AI quantification of inflammatory and architectural features in ulcerative colitis distinguishes active disease from remission

Short title: AI distinguishes active vs remission UC

Authors: Dylan Windell[1]-Dylan.Windell@Perspectum.com

Alastair Magness[1]-Alastair.Magness@Perspectum.com

Rongqi Li[1]-Ritchie.Li@Perspectum.com

Tom Davis[1]-Tom.Davis@perspectum.com

Helena Thomaides Brears[1]-Helena.Thomaides-Brears@perspectum.com

Sarah Larkin[1]-Sarah.Larkin@perspectum.com

Cayden Beyer[1]-Cayden.Beyer@perspectum.com

Paul Aljabar[1]-Paul.Aljabar@perspectum.com

Reema Kainth[1]-Reema.Kainth@perspectum.com

Phil Wakefield[1]-Philip.Wakefield@Perspectum.com

Caitlin Langford[1]-Caitlin.Langford@perspectum.com

Nicholas Powell[2]-nicholas.powell@imperial.ac.uk

Mark DeLegge[3]-mark.delegge@iqvia.com

Adrian C. Bateman[4]-adrian.bateman@uhs.nhs.uk

Roger Feakins[5]-roger.feakins@nhs.net

Eve Fryer[1,6]-eve.fryer@ouh.nhs.uk

Robert Goldin[1,7†]-r.goldin@imperial.ac.uk

Jonathan Landy[8†]-jonathan.landy@nhs.net

† Denotes equal senior co-authorship.

Affiliations:

1. Perspectum Ltd, Oxford, UK
2. Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, United Kingdom
3. IQVIA Inc., Falls Church, VA, USA.
4. University Hospital Southampton NHS Foundation Trust, Southampton, UK
5. Department of Cellular Pathology, Royal Free London NHS Foundation Trust, London, UK
6. Department of Cellular Pathology, John Radcliffe Hospital, OUH NHS FT, Oxford, UK
7. Section for Pathology, Imperial College, London, UK
8. St Mark's Hospital, London North West University Healthcare NHS Trust., UK

Authors' Contributions: Guarantor – CL. Conceptualisation – DW/AM (Equal), JL (Supporting). Writing (original draft)-DW (Lead), AM (Supporting). Data curation – AM (Lead), DW/RL/TD/CB/DW/EF/RG (Supporting). Investigation – DW (Lead), AM/TD/CB (Supporting). Formal analysis – AM/DW (Equal), TD (Supporting). Preparation – DW (Lead), AM/HTB/SL (Supporting), Visualisation – AM (Lead), DW (Supporting). Review & Editing – DW/HTB (Equal), AM/RL/TD/CD/SL/PA/RK/PW/CL/NP/MD/ACB/RF/EF/RG/JL (Supporting).

Funding Statement: The WHITS study (NCT05000242) was sponsored and funded by Perspectum Ltd.

Data availability statement: The PAIR-IBD detection system described for the first time in this study is the intellectual property of Perspectum Ltd and the subject of the Patent Cooperation Treaty: PCT/IB2025/062039. Summary data is included in the manuscript or uploaded as online supplemental information.

Conflicts of interest disclosure: DW/AM/RL/TD/CB/HTB/SL/PA/RK/PW and CL are employees at Perspectum Ltd. EF and RG are consultants for Perspectum Ltd. RF is a central histopathology reader for Alimentiv Clinical Trials; Lecture/expert consults for AbbVie, Eli Lilly, DeciBio, Janssen. All other co-authors have no conflicts of interest to declare relevant to this work. Manuscript includes use of data generated via collaboration with the West Hertfordshire NHS Trust.

Ethics approval statement: All parent studies were conducted in accordance with the ethical principles of the Declaration of Helsinki 2013 and the Good Clinical Practice guidelines and all studies approved by the London-Queen Square Research Ethics Committee (REC: 20/LO/0349).

Patient consent statement: Informed consent and assent (where required) was provided by all subjects and their caregivers (where required), respectively prior to their participation in the study.

Permission to reproduce material from other sources: Anonymised individual participant data can be shared upon request or as required by law and/or regulation with qualified external researchers. Approval of such requests is at the discretion of the study sponsors and is dependent on the nature of the request, the merit of the research proposed, the availability of the data, and the intended use of the data.

Clinical trial registration: The WHITS study was registered as NCT05000242.

ABSTRACT

Background and Aims: Artificial intelligence (AI) is increasingly applied to histological assessment in inflammatory bowel disease (IBD), but most approaches quantify features in isolation and ignore their anatomical location within the mucosa. We developed and validated PAIR-IBD (Perspectum AI Reading in IBD), an AI system that quantifies inflammatory cell populations, crypt injury, and epithelial damage within defined mucosal compartments to distinguish active disease, remission, and equivocal cases in ulcerative colitis (UC).

Methods: A deep learning ensemble was trained on three IBD biopsy datasets to identify lymphocytes, neutrophils, eosinophils, and plasma cells, and to segment crypts, lamina propria (LP), and muscularis mucosae. Inflammatory cell densities and crypt injury metrics (mucin depletion, solidity, roughness, branching, and abscess formation) were quantified. PAIR-IBD outputs were compared between histologically active and remissive UC, evaluated in inconclusive cases, and correlated with manual pathology grading.

Results: Neutrophil density increased 3.5-fold in the LP and 15-fold within crypts in active UC ($p<0.0001$). Eosinophil density doubled and LP lymphocytes increased 1.4-fold. Active UC showed increased mucin depletion, crypt branching, and crypt abscesses, with reduced crypt solidity ($p<0.0001$ for all). PAIR-IBD metrics correlated with manual inflammatory and crypt injury scores (r_s=0.23–0.72) and global indices (r_s=0.27–0.65). Up to 89% of inconclusive cases aligned with remission-like profiles based on multiple independent AI metrics.

Conclusion: PAIR-IBD provides spatially resolved, quantitative assessment of inflammation and epithelial injury in UC, improving disease stratification and resolution of equivocal histology, with potential to support scoring consensus and improve accuracy of histological endpoints in clinical trials.

Keywords

1. Artificial Intelligence
2. Digital Pathology
3. Histology

INTRODUCTION

Inflammatory Bowel Disease (IBD) is a chronic inflammatory disease with effective therapies but no cure to date. IBD affects more than 0.7% of the US population[1] and 0.8% of UK population[2,3] and is a globally expanding disease[4,5]. The two main subtypes, Crohn's Disease (CD) and Ulcerative Colitis (UC), cause debilitating symptoms and carry an increased risk of colorectal cancer.

As recently highlighted by the (European Crohn's and Colitis Organisation) ECCO statement[6], UC features include immune cell infiltrations, changes to crypt injury and epithelial damage in different tissue regions across the colon. The most severe histologic features are epithelial damage, erosion or ulceration[7] while neutrophils are considered the defining feature of active UC, appearing in the lamina propria (LP), crypt epithelium (cryptitis), or crypt lumens (crypt abscesses).

Histologic remission is associated with long-term clinical remission and a reduction in cancer risk[8,9]. Histology-based assessment can detect residual microscopic inflammation that may persist in up to 40% of patients exhibiting endoscopic healing[10]. Acute inflammatory indicators from histopathology are associated with a two- to three-fold increased risk of colitis relapse[11], and basal plasmacytosis predicts UC clinical relapse in patients with complete mucosal healing[12]. Consequently, histological remission is often recognised as a therapeutic goal in UC clinical trials[13-15], in addition to, the more established, endoscopic remission.

The most widely used histological indices used in clinical trials are the Nancy Index (NI), Robarts Histopathological Index (RHI) and the Geboes score (GS). The NI assesses LP chronic inflammation infiltrate, neutrophilic inflammation (cryptitis and crypt abscesses), and the presence of ulcers or erosions[16]. The GS incorporates inflammation features,

architectural distortion and graded epithelial injury, with the Simplified GS (SGS) providing a streamlined variant for routine use[17,18]. By contrast, the RHI prioritises LP, intraepithelial neutrophils and epithelial damage[19,20]. However, even among expert GI pathologists, both the GS and NI scoring systems mostly show only moderate inter-rater reliability[11,18,21] and the good inter-rater reliability of the RHI[22] requires further validation[24]. The fundamental differences between the >30 histological scoring indices in UC contributes to the variability in inter-rater reliability and divergent disease classification[11].

Multiple artificial intelligence (AI) models for histological analysis have been designed for UC[23,24]. AI systems have been designed to quantify various cell populations and crypt features[25-34] but these tools do not investigate both quantitative inflammatory and crypt morphological features in various degrees of disease severity. Accordingly, there is a need for a single quantitative support tool for pathologists that moves beyond coarse ordinal indices to accurately quantify both inflammatory indicators and architectural changes in a spatially resolved manner[35]. In this study we sought to develop and validate PAIR-IBD, a comprehensive AI tool that jointly quantifies regional inflammatory cell populations, crypt injury and epithelial damage in UC, we evaluated whether PAIR-IBD enabled 1) clear differentiation between histologically active and remissive UC across 3 histological indices (NI, RHI, and SGS), and 2) positioned inconclusive scored biopsies along a spectrum of mucosal injury and healing.

MATERIALS AND METHODS

Study design and populations

The West Hertfordshire IBD Technology Study (WHITS) is a prospective, longitudinal, observational, non-interventional study of individuals living with IBD (NCT05000242)[36]. We applied PAIR-IBD to baseline UC histology from WHITS participants recruited between 2021 and 2023 (**Supplementary Figure 1**). Inclusion criteria were digitised datasets above a quality-controlled threshold for biopsy staining and clinical confirmation of UC diagnosis (performed independently of this work) (**Supplementary Methods**).

