervisor may ameliorate the situation, it is time-consuming and cost-consuming. It is also difficult to implement in the clinical practice.⁷

With rapid progress, artificial intelligence (AI) is presenting as a promising tool used for medical fields, especially for diagnosis and differential diagnosis of diseases, monitoring of disease progression, and efficacy prediction.^{8,9} It uses deep learning methods to read amounts

of radiological, endoscopic, and pathologic images, producing proper machine learning (ML) algorithms. ML algorithms are capable of recognizing new images with quick response and high accuracy. Several studies have shown that well-trained ML can implement image interpretation comparable to specialist level, even better than humans.^{6,10} It is reported that AI systems significantly increased hyperplastic polyp detection rates from 35.07% to 64.93%.¹⁰

Recent studies have reported the possible roles of AI in IBD.^{6,11–14} However, accuracy, sensitivity, and specificity are the major concerns for the application of AI in medical fields. For image recognition, challenges lie in high-quality large-scale dataset, quality control, and precise ML algorithms. The convolutional neural network (CNN) is one of the most widely used ML algorithms.¹⁵ Many CNN architectures have been developed to carry out data classification.¹⁶ To date, there were no comprehensive reports comparing the performance of different CNN algorithms in evaluating endoscopic inflammation of UC based on a large dataset. Hence, we designed a study aimed to develop a new CNN-based algorithm to optimize the assessment of endoscopic activities for UC.

METHODS

Dataset generation

Three hundred eight patients with UC who underwent colonoscopy at Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, from January 2014 to December 2021 were enrolled in this study. All patients were diagnosed with UC based on comprehensive evaluation according to the IBD guidance and followed up until May 2022.¹ Colonoscopies were performed in four endoscopy centers in Tongji Hospital by experienced gastroenterologists. A total of 12,163 endoscopic images from 308 patients with UC were collected, all of which were obtained by Olympus colonoscopy. We first excluded possible interference, such as narrow-band images and unclear images with stool, blur, or halation, or those without sufficient

insufflation. The number of images after screening was 7978. Each image was reviewed by at least three gastroenterologists who were experts in IBD, and MES classifications were assigned for each image. Representative images for MES classification in our data are shown in [Figure 1](#). If there were different opinions among the reviews, they would gain consensus by discussion, and the final classifications of all images were reviewed by the IBD specialist. The training set and test set images were randomly divided into 4787 and 3191 according to the ratio of 3:2. Written informed consents were obtained from all patients prior to colonoscopy examination. The procedure was approved by the Institutional Ethics Board of Tongji Medical College.

In the preprocessing stage, the black frame surrounding each image was cropped on each side and the original endoscopic images were resized to 300×300. Several data augmentation strategies were applied to the dataset, such as rotation, flip, and brightness, contrast, sharpness, and scale transformation, whereas class balancing was also used. Finally, the number of training images was amplified to 37,515, including 9095 (24.2%) images with Mayo score of 0, and 9945 (26.5%) images with Mayo score of 1, and 9081 (24.2%) images with Mayo score of 2, and 9394 (25.0%) images with Mayo score of 3. The final dataset, which we named as Tongji Medical College-Ulcerative Colitis with Mayo scores (TMC-UCM), was open source (publicly accessible at [Baidu Drive](#)) and could be a new benchmark in the task of MES classification. Study flow diagram of ML predicting endoscopic inflammation in patients with UC is shown in [Figure 2](#).

CNN-based model and loss function

The model was constructed based on CNN. In the process of training, the network can continuously learn the features of input images and update the weight parameters according to the back propagation of loss. In addition, in the process of inferencing, the network can predict the category of input image according to the parameters obtained by training. Different from some classic CNNs in classification tasks, such as ResNet,¹⁷

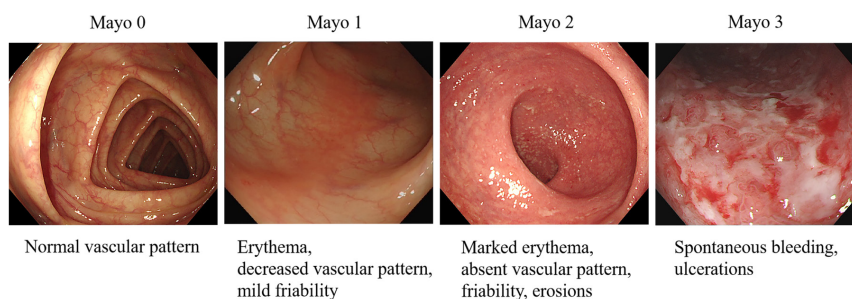


FIGURE 1 Representative images for Mayo scores classification in our data.

