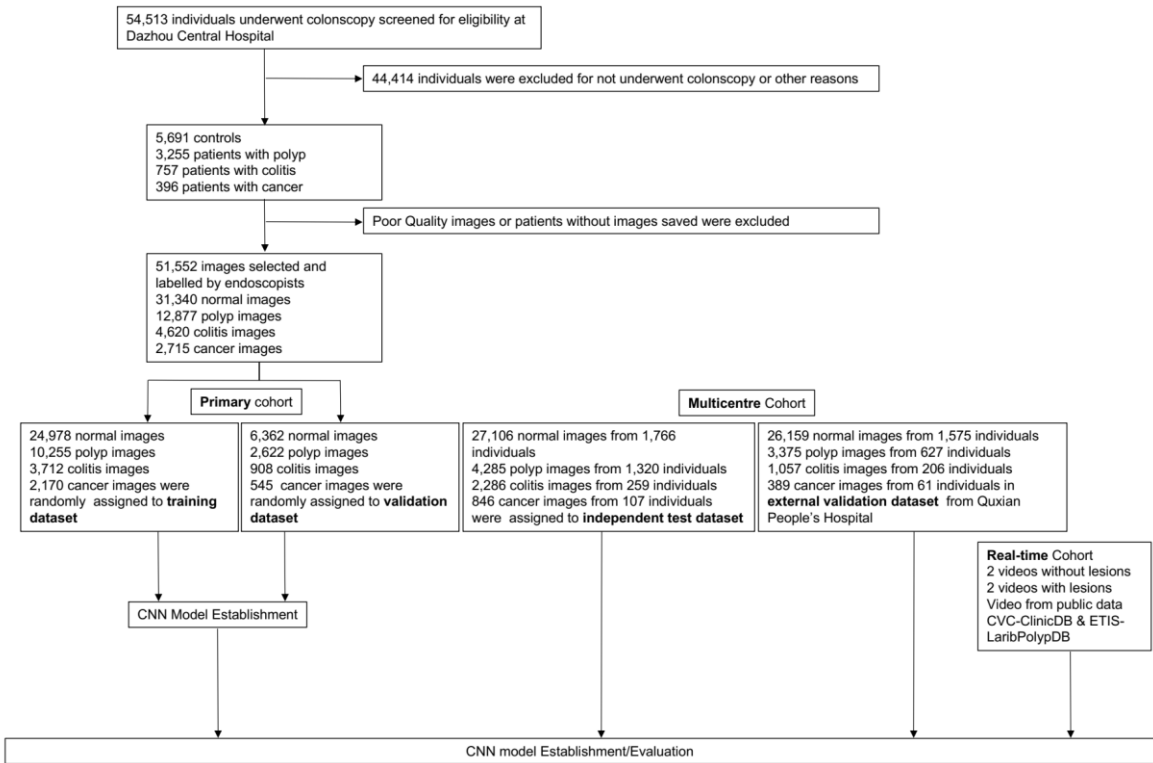
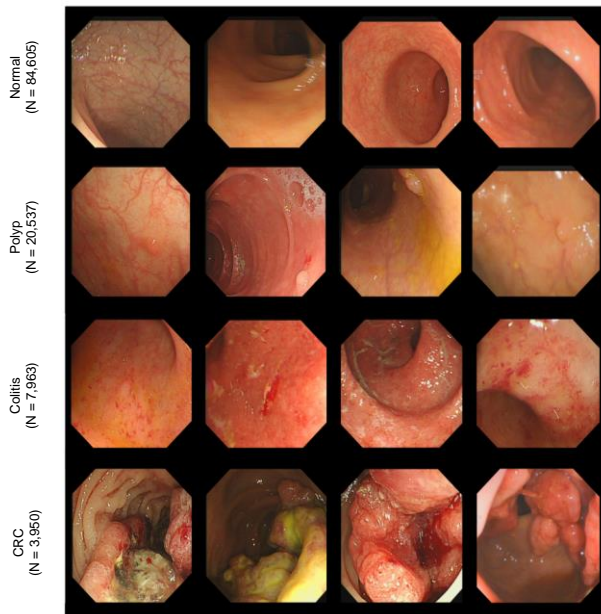


