



Diagnosis of ulcerative colitis from endoscopic images based on deep learning

Xudong Luo, Junhua Zhang*, Zonggui Li, Ruiqi Yang

School of Information Science and Engineering, Yunnan University, Kunming, Yunnan Province, China



ARTICLE INFO

Keywords:
Ulcerative colitis
DenseNet
Independently recurrent neural networks
Attention mechanism
Test time augmentation

ABSTRACT

Aims: Evaluating the endoscopic images of patients with ulcerative colitis can effectively determine a reasonable treatment plan. However, the endoscopic evaluation is usually affected by the doctor's subjective judgment. This research aims to develop a computer-assisted diagnosis system that can objectively diagnose the degree of endoscopic image activity in patients with ulcerative colitis.

Methods: We proposed a neural network named "Efficient Attention Mechanism Network" ("EAM-Net") which combines the efficient channel attention network and spatial attention module. The features extracted by convolutional neural network are divided into two branches and input into the recurrent neural network and EAM-Net modules to generate and splice the attention map. The proposed EAM-Net is formed into UC-DenseNet for the classification of ulcerative colitis. The proposed method was evaluated on two colonoscopy image datasets.

Results: By using UC-DenseNet for the diagnosis of endoscopic remission in patients with ulcerative colitis, the accuracy, precision, recall rate, F1-score, and area under curve are increased by 0.5% to 2%, 0.5% to 2%, 1% to 3%, 0.6% to 2%, and 0.5% to 1.8%, respectively. In addition, for the diagnosis of the severity of endoscopic inflammation in patients with ulcerative colitis, the accuracy, precision, recall rate, and F1-score are increased by 1.5% to 4%, 1% to 3.5%, 2% to 8%, 2% to 7%, respectively.

Conclusions: Experimental results show that the proposed UC-DenseNet can effectively diagnose ulcerative colitis. It can assist the endoscopist in formulating the best treatment strategy.

1. Introduction

Ulcerative colitis is a chronic intestinal inflammatory disease closely related to heredity, immunity, infection, environment, and other factors. It could easily cause persistent inflammation and ulcers in the digestive tract [1]. The main clinical manifestations are diarrhea, abdominal pain, mucus constipation, pus, and blood [2]. This disease mainly occurs in young and middle-aged people, with complex and long-lasting conditions, difficult to cure, recurring attacks, and a high possibility of further cancer. Therefore, early diagnosis and control of ulcerative colitis are essential. In addition to clinical diagnosis, the current diagnostic methods for ulcerative colitis are mainly evaluated by colonoscopy. Various colonoscopy evaluation methods have been proposed [3,4]. Among them, the Mayo endoscopy score has been proven to be one of the most reliable and frequently used metrics for evaluating ulcerative colitis activity [5]. As shown in Fig. 1, Small superficial ulcers, "Sandpaper" appearance of the mucosa, erythema, granular, and blood in the cavity are all signs of mucosal inflammation [6,7]. However, the

accurate assessment of endoscopic inflammation requires rigorous training. Considering the limited number of professionally trained endoscopists and the differences in evaluation between different endoscopists, using automatic methods such as deep learning to identify subtle abnormalities.

In recent years, researchers have applied deep learning methods to various fields, including medical image recognition [8,9,10], which can effectively diagnose diseases in medical images. Bhambhani et al. [11] constructed a deep learning model to classify endoscopic disease severity in patients. Takenaka et al. [12] constructed a deep neural network to diagnose endoscopic images and biopsy specimens, which achieved 90.1% and 92.9% accuracy to endoscopic diagnosis remission and histological remission. Similarly, Byrne et al. [13] automatically constructed a deep learning approach to improve ulcerative colitis disease detection using Mayo endoscopy score and ulcerative colitis endoscopic index of severity (UCEIS) [14]. Zeng et al. [15] proposed a multi-model network combining three pre-trained networks (Xception [16], ResNet [17], and DenseNet [18]). The proposed model has an

* Corresponding author.

E-mail address: jhzhang@ynu.edu.cn (J. Zhang).

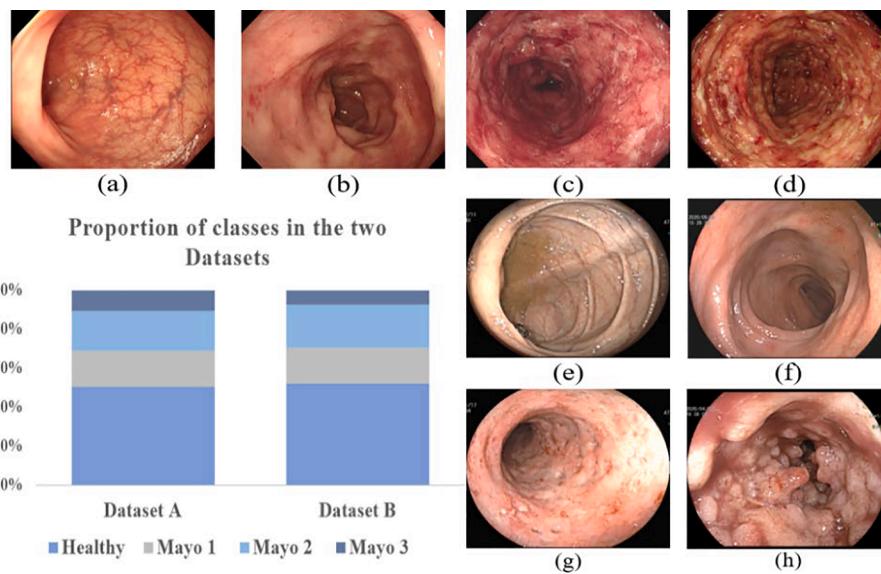


Fig. 1. Examples of ulcerative colitis images. (a), (b), (c), and (d) represent Mayo 0, Mayo 1, Mayo 2, and Mayo 3 from Dataset A, respectively. (e), (f), (g), and (h) represent Mayo 0, Mayo 1, Mayo 2, and Mayo 3 from Dataset B.

Table 1
Statistics of the Dataset.

Dataset	Slices	Healthy (Mayo 0)	Ulcerative Colitis		
			Mayo 1	Mayo 2	Mayo 3
Dataset A	9,928	5,022	1,831	2,025	1,050
Dataset B	4,378	2,289	799	970	320

accuracy rate of 97.93% in classifying ulcerative proctitis images. These methods can effectively identify the lesions of ulcerative colitis. However, to further improve the ability to extract image features, researchers try to combine Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) [19] to solve general image classification problems. The idea is first applied to the general image field. Guo et al. [20] proposed a network combining CNN and RNN, which has better performance than the wider-ResNet [21]. They applied the network to ImageNet [22] and significantly improved subcategory accuracy from 77.27% to 82%. In addition, Similar methods are also used in the field of medical image processing. Ayesha et al. [23] discussed the current research obstacles and emerging deep learning trends in medical imaging. At the same time, it is mentioned that combining CNN and RNN models is a feasible improvement direction. Coincidentally, Marentakis et al. [24] used CNN to extract the feature information of each tumor slice and employed an RNN network composed of SoftMax and Long Short Term Memory networks (LSTM) units [25] to classify tumor images. Yan et al. [26] used CNN to extract the features of the pathological image divided into small blocks and then used RNN to fuse each feature to make its classification. Yao et al. [27] introduced a special perceptron attention mechanism to fuse feature information extracted from two different neural networks. To better extract the features of the colonoscopy image, we designed a fusion network of CNN and RNN to recognize it. At the same time, we use Test Time Augmentation (TTA) to augment the test dataset. The main contributions of this paper are as follows:

1. We proposed a neural network called UC-DenseNet, which combines CNN and RNN to replace traditional methods in recognizing ulcerative colitis images.
2. We proposed an improved attention mechanism model, which emphasizes and locates feature information through cross-channel

communication to improve the diagnostic performance of ulcerative colitis images.

3. We summarized the performance improvement of the network after augmenting the test data and explored the impact of the kernel sizes on the network performance.
4. To train and evaluate the method we proposed, we collected and calibrated two ulcerative colitis image datasets, in which Dataset A contains 9,928 ulcerative colitis images, Dataset B contains 4,378.

2. Materials

2.1. Dataset

1) Ulcerative colitis dataset

The dataset used in this study is colonoscopy images obtained from the First Affiliated Hospital of Kunming Medical University. All images are collected by the Olympus electronic colonoscopy system and Japan's Fujinon electronic colonoscopy system. This dataset contains 14,306 colonoscopy images from 1317 consecutive patients with ulcerative colitis collected from January 2018 to August 2021. Unclear images with stool, blur, or halos were excluded. 9,928 and 4,378 images were recorded by Olympus electronic colonoscopy and Japan's Fujinon electronic colonoscopy, respectively. The collected images were reviewed by two experts and annotated with Mayo endoscopy scores (Mayo 0, Mayo1, Mayo 2, and Mayo 3). When two experts have inconsistent ratings of the same image, the controversial image is excluded from the dataset. The study was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University. Since this study used completely anonymous data for retrospective research, no informed consent was required. We named the two datasets, Dataset A and Dataset B. Table 1 shows the details of the datasets. We evaluated the proposed method using ulcerative colitis images taken by the above two machines.

