Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

JACC: ADVANCES VOL. 1, NO. 2, 2022

ORIGINAL RESEARCH

EMERGING TECHNOLOGIES AND INNOVATION

Real-Time Prediction of Mortality, Cardiac Arrest, and Thromboembolic Complications in Hospitalized Patients With COVID-19

Julie K. Shade, PhD,[a,b] Ashish N. Doshi, MD, PhD,[b,c] Eric Sung,[a,b] Dan M. Popescu, PhD,[b,d] Anum S. Minhas, MD,[e] Nisha A. Gilotra, MD,[e] Konstantinos N. Aronis, MD,[b,e] Allison G. Hays, MD,[e] Natalia A. Trayanova, PhD[a,b,e]

ABSTRACT

BACKGROUND COVID-19 infection carries significant morbidity and mortality. Current risk prediction for complications in COVID-19 is limited, and existing approaches fail to account for the dynamic course of the disease.

OBJECTIVES The purpose of this study was to develop and validate the COVID-HEART predictor, a novel continuously updating risk-prediction technology to forecast adverse events in hospitalized patients with COVID-19.

METHODS Retrospective registry data from patients with severe acute respiratory syndrome coronavirus 2 infection admitted to 5 hospitals were used to train COVID-HEART to predict all-cause mortality/cardiac arrest (AM/CA) and imaging-confirmed thromboembolic events (TEs) (n = 2,550 and n = 1,854, respectively). To assess COVID-HEART's performance in the face of rapidly changing clinical treatment guidelines, an additional 1,100 and 796 patients, admitted after the completion of development data collection, were used for testing. Leave-hospital-out validation was performed.

RESULTS Over 20 iterations of temporally divided testing, the mean area under the receiver operating characteristic curve were 0.917 (95% confidence interval [CI]: 0.916-0.919) and 0.757 (95% CI: 0.751-0.763) for prediction of AM/CA and TE, respectively. The interquartile ranges of median early warning times were 14 to 21 hours for AM/CA and 12 to 60 hours for TE. The mean area under the receiver operating characteristic curve for the left-out hospitals were 0.956 (95% CI: 0.936-0.976) and 0.781 (95% CI: 0.642-0.919) for prediction of AM/CA and TE, respectively.

CONCLUSIONS The continuously updating, fully interpretable COVID-HEART predictor accurately predicts AM/CA and TE within multiple time windows in hospitalized COVID-19 patients. In its current implementation, the predictor can facilitate practical, meaningful changes in patient triage and resource allocation by providing real-time risk scores for these outcomes. The potential utility of the predictor extends to COVID-19 patients after hospitalization and beyond COVID-19. (JACC Adv 2022;1:100043) © 2022 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA; [b]Alliance for Cardiovascular Diagnostic and Treatment Innovation, Johns Hopkins University, Baltimore, Maryland, USA; [c]Division of Pediatric Cardiology, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [d]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, Maryland, USA; and the [e]Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received February 23, 2022; revised manuscript received April 26, 2022, accepted April 27, 2022.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2022.100043

ABBREVIATIONS AND ACRONYMS

AM/CA = all-cause mortality/cardiac arrest

AUROC = area under the receiver operating characteristic curve

CV = cardiovascular

ICU = intensive care unit

ML = machine learning

SARS-CoV-2 = severe acute respiratory syndrome-coronavirus-2

TE = thromboembolic events

Patients with COVID-19, the disease caused by the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), often present with cardiovascular (CV) manifestations such as myocardial infarction, thromboembolism, and heart failure.[1] Clinically overt cardiac injury or cardiomyopathy is reported in 8% to 33% of hospitalized patients[2,3] and is associated with up to 50% mortality,[4] but imaging studies suggest the true incidence of cardiac involvement in all persons infected with SARS-CoV-2 could be as high as 60%.[5] Thromboembolic events (TEs) are also frequently reported in severe COVID-19 and are associated with mortality; 1 study found that 70.1% of nonsurvivors and 0.6% of survivors met criteria for disseminated intravenous coagulation.[6] Furthermore, thromboembolic complications are more pronounced in acute COVID-19 infection than in other viral illnesses and include pulmonary embolus and ischemic stroke, which can be fatal and are a significant cause of morbidity even as the infection resolves.[7] Despite the prevalence of thromboembolism and cardiac injury and their associations with poor outcomes,[2,6,8] no approach currently exists to forecast these types of adverse events in COVID-19 patients in real time.

Machine learning (ML) techniques are ideal for discovering patterns in high-dimensional biomedical data, especially when little is known about the underlying biophysical processes. ML is thus well positioned for applications in COVID-19 and indeed has been employed in screening, contract tracing, drug development, and outbreak forecasting.[9,10] ML approaches have been developed for prognostic assessment of hospitalized patients with COVID-19, including models that predict in-hospital mortality,[11-16] progression to severe disease,[13,17-20] and outcomes related to respiratory function.[9,14,21] A continuous remote monitoring system has been developed and validated,[22] but it is designed for outpatient use and does not include laboratory test results. An ML model was also proposed for prediction of TEs, but it required that all variables be present for all patients, did not provide dynamic risk updates, and was trained with data from only 76 patients.[23]

