Appendices

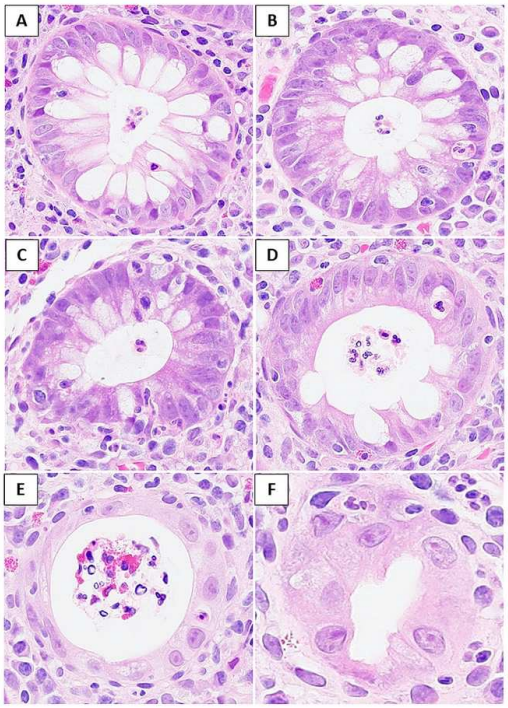
Statistical appendix

Fisher Z- transformation to compare Pearson correlations between endoscopic scores and between histological scores/features was performed using R-package cocor (http://CRAN.R-project.org/package=cocor).

CARRoT statistical package combines principles of good practice from machine learning, such as cross-validation and those in medical statistics, such as best subset regression restricted by the rule of ten events per variable ("one in ten rule").[35] CARRoT has been previously successfully used for outcome prediction in clinical studies,[36,37] in order to facilitate the choice of logistic regression model which leads to the best predictive power quantified as average AUROC. Once the best model was chosen, the corresponding regression model was fit to the whole dataset and the odds ratios (ORs) were computed.

Histological appendix

Some case were considered particularly challenging due to of 'cryptitis' and 'crypt abscess' on some atypical microscopic appearances, for example, when only one neutrophil was present within cryptal epithelium (for calling cryptitis), or only rare neutrophils were seen floating in the lumen of a crypt while there were no neutrophils within cryptal epithelium or no or only minimal damage of cryptal epithelium (as exampled in Supplementary Figure 5). Therefore, to strictly adhere to the predetermined criteria in scoring all cases is important to the pathologists.



Supplementary Figure 5

Artificial intelligence appendix

Two pathologists analysed the biopsies (75 remission, 63 with active disease) and annotated the presence of neutrophils in the target regions (lamina propria, surface epithelium, cryptal epithelium and cryptal lumen) using an in-house software. Whole-slide-images were then down-sampled to 20x resolution, divided into 512x512x3 patches with a 50% overlap, and patches with less than 20% of tissue were excluded. The development of the Convolutional Neural Network (CNN) framework was divided in two models. In the first model through a feature extractor (based on the VGG16 architecture) and a feature-refinement (based on a Squeeze and Excitation network) identified patches with neutrophils. This approach was implemented using Tensorflow 2.3.1 with Python 3.6. Experiments were conducted on the NVIDIA DGX A100 system. The model was trained for 20 epochs using a learning rate of 0.001 with a batch size of 64. Note that the number of epochs is not very high since we used a model pre-trained for histological imaging with a different task. Adadelta optimizer was applied to try to minimize the binary cross-entropy (BCE) loss function at each epoch. The second model combined the information previously linked to the patches to provide a prediction regarding presence or absence of active UC from each WSI. This model was obtained through a multiple instance learning (MIL), a weakly supervised type of learning, trained for 12 epochs with a learning rate of 0.001, minimizing the binary cross-entropy loss and with a batch size of 1 patient. All experiments were conducted on the NVIDIA DGX A100 system.

Supplementary Table 1. Modified scores for analysing correlations between histopathologic components and endoscopic scores / specified clinical outcomes

Histopathologic component	Score	Source score(s)
a. Neutrophil infiltration in lamina propria	0 1 2	If RHI-neutrophil-lamina propria = 0 If RHI-neutrophil-lamina propria ≤2 If RHI-neutrophil-lamina propria = 3
b. Neutrophil infiltration in cryptal epithelium	0 1 2	If RHI-neutrophil-epithelium = 0 If RHI-neutrophil-epithelium ≤2 If RHI-neutrophil-epithelium = 3
c. Neutrophil infiltration in total	= (a + b)	
d. Chronic inflammation	0 1 2	If RHI-neutrophil-chronic inflammation = 0 If RHI-neutrophil-chronic inflammation ≤2 If RHI-neutrophil-chronic inflammation = 3
e. Combined active and chronic inflammation	= (c + d)	
f. Basal plasmacytosis	0, 1, 2	Same subscore from ECAP
g. Eosinophil increase	0, 1, 2	Same subscore from ECAP
h. Crypt architecture distortion	0 1 2	

Supplementary table 2. PHRI correlation with endoscopic scores and specified clinical outcomes / relapse