ResNeXt,¹⁸ and EfficientNet,¹⁹ we used High-Resolution Network (HRNet),²⁰ a novel deep high-resolution representation learning network, which was sensitive to position information (Figure 3; See Appendix S1 for specific methods).

Loss function plays an important role in the training process of CNN. The commonly used loss function is cross entropy loss (CE). To avoid problems of overfitting and model overconfidence, label smoothing loss (LS) is used as a regularization method, which can make the clustering

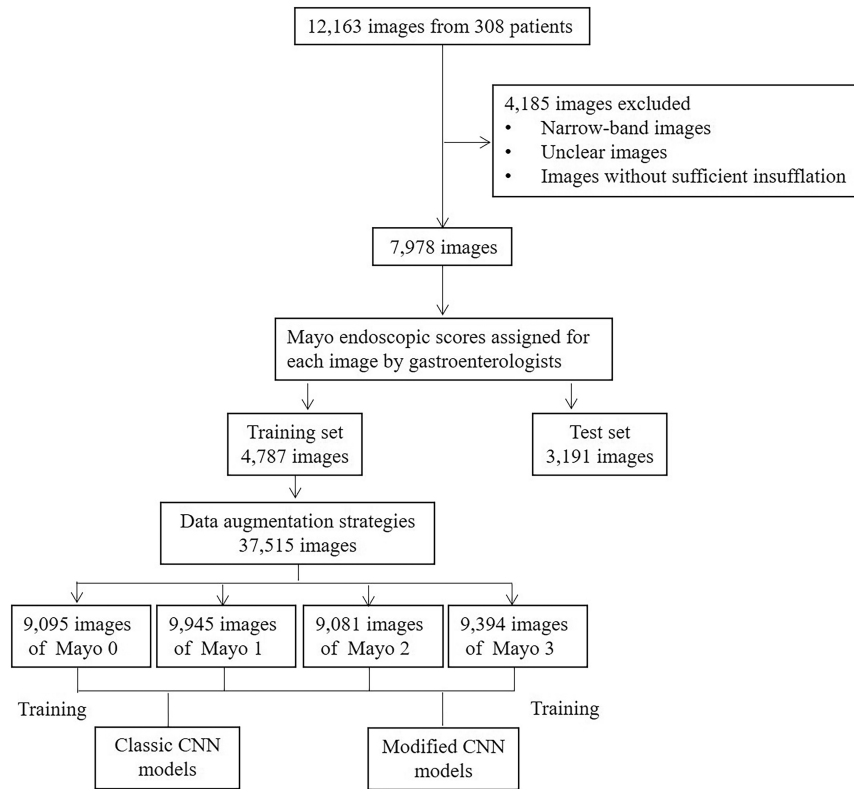


FIGURE 2 Study flow diagram of machine learning algorithms predicting endoscopic inflammation in patients with ulcerative colitis. CNN, convolutional neural network.