In this study, we develop and validate a prognostic ML model to forecast the *real-time* risk of all-cause mortality/cardiac arrest (AM/CA) and TEs in hospitalized patients with COVID-19. We term the model the COVID-HEART predictor. We focus on predicting 2 clinically important outcomes in COVID-19: in-hospital AM/CA and TEs. In-hospital AM/CA is a clearly identifiable outcome and is often CV-related, thus it was selected for proof-of-concept to demonstrate the potential utility of COVID-HEART. TEs are more difficult to identify and require imaging confirmation; thus, this outcome was selected to demonstrate the versatility of COVID-HEART in analyzing real-world clinical data and handling disease-specific outcomes. Finally, the predictor is tested in 2 different ways. First, it is tested with data from patients hospitalized after the end of data collection for patients in the development set, to ascertain that COVID-HEART can accurately predict risk in real time for new patients in the face of rapidly changing clinical treatment guidelines. The predictor is next tested with leave-hospital-out nested cross-validation to assess its performance when training and testing are done with data from different populations.

MATERIALS AND METHODS

PATIENT POPULATION. The COVID-HEART predictor was developed and validated in a retrospective cohort study approved by the Johns Hopkins University Institutional Review Board on May 21, 2020, under the protocol number IRB00249548: Prediction of Cardiac Dysfunction in COVID-19 Patients Using Machine Learning. The COVID-HEART study included adult patients (age ≥18 years at the time of COVID-19 diagnosis) admitted as inpatients to any of the following hospitals in the Johns Hopkins Health System: Howard County General Hospital, Suburban Hospital, Sibley Memorial Hospital, Johns Hopkins Bayview Medical Center, and Johns Hopkins Hospital. Patient data were collected in the retrospective COVID-19 Precision Medicine Analytics Platform Registry (JH-CROWN). For data from an admission to be included in this study, patients must have had SARS-CoV-2 infection confirmed by polymerase chain reaction within 14 days prior to the date of admission or during the admission. The minimum length of time from admission to discharge or death was 4 hours for AM/CA prediction and 72 hours for prediction of TEs, the difference being necessitated by the time granularity with which each outcome could be identified. Data were censored at the time of outcome or discharge.

Additional exclusion criteria were applied for prediction of each outcome separately. Patients were excluded from TE prediction if they experienced an imaging-confirmed TE or were suspected of

experiencing a TE immediately prior to admission, which was diagnosed on admission or within 24 hours of admission. For prediction of AM/CA, patients were excluded if they experienced cardiac arrest with return of spontaneous circulation immediately prior to admission or if the arrest was precipitated by an event not related to disease severity. These exclusion criteria mean that the scope of the predictor is limited to new, in-hospital events.

For prediction of both outcomes, patients were not excluded based on treatments received, disease severity, need for intensive care, missing clinical variables, or any other reason not listed. Although excluding patients for these reasons may have improved the ML models' performance, this would have resulted in a "clean" cohort not representative of real clinical data, making the risk predictor less useful in a real-world clinical setting. Outcome definition is discussed in Supplemental Methods.

MODEL SPECIFICATION. **Figure 1** presents a schematic of the COVID-HEART continuously updating risk predictor. The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines for development, validation, and presentation of a multivariable prediction model[24] were followed here (Supplemental Table 1). The model uses a selection of features extracted from up to 127 different clinical data inputs (shown schematically in **Figure 1A** and presented in detail in Supplemental Table 2), some of which are associated with CV complications in COVID-19 and in other severe respiratory illnesses and others with general physiologic function. To avoid bias, variables that were directly impacted by a physician's assessment of the patient's condition, such as the fraction of inspired oxygen set on a mechanical ventilator, are excluded. Definition of these predictors, how they were measured, and algorithmic preprocessing steps undertaken prior to dynamic feature extraction are provided in Supplemental Methods. No manual adjustments were made to raw clinical data, and algorithmic preprocessing steps were designed to apply the minimum necessary "corrections" to the raw clinical data inputs to ensure our development and validation data sets were realistic and that our model could be applied in a real-world clinical setting. Preprocessing steps included removal of features which were missing for >60% of time windows, mean value imputation for missing numerical features, and scaling all numerical features to zero mean and unit variance.

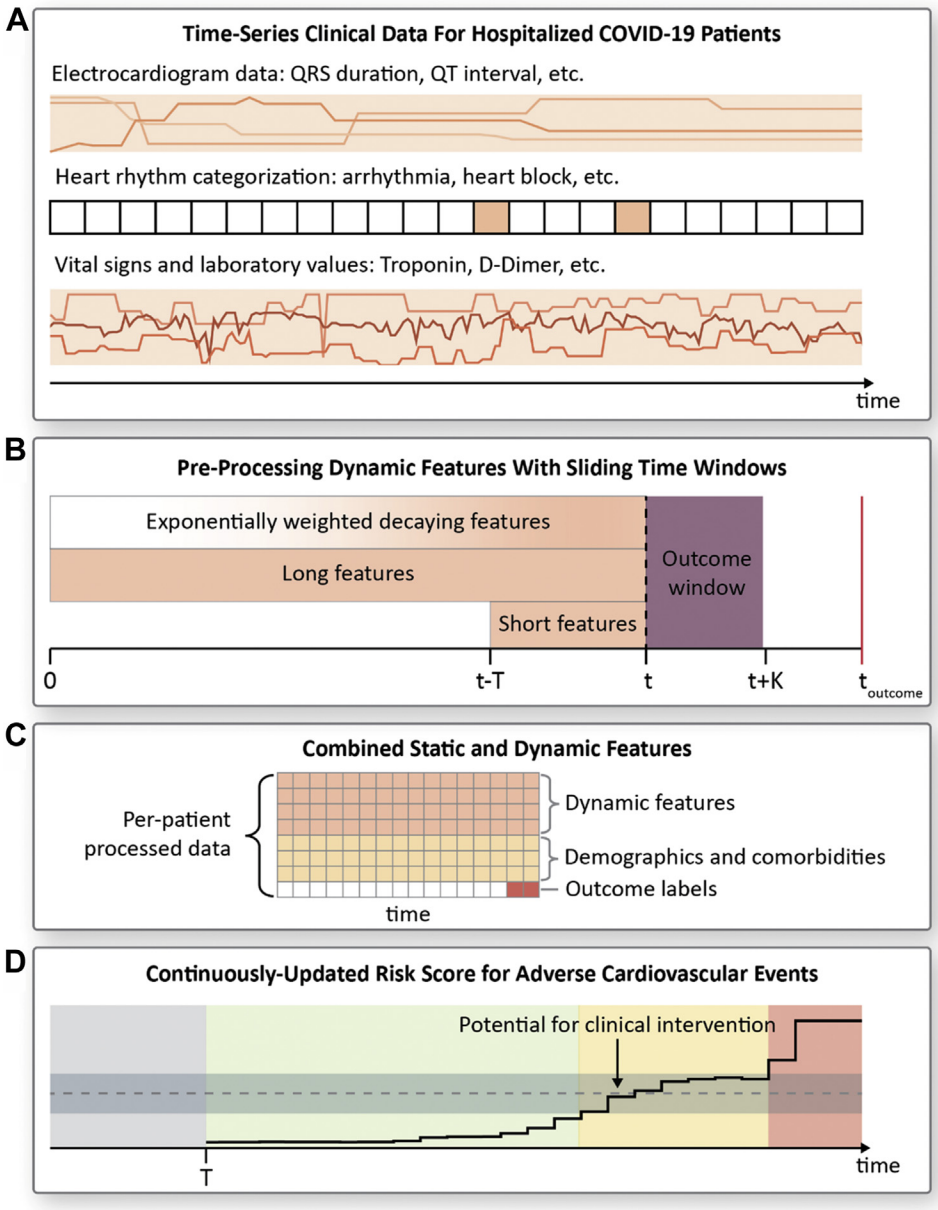
The COVID-HEART predictor was trained to estimate the probability that a patient will experience a particular event within a set number of hours (outcome window) after any point during the patient's hospitalization. It used static variables (demographics and comorbidities) and dynamic clinical data collected during time periods of markedly different duration prior to the time point of prediction. Dynamic features were calculated from the processed time-series clinical data inputs as illustrated in **Figure 1B**. Each time point was assigned a binary outcome label indicating whether the patient experienced the outcome of interest in an "outcome window" following the time point. **Figure 1C** schematically shows an array of processed data for a patient who experienced an adverse CV event. The outcome window for prediction of TEs was 24 hours as this was the minimum interval in which outcomes could be identified. Two hours was selected as the outcome window for prediction of AM/CA based on practical clinical considerations—this would provide health care personnel sufficient time for intervention if indicated. Multiple classifier configurations were investigated for prediction of each outcome, 2 linear and 1 nonlinear, detailed in Supplemental Methods.

CLASSIFIER DEVELOPMENT, OPTIMIZATION, AND TESTING. Eligible patients were divided into development and test sets according to the date of their first admission. The cutoff date was selected such that the development set for each outcome included 70% of eligible patients. Patients in the development set for prediction of AM/CA were admitted between March 1, 2020, and November 6, 2020; patients in the test set were admitted between November 7, 2020, and January 8, 2021. The cutoff date for prediction of TEs was November 5, 2020. Data collection ended on the respective cutoff dates for each set.

Classifier development began with 5-fold stratified patient-based cross-validation using the development set. We repeated this 20 times for each of the classifier configurations, each time progressively reducing the number of patients used for training and optimization from the full development set by moving the end cutoff date back 1 week (eg, November 6, October 30, October 23). At no point did the reduced training set include any patients from the separate test set. Hyperparameters were optimized through cross-validation with a Bayesian hyperparameter search strategy, and the optimal classifier configurations were selected based on the aggregated cross-validation area under the receiver operating characteristic curve (AUROC).

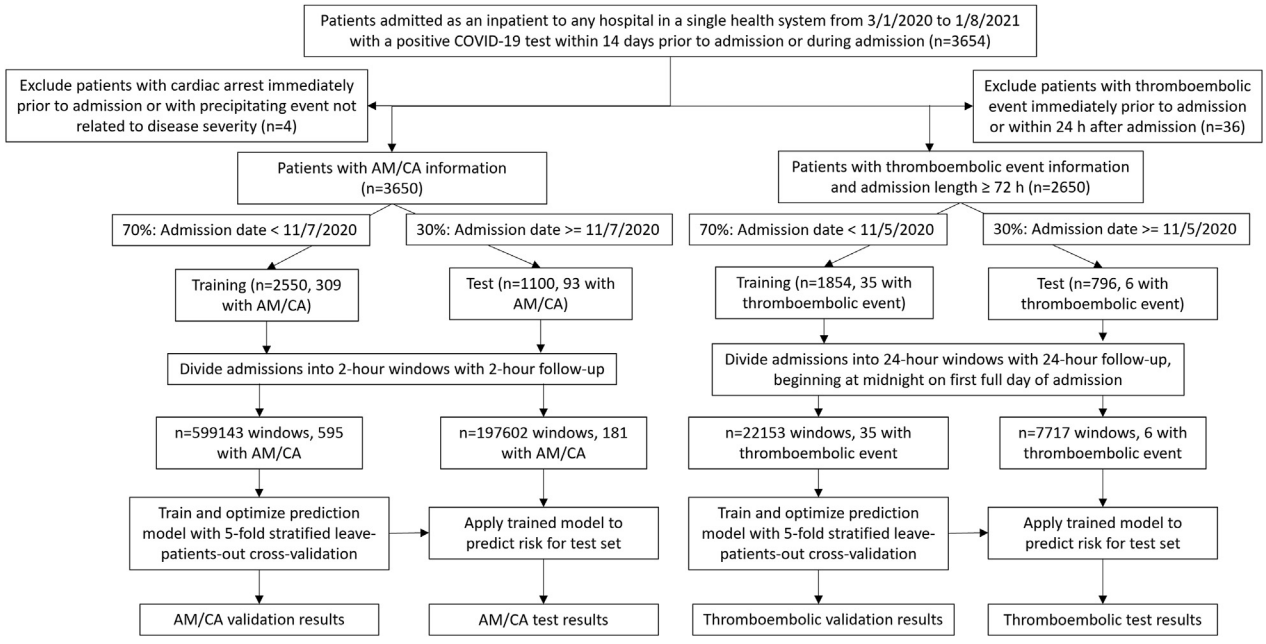
Following training and cross-validation of each classifier configuration for prediction of each outcome with the development set, the optimal

FIGURE 1 Schematic Overview of COVID-HEART Study



(A) Time-series clinical data used as input. Data shown here are representative and do not correspond with the risk score shown in **D**. **(B)** Dynamic features preprocessing with sliding time windows. Relative intensity levels within the 3 feature windows represent the weighting of values at each time; **darker colors** indicate higher weight. **(C)** Combined features. For each time window, the processed dynamic features are combined with static features including demographics and comorbidities. Outcome labels are assigned per-window. **(D)** Continuously-updating risk score. The COVID-HEART predictor provides a risk score (probability) for a given outcome in the K hours following a given time point. Shown is a sample risk score for a patient that experienced an event: **green** indicates a low risk score; **yellow** indicates a risk score within a predetermined range of a threshold value, and **red** indicates that the patient is at high risk for an event in the following K hours.

FIGURE 2 Participant Flow Diagram for Retrospective Study of COVID-HEART



Inclusion and exclusion criteria were applied separately for prediction of each outcome. The data were then temporally divided into development and test sets as shown. AM/CA = all-cause mortality/cardiac arrest.

classifier configuration was trained on the full development set and used to predict the time-series risk of each event for each patient in the respective temporally divided test set. A binary prediction was also made at each time point using the optimal threshold determined by the development data during training. Model performance was assessed by the following metrics: accuracy, balanced accuracy, sensitivity, specificity, and AUROC. As a secondary analysis, the number of time windows predicted positive for patients who eventually experienced events and for patients who did not were compared. Additional analyses to investigate the effects of missing features and the frequency of new clinical data measurements on testing performance were also performed.

Testing was repeated to obtain a 95% confidence interval for each testing performance metric using the final optimized model from each of the 20 iterations of cross-validation. To maintain the temporal nature of the development-test split, we selected an end cutoff date for the test set such that the development and test sets contained 70% and 30% of patients in the reduced data set, respectively. The earliest train-test cutoff date was June 25, 2020; we did not move the train-test cutoff beyond this date to ensure there were enough data to train the predictor. Since there were few events for each outcome, repeating the train-test split in this way provided an accurate estimate of the models' cross-validation performance and performance on a temporally separate test set. All test patient example predictions and data describing the characteristics of the development and testing sets were generated using the model trained with the full development and testing sets (March 1, 2020, to January 8, 2021).

Finally, to assess the predictor's performance when trained and tested with data of patients from different populations, we performed leave-hospital-out validation. This is justified by the fact that each of the 5 hospitals in the study has different characteristics and serves a different patient population (Supplemental Table 3). Leave-hospital-out validation was performed by removing all patients admitted to 1 of the 5 hospitals in the study, repeating the model training and optimization process using data from patients admitted to the remaining 4 hospitals, and testing the optimized model with data from

patients admitted to the left-out hospital. If a patient was transferred between hospitals or had multiple admissions to different hospitals, their admission to the left-out hospital was used in testing, and the rest of their data were removed from the training data set.

RESULTS

In total, 3,650 patients met eligibility criteria for prediction of AM/CA; 1,100 (30.1%) were assigned to the test set according to the date cutoff. In addition, 2,650 patients met eligibility criteria for prediction of TEs; 796 (30.0%) were assigned to the test set. **Figure 2** shows the flow of patients through the study. Supplemental Tables 4 and 5 provide demographic and clinical comparisons between patients who did and did not experience each outcome and between the development and test sets. Overall, 402 out of 3,650 patients (11.0%) experienced AM/CA, 26 of whom had subsequent return to spontaneous circulation. Of these, 18 occurred in the intensive care unit (ICU), 3 occurred in a non-ICU inpatient unit, 4 occurred in intermediate care/stepdown, and 1 occurred in long-term inpatient recovery care. Forty-one out of 2,650 (1.5%) eligible patients experienced imaging-confirmed TEs. Thirty-six additional patients either had an imaging-confirmed TE within 24 hours of admission or had clinical suspicion of a recent history of TEs prior to admission and were excluded for those reasons.

COVID-HEART performance for the 2 outcomes, in-hospital AM/CA and TEs, is summarized in **Figure 3**. Plots of the aggregated cross-validation AUROC are shown in **Figure 3A**. Linear models were optimal for prediction of both outcomes and included all features for prediction of AM/CA and short features only for prediction of TEs. The optimized COVID-HEART predictor achieved AUROCs of 0.918 and 0.771, sensitivities of 0.768 and 0.500, and specificities of 0.903 and 0.879 for the full test set for prediction of AM/CA and TEs, respectively (**Figure 3B**).

Following the initial development-test split, the results of which are further presented in **Figure 4** and Supplemental Table 6, the temporal development-test split was repeated, and results over 20 iterations were aggregated to obtain 95% confidence intervals for the performance metrics (**Figures 3C to 3E**). Mean cross-validation and test AUROCs were 0.917 (95% CI: 0.916-0.919) and 0.923 (95% CI: 0.918-0.927) for prediction of AM/CA and 0.757 (95% CI: 0.751-0.763) and 0.790 (95% CI: 0.756-0.824) for prediction of TEs, respectively.

Supplemental Table 7 presents leave-hospital-out cross-validation and testing results. For prediction of AM/CA, the mean test AUROC, sensitivity, and specificity for the left-out hospitals were 0.956 (95% CI: 0.936-0.976), 0.885 (95% CI: 0.838-0.933), and 0.887 (95% CI: 0.843-0.932), respectively. For prediction of imaging-confirmed TEs, the mean test AUROC, sensitivity, and specificity for the left-out hospitals were 0.781 (95% CI: 0.642-0.919), 0.453 (95% CI: 0.147-0.760), and 0.863 (95% CI: 0.822-0.904), respectively. There were 4 hospitals in the study at which fewer than 10 imaging-confirmed TEs were recorded, resulting in a wide confidence interval for sensitivity.

Supplemental Figure 1 illustrates the COVID-HEART's capability to accurately predict each outcome within outcome windows of different durations. This capability may provide significant clinical value in determining the patient's short-term and longer term risk, thus ensuring appropriate intervention and resource allocation. As the figures illustrate, cross-validation and test results are comparable, indicating the strong generalizability of the COVID-HEART despite statistically significant differences in demographics and prevalence of comorbidities between the development and test sets (Supplemental Table 5). **Figure 4** and Supplemental Figure 2 provide examples of time-series clinical data and resulting risk scores for "true-positive" and "true-negative" predictions for patients in the test set for each CV outcome. Supplemental Figure 3 illustrates 2 incorrect predictions; these are discussed in Supplemental Results.

For both outcomes, a larger number of time windows in the test set were predicted positive for patients that eventually experienced the outcome compared to those that did not: 38% vs 10% for AM/CA, 51% vs 12% for TEs. The 95% confidence intervals for these measurements over 20 iterations of temporally divided testing were 36% to 41% vs 9% to 11% for AM/CA and 68% to 82% vs 15% to 20% for TEs. This suggests that the ML model is sensitive in identifying warning signs of an impending adverse event earlier than the prespecified outcome window (Supplemental Figure 4). The interquartile ranges for the median early warning times over 20 iterations of temporally divided testing were 14 to 21 hours for AM/CA and 12 to 60 hours for TEs although the classifier was trained to predict outcomes within 2 hours for AM/CA and 24 hours for TEs. This could represent a clinically useful "early warning" system.

As it is essential for clinical decision-making to identify the features that most contribute to the predicted risk score for a particular outcome, the COVID-HEART predictor was designed to be fully

FIGURE 3 **The COVID-HEART Predictor Can Accurately Predict the Risk of All-Cause Mortality/Cardiac Arrest and Thromboembolic Events in Real Time**

A Validation (Development Set)