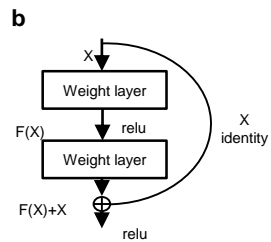
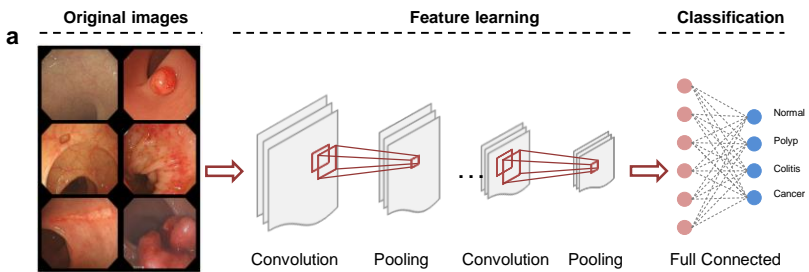
Supplementary files



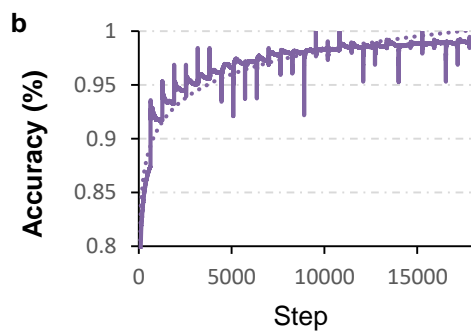
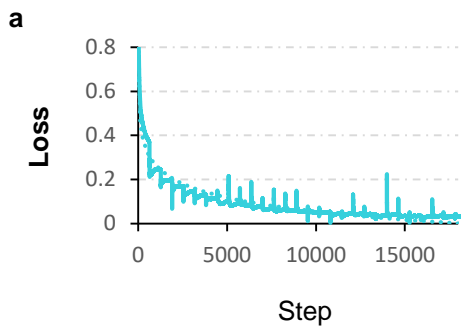
Supplementary Fig. 1 Flowchart of this study.



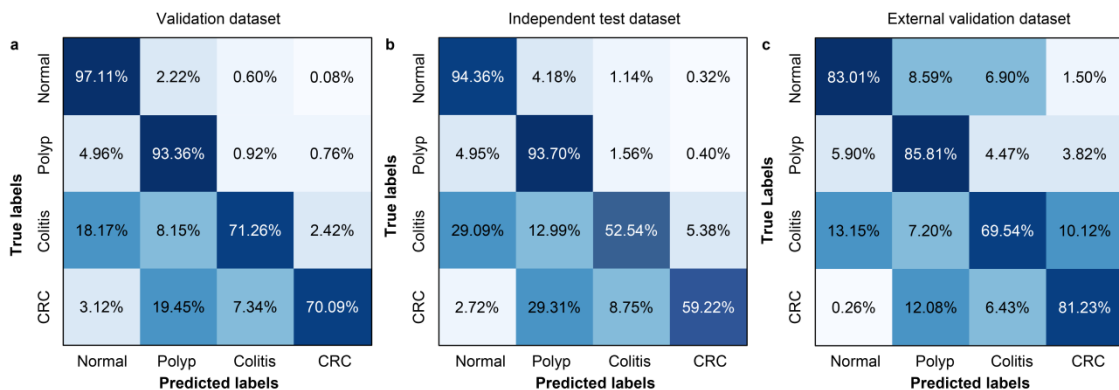
Supplementary Fig. 2 Typical pictures of colorectal diseases.



