5-1-2022

Development and structure of an accurate machine learning algorithm to predict inpatient mortality and hospice outcomes in the coronavirus disease 2019 era

Stephen Chi

Aixia Guo

Kevin Heard

Seunghwan Kim

Randi Foraker

See next page for additional authors

Authors

Stephen Chi, Aixia Guo, Kevin Heard, Seunghwan Kim, Randi Foraker, Patrick White, and Nathan Moore

OPEN

Development and Structure of an Accurate Machine Learning Algorithm to Predict Inpatient Mortality and Hospice Outcomes in the Coronavirus Disease 2019 Era

Stephen Chi, MD, Aixia Guo, PhD,† Kevin Heard, MS,‡ Seunghwan Kim, MS,§ Randi Foraker, PhD,† Patrick White, MD,‖ and Nathan Moore, MD¶*

Background: The coronavirus disease 2019 (COVID-19) pandemic has challenged the accuracy and racial biases present in traditional mortality scores. An accurate prognostic model that can be applied to hospitalized patients irrespective of race or COVID-19 status may benefit patient care.

Research Design: This cohort study utilized historical and ongoing electronic health record features to develop and validate a deep-learning model applied on the second day of admission predicting a composite outcome of in-hospital mortality, discharge to hospice, or death within 30 days of admission. Model features included patient demographics, diagnoses, procedures, inpatient medications, laboratory values, vital signs, and substance use history. Conventional performance metrics were assessed, and subgroup analysis was performed based on race, COVID-19 status, and intensive care unit admission.

Subjects: A total of 35,521 patients hospitalized between April 2020 and October 2020 at a single health care system including a tertiary academic referral center and 9 community hospitals.

Results: Of 35,521 patients, including 9831 non-White patients and 2020 COVID-19 patients, 2838 (8.0%) met the composite outcome. Patients who experienced the composite outcome were older (73 vs. 61 y old) with similar sex and race distributions between groups. The model achieved an area under the receiver operating characteristic curve of 0.89 (95% confidence interval: 0.88, 0.91) and an average positive predictive value of 0.46 (0.40, 0.52). Model performance did not differ significantly in White (0.89) and non-White (0.90) subgroups or when grouping by COVID-19 status and intensive care unit admission.

Conclusion: A deep-learning model using large-volume, structured electronic health record data can effectively predict short-term mortality or hospice outcomes on the second day of admission in the general inpatient population without significant racial bias.

Key Words: health informatics, palliative care, machine learning, clinical prediction rules

(*Med Care* 2022;60: 381–386)

The health care crisis caused by coronavirus disease 2019 (COVID-19) has highlighted deficiencies in current mortality prediction tools. As a critical endpoint for patients and physicians alike, mortality has been the subject of numerous clinical prediction models. Independent evaluation has found significant heterogeneity in accuracy and within-model variability, however, with many predictive tools demonstrating only modest discrimination, low clinical utility, and high risk of racial bias.[1–6] These limitations were further underscored by COVID-19, which forced many institutions to implement triage policies under crisis standards of care. Over 80% of hospitals with formal triage policies used a version of the Sequential Organ Failure Assessment score, which was later found to both systematically prioritize White patients over Black patients while also demonstrating lower discriminant accuracy than simply using patient age to predict mortality in ventilated COVID-19 patients.[7–9] Even as health care systems look to move past COVID-19, novel methods to estimate mortality without racial bias are clearly needed.

To address this prognostic challenge, researchers and clinicians have increasingly turned to machine learning (ML). With the ability to model complex interactions between diverse clinical datasets, ML models show superior performance to conventional mortality risk scores in diseases such as sepsis or heart failure.[10,11] Comparatively few studies attempt to apply ML to inpatient mortality, however, with limited generalizability driven by technical design,

From the *Division of Pulmonary and Critical Care Medicine; †Institute for Informatics, Washington University in St. Louis; ‡BJC HealthCare; §Division of General Medical Sciences, School of Medicine, Washington University in St. Louis; ‖Division of Palliative Medicine, Department of Medicine, Washington University in St. Louis; and ¶BJC Medical Group, St. Louis, MO.

This work was produced with the support of the Big Ideas Program, a BJC HealthCare, and Washington University internal grant program, hosted by the Healthcare Innovation Lab and the Institute for Informatics.