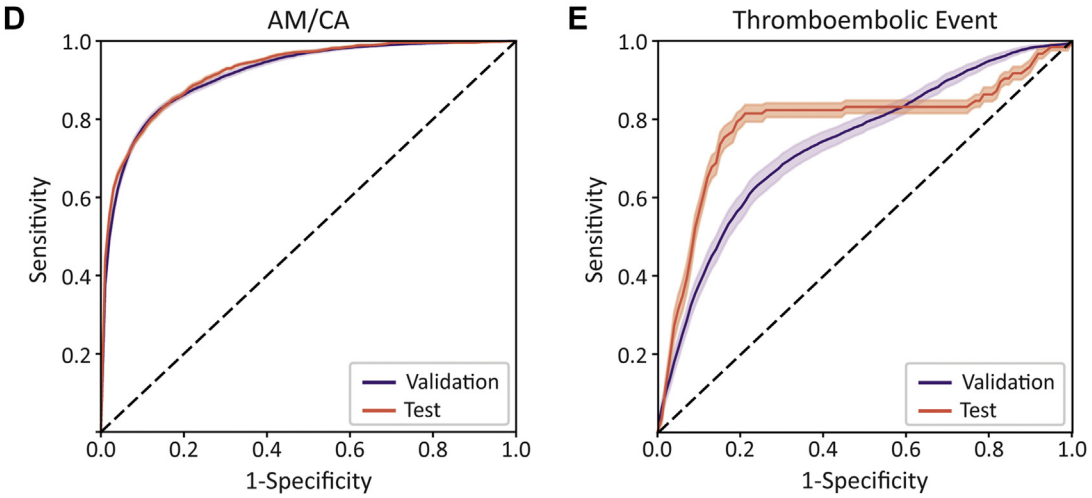
Outcome	Model Type	Dynamic Features	AUROC	Sensitivity	Specificity
AM/CA	Linear Classifier	Short only	0.913 [0.912, 0.915]	0.796 [0.788, 0.803]	0.875 [0.872, 0.877]
	Linear Classifier	**All**	0.917 [0.916, 0.919]	0.813 [0.811, 0.815]	0.869 [0.867, 0.871]
	Multilayer Perceptron	All	0.915 [0.914, 0.916]	0.800 [0.797, 0.803]	0.873 [0.871, 0.876]
Thromboembolic Event	**Linear Classifier**	**Short only**	0.757 [0.751, 0.763]	0.614 [0.597, 0.631]	0.777 [0.772, 0.782]
	Linear Classifier	All	0.745 [0.740, 0.750]	0.593 [0.579, 0.608]	0.768 [0.760, 0.776]
	Multilayer Perceptron	All	0.751 [0.746, 0.756]	0.569 [0.553, 0.584]	0.794 [0.790, 0.799]

B Test (Test Set)

Outcome	Model Type	Dynamic Features	AUROC	Sensitivity	Specificity
AM/CA	Linear Classifier	All	0.918	0.768	0.903
Thromboembolic Event	Linear Classifier	Short only	0.771	0.500	0.879

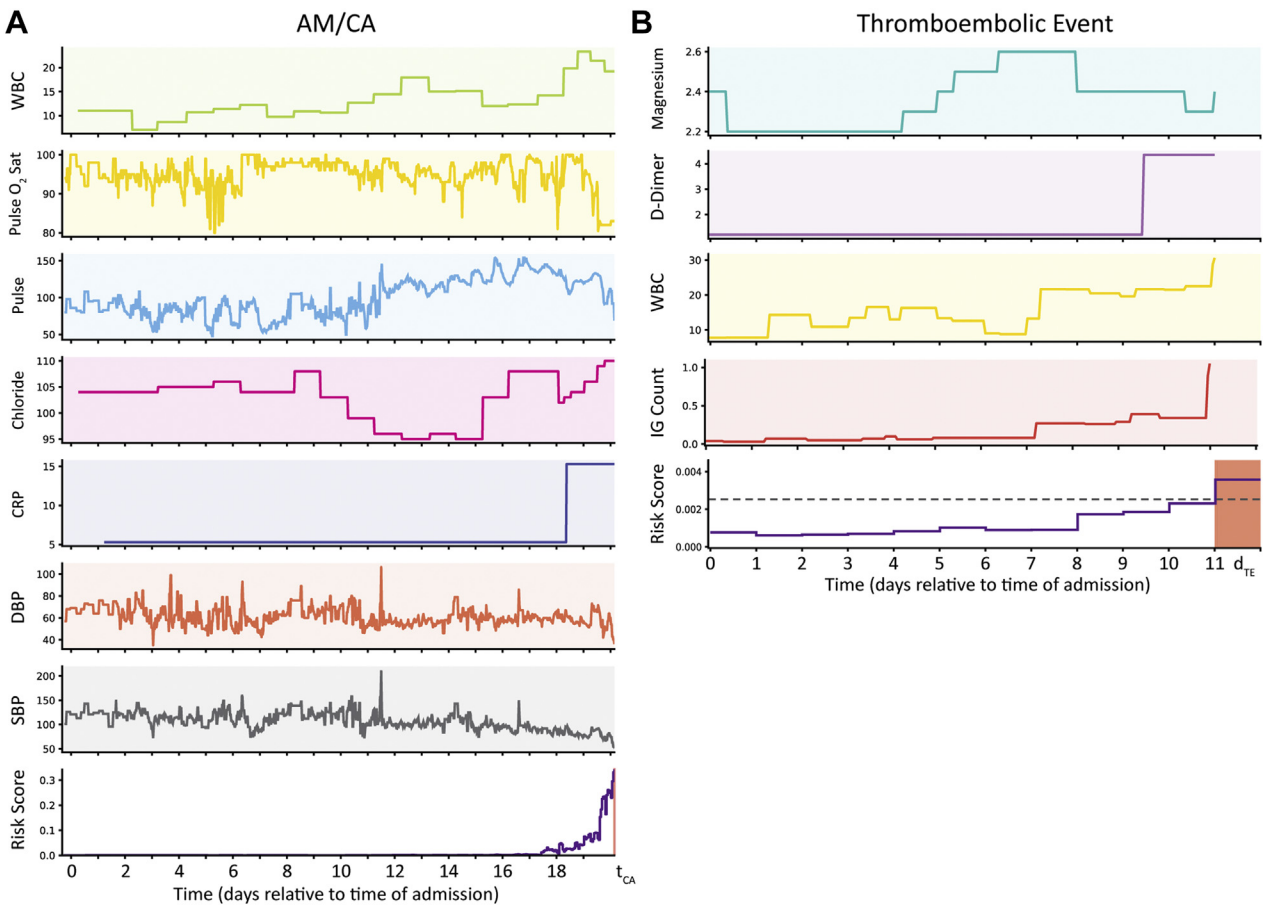
C Test (Repeated Temporal Split)

