Title

Machine Learning-Based Prediction of Pediatric Ulcerative Colitis Treatment Response using Diagnostic Histopathology

Journal

Gastroenterology, 166(5)

ISSN

0016-5085

Authors

Liu, Xiaoxuan
Prasath, Surya
Siddiqui, Iram
et al.

Publication Date

2024-02-01

DOI

10.1053/j.gastro.2024.01.033

Peer reviewed

Machine Learning–Based Prediction of Pediatric Ulcerative Colitis Treatment Response Using Diagnostic Histopathology

See editorial on page 730.

The initial presentation of ulcerative colitis within the pediatric population exhibits a degree of uniformity, the majority characterized by extensive colitis at the time of diagnosis. However, the response to initial therapy demonstrates marked heterogeneity.[1] It is challenging to discriminate which patients would successfully improve on corticosteroids followed by mesalamine therapy maintenance therapy and those who would benefit from early introduction of biologic therapy. The Clinical and Biological Predictors of Response to Standardized Pediatric Colitis Therapy (PROTECT) study, a multicenter inception cohort study, aimed to address this.[1,2] The study identified 3 early clinical features: pediatric ulcerative colitis activity index (PUCAI) of <45, hemoglobin of ≥ 10 g/dL at the time of diagnosis, and week 4 clinical remission as predictors of corticosteroid-free clinical remission on mesalamine maintenance therapy alone (CSFR) at 1 year. Moreover, PROTECT offered novel insights into the prognostic utility of histologic features assessed at disease onset—notably, surface villiform architectural abnormality[1,3] and rectal eosinophilia.

The manual evaluation of histologic slides remains indispensable. However, in the setting of a predictive tool that has the potential for wide adoption, such an approach would be restrictive. Alternatively, automated image processing can provide standardized, quantitative, and high-throughput analysis that has the potential to be widely implemented in clinical practice. In this study, we applied advanced computational approaches to the PROTECT diagnostic H&E-stained rectal biopsy specimens to develop an automated image analysis framework for patient classification.

A subcohort of 292 treatment-naive patients with rectal H&E biopsy samples available for digitization from PROTECT were included for model development. The external validation test cohort included 113 pediatric patients followed in the Canadian Children Inflammatory Bowel Disease Network inception cohort study at the Hospital for Sick Children (SickKids).[4,5] We used the PROTECT study primary outcome of CSFR at 1 year on mesalamine therapy alone and with no colectomy. Clinical remission was defined as a PUCAI score of <10 and with no corticosteroid use for 4 weeks or longer immediately before 1 year. The outcome measure for the external test cohort was analogous to PROTECT, except that the use of other mesalamine therapy in addition to Pentasa (Shire Pharmaceuticals/Pantheon) was permitted.

We first implemented a 2-step preprocessing strategy composed of stain normalization and patch generation to standardize the PROTECT and SickKids whole-slide images (WSIs).[6,7] We adopted a brightness ratio of 0.8 and overlap patch ratio of 0.25, generating 187,571 informative 512 $\times$ 512 patches from the 292 PROTECT WSI data (male, 53%; age: 12.7 years [interquartile range (IQR): 11–15]; White, 83%; PUCAI, 50 [IQR, 35–65]; CSFR, 41%) and 85,842 patches from the 113 SickKids WSI data (male, 60%; age, 13 year [IQR, 11–15]; White, 51%; PUCAI, 60 [IQR, 40–75]; CSFR, 40%). These patches were used to compute the histomic features for model training. Histomic features are objectively quantifiable and interpretable, representing various morphologic architectures within the tissue. The features capture the spatial arrangement, shape, color, intervoxel patterns, and orientations in a given image. We constructed 5 different classes of histomic features: nuclei, histogram based, and hue saturation value (HSV) color features as well as 2 texture features—gray-level co-occurrence matrix (GLCM) features and local binary pattern (LBP)—to capture information at the patch level. We computed 250 histomic input features at the patch level from the 5 classes: 11 nuclei (Otsu), 9 HSV color, 64 histogram-based, 156 GLCM, and 10 LBP features (Figure 1*A*).[8,9]

