Machine learning-based prediction of Geboes score and histologic improvement and remission thresholds in ulcerative colitis

STUDY BACKGROUND

Histology is emerging as a potential therapeutic endpoint for ulcerative colitis (UC) driven by associations between histologic response and long-term outcomes[1]. However, existing scoring systems are subjective and consequently have variable inter- and intra-reader variability[2]. Furthermore, manual histologic assessment is semi-quantitative and limited in the ability to capture spatial relationships.

Here we report the first machine learning (ML)-based prediction of the Geboes score, and GS-derived thresholds of histologic improvement and remission[3], directly from whole slide images (WSI) of hematoxylin and eosin (H&E)-stained mucosal biopsies. Together, PathAI models for characterization of the UC histology (IBD-Explore) and the PathAI algorithm for Geboes scoring (AI-measurement of histological improvement in UC (AIM-HI UC); for research use only) have the potential to identify clinically relevant histologic features, enable robust scoring, and ultimately advance precision medicine for patients with IBD.

MODEL-GENERATED OVERLAYS

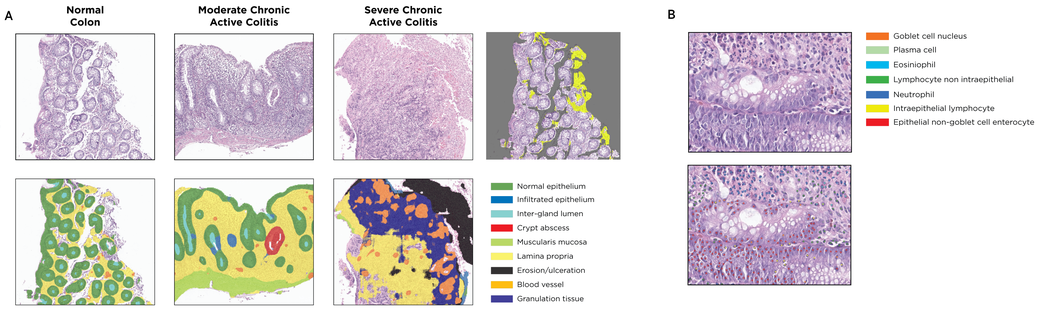

Figure 2. Example overlays generated by PathAI IBD-Explore tissue and cell models deployed on mucosal biopsy WSI. a) H&E-stained WSI and corresponding tissue overlay from normal colon, moderate chronic active colitis, and severe active chronic colitis samples. Artifact overlay shown for normal colon WSI (artifact, yellow; background, gray). b) H&E-stained WSI and corresponding cell overlay.

METHODS

- IBD-Explore tissue and cell prediction models based on convolutional neural networks (CNN) were developed using 8,245 WSI from clinical and commercial sources, split into training (5769, 70%), validation (1653, 20%) and test (823, 10%) sets. WSI were annotated by American Board of Pathology board-certified pathologists who routinely practice gastrointestinal pathology (Figures 1,2).

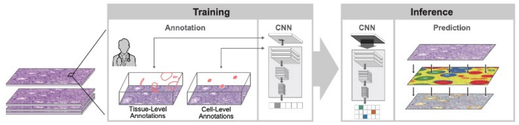

Figure 1. Schematic illustrating convolutional neural network (CNN) model development approach for pixel-level feature predictions.

- Tissue and cell predictions were used to extract human-interpretable features (HIFs) measuring tissue area proportions, cell count proportions, and cell densities.
- 4,310 WSI were then scored by three expert gastrointestinal pathologists and the median consensus score was used to determine GS for each slide. These slides were split into training (2,444, 57%), optimization (702, 16%) and test (1,164, 27%) to train a multi-task graph neural network model (GNN) to predict GS and GS subscores for each slide (Figure 3).

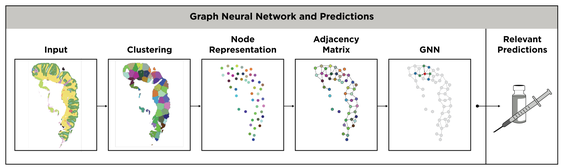

Figure 3. GNN training for prediction of Geboes scores. A GNN model learns to identify characteristic tissue phenotypes defined by cell and tissue features and the spatial arrangement of the tissue regions.

