1

Applying Deep Learning in Heart Failure: Hospital Readmission is Not Like Other Health Quality Metrics

Short Title: Predicting Readmission in Heart Failure

Authors: Hailey M. Shepherd, MD¹, Jeffrey T. Heaton, PhD², Theodore Marghitu, BA³, Yun Z.

Bai, MD¹, Melanie P. Subramanian, MD, MPHS¹, Sophia H. Roberts, MD¹, Martha M.O.

McGilvray, MD¹, Amit A. Pawale, MD¹, Gregory A. Ewald, MD⁴, Brian P. Cupps, PhD¹,

Michael K. Pasque, MD¹, Randi E. Foraker, PhD⁵

¹Division of Cardiothoracic Surgery, Department of Surgery, School of Medicine, Washington

University in St. Louis, St. Louis, Missouri

²Sever Institute, McKelvey School of Engineering, Washington University in St. Louis, St.

Louis, Missouri

³School of Medicine, Washington University in St. Louis, St. Louis, Missouri

⁴Division of Cardiovascular Medicine, Department of Medicine, School of Medicine,

Washington University in St. Louis, St. Louis, Missouri

⁵Institute for Informatics, Division of General Medical Sciences, School of Medicine,

Washington University in St. Louis, St. Louis, Missouri

Corresponding Author:

Hailey M. Shepherd, MD Washington University School of Medicine Campus Box 8109, 660 South Euclid Avenue St. Louis, MO 63110 Phone: (314) 482-0458 Fax: (314) 367-8459 Email: <u>hailey.shepherd@wustl.edu</u>

Manuscript Word Count: 2,349

2

Abstract

Background

Early identification of heart failure patients at increased risk for near-term adverse outcomes would assist clinicians in efficient resource allocation and improved care. Deep learning can improve identification of these patients.

Methods

This retrospective study examined adult heart failure patients admitted to a tertiary care institution between January 2009 and December 2018. A deep learning model was constructed with a dense input layer, three long short-term memory (LSTM) layers, and a dense hidden layer to cohesively extract features from time-series and non-time-series EHR data. Primary outcomes were all-cause hospital readmission or death within 30 days after hospital discharge.

Results

Among a final subset of 49,675 heart failure patients, we identified 171,563 hospital admissions described by 330 million EHR data points. There were 22,111 (13%) admissions followed by adverse 30-day outcomes, including 19,122 readmissions (87%) and mortality in 3,330 patients (15%). Our final deep learning model achieved an area under the receiver-operator characteristic curve (AUC) of 0.613 and precision-recall (PR) AUC of 0.38.

Conclusions

This EHR-based deep learning model developed from a decade of heart failure care achieved marginal clinical accuracy in predicting very early hospital readmission or death despite previous accurate prediction of 1-year mortality in this large study cohort. These findings suggest that

factors unavailable in standard EHR data play pivotal roles in influencing early hospital

readmission.

4

Abbreviations:

- AUC = Area under the receiver operating characteristic curve
- BMI = Body mass index
- ECMO = Extracorporeal membrane oxygenation
- EHR = Electronic health record
- ICD = International Classification of Diseases
- LOS = Length of stay
- LSTM = Long short-term memory
- PR AUC = Precision-recall area under the curve
- ReLU = Rectified linear activation unit function

Clinical Perspective

What is new? We developed an EHR-based deep learning model trained by 330 million data points from one of the largest cohorts of heart failure patients to date. Despite this model's highly accurate prediction of long-term outcomes, such as mortality and disease progression, our findings suggest that EHR data alone offers limited predictive power for predicting the shortterm outcomes of 30-day hospital readmission or death.

What are the clinical implications? Our study supports the notion that hospital readmission, in contrast to other health outcomes, is uniquely driven by additional factors beyond traditional EHR variables. Once identified, incorporation of these determinants into future deep learning models could allow for accurate heart failure risk-stratification at hospital discharge to facilitate more efficient allocation of limited resources to the most vulnerable patients.

Introduction

Heart failure remains a significant cause of morbidity and mortality in the United States with an annual mortality rate ranging from 22-30% (1-5). Heart failure patients represent the most rapidly growing and resource-intensive subset among patients with cardiovascular disease (6,7). Additionally, patients with heart failure have among the highest rates of unplanned 30-day hospital readmissions, ranging from 20-56% (5-7), many of which may be preventable (5-11). To this end, there has been increased emphasis on identifying heart failure patients who are at increased risk for adverse outcomes early after hospital discharge.

Deep learning models employ automated learning techniques to identify outcomepredictive data patterns in large datasets (12,13). Prior investigations using machine learning

6

models to predict hospital readmission in heart failure patients have reported inconsistent results (14-31). This may be attributed to difficulty in obtaining consistent, high-quality granular data for this heterogeneous patient group, or simply due to a lack of predictive signal in standard clinical metrics. Many other potential predictors, such as medication adherence and socioeconomic status, are not explicitly captured in EHR data but have been identified as critical determinants of care metrics including hospital readmission (32).

Although prior studies have not identified any single EHR variable capable of accurately predicting heart failure readmission with the precision required to drive clinical decision-making, predictive associations between groups of EHR variables and heart failure outcomes have been reported (17,24,32,33). As an alternative strategy to traditional regression analysis models, this investigation employs the pattern-recognition capabilities of deep learning to exploit the outcome-predictive capabilities associated with often complex and non-intuitive groupings of variables found in the medical records of heart failure patients.

EHR data platforms present patient-specific clinical variables in a uniquely structured format that may simplify and accelerate direct integration into machine learning models, facilitating application of their advanced analytic capabilities in risk prediction. These attributes give EHR-based deep learning models considerable potential for direct clinical utility in the identification of heart failure patients at high-risk for readmission or mortality. The accurate identification of these high-risk patients may improve distribution of available resources to the most vulnerable heart failure patients.

Methods

7

Data Source

The human studies *Institutional Review Board* at Washington University School of Medicine approved this study. All machine learning clinical variables analyzed in this study were obtained from the EHR of patients at Barnes-Jewish Hospital at Washington University Medical Center in St. Louis, Missouri.

Cohort and Study Design

The study population consisted of adult patients between the ages of 18 to 90 years old with any heart failure International Classification of Diseases (ICD) codes admitted between January 2009 and December 2018. Admissions within 30 days before the conclusion of data collection or a length of stay (LOS) under 48 hours were excluded. Admissions with no discharge date or ending in death were excluded as *index* hospitalizations, although not as subsequent *readmissions*. Finally, at least 5 unstructured records (e.g., diagnosis, treatment, procedures, etc.), 1 vital sign record, and 1 laboratory value were required to include a recorded medical encounter as a hospitalization. These criteria helped exclude brief admissions scheduled for diagnostic purposes, such as cardiac catheterization and transesophageal echocardiography, which are often ordered on an outpatient basis but performed in the hospital due to their invasive nature.

The unit of analysis used to train our machine learning model was hospital admission, in contrast to other studies that have performed analysis at a patient (29) or unspecified (27) level. In patients with serial hospitalizations, each admission may qualify as a unique index hospitalization as well as a potential readmission if within 30 days after a prior discharge. By analyzing each instance as a unique opportunity for learning, data utilized for model training can be maximized. Moreover, this approach more accurately represents clinical practice and

8

enhances prediction applicability, since heart failure clinicians are routinely tasked with evaluating the risk of readmission at the end of each hospitalization.

Our primary positive outcomes were all-cause readmission lasting \geq 48 hours or death within 30 days after discharge following an index hospitalization. These were classified as "positive events". The 30-day window after hospital discharge has been identified as a key health policy metric for health care quality improvement (34,35). Readmissions within this time frame often reflect poor patient risk evaluation and inadequate recognition of post-discharge needs (36). Ventricular assist device placement, extracorporeal membrane oxygenation (ECMO) support, or cardiac transplantation are not considered separately since all are captured as hospitalization events. Model inputs included EHR data prior to index hospitalization discharge date, including all EHR data obtained during prior and current hospitalizations and in the outpatient setting.

Feature Extraction and Deep Learning Network Design

The deep learning model in **Figure 1** is comprised of three LSTM layers to extract features from time-series EHR data, including vital signs, medication, laboratory, assessment, procedural, and diagnostic coding variables (37). Additionally, a dense layer extracts non-timeseries patient attributes, such as age, gender, race; and a dense hidden layer which extracts additional features from the output features generated by the three LSTM input layers and the dense input layer. The intermediate features are extracted and pooled to a final hidden dense layer containing 16 neurons and a bias neuron. The model employs the rectified linear activation unit function (ReLU) for all dense hidden layers, and sigmoid activation for the final output neuron.

9

Most patients had available data for a subset of the 70 laboratory values evaluated (**Table 1**). These values are encoded by a 70-length vector which initially starts as the mean data set value for each laboratory test. As new results are obtained over time, the previous value in the vector is replaced and remains in place until new testing is performed. Therefore, the vector contains the most recent test results or the mean value for the dataset if no results are available for a specific patient. Vectors are entered as input sequences for the LSTM layers in a sliding window fashion with the oldest vectors sliding off once they reach the end of the sequence. Vital signs are similarly encoded using a 7-length vector whose values are updated as more recent vital signs are recorded in the EHR.

