Abstract

Prospective Real-World Performance Evaluation of a Machine Learning Algorithm to Predict 30-Day Readmissions in Patients with Heart Failure Using Electronic Medical Record Data

Sujay S Kakarmath[1,2,3*], MBBS, MS; Neda Derakhshani[1*], MSc; Sara B Golas[1*], MA; Jennifer Felsted[1,3*], PhD; Takuma Shibahara[4*], PhD; Hideo Aoki[4*]; Mika Takata[5*]; Ken Naono[4*], PhD; Joseph Kvedar[1,2,3*], MD; Kamal Jethwani[1,2,3*], MD, MPH; Stephen Agboola[1,2,3*], MD, MPH

[1]Connected Health Innovation, Partners HealthCare, Boston, MA, United States

[2]Massachusetts General Hospital, Boston, MA, United States

[3]Harvard Medical School, Boston, MA, United States

[4]Research and Development Group, Hitachi, Ltd, Tokyo, Japan

[5]Big Data Laboratory, Hitachi America Ltd, Santa Clara, CA, United States

[*]all authors contributed equally

Corresponding Author:
Neda Derakhshani, MSc
Connected Health Innovation
Partners HealthCare
Boston, MA,
United States
Email: snderakhshani@partners.org

Abstract

Background: Heart failure (HF) patients have a high readmission rate with approximately 20% of patients being readmitted within 30-days after discharge. Hospital interventions to reduce HF readmissions are resource- and effort-intensive. Widespread availability of electronic medical record data has spurred interest in using machine learning-based techniques for risk stratification of heart failure patients. The predictive performance of machine learning-based predictive models is often evaluated solely using the Area Under the Receiver Operating Characteristic (AUROC) curve. However, the AUROC is independent of prevalence therefore predictive models with the same AUROC can have differential clinical utility. Furthermore, the AUROC does not provide any insight about the presence of overfitting or decay in predictive performance of a model over time, both of which can affect its real-world performance.

Objective: Our primary objective is to assess real-world performance of a 30-day readmission risk prediction model for HF patients, which had an AUROC of 0.71 in the training dataset.

Methods: Predictions for risk of 30-day readmissions in HF patients in the Partners Healthcare System were prospectively obtained from the model. We assessed the positive (PPV) and negative predictive value (NPV), in addition to sensitivity, specificity, accuracy, model calibration and Brier score.

Results: Four hundred twenty index admissions that were not part of the training dataset were included in this prospective evaluation. Readmission rate was 24% (101 30-day readmissions). The AUROC of the predictive model was 0.57. At a discrimination threshold of 0.2 for flagging high-risk index admissions, the sensitivity and specificity of the model were 53.46% and 63.32%, respectively. The PPV and NPV were 31.57% and 81.12%, respectively. The Brier score was 0.19.

Conclusions: Our analysis offers important insights about the real-world performance of this predictive model. The NPV suggests that the model's prediction about patients at low risk for readmission are reliable. This insight can be useful in optimizing resource allocation for patients with heart failure.

(iproc 2018;4(2):e11897) doi: 10.2196/11897

KEYWORDS

accuracy; machine learning; positive predictive value; validation

Edited by T Hale; this is a non–peer-reviewed article. Submitted 09.08.18; accepted 29.08.18; published 17.09.18.

Please cite as:
Kakarmath SS, Derakhshani N, Golas SB, Felsted J, Shibahara T, Aoki H, Takata M, Naono K, Kvedar J, Jethwani K, Agboola S Prospective Real-World Performance Evaluation of a Machine Learning Algorithm to Predict 30-Day Readmissions in Patients with Heart Failure Using Electronic Medical Record Data
iproc 2018;4(2):e11897
URL: http://www.iproc.org/2018/2/e11897/
doi: 10.2196/11897
PMID: