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 inconclusive 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 inconclusive histology, with potential to support scoring consensus and improve accuracy of histological endpoints in clinical trials.

Keywords

1. Artificial Intelligence
2. Digital Pathology
3. Ulcerative Colitis

INTRODUCTION

Inflammatory Bowel Disease (IBD) is a chronic inflammatory disease with effective therapies but no cure to date. IBD affects approximately 721 per 100,000 people in the US[1] and 781 per 100,000 people in the UK[2,3] and is increasing globally[4,5]. The main subtypes, Crohn's Disease (CD) and Ulcerative Colitis (UC), cause debilitating symptoms and carry an increased risk of colorectal cancer.

As highlighted by ECCO[6], UC features include immune cell infiltrations, crypt injury and epithelial damage 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 adult with UC 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 adults with UC with complete mucosal healing[12]. Histological remission is therefore an important therapeutic goal in UC clinical trials[13-15] alongside endoscopic remission.

The most widely used histological indices in clinical trials are the Nancy Index (NI), Robarts Histopathological Index (RHI) and the Geboes score (GS). These indices incorporate different combinations of inflammatory, epithelial and architectural disease features[16-20]. Even among expert GI pathologists, inter-rater reliability remains only moderate for GS and NI, while RHI requires further validation[11,18,21-23].

Differing criteria across the >30 UC histological indices contribute to variable inter-rater reliability and divergent disease classification[11].

Multiple artificial intelligence (AI) models for histological analysis have been designed for UC[23,24]. The AI systems quantify various cell populations and crypt features[25-34] but do not jointly assess inflammatory and crypt morphological features across the spectrum of UC severity or account for how these features are distributed through biopsy depth from the surface epithelium to the deeper mucosa. 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, including the use of biopsy material and associated clinical data for digital image analysis and secondary research. Ethical approval was provided by the London Queen Square research ethics committee (ethics reference 20/LO/0349).

Data collection

Endoscopy, Bloods, stool and biopsy sample collection

All individuals with UC attended a baseline assessment at West Hertfordshire Hospitals NHS Trust, including bloods, stool, medical history, treatment and anthropometric data. Endoscopic assessment was performed as part of standard clinical care, and where available, endoscopic activity was scored by a single expert endoscopist using the Mayo Endoscopic Score (MES). Mucosal biopsies were obtained during endoscopy for histological assessment, with an average of 4 biopsies per subject.

Biopsy digitization and selection

Whole slide images (WSIs) were digitised at 40× on two scanners and analysed on the biopsy section with the largest LP 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.

Remission cut-offs were NI≤0, SGS≤2A.1 and RHI≤3; discordant reads were classed as inconclusive[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 defined mucosal regions. Model performance was assessed on 30,000 held-out image-annotation pairs from 26 WSIs, with slide-level separation between training and testing datasets (**Supplementary Methods**).

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

AI-based quantification was performed on the biopsy section with the largest LP area (**Figure 1**). Histological analyses were performed at the biopsy (WSI) level rather than at the subject level. This was a retrospective analysis of prospectively acquired histological samples.

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

Cohort characteristics were summarised descriptively. Differences between AI-derived metrics across disease states were assessed using Wilcoxon rank-sum test, and associations with manual pathology scoring were assessed using Spearman's rank correlation (r_s). Inter-rater reliability was summarised using overall percentage agreement (OPA) and weighted Cohen's kappa. P values <0.05 were considered statistically significant, with Bonferroni correction for multiple comparisons.

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)**. Mean age was 42 years and 53% were male (**Table 1**). Baseline biomarkers suggested an overall active cohort, with elevated mean CRP and F-calprotectin, and low-normal albumin and haemoglobin. Most subjects were receiving treatment, predominantly topical therapy or biologic/small-molecule therapy. Eleven WSIs were excluded from the original 226 as non-colonic or other stains, two for poor quality, two by quality-control failure and two through pipeline failure.

Endoscopic scores were collected for 151 WSIs (70%) which is 46 subjects (86%). Overall, 62% of slides and 45% of subjects were in endoscopic remission, as per the MES 0-1.

Scoring system agreement

Agreement between pathologists across the universal template variables ranged from 66% to 84%. Final converted indices showed moderate to strong agreement, with NI agreement of 71% (κ=0.49), SGS agreement of 53% (κ=0.62), and RHI ICC of 0.88. At the subject level (n=55), using the worst-biopsy approach, active/remission/inconclusive distributions were 67%/29%/4% for NI, 38%/49%/13% for RHI and 60%/29%/11% for SGS.

MES also showed strong alignment with manual histology, with agreement across individual readers, consensus reads and all three histological indices ranging from 83% to 93%.

