Precision VISSTA: Machine Learning Prediction and Inference for Bring-Your-Own-Device (BYOD) mHealth Data

Tim Coleman[1], Lucas Mentch, PhD[1], Kimberly Glass, PhD[2-3], David Gotz, PhD[4], Nils Gehlenborg, PhD[3], Arlene E. Chung, MD MHA MMCi[4]
[1]University of Pittsburgh, Pittsburgh, PA, USA; [2]Brigham & Women's Hospital, Boston, MA, USA; [3]Harvard Medical School, Boston, MA, USA; [4]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Introduction

Precision VISSTA is a bring-your-own-device (BYOD) mobile health (mHealth) patient-powered research study focused on Inflammatory Bowel Diseases (IBDs). Participants report longitudinal survey data on outcomes such as disease activity along with patient-reported outcomes like as sleep disturbance, while also contributing mHealth lifestyle data from various wearable devices and apps (24 types). IBD patients have extremely heterogeneous phenotypes with symptoms that fluctuate. Prior work has suggested an association between increased self-reported physical activity and decreased disease activity[1], while self-reported sleep disturbance has been associated with increased disease activity[2]. However, the precise nature and quantity of activity and sleep associated with improved outcomes is not well established. Our mHealth dataset contains numerous features describing physical activity and a number of other lifestyle characteristics, allowing for large-scale analysis of the features most associated with IBD disease activity and symptoms. Because of the complex underlying relationships within the data, we considered a number of flexible machine learning (ML) approaches in order to avoid the rigid model structure imposed by most traditional statistical models. We leveraged recent theoretical results on inference for supervised learning ensembles to develop and implement permutation-style hypothesis tests for feature significance on these otherwise "black-box" models. The primary study objectives were: (1) to formally establish the predictive relevance of mHealth features in forming more accurate predictive models than could be obtained with survey data alone, and (2) to infer which specific mHealth lifestyle features are most predictive of outcomes for patients with IBDs.

Methods

To determine whether disease activity [Simplified Crohn's Disease Activity Index or SCDAI for Crohn's disease (CD); Simple Colitis Activity Index or SCAI for ulcerative colitis (UC)] and patient-reported outcomes (depression, anxiety, pain interference, sleep disturbance, social relationships) could be accurately predicted using mHealth lifestyle data, we constructed various machine learning models and conducted significance tests for features within this framework to account for potentially complex, nonlinear relationships. Our analysis takes data from three sources: (1) survey data on outcomes, (2) demographic information from surveys, and (3) mHealth data collected longitudinally as patients complete surveys, including device or app type. We aimed to determine whether mHealth data adds predictive power beyond what can be learned from survey data alone (data sources 1 and 2). For each outcome, we conducted a cross-validation (CV) analysis to select the optimal ML method for forecasting patient outcomes on their next survey. Models considered included best subset regression (Best Reg), elastic net, random forests (RF), conditional inference random forests (C-RF), gradient boosting (GrBoosting), partial least squares (PLS), and multivariate adaptive regression splines (MARS), with each model carefully tuned using the R package 'caret.' For each outcome, models were ranked according to their 5-fold cross-validated mean squared error (MSE) with better models attaining lower average rank in terms of generalization error. As a baseline, partial F-tests were conducted for each mHealth data type and outcome. We then conducted inference on those models identified as highly accurate in the CV analysis and tested whether a model trained on all features (i.e. demographic and prior survey data with mHealth data) attains equal generalization error to one trained using only prior survey outcomes and demographic information. Features with missing values were imputed using a RF model to iteratively predict missing values based on other complete feature information. The hypothesis tests for significance were carried out using a permutation-style procedure informed by recent work on inference with random forests.[4,5]

