Deep Learning Algorithm for Diagnose Endoscopic and Histological Images With Ulcerative Colitis

Yan Ye
The First Affiliated Hospital of Kunming Medical University
Xudong Luo
Yunnan University
Qiong Nan
The First Affiliated Hospital of Kunming Medical University
Yanhong Liu
Yunnan University
Yinglei Miao
The First Affiliated Hospital of Kunming Medical University
Jiarong Miao (miaojiarong60@163.com)
The First Affiliated Hospital of Kunming Medical University

Research Article

Keywords: Diagnose Endoscopic, Histological Images, Ulcerative Colitis

Posted Date: January 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1174665/v1

Abstract

The goal of treatment for ulcerative colitis is to achieve histological and endoscopic remission. Aiming at the problem that the observer will be affected by subjective factors in the endoscopic evaluation of ulcerative colitis and the cumbersome diagnosis process of histological images, this paper aims to develop a computer-assisted diagnosis system for real-time, objective diagnosis of endoscopic images and use the trained CNN model to predict histological images of patients with ulcerative colitis. Diagnosing endoscopic remission of ulcerative colitis, the accuracy of the CNN is 97.04% (95% CI,96.26%:97.62%). Diagnosing the severity of endoscopic inflammation in patients with ulcerative colitis, the accuracy of the CNN is 90.15% (95% CI, 89.49%:90.82%). The accuracy of predicting histological remission was 91.28%. The kappa coefficient between the CNN model and the biopsy results was 82.56%. The proposed computer-aided diagnosis system can effectively evaluate the inflammation of endoscopic images of patients with ulcerative colitis and predict the remission of histological images with high accuracy and consistency.

Introduction

Ulcerative Colitis (UC) is a chronic inflammatory disorder of the digestive tract. The lesions are mainly located in the colon and rectum, featuring recurrent inflammation of the colonic mucosa. UC manifests as diffuse, continuous superficial inflammation of the colorectal and corresponding histological changes[1]. It is more common in northern Europe and North America, but its incidence has been rising in our country[2–4]. Therefore, it is particularly important to achieve an early diagnosis of the disease. The goal of medical treatment in UC is to rapidly induce a steroid-free remission and prevent the disease itself to minimize its impact on the quality of life. The choice of treatment depends on the severity of disease activity. Mild to moderately active UC is usually treated with oral/topical 5-aminosalicylic acid or oral glucocorticoids. Severely active patients usually need to be admitted to the hospital for intravenous glucocorticoid therapy, even received immunosuppressants and expensive biological reagents[5,6]. In addition to clinical evaluation, complete blood count testing, erythrocyte sedimentation rate testing, and C-reactive protein testing, the severity of UC is mainly assessed by colonoscopy and histological slice. Several activity metrics for evaluating the endoscopic activity of UC have been proposed in the previous studies[7,8]. Among them, the Mayo Endoscopy Score has been proven to be one of the most widely used metrics in clinical practice to evaluate patients' disease activity with UC[9]. Mayo=0 is defined as endoscopic remission. However, interobserver variability of endoscopists is problematic for treatment planning of UC. The evaluation of histological sections is also considered critical, but different mucosal specimens are required, and pathologists may obtain different histological interpretations. Therefore, objective assessment of the disease condition of patients with UC can provide them with more effective treatment options.

Recently, artificial intelligence has made substantial progress in various medical and endoscopy fields [10–12]. Some studies have shown that deep learning models trained with specific medical images can achieve expert-level evaluation. This method does not need to set the features that need to be extracted manually but can automatically diagnose them end-to-end. Takenaka et al[13]. constructed a deep neural network system to diagnose endoscopic images and biopsy specimens of patients with UC. Byrne et al[14]. constructed a deep learning model based on InceptionV3[15] to assess the disease activity of patients with UC. Even though the existing research and applied works of artificial intelligence have made certain contributions, there is a lot more work to be done and problems to solve especially in UC. Therefore, this study collected 12257 endoscopic images of 1124 UC patients

with different endoscopic disease activity levels and 1763 histological slice images. Then, we developed a computer-assisted diagnosis system based on the CNN module to diagnose endoscopic images of patients with UC and predict the remission of histological slice images.

Materials And Methods

Data Collection. From January 2018 to November 2021, the clinical data of patients under endoscopic procedures were reviewed in the Department of Gastroenterology, the First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan Province, China. All images used standard colonoscopy and endoscope systems (Olympus, Japan). According to the Lennard-Jones criteria, a total of 1124 UC patients were diagnosed based on the typical clinical course of the disease and endoscopic examination with histologic confirmation[16]. The activity and severity of the disease were determined by Truelove and Witt's classification of UC[17]. Excluding unclear images with stool, blur, or halos, a total of 12257 endoscopic images of patients were collected. Each endoscopic image is reviewed by two experts and annotated according to the Mayo endoscopy score (Mayo=0, Mayo=1, Mayo=2, Mayo=3). In the evaluation criteria, Mayo=0 means no active lesions (remission); Mayo=1 means mild lesions, endoscopic features of erythema, and reduced blood vessel texture; Mayo=2 means moderate lesions, with obvious erythema and blood vessels under endoscopy Texture loss and erosion; Mayo=3 is a severe lesion, and its endoscopic features are spontaneous bleeding and ulcer formation. If two experts have different scores when reviewing the same endoscopic image, the final score will be determined after discussion by the endoscopy experts.

We performed targeted biopsies for surveillance and obtained at least one biopsy from the rectum. The endoscopists and pathologists were blinded to the clinical information. The histologic severity of inflammation was evaluated using the Geboes score[18]. Histologic remission (Geboes$\leq$3.0) and active (Geboes >3.0) evaluated[19]. Grade scoring for the histology was not performed because it was difficult to elucidate the grade from only endoscopic images accurately.

Firstly, we reviewed 1124 patients with UC. A total of 9807 images from 872 patients met our selection criteria collected from January 2018 to January 2021 were used as a training set. Secondly, to verify the effectiveness of the network, 2450 endoscopic images from 252 patients with UC and 1763 biopsy specimen images from 218 patients with UC from January 2021 to November 2021 were used as a verification set prospectively. The image is processed with data enhancement before training, including horizontal flip, vertical flip, random zoom, and randomly rotating. The study was approved by the Ethics Committee of The First Affiliated Hospital of Kunming Medical University (NO.2017-27). Written informed consent was obtained from all participants or their guardians (younger than age 18). Before data analysis, all accompanying patient information was annotated, and no identifiable patient information was obtained by the endoscopy experts participating in the study. All texts, images, and tables in the paper do not contain patient information, and all research protocols were conducted following relevant guidelines and regulations. According to the endoscopic images taken by different machines, it is named Dataset (Olympus). The detailed information of the Dataset is shown in Table 1.

Table 1
Statistics of the Dataset.

Dataset (Olympus)	Categories	Train(n%)	Validation(n%)	Verification (n%)
2 Categories	Remission	4350 (60.00)	1449 (19.98)	1450 (20.02)
	Active	3008 (60.06)	1000 (19.97)	1000 (19.97)
4 Categories	Mayo 0	4350 (60.00)	1449 (19.98)	1450 (20.02)
	Mayo 1	1131 (60.06)	376 (19.97)	376 (19.97)
	Mayo 2	1231 (60.02)	410 (19.99)	410 (19.99)
	Mayo 3	646 (60.14)	214 (19.93)	214 (19.93)

Construction of the CNN Models.

With the rapid development of high-performance computing equipment, convolutional neural networks are becoming more and more active in the field of computer vision. It plays a vital role in the field of medical image processing. This study uses the Inception-ResNet-v2[20] architecture as the skeleton network, composed of the stem, Inception-Resnet, Reduction, and SoftMax layers. Integrating the "residual" structure proposed in Resent[21] into the Inception module can speed up training and improve performance. The Inception module can obtain sparse or non-sparse features of the same layer, use 1×1 convolution to reduce the number of parameters, improve the recognition speed, and make the network converge faster. In addition, we use dropout[22] to reduce the weight to improve the robustness of the network. Finally, the corresponding probability is calculated by the SoftMax classifier. The Mayo score of each endoscopic image is judged. Construct a relationship with the histological image to predict the remission of the histological image. Figure 1 shows the overall network structure.

Results

Patient's characteristics and endoscopic and histologic results. After excluding patients with unclassified IBD, colorectal neoplasia, infectious diseases, and those who are contraindicated with colonoscopy, to verify the effectiveness of the network, we collected 2450 images from 252 patients from January 2021 to November 2021 as a verification set. The patient characteristics in the verification sets are shown in Table 2 and Table 3. We also analysed 1763 biopsy specimen images from 218 patients in the validation phase (biopsy samples were not obtained from 34 patients). The trained model evaluates endoscopic image remission, histological image remission, and endoscopic image severity.

Table 2
Results of endoscopic images and biopsy specimens in validation phase.

Endoscopic data (2450 images)				
Total Mayo score	0	1	2	3
Number (%):	1450 (59.18)	376 (15.35)	410 (16.73)	214 (8.74)
Histologic data (1763 specimen images)				
Histologic remission/active	Remission (Geboes ≤3.0)		Active (Geboes >3.0)	
Number (%):	981 (55.64)		782 (44.36)	

Table 3
Clinical characteristics of the study patients in validation phase.

Variables	All patients, n=252(%)
Sex, Male/ Female (%)	137(54.5%)/ 115(45.5%)
Median age at clinical onset (years)	45.5 ± 15.7
Location of disease	
E1—Proctitis	52(20.6%)
E2—Left-sided colitis	113(44.9%)
E3—Extensive/pancolitis	87(34.5%)
Median Hb(g/L) (mean ± SD)	129.8±24.4
Median CRP (mg/L) (mean ± SD)	15.8±28.1
Median ESR (mm/h) (mean ± SD)	14.8±16.3
Median ALB (g/L) (mean ± SD)	38.4±6.3
Disease activity	
Active	132(52.3%)
Remission	120(47.7%)
Disease severity	
Severe	20(15.1%)
Moderate	51(38.7%)
Mild	61(46.2%)
Concomitant treatment, n (%)	
5-aminosalicylic acid	204(80.6%)
steroids	103(40.7%)
Immunomodulators	34(13.4%)
Anti-tumor necrosis factor	15(5.9%)

Software and hardware environment and Evaluation index. The experiment uses a Windows10 system, Spyder editor, and SPSS data statistics software[23]. CPU model is AMD Ryzen 7, GPU model is NVIDIA GeForce RTX 2080Ti, all programs implemented by the open-source framework Keras[24] with TensorFlow as the backend and its Python port.

We use the following metrics to evaluate the performance of our proposed model. Accuracy represents the proportion of samples whose diagnostic predictions match the facts. Sensitivity is the percentage of patients with a positive test in the total number of patients. Specificity is the percentage of healthy people with negative test samples in the total number of healthy people. PPV and NPV are the proportions of positive and negative results in statistics and diagnostic tests that are true positive and true negative results, respectively. The kappa coefficient

is a method used in statistics to assess consistency. The higher the value of this coefficient, the better the classification accuracy achieved by the model. The formula is shown in Table 4 (TP: The recognition is correct; the sample is a positive sample. TN: The recognition is correct; the sample is negative. FP: Identify the sample as a positive sample, but the sample negative. FN: Identify the sample as negative, but the sample is positive).

Table 4. Formula of evaluation metrics.

Evaluated metrics	Formula	Evaluated metrics	Formula
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$	NPV	$\frac{TN}{FN+TN}$
Sensitivity	$\frac{TP}{TP+FN}$	PPV	$\frac{TP}{TP+FP}$
Specificity	$\frac{TN}{FP+TN}$	-	-
Kappa coefficient	$\frac{\frac{TP+TN}{N}-\frac{(TP+FP)(FN+TP)+(FN+TN)(FP+TN)}{N^2}}{1-\frac{(TP+FP)(FN+TP)+(FN+TN)(FP+TN)}{N^2}}$		

Experimental results. In this experiment, the training round is set to 30 epochs. After training, observe the changes in accuracy and loss. The model is saved on the validation set for the highest accuracy. Data augmentation is used to prevent overfitting and improve the model's generalization ability. Each image is adjusted to an input network of 299×299 size. Choose Adam optimizer[25]. The initial learning rate of the optimizer is set to 1e-4. The classification performance of the CNN model on this dataset is shown in Table 5.

Table 5
The result the CNN model (Dataset (Olympus)). The 95% CI values are listed in brackets

-	Evaluated metrics (%)					
-	Accuracy	Sensitivity	Specificity	NPV	PPV	Kappa coefficient
Endoscopic remission	97.04 [96.26:97.62]	98.43 [97.84:99.01]	94.57 [92.68:96.45]	97.65 [96.76:98.54]	96.36 [95.02:97.69]	93.49 [92.09:94.88]
Degree of disease	90.15 [89.49:90.82]	83.66 [82.94:84.38]	96.24 [95.99:96.49]	96.80 [96.42:97.18]	85.60 [83.39:87.81]	83.16 [82.04:84.27]

Table 6
Accuracy of different Mayo scores. We calculated the classification accuracy of the CNN model for each category of Mayo score. The brackets indicate the total number of images used to test the accuracy of the category, and the brackets indicate the number of correct classifications in this category.

-	Mayo 0 (Remission)	Active (Mayo 1-3)		
Olympus (2 categories)	1458(1432)	992(947)		
Accuracy	98.21%	95.46%		
-	Mayo 0	Mayo 1	Mayo 2	Mayo 3
Olympus (4 categories)	1450 (1427)	376 (270)	410 (352)	214 (166)
Accuracy	98.41%	71.80%	85.85%	77.57%

The trained model creates a probability score (with probability between 0 and 1) for each image, and uses the category with the highest probability as the final classification of the model. As shown in Figure 6.

Predict histological remission. We collected endoscopic images of 252 patients and histological images of 218 of the 252 patients and used the CNN model to train a diagnostic system for endoscopic images. The system was used to predict the remission of histological images (the histological image diagnosed by a Pathologist). Images with Mayo=0 showed endoscopic remission. Among 2450 images, 1450 images were Mayo=0, so 59.18% of the mucosa healed endoscopically. As shown in Table 7, 1546 images are consistent, and 221 images are inconsistent. The coincidence rate of endoscopy and histology is 87.46%.

Table 7
Diagnostic histological remission in images.

Mayo	Number of images	Number of images with histological	Coincidence rate of endoscopy and histology
Mayo=0	1450	874	100% (874/874) showed histological remission
Mayo =1	376	306	35.7% (109/306) showed histological remission
Mayo =2	410	378	27.2% (103/378) showed histological remission
Mayo =3	214	205	4.4% (9/205) showed histological remission

Table 8
Diagnostic histological remission in patients.

Patients (218 patients)	Endoscopic images			Kappa coefficient
	-	Remission	Active	82.56%
Histologic images	Remission	102	11	
	Active	8	97	
	Accuracy	91.28%		

Discussion

Recent studies have shown that making final judgments based on clinical results may not be enough to achieve long-term treatment success. Doctors should also evaluate the activity of histological images and treat histological remission as a treatment and clinical endpoint for UC[26,27]. Therefore, we constructed a computer-aided diagnosis system to automatically extract the endoscopic image features of patients with UC, judge the activity of the patient's endoscopic images according to the conditions of mucous membranes and blood vessels, and evaluate inflammation of UC endoscopic images and predict histological remission. Our research uses Inception-ResNet-v2 as a skeleton network to diagnose and classify the degree of activity of the endoscopic images. At the same time, because of the limited data set, a variety of data augmentation techniques are used to obtain images to extract more features in the image. We used 9807 endoscopic images of 872 patients with UC with different disease activity levels to train CNN. 2450 endoscopic images of 252 UC patients and 1763 biopsy specimen images of 218 UC patients were used as verification sets to evaluate the diagnostic and predictive capabilities of our model. The trained CNN model performs well in diagnosing endoscopic remission (mucosal healing). In Dataset, Accuracy, Sensitivity, Specificity, NPV, PPV and Kappa were 97.04%, 98.43%, 94.57%, 97.65%, 96.36% and 93.49%, respectively. The trained CNN model also performs well in diagnosing the severity of endoscopy (Mayo=0, Mayo=1, Mayo=2, and Mayo=3). In Dataset (Olympus), Accuracy, Sensitivity, Specificity, NPV, PPV and Kappa were 90.15%, 83.66%, 96.24%, 96.80%, 85.60% and 83.16%, respectively. The proposed computer-assisted diagnosis system has high accuracy and consistency, which can solve deviation in diagnosis by endoscopists. Figure 2 shows the Accuracy and Loss curves during model training. Figures 3-5 show the visualization results of box plot, confusion matrix, and Grad-CAM[28], respectively. At the same time, we compare the endoscopic images with the histological results in correspondence, observing whether the histological results are remission in the case of endoscopic remission, and its kappa was 82.56%. The results are shown in Table 7 and Table 8. The CNN model trained can predict the relief of histological images to a certain extent and provide doctors with reference.

The current research has some limitations. As can be seen from the confusion matrix in Figure 4 and Table 6, when diagnosing the severity of the Dataset (Olympus), the diagnosis accuracy rate for Mayo=1 is 71.80%, the diagnosis accuracy rate for Mayo=2 is 85.85%, the diagnosis accuracy rate for Mayo=3 is 77.57%. Mayo=1 and Mayo=3 have lower diagnostic accuracy. The endoscopic image with Mayo=1 is easily misdiagnosed as Mayo=0 and Mayo=2; the endoscopic image with Mayo=3 is easily misdiagnosed as Mayo=2. Because the similarity between adjacent images is relatively high and the number of images is small, in subsequent research, the Gan network can be used to generate images corresponding to Mayo scores to reduce the probability of

misclassification. In addition, such as Ozawa et al[29]. also use the Mayo score to distinguish endoscopic images, and its accuracy is comparable to ours. However, they further divided the colon region to verify the diagnostic effect and accuracy of endoscopic images of different colon regions. Finally, we just used the Mayo endoscopy evaluation index to judge the inflammatory activity in this article. Nevertheless, several other metrics can also effectively evaluate inflammation, such as the UCEIS and DUBLIN[13,30] evaluation indicator. The DUBLIN score is a simple clinical evaluation index that estimates the inflammatory activity according to the severity and scope of the disease. At the same time, it can establish a good relationship with clinical results, and the DUBLIN score of patients who need to continue colectomy tends to be higher. In the future, we intend to apply the DUBLIN indicator to this dataset to comprehensively evaluate the patient's disease status.

Conclusions

In conclusion, the computer-assisted diagnosis system proposed in this paper has very high stability when evaluating mucosal inflammation and predicting histological remission in UC patients. It can be used in various medical situations.

Declarations

Author contributions

Yan Ye and Xudong Luo co-designed the study, performing data extraction and analysis, and writing the manuscript. Qiong Nan assisted with data extraction and revised the manuscript. Yanhong Liu assisted with statistical analysis and comparison and revised the manuscript. Yinglei Miao and Jiarong Miao co- provided guidance on study design, advised on strategies for data preparation and analysis, and revised the manuscript.

Conflicts of Interest Competing interests

The authors declare no competing interests.

Acknowledgements

This research was supported by Yunnan health training project of high-level talents (H-2018040)

Data availability

The dataset used for the current study is available from the corresponding author on reasonable request.

References

1. DeRoche, C., Xiao, S., Liu, X. Histological evaluation in UC. Gastroenterology report. 3, 178-192 (2014).

2. Luo, C. et al. Chinese Research Into Severe UC Has Increased in Quantity and Complexity. World Journal of Clinical Cases. 6, 35-43 (2018).

3. Kaplan, G. G., Ng, S. C. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. Gastroenterology. 152, 313-321 (2017).

4. Kaplan, G. G., Windsor, J. W. The Four Epidemiological Stages in the Global Evolution of Inflammatory Bowel Disease. Nat Rev Gastroenterol Hepatol. 18, 56-66 (2021).

5. Osterman, M. T., Lewis, J. D. The Role and Importance of Endoscopic Mucosal Healing in UC. Techniques in Gastrointestinal Endoscopy. 6, 144-153 (2004).

6. Meier, J., Sturm, A. Current treatment of UC. World journal of gastroenterology. 17, 3204 (2011).

7. Osada, T. et al. Comparison of Several Activity Indices for the Evaluation of Endoscopic Activity in UC: Inter and Intraobserver Consistency. Inflamm. Bowel Dis. 16, 192-197 (2010).

8. Mohammed, V. N. et al. Endoscopic Scoring Indices for Evaluation of Disease Activity in UC. Cochrane Database Syst Rev. 1, D11450 (2018).

9. Lemmens, B. et al. Correlation Between the Endoscopic and Histological Score in Assessing the Activity of UC. Gastroenterology. 140, 421-422 (2011).

10. Chen, Z. et al. Artificial intelligence in medical applications. Journal of Healthcare Engineering. (2018).

11. Choi, J. et al. Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy. Clinical endoscopy. 53, 117-126 (2020).

12. Park, S. H., Han, K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology. 286, 800-809 (2018).

13. Takenaka, K. et al. Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients with UC. Gastroenterology (New York, N.Y. 1943). 158, 2150-2157 (2020).

14. Byrne, M. et al. DOP13 Artificial Intelligence (AI) in Endoscopy - Deep Learning for Detection and Scoring of UC (UC) Disease Activity Under Multiple Scoring Systems. Journal of Crohn's and colitis. 15, S51-S52 (2021).

15. Christian Szegedy, S. I. Rethinking the Inception Architecture for Computer Vision. Computer Vision and Pattern Recognition (cs.CV). (2015).

16.Lennard-Jones, J. E. Classification of inflammatory bowel disease. Scandinavian Journal of Gastroenterology. 24, 2-6 (1989).

17. Satsangi, J., Silverberg, M. S., Vermeire, S., Colombel, J. F. The Montreal Classification of Inflammatory Bowel Disease: Controversies, Consensus, and Implications. Gut. 55, 749-753 (2006).

18. Geboes, K. et al. A Reproducible Grading Scale for Histological Assessment of Inflammation in UC. Gut. 47, 404-409 (2000).

19. Bessissow, T. et al. Prognostic Value of Serologic and Histologic Markers On Clinical Relapse in UC Patients with Mucosal Healing. Am. J. Gastroenterol. 107, 1684-1692 (2012).

20. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning. Computer Vision and Pattern Recognition (cs.CV). (2016).

21. He, K., Zhang, X., Ren, S. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 770-778 (2016).

22. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. (2012).

23. Mishra, P. et al. Descriptive Statistics and Normality Tests for Statistical Data. Ann Card Anaesth. 22, 67-72 (2019).

24. Ketkar, N. Introduction to Keras. Deep learning with Python.97-111 (2017).

25. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. Machine Learning (cs.LG). (2016).

26. Travis, S. P. L. et al. Review Article: Defining Remission in UC. Aliment. Pharm. Ther. 34, 113-124 (2011).

27. Vuitton, L. et al. Defining Endoscopic Response and Remission in UC Clinical Trials: An International Consensus. Aliment Pharmacol Ther. 45, 801-813 (2017).

28. Ramprasaath, R. Selvaraju, M. C. A. D. Grad-CAM: Visual Explanations From Deep Networks Via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 618-626, (2017).

29. Ozawa, T. et al. Novel Computer-Assisted Diagnosis System for Endoscopic Disease Activity in Patients with UC. Gastrointest. Endosc. 89, 416-421 (2019).

30. Rowan, C. R. et al. DUBLIN [Degree of UC Burden of Luminal Inflammation] Score, a Simple Method to Quantify Inflammatory Burden in UC. J. Crohns Colitis. 13, 1365-1371 (2019).

Figures

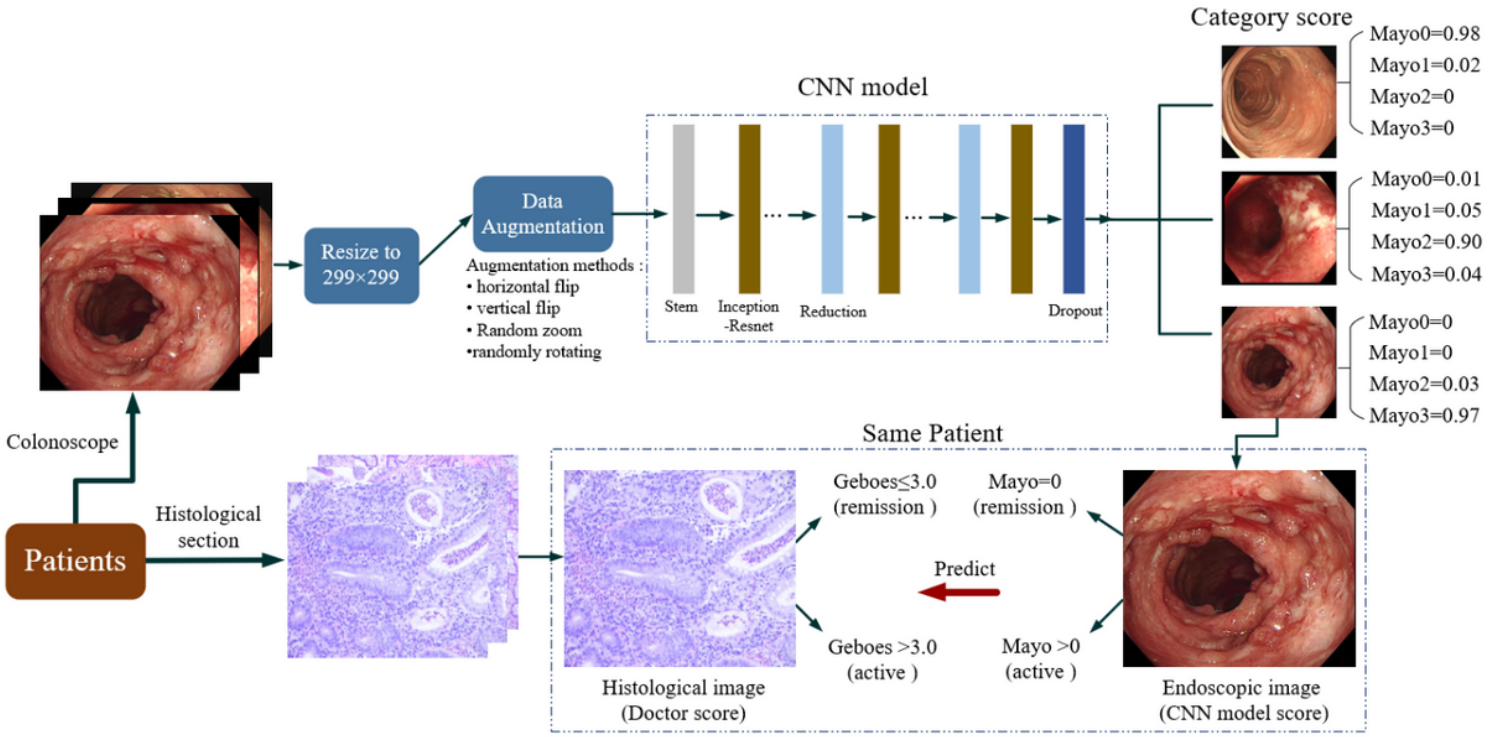

Figure 1

The overall network structure diagram.