AI model performance

PAIR-IBD model performance was assessed on 26 held-out, annotated WSIs sampling the IBD disease spectrum. Tile-based segmentation assessment across six tissue classes showed near-perfect area overlap for tissue foreground and strong performance for lamina propria and crypts, with lower performance for surface epithelium and muscularis mucosa, the latter most likely reflecting orientation and sectioning variability. Instance-level evaluation of crypts demonstrated balanced detection and high segmentation accuracy (**Supplementary Table 2-3**).

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 crypt injury and epithelial damage features including mucin depletion, crypt solidity, and crypt roughness. The strongest associations were seen for mucin depletion (rs=0.59–0.75), crypt abscesses (rs=0.50–0.59), and crypt roughness/solidity with architectural change (up to rs=0.68) (**Supplementary Figure 2-6**). Correlations with overall NI, RHI and SGS ranged from 0.25 to 0.64, supporting alignment with established histological frameworks while retaining additional quantitative granularity (**Figure 2**). LP neutrophils showed the strongest association with global inflammatory indices, particularly RHI and NI.

Characterisation of active UC with PAIR-IBD

Active UC showed increased inflammatory cell burden in both the LP and crypts. Compared with remission, LP neutrophil density was higher in active disease (19.1 ±21.9 vs 5.4 ±10.4 cells/mm²; p<0.0001; **Figure 3A**), with 91% of remission biopsies falling below the published 21.7 cells/mm² threshold. LP eosinophils (9.0 ±8.8 vs 4.0 ±4.2 cells/mm²; p<0.0001) and lymphocytes (2015 ±680 vs 1470 ±541 cells/mm²; p<0.0001; **Figure 4A**) were also increased, while LP plasma cell density did not differ significantly. Crypt neutrophil density was markedly higher in active UC (2.8 ±8.5 vs 0.2 ±0.9 cells/mm²; p<0.0001; **Figure 3B**). Active disease also showed greater mucin depletion (0.83 ±0.09 vs 0.70 ±0.07; **Figure 5A**), crypt abscesses (0.04 ±0.06 vs 0.00 ±0.01; **Figure 5B**), crypt branching (1.62 ±0.77 vs 1.17 ±0.83; **Figure 6A**), and crypt roughness (1.16 ±0.07 vs 1.09 ±0.06), together with reduced crypt solidity (0.85 ±0.06 vs 0.92 ±0.03; **Figure 6B**; all p<0.0001). Index-specific

metrics are shown in **Supplementary Table 4-6**. Among NI-derived metrics, the strongest discrimination of active versus remission disease was observed for mucin depletion (AUC 0.87), crypt roughness (AUC 0.86) and LP neutrophils (AUC 0.80), with good discrimination from crypt abscesses (AUC 0.77), lymphocyte density (AUC 0.73) and crypt branching (AUC 0.73). In contrast, the inverted AUC of 0.86 from reduced crypt solidity discriminated active from remission disease well.

Spatial distribution of inflammation in active UC

Across five 100 µm zones from the SE towards the MM, combined lymphocytes/plasma cell density was elevated in active UC than remission (1624 ±470 cells/mm^2 to 1553 ±797 cells/mm^2 vs 1311 ±431 to 965 ±743 cells/mm^2; $p<0.0001$) (**Figure 4B**). This pattern was consistence across NI grades (**Supplementary Figure 7A/B**). Active disease showed a flatter mucosal depth gradient than remission (−0.036 vs -0.086 $\log_{10}$ per 100 µm; 8% vs 18% decline respectively), indicating loss of normal mucosal zonation (interaction $p<0.0001$).

Mucosal spatial metrics demonstrated good discrimination between active and remission disease (**Supplementary Figure 7C**). Mean density from SE to MM yielded an AUC of 0.75 (95% CI 0.68–0.82), while the depth-profile AUC performed similarly (AUC 0.76, 95% CI 0.69–0.83). By contrast, the mucosal depth gradient showed more modest discrimination (AUC 0.63, 95% CI 0.55–0.71).

Characterisation of Inconclusive UC with PAIR-IBD

Inconclusive biopsies showed intermediate inflammatory profiles. LP neutrophils and lymphocytes lay between remission and active groups, whereas crypt neutrophils, mucin depletion and crypt solidity were closer to remission. Crypt branching and crypt roughness were closer to active disease. Pairwise ROC analyses supported this mixed pattern: inconclusive biopsies were better separated from remission by crypt roughness (AUC 0.71), crypt branching (AUC 0.71), crypt abscesses (AUC 0.66) and LP neutrophils (AUC 0.63), whereas separation from active disease was strongest in the remission-like direction for mucin depletion (AUC 0.14), LP neutrophils (AUC 0.33) and crypt neutrophils (AUC 0.39).

DISCUSSION

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 active disease. Firstly, active UC is characterised by marked increases in LP 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. Together, these quantitative outputs highlight disease features that may help biologically position borderline or inconclusive cases and potentially improve manual scoring, with mucin depletion providing the strongest support for remission-like positioning and crypt branching the clearest support for active-like positioning.

