Interpretable histopathology-based prediction of disease relevant features in Inflammatory Bowel Disease biopsies using weakly-supervised deep learning

Ricardo Mokhtari[1], Azam Hamidinekoo[1], Daniel Sutton[1], Arthur Lewis[1], Bastian R. Angermann[2], Ulf Gehrmann[2], Jessica Neisen[2], Pål Lundin[2], Hibret Adissu[1], Junmei Cairns[2], Emon Khan[2], Daniel Marks[2], Nia Khachapuridze[2], Talha Qaiser[1], Nikolay Burlutskiy[1]

[1]Clinical Pharmacology and Safety Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge UK ; [2]Respiratory & Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK & Gothenburg, Sweden

Abstract

Crohn's Disease (CD) and Ulcerative Colitis (UC) are the two main Inflammatory Bowel Disease (IBD) types. We developed deep learning models to identify histological disease features for both CD and UC using only endoscopic labels. We explored fine-tuning and end-to-end training of two state-of-the-art, self-supervised models for predicting three different endoscopic categories. We produced attention maps to interpret what the models learned and validated them with the support of a pathologist, where we observed a strong association between the models' predictions and histopathological inflammatory features of the disease. In parallel, we utilised a model trained on the Colon Nuclei Identification and Counting (CoNIC) dataset to predict and explore 6 cell populations. We observed correlation between areas enriched with the predicted immune cells in biopsies and the pathologist's feedback on the attention maps. These models can enhance our understanding about IBD pathology and can shape our strategies for patient stratification in clinical trials.

Introduction

- Histopathological endpoints are evolving as a clinical trial endpoint in Inflammatory Bowel Disease (IBD). Use of histology to screen potential clinical study subjects could add value in IBD clinical trials; for example, by refining eligibility criteria to ensure studies recruit patients with definitive active inflammation at the microscopic level.
- Several histopathological indices have been developed, but the relative complexity of available scores requires large-scale labour-intensive annotation by a pathologist.
- We aimed to develop computer vision tools to assist decoding the complex clinical disease features at the histological level for both Crohn's Disease (CD) and Ulcerative Colitis (UC). This will inform understanding of disease pathology and patient stratification to support clinical trial development strategies.

Methods

Dataset

A total of 1394 clinically annotated Haematoxylin & Eosin (H&E) stained whole slide images (WSIs) were included from 418 CD and 218 UC patients enrolled in a multicentred longitudinal Study of a Prospective Adult Research Cohort with IBD (SPARC IBD) obtained from the IBD Plexus program of the Crohn's & Colitis Foundation [1].

Pre-processing and Training

- To pre-process the data, we trained an image quality control (QC) model to automatically segment and remove imaging artefacts
- We trained 2 recently published, state-of-the-art predictive models: DSMIL [2] and HIPT [3] to predict 4 tasks in SPARC IBD (Table 1) and explored 2 training strategies: fine-tuning and end-to-end (E2E) training
- Fine-tuning involved using public models pre-trained on The Cancer Genome Atlas (TCGA) dataset, E2E training involved training all model components from scratch on SPARC IBD
- We generated visual attention maps to interrogate what the models learnt and validated them with a pathologist

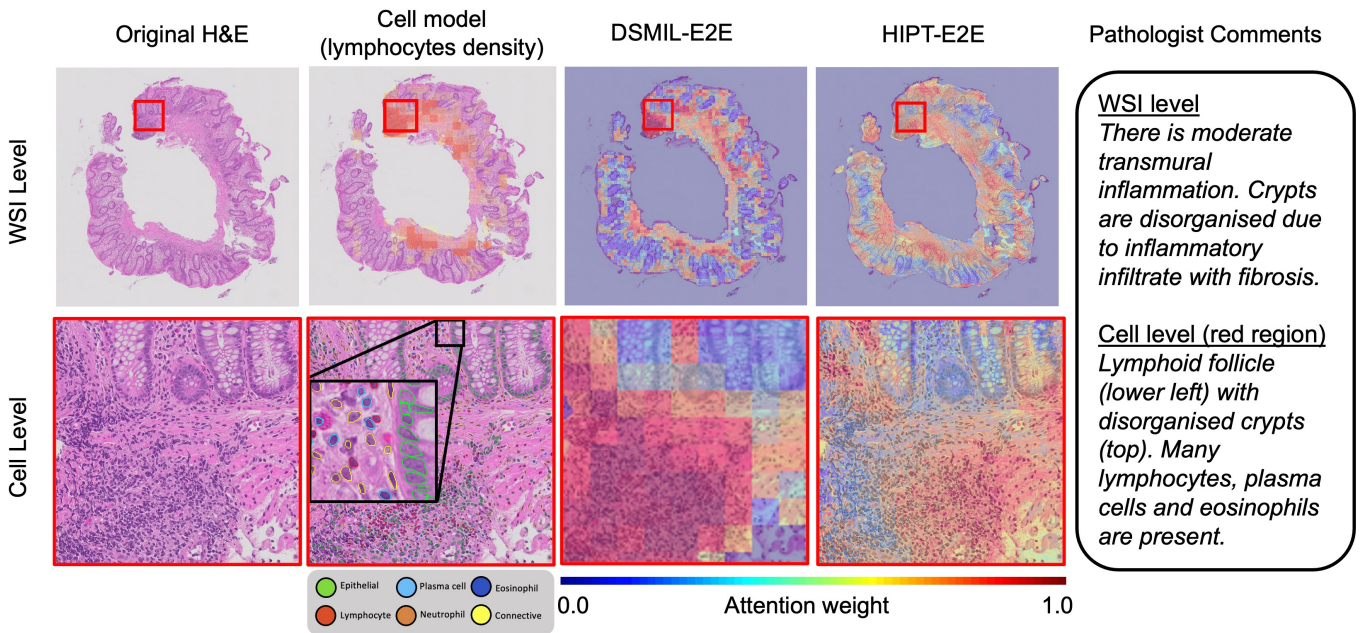

Figure 1: H&E image, DSMIL and HIPT attention maps for the prediction of lesional macroscopic appearance, cell model predictions (lymphocytes heatmap) and pathologist comments for UC patient with high endoscopic score.

