Original research

PICaSSO Histologic Remission Index (PHRI) in ulcerative colitis: development of a novel simplified histological score for monitoring mucosal healing and predicting clinical outcomes and its applicability in an artificial intelligence system

Xianyong Gui,1 Alina Bazarova,2,3 Rocío del Amor,4 Michael Vieth,5,6 Gert de Hertogh,7 Vincenzo Villanacci,8 Davide Zardo,9 Tommaso Lorenzo Parigi,10,11 Elin Synnøve Røyet,12 Uday N Shivaiah,2,13 Melissa Anna Teresa Monica,8 Giulio Mandelli,8 Pradeep Bhandari,14 Silvio Danese,15,16 Jose G Ferraz,17 Bu’Hussain Hayee,18 Mark Lazarov,19 Adolfo Parra-Blanco,20 Luca Pastorelli,21,22 Remo Panaccione,17 Timo Rath,23 Gian Eugenio Tontini,24 Ralf Kiesslich,25 Raf Bisschops,26 Enrico Grisano,27,28 Valery Naranjo,4 Subrata Ghosh,2,29 Marietta Iacucci2,13,30

ABSTRACT

Histological remission is evolving as an important treatment target in UC. We aimed to develop a simple histological index, aligned to endoscopy, correlated with clinical outcomes, and suited to apply to an artificial intelligence (AI) system to evaluate inflammatory activity.

Methods Using a set of 614 biopsies from 307 patients with UC enrolled into a prospective multicentre study, we developed the Paddington International virtual ChromoendoScopy ScOre Histologic Remission Index (PHRI). Agreement with multiple other histological indices and validation for inter-reader reproducibility were assessed. Finally, to implement PHRI into a computer-aided diagnosis system, we trained and tested a novel deep learning strategy based on a CNN architecture to detect neutrophils, calculate PHRI and identify active from quiescent UC using a subset of 138 biopsies.

Results PHRI is strongly correlated with endoscopic scores (Mayo Endoscopic Score and UC Endoscopic Index of Severity and PICaSSO) and with clinical outcomes (hospitalisation, colectomy and initiation or changes in medical therapy due to UC flare-up). A PHRI score of 1 could accurately stratify patients’ risk of adverse outcomes (hospitalisation, colectomy and treatment optimisation due to flare-up) within 12 months. Our inter-reader agreement was high (intraclass correlation 0.84). Our preliminary AI algorithm differentiated active from quiescent UC with 78% sensitivity, 91.7% specificity and 86% accuracy. Conclusions PHRI is a simple histological index in UC, and it exhibits the highest correlation with endoscopic activity and clinical outcomes. A PHRI-based AI system was accurate in predicting histological remission.

Significance of this study

► We developed a new simple histological index for UC, Paddington International virtual ChromoendoScopy ScOre Histologic Remission Index (PHRI), which could be successfully implemented into an artificial intelligence (AI) model to detect histological remission.

What is already known on this subject?

► Histological activity in UC is associated with poor outcomes and histological remission has been proposed as a treatment target in UC.

► Multiple histological indices have been developed to define disease activity, however they have not been widely adopted in clinical practice due to their complexity.

► Machine learning models are powerful tools that can complement and support pathologists in their histopathological evaluation.

What are the new findings?

► PHRI is a new score based simply on the presence or absence of neutrophils (yes/no) and it provides excellent diagnostic accuracy, the strongest correlation to endoscopic activity among several histological scores, minimal inter-rater variability and excellent prediction of long-term clinical outcome.

► An AI algorithm based on PHRI was able to accurately determine histological remission.

INTRODUCTION

Histological assessment plays a critical role in determining inflammatory activity and monitoring treatment response in UC. Histological remission (HR)
Significance of this study

How might it impact on clinical practice in the foreseeable future?

► PHRI can help standardise histological assessment of UC in a most practical and easy way.
► A machine learning model based on PHRI can further facilitate the histological reading and improve diagnostic performance.

(Also referred to as histological healing) is an emerging treatment target and is an important outcome in UC clinical trials due to its association with favourable outcomes. However, challenges remain on how to incorporate histology into clinical practice mainly due to: (1) the lack of a universal definition of HR to guide pathologists and (2) the lack of a sensitive, easily applicable histological score/index. Ideally, this index would be: (a) informative of and correlated with endoscopic assessment of disease activity, (b) representative of recovery/healing status of damaged mucosa and (c) predictive of disease outcomes.

The histopathological characteristics of UC are those of a chronic active colitis with a relapsing and remitting course, and consist of three fundamental components: (1) active inflammation (‘activity’), which is the neutrophil infiltration in cryptal epithelium as well as in the lamina propria; (2) chronic inflammation, characterised by expansion of mononuclear cell (lymphocyte and plasma cell) infiltrates in the lamina propria, often accompanied by basal plasmacytosis and eosinophilia and (3) cryptal architecture/structure distortion (‘chronicity’), characterised by irregularity and variation of crypts in size, shape, orientation and intercryptal distance, which is the result of mixed repetitive injury and regeneration of crypts.

Over the past decades, >30 histological scores have been developed, although their adoption in clinical practice remains modest. Similarly, different definitions and criteria of HR have been proposed, ranging from ‘elimination of mucosal ulceration/erosion’ to ‘complete histological normalisation’. Almost all investigators now agree that the absence of neutrophilic infiltration (‘neutrophil-free’ mucosa) is the key to a HR definition due to its association with favourable clinical outcomes. Indeed, two independent international expert panels recently recommended to define HR as the absence of neutrophil infiltration (ie, elimination of histological activity).

With the advent of digital pathology, artificial intelligence (AI) algorithms are increasingly employed into histopathological evaluation and diagnosis, as seen in many imaging-focused fields in medicine. For example, it is being widely introduced in oncopathology using convolutional neural network (CNN)-based learning. But, thus far, to the best of our knowledge, no computer-aided diagnosis (CAD) system has been developed to perform histological scoring and assess HR in UC. Part of the reason is that the complexity and mixed subjectivity of the existing histological scores makes it difficult to build and train deep learning algorithm, supervised and unsupervised.

Recently, we conducted a prospective international multicentre study to develop the Paddington International virtual ChromoendoScopy ScOre (PICaSSO) endoscopic score, a new tool for assessing endoscopic activity and remission in patients with UC by using high-definition virtual electronic chromoendoscopy (HD-VCE). The PICaSSO endoscopic score had better correlation than Mayo Endoscopic Score (MES) and UC Endoscopic Index of Severity (UCEIS) with multiple histological scores. The current study is distinct, and a step further, from all our previous published studies on PICaSSO endoscopic score, as it focuses on creating a new UC histological score that can be used quickly and easily by histopathologists in clinical practice, as well as in trials, and can be incorporated into an AI algorithm. Using the PICaSSO project as a platform, in the present study we meticulously analysed the mucosal biopsies taken from the same colonic areas assessed endoscopically, with a focus on identifying the specific histopathological component(s) associated with histological-endoscopic correlation and with the risk of adverse clinical outcomes. Ultimately, we aimed to develop a simplified and novel histological score that could accurately reflect microscopic mucosal inflammation and healing, predict clinical outcome, respond to therapy and be readily implementable into a machine learning algorithm and thus easily adopted into clinical practice and trials. Creating a simplified histological score, PHRI, that is an objective histological instrument was the main aim, as current use of histological scores in clinical practice is limited. The primary aim of PHRI was to create a simple ‘neutrophil-only’ histological evaluation that predicted specified clinical outcomes. An additional purpose was that an ideal histological index should go beyond the limit of endoscopic evaluation.

**PATIENTS AND METHODS**

**Study population, endoscopic evaluation and clinical follow-up**

A total of 307 patients with UC were prospectively enrolled from 11 centres in Europe and North America into the international multicentre PICaSSO study. The protocol for endoscopic and histological evaluation have been described in details in a previous publication. Briefly, each patient underwent white light HD colonoscopy to determine MES and UCEIS, followed by VCE (iSCAN, Pentax, Japan) to determine PICaSSO score, which comprises mucosal and vascular subscores (PICaSSO Mucosal Score (PMS) and PICaSSO Vascular Score (PVS)). In the same areas of rectum and sigmoid assessed and video recorded on endoscopy, at least two targeted mucosal biopsies were taken resulting in a total of 614 biopsies for histopathological analysis. Targeted biopsies were taken from the most inflamed area or showing the most representative features of endoscopic remission (ER) determined by PICaSSO. All patients were then followed up for at least 12 months with regular clinic visits to document the following prespecified adverse clinical outcomes: (1) hospitalisation as a result of UC relapse, (2) colectomy and (3) initiation or changes in medical therapy for UC flare-up including steroids, immunosuppressants and biologics (after excluding adverse effects, immunogenicity or low drug levels).