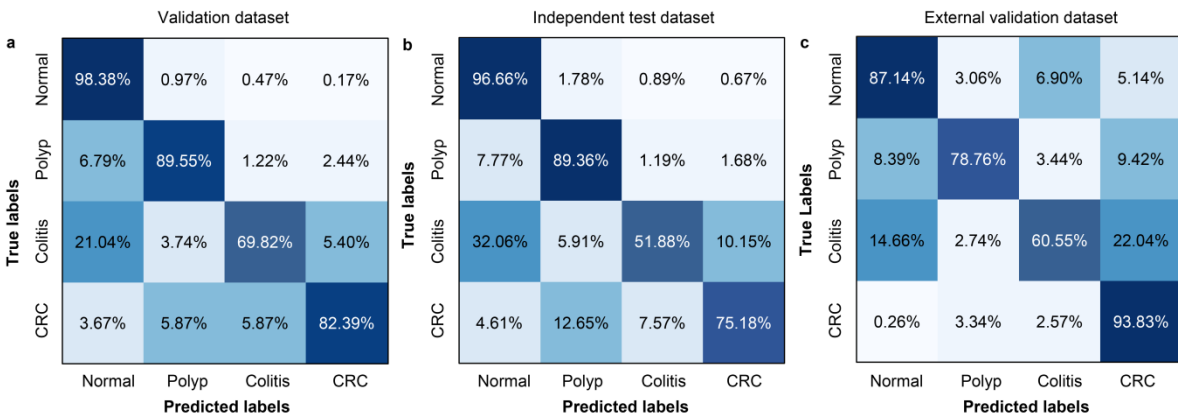
Supplementary Fig. 3 Patterns of the convolutional neural network.



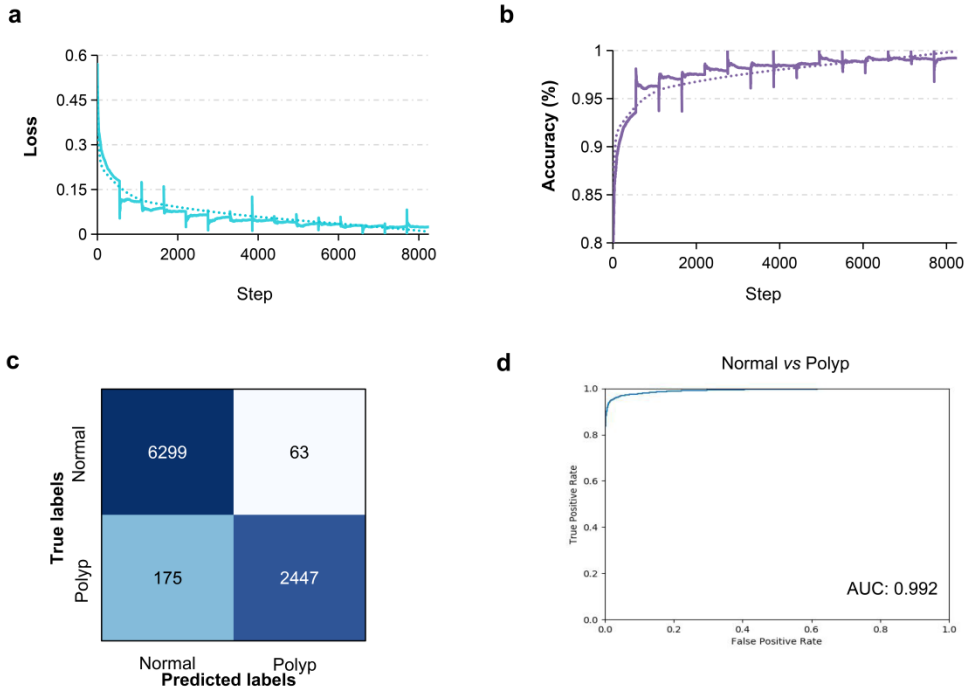
Supplementary Fig. 4 Performance development in the training dataset. In the 18,444 steps training process of the whole model, (a) shows the loss of cross entropy, (b) shows the improvement of accuracy.



