Gradient boosted decision tree to model ustekinumab trough levels in Crohn's Disease

Adam Saleh[1], Kerri Glassner[2], Bincy P. Abraham[2]

[1]Texas A&M University, Engineering Medicine, Houston, United States
[2]Division of Gastroenterology, Department of Medicine, Houston Methodist Hospital

BACKGROUND

- Strategies for predicting ustekinumab (UST) trough levels with machine learning techniques can improve personalized care and aid in decision making for UST initiation or scheduling.

AIMS

- The aim of this study was to identify variables capable of predicting an adequate UST response through a gradient boosted decision trees (GBDT) model.

METHODS

- A retrospective cohort of Crohn's disease (CD) patients from our quaternary referral center being treated with UST were reviewed for variables including age, gender, ethnicity, BMI, dosing schedule, time passed since starting UST, previously used biologics, disease duration, age of diagnosis, disease location, disease behavior, and measurements of inflammation.
- As part of feature selection, a univariate analysis was conducted to determine which features significantly correlated with UST trough levels.
- These features were then used to train a multivariate GBDT model, which was then evaluated using a nested cross-validation framework.
- The gini importance of the features included in each model was then ranked and then averaged across the different models.
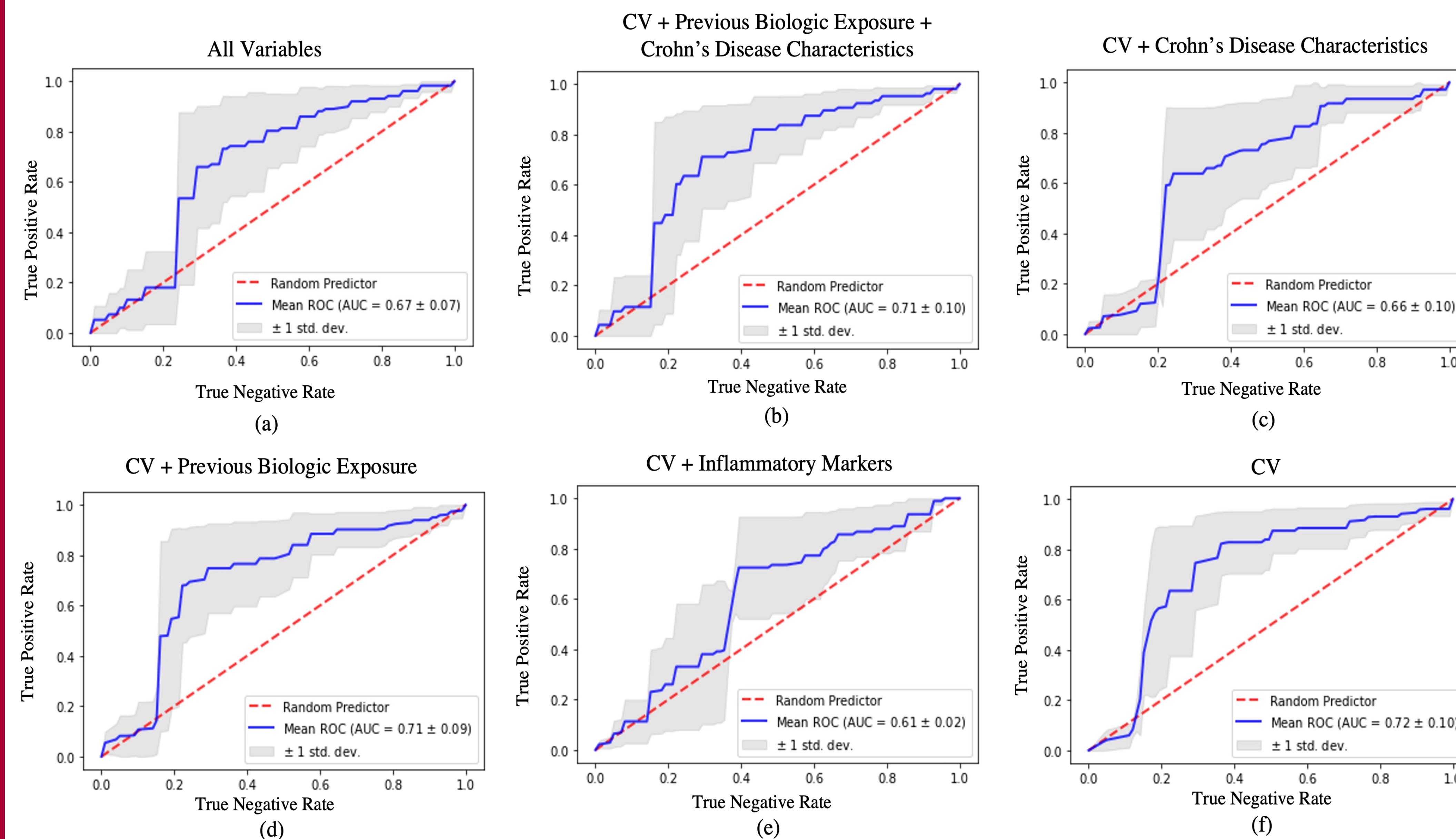
RESULTS