2) Other open medical image dataset

The Kvasir [28] dataset was a collection of gastrointestinal images taken with an endoscope and annotated and verified by a certified endoscopist. It was used to display anatomical landmarks and pathological findings in the gastrointestinal tract. Anatomical landmarks include Z-line, pylorus, cecum, etc. The pathological findings include esophagitis, polyps, and ulcerative colitis. It consists of 8,000 images in 8 categories, which is 1,000 images in each category. We divide the

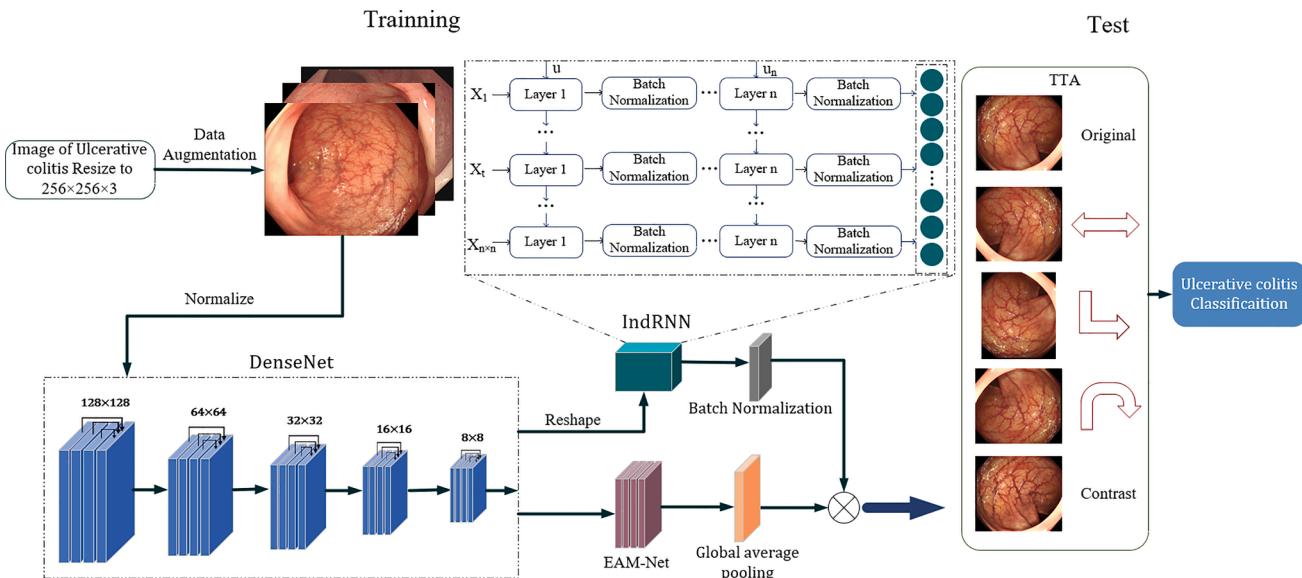


Fig. 2. Overview of the proposed framework.

Table 2
Algorithm pseudocode of the UC-DenseNet model.

Algorithm Pseudocode of applying proposed UC-DenseNet models to classification

Input: Dataset samples $S = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$. The S is classified into a training set (Train X , Train Y), a validation set (Val X , Val Y), a testing set (Test X , Test Y) in a ratio of 3:1:1 (X -endoscopic image, Y -the label of endoscopic image). The number of learning epochs is T .

Output: The optimal Model and its classification statistics

1. Load the training set and validation set;
2. Augmentation the training set;
3. Normalize the training set and validation set value to [0,1].
4. **Begin:**
5. Initialize all weights and biases.
6. **For** $m = 1, 2, \dots, T$ **do**
7. Extract features through CNN model (DenseNet201) $\rightarrow F_{\text{CNN}}$;
8. Reshape the feature map of F_{CNN} into sequence X , and input IndRNN;
9. Calculate the output of the IndRNN layer;
10. Input F_{CNN} to EAM-Net;
11. Calculate the output of the EAM-Net;
12. Concatenate the output features of IndRNN and EAM-Net;
13. Model Fit (Adam, (train X , train Y) $\rightarrow M(m)$;
14. Model Evaluate ($M(m)$, (Val X , Val Y) $\rightarrow R_{\text{acc}}(m)$.
15. **End For**
16. Save the optimal model which has max R_{acc} in T epochs.
17. **End**
18. Load the testing set;
19. Augmentation the testing set (Test Time Augmentation: zoom range, horizontal flip, vertical flip);
20. Load the optimal model in terms of classification performances.

Kvasir dataset into three subsets, which are 4800, 1600, and 1600 images for training, verification, and testing.

3. Methods

3.1. Overview

As shown in Fig. 2, DenseNet201 architecture is used as the backbone network. It comprises of five dense layers, and each layer is connected to the other layer in a feed-forward way. The features extracted by DenseNet are divided into two branches, inputting into Independently Recurrent Neural Networks (IndRNN) and an improved attention mechanism module to generate attention maps and stitch them to highlight the extracted features. The output of EAM-Net enters the

global average pooling layer to optimize the calculation space, thereby reducing the amount of calculation required. At the same time, it can also reduce the error caused by the increase in estimation variance to retain more background features. The function of Batch Normalization [29] is to regularize each feature separately to standardize data distribution in the network and reduce the model's sensitivity to network parameters, alleviating the problem of gradient disappearance. Finally, TTA is used to apply augmentation to the test image. We made predictions on both the original and augmented images, and the final predicted value is the average predicted value of all the images. For convenience, we named the classifier proposed in this paper UC-DenseNet. The pseudocode of UC-DenseNet is shown in Table 2.

3.2. Data Augmentation

Data augmentation is a method that can effectively improve the model's generalization performance [30,31]. In response to the limited images for ulcerative colitis, this paper applies the following data augmentation methods based on the original data. These methods include horizontal flip, vertical flip, random zoom, randomly rotating from 0 to 20 degrees.

3.3. RNN

Since RNN was first proposed, they have been widely used in various image processing fields [32]. Especially the research of medical image classification. Capturing the dynamic information in the serialized data can automatically locate the infected area and extract relevant features for disease classification. Therefore, to extract the feature information of colonoscopy images more effectively, we employed the IndRNN network [33] (a special RNN network) to process further the features extracted by CNN. Compared with traditional RNN and LSTM [34,35,36], it can prevent gradient vanishes and explosion problems more effectively.

At step t , the hidden layer h_t is updated based on X_t and the previously hidden layer h_{t-1} , as Eq.1:

$$h_t = \sigma(W_{X_t} + u_{t-1} \cdot h_{t-1} + b) \quad (1)$$

Among them, $X_t \in \mathbb{R}^M$ and $h_{t-1} \in \mathbb{R}^N$ are the input and the hidden state in the middle, \bullet represents the Hadamard product, σ is the neural activation function, $b \in \mathbb{R}^N$ is the bias term, $W \in \mathbb{R}^{M \times N}$ and $u \in \mathbb{R}^N$ are the weight matrices of the neuron's current input and cyclic input, respectively. N and M are the numbers of neurons and the dimension of the