Ethics and study consent

Informed consent was obtained from all study participants for the WHITS trial, with ethical approval provided by the London Queen Square research ethics committee (ethics reference 20/LO/0349).

Data collection

Bloods, stool and biopsy sample collection

All individuals with UC attended a baseline assessment (blood, stool, medical history, treatment and anthropometrics) and a histological assessment (average 4 biopsies per patient) at the West Hertfordshire Hospitals NHS Trust.

Biopsy digitization and selection

Whole slide images (WSIs) were digitised at 40× on two scanners and analysed on the biopsy section with the largest lamina propria area (Figure 1; Supplementary methods).

Histological scoring

WSIs were scored independently by two pathologists with 14 and 40+ years' experience, respectively, blinded to clinical data. All scoring was performed using a universal template (**Supplementary Table 1**), enabling subsequent conversion to the NI, SGS and RHI scores.

The remission cut-offs were set at NI ≤0, SGS ≤2A.1 and RHI ≤3; discordant reads were classed as inconclusive (equivocal)[37-39].

Development of PAIR-IBD for feature detection in IBD

PAIR-IBD is an ensemble of convolutional neural networks that segments key mucosal compartments and classifies inflammatory cell types on H&E-stained whole-slide images. The system quantifies inflammatory cell densities and crypt-level injury features within anatomically defined regions of the mucosa. Model performance was assessed on 30,000 held-out image-annotation pairs from 26 WSIs from all major colonic sites (**Supplementary Methods**).

Quantification of inflammatory cell, crypt injury and epithelial damage with PAIR-IBD

AI-based quantification of inflammatory cell detection, crypt injury and epithelial damage was performed on the biopsy section with the largest area of LP as determined by PAIR-IBD (**Figure 1**).

Five inflammatory metrics were derived:

1) LP lymphocyte density (cell/mm^2); 2) neutrophil density (cell/mm^2) across the i) LP and ii) within crypts; 3) LP eosinophil density (cell/mm^2); 4) LP plasma cell density (cell/mm^2) and 5) combined plasma cells and lymphocytes, as density (cell/mm^2) in five 100 µm zones from the surface epithelium (SE) towards the crypt bases and muscularis mucosa (MM).

In addition, the following 5 metrics of crypt injury and epithelial damage were derived:

1) mucin depletion, defined as loss of intraluminal mucin relative to total crypt area; 2) crypt abscess fraction, defined as the proportion of crypts containing intraluminal inflammatory cells; 3) crypt solidity and 4) crypt roughness, which captured deviations from normal crypt shape using area- and perimeter-based measures relative to the crypt convex hull; 5) crypt branching, a quantification of abnormal crypt architecture based on enumeration of branches within the crypt skeleton. All metrics were averaged across crypts within each analysed tissue section. Detailed definitions and equations are provided in the Supplementary Methods.

Statistical analyses

Descriptive statistics were used to summarise cohort characteristics. Mean and SD (standard deviation) were used to describe all normally distributed data.

Differences between AI-derived metrics across disease states (remission, active, inconclusive) were assessed using the Wilcoxon rank-sum test. Spearman's rank correlation (r_s) was used to assess associations between PAIR-IBD measurements and pathologist scoring.

Inter-rater reliability was calculated by pairwise overall percentage agreement (OPA) and weighted Cohen's kappa. OPA was applied to the assessment of raw agreement between categorical data and was defined as the percentage of cases in which both pathologists agreed. Fleiss-Cohen weights were used due to comparison enablement with Intraclass Correlation (ICC). SGS and NI were measured with Kappas and RHI with ICC.

The spatial decline in combined lymphocyte and plasma cell density across the mucosa was quantified as a gradient across five anatomical zones extending from the SE to the MM. Cell densities were log-transformed to stabilise variance and analysed as a function of tissue depth. Differences in gradient slope between histological activity states were assessed using linear mixed-effects models with fixed effects for depth, disease status, and their interaction, and a random intercept for slide to account for within-slide clustering. Statistical significance of fixed effects was evaluated using Type III ANOVA.

All analyses were performed in RStudio version 4.5.1 and p values <0.05 were considered statistically significant. Family-wise adjustments were made for multiple comparisons by utilising the Bonferroni methodology.

RESULTS

Study populations

Data from 55 individuals with confirmed UC with digitised WSI of sufficient quality, were recruited across two sites in the West Hertfordshire Hospitals NHS Trust (2021-2023) was included **(Supplementary Figure 1)**. Their mean age was 42 years and 53% were male (**Table 1**). Baseline non-invasive measurements suggested active disease, with elevated mean CRP (11.7 ±14.6 mg/L) and F-calprotectin (1349 ±1558 μg/g), alongside low-normal albumin (37.8 ±4.7 g/L) and haemoglobin (135 ±20 g/L) levels. 87% were on ongoing treatment for UC (58% on a topical, 25% on a biologic/small molecule and 4% only on steroids).

Excluded WSIs were non-colonic/other stains (n=11), poor quality (n=2), QC failures (n=2), or pipeline failures (n=2).

Scoring system agreement

Across the 55 individuals there were 4 WSI per person on average (total n=215 WSIs). Agreement between pathologists over 12 pathologist scores ranged from 66-84%: architectural changes (71%, 0.77κ), serrated architecture (82%, 0.16κ), mucin depletion (66%, 0.38κ), chronic mononuclear infiltrate (80%, 0.83κ), basal plasma cell increase (70%, 0.52κ), eosinophils increase in the LP (79%, 0.48κ), neutrophils increase in the LP (76%, 0.19κ), neutrophils increase in the crypts (84%, 0.73κ), overall crypt inflammation (78%, 0.63κ), crypt injury (79%, 0.30κ) and epithelial injury (82%, 0.64κ). Agreement between pathologists was moderate-strong in the final conversions to NI (71%, 0.49κ), RHI (0.88 ICC), and SGS (53%, 0.62κ). Cases assigned to active, remission and inconclusive for both pathologists for each scoring system were; NI (103, 85, 27), RHI (98, 100, 17) and SGS (93, 104, 18).

Model performance

PAIR-IBD model performance was assessed on 26 held-out, fully annotated WSIs sampling the IBD disease spectrum. We evaluated instance segmentation performance by comparing predicted polygon instances for crypts, lamina propria, muscularis mucosa, and overall tissue against expert ground-truth features. For each slide-region, predicted and ground-truth instances were matched for each feature one-to-one using an intersection-over-union

(IoU) criterion, with matches defined as pairs exceeding a specified IoU threshold (we evaluated thresholds of 0.25, 0.5, and 0.75); unmatched predictions and ground-truth instances were treated as false positives and false negatives, respectively. From the matched pairs we computed per-threshold precision/recall/F1, the mean IoU of matched instances, and Panoptic Quality (PQ), with PQ calculated as the product of segmentation quality (mean matched IoU) and recognition quality derived from the true positive, false positive and false negative counts. We report dataset level accuracy as macro-averages of slide metrics.

At an IoU threshold of 0.5, PAIR-IBD achieved strong instance-segmentation performance for tissue (PQ 0.942, 95% CI 0.866–0.982; F1 0.962, 0.846–1.000; mean matched IoU 0.979, 0.976–0.982), with similarly high F1 for lamina propria (PQ 0.823, 0.749–0.868; F1 0.962, 0.885–1.000; mean matched IoU 0.856, 0.838–0.871) and crypt features (PQ 0.795, 0.713–0.849; F1 0.962, 0.885–1.000; mean matched IoU 0.827, 0.793–0.858). Performance was lower for muscularis mucosa (PQ 0.475, 0.327–0.606; F1 0.654, 0.462–0.846; mean matched IoU 0.727, 0.675–0.776), indicating reduced boundary agreement for this structure, and likely reflecting high variation in feature appearance due to tissue slicing.

Characterisation of active UC with PAIR-IBD

Inflammation in LP

PAIR-IBD identified a total of 10,213,007 cells from 215 WSIs from patients with active UC from NI scoring (**Supplementary Table 2-4**).

Neutrophils Active UC demonstrated a marked increase in LP neutrophil density, compared to remission (19.1 ±21.9 vs 5.4 ±10.4 cells/mm^2, p<0.0001) (**Figure 2A**). Applying a cut-off of 21.7 cells/mm^2, 91% of WSIs classified as remission fell below this threshold in the LP[25].

Eosinophils Eosinophil density was substantially higher in active disease, compared to remission (9.0 ±8.8 vs 4.0 ±4.2 cells/mm^2, p<0.0001).

Lymphocytes Lymphocyte density was elevated in active UC, compared to remission (2,015 ±680 vs 1,470 ±541 cells/mm^2, p<0.0001) (**Figure 3A**).

Plasma Cells Plasma cell density in active UC did not differ from remission (119 ±143 vs 108 ±114 cells/mm^2, NS).

Spatial distribution from SE towards MM (Lymphocytes & Plasma Cells) Density of plasma cells, lymphocytes *or both* cell types combined was measured at each 100 um interval, up to 500 um from the SE towards the MM. Across all zones, the density of both lymphocytes and plasma cells was elevated in active UC 1624 ±470 cells/mm^2 at 0 µm, decreasing to 1553 ±797 cells/mm^2 at 500 µm compared with remission (1311 ±431 to 965 ±743 cells/mm^2; p<0.0001) (**Figure 3B**).