**Phase I: deep histological analyses and histological-endoscopic-clinical outcome correlations**

The H&E-stained glass slides of colorectal biopsies were scanned at 40× (0.25 μm per pixel) using Aperio Digital Pathology Scanning system (Leica Biosystem, Illinois, USA). The HD digitised slides were centrally hosted and read by a group of 6 GI pathologists (XG, MV, VV, DZ, GdH, ESR) experienced in IBD who were blinded to the endoscopic data. For each biopsy from each segment, the worst features were scored applying five different histological scoring schemes—Geboes Score (GS), Robarts Histological Index (RHI), Nancy Histological Index (NHI), extent, chronicity, activity and plus (ECAP) score and Villanacci Simplified Score (VSS). The average values of each score
and sub-score for both rectum and sigmoid were also separately analysed. The endoscopic-histological correlation was analysed in multiple steps in order to identify the histopathological features/components that specifically corresponded to the different endoscopic features/patterns of disease activity and remission, as well as predicted the risk of specified clinical outcomes at follow-up.

Phase II: development and assessment of PICaSSo Histologic Remission Index

Based on a solid conclusion from our multistep and comprehensive histological-endoscopic-clinical outcome correlation analyses in the phase I study and following a modified Delphi roundtable discussion between the expert pathologists, the presence of neutrophil infiltration was identified as the key element in UC histopathology that determines the disease activity, mucosal healing and clinical outcome. Subsequently, a novel simplified histological scoring scheme, PICaSSo Histologic Remission Index (PHRI), was proposed, as detailed in Table 1. This index takes into account solely the neutrophil infiltration in both epithelium and lamina propria, as illustrated in online supplemental figure 1. Ulcer and erosion were not included because the histological features of ulcers and/or erosions may not always be apparent in the biopsies due to sampling variation. We standardised particularly the criteria of cryptitis (any number of neutrophils infiltrating the epithelium of any number of crypt/gland) and crypt abscess (cryptitis with any number of neutrophils or any amount of neutrophilic exudate overflowing into cryptal lumen AND any degree of crypt epithelial cell injury).

<table>
<thead>
<tr>
<th>Histological finding</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophil infiltration in lamina propria</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Neutrophil infiltration in epithelum</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Surface epithelium*</td>
<td>1</td>
</tr>
<tr>
<td>Cryptal epithelium (cryptitis)</td>
<td>1</td>
</tr>
<tr>
<td>Crypt abscess</td>
<td>1</td>
</tr>
</tbody>
</table>

Total score = sum of all above (maximum 4)*

Criteria for the scoring histological components: (1) 'neutrophil infiltration' (in either lamina propria or epithelium): any number (even only one) of neutrophils is acceptable. (Evaluation under high power view at 40× is required if absence of neutrophil is determined. Neutrophil in the lamina propria must be outside of capillary lumina; (2) 'crypt abscess': cryptitis with any number of neutrophils or any amount of neutrophilic exudate overflowing into cryptal lumen AND any degree of crypt epithelial cell injury.

*If a biopsy has no intact surface epithelium but shows features of erosion/ ulceration (eg, granulation tissue and/or inflammatory exudates), also score 1.
†When there are multiple biopsies from different segments of bowel, the maximum/highest/worst score (PHRI_max) among all biopsy sites will be the preferred ‘global’ score.

Phase III: validation of PHRI

To validate PHRI, the same pathologists assessed 50 digital slides (about half quiescent and half active UC) and scored PHRI and the other five selected indexes. The validation cases were randomly selected and relabelled by a non-pathologist investigator from the same study group. The pathologists were blinded to clinical and endoscopic information and performed the histological scoring independently.

