Poster #074

Quantitative and explainable AI-powered approaches to predict ulcerative colitis disease activity from hematoxylin and eosin (H&E)-stained whole slide images (WSI)

STUDY BACKGROUND

Microscopic inflammation has been shown to be an important indicator of disease activity in ulcerative colitis (UC)[1]. However, manual histologic scoring is semi-quantitative and subject to interobserver variation, and AI-based solutions often lack interpretability[2].

Here we report two distinct quantitative approaches to predict disease activity scores and histological remission using AI-powered digital pathology. Both the random forest classifier (RFC) and graph neural network (GNN) further provide explainability and biological insight by identifying histological features informing model predictions.

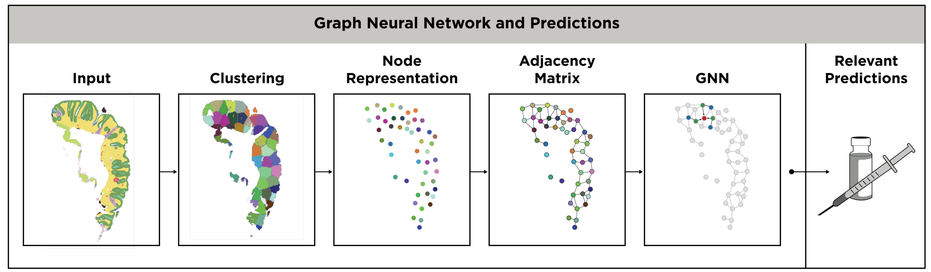

Figure 4. GNN training for prediction of NHI scores. To construct a graph in each WSI, we first identified the connected components for each tissue class on the CNN-generated tissue overlays. Connected components were clustered via Birch clustering. The final clusters defined the nodes of the graph, and directed edges were constructed from each node to its at most five nearest neighbors within a distance threshold. Both CNN-generated cell and tissue overlays were used to extract the features for each node. Once the graph structure for each slide was constructed, we trained a GNN model on 410 slides to predict the NHI scores.

METHODS

- Convolutional neural networks (CNNs) were developed using >162k annotations on 820 WSI of H&E-stained colorectal biopsies for pixel-level identification of tissue regions and cell types. All WSI were scored by 5 board-certified pathologists using the Nancy Histological Index (NHI) to establish consensus ground truth. (Figures 1,2).

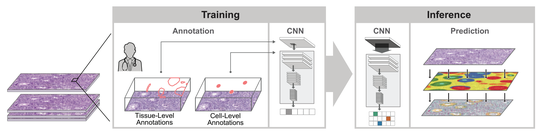

Figure 1. Schematic illustrating convolutional neural network (CNN) model development approach for pixel-level feature predictions.

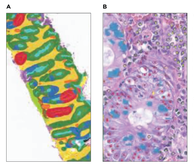

Figure 2. Model-generated overlays. (a) Tissue region overlay showing areas of crypt abscess, lamina propria, erosion/ulceration, etc. (b) Cell type overlay showing neutrophils, plasma cells, etc.

- Combining information from the tissue and cell model predictions, we extracted histological human-interpretable features (HIFs) such as cell densities and cell count proportions, and tissue area proportions (Figure 3). A rich, quantitative set of human interpretable features was then used to train a random forest classifier (RFC) to predict slide-level NHI score.

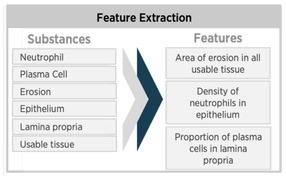

Figure 3. Extraction of human-interpretable features, Tissue and cell model predictions (substances) are used to generate higher-order features capturing tissue area proportions and cell count proportions and densities.

- To test the hypothesis that tissue region spatial relationships and cellular composition can inform AI-based predictions of disease activity, we also trained a separate graph neural network (GNN) to predict NHI score (Figure 4).

RESULTS (RFC)

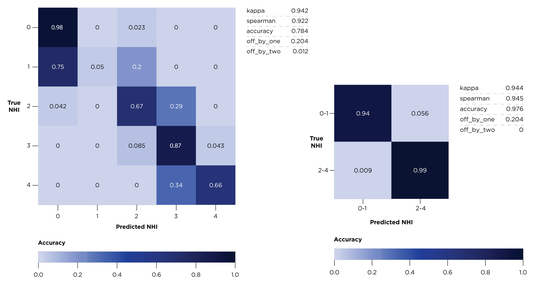

Figure 5. Accuracy of RFC predictions of NHI scores and histologic remission. Confusion matrices showing RFC-predicted vs consensus ground truth NHI scores for individual NHI scores (left) and binarized NHI scores (right). Weighted kappa, Spearman correlation, and accuracy of predictions were calculated. In addition, the percentage of predictions off by one, and off by two, are reported.