We first trained the histomic features on 14 machine learning models with 5-fold cross-validation for patch-level classification using the Scikit-learn library.[10] Feature importance was determined by the mean decrease in Gini (MDG), a measure of how each variable discriminates each image into its correct class, averaged across all decision trees (Figure 1*A*). We then selected the optimal features for classification and retrained the patch-level models. We undertook an alternative approach to further understand the impact of each feature class. We trained the 5-class features independently and determined the most discriminative features based on the MDG. We combined the optimal features into a single feature pool and retrained the machine learning classifier. Whole slide–level prediction was defined by threshold voting. We then applied the optimal histomic features on the independent real-world external pediatric SickKids cohort. The performances of patch-level and WSI models were evaluated using area under the receiver operating characteristic curve (AUROC), accuracy, precision, sensitivity, specificity, F1 score, and the DeLong test to assess the performance difference between the various predictive models.

We first trained the machine learning models on 250 histomic features at the patch level; the optimal model

Abbreviations used in this paper: **AUROC, area under the receiver operating characteristic curve; CI, confidence interval; CSFR, corticosteroid-free remission; GLCM, gray level co-occurrence matrix; HSV, hue saturation value; IQR, interquartile range; LBP, local binary pattern; MDG, mean decrease in Gini; PUCAI, pediatric ulcerative colitis activity index; RF, random forest; SickKids, Hospital for Sick Children; WSI, whole-slide images.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2024.01.033