Quantifying inflammatory and architectural features across mucosal compartments distinguishes PAIR-IBD from previous AI tools, which typically assess isolated 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

Neutrophil density was markedly higher in the LP and crypts in active UC compared with remission, consistent with the established role of neutrophils as recognised markers of histological remission[40]. Our findings align with previous AI systems showing that neutrophil densities drive activity scores within the Nancy and Geboes systems[32]. 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. Among biopsies classified as remission by both pathologists (RHI ≤3), 7% showed residual neutrophil activity detected by PAIR-IBD. These findings suggest that subtle inflammatory features may be overlooked in routine scoring and support the potential role of AI-assisted quantification as an adjunct to manual assessment.

Mucosal spatial analysis revealed that plasma and lymphocytes in LP are most abundant within the superficial LP (100 μm from SE) and progressively decline toward the MM, consistent with the expected zonation of immune cells in healthy and quiescent mucosa. We found 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[6].

Eosinophil density was more than twofold higher in active UC. Although eosinophils are less central to UC activity than neutrophils[41], these findings were lower than those reported in other cohorts[29].

Crypt injury and epithelial damage quantification

In active UC, PAIR-IBD quantified 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 associated with higher relapse rates even among subjects 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 as reported in another study[43]. 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 LP showed the strongest association with global histological indices, particularly the RHI (r_s=0.51-0.52) and NI (r_s=0.48). Features representing overt epithelial injury, such as LP neutrophils, crypt neutrophils, mucin depletion, and crypt abscesses, aligned closely with remission yet increased eosinophils and crypt roughness showed the opposite.

In biopsies with discordant manual reads, PAIR-IBD provided additional biological resolution. Inconclusive biopsies tended to align with remission for overt injury

features such as crypt neutrophils, abscesses and mucin depletion while remaining intermediate for other measures. This suggests that AI-derived continuous metrics may function as an adjudication aid in trials and potentially enabling improved manual scoring by highlighting the biologically continuous gradient of mucosal injury rather than forcing them into binary categories.

Clinical implications

Histological remission is increasingly embedded in UC drug development[44-50] requiring expert readers, central adjudication and ordinal indices for endpoint assessment. Continuous, spatially resolved measurements may improve endpoint sensitivity by detecting incremental changes in inflammatory burden and epithelial repair are absent from conventional methods. Such measurements, as shown here, could restore equivocal scoring and may also reduce adjudication burden and improve reproducibility by complementing rather than replacing current scoring systems.

The current results also suggest that PAIR-IBD may help distinguish “true” histological remission from score-level remission. Conventional remission thresholds often depend on absence of overt neutrophils, erosion or ulceration, but they do not quantify residual architectural abnormality or low-grade spatial disruption. In our study, biopsies classified as remission still harboured subtle epithelial injury or persistently abnormal cellular profiles in some cases. This may be particularly relevant where subtle residual abnormalities could mark incomplete healing despite formal remission.

Integration with endoscopy, molecular readouts and biochemical biomarkers may further support multimodal disease stratification and treatment-response modelling. It may also provide a practical bridge between pathology review, central trial adjudication and biomarker integration, particularly when borderline biopsies complicate binary remission decisions.

Challenges and limitations

Despite significant strengths, several challenges remain. The evaluation cohort comprised 55 subjects from a single clinical network, and broader external validation across laboratories, staining protocols, scanners and disease activity spectra is required before clinical implementation. Although stain-based colour augmentation was applied during training and file formats were harmonised, explicit stain normalisation was not used. Training annotations were produced by a single expert annotator and multi-annotator validation was not undertaken. Analyses were performed on the biopsy section with the largest LP area rather than a whole-slide or worst-site approach, and biopsies were evaluated at slide level rather than subject level. Vascular lumina were not explicitly segmented, although the likely impact of intravascular cells on LP density estimates is limited in mucosal pinch biopsies where large vessels are uncommon. Nevertheless, segmentation performance remained strong across the principal tissue classes, and most PAIR-IBD metrics continued to show moderate-to-strong correlation with manual pathology scoring. Finally, while ROC-based discrimination was assessed, formal classifier development with predefined thresholds and cross-validated sensitivity and specificity was beyond the scope of this study.

Our findings demonstrate that AI-based quantification of inflammatory and architectural features provides a robust and objective means of distinguishing pathologist-defined active UC from remission. By delivering quantitative measurements across the LP, crypt epithelium and deeper mucosal layers, PAIR-IBD complements conventional histological scoring with mucosal spatial metrics that remain biologically interpretable. As digital pathology matures, such tools may support histological standardisation, adjudication and treatment-response assessment in UC clinical trials and longitudinal disease monitoring.

Acknowledgments

We would like to thank the patients recruited into the study for their participation. We also like to thank the research nurses who collated the data at West Hertfordshire, particularly Alice Balaican. Finally, we acknowledge the IQPath facility at the 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. 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.
24. 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.
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.
51. Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*. 2005;85(3):257-268. doi:10.1093/ptj/85.3.257

Table 1: Study populations. Low κ values for certain features reflect prevalence effects due to the predominance of score 0 rather than true inter-reader disagreement, a recognised limitation of κ in highly imbalanced datasets[51].