The authors declare no conflict of interest.

Correspondence to: Stephen Chi, MD, 4523 Clayton Avenue #8052, Saint Louis, MO 63110. E-mail: chis@wustl.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

ISSN: 0025-7079/22/6005-0381

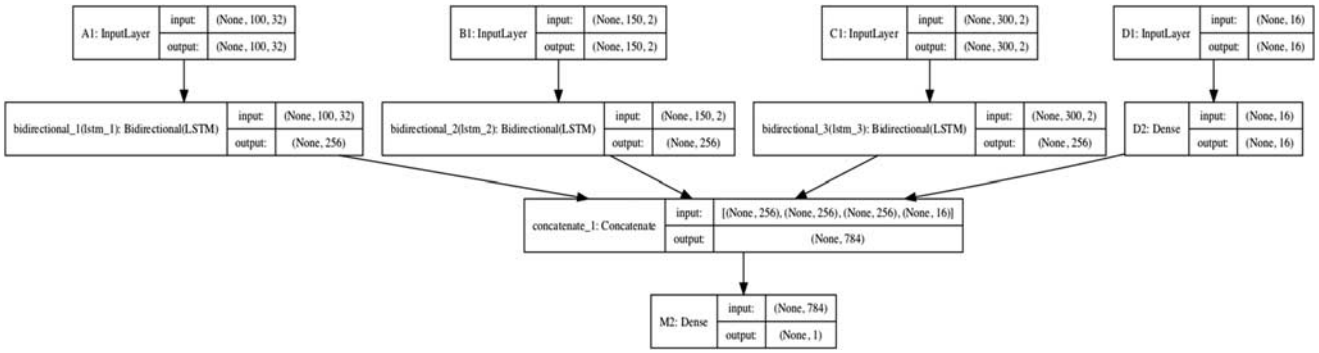

FIGURE 1. Deep learning model structure. Three bidirectional long short-term memory (LSTM) models were constructed to analyze clinical variables grouped by data type: 100 most recent diagnosis, procedure, and medication codes (A); 150 most recent laboratory test names and values (B); 300 most recent vital sign names and values (C). A fourth neural network model (D) was comprised of demographic and social history variables.

heterogenous populations, and potential for racial bias that has gone largely uninvestigated.[6,12–17] The data used to train these models also predate the COVID-19 pandemic, which impacted health care delivery and outcomes nationwide.[18–22] The utility of ML in predicting mortality outcomes of hospitalized patients in the modern COVID-19 era therefore remains to be demonstrated.

We have previously shown that a long short-term memory (LSTM) deep-learning model trained on electronic health record (EHR) and claims data was effective at predicting mortality relative to 3 other common ML approaches.[23] LSTM are a type of recurrent neural network that analyzes time-sorted data in the context of surrounding values; among other advantages, this property allows LSTM to learn from trends in patient data, thereby mimicking clinical practice. By consolidating the model to EHR-specific features and adding time-dependent variables, specifically vital signs and laboratory results from the patient's first hospital day, we sought to create an accurate, racially unbiased deep-learning model which would be readily integrated into the EHR for real-time clinical application.

TABLE 1. Study Population Characteristics

	n (%)		
Characteristics	No Mortality/ Hospice Outcome	Mortality/ Hospice Outcome	Total
No. patients	32,682 (92.0)	2839 (8.0)	35,521
Age [mean (SD)]	61.2 (17.5)	72.6 (15.0)	62.1 (17.6)
Sex			
Male	16,191 (49.5)	1451 (51.1)	17,642 (49.7)
Female	16,490 (50.5)	1388 (48.9)	17,878 (50.3)
Unknown	1 (0.0)	0 (0.0)	1 (0.0)
Race			
White	23,616 (72.3)	2074 (73.1)	25,690 (72.3)
Black	8438 (25.8)	669 (23.6)	9107 (25.6)
Asian	257 (0.8)	22 (0.8)	279 (0.8)
Other	371 (1.1)	74 (2.6)	445 (1.0)
COVID-19 status			
COVID-19 (+)	1587 (4.9)	433 (15.3)	2020 (5.7)
COVID-19 (–)	31,095 (95.1)	2406 (84.7)	33,501 (94.3)
ICU admission in first 24 h			
ICU (+)	4475 (13.7)	1089 (38.4)	5564 (15.7)
ICU (–)	28,207 (86.3)	1750 (61.6)	29,957 (84.3)

COVID-19 indicates coronavirus disease 2019; ICU, intensive care unit.