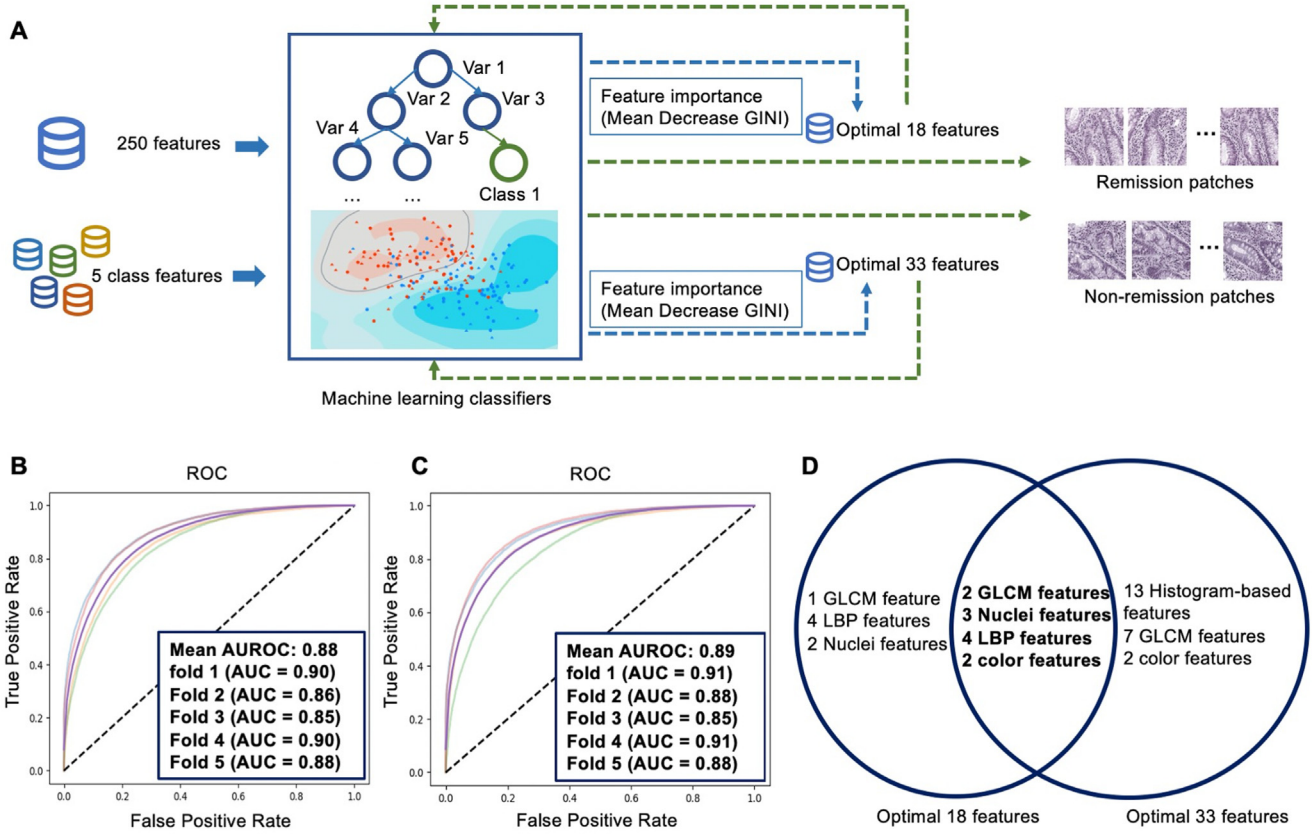


Figure 1. Overview of the machine learning approach and comparative patch-level predictive model performance. (*A*) Overview of the histomic predictive machine learning approach showing 2 parallel approaches. Thirteen machine learning models were trained using (*top*) the entire 250 features and (*bottom*) 5 class features independently. Feature importance was determined by the MDG for each approach and retrained to classify CSFR on mesalamine alone at 1 year. (*B*) The AUROC with 95% CI with 5-fold cross validation for patch-level performance using the optimal 18 features derived from the 250 features. (*C*) AUROC with 95% CI and 5-fold cross validation for patch-level performance using the optimal 33 features from the 5-class feature approach. (*D*) The Venn diagram shows shared histomic features between the 18 and 33 optimal features and includes 11 features: GLCM_contrast_3_2, GLCM_contrast_1_2, LBP_2_5, LBP_2_0, LBP_2_7, LBP_2_1, H_mean, H_thirdmoment, Otsu_equivalent_diameter, Otsu_area, and Otsu_perimeter.

trained was random forest (RF). The AUROC was 0.92 (95% confidence interval [CI], 0.89–0.95) and the accuracy was 0.92 (95% CI, 0.90–0.94), compared to 0.52 (95% CI, 0.44–0.60) and 0.53 (95% CI, 0.45–0.60), respectively, for logistic regression. Eighteen optimal features were selected based on MDG ranking, consisting of 3 GLCM, 8 LBP, 2 HSV, and 5 nuclei features. The best model trained on the 18 features at the patch level was RF, with an AUROC of 0.88 (95% CI, 0.85–0.92) and accuracy of 0.90 (95% CI, 0.80–1.00) (Figure 1*B*, Supplementary Table 1).

To evaluate the importance of each feature class, we also trained each of the 5 classes on the 2 top-performing models (Figure 1*A*). RF outperformed extra trees at the patch level for each class. The AUROC of histogram-based features was 0.85 (95% CI, 0.82–0.88), the AUROC of GLCM was 0.87 (95% CI, 0.84– 0.90), the AUROC of LBP features was 0.83 (95% CI, 0.80–0.86), the AUROC of color features was 0.80 (95% CI, 0.78–0.82), and the AUROC of nuclei features was 0.80 (95% CI, 0.76–0.83). For each class, the optimal features based on MDG were selected for a total of 33 features: 13 histogram-based features, 4 LBP features, 9 GLCM features, 4 color features, and 3 nuclei features. RF was the best model trained on the 33 features, with a patch-level AUROC of 0.89 (95% CI, 0.85–0.93) and accuracy of 0.90 (95% CI, 0.87–0.92) (Figure 1*C*).

The AUROC and accuracy at the WSI level from the patch-level model with the entire set of 250 histomic features were 0.87 (95% CI, 0.73–1.00) and 0.90 (95% CI, 0.80–1.00) and with 33 features were 0.89 (95% CI, 0.82–0.94) and 0.90 (95% CI, 0.80–1.00), respectively. The model trained using 18 optimal features was comparable with models trained on 250 features, with an AUROC of 0.89 (95% CI, 0.71–0.96) and accuracy of 0.90 (95% CI, 0.80–1.00). The DeLong test demonstrated no significant statistical difference between the predictive performance of the model using 18 vs 33 features ($P = 0.59$) (Figure 1*D*). Evaluation of the 18-histomic-features set on the real-world SickKids cohort demonstrated comparable performance. The AUROC and accuracy at the patch level were 0.88 (95% CI, 0.84–0.91) and 0.88 (95% CI, 0.82–0.92), respectively. Similarly, at the WSI level, the AUROC was 0.85 (95% CI, 0.75–0.95) and the accuracy was 0.85 (95% CI, 0.75–0.95) (Supplementary Table 2).

In the current study, we have validated 18 rectal histomic features that, when incorporated in a machine learning model, predicted steroid-free remission on mesalamine

alone in children with ulcerative colitis. These morphometric features capture the properties within the tissue and further support the development of an agnostic automated histopathology-based predictive tool using standard-of-care treatment biopsy specimens for classifying treatment response in patients with ulcerative colitis.

XIAOXUAN LIU
Division of Gastroenterology, Hepatology and Nutrition
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Division of Biomedical Informatics
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Department of Biomedical Informatics
University of Cincinnati
Cincinnati, Ohio

SURYA PRASATH
Division of Biomedical Informatics
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Department of Biomedical Informatics
University of Cincinnati
Cincinnati, Ohio

IRAM SIDDIQUI
Department of Paediatric Laboratory Medicine and Pathobiology
Division of Pathology
The Hospital for Sick Children
Toronto, Ontario, Canada

THOMAS D. WALTERS
SickKids IBD Centre
Division of Gastroenterology, Hepatology and Nutrition
The Hospital for Sick Children
Toronto, Ontario, Canada, *and*
Department of Paediatrics
University of Toronto
Toronto, Ontario, Canada

LEE A. DENSON
Division of Gastroenterology, Hepatology and Nutrition
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Department of Pediatrics
College of Medicine
University of Cincinnati
Cincinnati, Ohio

PROTECT CONSORTIUM

JASBIR DHALIWAL
Division of Gastroenterology, Hepatology and Nutrition
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Division of Biomedical Informatics
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio, *and*
Department of Biomedical Informatics
University of Cincinnati
Cincinnati, Ohio, *and*
Department of Pediatrics
University of Cincinnati
College of Medicine
Cincinnati, Ohio

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://doi.org/10.1053/j.gastro.2024.01.033.

References

1. Hyams JS, et al. Lancet Gastroenterol Hepatol 2017; 2:855–868.
2. Hyams JS, et al. Lancet 2019;393:1708–1720.
3. Boyle B, et al. Am J Surg Pathol 2017;41:1491–1498.
4. The Canadian Children Inflammatory Bowel Disease Network. https://cidscann.ca/.
5. Dhaliwal J, et al. J Crohns Colitis 2019;14:445–454.
6. Byfield P. https://github.com/Peter554/StainTools/tree/v2.1.3.
7. Vahadane A, et al. IEEE Trans Med Imaging 2016; 35:1962–1971.
8. van der Walt S, et al. PeerJ 2014;2:e453.
9. Lutnick B, et al. Proc SPIE Int Soc Opt Eng 2021; 11603:116030J.
10. Pedregosa F, et al. J Mach Learn Res 2011; 12:2825–2830.

Received October 16, 2023. Accepted January 26, 2024.

Correspondence
Address correspondence to: Jasbir Dhaliwal, MBBS, MRCPCH, MSc, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229. e-mail: Jasbir.dhaliwal@cchmc.org.

Acknowledgments
The PROTECT Consortium includes Jeffrey S. Hyams,[1] Subra Kugathasan,[2] Anne M. Griffiths,[3] Margaret H. Collins,[4] Robert N. Baldassano,[5] Brendan M. Boyle,[6] Melvin B. Heyman,[7] Neal S. Leleiko,[8,9] David Mack,[10] James Markowitz,[11] Joshua D. Noe,[12] Maria Oliva-Hemker,[13] Anthony Otley,[14] Ashish S. Patel,[15] Marian Pfefferkorn,[16] Paul A. Rufo,[17] Cary G. Sauer,[18] Jennifer Strople,[19] Boris Sudel,[20] Prateek Wali,[21] and David Ziring[22]; from the [1]Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, Connecticut; [2]Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, Georgia; [3]Department of Paediatric Laboratory Medicine and Pathobiology, Division of Pathology, The Hospital for Sick Children, Toronto, Ontario, Canada; [4]Division of Pathology and Laboratory Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [5]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [6]Nationwide Children's Hospital, Columbus, Ohio; [7]University of California–San Francisco, San Francisco, California; [8]Hasbro Children's Hospital, Providence, Rhode Island; [9]Alpert Medical School of Brown University, Providence, Rhode Island; [10]Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada; [11]Cohen Children's Medical Center, Queens, New York; [12]Medical College of Wisconsin, Milwaukee, Wisconsin; [13]Johns Hopkins Children's Center, Baltimore, Maryland; [14]IWK Health Centre, Halifax, Nova Scotia, Canada; [15]Phoenix Children's Gastroenterology, Phoenix, Arizona; [16]Riley Children's Hospital, Indiana University School of Medicine, Indianapolis, Indiana; [17]Harvard–Children's Hospital Boston, Massachusetts; [18]Emory Children's Center, Atlanta, Georgia; [19]Ann and Robert H. Lurie Children's Hospital of

Chicago, Chicago, Illinois; [20]University of Minnesota, Minneapolis, Minnesota; [21]Golisano Children's Hospital, Rochester, New York, and SUNY Upstate Medical University, Syracuse, New York; and [22]UCLA Medical Center, Los Angeles, California.

The authors would like to acknowledge Anil Jegga, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, and Department of Biomedical Informatics, University of Cincinnati, Cincinnati, Ohio; Oscar Lopez-Nunez, Division of Pathology and Laboratory Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; and James Reigle, Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.

CRediT Authorship Contributions

Xiaoxuan Liu, MSc (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Lead; Methodology: Supporting; Software: Lead; Validation: Equal; Visualization: Equal; Writing – original draft: Equal; Writing – review & editing: Supporting)

Surya Prasath, PhD (Formal analysis: Supporting; Methodology: Supporting; Validation: Supporting; Visualization: Supporting; Writing – review & editing: Supporting)

Iram Siddiqui, MBBS, MSc, FRCPC (Data curation: Supporting; Writing – review & editing: Supporting)

Thomas D. Walters, MBBS, MSc, FRACP (Data curation: Supporting; Writing – review& editing: Supporting)

Lee A. Denson, MD (Data curation: Supporting; Methodology: Supporting; Supervision: Supporting; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Jasbir Dhaliwal, MBBS, MRCPCH, MSc (Conceptualization: Lead; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Supervision: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Equal; Writing – review & editing: Lead)

Conflicts of interest

The authors disclose no conflicts.

Funding

Research reported here was supported by the Crohn's & Colitis Foundation's Clinical Research Investigator-Initiated Awards (award number 879083); National Institutes of Health (P30 DK078392) of the Digestive Diseases Research Core Center in Cincinnati; and PROCTER Scholar Award, Cincinnati Children's Hospital Medical Center. The data and biospecimens from the Predicting Response to Standardized Pediatric Colitis Therapy study (U01DK095745) reported here were supplied by the National Institute of Diabetes and Digestive and Kidney Diseases Central Repository. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The data and biospecimens from the external validation cohort were supplied from The Canadian Children Inflammatory Bowel Disease Network, supported by the Canadian Institutes of Health Research and the Children's Intestinal and Liver Disease Foundation (grant number 27862).

Data Availability

Data from the Predicting Response to Standardized Pediatric Colitis Therapy (https://doi.org/10.58020/p490-y092) reported here are available for request at the National Institute of Diabetes and Digestive and Kidney Diseases Central Repository website Resources for Research, https://repository.niddk.nih.gov/. The analytical approach and code are available on request.

Supplementary Materials and Methods

Study Participants

PROTECT was a multicenter inception cohort study based at 29 centers in the United States and Canada.[e1] A total of 400 children aged 4 to 17 years with a diagnosis of ulcerative colitis based on established clinical, endoscopic, and histologic parameters were included. Inclusion criteria included disease extension beyond the rectum, a baseline PUCAI score of at least 10, no previous therapy for colitis, and a stool culture result that was negative for enteric bacterial pathogens, including *Clostridium difficile* toxin. A detailed protocol and study description can be found in Hyams et al,[e1,e2] Depending on the initial PUCAI score (PUCAI of <10 denoted inactive disease or remission, $10-30$ denoted mild disease, $35-60$ denoted moderate disease, and 65 or higher denoted severe disease), patients received initial treatment with either mesalamine (mild disease) or corticosteroids (moderate and severe disease), with physician discretion permitted. A detailed description of treatment guidelines is provided in Hyams et al.[e1] All patients on mesalamine received the study drug in the form of Pentasa (Shire Pharmaceuticals/Pantheon).

The biopsy samples from both PROTECT and SickKids were taken from the most inflamed part of the rectosigmoid and were routinely processed and fixed in formalin and embedded in paraffin blocks from which 4- to 5-μm sections were cut and stained with H&E (Roche, HE600). All slides were scanned at $20\times$ with an Aperio T2 (AT2 DX) for digital analysis. The PROTECT biopsy specimens were all processed at Cincinnati Children's Medical Center, and the SickKids biopsy specimens at were processed at their clinical pathology laboratory.

Image Preprocessing

We undertook stain normalization by first applying the Python Staintools library,[e3] a structure-preserving color package on the WSI, to standardize the slides, with 1 WSI identified as the benchmark image. We undertook brightness normalization using Luminosity Standardizer,[e4] followed by stain normalization with the Vahadane method.[e4] The digitally driven stain normalization process allows standardizing the stain color appearance of a source image with respect to a reference image (also referred to as the target image), with no specific laboratory preanalytical or procedure protocols or other expertise required. We generated patches of 512×512 pixels and undertook experiments to determine the optimal overlap ratio and brightness threshold parameters. The overlap ratio indicates the overlap between patches, with the aim of providing sufficient coverage of the WSI, with the brightness threshold determining informative from noninformative patches. We applied the same imaging preprocessing parameters on both cohorts' WSIs.

Histomic Features

Algorithms have been manually engineered to extract distinctive characteristics and repeated patterns from histopathology images that can be used as input features in machine learning models. We constructed 5 different classes of histomic features: nuclei, histogram-based, and HSV color features as well as 2 texture features—GLCM and LBP features—to capture information at the patch level. The LBP feature creates a binary pattern by comparing the intensity of each pixel in an image to the intensity of its neighboring pixel and encodes whether it is darker or brighter.[e5] Histogram-based features represent the distribution of color or intensity values in an image using a histogram.[e6] The mean and standard deviation of the pixel values in a histogram can be used to represent the brightness and contrast of an image, whereas the skewness and kurtosis describe the texture. GLCM texture features describe the spatial relationship between pixel intensities in an image.[e7] Nuclei features were generated based on 3 different polygon methods (Otsu threshold, Delaunay triangulations, and Voronoi diagrams), and features were determined from the pixel value from each polygon. Five nuclei features used the Otsu algorithm, which is a thresholding method that can automatically separate an object of interest (eg, nuclei) at a given threshold from the background tissue.[e8] Delaunay triangulations and Voronoi diagram algorithms were applied to understand the spatial relationships between the nuclei. Imaging data were read by the SimpleITK package in Python,[e9] and GLCM and LBP features were generated by the skimage packages[e10] (graycomatrix, local_binary_pattern). We used HistomicsTK, a Python package for the analysis of digital pathology images, to count the number of nuclei.[e11] Features were implemented using self-developed functions without relying on pre-existing packages or libraries.