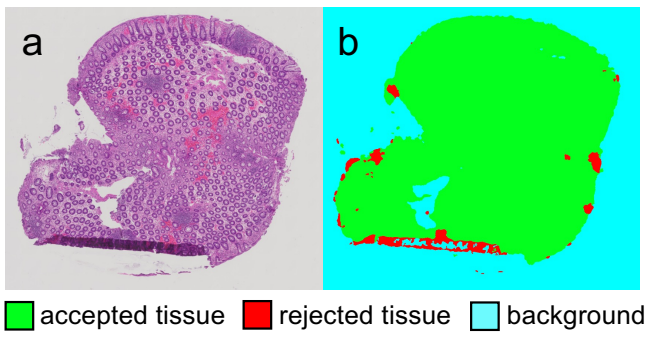

Figure 1: QC model applied to tissue correctly identifies artefacts such as tissue folding a) original slide b) QC model prediction

- We trained a cell prediction model and aggregated the predictions into heatmaps to quantify cell populations in the tissue

Results

QC Model

The QC model identified 11 slides that contained >50% rejected tissue, which were discarded (Fig. 2). For the remaining slides, we only saved patches with >50% accepted tissue.

DSMIL and HIPT Predictive Performance

- We found that HIPT significantly outperforms DSMIL on all prediction tasks and E2E training is the preferred strategy (Table 1) for the SPARC IBD dataset.
- We believe that the transfer of knowledge from models trained on TCGA is poor due to the differences in morphology between cancer and IBD as well as the difference between resection and biopsy sectioning methods
- We theorize that HIPT's transformer architecture better captures the spatial patterns among cells and IBD morphology as well as the multi-scale nature of IBD pathology, leading to improved performance over DSMIL

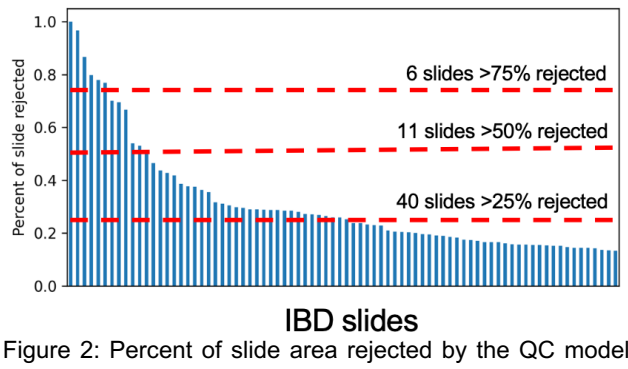

Figure 2: Percent of slide area rejected by the QC model for the top 80 slides in the SPARC IBD dataset.

Table 1: Mean AUROC ± standard error (5-fold cross validation) of trained models across different prediction tasks in the SPARC IBD dataset. DD=disease diagnosis, MA=Macroscopic Appearance, ES=Endoscopic Score, CD=Crohn's Disease, UC=Ulcerative Colitis, F=fine-tuned, E2E=end-to-end training

Model	DD	MA	ES (CD)	ES (UC)
DSMIL-F	0.656 ±0.007	0.522 ±0.008	0.528 ±0.016	0.592 ±0.007
DSMIL-E2E	0.692 ±0.010	0.750 ±0.006	0.740 ±0.009	0.634 ±0.017
HIPT-F	0.825 ±0.017	0.780 ±0.012	0.766 ±0.026	0.788 ±0.034
HIPT-E2E	**0.865 ±0.019**	**0.814 ±0.008**	**0.786 ±0.017**	**0.802 ±0.014**

Visual Attention Maps

The DSMIL and HIPT attention maps (Fig. 1) were compared with pathologist assessment and annotation. We found a strong association between the models' high attention areas and diseased tissue areas identified by the pathologist. The cell density heatmaps (Fig. 1) were confirmed to identify the correct cell populations at the WSI level, although further evaluation is needed at the cell level.

Conclusions

We have demonstrated that weakly-supervised learning can be used to build high performing predictive models for H&E stained IBD biopsies that identify disease relevant features such as inflammation. Further work is ongoing to distinguish other disease specific characteristics in IBD. In addition, we plan to validate the models and the attention map approach on IBD clinical trial datasets.

References

1. Raffals, L.E., Saha, S., Bewtra, M., Norris, C., Dobes, A., Heller, C., O'Charoen, S., Fehlmann, T., Sweeney, S., Weaver, A. and Bishu, S., 2022. The development and initial findings of a study of a prospective adult research cohort with inflammatory bowel disease (SPARC IBD). *Inflammatory bowel diseases*, *28*(2), pp.192-199.
2. Li, B., Li, Y. and Eliceiri, K.W., 2021. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14318-14328).
3. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G. and Mahmood, F., 2022. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16144-16155).

Acknowledgements

We would like to thank AstraZeneca and the Crohn's and Colitis Foundation for supporting this research.

Presented at Medical Imaging with Deep Learning (MIDL 23), Nashville, Tennessee, 10-12th July 2023

Read the full paper: