

Leveraging Machine Learning to Decrease In-Hospital Mortality Rates

Nathan Brajer*^{1,2}, Brian Cozzi*^{1,3}, Michael Gao¹, Marshall Nichols¹, Mike Revoir¹, Suresh Balu¹, Krista Whalen¹,
Joseph Futoma^{3,4}, Cara O'Brien⁵, Chetan B. Patel^{2,5}, Pooh Setji⁵, Adrian Hernandez*^{2,4}, Mark Sendak*¹

¹Duke Institute for Health Innovation; ²Duke University School of Medicine; ³Duke Department of Statistical Science; ⁴Harvard School of Engineering and Applied Sciences; ⁵Duke University Health System

Background

Each year, an average of 3% of patients admitted to Duke University Hospital die during the inpatient admission. Efforts to reduce inpatient mortality have focused on both improving the quality and safety of clinical care, and shifting the location of care when aggressive medical treatment in the hospital is no longer preferred. Early identification of patients at high risk of dying in the hospital can improve clinical and operational decision-making at the time of admission and improve outcomes for these patients. Specific opportunities identified at our institution are improving the efficiency of palliative care and support services, clinical documentation and coding, and bed assignment based on intensity of care required.

Methods.

We developed a machine learning model to estimate the risk of death during the inpatient encounter using 50 variables from our EHR (electronic health record), including patient demographics, lab results, vitals, and medication administrations that were collected before admission. We trained and tested a Gradient Boosting model using data from 43,137 hospitalizations occurring in 2014-2015, tested on additional data from 2018, and used the *CalibratedClassifierCV* in sklearn to calibrate the output so the predictions would correspond to probability values. Concordance between model and physician risk assignment was assessed via blinded review of admission notes by three physicians independently.

Results.

The area under the receiver operating characteristic curve (AUROC) was 0.86 for the 2014-2015 test data set and 0.84 for the 2018 test data set.

The Hosmer-Lemeshow test indicated adequate calibration ($p = 0.07$).

Table 1 shows concordance between physicians and model for assigning high (top 20%) vs. low (bottom 80%) risk for patients sampled from recent admissions (raters = 3, $n = 20$). Inter-physician rater variability was assessed with Fleiss' kappa ($\kappa = 0.325$, indicating "fair agreement"). Physicians agreed with 86% of patients that the model rated as high risk.

The model agreed with 89% of patients that the physicians rated as low risk. For the twenty sample patients, of which two had death outcomes, physician sensitivity and specificity were 100% and 28%, respectively, compared to 50% and 83% for the model. These results suggest the model output can be used to both ensure prioritization of high risk patients, and more efficiently prioritize empirically low-risk patients whose risk may be overestimated by physicians.

Clinical Integration.

The model was implemented on data extracted on a daily basis from Duke's EHR and monitored for two months. Initial workflows were designed, informed by the results above, to 1) ensure patients already enrolled in hospice are either re-routed to hospice from the emergency department or admitted under General Inpatient Care (GIP) status, 2) flag high risk patients for real-time coding review to increase accuracy of present on admission (POA) diagnoses, and 3) guide decision making around assigning a patient an ICU, intermediate, or floor bed and level of care. Threshold analysis was conducted to support clinical leaders in picking a threshold that optimally balances model sensitivity and positive predictive value for each workflow based on tolerance for false positives and operational resources available. A user interface, built as a simple patient list using Apache Superset, will be used to facilitate all initial workflows.

Conclusion.

We developed a high-performing machine learning model to predict in-hospital death at the time of admission. We have deployed this model on EHR data, performed a preliminary clinical validation, prototyped a user interface, and designed three clinical workflows targeting concrete operational decisions.

*these authors contributed equally to this work

Table 1: Physician/Model Concordance

Agreement between model and physicians in assigning patients into a high vs. low risk category		Low risk (model)	High risk (model)
		n = 15	n = 5
Low risk (physician)	avg. n = 5.33	4.67	0.67
High risk (physician)	avg. n = 14.67	10.33	4.33