A Collaboration with Johnson & Johnson: Comparing Population Level and Hospital Level Predictions of Adverse Outcomes Following Hip Fracture Surgeries

Data Science Institute
COLUMBIA UNIVERSITY

Thompson Bliss, Shreya Chaganti, Connor Daly, Zhongling Jiang, Patrick Lewis
Industry Mentors: Cindy Tong and Kade Etter

Johnson & Johnson

Rationale & Background

Hip fractures in adults over the age of 65 represent a major public health burden in the US, accounting for 72% of all fracture-related medical expenses. Hip fracture surgery frequently results in complications resulting in emergency room visits, hospital readmission, and mortality.

This project proposes and compares two modeling approaches to predict patient risk of an ER visit, hospital readmission, or mortality within a 90-day timeframe following hip fracture surgery. First, an automated machine learning algorithm is used to create different models for all hospitals across all three target variables. The results of these hospital-specific models are then compared to the results of a population model, in which a single model is produced for each of the three target variables across all individuals, regardless of hospital.

Data and Model Types

The Medicare dataset includes about 400,000 observations of patients from approximately 2000 different hospitals. The classes for all three target variables are unbalanced, as 15-20% of all observations are positive. For both approaches, feature selection and parameter tuning are automated across a variety of classification models, which include logistic regression, regularized logistic regression, XGBoost, and random forest.

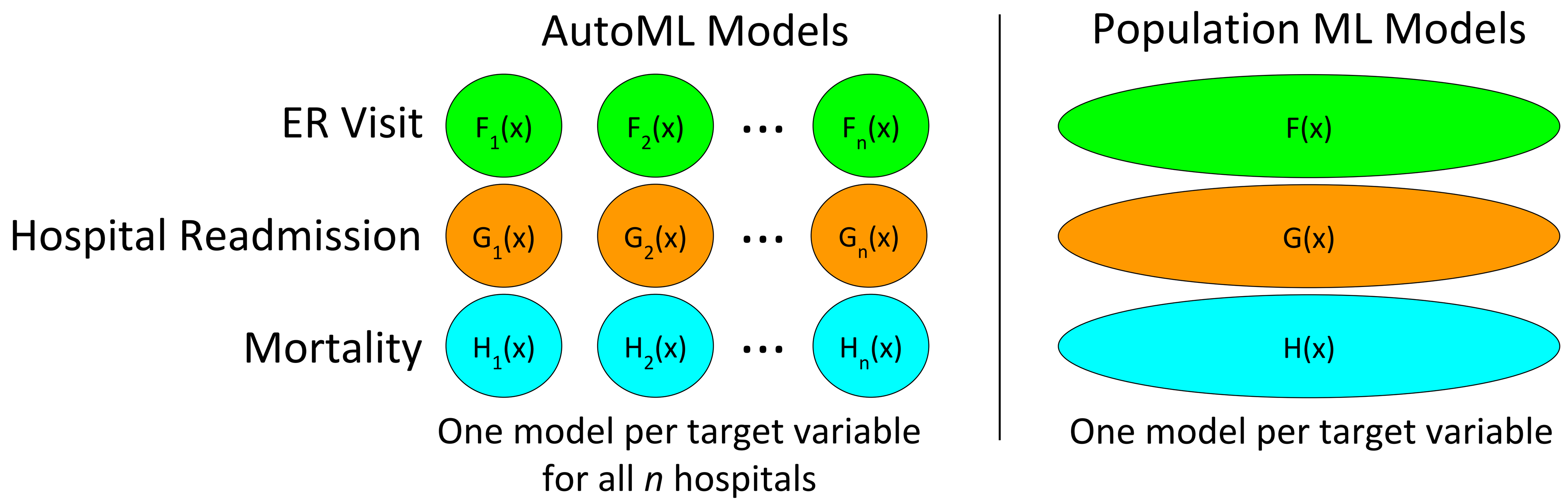


Figure 1. A visual comparison between the autoML hospital level models and the population level models.

AutoML Algorithm

The autoML algorithm iterates through a variety of steps which include feature creation, feature selection, hyperparameter tuning, and model selection to process and model data for each hospital. Due to a constraint of computation resources, only a random sample of 100 hospitals are used for this analysis, but the algorithm could trivially be extended to all hospitals for which data exists.

Population Model

The population model is created with similar steps to the autoML algorithm and serves as a baseline model to which every hospital's model can be compared. First, observations which are not from the 100 randomly chosen hospitals for autoML are filtered out, leaving approximately 35,000 observations.

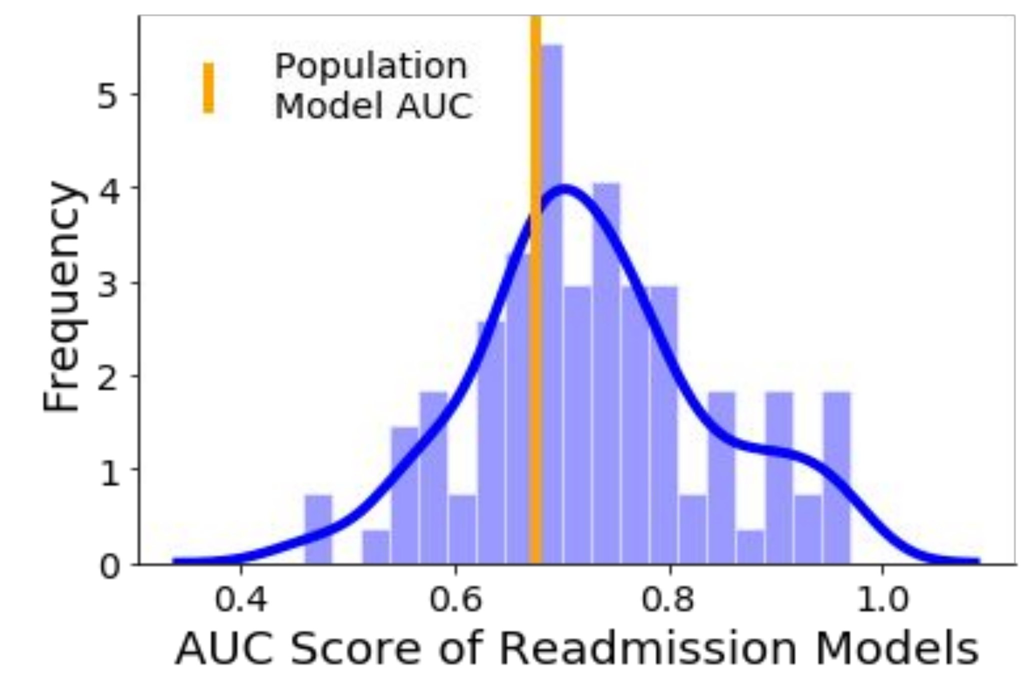
Modeling Process

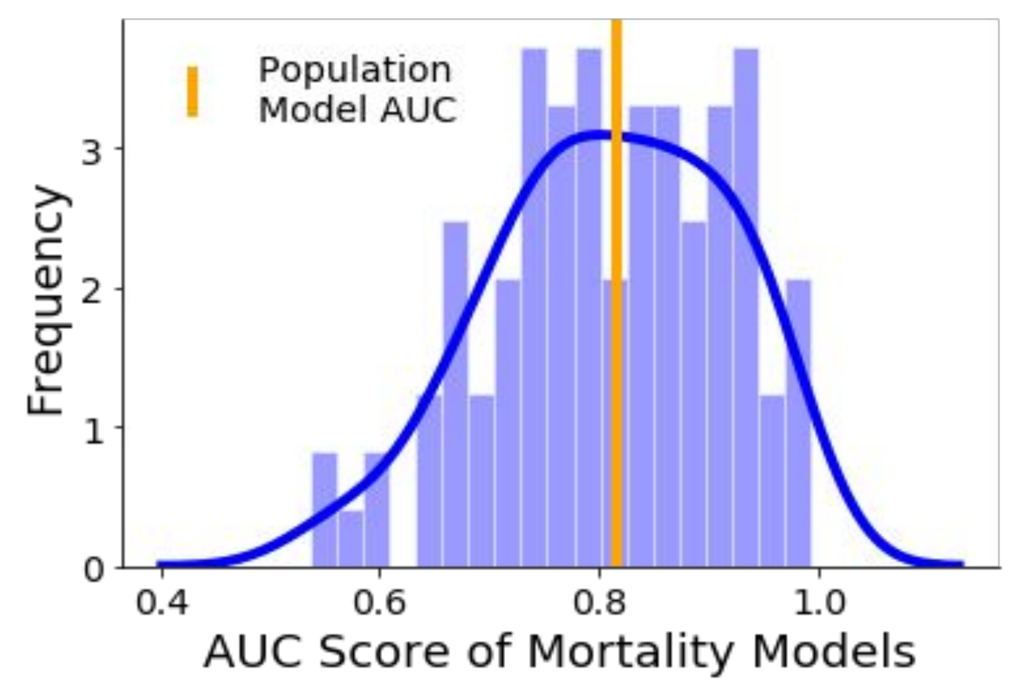
The processes for creating the population and hospital level models were very similar. The difference is that the process is applied 3 times for the population level data to fit across 3 target variables and 300 times for the hospital level data to fit across 3 variables and 100 hospitals. The process is as follows:

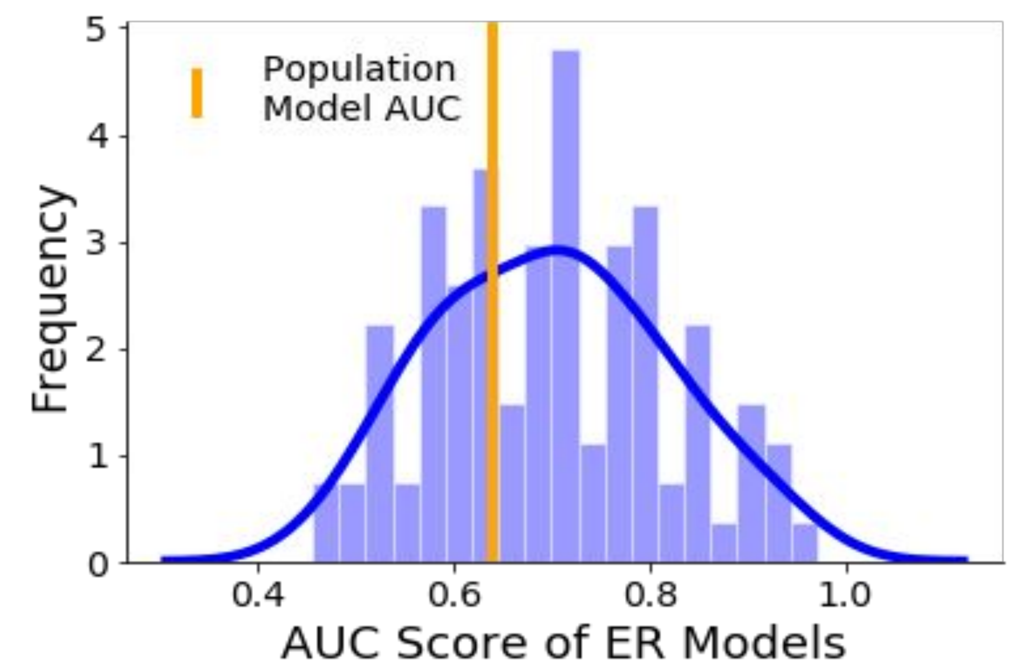
1. Remove features that do not have any variance, leak target information, contain null values, or cause the model to run out of memory.
2. Use the Boruta algorithm to facilitate feature selection.
3. Create a validation set by sampling 20% of the remaining data.
4. Use four-fold cross validation grid search on the training set to evaluate the performance of the various models and parameters.
5. Calculate the AUC, log-loss and accuracy for each model on the validation set.
6. Select the best model by comparing AUC scores across different model types.

Results and Conclusion

Outcome Variable	Percent of Hospitals with better max AUC than population model
Emergency Room Visit	65%
Hospital Readmission	70%
Mortality	49%







Figures 2-4. AUC Scores of hospital models compared to population model.

From the charts above, we see that the hospital level models generally outperform the population model in predicting ER visits and hospital readmissions. The population level model slightly outperforms the hospital level models in predicting mortality. However, it is not conclusive that either model is definitively better than the other.

Acknowledgments

Many thanks to the Johnson & Johnson (Kade Etter, Cindy Tong, Chin-Wen Chang and Neha Prabhakaran) and our faculty advisor, Adam Kelleher.