METHODS

Data Source

Patient data were collected from admissions to 10 academic and community-based hospitals within the BJC HealthCare system from April 2020 to October 2020. This health care system covers a diverse catchment area across mid-Missouri, Southern Illinois, and greater Saint Louis regions, and includes a 1250-bed tertiary referral center as well as multiple community hospitals. Hospitalizations longer than 24 hours were included for feature extraction. Admissions to psychiatry, labor/delivery, and bone marrow transplant units were excluded. This project was approved by the Washington University in Saint Louis Institutional Review Board; need for informed consent was waived.

Cohort and Study Design

Of 46,206 admissions with hospitalizations longer than 24 hours and identifiable mortality outcomes within 30 days, a total of 35,521 unique patients were included in the analysis. For patients with multiple admissions, one admission was selected randomly for feature extraction to reduce selection bias. The primary outcome of interest was in-hospital mortality, discharge to hospice, or death within 30 days of admission.

Subgroup analysis was performed in 3 patient subgroups to evaluate model accuracy and bias. These subgroups were defined by COVID-19 status, intensive care unit (ICU) admission within the first 24 hours of hospitalization, and race (White and non-White). To account for external COVID-19 testing and processing time for internal COVID-19 assays, COVID-19 status was determined retrospectively based on either a positive test result for COVID-19 or infection prevention flags specifying confirmed COVID-19 infection during the index admission.

Feature Extraction

All encounter records for each patient available up to 24 hours after time of admission were extracted from the EHR. Features included demographics, diagnosis codes, procedure codes, inpatient medication lists, laboratory results, vital signs, and social history. All features were sorted in a time increasing order.

Diagnosis, procedure, and medication codes were mapped to a 32-dimensional vector space using the Word2Vec technique.[24] The Python Genism Word2Vec model employed the following hyperparameters: size (embedding dimension) of 32, window (the maximum distance between a target word and all words around it) of 5, min_count (the minimum number of words counted when training the model) of 1, and the sg (training algorithm) was CBOW (continuous bag of words). Each feature was then represented by a 32-dimensional numerical vector.

Deep-learning Model Development

Patient features were structured into 4 groups for input into the deep-learning model: (1) embedding vectors of diagnosis codes, procedure codes, and medication codes with a dimension of (100, 32), where 100 denoted the most recent 100 codes and 32 was the dimension of embedding vectors; (2) numerical variables from laboratory results with a dimension of (150, 2), where 150 was the number of most recent laboratory records, and 2 was the test names and related values; (3) numerical variables from vital signs with a dimension of (300, 2), where 300 was the number of most recent vital sign records, and 2 was the vital sign names and related values; and (4) demographic and social history variables with a dimension of 16, which represented age and categorical variables such as sex, COVID-19 infection status, ICU admission, and substance use history.

The deep-learning model was comprised of 3 bidirectional LSTM models and a fourth neural network model representing the above groups (Fig. 1). A binary cross-entropy loss function was employed as the output layer and a Sigmoid function was used as the activation function for the hidden layer. An Adam optimizer was used to optimize the model with a mini-batch size of 256 samples.

Data Splits and Model Evaluation

Patients were randomly divided into training (80%), validation (10%), and testing (10%) datasets for a total of 28,417 patients in the training data set and 3552 patients each in the validation and testing datasets. Model performance was evaluated in the overall cohort and each subgroup using standard performance metrics as well as receiver operating characteristic (ROC) and precision-recall curves. Precision-recall curves compare a model's positive predictive value (PPV or precision) against its sensitivity (recall) as the discrimination threshold is varied. In contrast to ROC curves which utilize the false positive rate instead of PPV, precision-recall curves are not dependent on the number of true negative cases and may therefore be more informative for imbalanced datasets with a low expected outcome rate. Ninety-five percent confidence intervals were calculated for each metric. The model was also tested under a series of discrimination thresholds or cutoffs, ranging from 0.1 to 0.9.