When expressed as change in $\log_{10}$-transformed cell density per micron, the gradient in active disease corresponded to −0.036 $\log_{10}$ per 100 µm, equivalent to an 8% decline per 100 µm, compared with a substantially steeper −0.086 $\log_{10}$ per 100 µm (18% decline) observed in remission. This marked flattening indicated loss of normal mucosal zonation in active UC. Cell density declined significantly with depth across all samples (main effect of zone: F = 68.2, $p<10^{-15}$) with disease activity significantly modifying this gradient (zone x disease interaction: F = 17.2, p<0.0001).

Inflammation in crypts

Active UC demonstrated a substantial elevation in crypt neutrophil density, compared to remission (2.8 ±8.5 vs 0.2 ±0.9 cells/mm^2, p<0.0001) (**Figure 2B**).

Crypt injury and epithelial damage

Mucin Depletion Active UC displayed significantly increased mucin depletion, compared to remission (0.83 ±0.09 vs 0.70 ±0.07, p<0.0001) (**Figure 4A**).

Crypt Abscesses Crypt abscess formation was markedly increased in active disease, compared to remission (0.04 ±0.06 vs 0.00 ±0.01, p<0.0001) (**Figure 4B**).

Crypt Branches Crypt branching was greater in active disease, compared to remission (1.62 ±0.77 vs 1.17 ±0.83, p<0.0001) (**Figure 5A)**.

Crypt Solidity Crypts in active UC exhibited reduced solidity, compared to remission (0.85 ±0.06 vs 0.92 ±0.03, p<0.0001) **(Figure 5B)**.

Crypt Roughness Crypt roughness was increased in active disease, compared to remission (1.16 ±0.07 vs 1.09 ±0.06, p<0.0001).

Characterisation of inconclusive UC with PAIR-IBD

Inflammation in LP

Neutrophils LP neutrophil density in inconclusive cases was between values for remission (5.4 ±10.4) and active UC (19.1 ±21.9), measuring 10.4 ±16.3 cells/mm² (NS vs remission; $p<0.05$ vs active).

Eosinophils Eosinophil density in inconclusive cases (7.4 ±9.7) ranged between active (9.0 ±8.8) and remission (4.0 ±4.2) (NS vs both).

Lymphocytes Lymphocyte density in inconclusive cases was between values for (inconclusive 1,704 ±454 vs remission 1,470 ±541 and active 2,015 ±680; NS vs both).

Plasma Cells Plasma cell density in inconclusive UC (104 ±91) most closely resembled remission (108 ±114) and remained lower than active (119 ±143) (NS vs both).

Spatial distribution (Lymphocytes & Plasma Cells) Across all measurement zones, inconclusive cases showed combined lymphocyte–plasma cell densities ranging 1,335-1,069 cells/mm^2 between 0 and 500 μm, lying between values for remission (1311–965 cells/mm^2) and active UC (1623–1553 cells/mm^2) ($p<0.005$ vs active at 300-500 μm).

The spatial gradient in inconclusive UC measured −0.054 $\log_{10}$ per 100 μm (~10% decline), intermediate between remission (−0.086; ~18% decline) and active UC (−0.036; ~8% decline). Disease activity continued to influence the gradient (overall slopes: $F = 7.51$, $p<0.001$).

Inflammation in Crypts

Neutrophils Inconclusive crypt neutrophils (0.45 ±1.02) were much closer to remission (0.20 ±0.94) than active (2.81 ±8.48) (NS vs both).

Crypt injury and epithelial damage

Mucin Depletion Inconclusive mucin depletion (0.71 ±0.06) resembled remission (0.70 ±0.07) and differed substantially from active (0.83 ±0.09) (NS vs remission; $p<0.0001$ vs active).

Abscess Inconclusive mucin depletion (0.02 ±0.04) resembled remission (0.00 ±0.01) and differed substantially from active (0.04 ±0.06) (NS vs remission; $p<0.0001$ vs active).

Branches Branching was slightly higher in inconclusive UC (1.67 ±0.99) than remission (1.17 ±0.83) and comparable to active (1.62 ±0.77) ($p<0.01$ vs remission; NS vs active).

Solidity Crypt solidity in inconclusive cases (0.89 ±0.05) aligned closer remission (0.92 ±0.03) but remained higher than active (0.85 ±0.06) ($p<0.05$ vs remission; $p<0.01$ vs active).

Roughness Crypt roughness in inconclusive UC (1.14 ±0.09) was similar with active (1.16 ±0.07) and remission (1.09 ±0.06) ($p<0.01$ vs remission; $p<0.05$ vs active).

Comparison of PAIR-IBD with manual pathology reads

Overall, 10 of 11 PAIR-IBD metrics demonstrated moderate correlations with corresponding pathologist scores ($r_s \geq 0.3$), with strong correlations ($r_s > 0.6$) observed for key crypt injury and epithelial damage features including mucin depletion, crypt solidity, and crypt roughness.

*Inflammation in LP (**Supplementary Figure 2**).*

Neutrophils AI-derived LP neutrophil density showed weak correlation with manual LP neutrophil increase scores (r_s=0.30–0.31; $p<0.01$-$p<0.001$).

Eosinophils LP eosinophil density exhibited weak correlation with eosinophil increase scores (r_s=0.30–0.39; $p<0.01$-$p<0.0001$).

Lymphocytes Lymphocyte density demonstrated moderate correlation with chronic mononuclear infiltrate scores (r_s=0.37–0.42; $p<0.0001$).

Plasma Cells No significant correlation was observed between plasma cell increase scores and plasma cell density.

Lymphocytes & Plasma Cells Combined metrics also correlated moderately with chronic mononuclear infiltrate at the MM (r_s=0.39–0.44; $p<0.0001$).

*Inflammation in Crypt (**Supplementary Figure 2**).*

Neutrophils AI crypt neutrophil density correlated moderately with pathologist crypt neutrophil grading (r_s=0.39–0.44; $p<0.0001$).

Histological scoring systems (NI, RHI and SGS) Inflammation-related PAIR-IBD metrics showed low–moderate correlations with NI (r_s=0.25–0.48; $p<0.0001$), RHI (r_s=0.28–0.52; $p<0.0001$) and SGS (r_s=0.26–0.47; $p<0.0001$) (**Figure 6**).

*Crypt injury and epithelial damage (**Supplementary Figure 3-6**)*

Mucin Depletion The mucin depletion metric showed strong correlation with pathologist-assigned mucin loss scores (r_s=0.59–0.75; $p<0.0001$).

Abscess AI-detected crypt abscess frequency correlated moderately with both overall crypt inflammation (r_s=0.50–0.53; p<0.0001) and crypt abscess scores (r_s=0.57–0.59; p<0.0001).

Branches Crypt branching correlated moderately with architectural change (r_s=0.40–0.48; p<0.0001) but weakly with other crypt-injury pathological features.

Solidity Crypt solidity correlated from low to strong across serrated architecture (r_s=0.17–0.24; NS-p<0.05), crypt injury (r_s=0.25–0.37; p<0.05-p<0.0001), surface epithelial injury (r_s=0.34–0.42; p<0.0001) and architectural change (r_s=0.68; p<0.0001).

Roughness Crypt roughness correlated significantly with serrated architecture (r_s=0.21–0.25; NS – p<0.05), crypt injury (r_s=0.22–0.37; NS-p<0.0001), surface epithelial injury (r_s=0.34–0.40; p<0.0001) and architectural change (rs=0.68; p<0.0001).

Histological scoring systems (NI, RHI and SGS) Crypt injury and epithelial damage related PAIR-IBD metrics showed low–strong correlations with NI (r_s=0.28–0.64; p<0.0001), RHI (r_s=0.33–0.6; p<0.0001) and SGS (r_s=0.30–0.64; p<0.0001) (**Figure 6**).

DISCUSSION

In this evaluation of a new AI-system for inflammatory cell populations, crypt injury and epithelial damage, PAIR-IBD, applied to 215 UC biopsy samples, revealed three features of disease. Firstly, active UC is characterised by marked increases in lamina propria and crypt neutrophils, together with elevated eosinophils and increased basal lymphoplasmacytosis. Secondly, epithelial and architectural integrity are reduced in active disease, reflected by greater mucin depletion, increased crypt roughness, and loss of crypt solidity. Thirdly, crypt-level injury, such as branching abnormalities and abscess formation, was prevalent in active UC but largely absent in histological remission. Almost all PAIR-IBD metrics demonstrated moderate-to-strong correlation with manual pathology scoring. This enabled PAIR-IBD to position borderline or inconclusive cases along a biologically continuous gradient of mucosal injury potentially enabling improved manual scoring.

The ability to quantify both inflammatory and architectural features and their spatial relationship in anatomical regions distinguishes PAIR-IBD from previous AI tools for IBD, which typically assess isolated histological metrics or overall activity scores[25-34]. Such a quantitative and spatially resolved framework is well suited to clinical trial settings, where objective, reproducible assessment of histological activity and mucosal healing is required.

Inflammatory cell quantification

Overall, neutrophil density was markedly higher in the lamina propria and crypts in active UC compared with remission. The absence of neutrophils is a recognised marker of histological remission[40], and that our findings align with previous AI systems showing neutrophil densities drive activity scores within the Nancy and Geboes systems[32]. Other studies have similarly reported that persistent increases in neutrophil counts are associated with recurrence when inflammation despite endoscopic remission[27]. A recent analysis proposed a neutrophil density threshold (21.7 cells/mm^2) as a reliable indicator of treatment response[25] which, when applied to our data, classified the majority of remission cases below this threshold.

Our spatial analysis revealed that plasma and lymphocytes in LP are most abundant within the superficial lamina propria (100 µm from SE) and progressively declined toward the MM, consistent with the expected zonation of immune cells in healthy and quiescent mucosa, but that in active UC, this organisation is markedly disrupted. The AI-based spatial profiling

captured spatial and architectural signatures of disease activity that conventional histological scores do not quantify, particularly given the lack of clinical consensus on how to measure mucosal chronic inflammatory cell density[6].

Although the role of eosinophils in IBD activity is uncertain, mucosal eosinophilia has been associated with disease activity and poorer treatment response[41]. Here, eosinophil density was more than twofold higher in active UC compared with remission, although, absolute eosinophil densities were substantially lower than those reported in cohorts with extensive or severe IBD[29].

Crypt injury and epithelial damage quantification

Our crypt metrics showed increases in mucin depletion, crypt roughness, crypt branching and crypt abscess formation with reductions in crypt solidity. Crypt distortion, including branching, irregularity, dilation, and variation in size and shape, is characteristic of active UC and is associated with higher relapse rates even among patients in endoscopic remission[42].

Goblet cell mucin has been shown to outperform manual scoring as a relapse predictor[31]. Our mucin depletion metric showed an increased loss of mucin content in active disease (~20%) consistent with a goblet cell ratio decline in active disease vs reported in another study[43]. This metric, although not typically scored, predicted relapse in patients exhibiting mucin depletion with hazard ratios of 2.18, highlighting the need to accurately capture clinically relevant epithelial dysfunction that manual scoring frequently misses.

Among global indices, the strongest crypt-related PAIR-IBD metrics were observed for mucin depletion, with crypt roughness and solidity also showing strong correlations with both RHI and NI histological indices.

Associations with manual scoring

Neutrophil density in the lamina propria showed the strongest association with global histological indices, particularly the RHI (r_s=0.51-0.52) and NI (r_s=0.48). In biopsies discordant or equivocal manual scores, PAIR-IBD provided additional biological resolution

with a tendency of inconclusive cases to lean towards a remission profile. Features representing overt epithelial injury, such as LP neutrophils, crypt neutrophils, mucin depletion, and crypt abscesses, aligned closely with remission. To a lesser extent, increased eosinophils and crypt roughness also tended to draw the inconclusive group slightly toward the remission end of the spectrum. In contrast, LP lymphocytes, crypt branching and crypt solidity showed intermediate values. This interpretation was reinforced by the zonal analysis where remission biopsies preserved the expected decline in immune cell density from SE to MM, whereas inconclusive cases showed only a partial attenuation of this gradient, indicating mild extension of inflammation into deeper layers. Collectively, these patterns suggest that PAIR-IBD can locate inconclusive WSIs along a continuum of mucosal injury and healing, rather than forcing them into arbitrary, ordinal categories. These biological borderline states illustrate how PAIR-IBD may support pathologists in resolving diagnostically challenging cases or adjudicating between discrepant reads.

Clinical Implications

Histological remission is becoming a key therapeutic objective in UC and is embedded across modern IBD drug development. Recent phase 2b and phase 3 IBD trials[44-50] have consistently incorporated histological assessment alongside endoscopic and clinical endpoints. These programmes predominantly rely on the GS, with increasing adoption of the NI and RHI, and most define histologic remission by the absence of neutrophils, ulceration and erosion (e.g., GS ≤3.1 or GS <2.0). However, all current scoring systems compress the complexity of mucosal inflammation and epithelial injury into coarse ordinal categories and require centralised interpretation by expert GI pathologists, resulting in only moderate inter-reader agreement.

Unlike conventional indices, PAIR-IBD continuous measurements result in interpretable components of inflammation and architecture. A potential application would be to enable gastroenterologists and pathologists to distinguish true, complete histological remission (absent neutrophils, preserved crypt architecture, restored mucin) from “score-level remission” that may still harbour subtle epithelial injury or persistently abnormal cellular profiles. The regional differentiation of PAIR-IBD could support pathologist decision-making to reduce inter-rater variability and distinguish IBD subtypes whilst monitoring disease progression or treatment response within clinical trials. In the context of mucosal healing,

these measurements may enable precise quantification of key histological drivers used in trial endpoints including neutrophils, mucin loss, crypt abscesses, and epithelial injury. In addition, they capture spatial and architectural signatures that lie beyond the resolution of conventional scoring systems but are consistently associated with favourable long-term outcomes across trials, such as reduced relapse, steroid-free remission, and deeper mucosal healing.

Given the increasing emphasis on microscopic remission in IBD, such metrics could enhance the precision and sensitivity of histological endpoints and reduce adjudication requirements. As the field moves toward integrated strategies, combining endoscopy, symptoms, biomarkers, and histology, these tools may ultimately improve patient stratification, accelerate trial recruitment and readouts, and enable earlier detection of residual disease. Collectively, AI-based digital pathology has the potential to shift histology from a subjective, reader-dependent modality to a standardised decision support tool for assessing mucosal healing in IBD.

Challenges and Limitations

Despite significant progress, several challenges remain. The field still lacks consensus on histologic ground truths for training data, and definitions of key features such as mucin depletion, crypt branching, erosion, and basal/lymphoplasmacytosis vary widely[6]. As a result, AI systems inherit an element of pathologist subjectivity, particularly when labels originate from a limited number of expert readers. In addition, the inclusion of well-characterised normal colonic biopsies would be valuable to establish reference ranges for inflammatory cell populations and crypt architecture and to more clearly distinguish true pathological change from mild increases. IBD samples are manually scored across 3 serial sections, making it difficult to discern which tissue regions are used in the final score. Not only may there be multiple biopsies from a particular segment of the bowel, but the biopsies from multiple segments may also be placed on a single strip. In this study, PAIR-IBD analysed the tissue region of largest LP with no indication given to either pathologist of which biopsy section to score, thus requiring further analysis on a whole-slide or “worst-site” approach. Despite being trained on 3 different datasets, this study draws upon a single-centre cohort, and broader multi-institution validation will be required to confirm robustness and reproducibility across diverse technical (e.g. staining and scanners) and demographic settings. While AI outputs correlated strongly with histological scores, their true clinical

value lies in predicting outcomes such as relapse, treatment response, or mucosal healing which will require longitudinal and multimodal datasets integrating endoscopy, molecular readouts and patient level treatment responses. Finally, although our model delineates zonal inflammatory patterns, refinement of zonal plasma cell detection is required to capture basal plasmacytosis accurately.

Our findings demonstrate that AI-based quantification of inflammatory and architectural features provides a robust and objective means of distinguishing pathologist-defined active from remission ulcerative colitis. By delivering quantitative measurements across the lamina propria, crypt epithelium, and deeper mucosal layers, PAIR-IBD overcomes many of the limitations of conventional histological scoring systems, which continue to suffer from moderate interobserver agreement even among expert GI pathologists. Unlike traditional indices, PAIR-IBD enables spatially resolved assessment of mucosal inflammation and structural injury, offering interpretable metrics that capture disease biology not reflected in categorical grade-based systems. As digital pathology matures, the integration of AI into clinical trials and longitudinal disease monitoring has the potential to support pathologist scoring procedures, standardise histological evaluation, and enhance sensitivity to treatment response. Together, these advances signal a shift toward precision histopathology in IBD, where every cellular and architectural feature contributes to a more consistent and data-driven assessment of disease activity.

ACKNOWLEDGEMENTS

We would like to thank the patients recruited into the study used in this article for their participation. We would also like to thank the IQPath facility at UCL Queen Square Institute of Neurology for scanning the biopsy slides.

REFERENCES

1. Lewis JD, Parlett LE, Funk MLJ, et al. Incidence, prevalence, and racial and ethnic distribution of inflammatory bowel disease in the United States. *Gastroenterology*. 2023;165(5):1197-1205. e2.
2. Jones G-R, Lyons M, Plevris N, et al. IBD prevalence in Lothian, Scotland, derived by capture–recapture methodology. *Gut*. 2019;68(11):1953-1960.
3. Hamilton B, Green H, Heerasing N, et al. Incidence and prevalence of inflammatory bowel disease in Devon, UK. *Frontline Gastroenterology*. 2021;12(6):461-470.
4. Zhou J-L, Bao J-C, Liao X-Y, et al. Trends and projections of inflammatory bowel disease at the global, regional and national levels, 1990–2050: a bayesian age-period-cohort modeling study. *BMC Public Health*. 2023;23(1):2507.
5. Hracs L, Windsor JW, Gorospe J, et al. Global evolution of inflammatory bowel disease across epidemiologic stages. *Nature*. 2025:1-9.
6. Feakins R, Borralho Nunes P, Driessen A, et al. Definitions of histological abnormalities in inflammatory bowel disease: an ECCO position paper. *Journal of Crohn's and Colitis*. 2024;18(2):175-191.
7. Pai RK, Lauwers GY, Pai RK. Measuring histologic activity in inflammatory bowel disease: why and how. *Advances in anatomic pathology*. 2022;29(1):37-47.
8. Park S, Abdi T, Gentry M, Laine L. Histological disease activity as a predictor of clinical relapse among patients with ulcerative colitis: systematic review and meta-analysis. *Official journal of the American College of Gastroenterology| ACG*. 2016;111(12):1692-1701.
9. Bryant RV, Burger DC, Delo J, et al. Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. *Gut*. 2016;65(3):408-414.
10. Pai RK, Jairath V, Casteele NV, Rieder F, Parker CE, Lauwers GY. The emerging role of histologic disease activity assessment in ulcerative colitis. *Gastrointestinal endoscopy*. 2018;88(6):887-898.
11. Mosli MH, Feagan BG, Zou G, et al. Reproducibility of histological assessments of disease activity in UC. *Gut*. 2015;64(11):1765-1773.
12. Riley S, Mani V, Goodman M, Dutt S, Herd M. Microscopic activity in ulcerative colitis: what does it mean? *Gut*. 1991;32(2):174-178.
13. Langner C, Magro F, Driessen A, et al. The histopathological approach to inflammatory bowel disease: a practice guide. *Virchows Archiv*. 2014;464:511-527.
14. Lemmens B, Arijs I, Van Assche G, et al. Correlation between the endoscopic and histologic score in assessing the activity of ulcerative colitis. *Inflammatory bowel diseases*. 2013;19(6):1194-1201.
15. Peyrin–Biroulet L, Bressenot A, Kampman W. Histologic remission: the ultimate therapeutic goal in ulcerative colitis? *Clinical Gastroenterology and Hepatology*. 2014;12(6):929-934. e2.

16. Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, et al. Development and validation of the Nancy histological index for UC. *Gut*. 2017;66(1):43-49.
17. Geboes K, Riddell R, Öst A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut*. 2000;47(3):404-409.
18. Jauregui-Amezaga A, Geerits A, Das Y, et al. A simplified Geboes score for ulcerative colitis. *Journal of Crohn's and Colitis*. 2017;11(3):305-313.
19. Magro F, Lopes J, Borralho P, et al. Comparing the continuous Geboes score with the Robarts histopathology index: definitions of histological remission and response and their relation to faecal calprotectin levels. *Journal of Crohn's and Colitis*. 2020;14(2):169-175.
20. Pai RK, Khanna R, D'Haens GR, et al. Definitions of response and remission for the Robarts Histopathology Index. *Gut*. 2019;68(11):2101-2102. doi:10.1136/gutjnl-2018-317547
21. Hutchings MC, Bateman AC. Nancy histological index in ulcerative colitis: an interobserver study. *Journal of Clinical Pathology*. 2025;
22. Arkteg CB, Wergeland Sørbye S, Buhl Riis L, Dalen SM, Florholmen J, Goll R. Real-life evaluation of histologic scores for Ulcerative Colitis in remission. *PLoS One*. 2021;16(3):e0248224.
23. Maeda Y, Kudo S-e, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointestinal endoscopy*. 2019;89(2):408-415.
24. Ahmed M, Stone ML, Stidham RW. Artificial intelligence and IBD: where are we now and where will we be in the future? *Current gastroenterology reports*. 2024;26(5):137-144.
25. Iacucci M, Vadori V, Meseguer P, et al. Detection, localisation, and quantification of neutrophils to assess disease activity and early response to therapy in ulcerative colitis: a novel AI-driven model. *eClinicalMedicine*. 2025;90:103658.
26. Zhang C, Zhang J, Zhang Y, et al. Identifying neutrophil-associated subtypes in ulcerative colitis and confirming neutrophils promote colitis-associated colorectal cancer. *Frontiers in Immunology*. 2023;14:1095098.
27. Ohara J, Maeda Y, Ogata N, et al. Automated neutrophil quantification and histological score estimation in ulcerative colitis. *Clinical Gastroenterology and Hepatology*. 2025;23(5):846-854. e7.
28. Gui X, Bazarova A, Del Amor R, et al. 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. *Gut*. 2022;71(5):889-898.
29. Vande Casteele N, Leighton JA, Pasha SF, et al. Utilizing deep learning to analyze whole slide images of colonic biopsies for associations between eosinophil density and clinicopathologic features in active ulcerative colitis. *Inflammatory bowel diseases*. 2022;28(4):539-546.
30. Furlanello C, Bussola N, Merzi N, et al. The development of artificial intelligence in the histological diagnosis of Inflammatory Bowel Disease (IBD-AI). *Digestive and Liver Disease*. 2025;57(1):184-189.
31. Ohara J, Nemoto T, Maeda Y, Ogata N, Kudo S-e, Yamochi T. Deep learning-based automated quantification of goblet cell mucus using histological images as a predictor of clinical relapse of ulcerative colitis with endoscopic remission. *Journal of Gastroenterology*. 2022;57(12):962-970.

32. Najdawi F, Sucipto K, Mistry P, et al. Artificial intelligence enables quantitative assessment of ulcerative colitis histology. *Modern Pathology*. 2023;36(6):100124.
33. Rymarczyk D, Schultz W, Borowa A, et al. Deep learning models capture histological disease activity in Crohn's disease and ulcerative colitis with high fidelity. *Journal of Crohn's and Colitis*. 2024;18(4):604-614.
34. Das A, Shukla T, Tomita N, et al. Deep Learning for Classification of Inflammatory Bowel Disease Activity in Whole Slide Images of Colonic Histopathology. *The American Journal of Pathology*. 2025;195(4):680-689.
35. Iacucci M, Santacroce G, Zammarchi I, et al. Artificial intelligence and endo-histo-omics: new dimensions of precision endoscopy and histology in inflammatory bowel disease. *The Lancet Gastroenterology & Hepatology*. 2024;
36. West Hertfordshire Inflammatory Bowel Disease Technology Study. National Library of Medicine (US). https://clinicaltrials.gov/study/NCT05000242
37. Pavel C, Diculescu MM, Stepan A-E, et al. Considering histologic remission in ulcerative colitis as a long-term target. *Journal of Clinical Medicine*. 2024;13(1):289.
38. Pai RK, D'Haens G, Kobayashi T, et al. Histologic assessments in ulcerative colitis: the evidence behind a new endpoint in clinical trials. *Expert Review of Gastroenterology & Hepatology*. 2024;18(1-3):73-87.
39. Vespa E, D'Amico F, Sollai M, et al. Histological scores in patients with inflammatory bowel diseases: the state of the art. *Journal of Clinical Medicine*. 2022;11(4):939.
40. Magro F, Doherty G, Peyrin-Biroulet L, et al. ECCO position paper: harmonization of the approach to ulcerative colitis histopathology. *Journal of Crohn's and Colitis*. 2020;14(11):1503-1511.
41. Kim EM, Randall C, Betancourt R, et al. Mucosal eosinophilia is an independent predictor of vedolizumab efficacy in inflammatory bowel diseases. *Inflammatory Bowel Diseases*. 2020;26(8):1232-1238.
42. Gupta A, Yu A, Peyrin-Biroulet L, Ananthakrishnan AN. Treat to target: the role of histologic healing in inflammatory bowel diseases: a systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology*. 2021;19(9):1800-1813. e4.
43. Ozaki R, Kobayashi T, Okabayashi S, et al. Histological risk factors to predict clinical relapse in ulcerative colitis with endoscopically normal mucosa. *Journal of Crohn's and Colitis*. 2018;12(11):1288-1294.
44. Sandborn WJ, Peyrin-Biroulet L, Zhang J, et al. Efficacy and safety of etrasimod in a phase 2 randomized trial of patients with ulcerative colitis. *Gastroenterology*. 2020;158(3):550-561.
45. Sandborn WJ, Vermeire S, Peyrin-Biroulet L, et al. Etrasimod as induction and maintenance therapy for ulcerative colitis (ELEVATE): two randomised, double-blind, placebo-controlled, phase 3 studies. *The Lancet*. 2023;401(10383):1159-1171.
46. Magro F, Peyrin-Biroulet L, Biedermann L, et al. P0636 Impact of obefazimod treatment on histologic and combined histologic and endoscopic outcomes in patients with moderately to severely active ulcerative colitis: results from the Phase 2b open-label maintenance study. *Journal of Crohn's and Colitis*. 2025;19(Supplement_1):i1251-i1252.
47. Sandborn WJ, Feagan BG, D'Haens G, et al. Ozanimod as induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine*. 2021;385(14):1280-1291.
48. Peyrin-Biroulet L, Loftus Jr EV, Colombel J-F, et al. Histologic outcomes with vedolizumab versus adalimumab in ulcerative colitis: results from an efficacy and safety study of vedolizumab intravenous compared to adalimumab subcutaneous in participants with ulcerative colitis (VARSITY). *Gastroenterology*. 2021;161(4):1156-1167. e3.

49. D'haens G, Dubinsky M, Kobayashi T, et al. Mirikizumab as induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine*. 2023;388(26):2444-2455.
50. Danese S, Vermeire S, Zhou W, et al. Upadacitinib as induction and maintenance therapy for moderately to severely active ulcerative colitis: results from three phase 3, multicentre, double-blind, randomised trials. *The lancet*. 2022;399(10341):2113-2128.

Table 1: Study populations

*Shown as n(%). #Shown as Mean± SD. P1 – Pathologist 1, P2 – Pathologist 2.

