Predicting Hypertension Control Status using a Propensity-Adjusted Neural Network

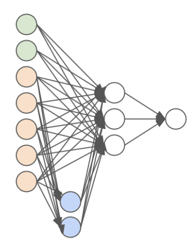
John Schrom, MPH, Tsz Kiu Liu, MPH, Allison Gilmore, PhD, Lenard Lesser, MD, MSHS, Raj Behal, MD, MPH

One Medical, San Francisco, CA

Introduction: Hypertension is a prevalent risk factor for cardiovascular disease; approximately one-third of adults in the United States have high blood pressure, with another one-third having pre-hypertension[1]. Only 54% of those with hypertension have it under control[2], leaving millions of patients unnecessarily at risk for a cardiovascular event. Meanwhile, clinicians are overburdened with desktop medicine tasks, spending over 6 hours per day in their electronic health record[3]. This leaves minimal time for chronic disease and population health management. This preliminary work assesses the feasibility of using machine learning to aid clinicians in the prioritization of their uncontrolled hypertensive patients by predicting a patient's future hypertension control status.

Methods: *Study Design.* Secondary data was retrospectively collected from the electronic health record of a large, national, primary care clinic system. Patients with diagnosed hypertension who were not in-control (defined using the 2018 HEDIS specification) at some point between 2017 and 2019 were identified. Patients were enrolled in the study on the last date that they were not in-control or, if they had never been in-control, six months prior to this analysis. Control status was recorded six months after enrollment ("outcome"). Demographics, diagnoses, and medications were collected for the one year prior to enrollment ("baseline"). New medications in the three months post-enrollment were collected ("treatment"). Medications were coded by their component active ingredients, and diagnosis data was grouped to ICD-10 categories.

Model Design. A neural network was designed to include every baseline (inset, colored orange) and treatment variable (green). To account for confounding in treatments within the population[4], propensity nodes (blue) were trained using just baseline variables to predict the probability of the patient having each treatment variable. The baseline, treatment, and propensity nodes were fed into a feed-forward multilayer perceptron (representative hidden and output layers shown in white). Network layout and hyperparameters were tuned using grid search. Networks were trained using Keras version 2.2.4-tf, data processed using Pandas version 0.20.3 and sklearn version 0.19.0. Thirty percent of data was withheld for validation, and training samples were re-weighted to account for class imbalance.



Results: In total, 11,144 patients were included in this study, with 44.6% in-control at the end of follow-up. The population had an average age of 52 (range: 21 - 87) and was 65% male. There were 775 features identified, with an additional 58 propensity nodes included. The best performing network was three layers deep with successively fewer nodes (100 nodes - 50 nodes - 25 nodes), linear kernel with sigmoid activation function, and L1 normalized with lambda 0.001. This model performed with AUC of 0.62, positive predictive value of 0.24, negative predictive value of 0.87, and F1 score of 0.64. The predicted probability of being in-control with no treatment was 34% (range: 18% - 72%).

Discussion: While the overall performance of this model stands to be improved, the negative predictive value suggests there may be utility in specifically identifying patients not likely to have their hypertension in control in six months. Further work is necessary for improving and validating this approach.

References

1. High Blood Pressure Fact Sheet. https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_bloodpressure.htm. May 10, 2019.
2. Farley TA, Dalal MA, Mostashari F, Frieden TR. Deaths preventable in the U.S. by improvements in the use of clinical preventive services. Am J Prev Med. 2010;38(6):600–9
3. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. Ann Fam Med. 2017; 15(5): 419-426.
4. Sturmer T, Wyss R, Glynn RJ, Brookhard MA. Propensity Scores for confounder adjustment when assessing the effects of medical interventions using non experimental study designs. J Internal Medicine 2014;275(6): 570-580.