*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%)
Endoscopy (Mayo Endoscopic Score)*	
Slides with endoscopy	151 (70%)
Subjects with endoscopy	47 (86%)
MES 0 (Slides/Subjects)	71 (47%) / 15 (32%)
MES 1 (Slides/Subjects)	23 (15%) / 6 (13%)
MES 2 (Slides/Subjects)	42 (28%) / 16 (34%)
MES 3 (Slides/Subjects)	15 (10%) / 10 (21%)
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%)
Subject level Manual Histological Assessment - Active/Remission/Inconclusive*	
Nancy Index	37 (67%) / 16 (29%) / 2 (4%)
Robarts Histopathology Index	21 (38%) / 27 (49%) / 7 (13%)
Geboes Score	33 (60.0%) / 16 (29.1%) / 6 (10.9%)

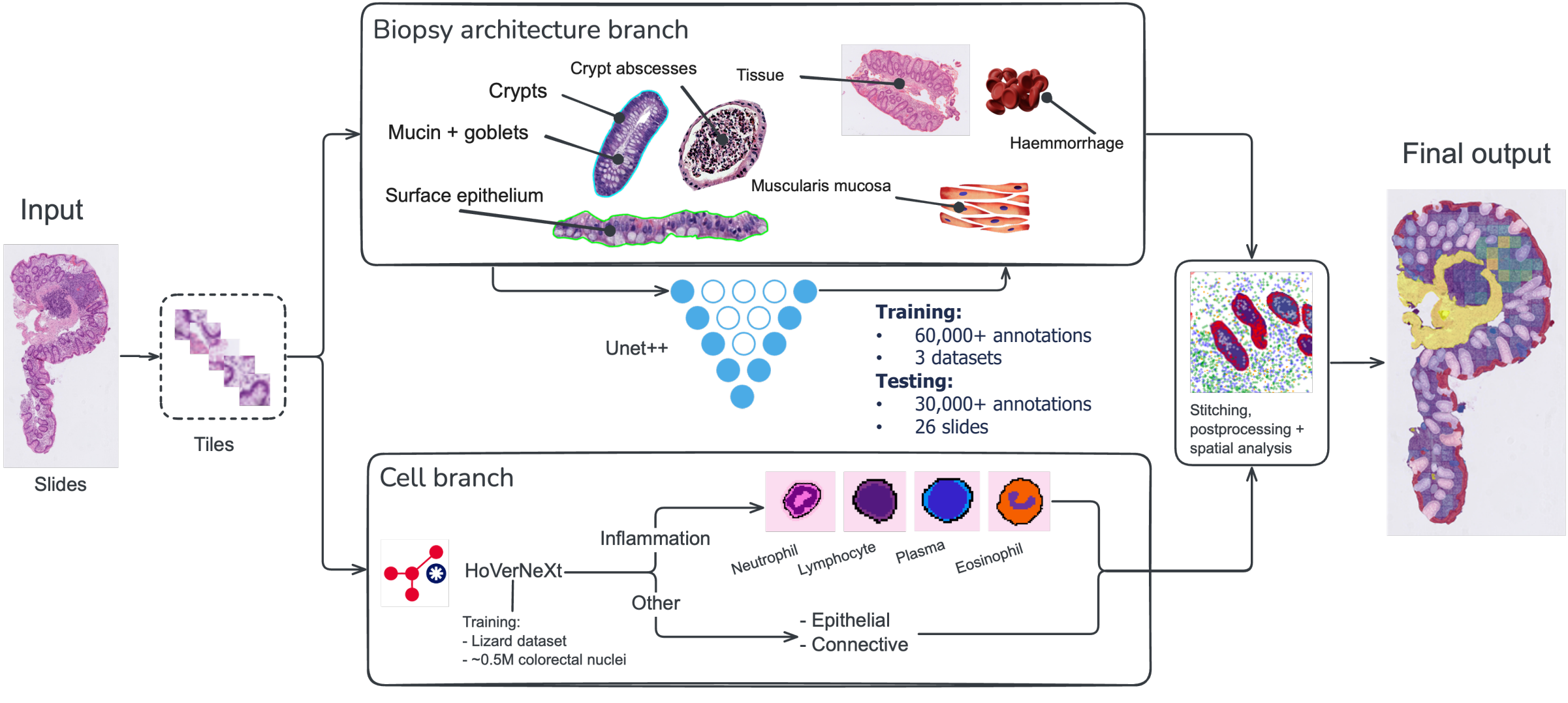

Figure 1 – Integrated tissue segmentation and cell classification enables quantitation of IBD phenotypes and highlighting of features in whole slides for pathologist decision support.