RESULTS

The observed rate of the composite outcome of in-hospital mortality, discharge to hospice, or death within 30 days of admission was 8%. Patients who met the composite outcome were older (73 vs. 61 y) with similar sex and race distributions between groups (Table 1). Patients with COVID-19 infection were more likely to experience the mortality/hospice outcome than patients without COVID-19 (21% vs. 7%). Thirty-eight percent of patients who met the mortality/hospice outcome were admitted to an ICU within the first 24 hours of admission.

ROC and precision-recall curves were constructed to evaluate model performance in the overall cohort and patient subgroups

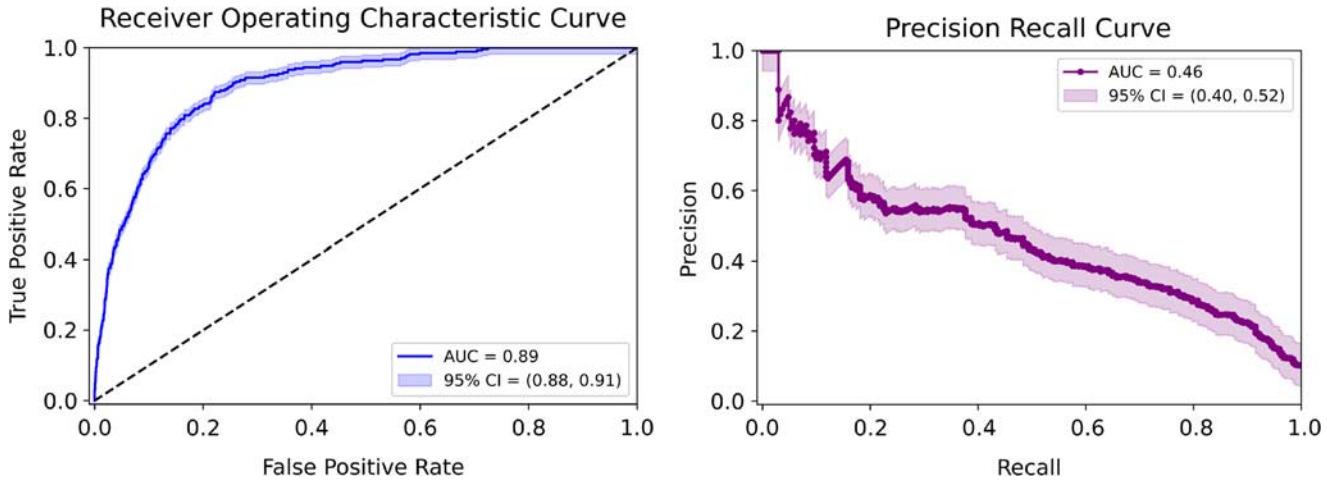

FIGURE 2. Deep learning model prediction performance. Area under the receiver operating characteristic and precision-recall curves for model performance in the overall cohort. Shaded areas denote 95% CIs. AUC indicates area under the curve; CI, confidence interval.

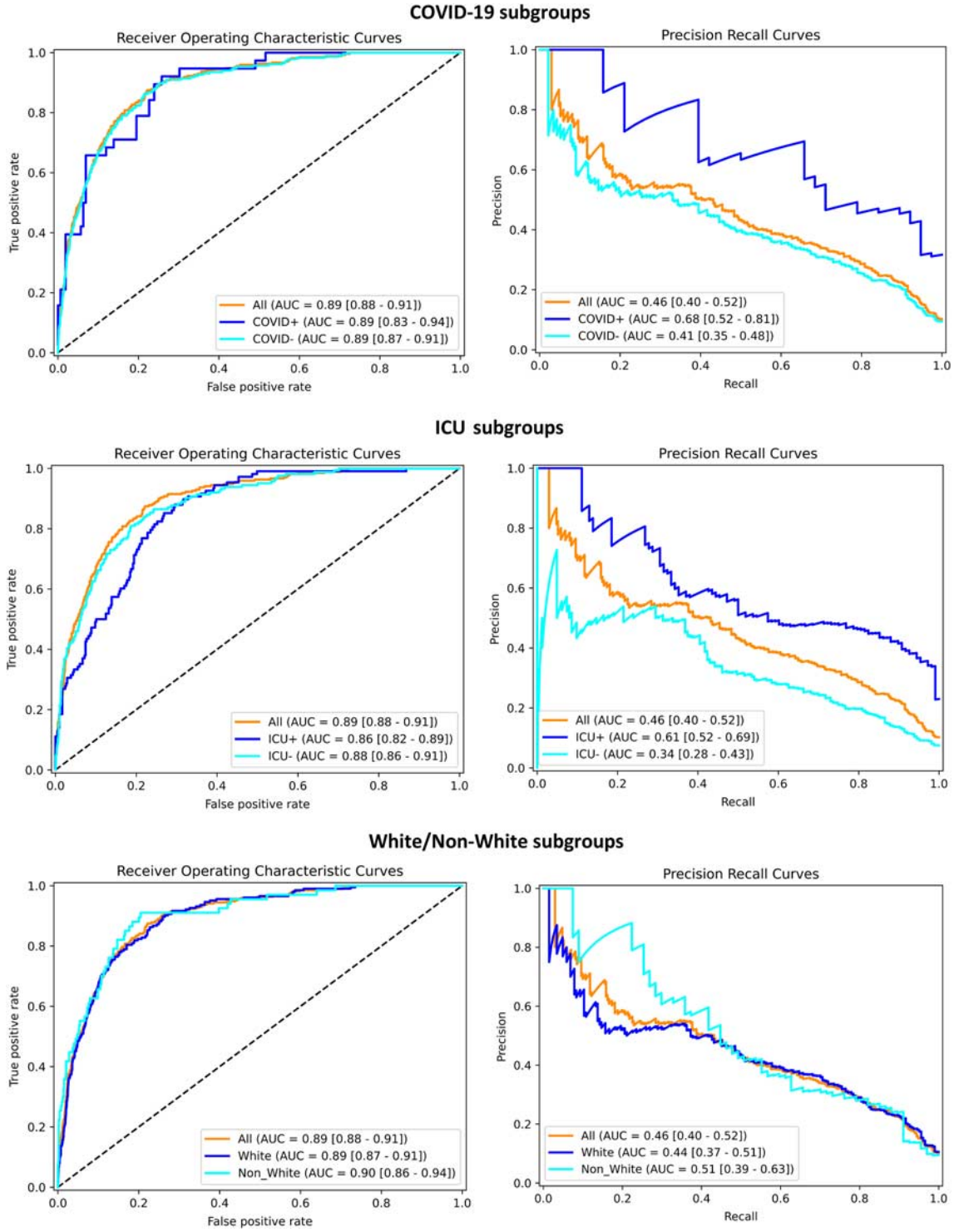

FIGURE 3. Deep learning model performance in clinical subgroups. Area under the receiver operating characteristic and precision-recall curves for model performance in clinical subgroups. COVID-positive patients were classified retrospectively based on either a positive COVID-19 test result or infection prevention flags specifying confirmed COVID-19 infection during the index admission. ICU+ was defined by patients admitted to an ICU within the first 24 hours of hospitalization. Brackets indicate 95% confidence intervals. AUC indicates area under the curve; COVID-19, coronavirus disease 2019; ICU, intensive care unit.

(Figs. 2, 3). In the overall study cohort, the area under the ROC curve was 0.89 (0.88, 0.91). The area under the precision-recall curve was 0.46 (0.40, 0.52), corresponding to an average model PPV of 46% compared with the observed outcome rate of 8%. Table 2 lists the model's performance metrics at different thresholds of predicted probabilities. The F1 score, which is the harmonic mean of PPV and sensitivity, plateaued at thresholds up to 0.3; this cutoff corresponded to a sensitivity of 0.38 (0.32, 0.43), specificity of 0.97 (0.97, 0.97), and PPV of 0.54 (0.46, 0.61).

In the prespecified patient subgroups, area under the ROC curve did not differ significantly by COVID-19 status [0.89 (0.83, 0.94) vs. 0.89 (0.87, 0.91)], ICU [0.86 (0.82, 0.89)] versus non-ICU admission [0.88 (0.86, 0.91)], or Whites [0.89 (0.87, 0.91)] compared with non-Whites [0.90 0.86, 0.94)]. Average model PPV was higher in the COVID-19 positive [0.68 (0.52, 0.81)] and ICU admission [0.71 (0.52, 0.69)] subgroups. Subgroup model performance by cutoffs was also evaluated (Supplemental Table 1, Supplemental Digital Content 1, http://links.lww.com/MLR/C412); when using a threshold risk score of 0.3, positive predictive value exceeded 50% and specificity was >90% in every subgroup.

DISCUSSION

In this study, we developed and validated a structured deep-learning model using EHR-specific data from academic and community hospitals to predict in-hospital mortality, discharge to hospice, or death within 30 days of admission. The model incorporated both historical and acute variables in the patient's medical record to generate a numerical risk score on the second day of admission for clinical use. This ML approach was shown to have excellent predictive value with no significant differences in performance based on race or COVID-19 status.

Our investigation has several unique strengths and limitations. The model was trained on data from the first 7 months of the COVID-19 pandemic, during which health care delivery was severely impacted in the form of treatment delays, experimental interventions, and crisis standards of care seen at many facilities.[18–22,25–29] This clinical variability would be expected to diminish the accuracy of any prediction tool, yet our model showed favorable performance metrics and no significant racial bias in comparison to many previously published traditional and ML mortality scores.[1,6,9] This finding may be attributed to the structure and scope of our deep-learning model. An average of over 500continuous, categorical, and multidimensional variables per patient were incorporated into a framework of LSTM models, which may be more effective than other ML techniques at predicting mortality.[23] Compared with other mortality scores that include only acute gical or disease-specific data, this multifaceted approach to feature inclusion leverages the full breadth of data available in modern EHRs, thereby providing a more comprehensive picture of patient health.[6,10,30] This reliance on "big data" decreases our model's transparency, however, as the factors contributing to a patient's score within the LSTM hierarchy are not readily visible. This lack of interpretability is a significant limitation shared by other ML models intended for clinical implementation.[31] Future steps to improve model transparency, such as feature importance analysis, will be required to maximize clinical utilization.

Our model was developed using data from a single health care system, which poses a significant challenge to portability due to EHR and population heterogeneity. While direct replication of our model may not be feasible, however, we would argue that our study population and model design support generalization of similarly structured ML models to other centers. The training data set included a diverse patient population spread across medical, surgical, and subspecialty floors and ICUs at academic and community hospitals encompassing a large geographic area. Urban, suburban, and rural counties were included along with significant non-White representation. The model structure does not require any manual curation or preprocessing of patient features, which can dramatically increase cost and computational time, and the use of EHR-specific data has enabled complete integration of the model into our institution's EHR.[32] Risk scores are now generated automatically on all inpatients without any user input for real-time clinical use. While implementation will necessarily vary between centers, we hope other institutions see this model as a proof-of-concept for the potential of LSTM models to yield accurate, racially unbiased predictions in the COVID-19 era.

Clinical applications of this mortality model are the subject of ongoing investigations. Our model demonstrated excellent positive predictive value in all clinical subgroups, supporting its clinical relevance as a mortality screening tool.[33] For example, a screening threshold of 0.3, corresponding to a minimum 30% risk of inpatient mortality, hospice discharge, or death within 30 days of admission, would identify at-risk patients with >50% PPV and 90% specificity. Clinical care can then be enhanced for those patients through increased clinical attention or advance care planning discussions. One promising application is through directed palliative care, which has been previously shown to decrease ICU transfers, increase advance care planning, and facilitate goal-concordant limitations in care.[34–36] With multiple recent studies highlighting the potential of improving palliative care

TABLE 2. Performance Metrics by Cutoff Values (95% Confidence Interval)