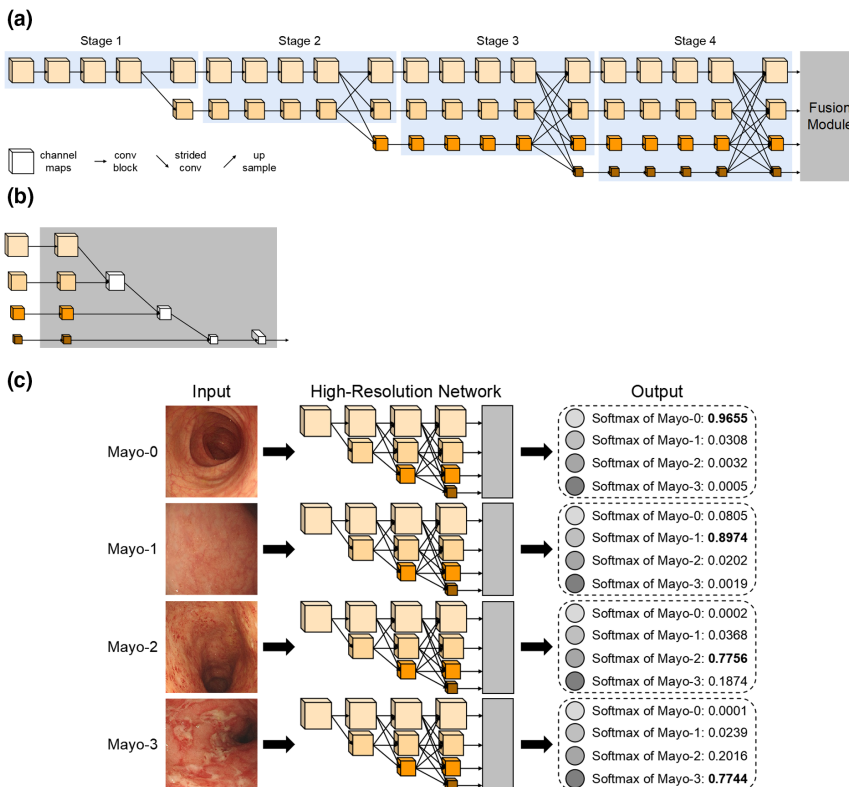


FIGURE 3 Architecture of Class-Balanced High-Resolution Network (CB-HRNet).

between classifications more compact and increase the distance between different classes while decreasing the distance within the class.²¹

In addition, the balanced data was still randomly expanded from the original unbalanced data, without essentially solving the problem of class imbalance in the dataset. Therefore, some loss functions that could optimize this problem were also considered for our classification task, such as focal loss (FL)²² and class-balanced loss (CB²³; see Appendix S1 for specific methods).

Experiments and statistical analysis

We used PyTorch²⁴ to implement the above methods on a server with GeForce RTX 3090 and 24G memory. Model parameters were initialized using weights pretrained on ImageNet²⁵ followed by end-to-end training. Batch size, learning rate, weight decay, and momentum were set to 64, 0.002, 0.0005, and 0.9, respectively. Then, model output generated the probabilities of each MES for each image. Test-time augmentation²⁶ was also applied to the inferencing process.

To prove the effectiveness of our methods, several evaluation metrics were used to quantify the results, including overall classification accuracy (ACC), area value under the receiver operating characteristic (ROC) curve (AUROC), sensitivity (SES), specificity (SPC), positive predictive value (PPV), negative predictive value (NPV), and Kappa coefficient (K).

To facilitate the comparison, the MES classification task was further divided into four subtasks, namely MES 0/1/2/3, 0/1/23, 0/123, and 01/23. The performances of CNN-based models were validated by evaluating whether it could classify each image into the right MES. The evaluation metrics ACC and K were calculated in all subtasks, SES, SPC, PPV, NPV and AUROC were only calculated in binary classification tasks (i.e., MES 0/123 and 01/23). AUROC was calculated using Python. The performances of CNN-based models with different strategies were compared.

Besides, in order to verify the effectiveness and robustness of our method, we conducted additional experiments

using nested K-fold cross-validation. We also compared our method with other prediction methods for MES in each round of nested K-fold cross-validation.

RESULTS

There were 3191 images in the test set, including 1177 (36.9%) images with Mayo score of 0, 695 (21.8%) images with Mayo score of 1, 712 (22.3%) images with Mayo score of 2, and 607 (19.0%) images with Mayo score of 3. It cost ML algorithms the mean time of 0.18 s to generate a Mayo score for each endoscopic image.