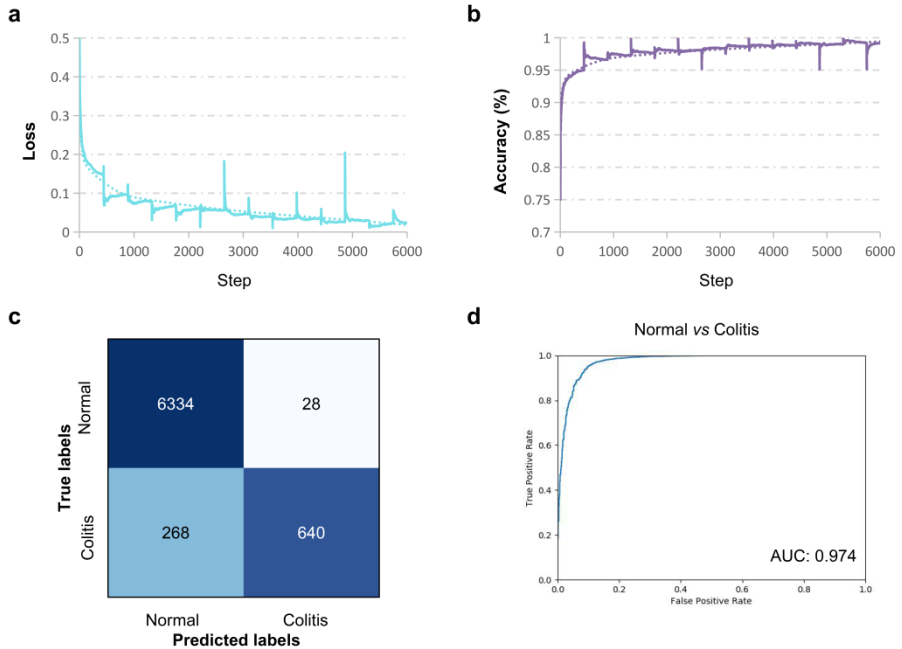
Supplementary Fig. 5 Diagnostic performance of CNN2 model in multi-class classification. **a** Confusion Matrix of the CNN2 model in validation dataset. **b** Confusion matrix of the CNN2 model in independent test dataset. **c** Confusion matrix of the CNN2 model in external validation dataset.



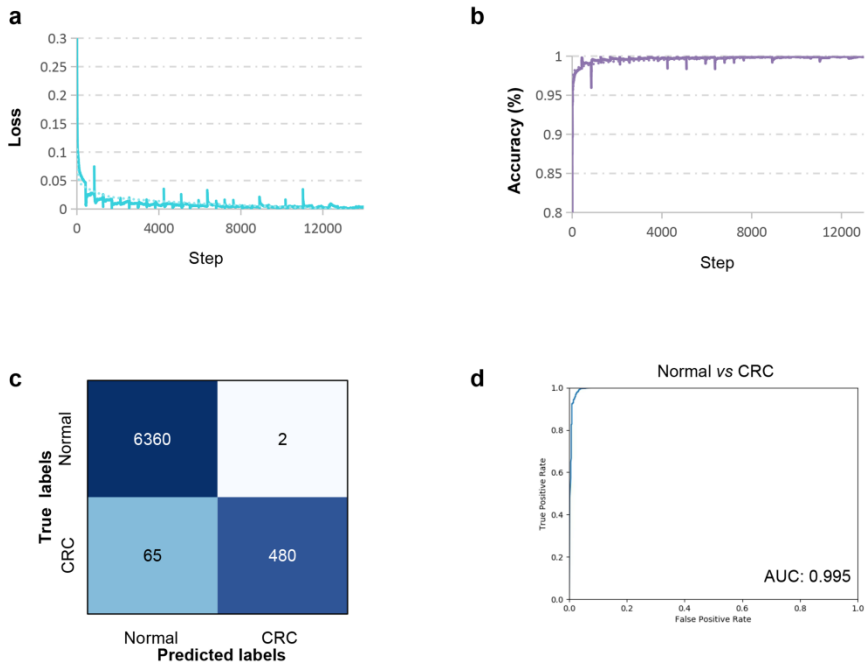
Supplementary Fig. 6 Diagnostic performance of CNN3 model in multi-class classification. **a** Confusion matrix of the CNN3 model in validation dataset. **b** Confusion matrix of the CNN3 model in independent test dataset. **c** Confusion matrix of the CNN3 model in external validation dataset.



Supplementary Fig. 7 Performance of polyp classification in CNN1-NPo model. In the 8,250 steps training process of the whole model, **(a)** shows the loss of cross entropy, **(b)** shows the improvement of accuracy. **c** Confusion matrix of the CNN1-NPo model in classifying of normal and polyp in the validation dataset. **d** Corresponding ROC curves. CNN1-NPo model represents the ResNet-50 classifying the normal and polyp.



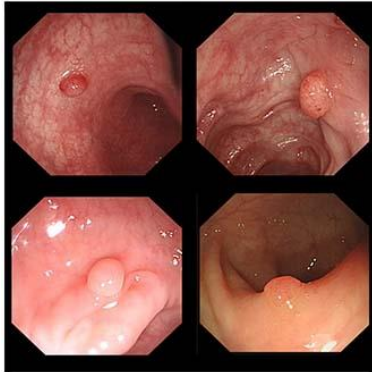
Supplementary Fig. 8 Diagnostic performance of normal vs colitis in CNN1-NCo model. In the 6,188 steps training process of the whole model, **(a)** shows the loss of cross entropy, **(b)** shows the improvement of accuracy. **(c)** Confusion matrix of the CNN1-NCo model in classifying of normal and colitis in the validation dataset. **(d)** Corresponding ROC curves. CNN1-NCo model represents the ResNet-50 classifying the normal and colitis.