Phase IV: development of AI algorithm

In this exploratory study we included 138 biopsies, randomly selected from the study collection, that were representative of different grades of inflammation from the 614 collected in the whole study. We developed a CNN classifier to detect the neutrophils in whole slide images (WSIs) and classify them into either histological remission or non-remission based on the presence of neutrophils. The detailed design of the CNN is reported in the AI online supplemental appendix. Briefly, a first model identified patches (areas of the WSI) containing neutrophils, while a second model, using a multiple instance learning approach, combined the features of each patch in the slide into a final dichotomous result (presence or absence of active disease) following the PHRI (figure 1).

Table 1: PICaSSo Histologic Remission Index (PHRI)

<table>
<thead>
<tr>
<th>Histological finding</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophil infiltration in lamina propria</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Neutrophil infiltration in epithelium</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Surface epithelium*</td>
<td>1</td>
</tr>
<tr>
<td>Cryptal epithelium (cryptitis)</td>
<td>1</td>
</tr>
<tr>
<td>Crypt abscess</td>
<td>1</td>
</tr>
</tbody>
</table>

Total score = sum of all above (maximum 4)*

Table 1

<table>
<thead>
<tr>
<th>Histological finding</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophil infiltration in lamina propria</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Neutrophil infiltration in epithelium</td>
<td></td>
</tr>
<tr>
<td>Absent (no)</td>
<td>0</td>
</tr>
<tr>
<td>Present (yes)</td>
<td>1</td>
</tr>
<tr>
<td>Surface epithelium*</td>
<td>1</td>
</tr>
<tr>
<td>Cryptal epithelium (cryptitis)</td>
<td>1</td>
</tr>
<tr>
<td>Crypt abscess</td>
<td>1</td>
</tr>
</tbody>
</table>

Total score = sum of all above (maximum 4)*

Criteria for the scoring histological components: (1) 'neutrophil infiltration' (in either lamina propria or epithelium): any number (even only one) of neutrophils is acceptable. (Evaluation under high power view at 40× is required if absence of neutrophil is determined. Neutrophil in the lamina propria must be outside of capillary lumina; (2) 'crypt abscess': cryptitis with any number of neutrophils or any amount of neutrophilic exudate overflowing into cryptal lumen AND any degree of crypt epithelial cell injury.

*If a biopsy has no intact surface epithelium but shows features of erosion/ ulceration (eg, granulation tissue and/or inflammatory exudates), also score 1.
†When there are multiple biopsies from different segments of bowel, the maximum/highest/worst score (PHRI_max) among all biopsy sites will be the preferred ‘global’ score.

Phase III: validation of PHRI

To validate PHRI, the same pathologists assessed 50 digital slides (about half quiescent and half active UC) and scored PHRI and the other five selected indexes. The validation cases were randomly selected and relabelled by a non-pathologist investigator from the same study group. The pathologists were blinded to clinical and endoscopic information and performed the histological scoring independently.

Phase IV: development of AI algorithm

In this exploratory study we included 138 biopsies, randomly selected from the study collection, that were representative of different grades of inflammation from the 614 collected in the whole study. We developed a CNN classifier to detect the neutrophils in whole slide images (WSIs) and classify them into either histological remission or non-remission based on the presence of neutrophils. The detailed design of the CNN is reported in the AI online supplemental appendix. Briefly, a first model identified patches (areas of the WSI) containing neutrophils, while a second model, using a multiple instance learning approach, combined the features of each patch in the slide into a final dichotomous result (presence or absence of active disease) following the PHRI (figure 1).

Statistics

Statistical Software R (R Core Team, https://www.R-project.org/) was used. The strength of the correlation of continuous and categorical variables was measured with Spearman’s (p) correlation coefficient. Coefficients of 0.8–1.0 were considered as ‘very strong’, 0.6–0.79 as ‘strong’, 0.4–0.59 as ‘moderate’ and 0.2–0.39 as ‘weak’. Spearman’s correlations were compared by drawing 100 bootstrap samples for each pair of variables and computing the corresponding quantiles. Wilcoxon and Fish- er’s exact tests were used to determine the differences between continuous and binary distributions, respectively. For area under the receiver operating characteristic curve analysis, we used R-package pROC (https://CRAN.R-project.org/package=pROC). Predictive modelling was performed by R-package CARRoT (https://CRAN.R-project.org/package=CARRoT). Details are reported in the statistical online supplemental appendix.