Performance comparison of different backbone networks

The performance of HRNet compared to other prediction methods of MES with three common backbone networks (CE is used as the loss function) upon endoscopic images recognition of Mayo scores was shown in Table 1. All backbones had a high accuracy from 92.17% to 94.36% upon images recognition of MES 0/123 and MES 01/23 tasks, better than MES 0/1/2/3 and MES 0/1/23 tasks, upon which the accuracy differed from 82.45% to 87.37%. The accuracy of ResNet, ResNeXt, EfficientNet, and HRNet in predicting endoscopic remission was 93.54%, 93.58%, 93.70%, and 94.36%, respectively.

HRNet achieved the best performance in each subtask, which obtained the accuracy of 83.92%, 87.37%, 92.60%, and 94.36%, respectively. Besides, the AUROCs of HRNet upon MES 01/23 and MES 0/123 evaluation was high up to 0.9698 and 0.9806, respectively.

Performance comparison of different loss functions

To further optimize the performance of HRNet in image recognition, different loss functions (HRNet is used as the

TABLE 1 Performance comparison of different backbone networks.

MES	Metrics	ResNet	ResNeXt	EfficientNet	HRNet
0/1/2/3	ACC, %	82.45	82.92	83.05	83.92
0/1/23	ACC, %	86.05	86.56	86.87	87.37
0/123	ACC, %	92.17	92.64	92.95	92.60
	AUROC	0.9683	0.9685	0.9587	0.9698
01/23	ACC, %	93.54	93.58	93.70	94.36
	AUROC	0.9768	0.9782	0.9688	0.9806

Abbreviations: ACC, accuracy; AUROC, area value under the receiver operating characteristic (ROC) curve; MES, Mayo Endoscopic Subscores.

backbone network) were used. The performance comparison of four loss functions is shown in Table 2. The accuracy of the other three loss functions (LS, FL, and CB) in distinguishing all MES subtasks were all slightly superior or equal to CE. Besides, the AUROCs of these four strategies in predicting endoscopic remission were 0.9806, 0.9770, 0.9786, and 0.9834, respectively.

After comparing the results of different CNN-based models with different loss functions, High-Resolution Network with Class-Balanced Loss achieved the best performance in all MES classification subtasks. We named this new CNN-based algorithm Class-Balanced High-Resolution Network (CB-HRNet). Its specific results for evaluation of endoscopic activity are shown in Table 3. The SES, SPC, PPV, and NPV of recognizing MES 0/123 were 92.25%, 93.20%, 95.87%, and 87.55%, respectively. For the MES 01/23 recognition, CB-HRNet obtained the performance with SES, SPC, PPV, and NPV of 92.87%, 95.41%, 93.44%, and 95.00%, respectively. Kappa coefficient values of CB-HRNet in the MES 0/123 and MES 01/23 classification were 0.8433 and 0.8836, respectively. The ROC curves and corresponding AUROCs of binary classification tasks are shown in Figure 4.

Performance of nested K -fold cross-validation

Specifically, we divided the original 7978 images into K equal parts (here, $K=5$), took the $K-1$ part each time as

the training set, and the rest as the test set, and conducted cross-validation experiments for K times. In order to facilitate comparison, the cross-validation experiment was carried out for the classification task of MES 01/23, and the specific results are shown in Tables S1 and S2. For different test sets divided each time, CB-HRNet could achieve good performance, and the ACC reached 93.22%, 94.05%, 94.67%, 94.36%, and 93.73%, respectively. Compared with other prediction methods (including ResNet with CE loss, ResNeXt with CE loss, and HRNet with CE loss), the cross-validation results of our CB-HRNet in each round were superior to those of other methods, with the highest accuracy increased by 10.86% and the highest AUROC increased by about 0.17.

DISCUSSION

In the present study, we for the first time compared the performance of different CNN-based models on a large dataset of endoscopic images from patients with UC via comprehensive tasks. Furthermore, we proposed a new CNN-based algorithm, CB-HRNet, capable of evaluating endoscopic activities of UC images accurately, especially in distinguishing endoscopic activities of MES 0 from MES 1-3 and MES 0-1 from MES 2-3. We also described the AI-assisted image recognition in detail and made an open-source dataset, which could be a new benchmark in the task of MES classification.