	All (n=215)
Demographics#	
Age (yr)	42 ±15
Male	29(53%)
Female	26(47%)
BMI	26.4 ±4.6
Non-invasive tests#	
Haemoglobin	135 ±20.2 (n=41)
C-Reactive Protein	11.7 ±14.6 (n=35)
Faecal Calprotectin	1349 ±1558 (n=17)
Albumin	37.8 ±4.74 (n=39)
Treatment*	
Topical (5-ASA)	32(58%)
Biologic/Small Molecule (TNF-α, Ustekinumab, Vedolizumab, Tofacitinib)	14(25%)
Steroid	2(4%)
Increase in Neutrophils in the LP*	
Score 0 – (P1-P2)	159(74%) – 192(89%)
Score 1 – (P1-P2)	51(24%) – 21(10%)
Score 2 – (P1-P2)	4(2%) – 1(<1%)
Score 3 – (P1-P2)	1(<1%) – 1(<1%)
Increase in Neutrophils in Crypt Epithelium*	
Score 0 – (P1-P2)	123(57%) – 114(53%)
Score 1 – (P1-P2)	54(25%) – 68(32%)
Score 2 – (P1-P2)	34(16%) – 20(9%)
Score 3 – (P1-P2)	4(2%) – 13(6%)
Increase in Basal Plasma Cells*	

Score 0 – (P1-P2)	135(63%) – 103(48%)
Score 1 – (P1-P2)	46(21%) – 55(26%)
Score 2 – (P1-P2)	22(10%) – 20(9%)
Score 3 – (P1-P2)	12(6%) – 37(17%)
Increase in Eosinophils in the LP*	
Score 0 – (P1-P2)	151(70%) – 166(77%)
Score 1 – (P1-P2)	61(28%) – 41(19%)
Score 2 – (P1-P2)	3(2%) – 6(3%)
Score 3 – (P1-P2)	0 – 2(1%)
Chronic Mononuclear Infiltrate*	
Score 0 – (P1-P2)	87(41%) – 103(48%)
Score 1 – (P1-P2)	62(29%) – 35(16%)
Score 2 – (P1-P2)	50(23%) – 37(17%)
Score 3 – (P1-P2)	16(7%) – 40(19%)
Architectural Changes*	
Score 0 – (P1-P2)	69(32%) – 102(48%)
Score 1 – (P1-P2)	102(47%) – 50(23%)
Score 2 – (P1-P2)	32(15%) – 26(12%)
Score 3 – (P1-P2)	12(6%) – 37(17%)
Serrated Architecture*	
Score 0 – (P1-P2)	200(93%) – 182(85%)
Score 1 – (P1-P2)	12(6%) – 22(10%)
Score 2 – (P1-P2)	3(1%) – 6(3%)
Score 3 – (P1-P2)	0(0%) – 5(2%)
Crypt Injury*	
Score 0 – (P1-P2)	174(80%) – 183(85%)
Score 1 – (P1-P2)	10(5%) – 10(5%)
Score 2 – (P1-P2)	21(10%) – 8(4%)
Score 3 – (P1-P2)	10(5%) – 14(6%)
Overall Crypt Inflammation*	
Score 0 – (P1-P2)	123(57%) – 117(54%)

Score 1 – (P1-P2)	45(21%) – 55(26%)
Score 2 – (P1-P2)	37(17%) – 29(14%)
Score 3 – (P1-P2)	10(5%) – 14(6%)
Mucin Depletion*	
Score 0 – (P1-P2)	103(48%) – 141(65%)
Score 1 – (P1-P2)	59(27%) – 41(19%)
Score 2 – (P1-P2)	41(19%) – 23(11%)
Score 3 – (P1-P2)	12(6%) – 10(5%)
Crypt Abscess*	
Score 0 – (P1-P2)	168(78%) – 163(76%)
Score 1 – (P1-P2)	27(13%) – 32(15%)
Score 2 – (P1-P2)	18(8%) – 18(8%)
Score 3 – (P1-P2)	2(1%) – 2(1%)
Epithelial Injury*	
Score 0 – (P1-P2)	168(78%) – 187(87%)
Score 1 – (P1-P2)	26(12%) – 13(6%)
Score 2 – (P1-P2)	4(2%) – 2(1%)
Score 3 – (P1-P2)	13(6%) – 10(5%)
Score 4 – (P1-P2)	4(2%) – 3(1%)
Simplified Geboes Score*	
Score 0 – (P1-P2)	98(46%) – 99(46%)
Score 1 – (P1-P2)	4(2%) – 8(4%)
Score 2 – (P1-P2)	14(6%) – 6(3%)
Score 3 – (P1-P2)	52(24%) – 74(34%)
Score 4 – (P1-P2)	47(22%) – 28(13%)
Nancy Index*	
Grade 0 – (P1-P2)	86(40%) – 111(52%)
Grade 1 – (P1-P2)	32(15%) – 0
Grade 2 – (P1-P2)	59(27%) – 69(32%)
Grade 3 – (P1-P2)	21(10%) – 22(10%)
Grade 4 – (P1-P2)	17(8%) – 13(6%)
Robarts Histopathological Index*	

Score 0-3 – (P1-P2)	103(48%) – 110(51%)
Score 6-9 – (P1-P2)	46(21%) – 44(21%)
Score 12-15 – (P1-P2)	36(17%) – 41(19%)
Score 18-30 – (P1-P2)	30(14%) – 20(9%)

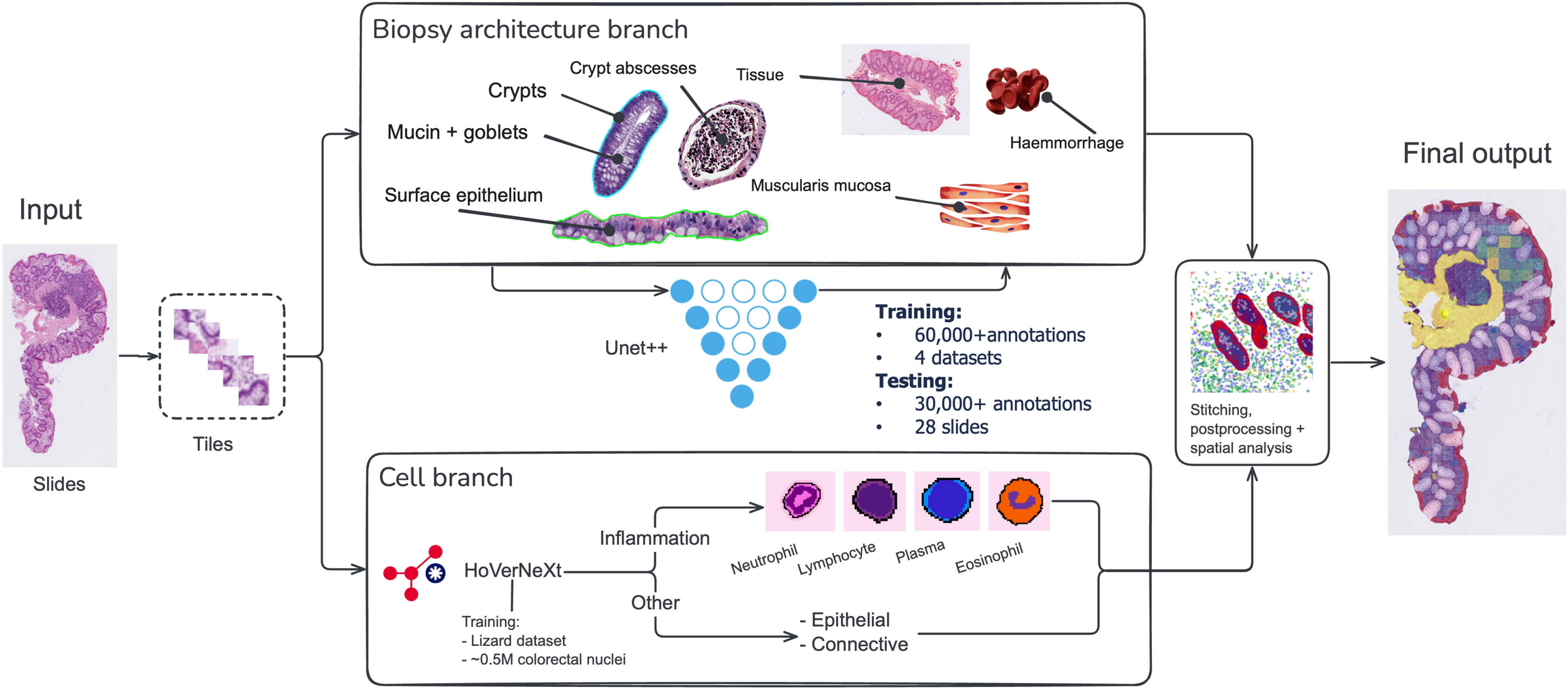

Figure 1 – Integrated tissue segmentation and cell classification enables quantitation of IBD phenotypes and highlighting of features in whole slides for pathologist decision support.