(a) All Variables — Mean ROC (AUC = 0.67 ± 0.07)
(b) CV + Previous Biologic Exposure + Crohn's Disease Characteristics — Mean ROC (AUC = 0.71 ± 0.10)
(c) CV + Crohn's Disease Characteristics — Mean ROC (AUC = 0.66 ± 0.10)
(d) CV + Previous Biologic Exposure — Mean ROC (AUC = 0.71 ± 0.09)
(e) CV + Inflammatory Markers — Mean ROC (AUC = 0.61 ± 0.02)
(f) CV — Mean ROC (AUC = 0.72 ± 0.10)

TABLE 1. Gini importance ranking of variables from different models

Variable	All	CV + previous biologic exposure + CD characteristics	CV + CD Characteristics	CV + previous biologic exposure	CV + Inflammatory Markers	CV	Average
Core Variables (CV)							
Gender	9	4	4	6	5	3	4
Dosing Schedule	5	2	3	3	4	2	2
Time on UST	1	1	1	1	1	1	1
Previous Biologic Exposures							
Failed Adalimumab	6	7	-	4	-	-	6
Failed Infliximab	10	6	-	2	-	-	5
Failed Certolizumab	7	8	-	5	-	-	7
CD Characteristics							
Non-stricturing, non-penetrating	12	10	6	-	-	-	11
Penetrating	11	5	5	-	-	-	10
Ileal Disease	8	9	7	-	-	-	12
Ileocolonic Disease	4	3	2	-	-	-	3
Inflammatory Markers							
CRP	2	-	-	-	2	-	8
ESR	3	-	-	-	3	-	9

RESULTS

- 155 CD patients were identified in our cohort with UST trough levels obtained.
- Univariate analysis determined the following variables to be significant positive predictors of adequate UST response: female gender, higher frequency of dosing, time on UST, and the Montreal classifications B1, B3, L1, L3.
- The following were negative predictors of adequate UST response: ESR, CRP, failed adalimumab, failed infliximab, failed certolizumab
- Results of various input variable combinations are outlined in figure 1.
- The gini importance of features in each model is included in table 1.
- Of the generated GBDT models, core variables only, and core variables with previous biologic exposure and CD characteristics were the best performing models with mean AUC of 0.72 ± 0.10 and 0.71 ± 0.09 respectively.

CONCLUSIONS

- **This proof-of-concept study demonstrates how predictive models can be used to understand variables important for UST response, and when additional doses of UST might be necessary to achieve therapeutic levels.**
- **Our proof-of-concept models seem to illustrate that the most predictive variables for UST trough levels were time passed since starting UST, dosing schedule, ileocolonic disease, and previously failed anti-tumor necrosis factor agents.**