

Leveraging Health Data to Predict Colonoscopy Findings in IBD Patients: A Visualisation and Machine Learning Approach

Author: Tobi Dairo

Introduction

Inflammatory Bowel Disease (IBD) affects thousands of lives across the UK, often requiring invasive procedures. Biomarker test results have demonstrated potential in predicting IBD outcomes (Kraszewski et al., 2021; Lewis, 2011).

This project aims to enhance the management and treatment of IBD by:

- 1) Creating a unified patient event log that consolidates diverse data sources into a single, comprehensive file.
- 2) Developing an interactive timeline for clinicians to visualise and track patient histories efficiently.
- 3) Utilising machine learning to predict colonoscopy findings based on biomarker test results, potentially reducing the need for invasive procedures.

Methodology

Patient Event Log

NHS Trust data on medication events, tests, discharges, and colonoscopies were consolidated into a single event log using Python.

Interactive Timeline

A visualisation of patient histories was created using Power BI to develop an interactive timeline.

Prediction Model

Key biomarkers (e.g. IgG, faecal calprotectin, CRP) were used in a Gradient Boosting model to predict colitis or ileitis in colonoscopy patients. Data preparation included feature engineering, followed by model training, validation and testing. Hyperparameter tuning was performed using randomised search, grid search, and Bayesian optimisation, all with cross-validation (Figure 2).

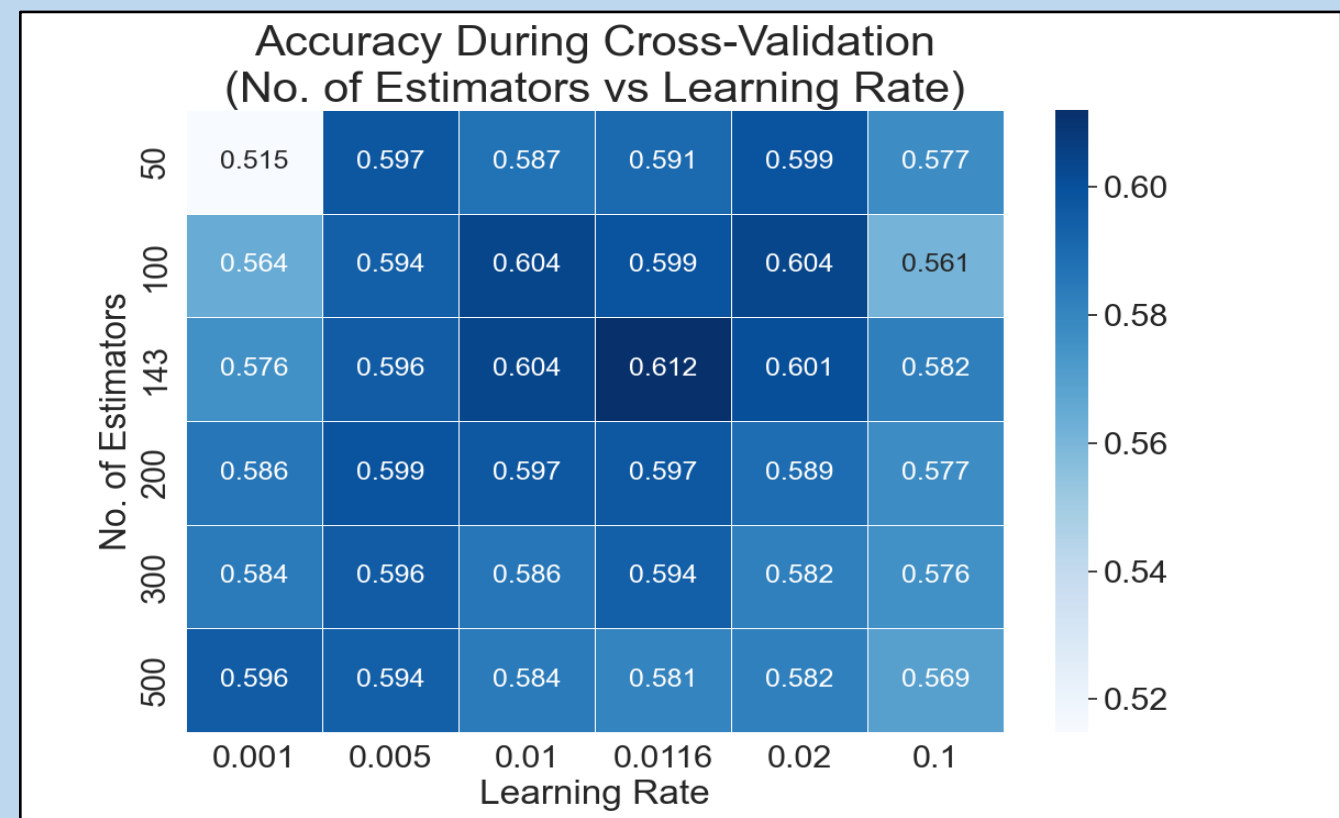


Figure 2: Gradient Boosting Model Cross-Validation Scores

Feature Importance Summary (SHAP)