- Model performance compared to pathologist was evaluated on a held-out test set (n=1,164, including 888 WSI from an external lab not used for model training or optimization) by assessing positive percent agreement (PPA; agreement below cutoffs) and negative percent agreement (NPA; agreement above cutoffs) for histologic improvement (HI, GS ≤ 3.1) and remission (HR, GS < 2) cutoffs, and linear kappa for Geboes grade-level predictions (GS 0,1,2,3,4,5).

RESULTS

- To assess our CNN models, we used CNN tissue and cell model predictions to extract quantitative human interpretable features (HIFs) measuring tissue area proportions, cell count proportions, and cell densities. We then carried out Spearman's correlation analysis between HIFs and median consensus Geboes scores.

GS grade	Features	Correlation	p-value
GS0	AREA PROPORTION OF NORMAL EPITHELIUM OVER LAMINA PROPRIA	-0.603	< 0.001
	COUNT PROPORTION OF NON-INTRAEPITHELIAL LYMPHOCYTE CELLS OVER ALL CELLS IN MUCOSA	0.57	< 0.001
GS1	DENSITY OF LYMPHOCYTE AND PLASMA CELLS IN LAMINA PROPRIA	0.485	< 0.001
	DENSITY OF EOSINOPHIL CELLS IN LAMINA PROPRIA	0.54	< 0.001
GS2A	COUNT PROPORTION OF EOSINOPHIL CELLS OVER ALL CELLS IN LAMINA PROPRIA	0.519	< 0.001
	DENSITY OF NEUTROPHIL CELLS IN BLOOD VESSELS	0.602	< 0.001
GS2B	COUNT PROPORTION OF NEUTROPHIL CELLS OVER ALL CELLS IN MUCOSA	0.565	< 0.001
	DENSITY OF NEUTROPHIL CELLS IN LAMINA PROPRIA	0.457	< 0.001
GS3	AREA PROPORTION OF INFILTRATED EPITHELIUM AND CRYPT ABSCESS OVER NORMAL EPITHELIUM	0.773	< 0.001
	AREA PROPORTION OF INFILTRATED EPITHELIUM AND CRYPT ABSCESS OVER MUCOSA	0.769	< 0.001
	DENSITY OF NEUTROPHIL CELLS IN INFILTRATED EPITHELIUM AND CRYPT ABSCESS	0.754	< 0.001
	COUNT PROPORTION OF NEUTROPHIL CELLS OVER ALL CELLS IN ALL EPITHELIUM AND CRYPT ABSCESS	0.676	< 0.001
GS4	AREA PROPORTION OF INFILTRATED EPITHELIUM AND CRYPT ABSCESS OVER MUCOSA	0.59	< 0.001
	AREA PROPORTION OF CRYPT ABSCESS OVER MUCOSA	0.586	< 0.001
	AREA PROPORTION OF INFILTRATED EPITHELIUM OVER MUCOSA	0.544	< 0.001
	AREA PROPORTION OF INFILTRATED EPITHELIUM OVER NORMAL EPITHELIUM AND INFILTRATED EPITHELIUM	0.535	< 0.001
	DENSITY OF NEUTROPHIL CELLS IN ALL EPITHELIUM AND CRYPT ABSCESS	0.511	< 0.001
GS5	AREA PROPORTION OF GRANULATION TISSUE AND EROSION OR ULCERATION OVER MUCOSA	0.712	< 0.001
	DENSITY OF NEUTROPHIL CELLS IN GRANULATION TISSUE AND EROSION OR ULCERATION	0.699	< 0.001

Correlation Strength: 0.8, 0, -0.8

Table 1. Correlation of consensus Geboes scores with human interpretable features extracted from CNN-generated tissue and cell predictions.

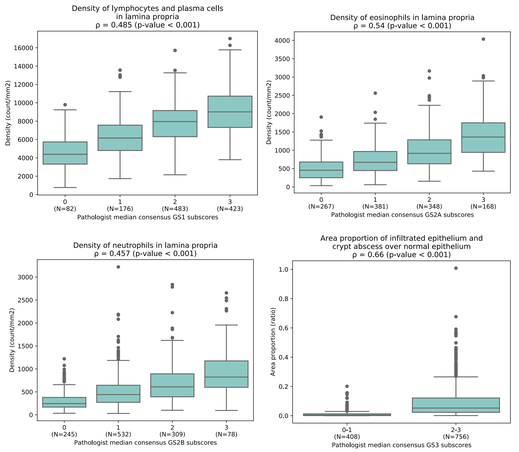

Figure 4. Correlation of GS1, GS2A, GS2B and GS3 subscores with HIFs reflecting histologic features used by pathologists for scoring.

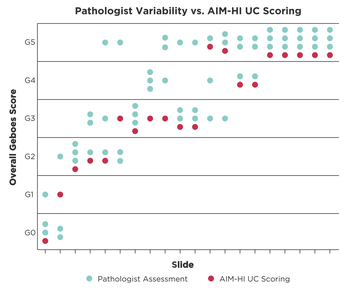

Figure 5. Variability in Geboes scoring among pathologists and AIM-HI UC. Scores from 3 pathologists and the AIM-HI UC algorithm shown for 20 slides randomly chosen from the test set.

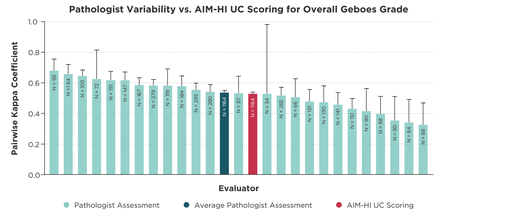

Figure 6. Average pairwise agreement among individual pathologists for overall Geboes score. The average agreement among pathologists is also computed, along with the corresponding average agreement between AIM-HI UC and pathologists.

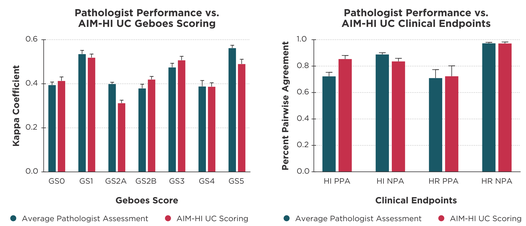

Figure 7. Average agreement between AIM-HI UC and pathologists compared to average agreement among pathologists for Geboes scores and clinically relevant endpoints. Geboes overall grade linear kappa, and PPA and NPA for histologic improvement and remission are reported.

CONCLUSIONS

We report a ML-based approach for predicting Geboes score, and key GS-based thresholds of histological improvement and histologic remission. Average agreement between AIM-HI UC and pathologists is comparable to average agreement among pathologists on the held-out test set for Geboes scores and clinically relevant endpoints.

AIM-HI UC, the PathAI Geboes scoring algorithm, can be integrated into a clinical trial workflow facilitated by the PathAI AISight Clinical Trials Platform, which enables intake of WSIs, deployment of ML models, pathologist review, and reporting of case-level results. This approach may enable standardized, reproducible, and accurate prediction of clinically relevant thresholds to better measure histologic disease activity and treatment response in clinical trials.

AUTHORS

Carlos Gaitán[1], Zahil Shanis[1], Kathleen Sucipto[1], John Shamshoian[1], Jin Li[1], George Hu[1], Harshith Padigela[1], Andrew Walker[1], Harsha Pokkalla[1], Darren Fahy[1], Geetika Singh[1], Chinmay Surve[2], Natoria Wade[2], Archit Khosla[1], Mary Lin[1], Michael Montalto[1], Andrew Beck[1], Jimish Mehta[1] , Ilan Wapinski[1], Michael Drage[1], Fedaa Najdawi[1], Christina K.B. Jayson[1]

[1]PathAI, Boston, Massachusetts
[2]PathAI Diagnostics, Memphis, Tennessee

contact:
kathleen.sucipto@pathai.com
john.shamshoian@pathai.com

REFERENCES

1. Gonzalez-Partida, I. et al. *Eur J Gastroenterol Hepatol*. 2021; 33:e796-e802.
2. Romkens, TH. et al. *J Crohns Colitis*. 2018; 12:425-431.
3. Li, K. et al. *Gastroenterol*. 2020; 159:P2052-P2064.

ACKNOWLEDGMENTS

We would like to thank the members of PathAI Pathology Network who contributed towards this work.

PathAI