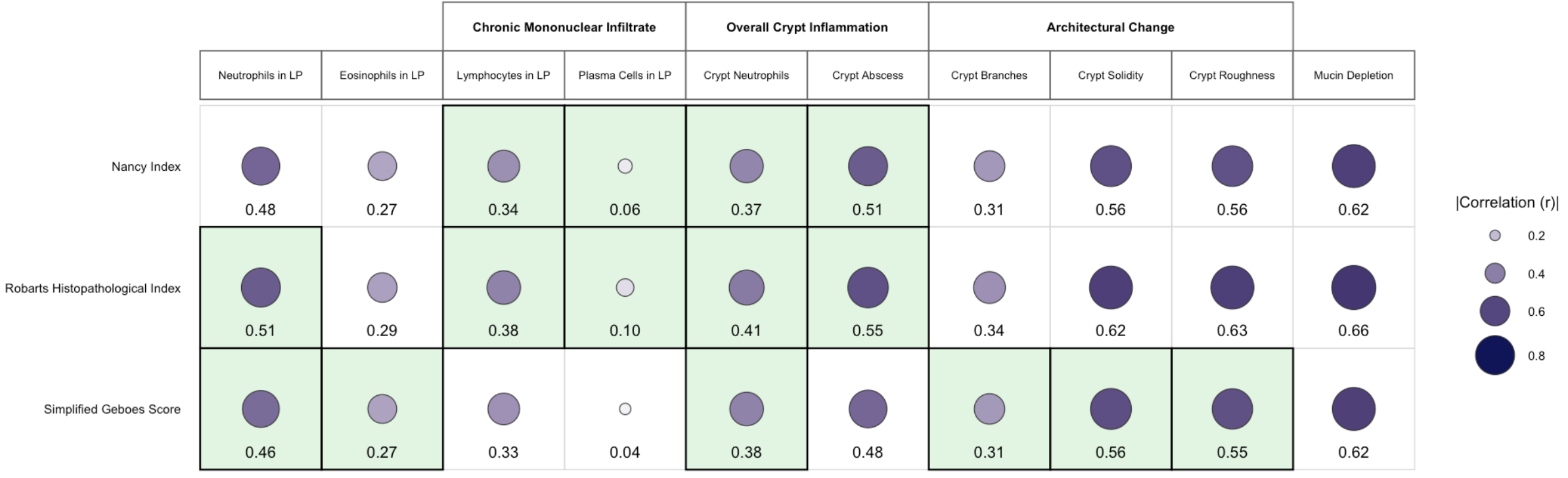

Figure 2 – Correlation map showing individual AI-derived histologic features related to the three major IBD activity scoring systems: Nancy Index, Robarts Histopathological Index, and Simplified Geboes Score. Larger, darker bubbles indicate stronger correlations between each AI feature and the corresponding 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).