The Cox proportional hazard model was used to calculate the probability of survival without specified clinical outcomes for different cut-offs of PHRI. The difference between groups of patients was assessed by HR test and survival analysis implemented via R-package survival (https://CRAN.R-project.org/package=survival).

To assess the inter-rater agreement of the histological scorings, we used one-way intraclass correlation (ICC) coefficient by means of R package irr (http://cran.r-project.org/package=irr). In order to test the hypothesis of the ICC being >0.5 against the alternative, we needed a minimum of 40 histology images to reach the power of 0.8 with a type I error of 0.05. According to Landis and Koch benchmarks, ICC of <0.2, 0.2 to 0.4, >0.4 to ≤0.6, >0.6 to 0.8 and >0.8 was considered ‘poor’, ‘fair’, ‘moderate’, ‘good’, ‘substantial’ and ‘almost perfect’, respectively. Results of all statistical tests were considered significant at p<0.05. Statistical power was computed in the PICaSSo...
endoscopic and histological study recently published based on correlation of PICaSSO endoscopic score and histological scores compared with standard MES and also on specified clinical outcome rates and a sample size of 302 was determined.

The diagnostic performance of the AI CAD for the detection of active UC was reported as sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV) and accuracy (ACC).

RESULTS
Six hundred fourteen biopsies from 307 patients with UC were analysed. One hundred sixty-eight (54.7%) patients were in ER as defined by MES 0, while the others had endoscopically active disease at the time of study. None of the patients was on topical therapy or had Montreal E1 disease. Two hundred seventy (88%) patients completed follow-up for 12 months. The detailed demographic data of the study subjects are shown in table 2.

Phase I: neutrophils as the key determinant in histological-endoscopic-clinical correlation
All five histological indices (VSS, RHI, NHI, GS and ECAP) correlated strongly with all of the endoscopic scores in the same regions of bowel (rectum and sigmoid colon) (Spearman’s ρ=0.55–0.78), as illustrated by the heatmaps (figure 2). All histological indices also showed a weak to moderate correlation with the prespecified adverse clinical outcomes at 12 months (ρ=0.34–0.42) (figure 2).

Looking further into the correlations between the various histopathological components (online supplemental table 1) and endoscopic scores (represented by the mucosal and vascular subcores of PICaSSO score), the neutrophil infiltration in the lamina propria and in epithelium, especially that in lamina propria and the combination of both, generally showed the strongest correlation (ρ=0.60–0.76), as compared with other histological features that also showed correlation to some degree (moderate to strong, ρ=0.43–0.64) (p<0.05) (figure 3). Similarly, neutrophil infiltration also showed a stronger correlation, although overall weak/moderate (ρ=0.40–0.45), with clinical outcomes at 12 months compared with other histological features (ρ=0.24–0.37) (p<0.05) (figure 3).

Phase II: PHRI for histological-endoscopic and clinical outcome correlations
PHRI correlated best with endoscopic disease activity
PHRI correlated strongly with the endoscopic scores, and the strength of its correlation was the best among all the histological indices (p<0.05) (figure 2).

Correlation of PHRI with specified clinical outcomes and relapse risk
For the entire cohort, the PHRI showed a similar moderate correlation with the specified adverse clinical outcomes at 12 months (ρ values around 0.4). Additionally, the average PHRI scores were significantly higher in those who had specified adverse clinical outcomes at 12 months than in those with no events (online supplemental table 2, online supplemental figure 4).

Furthermore, we performed a multivariable logistic regression to explore whether other histological features (chronic
inflammation, basal plasmacytosis and eosinophilia) could improve PHRI prediction of specified clinical outcomes. We found that the addition of none of these histological features further improved PHRI prognostic outcome ability (figure 4).

Patients with PHRI >0 compared with those with PHRI=0 had significantly more negative clinical events (outcomes) at 12 months (48.65% (54/111) vs 13.91% (21/151), p<0.00001), as shown in online supplemental figure 2. In addition, analysis of specified clinical outcomes during 12 months follow-up were significantly stratified, as shown in figure 5A and B. The predictive power of PHRI in any form were almost the same.

