# Predicting hospital admission using machine learning

# Abstract

**Objective:**To predict hospital admission at the time of ED triage using patient history in addition to information collected at triage.

**Methods:**This retrospective study included all adult ED visits between March 2014 and July 2017 from one academic and two community that resulted in either admission or discharge. A total of 972 variables were extracted per patient visit. Samples were randomly partitioned into training (80%), validation (10%), and test (10%) sets. We trained a series of nine binary classifiers using logistic regression (LR), gradient boosting (XGBoost), and deep neural networks (DNN) on three dataset types: one using only triage information, one using only patient history, and one using the full set of variables. Next, we tested the potential benefit of additional training samples by training models on increasing fractions of our data. Lastly, variables of importance were identified using information gain as a metric to create a low-dimensional model.

**Results:**A total of 560,486 patient visits were included in the study, with an overall admission risk of 29.7%. Models trained on triage information yielded a test AUC of 0.87 for LR (95% CI 0.86-0.87), 0.87 for XGBoost (95% CI 0.87-0.88) and 0.87 for DNN (95% CI 0.87-0.88). Models trained on patient history yielded an AUC of 0.86 for LR (95% CI 0.86-0.87), 0.87 for XGBoost (95% CI 0.87-0.87) and 0.87 for DNN (95% CI 0.87-0.88). Models trained on the full set of variables yielded an AUC of 0.91 for LR (95% CI 0.91-0.91), 0.92 for XGBoost (95% CI 0.92-0.93) and 0.92 for DNN (95% CI 0.92-0.92). All algorithms reached maximum performance at 50% of the training set or less. A low-dimensional XGBoost model built on ESI level, outpatient medication counts, demographics, and hospital usage statistics yielded an AUC of 0.91 (95% CI 0.91-0.91).

**Conclusion:**Machine learning can robustly predict hospital admission using triage information and patient history. The addition of historical information improves predictive performance significantly compared to using triage information alone, highlighting the need to incorporate these variables into prediction models.