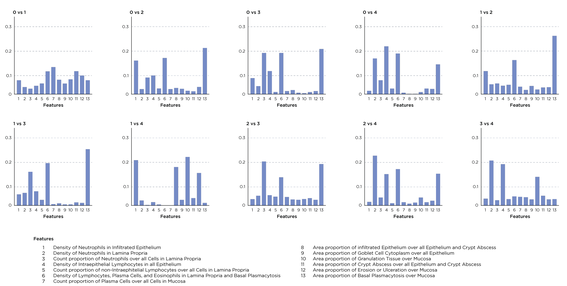

Figure 6. Histologic features contributing to RFC predictions of NHI scores. The RFC model identified features known to be relevant to how NHI is scored, such as cell features of plasma cells and lymphocytes that distinguish NHI 1 cases, and features known to be relevant to disease activity, such as infiltrated epithelium or neutrophil cell features that distinguish higher NHI scores (NHI 2, 3 and 4). The model also identified features beyond those assessed by the NHI, such as area proportion of basal plasmacytosis associated with predictions of NHI 2 and 3. Other features not previously implicated in UC disease activity were also highlighted, including intraepithelial lymphocytes associated with higher NHI scores (NHI 3 and 4).

RESULTS (GNN)

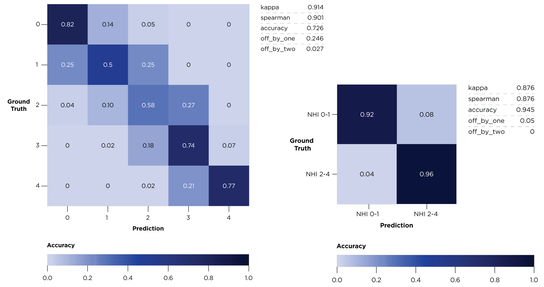

Figure 7. Accuracy of GNN predictions of NHI scores and histologic remission. Confusion matrices showing GNN-predicted vs consensus ground truth NHI scores for individual NHI scores (left) and binarized NHI scores (right). Weighted kappa, Spearman correlation, and accuracy of predictions were calculated. In addition, the percentage of predictions off by one, and off by two, are reported.

NHI scores	Important features
0	Normal epithelium mean
1	Infiltrated epithelium mean
2	Non-intraepithelial lymphocyte std, basal plasmacytosis mean, infiltrated epithelium mean
3	Intra-epi lymphocyte std, basal plasmacytosis mean, crypt abscess mean, infiltrated epithelium mean, inter gland lumen mean
4	Neutrophil std, erosion or ulceration mean, granulation tissue mean

Figure 8. Histologic features contributing to GNN predictions of NHI scores. Example application of GNNExplainer[3] to identify features and locate subgraphs that significantly contribute to the scoring prediction (left). Features of importance identified by GNNExplainer (right). Features that are consistent with how NHI is scored were identified, including infiltrated epithelium for NHI 2 and NHI 3 or area proportion features of erosion, ulceration, and granulation tissue for NHI 4. The GNN-based approach also identified features beyond those assessed by the NHI, including area proportion of basal plasmacytosis associated with predictions of NHI 2 and 3. Other features not previously implicated in UC disease activity were highlighted, such as cell features of intraepithelial lymphocytes associated with predictions of NHI 3.

CONCLUSIONS

We report quantitative and interpretable AI-powered approaches for UC histological assessment. CNN identification of UC histology was used as input to two distinct disease activity classifiers that showed strong concordance with consensus pathologist scoring. Both approaches provide interpretable features that explain model predictions and that may be used to inform biomarker selection and clinical development efforts.

AUTHORS

Kathleen Sucipto[1], Archit Khosla[1], Michael G. Drage[1], Yilan Wang[1], Darren Fahy[1], Mary Lin[1], Murray B. Resnick[1], Michael Montalto[1], Andrew Beck[1], Ilan Wapinski[1], Stephanie Hennek[1], Christina K.B. Jayson[1], & Fedaa Najdawi[1]

[1]PathAI, Boston, Massachusetts

contact:
fedaa.najdawi@pathai.com
archit.khosla@pathai.com

**Disclosure: All authors are current employees of PathAI, or were formerly employed by PathAI during the research period, and may own company stock.*

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. Ying, R. *arXiv*. 2019; 1903.03894 (cs.LG).

ACKNOWLEDGEMENTS

This poster template was developed by SciStories LLC.
https://scistories.com/

PathAI

Crohn's and Colitis Congress, 19-21 January 2023, Denver, CO, USA