Predicting Next 7-day Discharges of Hospitalized COVID-19 Patients Using Ensemble Learning

Xiangyang Ye[1], Jincheng Shen[2], Matthew Samore[1], Mingyuan Zhang[3], Julia Bohman[1], Jeffrey Humpherys[1], Yue Zhang[1]

1. University of Utah Department of Epidemiology, Salt Lake City, UT; 2. University of Utah Department of Health Sciences, Salt Lake City, UT; 3. University of Utah School of Medicine, Salt Lake City, UT

HEALTH UNIVERSITY OF UTAH

Introduction

- COVID-19 pandemic, especially during resurgences of cases in hard-hit areas, led to significant shortage of hospital beds. Such shortages may be alleviated through timely and effective forecasting of hospital discharges.

Objectives

- The objective of this study is to predict next 7-day discharges of hospitalized COVID-19 patients using daily-based electronic health records (EHR) data.

Methods

- Using EHR data of hospitalized COVID-19 patients from 03/2020-12/2021, we employed ensemble learning to predict next 7-day discharges of individual patients. We used both baseline and daily inpatient features for model training, validation, and test. Baseline features include demographic and clinical characteristics, and comorbidities. The daily inpatient features were vital signs, laboratory tests, medications administered, acute physiological scores, use of ventilator, and use of intensive care unit.
- 2308 hospitalized patients were identified (14,644 hospital days). Samples were randomly split at patient level (5:2:1:2) into training set (N=1,153), validation set (N=462), weights determination set (N=231), and test/holdout set (N=462).
- We conducted the model training on the samples of admission day and the samples of days after admission day, respectively

Table 1. Samples for Model Training

	Patient Number	% Total samples	Hospital Days
Training set	1,153	50	7,463
Validation set	462	20	2,691
Weights determination set	231	10	1,582
Test set (holdout)	462	20	2,908
Overall	2,308	100	14,644

- Prediction models were trained on the training set and the validation set.
- We used weighted average of the predictions by the individual models as our ensemble approach**.** Weights were calculated based on the Area under the ROC Curve (AUC) of the models of admission and after admission, respectively.
- The predictions were based on the ensemble learning from decision tree, XGBoost, logistic regression, and multilayer perceptron (MLP), long short-term memory (LSTM), bi-directional LSTM (Bi-LSTM), and convolutional neural network (CNN).

Methods

Figure 1. Data Process Pipeline

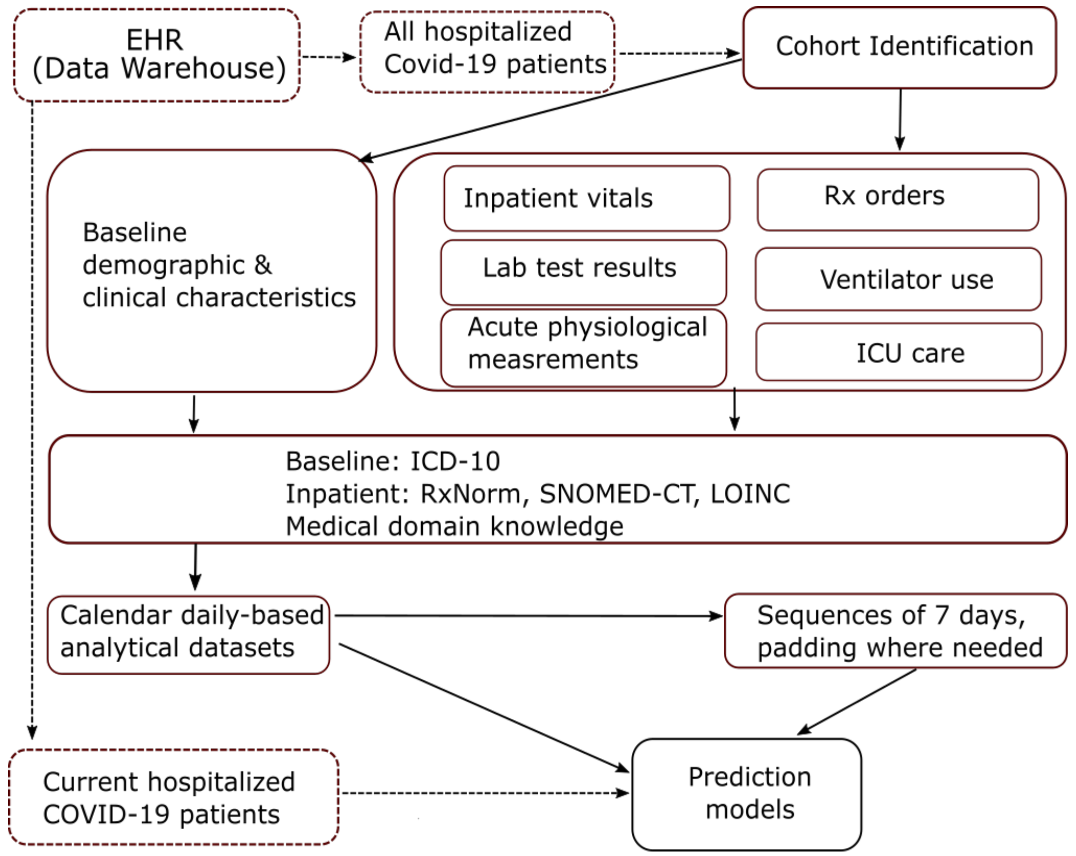
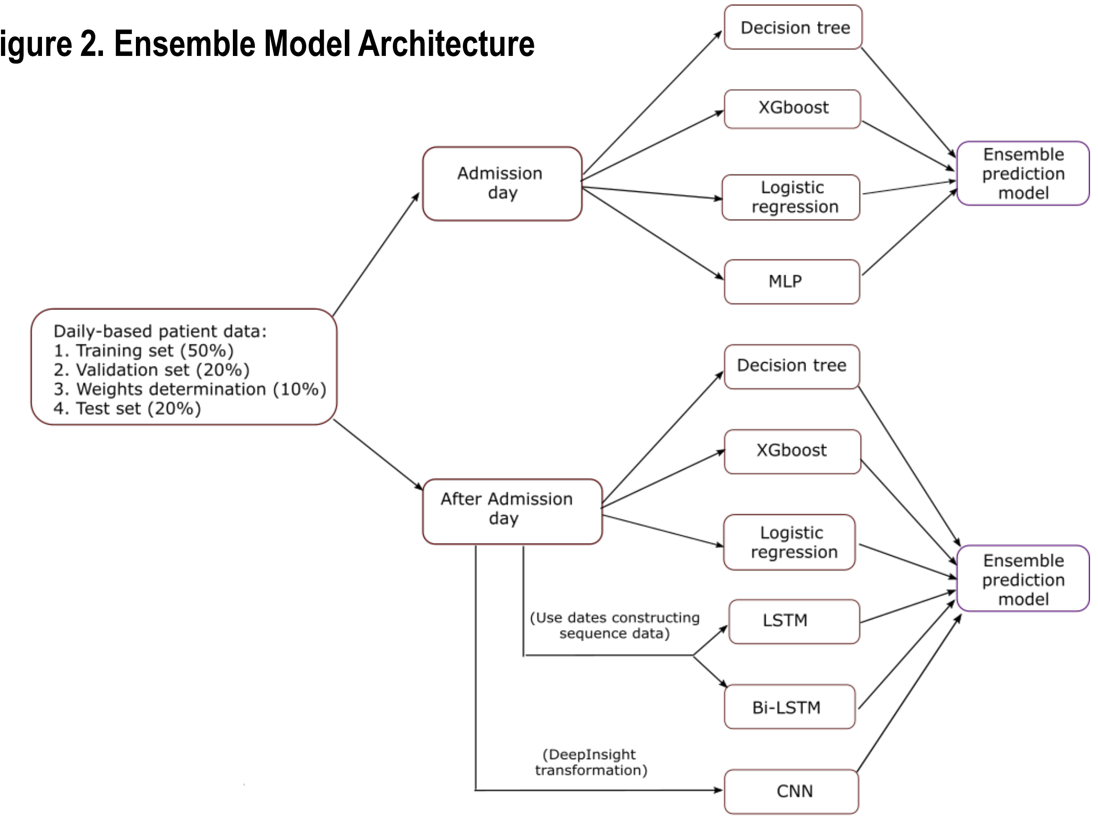


Figure 2. Ensemble Model Architecture



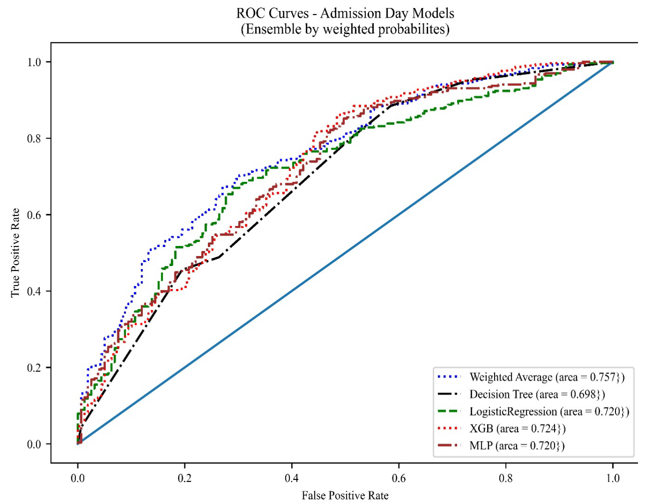
Presented at the IDWeek, Washington DC, October 19-23, 2022

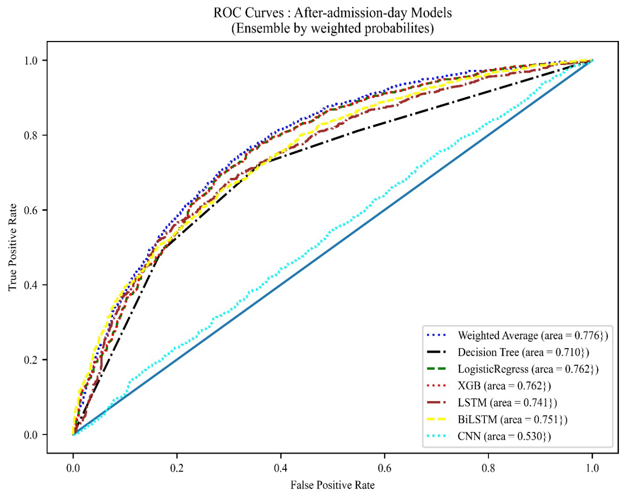
Results

- The overall average hospital length of stay was 9.6 (SD=10.8) days.
- The ensemble learning accuracy for admission-day samples was 0.729, and the F1-score for was 0.810.
- The ensemble learning accuracy for after-admission-day samples was 0.720, and the F1-score was 0.782.

Table 2. Performance Metrics of Final Ensemble Models

Model	Accuracy	Precision	Recall	F1-score	ROC AUC
Admission day					
Training	0.827	0.808	0.956	0.876	0.908
Validation	0.742	0.797	0.870	0.832	0.659
Test	0.729	0.751	0.878	0.810	0.757
After admission					
Training	0.772	0.743	0.915	0.820	0.871
Validation	0.716	0.728	0.889	0.800	0.743
Test	0.720	0.705	0.877	0.782	0.776





Conclusion

- EHR data of hospitalized COVID-19 patients can be used to predict next 7-day hospital discharges. Additional inpatient features and more advanced machine learning techniques are needed for prediction accuracy improvement.

Acknowledgements

The study was funded by CDC, Modeling Infectious Diseases in Healthcare Network (MInD - Healthcare)