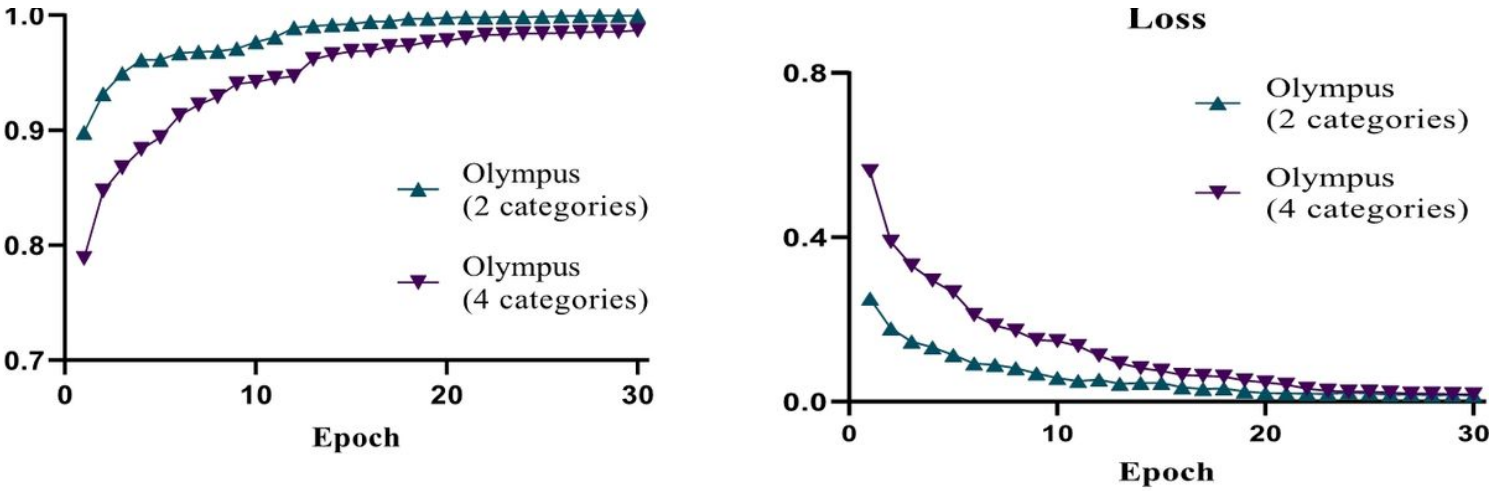

Figure 2

Accuracy and loss of Dataset (Olympus).

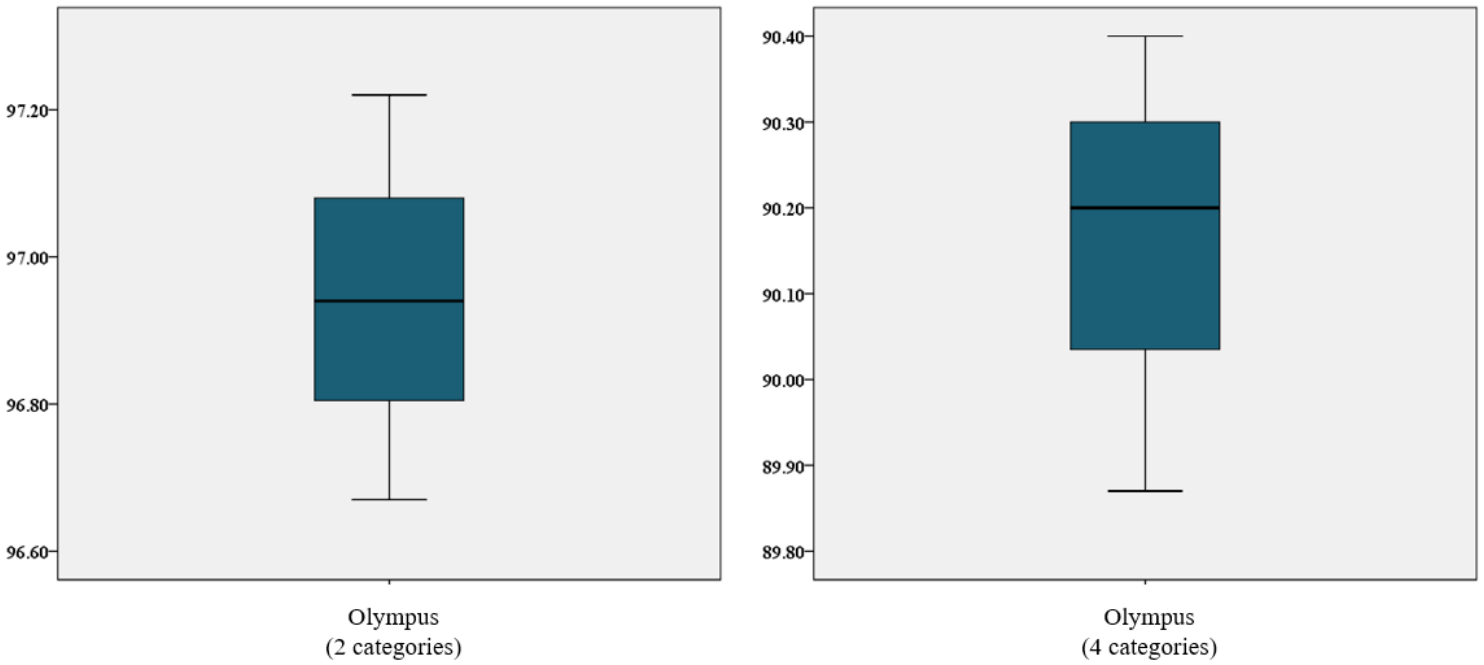

Figure 3

Box plot of endoscopic image diagnosis and classification (Accuracy).