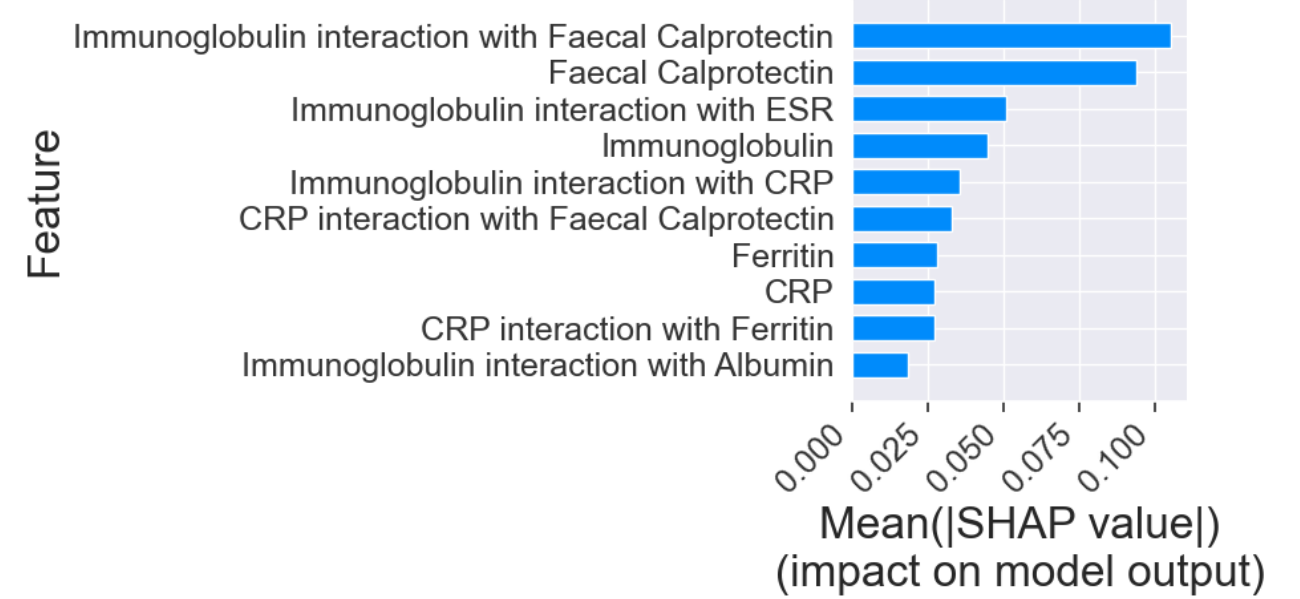


Figure 3: Summary of Feature Importance for Prediction Model

Outcomes

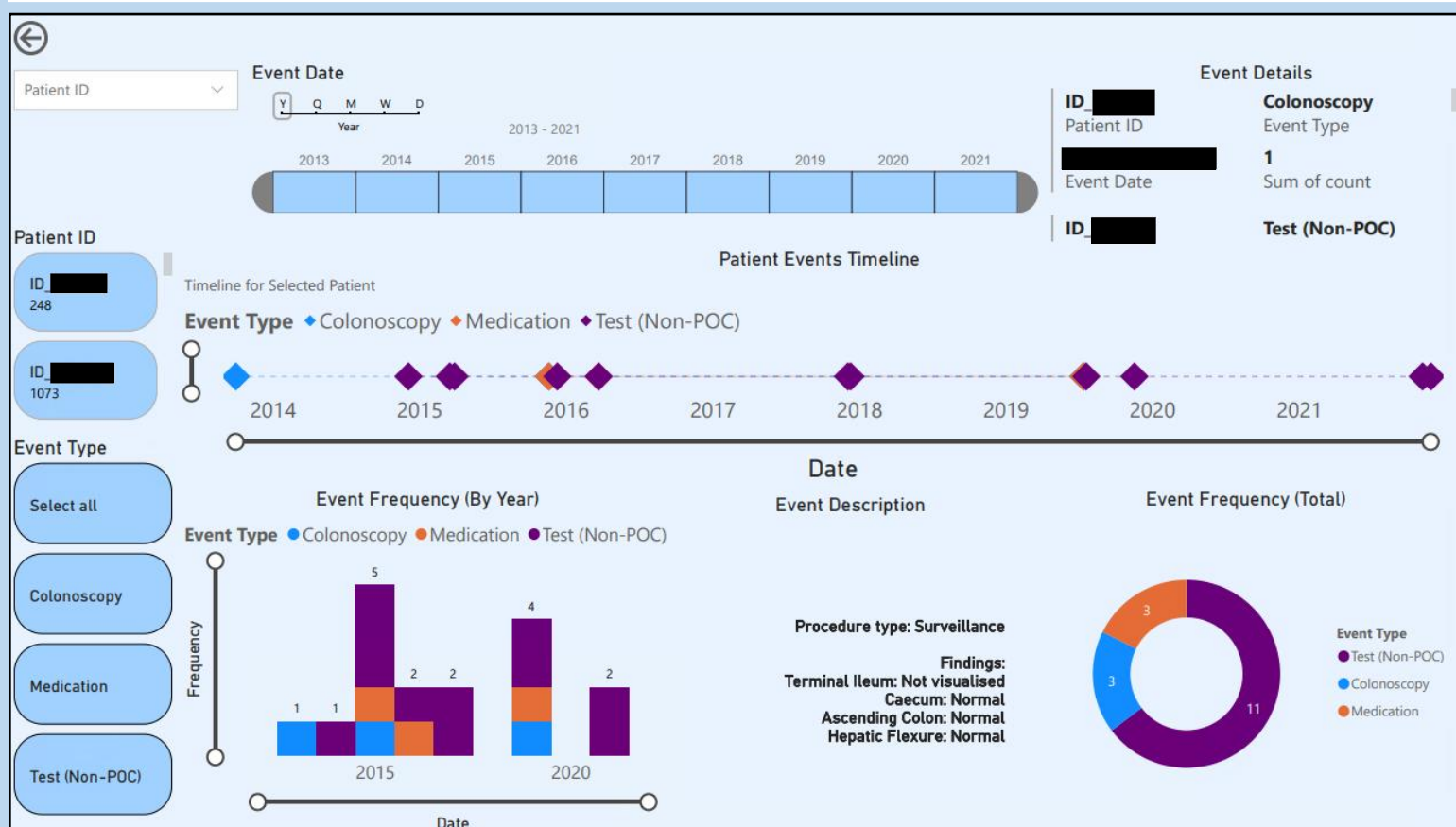


Figure 1: Single Patient Event Timeline

1) A Comprehensive Patient Event Log

A complete event log was generated as a .csv file, detailing every patient event with full descriptions.

2) An Interactive Timeline of Patient Events

Figures 1 and 2 showcase the interactive dashboard, which visualises patient trajectories. Users can filter by patient, event type, and time period for a customised view of the data and can hover over each event for more details.

3) A Model for Predicting Colonoscopy Findings

The model demonstrated a sensitivity of 78% - it correctly identified 78% of cases where either colitis or ileitis was present. However, it identified just 52% of cases where either finding was absent, indicated by its specificity score (Table 1). Figure 3 shows a summary of the most important features.

	Sensitivity (True Cases)	Specificity (False Cases)	Support
Colitis/Ileitis (True)	78%		85
No Colitis/Ileitis (False)		52%	67
Overall Accuracy			66%

Table 1: Model Performance on Testing Dataset

Conclusions

The timeline visualisations provide clinicians and researchers with a powerful tool to interactively explore patient histories, aiding in their assessment of past events and potentially uncovering patterns in patient trajectories.

Although the prediction model shows promise in assisting clinicians with decision-making, its limited accuracy highlights significant room for improvement. The model's performance suggests that while biomarker test results, and particularly immunoglobulin levels, hold predictive value, they are not sufficient alone to reliably predict the presence of colitis or ileitis.

A key limitation of this project was the restricted dataset used for model training. The predictive accuracy of the model could be strengthened by providing the model with other aspects of patients' prior medical history, such as demographic details, and by selecting only the high-impact features identified in Figure 3.

References

- Kraszewski S, Szczurek W, Szymczak J, Reguła M, Neubauer K. Machine Learning Prediction Model for Inflammatory Bowel Disease Based on Laboratory Markers. Working Model in a Discovery Cohort Study. Journal of Clinical Medicine. 2021; 10(20):4745. <https://doi.org/10.3390/jcm10204745>
- Lewis, J. D. (2011). The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. Gastroenterology, 140(6), 1817-1826.