The coding variables for basic procedures, surgical procedures, diagnostics, assessments, and medications were transformed into 100-length sequences of 32-dimensional vectors using Python's Genism Word2Vec (38). These sequences were then entered into an LSTM layer. Since the data were high-cardinality and sparse, no missing data management was required. The encoding process only considered present values and ignored those that were not present in the time series.

The data were randomly divided into three sets: training (80%), validation (10%), and testing (10%). The model's performance was evaluated using AUC, PR AUC, log loss, and Brier score. Feature importance was determined using the permutation feature importance method, where each feature in the trained neural network was permuted and the change in AUC analyzed. Additionally, multiple features were permuted together to analyze grouped feature importance.

Results

Heart Failure Dataset Baseline Characteristics

Our initial patient cohort comprised 52,806 heart failure patients. After the previously described exclusions, our final study set consisted of 49,675 patients, of whom, 20% (9,975) experienced positive events during the 10-year study period. Of those who experienced positive events, 84% (8,419) experienced 30-day readmissions and 28% (2,751) suffered 30-day all-cause mortality. The baseline demographic characteristics and vital signs of our training and internal validations cohorts are listed in **Table 2**.

Heart Failure Dataset 30-day Adverse Events

Of 171,563 study hospitalizations, 22,111 (13%) met our primary positive outcome (**Table 3**). Of these positive events, 19,122 (11%) involved 30-day all-cause readmission and 3,330 (2%) involved 30-day all-cause mortality. Some patients experienced both mortality and re-admission within 30 days (**Table 3**).

Heart Failure Dataset Hospital LOS Data

The overall average LOS for the index hospitalizations was 122 hours. The average LOS of positive versus negative observations was 148 hours versus 133 hours, respectively.

Heart Failure Dataset Variable Importance

Table 1 shows the importance of the 70 variables evaluated for individual impact on 30day all-cause readmission and mortality. Patient height, C-reactive protein level, and unstructured row count (which included variables such as diagnosis, treatment, procedures, etc.) demonstrated the highest importance at a univariate level. **Figure 2** demonstrates the relative predictive importance of vital sign and laboratory test variables depicted on the x and y axes,

11

respectively. Variables are listed in descending order of univariate importance. Red hues represent those with highest impact on model prediction (darker red hues representing factors with the most important impact) and blue hues highlighting low impact variables (dark blue hues representing the least important impact). Brain natriuretic peptide levels and serum sodium levels exhibited the highest impact among the clinical variables evaluated (vertical gradient), and there was no apparent association between clinical variable importance and vital sign values (horizontal gradient).

Model Performance

The performance of the model for predicting 30-day readmission or death in heart failure patients was measured using the ROC curve with a test set ROC AUC of 0.613 (95% CI: 0.59-0.64) (**Figure 3**). **Table 4** shows additional performance metrics, including PR AUC of 0.38 (CI: 0.33-0.41) shown in **Figure 4**. The model's predictive performance remained stable across a range of hospitalization durations, including hospitalizations with a brief LOS (**Table 1**). The log loss was 0.57 with a Brier Score of 0.19

Discussion

This deep learning model predicting 30-day all-cause hospital readmission or death was developed using a decade of EHR data from a tertiary care hospital and included 330 million data points from 171,563 hospital admissions in 49,675 heart failure patients. The model employed time-sensitive LSTM networks to analyze subtle fluctuations in clinical EHR variables over time and utilized densely connected networks for fixed, non-time-dependent demographics. Notably, we previously demonstrated that a nearly identical model successfully generated

12

accurate predictions of long-term adverse outcomes in the same cohort of heart failure patients, achieving an AUC of 0.91 (39). Our current findings indicate that EHR data alone is insufficient to generate similar predictive accuracy of 30-day hospital readmission or death in heart failure patients. This study underscores the unique nature of hospital readmission in comparison to other health quality metrics and the need to identify additional factors beyond traditional EHR variables that influence preventable hospital readmissions in this patient population.

Our study supports a growing body of literature heightening awareness of socioeconomic factors (such as income, education, place of residence, disability, ethnicity, race) as pivotal determinants contributing to hospital readmission in heart failure patients (40,41). While it is unlikely that targeting individual socioeconomic factors alone will be enough to reduce readmissions (41), the predictive model we seek likely requires cohesive integration of a cumulative burden of socioeconomic determinants in combination with key clinical variables that can be extracted from the EHR. This task is one for which machine learning is uniquely suited, given its unique proficiency at identifying non-intuitive, multiparametric outcome predictive patterns, and its seemingly boundless capability for comprehensive data analysis and integration (14,15,17,19,22-24). Additionally, incorporating time-based metrics through recurrent neural network units allows for the detection of even subtle progression in heart failure status, which can uncover time-dependent relationships between variables to enhance prediction and accelerate learning. These results suggest only modest capability of time-sensitive neural networks for predicting post-discharge adverse outcomes within the framework of each individual patient's unique heart failure course.

Prior research utilizing machine learning to predict hospital readmission in heart failure patients has yielded varying degrees of accuracy, with predictive AUC values ranging from 0.62

13

to 0.96 (14,15,17,19,22-24,26-29). Our final deep learning model was developed in one of the largest cohorts of heart failure patients to date and lends support to previous model results at the lower end of that spectrum of predictive accuracy. Although our model AUC does not attain the high level of accuracy achieved by Lv et al in a smaller cohort of 13,600 patients (17), our results are comparable to those reported by larger models and may accurately represent current potential of large scale implementation of patient-specific outcome prediction using an embedded EHR model.

By excluding hospitalizations with $LOS \le 48$ hours as index admissions or readmissions, our model attempts to exclude routine follow-up admissions, such as those related to elective, regularly scheduled endomyocardial biopsy and other surveillance investigations, many of which are unique to this complex patient subpopulation. This type of readmission may have contributed to overestimation of readmission for actual heart failure decompensation in previously published investigations (14,24,28). Instead, our model focuses upon longer hospitalizations, such as those for severe heart failure decompensation.

Several limitations must be considered when interpreting the findings of this study. Its retrospective design and the use of data from a single institution may restrict the generalizability of the model. Additionally, retrograde leakage of data that may falsely improve predictive accuracy is always a possibility (42-44). Our definition of a positive event as all-cause readmission or mortality may thereby include a small proportion of patients for whom 30-day readmission or death was not related to heart failure. Although not ideal, this primary outcome definition is the most effective and clinically applicable in our large population of medically complex heart failure patients. The inclusion of multiple hospitalizations from each patient in predictive model development may potentially skew results to the subpopulation with heightened

14

disease severity and worse overall health status. However, multiple readmissions are common in heart failure patients and their impact should be represented in predictive models. Clinical models assist but should not substitute clinical judgement. Additional validation is essential to verify the applicability of this model in other heart failure patient populations.

Our results suggest that EHR data alone offers limited predictive power for certain care quality metrics, such as early hospital readmission or death, and support the notion that other factors not routinely recorded in hospital medical records may play a pivotal role in determining these outcomes. To this end, further investigation is necessary to determine which of these nonstandard metrics are most important in this patient population. Once identified, incorporation of these variables into future deep learning models could allow for accurate heart failure riskstratification at hospital discharge to facilitate more efficient allocation of limited resources to the most vulnerable patients.

15

References

- Husaini BA, Mensah GA, Sawyer D et al. Race, sex, and age differences in heart failure-related hospitalizations in a southern state: implications for prevention. Circ Heart Fail 2011;4:161-9.
- 2. Johansen H, Strauss B, Arnold JM, Moe G, Liu P. On the rise: The current and projected future burden of congestive heart failure hospitalization in Canada. Can J Cardiol 2003;19:430-5.
- McCullough PA, Philbin EF, Spertus JA et al. Confirmation of a heart failure epidemic: findings from the Resource Utilization Among Congestive Heart Failure (REACH) study. J Am Coll Cardiol 2002;39:60-9.
- 4. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, 2019.
- Chen J, Normand SL, Wang Y, Krumholz HM. National and regional trends in heart failure hospitalization and mortality rates for Medicare beneficiaries, 1998-2008. JAMA 2011;306:1669-78.
- Heidenreich PA, Albert NM, Allen LA et al. Forecasting the impact of heart failure in the United
 States: a policy statement from the American Heart Association. Circ Heart Fail 2013;6:606-19.
- Virani SS, Alonso A, Aparicio HJ et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation 2021;143:e254-e743.
- Yancy CW, Jessup M, Bozkurt B et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 2013;128:1810-52.
- 9. Krumholz HM, Merrill AR, Schone EM et al. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes 2009;2:407-13.

- Khan MS, Sreenivasan J, Lateef N et al. Trends in 30- and 90-Day Readmission Rates for Heart Failure. Circ Heart Fail 2021;14:e008335.
- Unplanned Hospital Visits National. 7/27/22 ed: The Centers for Medicare & Medicaid Services,
 2022.
- 12. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol 2020;9:14.
- 13. Deo RC. Machine Learning in Medicine. Circulation 2015;132:1920-30.
- 14. Sharma V, Kulkarni V, McAlister F et al. Predicting 30-Day Readmissions in Patients With Heart Failure Using Administrative Data: A Machine Learning Approach. J Card Fail 2022;28:710-722.
- 15. Wang Z, Chen X, Tan X et al. Using Deep Learning to Identify High-Risk Patients with Heart Failure with Reduced Ejection Fraction. J Health Econ Outcomes Res 2021;8:6-13.
- 16. Sarijaloo F, Park J, Zhong X, Wokhlu A. Predicting 90 day acute heart failure readmission and death using machine learning-supported decision analysis. Clin Cardiol 2021;44:230-237.
- Lv H, Yang X, Wang B et al. Machine Learning-Driven Models to Predict Prognostic Outcomes in Patients Hospitalized With Heart Failure Using Electronic Health Records: Retrospective Study. J Med Internet Res 2021;23:e24996.
- Najafi-Vosough R, Faradmal J, Hosseini SK, Moghimbeigi A, Mahjub H. Predicting Hospital Readmission in Heart Failure Patients in Iran: A Comparison of Various Machine Learning Methods. Healthc Inform Res 2021;27:307-314.
- Shameer K, Johnson KW, Yahi A et al. Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort. Pac Symp Biocomput 2017;22:276-287.
- 20. Ashfaq A, Sant'Anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. J Biomed Inform 2019;97:103256.

- 21. Park J, Zhong X, Babaie Sarijaloo F, Wokhlu A. Tailored risk assessment of 90-day acute heart failure readmission or all-cause death to heart failure with preserved versus reduced ejection fraction. Clin Cardiol 2022;45:370-378.
- Beecy AN, Gummalla M, Sholle E et al. Utilizing electronic health data and machine learning for the prediction of 30-day unplanned readmission or all-cause mortality in heart failure.
 Cardiovasc Digit Health J 2020;1:71-79.
- 23. Jiang W, Siddiqui S, Barnes S et al. Readmission Risk Trajectories for Patients With Heart Failure Using a Dynamic Prediction Approach: Retrospective Study. JMIR Med Inform 2019;7:e14756.
- 24. Golas SB, Shibahara T, Agboola S et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 2018;18:44.
- 25. Mahajan SM, Ghani R. Using Ensemble Machine Learning Methods for Predicting Risk of Readmission for Heart Failure. Stud Health Technol Inform 2019:243-247.
- 26. Mortazavi BJ, Downing NS, Bucholz EM et al. Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes 2016;9:629-640.
- 27. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Chow BJ, Dwivedi G. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death. PLoS One 2019;14:e0218760.
- Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. Sci Rep 2019;9:9277.
- 29. Frizzell JD, Liang L, Schulte PJ et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. JAMA Cardiol 2017;2:204-209.

- 18
- Turgeman L, May JH. A mixed-ensemble model for hospital readmission. Artif Intell Med
 2016;72:72-82.
- 31. Xiao C, Ma T, Dieng AB, Blei DM, Wang F. Readmission prediction via deep contextual embedding of clinical concepts. PLoS One 2018;13:e0195024.
- Segar MW, Hall JL, Jhund PS et al. Machine Learning-Based Models Incorporating Social
 Determinants of Health vs Traditional Models for Predicting In-Hospital Mortality in Patients
 With Heart Failure. JAMA Cardiol 2022;7:844-854.
- Dharmarajan K, Hsieh AF, Lin Z et al. Diagnoses and timing of 30-day readmissions after
 hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA
 2013;309:355-63.
- Vader JM, LaRue SJ, Stevens SR et al. Timing and Causes of Readmission After Acute Heart
 Failure Hospitalization-Insights From the Heart Failure Network Trials. J Card Fail 2016;22:875-883.
- Leppin AL, Gionfriddo MR, Kessler M et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Intern Med 2014;174:1095-107.
- McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015;131:1796-803.
- Mpanya D, Celik T, Klug E, Ntsinjana H. Machine learning and statistical methods for predicting mortality in heart failure. Heart Fail Rev 2021;26:545-552.
- Choi E, Xiao C, Stewart W, Sun J. MiME: Multilevel medical embedding of electronic health records for predictive healthcare. arXiv 2018;1810.09593v1 [cx.LG].

- McGilvray MMO, Heaton J, Guo A et al. Electronic Health Record-Based Deep Learning Prediction of Death or Severe Decompensation in Heart Failure Patients. JACC Heart Fail 2022;10:637-647.
- 40. Murray F, Allen M, Clark CM, Daly CJ, Jacobs DM. Socio-demographic and -economic factors associated with 30-day readmission for conditions targeted by the hospital readmissions reduction program: a population-based study. BMC Public Health 2021;21:1922.
- 41. Sterling MR, Ringel JB, Pinheiro LC et al. Social Determinants of Health and 30-Day Readmissions Among Adults Hospitalized for Heart Failure in the REGARDS Study. Circ Heart Fail 2022;15:e008409.
- 42. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735-80.
- 43. Ma M, Liu C, Wei R, Liang B, Dai J. Predicting machine's performance record using the stacked long short-term memory (LSTM) neural networks. J Appl Clin Med Phys 2022;23:e13558.
- 44. Yu Y, Si X, Hu C, Zhang J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Comput 2019;31:1235-1270.

Figure 1

A deep learning model was constructed with a dense input layer, three long short-term memory

(LSTM) layers, and a dense hidden layer to cohesively extract features from time-series and non-

time-series EHR data.

Figure 2

Heat map demonstrating the relative predictive importance of vital sign and laboratory test variables depicted on the x and y axes, respectively. Variables are listed in descending order of univariate importance, with red hues representing those with highest impact on model prediction and blue hues highlighting low impact variables.

Figure 3

The area under the receiver-operating characteristic curve (AUC) (or C-statistic) of 0.61 with a 95% CI of 0.59-0.64 demonstrates marginal clinical accuracy in short-term prediction of adverse 30-day outcomes in this large study cohort.

Figure 4

The precision recall area under the curve (PR AUC) of 0.37 with a 95% CI of 0.33-0.41 further supports the notion that standard EHR data lacks crucial factors, such as socioeconomic

determinants, that lay pivotal roles in influencing early readmission and mortality in this cohort.

Table 1

Univariate importance of the 70 clinical variables evaluated for individual impact on 30-day all-

cause readmission and mortality.

Variable	AUC Without Variable	Importance				
Vital Signs						
Height	0.593	1.000				
Weight	0.594	0.996				
Heart rate	0.598	0.987				
Diastolic blood pressure	0.598	0.987				
Oxygen flow rate	0.599	0.984				
Systolic blood pressure	0.603	0.974				
Oxygen saturation	0.603	0.974				
Laboratory Values	1					
C-reactive protein	0.503	1.000				
Sodium	0.510	0.987				
Lactate dehydrogenase	0.513	0.980				
Brain natriuretic peptide	0.517	0.973				
Hematocrit	0.521	0.966				
Fasting glucose	0.521	0.965				
Monocytes	0.524	0.958				
Albumin	0.525	0.956				
Fraction of inspired oxygen	0.526	0.954				
Lactic acid	0.529	0.949				
Uric acid	0.529	0.948				
Plasma potassium	0.530	0.946				

Ammonia	0.531	0.945
Cannabinoids	0.532	0.943
Hemoglobin	0.533	0.940
Amylase	0.534	0.939
Total hemoglobin	0.535	0.936
Venous partial pressure of oxygen	0.537	0.932
Triglycerides	0.541	0.925
Hemoglobin A1c	0.541	0.925
Chloride	0.541	0.925
Potassium	0.541	0.924
Carbon dioxide	0.545	0.916
Carboxyhemoglobin	0.546	0.915
Calculated oxygen saturation	0.546	0.914
Venous oxygen content	0.546	0.914
Oxygen gradient	0.547	0.912
Ketones	0.550	0.907
Alanine aminotransferase	0.550	0.906
Basophils	0.552	0.903
Base excess	0.552	0.901
Methemoglobin	0.553	0.901
Blood urea nitrogen	0.553	0.899
Troponin I	0.555	0.897
Benzodiazepines	0.556	0.894
Pulmonary artery oxygen	0.558	0.891
Partial thromboplastin time	0.558	0.891
Magnesium	0.558	0.890
Alkaline phosphatase	0.559	0.888

Glucose	0.560	0.887		
Partial pressure of carbon dioxide	0.561	0.885		
Pulmonary artery oxyhemoglobin	0.563	0.880		
International normalized ratio	0.563	0.879		
Cocaine	0.564	0.878		
Creatinine	0.569	0.867		
Direct bilirubin	0.572	0.862		
High-density lipoprotein	0.574	0.858		
Prothrombin time	0.575	0.856		
Phosphorous	0.575	0.855		
D dimer	0.584	0.839		
Aspartate aminotransferase	0.586	0.833		
Fraction of inspired oxygen	0.595	0.816		
Plasma protein	0.595	0.815		
Anion gap	0.596	0.813		
Calcium	0.598	0.809		
рН	0.599	0.807		
Eosinophils	0.599	0.807		
Fibrinogen	0.600	0.806		
Urine specific gravity	0.600	0.806		
Lymphocytes	0.600	0.806		
Indirect Coombs	0.600	0.806		
Lipase	0.600	0.805		
Bicarbonate	0.602	0.802		
Creatine kinase	0.602	0.802		
Platelets	0.602	0.802		
Bilirubin	0.603	0.800		
L				

Partial pressure of oxygen	0.603	0.800		
Thyroid stimulating hormone	0.603	0.799		
Neutrophils	0.603	0.799		
Cholesterol	0.604	0.798		
Oxygen saturation	0.604	0.798		
Static Variables				
Unstructured row count	0.540	1.000		
Laboratory test row count	0.567	0.940		
Length of stay (LOS) hours	0.591	0.889		
Vital sign row count	0.597	0.875		
Gender	0.602	0.865		
Race	0.604	0.859		
Age at admission	0.606	0.857		

28

Table 2

Baseline characteristics in training and internal validation cohorts.

	Training Cohort	Internal Validation Cohort		
Number of participants				
Number of participants	19,884	2,480		
Age at admission, years	63 (SD:15)	63 (SD:15)		
Gender				
Female, %	9,052 (46)	1,095 (44)		
Male, %	10,832 (55)	1,390 (46)		
Unknown, %	0 (0)	1 (0)		
Ethnicity				
White, %	9,036 (45)	1,093 (44)		
Black, %	9,313 (47)	1,202 (48)		
Asian, %	87 (0)	7 (0)		
Other, %	1,448 (7)	184 (7)		
LOS, hours	268 (SD:718)	290 (SD:832)		
Systolic blood pressure, mmHg	125 (SD:15)	125 (SD:14)		
Diastolic blood pressure, mmHg	70 (SD:9)	71 (SD:9)		
Heart rate, bpm	82 (SD:8)	82 (SD:8)		
O2 saturation, %	97 (SD:1)	97 (SD:1)		

Table 3

	Negative	Positive
	Observation	Observation
Number of participants	49,675	9,975
Age at admission, years	62 (SD:13)	62 (SD:15)
Gender		
Female, %	22,956 (46)	4,488 (45)
Male, %	26,717 (54)	5,487 (55)
Unknown, %	2 (0)	0 (0)
Ethnicity		
White, %	19,373 (39)	3,591 (36)
Black, %	27,321 (55)	5,386 (54)
Asian, %	496 (1)	1 (0)
Other, %	2,485 (5)	997 (10)
LOS, hours	98 (SD:453)	128 (SD:439)
Systolic blood pressure, mmHg	127 (SD:20)	125 (SD:21)
Diastolic blood pressure, mmHg	71 (SD:13)	71 (SD:13)
Heart rate, bpm	81 (SD:13)	82 (SD:13)
O2 saturation, %	97 (SD:2)	97 (SD:2)

Comparison of positive and negative observations.

30

Table 4

Predictive model performance results.

Measure	Result
ROC AUC	0.613
Log Loss	0.570
PR AUC	0.380
Brier Score	0.190



A1: Input Layer

Unstructured Data

32 vector x 100

sequences

Dense 2: Model Output Input: 256+256+256+16 Output: 1



	SYSTOLIC	WEIGHT	PULSE	DIASTOLIC	HEIGHT	02 5AT	OF FLOW
Brein netrianetic peptide	0.967	0.997	0.965	0.972	0.958	0.958	1.000
Sotian	0.863	0.863	0.844	0.843	0.859	0.858	0.865
Vendus periol pressure of oxygen	0.852	0.650	0.862	0.841	0.896	0.868	0.896
Lactité delyérogenase	0.830	0.816	0.804	0.801	0.816	0.831	0.820
Hernelocrit	9.767	0.761	0.768	0.781	0.798	0.741	0.787
Albuma Compliant emplois	0.762	0.770	0.754	0.744	0.785	0.112	0.700
C-reative protein	0.767	0.770	0.754	0.791	0.781	0.758	0.100
Fraction of Imaliani causes	0.733	0.747	0.738	0.727	0.749	0.752	0.749
Hereoglobin	0.756	0.747	0.787	0.751	0.746	0.727	0.739
Cannabinoids	0.707	0.690	0.701	0.722	0.725	0.095	0.719
Calculated axygen saturation	0.728	0.710	0.745	0.739	0.745	0.740	0.713
Total hersoglobis	0.735	0.722	0.705	0.735	0.732	0.728	0.713
Arryine	0.680	0.696	0.684	0.691	0.694	0.702	0.701
atensational normalized ratio	0.691	0.726	0.689	0.098	0.705	0.094	0.695
Potentiam	0.694	0.700	0.709	0.700	0.696	0.681	0.689
Benzodiazepines	0.673	0.664	0.682	0.692	0.097	0.677	0.687
Onygen gradient	0.680	0.678	0.683	0.660	0.685	0.690	0.681
Venaus cayges cantent	0.004	0.665	0.679	0.005	0.335	0.678	0.678
Nonotyses	9.661	0.660	0.662	0.999	0.109	0.011	0.678
Page a colonalum	0.000	0.676	0.666	0.440	0.693	0.004	0.676
Lactic acid	0.659	0.664	0.660	0.653	0.677	0.670	0.057
Methemoglobin	0.664	0.648	0.660	0.665	0.647	0.657	0.664
Carbonyhemosisbin	0.654	0.679	0.673	0.679	0.674	0.675	0.660
Partial theorybopiastic time	0.665	0.659	0.655	0.658	D.954	0.059	0.659
Banophila	0.668	0.675	0.662	0.664	0.661	0.659	0.661
Pulmonary artery sxyhemoglobin	0.636	0.656	0.665	0.655	0.655	0.659	0.649
Ammonia	0.654	0.660	0.653	0.663	0.667	0.654	0.636
Partial pressure of carbon dioxide	0.643	0.621	0.645	0.622	0.632	0.630	0.636
Glucose	0.623	0.609	0.629	0.625	0.616	0.614	0.636
Carbon dioxide	0.639	0.630	0.647	0.631	0.632	0.639	0.635
Triglycerides	0.626	0.635	0.646	0.642	0.626	0.641	0.635
Produceros bree	0.091	0.697	0.615	0.442	0.023	0.040	0.030
Transair I	0.635	0.623	0.642	0.629	0.645	0.635	0.633
Fasting diagona	0.628	0.617	0.623	0.630	0.633	0.635	0.631
Hemoglobin Allo	0.631	0.631	0.626	0.624	0.646	0.635	0.629
Netones	0.634	0.643	0.635	0.643	D.647	0.611	0.628
Bane encene	0.655	0.652	0.656	0.649	0.646	0.634	0.625
Riccd urea nitragen	0.605	0.606	0.625	0.600	0.903	0.903	0.618
Alkaline phosphetese	0.620	0.611	0.634	0.631	0.648	0.625	0.617
Pulmonary artery soygen	0.591	0.597	0.586	0.600	0.933	0.593	0.605
Alamino aminofransferase	0.587	0.601	0.584	0.589	0.592	0.590	0.601
Cacaine	0.605	0.599	0.605	0.601	0.592	0.005	0.601
high-density lipopratain	0.575	0.562	0.503	0.000	0.670	0.982	0.587
Associate and other design	0.547	0.576	0.578	0.583	0.558	0.591	0.567
Phoethorous	0.563	0.562	0.561	0.565	0.555	0.545	0.565
Creatinine	0.543	0.535	0.536	0.544	0.540	0.529	0.544
Planma protein	0.613	0.628	0.524	0.525	0.624	0.520	0.525
D dimer	0.504	0.511	0.512	0.514	D.514	0.511	0.514
Praction of implined oxygen	0.909	0.506	0.501	0.514	0.509	0.509	0.611
рН	0.502	0.505	0.505	0.507	0.505	0.907	0.505
Calolam	4.607	0.507	0.567	0.509	0.508	0.909	4,605
Thyroid stimulating homone	0.501	0.503	0.503	0.505	0.503	0.500	0.504
Indirect Coamba	9.502	0.602	0.505	0.596	0.903	0.501	0.504
Choeldeloi	0.5002	0.503	0.502	0.503	0.503	0.001	0.504
Choattho Kinase	0.601	0.503	0.903	0.994	0.904	0.900	0.503
Non a spikilik gravity Rinationala	0.001	0.509	0.500	0.604	0.900	0.601	0.909
Platelets	0.501	0 500	0.503	0.503	0.901	P 501	0.503
Licese	0.501	0 500	0.500	0.504	0.972	0.501	0 502
Anion geg	4.902	0 504	0.504	0.503	0.904	0.494	0.902
Partial pressure of oxyges.	0.502	0.504	0.504	0.505	0.502	0.501	0.502
Blindin	0.601	0.601	0.508	0.504	0.901	9.500	9.502
Gosinophils	0.498	0.503	0.500	0.502	D.4999	0.499	0.502
Ne.exphile	0.500	0.501	0.502	0.504	0.900	0.500	0.502
Fibringes	0.500	0.503	0.501	0.505	0.902	0.503	0.500
Lamphoestas	0.600	0.500	0.501	0.503	0.502	0.500	0.500