Supplementary Fig. 9 Diagnostic performance of normal vs cancer in CNN1-NCa model. In the 13,992 steps training process of the whole model, **(a)** shows the loss of cross entropy, **(b)** shows the improvement of accuracy.

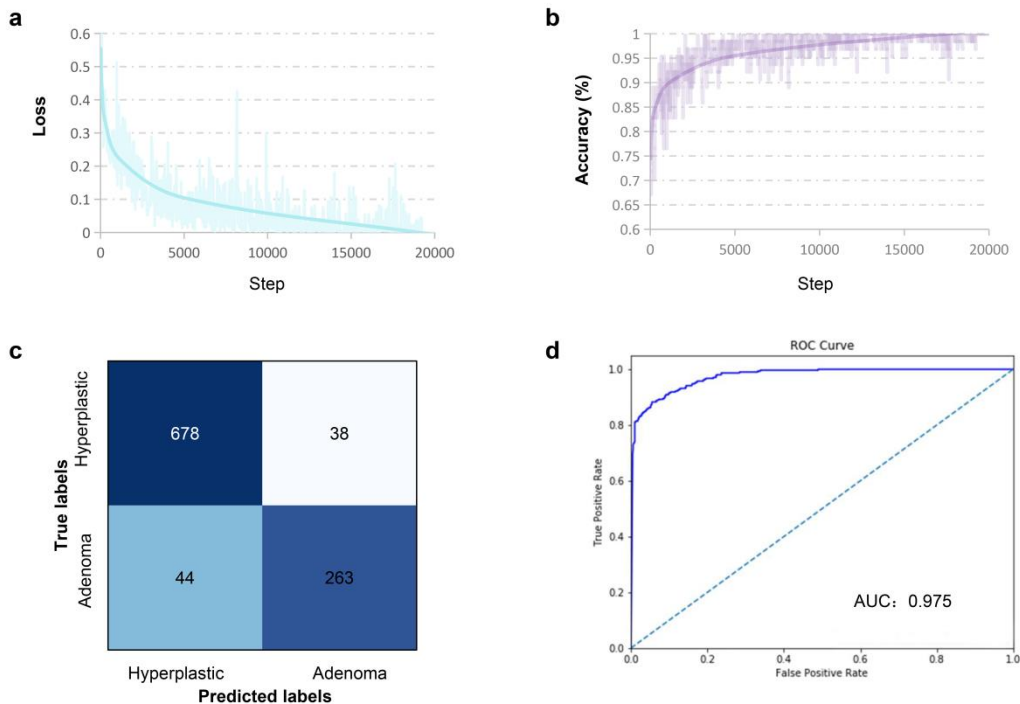
c Confusion matrix of the CNN1-NCa model in classifying of normal and cancer in the validation dataset. **d** Corresponding ROC curves. CNN1-NCa model represents the ResNet-50 classifying the normal and CRC.

Adenoma
N = 1536



Hyperplastic
N = 3580

Supplementary Fig. 10 Typical picture of adenoma and hyperplastic.



Supplementary Fig. 11 Performance of polyp subgroup classification in CNN1-HA model. In the 19,995 steps training process of the whole model, **(a)** shows the loss of cross entropy, **(b)** shows the improvement of accuracy. **c** Confusion matrix of the CNN1-HA model in classifying of hyperplastic polyp and adenoma polyp in the validation dataset. **d** Corresponding ROC curves. CNN1-HA model represents the ResNet-50 classifying the hyperplastic and adenomatous polyp.