MES	Metrics	CE	LS	FL	CB
0/1/2/3	ACC, %	83.92	84.17	84.51	84.64
0/1/23	ACC, %	87.37	87.72	88.05	88.12
0/123	ACC, %	92.60	92.79	93.54	93.73
	AUROC	0.9698	0.9728	0.9716	0.9754
01/23	ACC, %	94.36	94.51	94.89	95.07
	AUROC	0.9806	0.9770	0.9786	0.9834

TABLE 2 Performance comparison of different loss functions.

Abbreviations: ACC, accuracy; AUROC, area value under the receiver operating characteristic (ROC) curve; CB, class-balance loss; CE, cross entropy loss; FL, focal loss; LS, label smoothing loss; MES, Mayo Endoscopic Subscores.

TABLE 3 Specific results of CB-HRNet.

MES	ACC (%)	AUROC	SES (%)	SPC (%)	PPV (%)	NPV (%)	K
0/1/2/3	84.64	/	/	/	/	/	0.7788
0/1/23	88.12	/	/	/	/	/	0.8034
0/123	93.73	0.9754	92.25	93.20	95.87	87.55	0.8433
01/23	95.07	0.9834	92.87	95.41	93.44	95.00	0.8836

Abbreviations: ACC, accuracy; AUROC, area value under the receiver operating characteristic (ROC) curve; CB-HRNet, Class-Balanced High-Resolution Network; K, Kappa coefficient; MES, Mayo Endoscopic Subscores; NPV, negative predictive value; PPV, positive predictive value; SES, sensitivity; SPC, specificity.

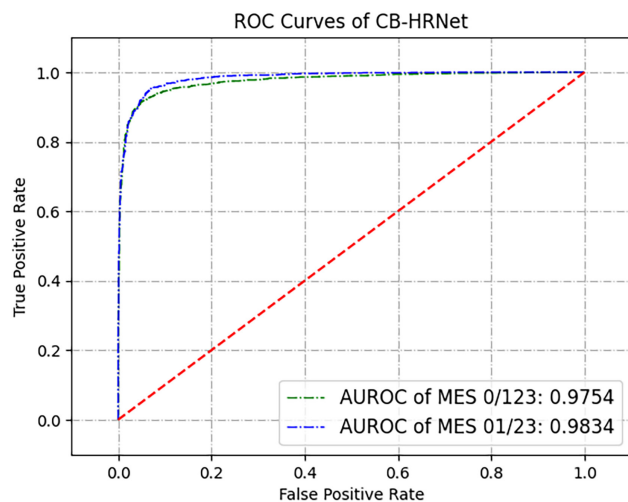


FIGURE 4 ROC curves of Class-Balanced High-Resolution Network (CB-HRNet). AUROC, area value under the receiver operating characteristic; MES, Mayo Endoscopic Subscores; ROC, receiver operating characteristic.

There were several studies demonstrating the ability of AI assessing endoscopic disease activity in patients with IBD. A study trained the CNN with 26,304 colonoscopy images from 841 patients with UC and then validated the model with other 3981 images from 114 patients with UC, showing a high level of performance with AUROCs of 0.86 and 0.98 to identify MES 0 and MES 0-1, respectively.¹¹ It demonstrated the better performance when recognizing a mucosal healing state (MES 0-1) than MES 0, which was consistent with our findings, regardless of different CNN-based models. In another prospective validation study, the deep neural network for evaluation for UC endoscopic index of severity (UCEIS) identified 875 patients with endoscopic remission with 90.1% accuracy and a Kappa coefficient of 0.798.²⁷ Similarly, using the CNN to learn 14,862 colonoscopy images from patients with UC and test 1652 images from 304 patients with UC, Stidham et al.⁶ found ML could distinguish MES 0-1 from MES 2-3 with a sensitivity of 83.0% and PPV of 0.86. Another recent study using 777 images from the publicly available Hyper-Kvasir dataset showed across MES 1, 2, and 3, average SPC was 85.7%, average SES was 72.4%, average PPV was 77.7%, and average NPV was 87.0%.²⁸ In our study, the results showed HRNet could identify endoscopic remission with 94.36% accuracy and AUROC of 0.98. Meanwhile, HRNet obtained the highest accuracy and AUROC in the Mayo 01/23 task compared with three classic CNNs. To be noted, CB-HRNet showed excellent diagnostic ability of endoscopic remission with a sensitivity of 92.87%, PPV of 0.93, and Kappa of 0.88. Together, AI based CNN algorithms have shown the great potential and clinical value in the future.