In correlation with	PHRI	
	All Patients	MES 0 Patients
	Spearman ρ	*Spearman ρ*
PICaSSO mucosal score	0.809705	0.243540
PICaSSO vascular score	0.715309	0.100339
12-months specified clinical outcomes	0.403166	0.031718

Supplementary Table 3

	PHRI	
	12 months outcomes	
	All patients	MES 0 patients
w/ specified clinical outcomes	2.36 ± 1.67	0.36 ± 0.93
w/o specified clinical outcomes	0.84 ± 1.46	0.22 ± 0.74
Wilcoxon test *p* value	7.399e-11	0.7069

Supplementary Table 4. Best cut-off values of PHRI in predicting specified clinical outcomes / relapse

A. For all patients				
For 12-month specified clinical outcomes				
Best value	**AUROC (95% CI)**	**Specificity (95% CI)**	**Sensitivity (95% CI)**	**Accuracy (95% CI)**
PHRI_max ≤1	0.730 (0.665, 0.795)	0.781 (0.566, 0.845)	0.680 (0.462, 0.777)	0.752 (0.603, 0.794)
PHRI_sigmoid ≤1	0.693 (0.628, 0.759)	0.868 (0.669, 0.915)	0.526 (0.320, 0.636)	0.770 (0.634, 0.804)
PHRI_rectum ≤0	0.722 (0.659, 0.785)	0.814 (0.628, 0.888)	0.632 (0.471, 0.731)	0.761 (0.623, 0.814)

B. For patients in ER (MES 0)				
For 12-month specified clinical outcomes				
Best value	**AUROC (95% CI)**	**Specificity (95% CI)**	**Sensitivity (95% CI)**	**Accuracy (95% CI)**
PHRI_max ≤1	0.517 (0.415, 0.619)	0.939 (0.566, 0.845)	0.143 (0.000, 0.357)	0.863 (0.683, 0.904)
PHRI_sigmoid ≤2	0.506 (0.432, 0.581)	0.977 (0.844, 1.000)	0.071 (0.000, 0.214)	0.890 (0.770, 0.911)
PHRI_rectum ≤1	0.539 (0.440, 0.639)	0.970 (0.780, 0.992)	0.143 (0.000, 0.286)	0.891 (0.725, 0.912)

Supplementary table 4. Best cut-off values of PHRI in predicting specified clinical outcomes / relapse

Supplementary Table 5. Inter-rater agreement on different histological scores and their components

Histologic Index/Score		Intra-Class Correlation (ICC) Coefficient, overall (95% CI)	
Name	Histologic Components	ICC	Total ICC
PHRI			0.84 (0.78, 0.90)
	Neutrophils in lamina propria	**0.85** (0.78, 0.90)	
	Neutrophils in surface epithelium	**0.70** (0.60, 0.79)	
	Cryptitis	0.66 (0.55, 0.76)	
	Crypt abscess	0.63 (0.51, 0.73)	
RHI			0.77 (0.69, 0.85)
	Chronic inflammation	0.68 (0.57, 0.76)	
	Neutrophils in lamina propria	**0.80** (0.72, 0.87)	
	Neutrophils in epithelium	**0.72** (0.63, 0.81)	
	Erosion / Ulceration	0.62 (0.51, 0.73)	
NHI*			0.85 (0.79, 0.90)
VSS			0.77 (0.68, 0.84)
	Crypt abscess	0.51 (0.39, 0.64)	
	Erosion / Ulceration	0.44 (0.32, 0.58)	
	Neutrophils in lamina propria	**0.79** (0.71, 0.86)	
ECAP			0.87 (0.82, 0.92)
	Extent	0.55 (0.43, 0.67)	
	Crypt architectural alteration	0.44 (0.32, 0.58)	
	Paneth cell metaplasia/hyperplasia	0.41 (0.29, 0.55)	
	Surface epithelium (reactive, neutrophils, erosion /ulceration)	0.73 (0.64, 0.82)	
	Cryptitis	0.69 (0.58, 0.78)	
	Crypt abscess	0.55 (0.44, 0.68)	
	Crypt destruction	0.47 (0.34, 0.60)	
	Mononuclear cells in lamina propria	0.74 (0.64, 0.82)	
	Basal plasmacytosis	0.68 (0.58, 0.78)	
	Neutrophils in lamina propria	**0.85** (0.78, 0.90)	
	Eosinophils in lamina propria	0.42 (0.30, 0.56)	
	Lymphoid follicles or aggregates	0.43 (0.31, 0.57)	
GS			0.82 (0.75, 0.88)
	Grade 0 (architectural change)	0.59 (0.47, 0.71)	
	Grade 1 (chronic inflammatory infiltrate)	0.68 (0.57, 0.77)	
	Grade 2a (eosinophils in lamina propria)	0.40 (0.28, 0.53)	
	Grade 2b (neutrophils in lamina propria)	**0.80** (0.62, 0.87)	
	Grade 3 (neutrophils in epithelium)	**0.71** (0.61, 0.80)	
	Grade 4 (crypt destruction)	0.55 (0.43, 0.67)	
	Grade 5 (erosion / ulceration)	0.60 (0.49, 0.71)	

*Breakdown ICC data for Nancy index was not computed