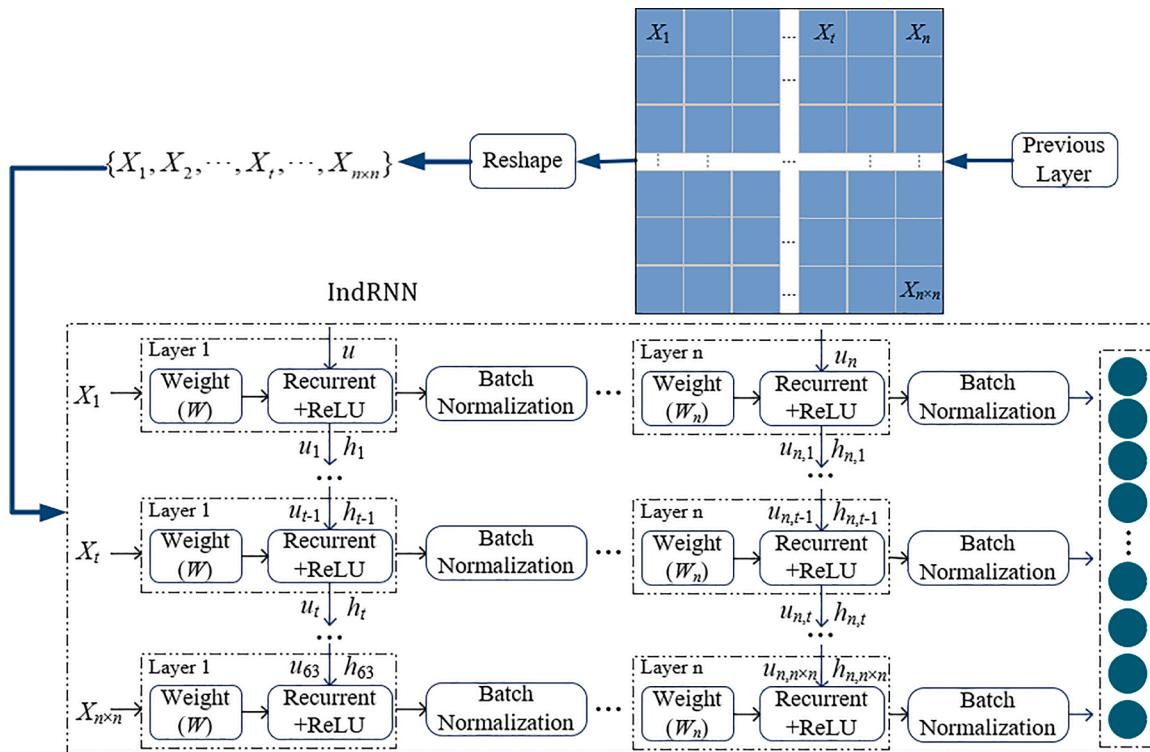


Fig. 3. Basic structure of the IndRNN.

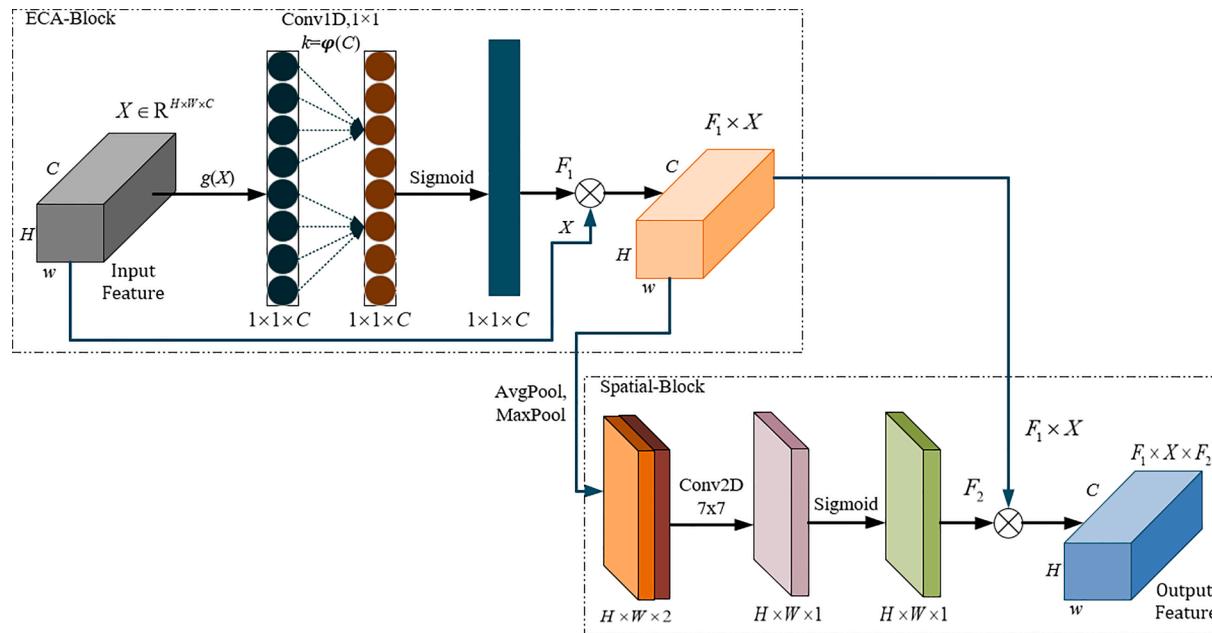


Fig. 4. Layered architecture of the EAM-Net.

Table 3

Experiment hyperparameters of the UC-DenseNet model.

Hyperparameters	Exp1	Exp2	Exp3
Input size	256×256	256×256	256×256
Batch size	16	16	16
optimizer	RMSprop	Adam	SGD
-	Exp4	Exp5	Exp6
Input size	112×112	112×112	256×256
Batch size	8	32	8
optimizer	Adam	Adam	Adam

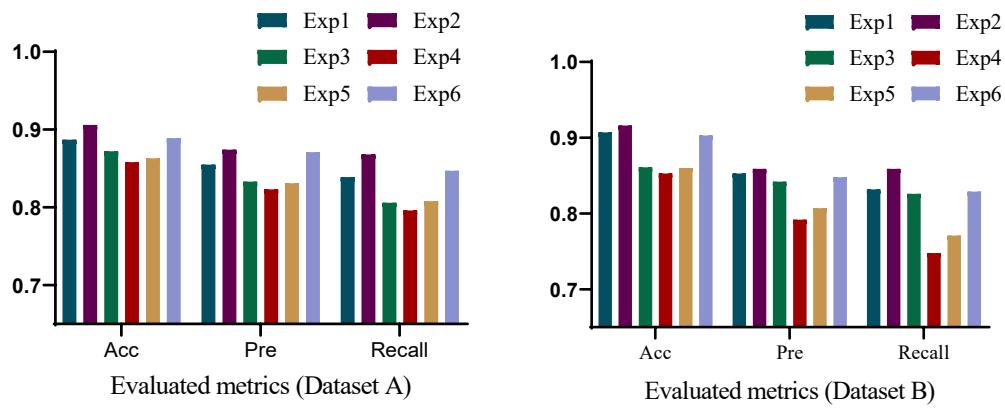


Fig. 5. Results of different hyperparameters of the UC-DenseNet model.

Table 4
Parameter settings.

Hyper parameter	VGG19	ResNet152	InceptionV3	DenseNet201	UC-DenseNet
Optimizer	Adam	SGD	Adam	Adam	Adam
Batch size	32	16	16	16	16
Input size	256 × 256	256 × 256	256 × 256	256 × 256	256 × 256
parameter	2.00 × 10 ⁷	5.83 × 10 ⁷	2.18 × 10 ⁷	1.83 × 10 ⁷	2.22 × 10 ⁷

input.

For the n^{th} neuron, the hidden state $h_{n,t}$ can be obtained by Eq. 2:

$$h_{n,t} = \sigma(W_{nX_t} + u_{n,t-1} \cdot h_{n,t-1} + b_n) \quad (2)$$

W_n and h_n are the n^{th} row of the input weight and recurrent weight. b_n is the n^{th} row of bias.

At the end of each propagation, the loss function (L) of IndRNN is calculated by Eq. 3:

$$L = \frac{1}{M} \sum_i^M |y_i - \hat{y}_i| \quad (3)$$

Among them, y_t and \hat{y}_t are the actual and estimated value at the time step t , respectively, and M is the input dimension. Since the neurons in each layer are independent, the gradient backpropagation of the neurons in a layer can also be performed independently. For the n^{th} neuron as Eq. (2), in the case of ignoring the deviation, suppose that the goal to be minimized at time step T is J_n . The gradient backpropagated to the time step t is calculated by Eq. 4:

$$\begin{aligned} \frac{\partial J_n}{\partial h_{n,T}} &= \frac{\partial J_n}{\partial h_{n,T}} \frac{\partial h_{n,T}}{\partial h_{n,t}} = \frac{\partial J_n}{\partial h_{n,T}} \prod_{k=t}^{T-1} \frac{\partial h_{n,k+1}}{\partial h_{n,k}} = \frac{\partial J_n}{\partial h_{n,T}} \prod_{k=t}^{T-1} \sigma'_{n,k+1} u_n \\ &= \frac{\partial J_n}{\partial h_{n,T}} u_n^{T-t} \prod_{k=t}^{T-1} \sigma'_{n,k+1} \end{aligned} \quad (4)$$

$\sigma'_{n,k+1}$ is the derivative of the element-wise activation function. By calculating Eq. (1) to Eq. (4), updating the hidden state, calculating the loss function through time backpropagation and backpropagation error loss constitute a complete IndRNN training process. The overall structure is shown in Fig. 3.