Subgroup analysis of patients with only endoscopic remission

When we singled out the patients who were in ER as defined by MES 0, the histological-endoscopic-clinical outcome correlations became weak in all aspects. In the phase I of the study, for this particular subpopulation of patients, of whom only a few had residual mild neutrophil infiltration in colorectal biopsies (5.7% with neutrophils in lamina propria and 5.4% in epithelium), the correlations between histological and endoscopic scores (represented by PICaSSO mucosal score and PICaSSO vascular score) (p<0.30) and between histological scores and specified clinical outcome (indicative of relapse in this particular patient population) both became weak or near zero (p=0.0–0.12) (online supplemental figure 2). Nevertheless, neutrophil infiltration was the single histological feature that remained correlated, although weakly (slightly over 0.1) (online supplemental figure 3).

In the phase II of the study, in patients in ER (MES 0), of which only 10.9% had PHRI >0 (presence of neutrophil infiltration) and 89.1% had PHRI of 0 (no neutrophil infiltration), the correlation between PHRI and endoscopic scores also turned to be much weaker (p=0.24–0.36) (online supplemental table 4). However, PHRI still appeared generally superior to most of the other histological indices (p<0.05), as represented by their correlation with PICaSSO score and its mucosal and vascular subscores (online supplemental figure 2). Moreover, the correlation between PHRI scores and prespecified clinical outcomes was also very weak, but still performed better than the other histological scores (p<0.05) (online supplemental figure 2 and online supplemental table 4). Consistent with this, patients with PHRI >0 seemed to have a higher disease relapse rate at 12 months, as compared with those with PHRI 0 (11.76% (2/17) vs 9.3% (12/129)), although the differences did not reach statistical significance (p>0.05) (online supplemental figure 4D). Lastly, the best cut-off value of PHRI for predicting the relapse at 12 months in patients in ER seemed to be 1 (≤1 vs >1), although further analysis with Cox proportional hazards curves failed to satisfactorily stratify the patients’ relapse risk (figure 5C).

Phase III: validity and reliability of PHRI

The inter-rater agreement among pathologists on all of the histological scores was excellent, as reflected by ICCs: RHI 0.77 (95% CI 0.69 to 0.85), NHI 0.85 (95% CI 0.79 to 0.90), GS 0.82 (95% CI 0.75 to 0.88), ECAP 0.87 (95% CI 0.82 to 0.92), VSS 0.77 (95% CI 0.71 to 0.86) and PHRI 0.84 (95% CI 0.78 to 0.90). The differences between the ICCs of each index were not statistically significant. Overall, interobserver agreement for PHRI was almost perfect, although not necessarily significantly superior to the other histological indices. The breakdown of ICC on each of the histological components of different histological indices were also analysed. For any given histological score, we had the best agreement on assessment for the neutrophil-related parameters, as shown in online supplemental table 5.

Phase IV: convolutional neural network classifier able to detect neutrophils

We divided our cohort in two sets, training and testing, with similar patient characteristics to avoid overfitting our system and ensuring its generalisability. Seventy per cent of the biopsies were used to train the model and 30% to test it. To train the proposed
models and optimise the hyperparameters involved, 15% of the training set was used as validation. In the testing set, our CAD to detect neutrophils had SE 0.71, SP 0.95, PPV 0.85, NPV 0.89 and accuracy 0.88, these results were in line with those of the validation cohort (see table 3). Figure 6 shows the class activation maps to highlight the regions of interest at patch-level in which the proposed model focused to predict the samples. The highlighted regions match with the areas containing neutrophils. For the histological remission prediction, the diagnostic performance, expressed as the same characteristics, was 0.78, 0.92, 0.88, 0.85 and 0.86, respectively (see table 3).