Supplementary Table 1. Characteristics of the subjects

Characteristics	Dazhou Central Hospital cohort		Quxian People's Hospital cohort
	Training and Validation dataset	Independent test dataset	External validation dataset
Age, mean \pm SD, years	52.4 \pm 13.2	52.7 \pm 12.7	53.3 \pm 13.4
Sex, n (%)			
Female	5,031 (49.8%)	1,732 (50.2%)	1,152 (47.0%)
Male	5,068 (50.2%)	1,720 (49.8%)	1,301 (53.0%)
Normal, n (%)	5,691 (56.4%)	1,766 (51.2%)	1,575 (63.8%)
Polyp, n (%)	3,255 (32.2%)	1,320 (38.2%)	627 (25.4%)
Colitis, n (%)	757 (7.5%)	259 (7.5%)	206 (8.3%)
Cancer, n (%)	396 (3.9%)	107 (3.1%)	61 (2.5%)

Supplementary Table 2. Summary of CNN1 model performance in discriminating multi-class in validation dataset

Items	Support Images	Precision	Recall	F1-score	Accuracy
Normal	6,362	0.960 (0.955, 0.965)	0.975 (0.972, 0.978)	0.967 (0.963, 0.971)	
Polyp	2,622	0.938 (0.929, 0.947)	0.911 (0.900, 0.922)	0.924 (0.914, 0.934)	
Colitis	908	0.801 (0.775, 0.827)	0.794 (0.768, 0.820)	0.797 (0.771, 0.823)	
Cancer	545	0.820 (0.788, 0.852)	0.794 (0.760, 0.828)	0.807 (0.774, 0.840)	0.934 (0.929, 0.939)
Macro avg	10,437	0.880 (0.874, 0.886)	0.869 (0.863, 0.875)	0.874 (0.868, 0.880)	
Weighted avg	10,437	0.933 (0.928, 0.938)	0.934 (0.929, 0.939)	0.933 (0.928, 0.938)	

95% confidence intervals are included in brackets

Supplementary Table 3. Summary of CNN1 model performance in discriminating multi-class in independent test dataset

Items	Support Images	Precision	Recall	F1-score	Accuracy
Normal	27,106	0.981 (0.979, 0.983)	0.838 (0.834, 0.842)	0.904 (0.901, 0.907)	
Polyp	4,285	0.737 (0.727, 0.747)	0.903 (0.894, 0.192)	0.812 (0.800, 0.824)	
Colitis	2,286	0.330 (0.311, 0.349)	0.735 (0.717, 0.753)	0.455 (0.435, 0.475)	
Cancer	846	0.600 (0.567, 0.633)	0.710 (0.680, 0.740)	0.650 (0.628, 0.682)	0.836 (0.832, 0.840)
Macro avg	34,523	0.662 (0.657, 0.667)	0.797 (0.793, 0.801)	0.705 (0.700, 0.710)	
Weighted avg	34,523	0.898 (0.895, 0.901)	0.836 (0.832, 0.840)	0.856 (0.852, 0.860)	

95% confidence intervals are included in brackets

Supplementary Table 4. Summary of CNN1 model performance in discriminating multi-class in external validation dataset

Items	Support Images	Precision	Recall	F1-score	Accuracy
Normal	26,159	0.990 (0.989, 0.991)	0.973 (0.971, 0.975)	0.981 (0.962, 1.000)	
Polyp	3,375	0.814 (0.801, 0.827)	0.976 (0.971, 0.981)	0.888 (0.878, 0.898)	
Colitis	1,057	0.918(0.902, 0.934)	0.793 (0.769, 0.817)	0.851 (0.830, 0.872)	
Cancer	389	0.991 (0.982, 1.000)	0.810 (0.771, 0.849)	0.891 (0.860, 0.922)	0.965 (0.963, 0.967)
Macro avg	30,980	0.928 (0.925, 0.931)	0.888 (0.891, 0.891)	0.903 (0.900, 0.906)	
Weighted avg	30,980	0.968 (0.966, 0.970)	0.965 (0.963, 0.967)	0.965 (0.963, 0.967)	

95% confidence intervals are included in brackets

Supplementary Table 5. Performance of different model in classification of normal, polyp, colitis and CRC in validation dataset