Cutoffs	Accuracy	Sensitivity	Specificity	Precision	F1 Score	Negative Predictive Value
0.1	0.86 (0.84, 0.87)	0.76 (0.71, 0.81)	0.86 (0.86, 0.86)	0.32 (0.28, 0.35)	0.45 (0.4, 0.49)	0.98 (0.98, 0.98)
0.2	0.91 (0.9, 0.92)	0.52 (0.46, 0.57)	0.94 (0.94, 0.94)	0.42 (0.37, 0.47)	0.46 (0.41, 0.51)	0.96 (0.96, 0.96)
0.3	0.93 (0.92, 0.94)	0.38 (0.32, 0.43)	0.97 (0.97, 0.97)	0.54 (0.46, 0.61)	0.44 (0.39, 0.5)	0.95 (0.95, 0.95)
0.4	0.93 (0.92, 0.93)	0.23 (0.18, 0.27)	0.98 (0.98, 0.98)	0.55 (0.46, 0.65)	0.32 (0.26, 0.38)	0.94 (0.94, 0.94)
0.5	0.93 (0.92, 0.94)	0.13 (0.09, 0.16)	0.99 (0.99, 0.99)	0.64 (0.51, 0.78)	0.21 (0.15, 0.27)	0.93 (0.93, 0.93)
0.6	0.93 (0.92, 0.94)	0.1 (0.06, 0.13)	1.0 (1.0, 1.0)	0.76 (0.61, 0.89)	0.17 (0.11, 0.22)	0.93 (0.93, 0.93)
0.7	0.93 (0.92, 0.93)	0.05 (0.02, 0.07)	1.0 (1.0, 1.0)	0.81 (0.6, 1.0)	0.09 (0.05, 0.14)	0.93 (0.93, 0.93)
0.8	0.93 (0.92, 0.93)	0.02 (0.0, 0.04)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)	0.04 (0.01, 0.07)	0.93 (0.93, 0.93)
0.9	0.93 (0.92, 0.93)	0.0 (0.0, 0.0)	1.0 (1.0, 1.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.93 (0.93, 0.93)

through deep learning, we plan to implement a targeted palliative care intervention through the hospital-wide screening potential of our model.[37–39]

CONCLUSIONS

A structured deep-learning model developed during the COVID-19 pandemic accurately predicted in-hospital mortality, discharge to hospice, or death within 30 days of admission among inpatients at a large academic and community-based health care system. Our study suggests that a single model can predict short-term mortality outcomes of patients across multiple clinical subgroups with excellent predictive value and minimal racial bias. Clinical applications of this inpatient mortality model are the targets of ongoing investigation.

ACKNOWLEDGMENTS

The authors thank Joshua Landman for his assistance with data management and analysis.