DISCUSSION
We developed a novel and simpler HR index for UC, the PHRI, that correlates well with endoscopic disease activity and with clinical outcomes and it can be easily implemented into a CNN model. The development process of this histological index differs from that of existing scores. PHRI was the result of a joint collaboration between pathologists and endoscopists aiming to develop a histological score aligned to the endoscopic score and going beyond endoscopic evaluation.27 Our work has several strengths. First, the histological study was part of a large international multicentre prospective study with the precise focus on endoscopic-histological-clinical correlation. We included a large number of matching biopsies taken immediately after and exactly from the same areas where endoscopic assessment was performed, rather than limiting the comparison to a patient-level. Second, instead of including multiple diagnostic features as in other histological indices, we limited the PHRI to one parameter only, neutrophil infiltration (active inflammation), the single
factor identified by multiple comparative analyses, to be most relevant to both endoscopic features and clinical outcomes. Our independent finding echoes the study by Pai et al. and is consistent with a gathering consensus on the importance of neutrophils in the definition of disease activity and HR.

The most notable advantage of PHRI is its simplicity. PHRI requires only identifying the presence or absence of infiltrating neutrophils within the lamina propria and glandular epithelium, in a straightforward dichotomous way of ‘yes or no’ (present or absent). It also avoids the usual activity grading (eg, mild, moderate and severe) by arbitrary visual scale or estimate of percentage values, which is somewhat subjective. As found by other investigators and shown by our own inter-rater agreement data, the assessment for neutrophils has always been the most reproducible characteristic. Presence of ulceration/erosion, often included in other indexes, was eliminated from PHRI as we considered it a potential source of variability with little contribution to the score’s accuracy. Indeed erosions/ulcers might not be visible on biopsy histology, the distinction between the two is not always possible and, more importantly, patients with erosions/ulcers inevitably have more extensive neutrophilic infiltration anyway. Adopting this simplified ‘neutrophil-only’ approach, we expect that the histological readings would be maximally objective and reproducible. The addition of other histological components that also had some degree of impact on endoscopic features and/or clinical outcomes did not add significant benefit from a practical point of view and would have instead complicated the development of the AI algorithm. We feel that compared with the other currently available histological scores, PHRI is the easiest to apply in daily practice as a universal histological indicator and quantitative measurement (grading tool) of disease activity in UC.

Another advantage is that PHRI makes it easier to perform histological scoring on multiple biopsies from different segments of colon in patients with extensive colitis, to achieve an entire assessment and generate a global (total, maximum, or average) score per colon. This approach would appreciate the globality and increase the overall accuracy of the histological assessment. In our analysis, we found that the PHRI scores of rectum and sigmoid were similar in terms of their correlation with endoscopy and prediction of clinical outcome. In addition, the highest score and the total score of PHRI (PHRI_max and PHRI_total) had the same value of application and significance. Therefore, it is our preference to set the global score as the highest/worst score among all biopsies (PHRI_max) or simply the score of the histologically worst biopsy only, considering that the total number of biopsies being taken and the extent of disease vary in different patients and different clinicians. Finally, the successful
Artificial intelligence

Table 3  Classification results reached during the validation and the test stage with the neutrophil identification model and the activity of UC prediction

<table>
<thead>
<tr>
<th>Neutrophil infiltration</th>
<th>UC activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Validation set</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.8136</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9253</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.8863</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.9076</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.8108</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.8783</td>
</tr>
</tbody>
</table>

Figure 6  Original images (first column), annotation of the pathologist (second column) and class activation maps (CAMS) (third column). Note that in this case, the first row corresponds to the lamina propria while the second row corresponds to the surface of the epithelium.

development of a computer-aided UC histological diagnosis and scoring system based on PHRI, to the best of our knowledge the first in the field of IBD, supports the notion that a simplified score is readily implementable into an AI model. This may complement rapidly advancing development of AI systems for endoscopic scoring of UC, including prediction of histology from endoscopic scores by a number of authors including us. 40–46

Although preliminary, these findings are particularly promising in light of the rapid integration of CAD systems into clinical practice. The potential benefits of this change are extraordinary, but their discussion exceeds the objectives of our study.

Admittedly, our work has a few limitations. First, our patients’ follow-up protocol did not include the endoscopic and histological reassessments at 12 months (not standard of care in all centres), and second it only lasted 12 months, whereas some clinical outcomes might be observed even after 36 months.49

Third, we did not follow-up patients using patient-reported outcomes similar to other studies 1 15 as symptoms do not relate well to histology or endoscopy. Fourth, we have not yet tested sufficiently the global score of PHRI throughout different regions of the entire colon in patients with pan-colitis, although we did include two sites (rectum and sigmoid) in the present study which recruited a diverse cohort of patients. We recruited patients as part of standard of care and this included flexible sigmoidoscopy or colonoscopy. External validation in large cohort in the context of specific clinical trials is necessary and will be conducted as the next step. Lastly, some histological interpretations were challenging and required further discussion (details in histological appendix). An unsolved issue remains the suboptimal performance of all the current histological indices for patients who have reached ER where residual neutrophils are lacking or scarce. In this particular patient population histological indices, including our new PHRI, lose correlation with either endoscopic scores or with relapse rate. The predictive power of PHRI in assessing the relapse risk in these patients was also limited, although a biopsy finding of PHRI >0 in patients who are otherwise in ER would still be of interest. By using a different histological scoring, Narula et al also failed to show the significance of the impact of histological activity on the relapse in this subpopulation of patients. 49 The reasons for this shortfall may be several. First, the small number of cases with histological but not endoscopic activity (only 10% had PHRI >0 in our patients) underpowered the analysis. Second, the heterogeneous distribution of residual inflammation in treated UC might have generated an underestimation of disease activity. Third, the recurrence of UC is not simply arising from the minimal residual inflammation but is the result of the reactivation of dysregulated mucosal immunological mechanisms.

Nonetheless, in our opinion, as compared with any of the other histological indices, PHRI is the simplest one, while it is also most objective and sensitive. Since a pathologist needs only to identify neutrophils, which is a part of routine in reading biopsy slides as clinical histopathological evaluation, one can have the PHRI score immediately without making additional effort and spending extra time. Thus, the PHRI score can also be easily included into the pathology reports, which would be something that infrequently happen at present. Therefore, we believe that PHRI can be applied efficiently in clinical practice.

In conclusion, PHRI is a simple and reproducible histological index that correlates strongly with endoscopic activity and predicts clinical outcomes in UC. It is therefore ideally suited for adoption in clinical practice as well as for consideration in clinical trials and central readouts if further validated to fulfill requirements of US Food and Drugs Administration or European Medicines Agency requirements. We suggest using a PHRI cutoff of 0 to define HR, and a cut-off of 1 to stratify low vs high risk of adverse outcomes. The dichotomous nature of PHRI (ie, presence or absence of neutrophils) allowed the development of machine learning algorithm with high diagnostic accuracy for detection of the disease activity and HR in UC. Further studies are ongoing to validate the deep learning-based computer-aided classifier before it can be adopted in clinical practice.

Author affiliations

1Pathology, University of Washington School of Medicine, Seattle, WA, USA
2Institute of Translational Medicine, University of Birmingham, Birmingham, UK
3Institute for Biological Physics, University of Cologne, Koln, Germany
4Instituto de Investigación e Innovación en Bioingeniería, I3B, Universitat Politècnica de Valencia, Valencia, Spain
5Institute of Pathology, Klinikum Bayreuth GmbH, Bayreuth, Germany
6Institute of Pathology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
7Department of Pathology, KU Leuven University Hospitals Leuven, Leuven, Belgium
8Department of Pathology, ASST Spedali Civili di Brescia, Brescia, Italy
9Department of Cellular Pathology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
10Department of Biomedical Sciences, Humanitas University, Milan, Italy
11Immunology and Immunotherapy, University of Birmingham, Birmingham, UK
12Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway
13Gastroenterology, National Institute of Health Research Birmingham Biomedical Research Unit, Birmingham, UK
14Department of Gastroenterology, Queen Alexandra Hospital, Portsmouth, UK
15Department of Gastroenterology and Endoscopy, Università Vita Salute San Raffaele, Milano, Italy
Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Tommaso Lorenzo Parigi http://orcid.org/0000-0002-3388-0231
Silvio Danese http://orcid.org/0000-0001-8667-8861
Bu'Hussain Hayee http://orcid.org/0000-0003-1670-8815
Raf Bisschops http://orcid.org/0000-0002-9994-8126
Subrata Ghosh http://orcid.org/0000-0002-1713-7979

REFERENCES

Gut: first published as 10.1136/gutjnl-2021-326376 on 16 February 2022. Downloaded from http://gut.bmj.com/ on August 21, 2022 by guest. Protected by copyright.