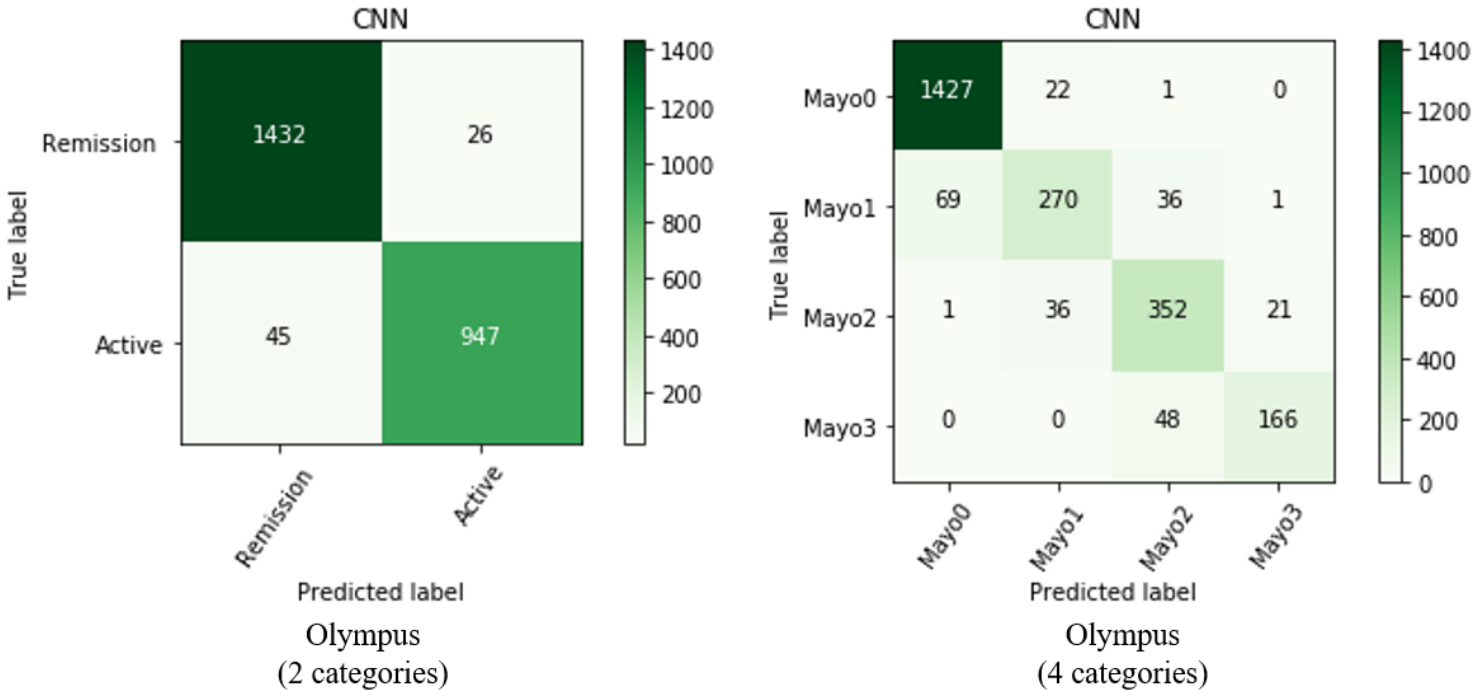

Figure 4

Confusion matrix of Dataset (Olympus).

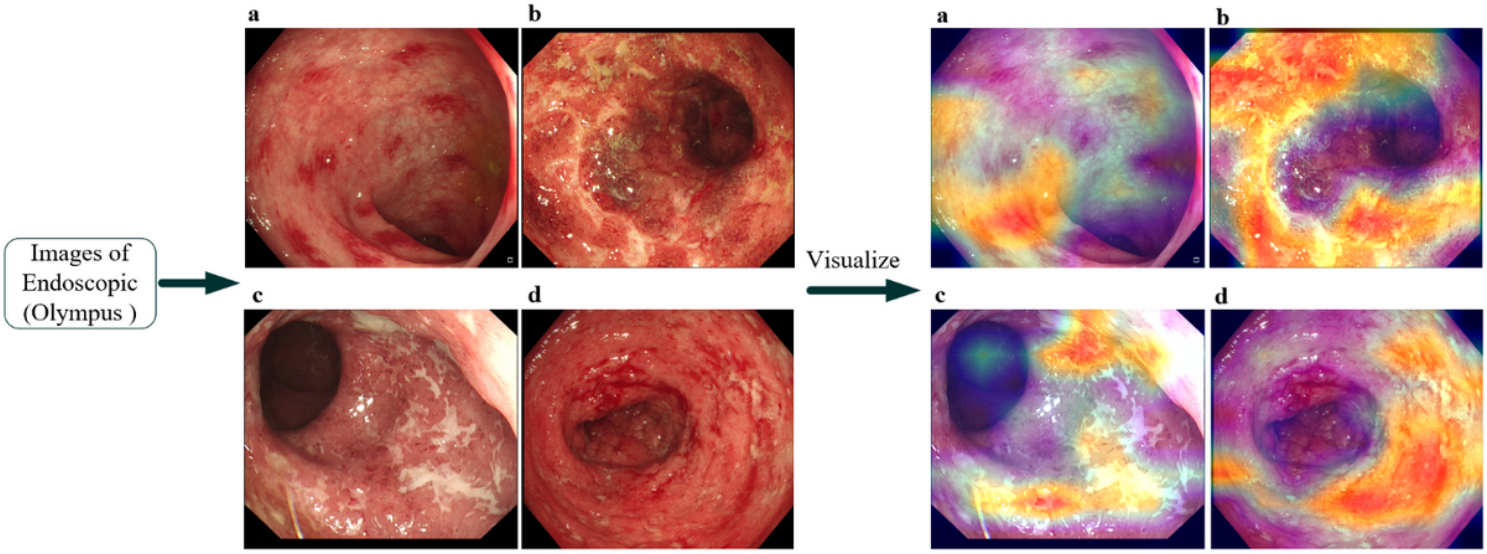

Figure 5

The Grad-CAM visualizations of endoscopic image.

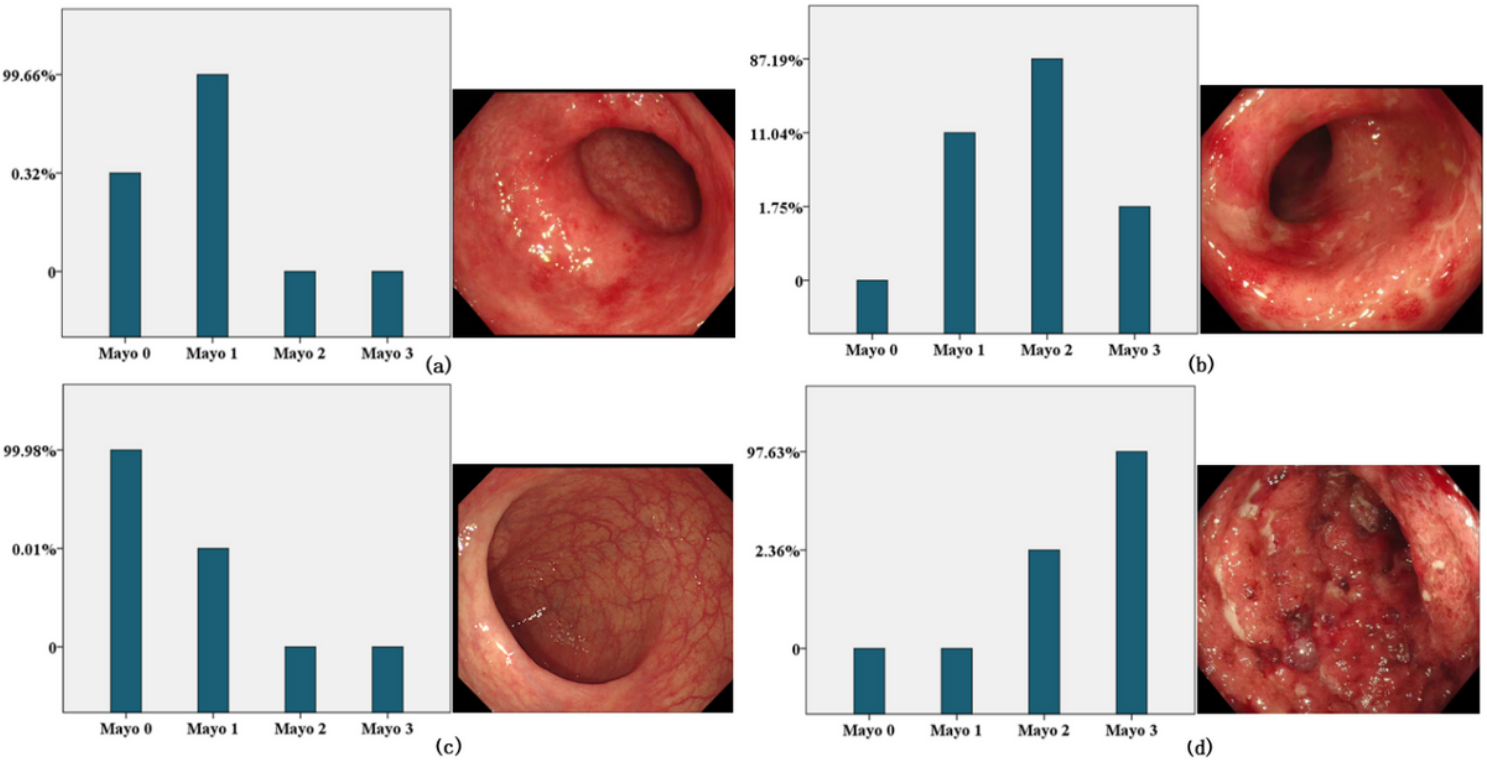

Figure 6

Probability score of a single endoscopic image of UC. (a)It is an image with a Mayo endoscope score of 1 in the Dataset (Olympus). The trained CNN network evaluates it as Mayo 1 with a probability of 0.99. (b) It is an image with a Mayo endoscope score of 2 in the Dataset (Olympus). The trained CNN network evaluates it as Mayo 2 with a probability of 0.87. (c) It is an image with a Mayo endoscope score of 0 in the Dataset (Olympus). The trained CNN network evaluates it as Mayo 0 with a probability of 0.99, (d) It is an image with a Mayo endoscope score of 3 in the Dataset (Olympus). The trained CNN network evaluates it as Mayo 3 with a probability of 0.97.