	Items	Normal	Polyp	Colitis	CRC
Precision	ResNet-50	0.960 (0.955,0.965)	0.938 (0.928, 0.948)	0.801 (0.775, 0.827)	0.820 (0.790, 0.850)
	VGG16	0.952 (0.947, 0.957)	0.884 (0.872, 0.896)	0.864 (0.842, 0.886)	0.890 (0.864, 0.916)
	VGG19	0.941 (0.935, 0.947)	0.944 (0.934, 0.954)	0.871 (0.849, 0.893)	0.784 (0.749, 0.819)
Recall	ResNet-50	0.975 (0.972, 0.978)	0.911(0.901, 0.921)	0.794 (0.768, 0.820)	0.794 (0.760, 0.828)
	VGG16	0.971 (0.967, 0.975)	0.934 (0.924, 0.944)	0.713 (0.683, 0.743)	0.701 (0.663, 0.739)
	VGG19	0.984 (0.981, 0.987)	0.895 (0.884, 0.906)	0.698 (0.669, 0.727)	0.824 (0.792, 0.856)
F1-score	ResNet-50	0.967 (0.963, 0.971)	0.924 (0.914, 0.934)	0.797 (0.771, 0.823)	0.807 (0.774, 0.840)
	VGG16	0.961 (0.956, 0.966)	0.908 (0.897, 0.919)	0.781 (0.754, 0.808)	0.784 (0.749, 0.819)
	VGG19	0.962 (0.957, 0.967)	0.919 (0.909, 0.929)	0.775 (0.748, 0.802)	0.804 (0.771, 0.837)
Accuracy	ResNet-50	0.934 (0.929, 0.939)			
	VGG16	0.925 (0.920, 0.930)			
	VGG19	0.928 (0.923, 0.933)			

95% confidence intervals are included in brackets

Supplementary Table 6. Summary of CNN1 model performance in discriminating normal and different large intestinal diseases

	Items	Support Images	Precision	Recall	F1-score	AUC	Accuracy
CNN1-NPo	Polyp	2,622	0.975 (0.968, 0.982)	0.933 (0.923, 0.943)	0.954 (0.946, 0.962)	0.992 (0.990, 0.994)	0.966 (0.963, 0.969)
	Normal	6,362	0.973 (0.969, 0.977)	0.990 (0.988, 0.992)	0.981 (0.978, 0.984)		
	Macro avg	8,984	0.974 (0.971, 0.977)	0.962 (0.958, 0.966)	0.968 (0.965, 0.971)		
	Weighted avg	8,984	0.974 (0.971, 0.977)	0.973 (0.970, 0.976)	0.973 (0.970, 0.976)		
CNN1-NCo	Colitis	908	0.958 (0.945, 0.971)	0.705 (0.675, 0.735)	0.812 (0.786, 0.838)	0.974 (0.970, 0.978)	0.959 (0.954, 0.964)
	Normal	6,362	0.959 (0.954, 0.964)	0.996 (0.995, 0.997)	0.977 (0.974, 0.980)		
	Macro avg	7,270	0.959 (0.954, 0.964)	0.851 (0.843, 0.859)	0.895 (0.888, 0.902)		
	Weighted avg	7,270	0.958 (0.953, 0.963)	0.960 (0.955, 0.965)	0.956 (0.951, 0.961)		
CNN1-NCa	CRC	545	0.996 (0.992, 1.000)	0.881 (0.854, 0.908)	0.935 (0.914, 0.956)	0.995 (0.993, 0.997)	0.990 (0.988, 0.992)
	Normal	6,362	0.990 (0.988, 0.992)	0.999 (0.998, 1.000)	0.994 (0.992, 0.996)		
	Macro avg	6,907	0.993 (0.991, 0.995)	0.940 (0.934, 0.946)	0.965 (0.960, 0.970)		
	Weighted avg	6,907	0.990 (0.988, 0.992)	0.990 (0.988, 0.992)	0.989 (0.987, 0.991)		

95% confidence intervals are included in brackets

Abbreviations: AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value

Supplementary Table 7. All CNN models in this study and their training steps

Model	Algorithm	Object	Training steps (Steps)
CNN1	ResNet-50	discriminate normal, polyp, colitis, and CRC	18,444
CNN1-NPo	ResNet-50	discriminate normal and polyp	8,250
CNN1-NCo	ResNet-50	discriminate normal and colitis	6,188
CNN1-NCa	ResNet-50	discriminate normal and CRC	13,992
CNN1-HA	ResNet-50	discriminate hyperplastic and adenomatous polyp	19,995
CNN2	VGG16	discriminate normal, polyp, colitis, and CRC	16,536
CNN3	VGG19	discriminate normal, polyp, colitis, and CRC	13,356