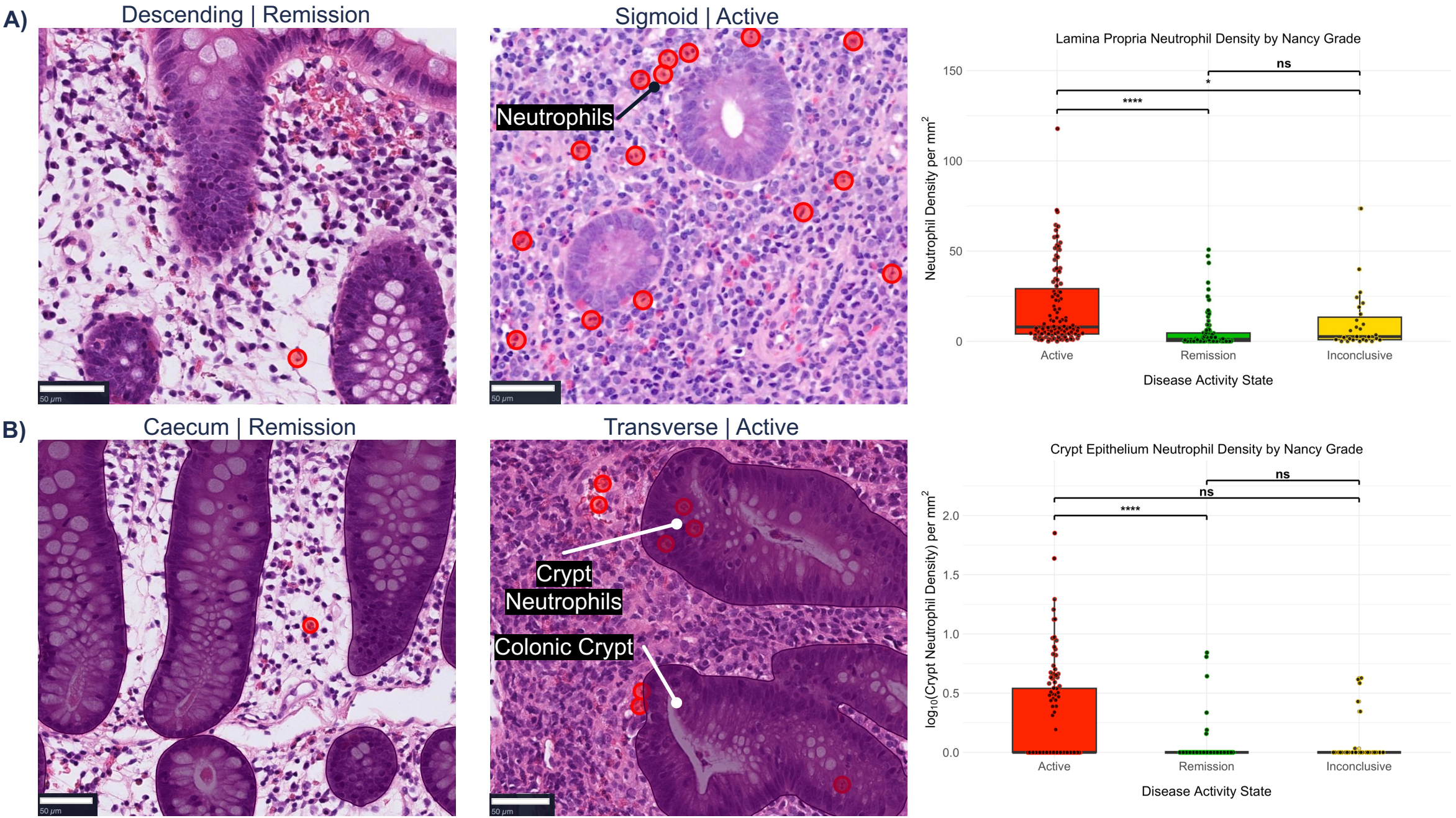

Figure 3 – Representative AI-generated overlays showing neutrophil changes in UC and their comparison with two pathologist scores. (A) Spatial distribution of neutrophils within the lamina propria and (B) neutrophils within the crypts visualised in remission and active UC. Box-and-whisker plots show 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.