Model Training and Identification of Optimal Features

We first trained histomic features on 14 machine learning models (naive Bayes–based model; CatBoost; AdaBoost; tree-based models—extra trees, random forest, and decision trees; Bagging; GradientBoosting; and logistic regression (reference standard) with 5-fold cross-validation for patch-level classification. We grouped patches at the patient level (WSI level) for each fold, using the StratifiedGroupKFold function. We fine-tuned the hyperparameters by grid search.[e12] Models were implemented and built by Scikit-learn library.[e13] Feature importance was determined by the MDG. Features with higher MDG have the greatest predictive power and are most important for classification.[e14] The feature importance was computed using the Scikit-Learn features_importance function and was normalized. We selected the optimal features for classification and retrained the patch-level models.

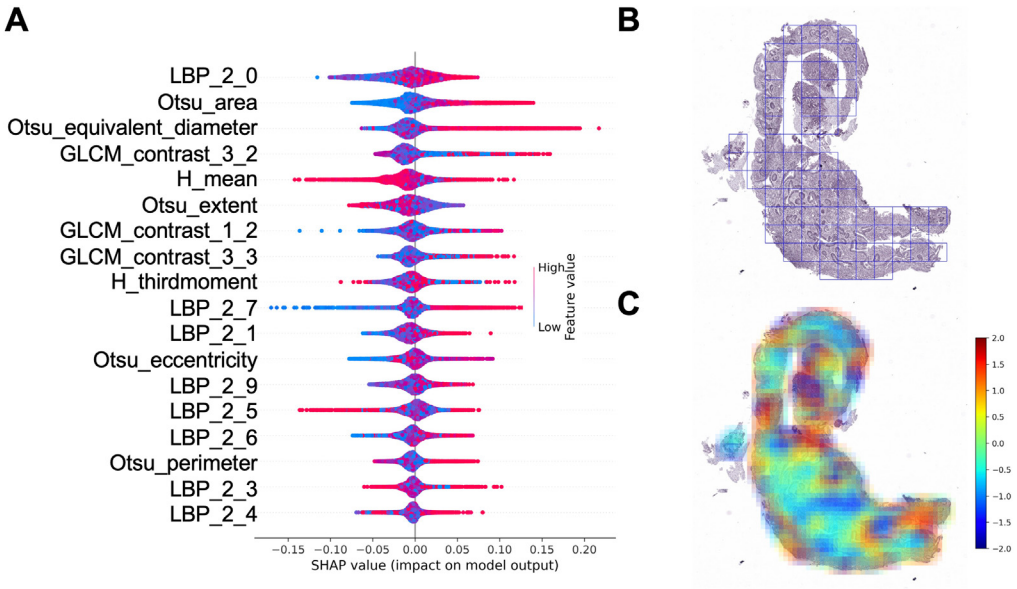
Interpretability of Histomic Features

To evaluate the impact and importance of the optimal features on remission prediction, we generated the Shapley additive explanation (SHAP) value.[e14] The SHAP value measures the contribution of the feature value to the prediction value, as illustrated in Supplementary Figure 1*A*. The *y*-axis indicates the feature. The *x*-axis indicates the SHAP

value, where a positive value is indicative of CSFR and a negative value of non-CSFR. The magnitude of the feature value is represented by a color bar, with red indicating high and blue representing low. Each individual patient is represented by a dot. We noted nucleus features with a higher value—Otsu_equivalent_diameter, Otsu_area, and Otsu_perimeter—have a positive impact on the likelihood of CSFR. Conversely, a high value of LBP_2_5 has a negative impact (negative SHAP value) on the likelihood of CSFR. Supplementary Figure 1*B* visualizes the feature Otsu_equivalent_diameter at the slide level.