3.4. Improved attention mechanism

Although CNN can obtain feature information of colonoscopy images, some critical information cannot be extracted. The attention mechanism model is widely used in various deep learning tasks, such as segmentation, classification, and other fields. It quickly scans images similar to how the human eye observes objects, capturing the area that needs to be locked. Various attention mechanisms with different structures and characteristics have been developed and applied to image processing tasks with excellent results [37,38,39]. In recent years, researchers have used the attention mechanism to embed in the network to improve the classification performance of CNN. For example, Qin et al. [40] proposed a multi-dimensional attention mechanism model. The classification effect is further improved by embedding it into a fine-path network to extract fine-grained features more effectively. Zeng et al. [41] proposed a Visual Geometry Group (VGG) [42] model with a single-channel attention mechanism. It can stably and effectively improve the classification effect of a variety of lung perfusion images. Moreover,

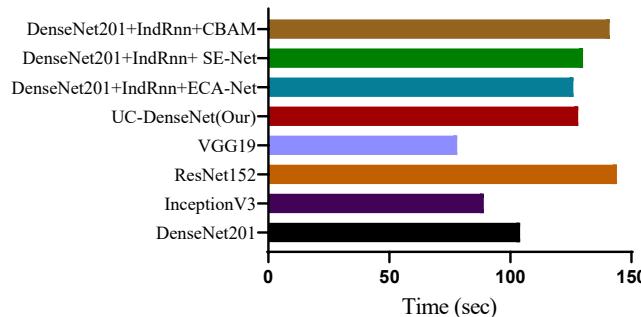
Table 5
Results of five metrics using different classifiers on the Dataset A.

Dataset	-	Evaluated metrics	Model				
			VGG19	ResNet152	InceptionV3	DenseNet201	UC-DenseNet
Dataset A	Endoscopic remission (yes/no)	Acc	0.962	0.974	0.971	0.970	0.976
		Pre	0.962	0.975	0.972	0.970	0.975
		Recall	0.946	0.965	0.965	0.956	0.980
		F1-score	0.962	0.974	0.971	0.970	0.976
		AUC	0.962	0.974	0.968	0.972	0.975
	Degree of disease (Mayo 0–3 points)	Acc	0.863	0.887	0.882	0.890	0.906
		Pre	0.836	0.852	0.843	0.858	0.874
		Recall	0.796	0.845	0.839	0.839	0.863
		F1-score	0.807	0.848	0.841	0.847	0.868
		AUC	—	—	—	—	—

Table 6

Results of five metrics using different classifiers on the Dataset B.

Dataset	-	Evaluated metrics	Model				
			VGG19	ResNet152	InceptionV3	DenseNet201	UC-DenseNet
Dataset B	Endoscopic remission (yes/no)	Acc	0.962	0.981	0.971	0.983	0.989
		Pre	0.961	0.981	0.971	0.983	0.989
		Recall	0.962	0.976	0.961	0.988	0.986
		F1-score	0.963	0.981	0.971	0.983	0.989
		AUC	0.966	0.980	0.970	0.984	0.988
	Degree of disease (Mayo 0–3 points)	Acc	0.881	0.898	0.883	0.902	0.916
		Pre	0.857	0.855	0.821	0.871	0.859
		Recall	0.755	0.816	0.816	0.839	0.859
		F1-score	0.778	0.830	0.818	0.852	0.858
		AUC	—	—	—	—	—

**Fig. 6.** The running time of the different algorithms in an epoch.

some researchers have proposed the multi-path attention module to improve the ability to capture feature information in the feature map's spatial and channel dimensions. Such as, Chen et al. [43] Embed the dual attention mechanism into the expanded residual network [44], improving the ability to recognize pathological images. Jiang et al. [45] proposed two attention mechanisms to embed them in a 3D deep dual-path [46] to improve the performance of lung nodule classification. Wang et al. [47] proposed a model of triple attention mechanism and embedded it in the network to improve the ability to extract channel and spatial features in chest images. The model achieved the highest average (Area Under Curve) AUC in the ChestX ray14 dataset. Therefore, we designed an Efficient Attention Mechanism Network (EAM-Net) and applied it to DenseNet for colonoscopy image classification. EAM-Net comprises the Efficient Channel Attention Network (ECA-Net) [48] and the Convolutional Block Attention Module (CBAM) [49]. ECA-Net is based on the squeeze-and-excitation networks [50]. It can realize a

cross-channel interaction strategy without dimensionality reduction through one-dimensional convolution. Moreover, there is also a significant performance improvement when only a small number of parameters are involved.

Feature map $X \in \mathbb{R}^{W \times H \times C}$ first generates a $1 \times 1 \times C$ feature map through channel-wise global average pooling and adaptively calculates the kernel size according to the channel dimension of the feature map. After that, the kernel size is applied to the one-dimensional convolution to obtain the weight of each channel. Such an operation can effectively reduce feature dimensions and extract useful features related to the target feature class. The channel-wise global average pooling $g(X)$ as Eq. 5, sigmoid $\sigma(x)$ as Eq. 6. The channel feature extraction process can be calculated by Eq. 7 and Eq. 8.

$$g(X) = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H X_{ij} \quad (5)$$

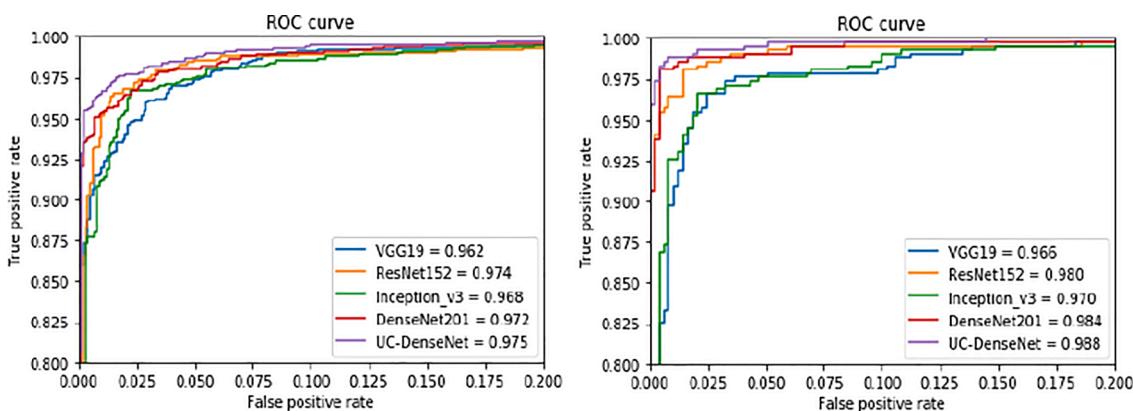
$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

$$k = \varphi(C) = \left\lceil \frac{\log_2(C) + b}{\gamma} \right\rceil \quad (7)$$

$$F_1 = \sigma(\text{Conv1D}_k(g(X))) \quad (8)$$

Conv1D represents one-dimensional convolution, k is the kernel size, C is the network channel dimension, $b = 1$, $\gamma = 2$, and the function of b and γ is to ensure $k > 1$.

After passing ECA-Net, the spatial attention module is introduced to lock the location of the feature information. The channel-enhanced feature map $F_1 \times X$ first applies global average pooling and global max pooling to generate two $H \times W \times 1$ feature maps and connect them

**(a) Endoscopic remission (yes/no) of Dataset A (b) Endoscopic remission (yes/no) of Dataset B****Fig. 7.** Receiver operating characteristic curve.

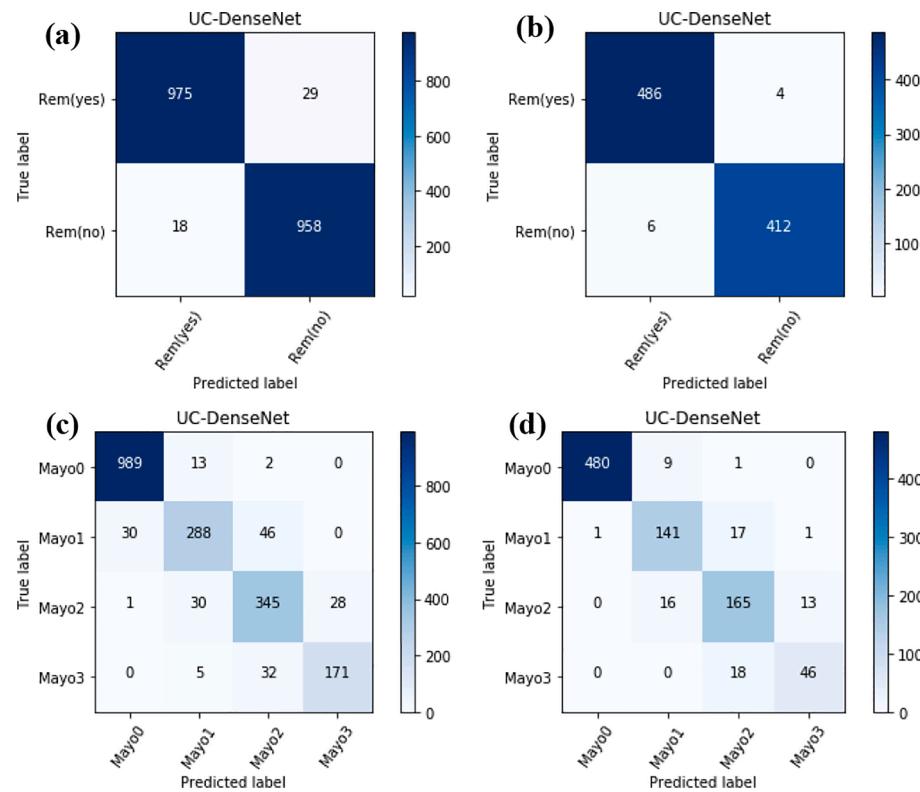


Fig. 8. The confusion matrix of the UC-DenseNet model. (a) and (b) is Endoscopic remission (yes/no) of Dataset A and Dataset B; (c) and (d) is Degree of disease (Mayo 0–3 points) of Dataset A and Dataset B.