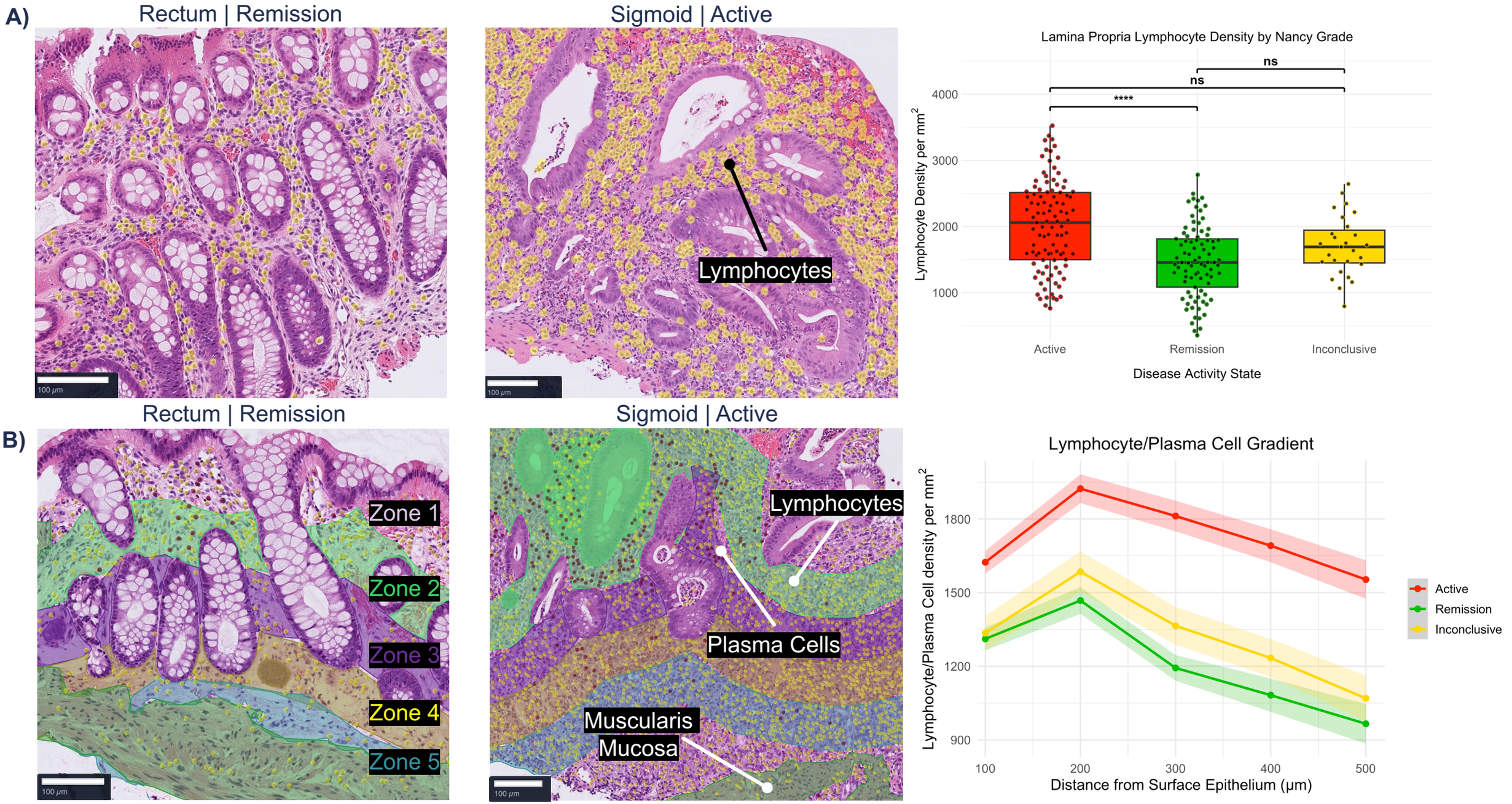

Figure 4 – Representative AI-generated overlays showing plasma cell and lymphocyte changes in UC and their comparison with two pathologist scores. (A) Regional distribution of both plasma cell and lymphocytes in zonal regions from the surface epithelium to the muscularis mucosa. Densities of plasma cells 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.

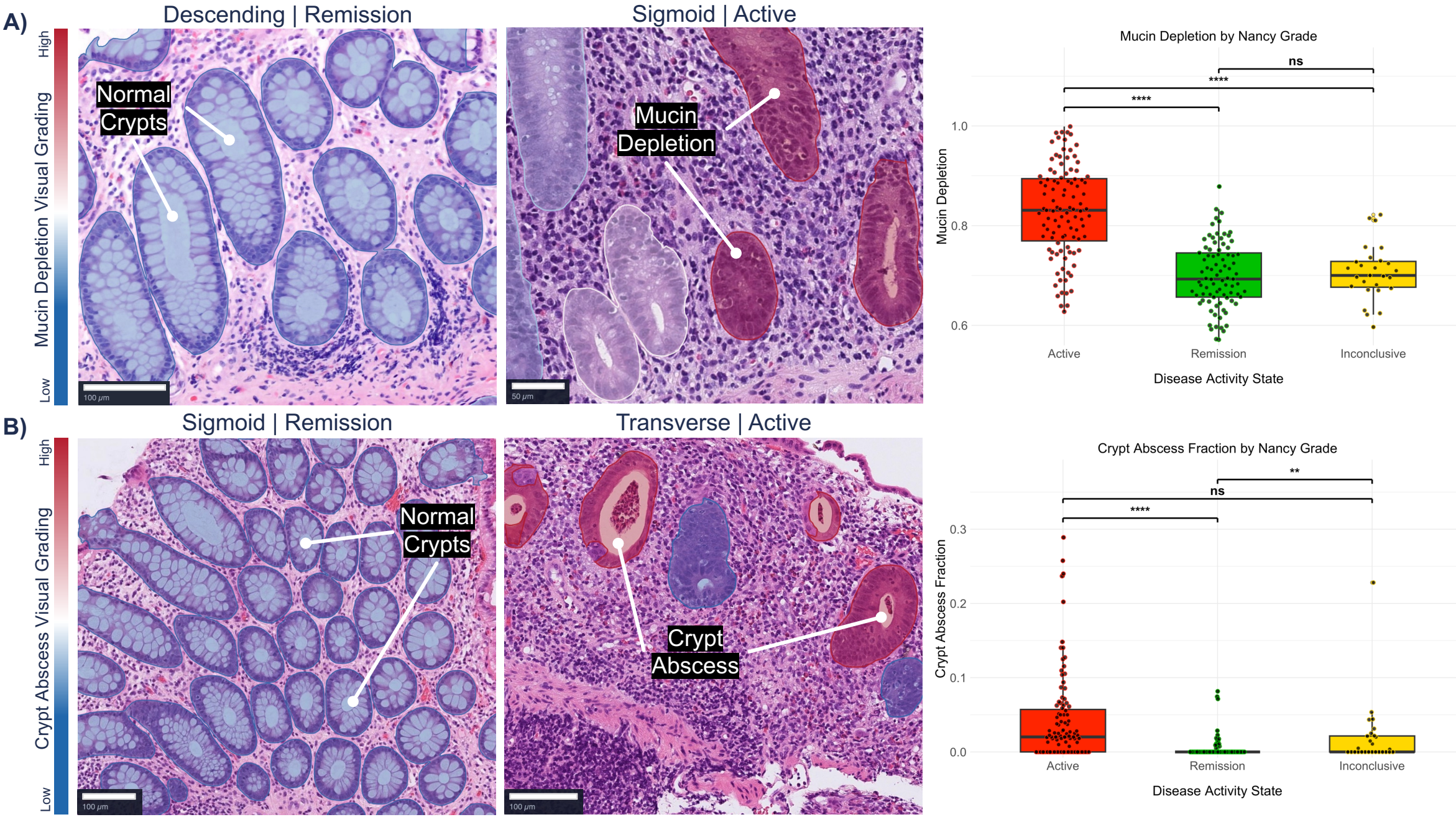

Figure 5 – Representative AI-generated overlays showing crypt morphology metric changes in UC and their comparison with two pathologist scores. (A) Crypt abscess presence is highlighted in active disease cases. (B) Mucin depletion is displayed as a loss of mucin content across crypts and in remission 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.