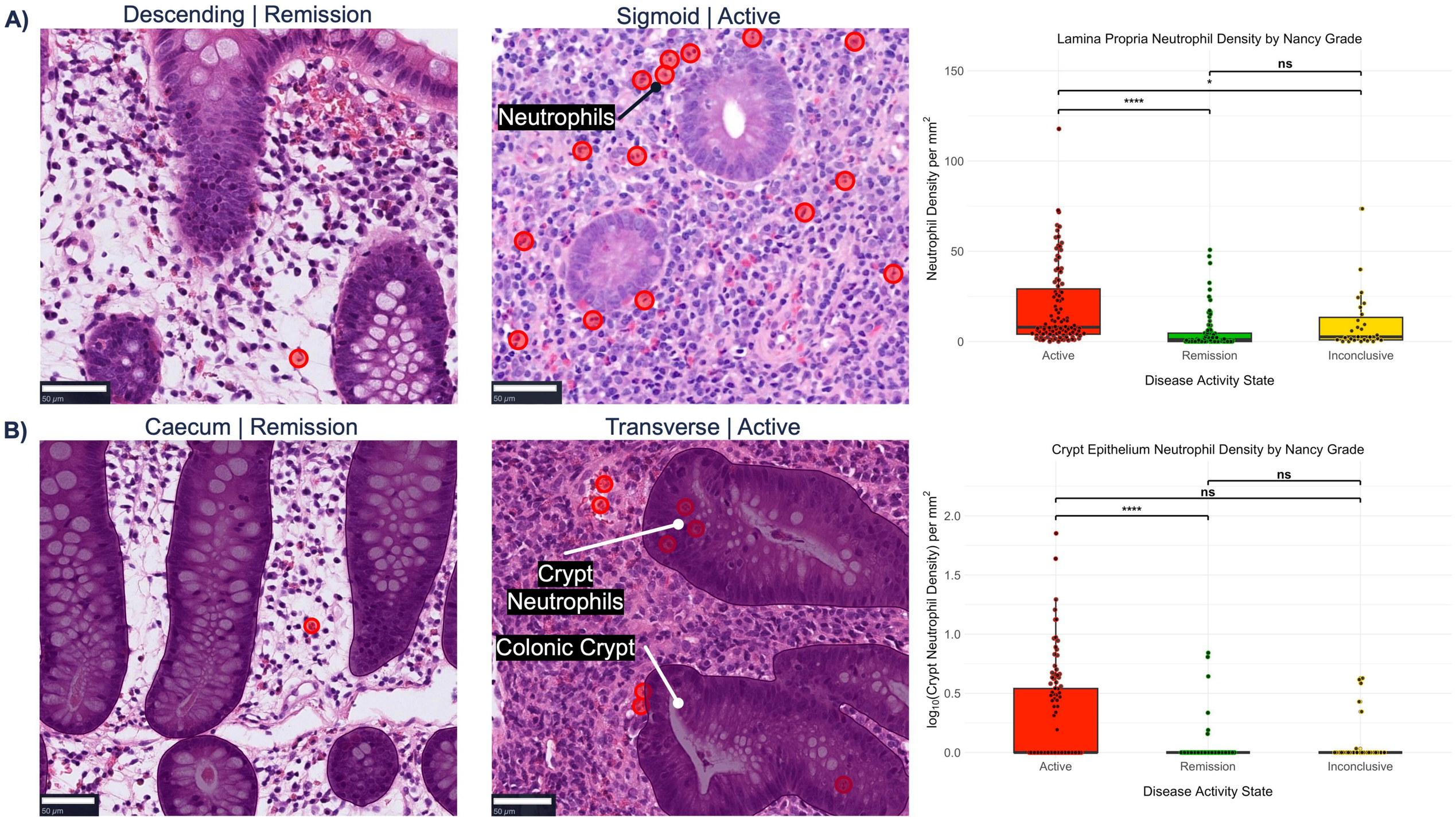

Figure 2 – Representative AI-generated overlays showing neutrophil changes in UC and their comparison with two pathologist scores. (A) Spatial distribution of neutrophils within the lamina propria and (B) neutrophils within the crypts visualised in remission and active UC. Box-and-whisker plots show corresponding quantitative AI measurements for neutrophil and neutrophil crypt metrics within manual histological consensus scoring: Nancy Index remission (≤0), active disease (≥1), and inconclusive (non-agreed) cases. Log-transformation (log10) was applied with a +1 offset to accommodate zero values.

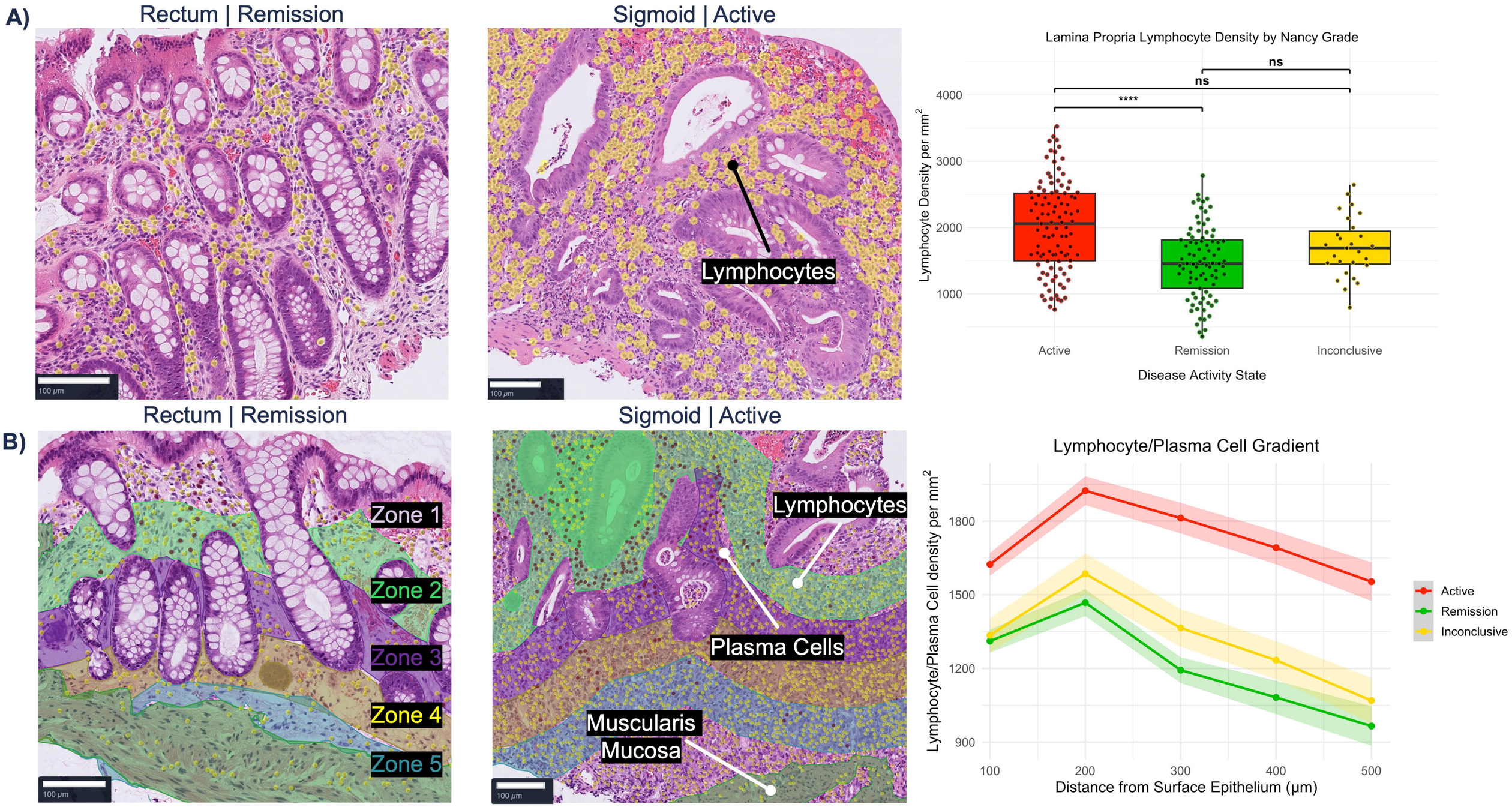

Figure 3 – Representative AI-generated overlays showing plasma cell and lymphocyte changes in UC and their comparison with two pathologist scores. (A) Regional distribution of both plasma cell and lymphocytes in zonal regions from the surface epithelium to the muscularis mucosa. Densities of plasma cells and lymphocytes were measured across the lamina propria within each 100 µm zone. (B) Lymphocyte densities within the lamina propria visualised in remission and active UC. Box-and-whisker plots show corresponding quantitative AI measurements for plasma cells and lymphocytes within manual histological consensus scoring: Nancy Index remission (≤0), active disease (≥1), and inconclusive (non-agreed) cases.