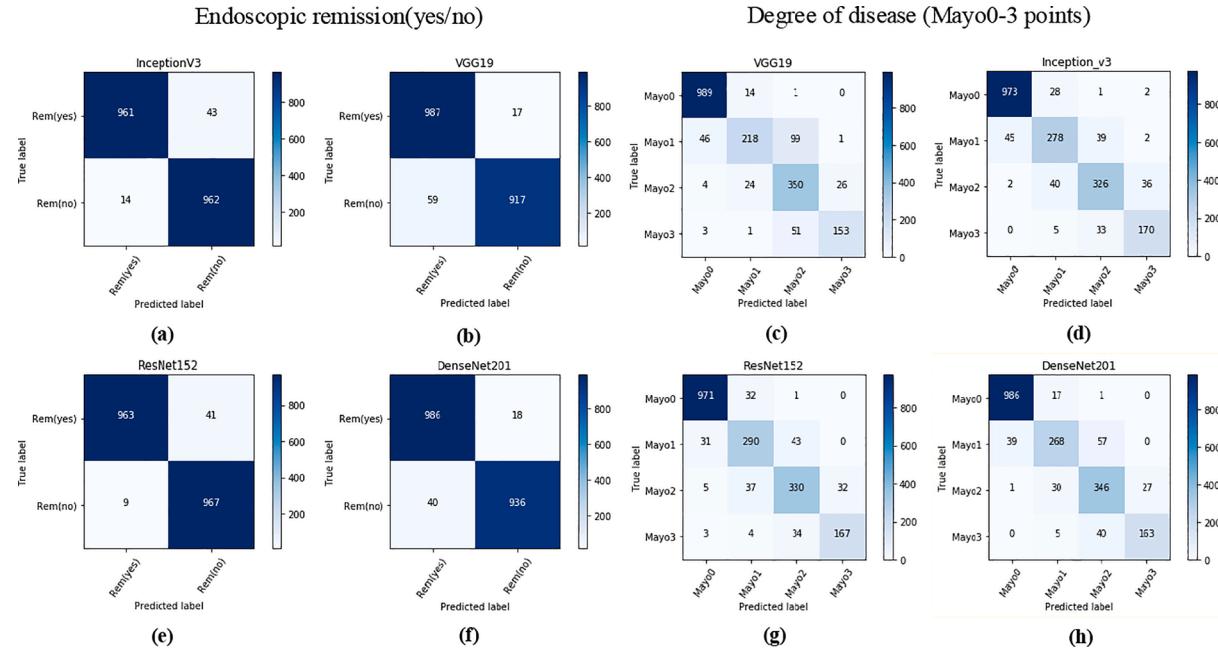


Fig. 9. The confusion matrix of other models (Dataset A).

to aggregate the information of the feature maps. Then use a 7×7 two-dimensional convolution layer to convolve the feature map to obtain a spatial matrix with spatial attention weights. Finally, the obtained spatial attention matrix is multiplied by the original feature map. It can be calculated by Eq.9:

$$F_2 = \sigma(f^{7 \times 7}([\text{AvgPool}(F_1 \times X), \text{MaxPool}(F_1 \times X)])) \quad (9)$$

The feature extraction process of the entire EAM-Net module can be calculated by Eq.10:

$$X_F = X \times F_1 \times F_2 \quad (10)$$

The new feature map X_F is input to the next layer. The overall structure is shown in Fig. 4.

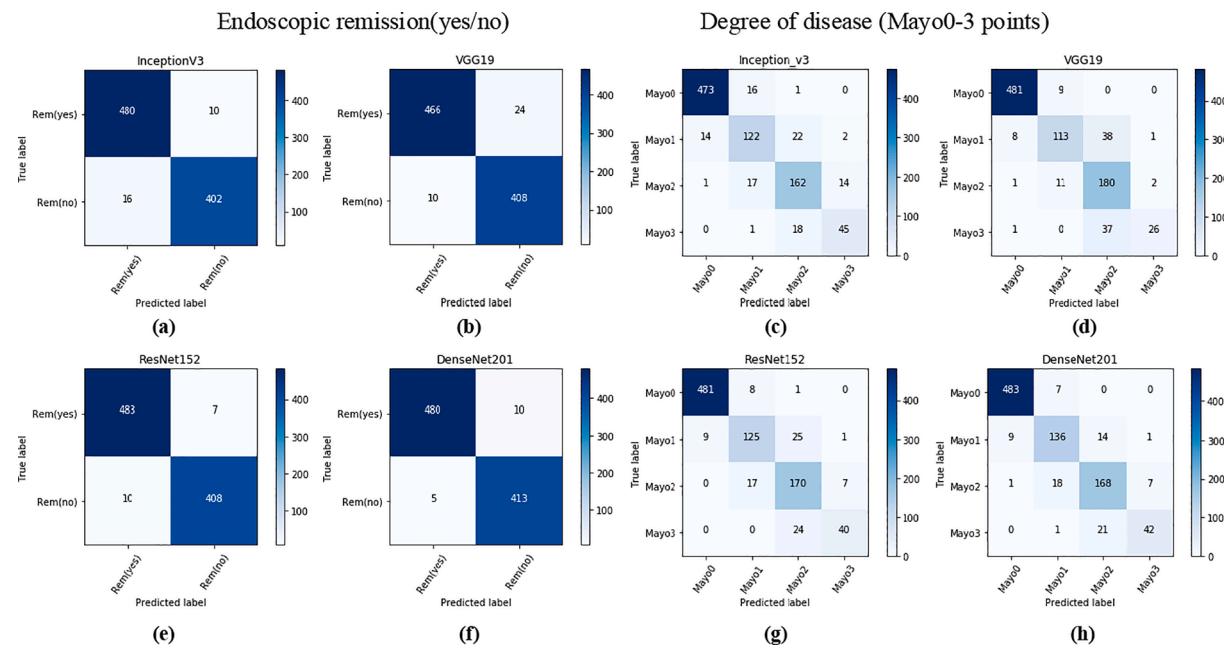


Fig. 10. The confusion matrix of other models (Dataset B).

Table 7
Results of the Test Time Augmentation method.

	Dataset	TTA	Evaluated metrics				
			Acc	Pre	Recall	F1-score	AUC
Endoscopic remission (yes/no)	Dataset A	×	0.976	0.975	0.980	0.976	0.975
		✓	0.977	0.976	0.977	0.977	0.973
	Dataset B	×	0.989	0.989	0.986	0.988	0.988
		✓	0.990	0.992	0.985	0.989	0.989
Degree of disease (Mayo 0–3 points)	Dataset A	×	0.906	0.874	0.863	0.868	—
		✓	0.913	0.887	0.870	0.878	—
	Dataset B	×	0.916	0.859	0.859	0.858	—
		✓	0.919	0.876	0.860	0.867	—

Table 8
Comparison of different attention modules (Endoscopic remission (yes/no)).

Method	Dataset	Evaluated metrics					Parameter
		Acc	Pre	Recall	F1-score	AUC	
DenseNet201 + IndRnn + CBAM	A	0.976	0.976	0.964	0.973	0.976	2.57×10^7
	B	0.986	0.986	0.982	0.985	0.984	—
DenseNet201 + IndRnn + SE-Net	A	0.973	0.970	0.968	0.967	0.970	2.24×10^7
	B	0.983	0.983	0.985	0.983	0.982	—
DenseNet201 + IndRnn + ECA-Net	A	0.974	0.974	0.973	0.974	0.973	2.22×10^7
	B	0.986	0.986	0.983	0.986	0.986	—
UC-DenseNet (DenseNet201 + IndRnn + EAM-Net)	A	0.976	0.975	0.980	0.976	0.975	2.22×10^7
	B	0.989	0.989	0.986	0.989	0.988	—

Table 9
Comparison of different attention modules (Degree of disease (Mayo 0–3)).

Method	Dataset	Evaluated metrics					Parameter
		Acc	Pre	Recall	F1-score	AUC	
DenseNet201 + IndRnn + CBAM	A	0.904	0.870	0.850	0.863	—	2.57×10^7
	B	0.913	0.856	0.852	0.856	—	—
DenseNet201 + IndRnn + SE-Net	A	0.891	0.865	0.846	0.852	—	2.24×10^7
	B	0.906	0.858	0.821	0.834	—	—
DenseNet201 + IndRnn + ECA-Net	A	0.892	0.882	0.838	0.847	—	2.22×10^7
	B	0.911	0.851	0.847	0.848	—	—
UC-DenseNet (DenseNet201 + IndRnn + EAM-Net)	A	0.906	0.874	0.863	0.868	—	2.22×10^7
	B	0.916	0.859	0.859	0.858	—	—

Table 10
Comparison of classification performance on the Kvasir dataset.