REFERENCES

1. Siontis GCM, Tzoulaki I, Ioannidis JPA. Predicting death: an empirical evaluation of predictive tools for mortality. *Arch Intern Med*. 2011;171:1721–1726.
2. Yourman LC, Lee SJ, Schonberg MA, et al. Prognostic indices for older adults: a systematic review. *JAMA*. 2012;307:182–192.
3. Vigil JM, Coulombe P, Alcock J, et al. Patient ethnicity affects triage assessments and patient prioritization in U.S. Department of Veterans Affairs Emergency Departments. *Medicine (Baltimore)*. 2016;95:e3191.
4. Ho LO, Li H, Shahidah N, et al. Poor performance of the modified early warning score for predicting mortality in critically ill patients presenting to an emergency department. *World J Emerg Med*. 2013;4:273–278.
5. Schrader CD, Lewis LM. Racial disparity in emergency department triage. *J Emerg Med*. 2013;44:511–518.
6. Allen A, Mataraso S, Siefkas A, et al. A racially unbiased, machine learning approach to prediction of mortality: algorithm development study. *JMIR Public Health Surveill*. 2020;6:e22400.
7. Antommaria AHM, Gibb TS, McGuire AL, et al. Ventilator triage policies during the COVID-19 pandemic at U.S. hospitals associated with members of the Association of Bioethics Program Directors. *Ann Intern Med*. 2020;173:188–194.
8. Raschke RA, Agarwal S, Rangan P, et al. Discriminant accuracy of the SOFA score for determining the probable mortality of patients with COVID-19 pneumonia requiring mechanical ventilation. *JAMA*. 2021;325:1469–1470.
9. Ashana DC, Anesi GL, Liu VX, et al. Equitably allocating resources during crises: racial differences in mortality prediction models. *Am J Respir Crit Care Med*. 2021;204:178–186.
10. Guo A, Pasque M, Loh F, et al. Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models. *Curr Epidemiol Rep*. 2020;7:212–219.
11. van Doorn WPTM, Stassen PM, Borggreve HF, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. *PLoS One*. 2021;16:e0245157.
12. Mohamadlou H, Panchavati S, Calvert J, et al. Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction. *Health Informatics J*. 2020;26:1912–1925.
13. Soffer S, Klang E, Barash Y, et al. Predicting in-hospital mortality at admission to the medical ward: a big-data machine learning model. *Am J Med*. 2021;134:227–234.e4.
14. Major VJ, Aphinyanaphongs Y. Development, implementation, and prospective validation of a model to predict 60-day end-of-life in hospitalized adults upon admission at three sites. *BMC Med Inform Decis Mak*. 2020;20:214.
15. Brajer N, Cozzi B, Gao M, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. *JAMA Netw Open*. 2020;3:e1920733.
16. Escobar GJ, Gardner MN, Greene JD, et al. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care*. 2013;51:446–453.
17. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453.
18. Rosenbaum L. The untold toll—the pandemic's effects on patients without COVID-19. *N Engl J Med*. 2020;382:2368–2371.
19. Satty T, Ramgopal S, Elmer J, et al. EMS responses and non-transports during the COVID-19 pandemic. *Am J Emerg Med*. 2021;42:1–8.
20. Ueberham L, König S, Pellissier V, et al. Admission rates and care pathways in patients with atrial fibrillation during the COVID-19 pandemic—insights from the German-wide Helios Hospital Network. *Eur Heart J Qual Care Clin Outcomes*. 2021;7:257–264.
21. Baum A, Schwartz MD. Admissions to Veterans Affairs hospitals for emergency conditions during the COVID-19 pandemic. *JAMA*. 2020;324:96–99.
22. Rizzi A, Mammarella L, Necozione S, et al. Looking at intra-hospital non COVID-19 mortality among elderly patients during COVID-19 pandemic. *Arch Gerontol Geriatr*. 2020;90:104173.
23. Guo A, Foraker R, White P, et al. Using electronic health records and claims data to identify high-risk patients likely to benefit from palliative care. *Am J Manag Care*. 2021;27:e7–e15.
24. Mikolov T, Chen K, Corrado G, et al. Word2vec. Proc Int Conf Learn Represent; 2013.
25. El-Solh AA, Lawson Y, Carter M, et al. Comparison of in-hospital mortality risk prediction models from COVID-19. *PLoS One*. 2020;15:e0244629.
26. Stasi C, Fallani S, Voller F, et al. Treatment for COVID-19: an overview. *Eur J Pharmacol*. 2020;889:173644.
27. Zhang JJY, Lee KS, Ang LW, et al. Risk factors for severe disease and efficacy of treatment in patients infected with COVID-19: a systematic review, meta-analysis, and meta-regression analysis. *Clin Infect Dis*. 2020;71:2199–2206.
28. RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with COVID-19. *N Engl J Med*. 2021;384:693–704.
29. Yokoyama Y, Briasoulis A, Takagi H, et al. Effect of remdesivir on patients with COVID-19: a network meta-analysis of randomized control trials. *Virus Res*. 2020;288:198137.
30. Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiol*. 2021;6:633–641.
31. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. 2021;113:103655.
32. Sendak MP, Balu S, Schulman KA. Barriers to achieving economies of scale in analysis of EHR data. *Appl Clin Inform*. 2017;08:826–831.
33. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21:128–138.
34. Picker D, Dans M, Heard K, et al. A randomized trial of palliative care discussions linked to an automated early warning system alert. *Crit Care Med*. 2017;45:234–240.
35. Deptola AZ, Riggs J. Inpatient goals-of-care conversations reduce intensive care unit transfers in high-risk patients. *Am J Hosp Palliat Care*. 2019;36:583–586.
36. Ma J, Chi S, Buettner B, et al. Early palliative care consultation in the medical ICU: a cluster randomized crossover trial. *Crit Care Med*. 2019;47:1707–1715.
37. Avati A, Jung K, Harman S, et al. Improving palliative care with deep learning. *BMC Med Inform Decis Mak*. 2018;18(suppl 4):122.
38. Courtright KR, Chivers C, Becker M, et al. Electronic health record mortality prediction model for targeted palliative care among hospitalized medical patients: a pilot quasi-experimental study. *J Gen Intern Med*. 2019;34:1841–1847.
39. Murphree DH, Wilson PM, Asai SW, et al. Improving the delivery of palliative care through predictive modeling and healthcare informatics. *J Am Med Inform Assoc*. 2021;28:1065–1073.