ResNet, a more frequently used CNN architecture in IBD, has been used for differential diagnosis of CD, grading CD ulcerations, and grading UC for MES, but with variable performance and limited reproducibility.²⁹⁻³³ Specifically, the accuracy of ResNet predicting endoscopic remission was 72.02%, compared with 84.52% of InceptionV3, 73.81% of VGG19, and 87.50% of DenseNet, respectively. To the best of our knowledge, although it has been the only study reported on the comparison of different CNN architectures in grading endoscopic images of UC so far, the limitations were marked with small samples, unsatisfactory performance, class imbalance among more moderate/severe cases, selection bias, and lack of multiclass classification.³³ We used a larger dataset and we introduced HRNet, which was one of novel CNNs and initially applied to the task of human pose estimation. Although there is a way for classic CNNs to obtain strong semantic information through downsampling, and then upsampling to recover high-resolution position information. However, it will result in the loss of a large amount of valid information in the continuous up-down sampling process. HRNet can reduce the loss and maintain high-resolution during the whole process.²⁰ In the MES classification, lesions tend to only exist in a certain area of endoscopic images, and the granularity of differentiation between different categories is small. Hence, such a position-sensitive high-resolution network can be more informative and bring more interpretability in our task. The reduction of accuracy is a major concern limiting the AI models at the time of transferring them to a real environment.³⁴ The cross-validation results demonstrated the superiority of our method in image interpretation, especially compared with current common AI methods used in the evaluation tasks of endoscopic images, implying the possibility of its implementation in clinical practice.

Furthermore, to avoid the selection bias as far as possible, we include all images taken by a single colonoscopy examination. To assess the identifying ability of our new CNN algorithm, we included a larger proportion of validation samples and more endoscopic images of MES 1, MES 2, and MES 3. To include more information of endoscopic images and generate a balanced dataset, we used an augmentation strategy in the stage of image processing. Noticeably, few studies were reported about the separate Mayo 0/1/2/3 recognition of endoscopic images of UC.⁶ Stidham et al. showed that the accuracy of identifying MES 0, MES 1, MES 2, and MES 3 by using Inception V3 architecture were 89.6%, 54.5%, 69.9%, and 74.3%, respectively.⁶ However, the accuracy of identifying MES 0/1/2/3 subtask by CB-HRNet can reach 84.64%, and the AUROC obtained by CB-HRNet in the Mayo 0/123 task was 0.9754. As for implementing

the Mayo 0/1/23 task, the accuracy of each CNN-based model was able to maintain from 86.05% to 87.37%. Their declined distinguishing abilities might be mainly caused by difficulties to divide a true Mayo 1 into the Mayo 1 task accurately, especially when considering ambiguous endoscopic characteristics, such as scars, little red spots, and the extent of erythema and congestion. Consistent with the reported study, we also found that it was more difficult to precisely sort endoscopic images scored with MES 1 and MES 2 compared with MES 0 and MES 3.⁶ However, we introduced three other loss functions to regularize the classifications and reduce the class imbalance of datasets in the CNN training process. The results demonstrated that they could improve the performance of HRNet identifying endoscopic inflammation, among which CB-HRNet was able to optimize the results of each Mayo task.

When evaluating the endoscopic inflammation status of UC, the entire video is superior to still images for the ability to understand the comprehensive and informative impression of endoscopy. The neural network gained excellent agreement with central readers identifying endoscopic videos from patients with UC, with Kappa value of 0.844 for MES.¹² Yao et al. demonstrated that the CNN model could exactly distinguish remission versus active disease in 83.7% videos, whereas only predicting accurately the MES in 57.1% unaltered clinical trial videos.³⁵ Challenges and difficulties of automated video recognition are the same as still image recognition, which are the accurate assessments of MES 1 and MES 2. It is believed that a fine designed ML algorithm will improve the performance and reliability of AI. In this study, we proposed a new CNN-based algorithm which might have great potential for broad applications in IBD.