	Methods	Evaluated metrics			
		Acc	Pre	Recall	F1-score
Pogorelov et al [28]	6 GF Logistic Model Tree	0.937	–	–	0.747
Naqvi et al [52]	Ensemble model	0.942	–	–	0.767
Ayidzoe et al [53]	Caps Nets	0.910	–	–	–
Şaban et al [54]	ResNet50 + Residual LSTM	0.980	0.980	–	0.980
Agrawal et al [55]	Baseline + Inception-V3 + VGGNet	0.961	–	–	0.847
Yogapriya et al [56]	VGG16 + ResNet-18 + GoogLeNet	0.963	0.965	0.963	0.965
Gamage et al [57]	DenseNet-201 + ResNet-18 + VGG-16	0.973	0.971	0.972	0.972
Our		0.983	0.983	0.983	0.983

Table 11
Ablation experiment of the UC-DenseNet.

Model	Evaluated metrics	Endoscopic remission (yes/no)		Degree of disease (Mayo 0–3 points)	
		Dataset A	Dataset B	Dataset	Dataset
				A	B
DenseNet201	Acc	0.970	0.983	0.890	0.902
	Pre	0.970	0.983	0.858	0.871
	Recall	0.956	0.988	0.839	0.839
	F1-score	0.970	0.983	0.847	0.852
	AUC	0.970	0.983	–	–
DenseNet201 + IndRNN	Acc	0.971	0.984	0.892	0.911
	Pre	0.971	0.984	0.859	0.863
	Recall	0.969	0.988	0.842	0.854
	F1-score	0.971	0.984	0.848	0.858
	AUC	0.971	0.984	–	–
DenseNet201 + IndRNN + EAM-Net (UC-DenseNet)	Acc	0.976	0.989	0.906	0.916
	Pre	0.975	0.989	0.874	0.859
	Recall	0.980	0.986	0.863	0.859
	F1-score	0.976	0.989	0.868	0.858
	AUC	0.975	0.988	–	–

4. Experiment and results

4.1. Experimental environment

We implemented the network architecture of UC-DenseNet in Keras [51]. The Windows10 system and Spyder editor were used in the experiment, running on the GPU computing platform. CPU is AMD Ryzen7 3800X 8-Core Processor, 8G × 2 memory, and GPU is NVIDIA GeForce RTX 2080Ti. All the programs are implemented by the open-source framework Keras with TensorFlow as the backend.

4.2. Evaluation metrics

Accuracy (Acc), Precision (Pre), Recall, F1-score, and AUC are used as metrics in our experiment. True Positive (TP) is correctly diagnosed as diseased. True Negative (TN) is correctly diagnosed as not diseased. False Positive (FP) is incorrectly diagnosed as diseased. False Negative (FN) is incorrectly diagnosed as not diseased.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$Pre = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F1 = \frac{2Pre \times Recall}{Pre + Recall} \quad (14)$$

4.3. Classification performance

We evaluated the proposed method using two different ulcerative colitis data sets, Datasets A and B (Dataset A and Dataset B are two independent data sets, which do not affect each other). The ratio of the training set, validation set, and test set is 3:1:1, and there were no duplicate images between the training set, the validation set, and the test set. Ulcerative colitis image data is converted into an array and then input into the neural network we designed. The augmentation method used was as follows: horizontal flip, vertical flip, Random zoom, randomly rotating from 0 to 20 degrees. We also compared the effects of different parameters on the performance of the UC-DenseNet model and selected the best one, shown in Table 3. We used two different evaluation methods for Dataset A and B. The first is to evaluate whether the condition is in remission. That is, according to the Mayo score of the ulcerative colitis image, mayo = 0 (remission) or mayo ≥ 1 (not remission) is divided into two categories. The other is to evaluate the severity of the disease based on the Mayo score of each image (Mayo = 0, Mayo = 1, Mayo = 2, and Mayo = 3 points. There are four categories, among which Mayo = 3 is serious). Fig. 5 shows the performance of UC-DenseNet under different parameter configurations.

In order to test the classification performance of the UC-DenseNet model, the common network models of VGG19, ResNet152, InceptionV3, and DenseNet201 were compared with the UC-DenseNet model for experimental analysis. The parameter configuration of the optimal performance of each network model on the ulcerative colitis dataset is shown in Table 4.

Table 5 and Table 6 show the results of five metrics using different classifiers. Fig. 6 shows the training time required for different models. Among them, evaluating the severity of the disease includes four classes, so we did not calculate its AUC. The Receiver Operating Characteristic (ROC) curve in Fig. 7 shows that the deep learning model has different capabilities in classifying ulcerative colitis images. The closer the curve is to the upper left corner, the better the effect. The ROC curve of UC-DenseNet is higher than other models. It can be seen from the comprehensive results that, compared with other CNN-based classifier networks, our proposed UC-DenseNet achieves the most advanced classification performance on two different datasets. Fig. 8,9,10 show the confusion matrix of the deep learning model. The confusion matrix based on different models shows that when judging the severity of the disease (4 categories), the misclassification mainly occurs between Mayo 1 and Mayo 2, and Mayo 2 and Mayo 3 (Fig. 8(c), (d), Fig. 9(c), (d), (g), (h), Fig. 10(c), (d), (g), (h)). The main reason is that Mayo 1 and Mayo 2 are relatively similar, and Mayo 2 and Mayo 3 are relatively similar. At the same time, Mayo 1, Mayo 2, and Mayo 3 contain fewer images, far lower than Mayo 0. If these three classes of images can be added later, the accuracy of judging the severity of the disease (4 categories) may be improved.

To improve the diagnosis and classification performance of endoscopic images of patients with ulcerative colitis, we added the TTA data augmentation method to the experiment. Comparative experiments are carried out in Table 7 to verify the effectiveness of this module. The area shown in the original image may lack some crucial features. Therefore, considering the single shape of the test set data, adding TTA for augmentation can improve the model's ability to recognize test images, improving classification accuracy. Experimental results show that adding the above strategy has a good effect on diagnosing the remission and severity of ulcerative colitis.

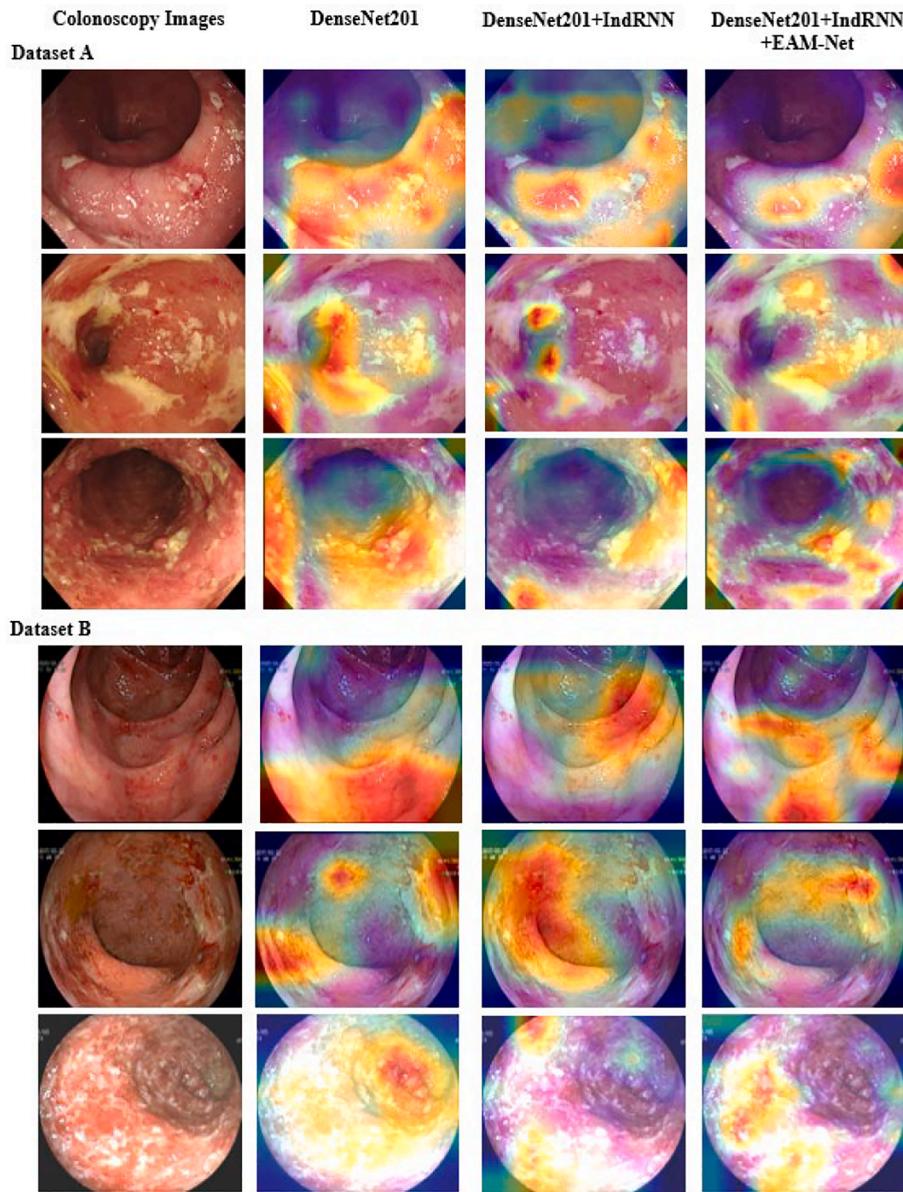


Fig. 11. The Grad-CAM visualizations of the Dataset A and Dataset B.