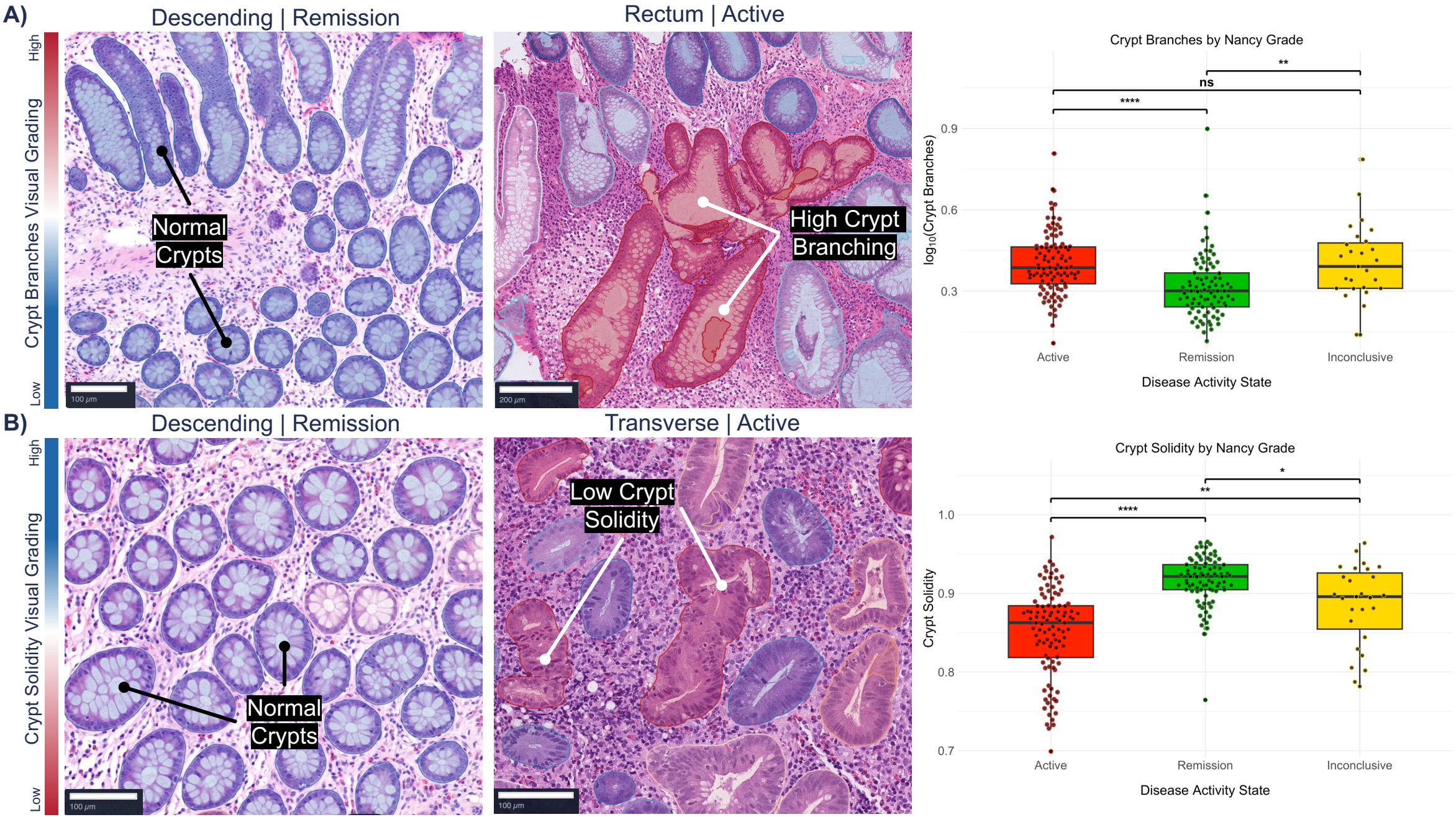

Figure 6 – Representative AI-generated overlays showing crypt morphology metric changes in UC and their comparison with two pathologist scores. (A) Crypt branching is represented across groups of branched crypts in active disease cases. (B) Crypt solidity is displayed as whether crypt shape is deformed 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.