Besides qualified recognition of fine lesions and delicate calculation of image scoring with less variance, another big advantage of AI application is saving time and labor. The CNN-based algorithm was capable of reducing the average time required for reading images performed by capsule endoscopy from 96.6 min to 5.9 min,³⁶ or reading a 20-min colonoscopy video in 18 s.⁶ It even took less than 0.2 s to obtain a result from the input of an endoscopic image.²⁷ Compared with unsupervised learning, supervised learning is faster to finish tasks because data are labeled previously. In our algorithm, the mean time of reading an endoscopic image and giving a Mayo score was just 0.18 s.

There were several limitations in this study. First, this was a retrospective study with data obtained from four centers in one hospital. The model we developed has not undergone rigorous external validation, only cross-validation instead. Second, we only detected the performance of CNNs when evaluating the inflammation

with MES system. However, UCEIS or a newly score system, Degree of Ulcerative colitis Burden of Luminal Inflammation (DUBLIN) score, are more refined methods with more items to evaluate inflammation severity.³⁷ There lacked the comparison of CNN performance upon the evaluation of these indices. Third, we did not develop the AI system based on real-time video, although application of similar methods to videos is early in development. Finally, recognition of the separate feature, such as vascular pattern, bleeding, and erosions and ulcers, was not designed in the study. We did not observe what features were aiding in predicting endoscopic activity and how it matched up with expert graders of images. But importantly, the study highlighted the performance of a new CNN algorithm and its promising role in transforming AI studies into clinic care. Future strategies, including expanding samples, detailed UC characteristics of endoscopic and histologic interpretation, and assessment of CD both in still images and videos, have been under consideration to validate the capability of this new ML algorithm.

CONCLUSION

We proposed a new CNN-based algorithm, CB-HRNet, to evaluate endoscopic activity of UC with excellent performance, showing great distinguishing ability of all four MES and especially good at determining endoscopic remission in UC. CB-HRNet will greatly help to improve the accuracy and consistency of endoscopic evaluation in all IBD centers and hospitals. Wide application of this CNN-based algorithm for endoscopic image recognition in patients with UC warrants further exploration.

AUTHOR CONTRIBUTIONS

G.W., S.J.Z., and Q.Y. wrote the manuscript. G.W., Q.Y., and F.H.Z. designed the research. G.W., J.L., Q.D., D.A.T., S.J.Z., K.Z., and Q.Y. performed the research. G.W., J.L., S.J.Z., F.H.Z., and Q.Y. analyzed the data. R.X.L. and F.H.Z. contributed analytical tools.

ACKNOWLEDGMENTS

Unfortunately, Professor Fuhao Zou died of cancer during the review process of this paper. We thank him immensely for his contributions in this work and thank Professor Zaobin Gan for his great help in revising this article.

FUNDING INFORMATION

This study was supported by National Natural Science Foundation of China grant nos. 81800493, 81974063, and 81770528 and the Hubei Pathophysiology Society Foundation grant no. 2021HBAP011.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest for this work.

DATA AVAILABILITY STATEMENT

The image sets that support the findings of this study are open source (publicly accessible at [Baidu Drive](https://www.baidu.com/)).

ORCID

Qin Yu  <https://orcid.org/0000-0002-4342-9398>

REFERENCES

- Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis*. 2019;13:144-164.
- Brandse JF, Bennink RJ, van Eeden S, Löwenberg M, van den Brink GR, D'Haens GR. Performance of common disease activity markers as a reflection of inflammatory burden in ulcerative colitis. *Inflamm Bowel Dis*. 2016;22:1384-1390.
- Mohammed Vashist N, Samaan M, Mosli MH, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database Syst Rev*. 2018;1:CD011450.
- Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med*. 1987;317:1625-1629.
- Travis SP, Schnell D, Krzeski P, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the ulcerative colitis endoscopic index of severity (UCEIS). *Gut*. 2012;61:535-542.
- Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open*. 2019;2:e193963.
- Feagan BG, Sandborn WJ, D'Haens G, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology*. 2013;145:149-157.e2.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
- 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;169:357-366.
- Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68:1813-1819.
- Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc*. 2019;89:416-421.e1.
- Gottlieb K, Requa J, Karnes W, et al. Central Reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology*. 2021;160:710-719.e2.
- Klein A, Mazor Y, Karban A, Ben-Itzhak O, Chowers Y, Sabo E. Early histological findings may predict the clinical phenotype in Crohn's colitis. *United Eur Gastroenterol J*. 2017;5:694-701.
- Stidham RW, Enchakalody B, Waljee AK, et al. Assessing small bowel stricturing and morphology in Crohn's disease using semi-automated image analysis. *Inflamm Bowel Dis*. 2020;26:734-742.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
- Ananda A, Ngan KH, Karabağ C, et al. Classification and visualisation of normal and abnormal radiographs; a comparison between eleven convolutional neural network architectures. *Sensors (Basel)*. 2021;21:5381.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2016:770-778.
- Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. *Computer Vision and Pattern Recognition*. 2016:1492-1500. arXiv:1611.05431.
- Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*. 2019. arXiv:1905.11946.
- Sun K, Xiao B, Liu D, Wang JJ. Deep high-resolution representation learning for human pose estimation. *Computer Vision and Pattern Recognition*. 2019. arXiv:1902.09212.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Machine Learning*. 2015;1-11. arXiv:1502.03167.
- Lin TY, Goyal P, Girshick R, et al. *Focal Loss for Dense Object Detection*. IEEE; 2017:2999-3007.
- Cui Y, Jia M, Lin TY, Song Y, Belongie S. Class-balanced loss based on effective number of samples. *Computer Vision and Pattern Recognition. Machine Learning*. 2019. arXiv:1901.05555.
- Paszke A, Gross S, Massa F, Lerer A, Chintala S. PyTorch: an imperative style, high-performance deep learning library. *Machine Learning*. 2019. arXiv:1912.01703.
- Jia D, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009:248-255.
- Ayhan MS, Berens P. Test-Time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. *Computer Science*. 2018.
- Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology*. 2020;158:2150-2157.
- Bhambhani HP, Zamora A. Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis. *Eur J Gastroenterol Hepatol*. 2021;33:645-649.
- Ruan G, Qi J, Cheng Y, et al. Development and validation of a deep neural network for accurate identification of endoscopic images from patients with ulcerative colitis and Crohn's disease. *Front Med (Lausanne)*. 2022;9:854677.
- Barash Y, Azaria L, Soffer S, et al. Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc*. 2021;93:187-192.
- Gutierrez Becker B, Arcadu F, Thalhammer A, et al. Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data. *Ther Adv Gastrointest Endosc*. 2021;14:2631774521990623.
- Kim JM, Kang JG, Kim S, Cheon JH. Deep-learning system for real-time differentiation between Crohn's disease, intestinal

- Behcet's disease, and intestinal tuberculosis. *J Gastroenterol Hepatol*. 2021;36:2141-2148.
33. Sutton RT, Zai Ane OR, Goebel R, Baumgart DC. Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images. *Sci Rep*. 2022;12:2748.
34. Vazquez-Zapien GJ, Mata-Miranda MM, Garibay-Gonzalez F, Sanchez-Brito M. Artificial intelligence model validation before its application in clinical diagnosis assistance. *World J Gastroenterol*. 2022;28:602-604.
35. Yao H, Najarian K, Gryak J, et al. Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc*. 2021;93:728-736.e1.
36. Ding Z, Shi H, Zhang H, et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology*. 2019;157:1044-1054 e1045.
37. Rowan CR, Cullen G, Mulcahy HE, et al. DUBLIN [degree of ulcerative colitis burden of luminal inflammation] score, a

simple method to quantify inflammatory burden in ulcerative colitis. *J Crohns Colitis*. 2019;13:1365-1371.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang G, Zhang S, Li J, et al. CB-HRNet: A Class-Balanced High-Resolution Network for the evaluation of endoscopic activity in patients with ulcerative colitis. *Clin Transl Sci*. 2023;16:1421-1430. doi:[10.1111/cts.13542](https://doi.org/10.1111/cts.13542)