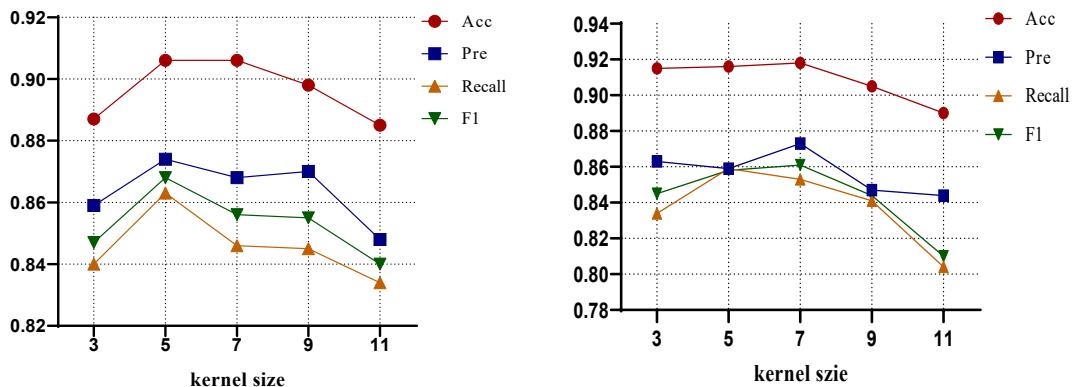


Fig. 12. Network performance under different kernel size. (a) Degree of disease (Mayo 0–3 points) of Dataset A (b) Degree of disease (Mayo 0–3 points) of Dataset B.

4.4. Comparison of different attention modules

In order to prove the effectiveness of the proposed improved attention mechanism, we added different attention modules to the network to compare the performance and parameter sizes of various indicators. Among them, the attention modules compared with EAM-Net are CBAM, SE-Net, and ECA-Net. As shown in [Table 8](#) and [Table 9](#), the combination of EAM-Net and the model in this paper has higher performance on multiple indicators. In addition, EAM-Net has lower parameters than SE-Net and CBAM, saving more computing resources to a certain extent.

4.5. Comparison with related methods on the Kvasir dataset

In the end, in order to further prove the universal applicability of UC-DenseNet, we apply our method to another medical image classification dataset, the Kvasir. It is also compared with the popular deep learning image classification model in the literature. The model training on the dataset follows the hyperparameter settings of UC-DenseNet described in [Table 4](#). [Table 10](#) lists the detailed classification results of the Kvasir dataset.

5. Discussion

To acquire end-to-end ulcerative colitis image classification and diagnosis, we proposed a convolutional network that combines CNN and RNN. We conducted an EAM-Net including ECA-Net and CBAM to improve the ability of key feature extraction. The final UC-DenseNet network is based on the DenseNet-201 model, IndRNN, and an improved attention mechanism. We used images from two different colonoscopy machines to evaluate our method.

[Table 5](#) and [Table 6](#) show the experimental results of the metrics. UC-DenseNet achieves better performance for both ulcerative colitis images. Compared with other networks, on average, the UC-DenseNet shows improvement (Endoscopic remission: yes/no) of approximately 1.08%, 1.06%, 1.81%, 1.06%, 0.94% for Acc, Pre, Recall, F1-score, AUC, respectively. Furthermore, the UC-DenseNet shows improvement (Degree of disease: Mayo 0–3 points) of approximately 2.53%, 1.74%, 4.29%, 3.54% for Acc, Pre, Recall, F1-score, respectively.

To further verify this method, we conducted ablation experiments on the proposed UC-DenseNet. [Table 11](#) shows the detailed experimental results. Some studies have shown that [26,27] combined CNN and RNN to solve medical image classification is an effective method. The most commonly used is to send the pathological image feature information extracted by CNN into RNN. Processing image features through simple connection operations can effectively reduce the amount of calculation and have higher performance. However, some detailed information would be lost. Therefore, we proposed an improved attention mechanism module to retain more detailed features. [Fig. 11](#) show the Grad-CAM [58] visualization of the UC-DenseNet ablation experiment. The feature map of the original model is slightly rough, focusing on more irrelevant areas. In contrast, UC-DenseNet can more accurately focus the position of the lesion in the endoscopic image.

The proposed attention mechanism adaptively determines the kernel size according to the number of channels. However, some studies have shown that the performance of different network models will change with the different kernel size [48]. In order to more accurately determine the most suitable kernel size for different data sets and models, manually adjust kernel size to observe the classification performance. As shown in [Fig. 12](#), the performance is better when the kernel size is 5 (adaptively) or 7. If the kernel is too small, complete features cannot be extracted, and if the kernel is too large, the noise will be introduced, and critical information will be discarded.

The system we propose has some limitations. As shown in [Fig. 5](#) (Exp2, Exp5 and Exp6), when the Batch size or Input size is increased, the Acc in the evaluation index increases. However, because the experiment is limited by GPU performance and memory, if a larger batch

size or input size is set, the model will be better by debugging the parameter configuration. In the future, different technologies will be further used to improve the optimization model. For example, the self-attention negative feedback model is used to select the critical information of the image, and the low-resolution real-time image is reconstructed into a high-resolution image [59]. In addition, we can also use different scoring methods, such as UCEIS [14], to evaluate the inflammation of ulcerative colitis and add the diagnosis of histological images of ulcerative colitis.

6. Conclusions

This research develops and validates a new image classification method for ulcerative colitis. We propose an improved attention mechanism that can generate channel weights through fast one-dimensional convolution, introducing only a small number of parameters and avoiding dimensionality reduction. The features extracted by CNN are input in IndRNN and improved attention mechanism modules, and attention maps are generated and stitched together to highlight the extracted features. During the testing phase, the TTA method is introduced into our model. On three different colonoscopy datasets, our method has achieved superior results, which is competitive in the field of medical image classification.

CRediT authorship contribution statement

Xudong Luo: Conceptualization, Methodology, Formal analysis, Writing-original draft, Software. **Junhua Zhang:** Conceptualization, Methodology, Data curation, Funding acquisition, Supervision, Writing-review & editing. **Zonggui Li:** Formal analysis, Validation. **Ruiqi Yang:** Formal analysis, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research work was supported by National Natural Science Foundation of China (No. 62063034). The authors acknowledge the First Affiliated Hospital of Kunming Medical University in China for providing the colonoscopy images used in this work.

References

- [1] J. Torres, J. Halfvarson, I. Rodríguez-Lago, et al., Results of the Seventh Scientific Workshop of ECCO: Precision medicine in IBD- prediction and prevention of inflammatory bowel disease, *J. Crohn's Colitis* (2021).
- [2] Y.R. Yu, J.R. Rodriguez, Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes, *Semin. Pediatr. Surg.* 26 (2017) 349–355.
- [3] S.N. Ket, R. Palmer, S. Travis, Endoscopic Disease Activity in Inflammatory Bowel Disease, *Curr. Gastroenterol. Rep.* 17 (12) (2015), <https://doi.org/10.1007/s11894-015-0470-0>.
- [4] V.N. Mohammed, M. Samaan, M.H. Mosli, et al., Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis, *Cochrane Database Syst. Rev.* 1 (2018) CD011450.
- [5] G. D'Haens, W.J. Sandborn, B.G. Feagan, K. Geboes, S.B. Hanauer, E.J. Irvine, M. Lémann, P. Marteau, P. Rutgeerts, J. Schölmerich, L.R. Sutherland, A Review of Activity Indices and Efficacy End Points for Clinical Trials of Medical Therapy in Adults With Ulcerative Colitis, *Gastroenterology* 132 (2) (2007) 763–786.
- [6] T. de Lange, S. Larsen, L. Aabakken, Inter-observer agreement in the assessment of endoscopic findings in ulcerative colitis, *BMC Gastroenterol.* 4 (1) (2004), <https://doi.org/10.1186/1471-230X-4-9>.
- [7] Y.R. Yu, J.R. Rodriguez, Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes, *Semin. Pediatr. Surg.* 26 (2017) 349–355.
- [8] D.S. McKeown, M. Goldbaum, W. Cai, C.C.S. Valentim, H. Liang, S.L. Baxter, A. McKeown, G.e. Yang, X. Wu, F. Yan, J. Dong, M.K. Prasadha, J. Pei, M.Y.L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi,