Supplementary References

e1. Hyams JS, et al. Lancet 2019;393:1708–1720.
e2. Hyams JS, et al. Lancet Gastroenterol Hepatol 2017; 2:855–868.
e3. Byfield P. https://github.com/Peter554/StainTools/tree/v2.1.3.
e4. Vahadane A, et al. IEEE Trans Med Imaging 2016; 35:1962–1971.
e5. Ojala T, et al. Pattern Recognition 1996;29:51–59.
e6. Chapelle O, et al. IEEE Trans Neural Netw 1999; 10:1055–1064.
e7. Mohanaiah P, et al. IJSRP 2013;3:1–5.
e8. Linares OC, et al. In: 2020 IEEE 33rd International Symposium on CMBS.
e9. Lowekamp B, et al. Frontiers. Neuroinformatics 2013;7.
e10. van der Walt S, et al. PeerJ 2014;2:e453.
e11. Lutnick B, et al. Proc SPIE Int Soc Opt Eng 202111603.
e12. Probst P, et al. WIREs 2019;9:e1301.
e13. Pedregosa F, et al. JMLR 2011;12:2825–2830.
e14. Lundberg SM, Lee S-I. In: Advances in Neural Information Processing Systems 30 (NIPS 2017).



Supplementary Figure 1. Histomic feature importance represented by the SHAP values. (*A*) The relationship between the 18 histomic features and the outcome of CSFR with mesalamine alone at 1 year. Positive SHAP values (*x*-axis) are indicative of clinical remission, and negative values represent nonremission. The magnitude of each feature value is represented by the color bar, with red being high and blue representing low. Features include LBP and GLCM; histogram features include H_third moment and H_mean; and nuclei features include Otsu_perimeter, Otsu_eccentricity, Otsu_area, and Otsu_equivalent_diameter. (*B*) Image 1 shows nonoverlapping patches of H&E stain–normalized WSIs. (*C*) A heatmap of the Otsu equivalent diameter, a nuclei feature, with the color bar representing feature values. High values are in red/brown, and low values are in blue.

Supplementary Table 1. Patch-Level Image Model Performance Metrics Using the 18 Optimal Histomic Features

Model type	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC
PROTECT cohort						
Random forest	0.85 (0.78–0.93)	0.91 (0.87–0.95)	0.90 (0.85–0.96)	0.90 (0.8– 0.91)	0.89 (0.86–0.91)	0.88 (0.85–0.92)
Logistic regression	0.44 (0.33–0.55)	0.57 (0.44–0.69)	0.60 (0.49–0.71)	0.52 (0.29–0.54)	0.51 (0.47–0.55)	0.50 (0.45–0.56)
SickKids cohort						
Random forest	0.90 (0.85–0.95)	0.86 (0.82–0.9)	0.91 (0.86–0.96)	0.86 (0.79–0.9)	0.88 (0.85–0.92)	0.88 (0.84–0.91)
Logistic regression	0.84 (0.78–0.89)	0.79 (0.73–0.84)	0.86 (0.8–0.93)	0.79 (0.69–0.84)	0.82 (0.78–0.86)	0.81 (0.77–0.85)

NOTE. Values are averages with 95% CI (point estimate and 95% CI).

Supplementary Table 2. Whole-Slide Image Model Performance Metrics Using the 18 Optimal Histomic Features

Model type	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC
PROTECT cohort						
Random forest	0.84 (0.78–0.90)	0.94 (0.82–1.00)	0.91 (0.82–1.00)	0.87 (0.82–0.92)	0.90 (0.80–1.00)	0.89 (0.71–0.96)
Logistic regression	0.47 (0.30–0.63)	0.50 (0.41–0.58)	0.59 (0.50–0.67)	0.48 (0.40–0.56)	0.48 (0.42–0.54)	0.48 (0.41–0.56)
SickKids cohort						
Random forest	0.78 (0.49–1.00)	0.91 (0.81–1.00)	0.84 (0.57–1.00)	0.82 (0.59–1.00)	0.85 (0.63–1.00)	0.85 (0.75– 0.95)
Logistic regression	0.42 (0.34–0.50)	0.43 (0.39–0.49)	0.53 (0.50–0.59)	0.43 (0.40–0.47)	0.48 (0.44–0.51)	0.47 (0.40–0.55)

NOTE. Values are averages with 95% CI (point estimate and 95% CI).