Results

The analytic sample had 539 inter-survey observations from 371 unique patients, of which 332 observations came from CD patients and 228 from UC patients. The cross-validation analysis consistently suggested that C-RF and elastic net models generated the most accurate predictions across outcomes. The overall tests for significance indicate that, in aggregate, mHealth data was predictive of pain interference across all three tests with more modest evidence for an effect on SCDAI disease activity (CD), social relationship, and depression scores (Table 1). The more granular tests for predictive significance of individual mHealth features suggest that time spent in moderate-to-vigorous activity (active_duration) was predictive of pain interference and disease activity for patients with either CD (SCDAI) and UC (SCAI) in the elastic net models (Figure 1). The tests on the elastic net model also suggested that distance traveled throughout the day was predictive of disease activity for UC patients and also predictive of sleep disturbance, fatigue, and depression scores across patients. While total hours of sleep (total_sleep) was only predictive of disease activity for patients with UC (SCAI), it was also predictive for depression and pain. Total steps per day was only strongly predictive for disease activity in patients with CD. Tests conducted on the conditional inference random forest model detected fewer significant results, and water consumption was the only mHealth feature consistently shown to improve predictions across outcomes. We suspect, however, that this may be an artifact of its low correlation with other features rather than an indication of strong predictive power.

Test\Outcome	Anxiety	Depression	Fatigue	Sleep	Social	SCAI	SCDAI	Pain
F Test	0.0498*	0.5428	0.0374*	0.5114	0.1571	0.3969	0.0045**	0.0017**
C-Forest Test	0.2298	0.6913	0.2138	0.6993	0.005**	0.8581	0.6124	0.001**
Elastic Net Test	0.8991	0.008**	0.4256	0.6663	0.0689	0.0889	0.2877	0.03*

Table 1. P-values for overall effect of aggregate mHealth features by outcome and test procedure. * p=0.05, **p=0.01

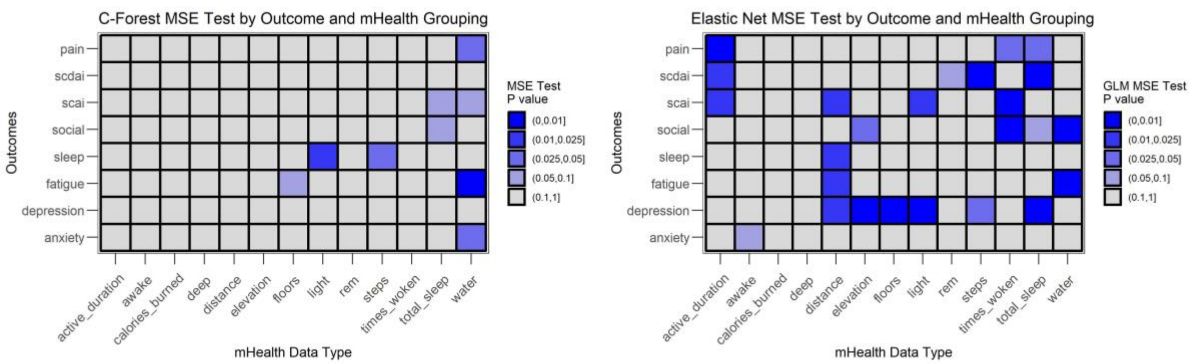

Figure 1. Conditional inference random forest and bagged elastic net results by mHealth data type and outcome.

Conclusion

Elastic Net and C-RF were identified as the top performing ML models with improved predictive accuracy when prior outcomes were included as features. C-RF models are inherently more flexible, and thus, we found fewer marginally significant predictive features since the models can better account for lost signal by detecting more complex signals from other available features. Best subset regression models performed poorly, demonstrating that traditional linear models fail to capture substantial predictive information contained within the mHealth data. Significance tests suggested that mHealth data are predictive of pain interference, even when accounting for prior patient-reported outcomes data. Moderate to vigorous physical activity (active_duration) was also highly predictive of IBD disease activity in both UC and CD patients in the elastic net models, while steps were predictive only for those with CD.

References

1. Jones PD, Kappelman MD, Martin CF, Chen W, et al. Exercise decreases risk of future active disease in patients with inflammatory bowel disease in remission. Inflammatory bowel diseases. 2015;21:1063-71.
2. Ananthakrishnan AN, Long MD, Martin CF, et al. Sleep disturbance and risk of active disease in patients with Crohn's disease and ulcerative colitis. Clinical Gastroenterology and Hepatology. 2013;11:965-71.
3. Mentch L, Hooker G. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. The Journal of Machine Learning Research. 2016 Jan 1;17(1):841-81.
4. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association. 2018 Jul 3;113(523):1228-42.
5. Coleman, T., Peng, W., & Mentch, L. (2019). Scalable and Efficient Hypothesis Testing with Random Forests. arXiv preprint arXiv:1904.07830.