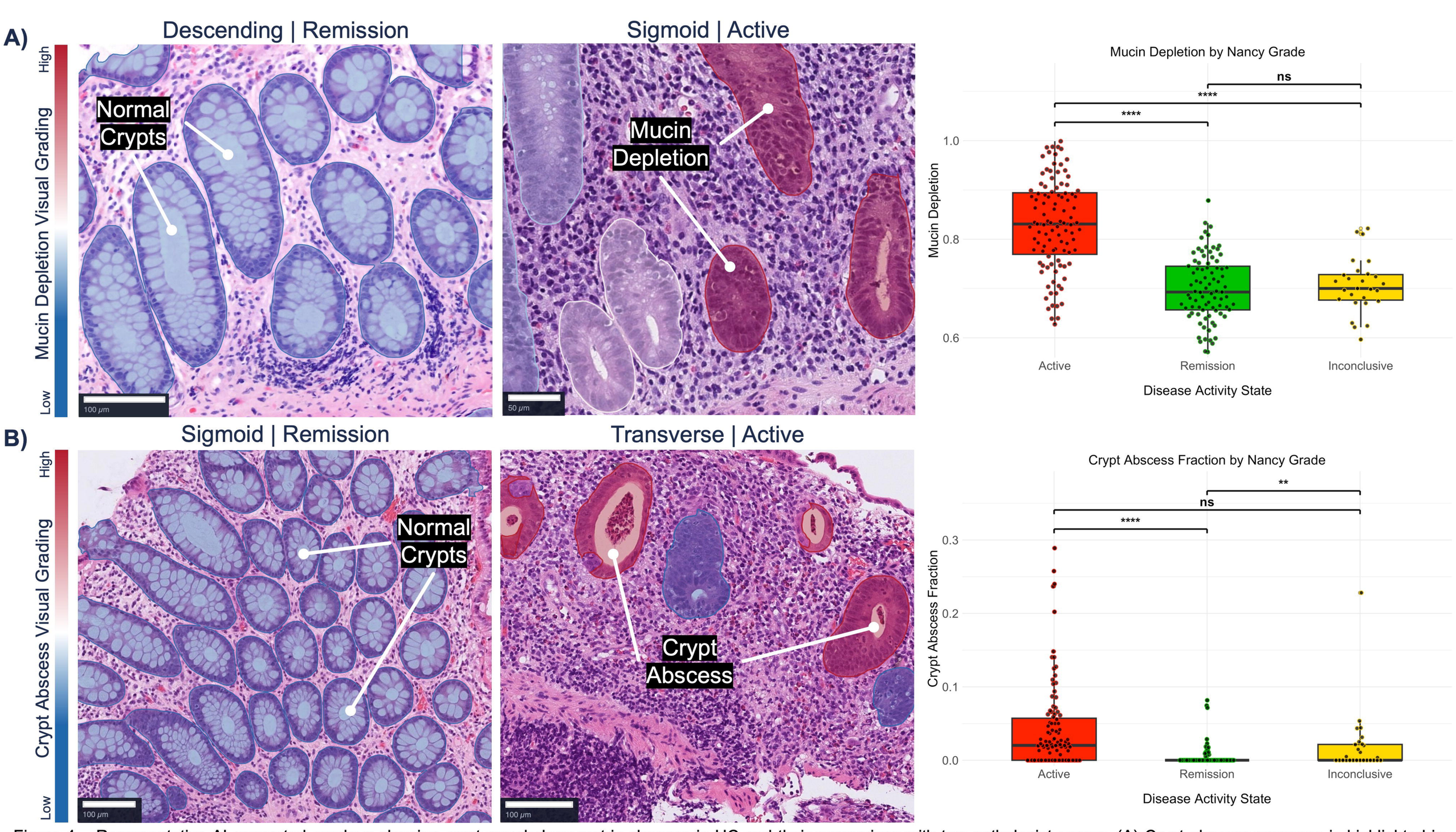

Figure 4 – Representative AI-generated overlays showing crypt morphology metric changes in UC and their comparison with two pathologist scores. (A) Crypt abscess presence is highlighted in active disease cases. (B) Mucin depletion is displayed as a loss of mucin content across crypts and in remission and active UC. Box-and-whisker plots show corresponding quantitative AI measurements for plasma cells and lymphocytes within manual histological consensus scoring: Nancy Index remission (≤0), active disease (≥1), and inconclusive (non-agreed) cases. Visual scale showing increase disease activity in red and lower disease activity in blue.

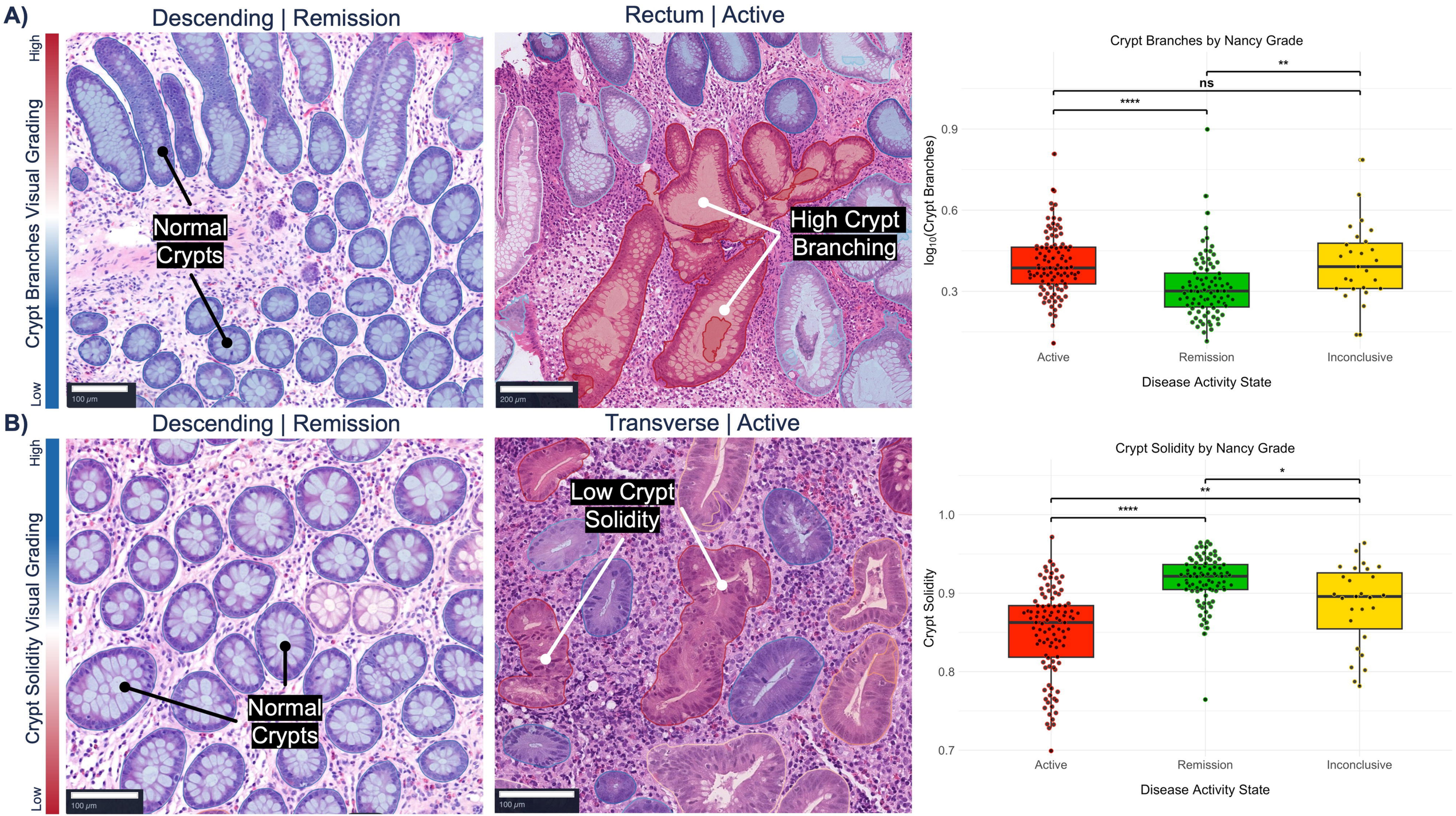

Figure 5 – Representative AI-generated overlays showing crypt morphology metric changes in UC and their comparison with two pathologist scores. (A) Crypt branching is represented across groups of branched crypts in active disease cases. (B) Crypt solidity is displayed as whether crypt shape is deformed or altered in active UC with low crypt solidity highlighting disease activity. Box-and-whisker plots show corresponding quantitative AI measurements for plasma cells and lymphocytes within manual histological consensus scoring: Nancy Index remission (≤0), active disease (≥1), and inconclusive (non-agreed) cases. Visual scale showing increase disease activity in red and lower disease activity in blue. Log-transformation (log10) was applied with a +1 offset to accommodate zero values.

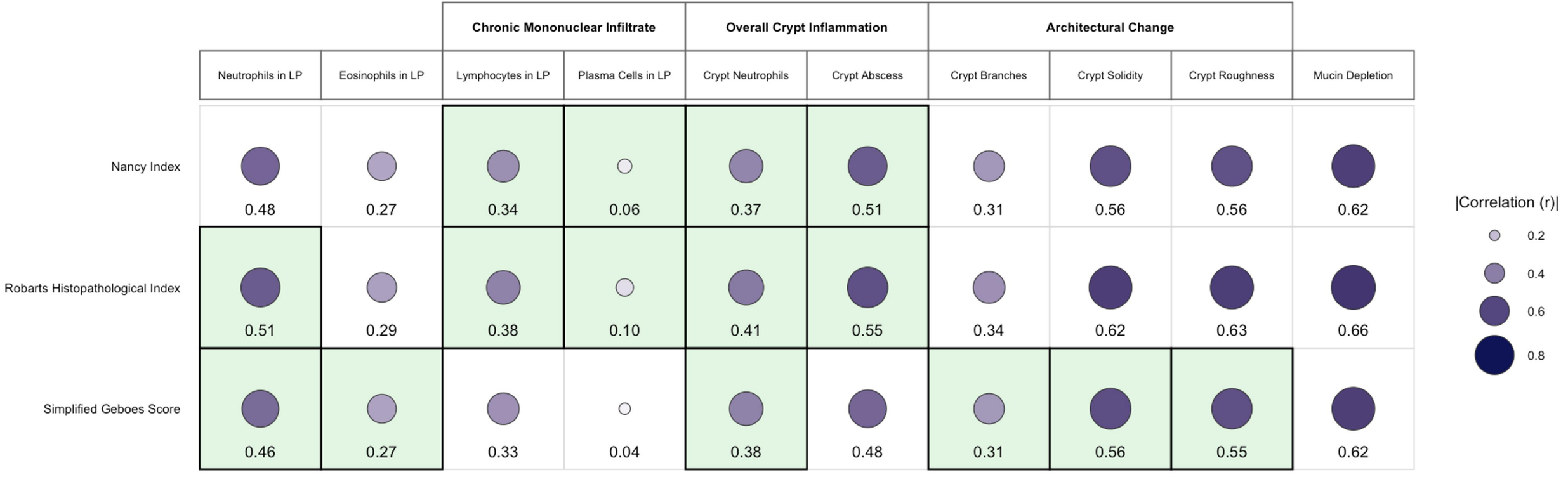

Figure 6 – Correlation map showing individual AI-derived histologic features related to the three major IBD activity scoring systems: Nancy Index, Robarts Histopathological Index, and Simplified Geboes Score. Larger, darker bubbles indicate stronger correlations between each AI feature and the corresponding pathologist rating. Features are grouped into clinically relevant domains, including chronic inflammatory infiltrates, crypt-related inflammation, and architectural distortion. Highlighted cells denote the specific tissue features that are formally assessed within each scoring system itself (e.g., neutrophils and eosinophils in the Geboes Score; chronic mononuclear infiltrate and plasma cells in the Nancy and Robarts indices; crypt neutrophils and abscesses in systems that evaluate crypt inflammation).