- X. Fu, Y. Duan, V.A.N. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M.A. Lewis, H. Xia, K. Zhang, Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, *Cell* 172 (5) (2018) 1122–1131.e9.
- [9] A. Nogales, Á.J. García-Tejedor, D. Monge, et al., A survey of deep learning models in medical therapeutic areas, *Artif. Intell. Med.* 112 (2021), 102020.
- [10] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88.
- [11] H.P. Bhamhani, A. Zamora, Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis, *Eur. J. Gastroenterol. Hepatol.* 33 (2021) 645–649.
- [12] K. Takenaka, K. Ohtsuka, T. Fujii, et al., Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis, *Gastroenterology* (New York, N.Y. 158 (2020) (1943) 2150–2157.
- [13] M. Byrne, J. East, M. Iacucci, et al., DOP13 Artificial Intelligence (AI) in endoscopy - Deep learning for detection and scoring of Ulcerative Colitis (UC) disease activity under multiple scoring systems, *J. Crohn's Colitis* 15 (2021) S051–S052.
- [14] S.P. Travis, D. Schnell, P. Krzeski, et al., Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity, *Gut* 61 (2012) 535–42.
- [15] F. Zeng, X. Li, X. Deng, L. Yao, G. Lian, An image classification model based on transfer learning for ulcerative proctitis, *Multimedia Syst.* 27 (4) (2021) 627–636.
- [16] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.
- [17] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [18] G. Huang, Z. Liu, L. Maaten, et al., Densely Connected Convolutional Networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.
- [19] M.I. Jordan, Chapter 25 - Serial Order: A Parallel Distributed Processing Approach, *Advances in Psychology*, North-Holland. (1997) 471–495.
- [20] Y. Guo, Y. Liu, E.M. Bakker, et al., CNN-RNN: a large-scale hierarchical image classification framework, *Multimed Tools Appl.* 77 (8) (2018) 10251–10271.
- [21] S. Targ, D. Almeida, K. Lyman, Resnet: Generalizing Residual Architectures, 2016.
- [22] J. Deng, W. Dong, R. Socher, et al., ImageNet: A large-scale hierarchical image database, *IEEE* (2009) 248–255.
- [23] H. Ayesha, S. Iqbal, M. Tariq, M. Abrar, M. Sanaullah, I. Abbas, A. Rehman, M.F. K. Niazi, S. Hussain, Automatic medical image interpretation: State of the art and future directions, *Pattern Recogn.* 114 (2021) 107856, <https://doi.org/10.1016/j.patcog.2021.107856>.
- [24] P. Marentakis, P. Karaiskos, V. Kouloulias, et al., Lung cancer histology classification from CT images based on radiomics and deep learning models, *Med. Biol. Eng. Comput.* 59 (2021) 215–226.
- [25] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [26] R. Yan, F. Ren, Z. Wang, et al., A Hybrid Convolutional and Recurrent Deep Neural Network for Breast Cancer Pathological Image Classification, *IEEE*. (2018) 957–962.
- [27] H. Yao, X. Zhang, X. Zhou, S. Liu, Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification, *Cancers* 11 (2019) 1901.
- [28] K. Pogorelov, K. Randel, C. Griwodz, et al., KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. (2017) 164–169.
- [29] J. Feng, S. Lu, Performance Analysis of Various Activation Functions in Artificial Neural Networks, *J. Phys. Conf. Ser.* 1237 (2019) 22030.
- [30] M. Ebrahim, M. Alsmirat, M. Al-Ayyoub, Performance study of augmentation techniques for HEp2 CNN classification, *IEEE*. (2018) 163–168.
- [31] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, S.W. Baik, Multi-grade brain tumor classification using deep CNN with extensive data augmentation, *J Comput Sci-Neth.* 30 (2019) 174–182.
- [32] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-Term Recurrent Convolutional Networks for Visual Recognition and Description, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (4) (2017) 677–691.
- [33] S. Li, W. Li, C. Cook, et al., Independently Recurrent Neural Network, IndRNN), Building A Longer and Deeper RNN., 2018.
- [34] B. Zhao, S. Li, Y. Gao, IndRNN based long-term temporal recognition in the spatial and frequency domain, in: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, 2020, pp. 368–372.
- [35] P. Zhang, J. Meng, Y. Luan, C. Liu, Plant miRNA-LncRNA Interaction Prediction with the Ensemble of CNN and IndRNN, *Interdiscip Sci.* 12 (1) (2020) 82–89.
- [36] H. Liao, Image Classification Based on IndCRNN Module, in: *Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing*, 2020, pp. 1–6.
- [37] L. Xu, J. Huang, A. Nitanda, et al., A Novel Global Spatial Attention Mechanism in Convolutional Neural Network for Medical Image Classification. (2020).
- [38] F. Wang, M. Jiang, C. Qian, et al., Residual Attention Network for Image Classification, 2017.
- [39] C. Liu, L. Huang, Z. Wei, W. Zhang, Subtler mixed attention network on fine-grained image classification, *Appl. Intell.* 51 (11) (2021) 7903–7916.
- [40] R. Qin, Z. Wang, L. Jiang, et al., Fine-Grained Lung Cancer Classification from PET and CT Images Based on Multidimensional Attention Mechanism, *Complexity* (New York, N.Y.) 2020 (2020) 1–12.
- [41] S. Zeng, Y. Cao, Q. Lin, Z. Man, T. Deng, R. Wang, Deep learning SPECT lung perfusion image classification method based on attention mechanism, *J. Phys. Conf. Ser.* 1748 (4) (2021) 042050, <https://doi.org/10.1088/1742-6596/1748/4/042050>.
- [42] K. Simonyan, A. Zisserman, in: *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2014, pp. 1409–1556.
- [43] X. Chen, J. Wu, L. Lin, et al., A Dual-Attention Dilated Residual Network for Liver Lesion Classification and Localization on CT Images, *IEEE*. (2019) 235–239.
- [44] F. Yu, V. Koltun, T. Funkhouser, Dilated Residual Networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 472–480.
- [45] H. Jiang, F. Gao, X. Xu, F. Huang, S. Zhu, Attentive and ensemble 3D dual path networks for pulmonary nodules classification, *Neurocomputing (Amsterdam)*. 398 (2020) 422–430.
- [46] Y. Chen, J. Li, H. Xiao, et al., Dual path networks. *arXiv preprint arXiv.* 1707 (01629) (2017).
- [47] H. Wang, S. Wang, Z. Qin, et al., Triple attention learning for classification of 14 thoracic diseases using chest radiography, *Med. Image Anal.* 67 (2021), 101846.
- [48] Q. Wang, B. Wu, P. Zhu, et al., ECA-Net: efficient channel attention for deep convolutional neural networks, *arXiv preprint arXiv.* 1910 (03151) (2020).
- [49] S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module. (2018) 3–19.
- [50] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-Excitation Networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (8) (2020) 2011–2023.
- [51] N. Ketkar, Introduction to Keras, Apress, Berkeley, CA, 2017, pp. 97–111.
- [52] S.S.A. Naqvi, S. Nadeem, M.A. Tahir, et al., *Ensemble of Texture and Deep Learning Features for Finding Abnormalities in the Gastro-Intestinal Tract*, Springer International Publishing, Cham, 2018, pp. 469–478.
- [53] M.A. Ayidzoe, Y. Yu, P.K. Mensah, et al., Gabor capsule network with preprocessing blocks for the recognition of complex images, *Mach. Vis. Appl.* 32 (2021) 1–16.
- [54] Ö. Şaban, U. Özkaya, Residual LSTM layered CNN for classification of gastrointestinal tract diseases, *J. Biomed. Inform.* 113 (2021), 103638.
- [55] T. Agrawal, R. Gupta, S. Sahu, et al., SCL-UMD at the Medico Task-MediaEval 2017: Transfer Learning based Classification of Medical Images, *MediaEval*. (2017).
- [56] J. Yogapriya, V. Chandran, M. G. Sumithra, et al., Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model, *Computational and mathematical methods in medicine*. (2021) 1–12.
- [57] C. Gamage, I. Wijesinghe, C. Chitraranjan, et al., GI-Net: anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning, in: *2019 Moratuwa Engineering Research Conference*, 2019, pp. 66–71.
- [58] R. Ramprasaath, M.C. Selvaraju, Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.
- [59] X. Liu, S. Chen, L. Song, et al., Self-attention negative feedback network for real-time image super-resolution, *J. King Saud Univ. Comput. Inf. Sci.* (2021).