Outcome	Model Type	Dynamic Features	AUROC	Sensitivity	Specificity
AM/CA	Linear Classifier	All	0.923 [0.918, 0.927]	0.772 [0.750, 0.795]	0.896 [0.886, 0.906]
Thromboembolic Event	Linear Classifier	Short only	0.790 [0.756, 0.824]	0.740 [0.678, 0.801]	0.822 [0.799, 0.844]



(A) COVID-HEART 5-fold cross-validation performance metrics for AM/CA and thromboembolic events. Values shown are the mean [95% confidence interval] for each metric over 20 full iterations of cross-validation. AM/CA predictions presented here are for an outcome window of 2 hours, short-time feature window of 2 hours, and time-step of 1 hour. Thromboembolic event predictions shown here are for an outcome window of 24 hours, short-time feature window of 24 hours, and time-step of 24 hours. **(B)** COVID-HEART test performance metrics for temporally divided test set. Characteristics of this set are provided in Supplemental Table 5. **(C)** COVID-HEART test performance metrics over 20 iterations of repeated temporally divided testing. **(D)** Risk of cardiac arrest prediction. Cross-validation (purple) and testing (orange) receiver operating characteristic (ROC) curves for prediction of AM/CA using the optimal classifier configuration: a linear classifier with all feature types. To generate the ROC curves, 20 iterations of 5-fold temporal patient-based cross-validation were run resulting in a total of 20 test sets and 100 internal loops of cross-validation. Shaded regions represent the 95% confidence interval of each ROC curve. **(E)** Risk of thromboembolic event prediction. AM/CA = all-cause mortality/cardiac arrest; AUROC = area under the receiver operating characteristic curve.

FIGURE 4 Examples of "True-Positive" Predictions for 2 Different Patients: 1 From the All-Cause Mortality/Cardiac Arrest Test Set and 1 From the Thromboembolic Event Test Set



(A) Clinical time-series inputs **(top 7 rows)** from which the features with the largest coefficients were derived for prediction of AM/CA, and time-series risk score **(bottom row)** for a patient who experienced AM/CA during their hospitalization, and for whom the classifier's prediction was correct prior to the AM/CA. A new prediction is generated every hour. The binary risk threshold is 0.0008; the red bar indicates the hour during which the patient experienced AM/CA. Units for each predictor are as follows: WBC (cells/mm^3), pulse O_2 saturation (%), pulse (beats/min), chloride (mEq/L), CRP (mg/L), DBP (mm Hg), SBP (mm Hg). **(B)** Clinical time-series inputs **(top 4 rows)** from which the selected features were derived for prediction of thromboembolic events, and time-series risk score **(bottom row)** for a patient who experienced a thromboembolic event during their hospitalization. A new prediction is generated every 24 hours. Units for each predictor are as follows: magnesium (mEq/L), D-dimer (nmol/L), WBC (cells/mm^3), IG count (%). AM/CA = all-cause mortality/cardiac arrest; CRP = C-reactive protein; DBP = diastolic blood pressure; IG = immature granulocyte; SBP = systolic blood pressure; WBC = white blood cell count.

interpretable. Supplemental Table 6 lists up to 20 features with the largest coefficients in the optimal classifier for each of the 2 outcomes. The final COVID-HEART predictor includes 61 features for prediction of AM/CA, many of which are routinely and continuously acquired vital signs and basic metabolic tests. COVID-HEART includes 9 features for prediction of TEs. These features are extracted from 39 and 5 unique clinical data inputs, for the 2 models, respectively. Note that features were normalized prior to classifier training, and that models are not simple logistic regressions, thus interpretation of the coefficients is not straightforward. Many of these features confirm previous observations in cohorts of severely ill COVID-19 patients. For example, lower O_2 saturation[14] is associated with AM/CA, and multiple coagulation-related lab results are associated with TEs.[25,26] Finally, the effects of missing variables on testing results are presented in Supplemental Figure 5 and are discussed in Supplemental Results.

DISCUSSION

In this study, we developed and validated the COVID-HEART predictor (**Central Illustration**), a real-time model that can forecast multiple adverse events in hospitalized patients with COVID-19. The COVID-HEART predictor is robust to missing data and can be updated each time new data become available, representing a continuously evolving warning system for an impending event. It can also predict the likelihood of an adverse event within multiple timeframes (eg, 2 hours, 8 hours, 24 hours). Although predictions were made for patients in the test set at the same time steps as in the development set for consistency, it is possible to apply the model at any arbitrary time during a patient's hospitalization. We envision that in practice, it could provide the physician with an updated risk score each time any new clinical data input becomes available or only after passing a certain "high-risk" threshold, to reduce health care provider "alert fatigue." The COVID-HEART predictor is thus anticipated to be of great clinical use in triaging patients and optimizing resource utilization by identifying at-risk patients in real time. Finally, COVID-HEART identifies dynamic predictive features that have not previously been investigated for prediction of these outcomes in patients with COVID-19; these may suggest avenues for future research and personalized targets for clinical intervention.

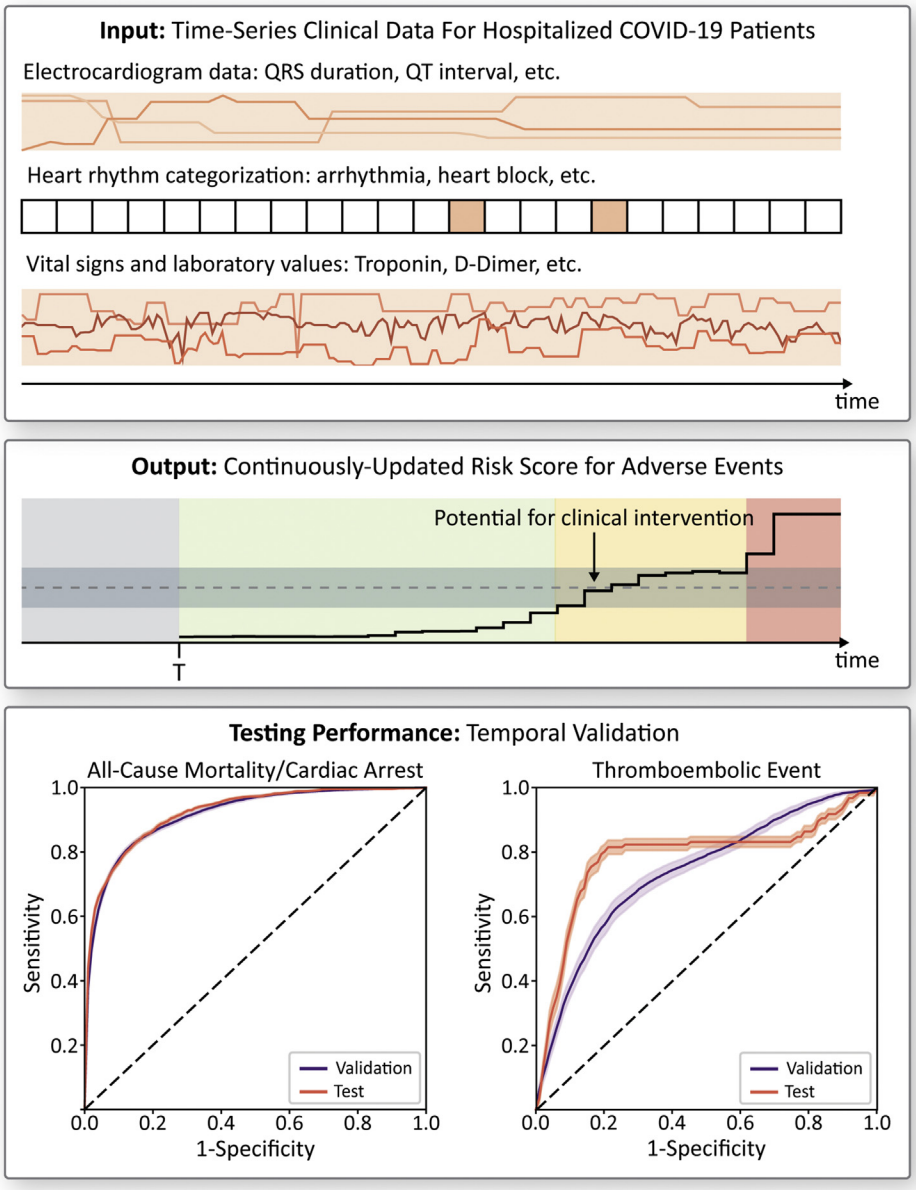
The COVID-HEART risk-prediction approach provides transparency and clinical intepretability, including the ability to determine which features are dominant contributors to a patient's risk level at a particular time, which may suggest potential patient-specific targets for clinical intervention. Prediction models for CV adverse events in patients with COVID-19 have been limited by lack of sufficient data, impractical requirements for use (eg, that all data be available for all patients or that measurements are taken at the same time relative to time of admission), and overly restrictive inclusion/exclusion criteria that result in idealistic training and testing cohorts not representative of real patient data.[23,27] Our model is designed to handle real-world data, which may include noise, missing variables, and data collected at different points in a patient's hospitalization. The validation and test results indicate strong generalizability despite statistically significant differences between the temporally divided development and test sets and between hospitals in the health system. Finally, the inclusion of multiple time-duration features gives the model the "memory" advantages of a long short-term memory neural network without compromising interpretability or becoming a "black box." It is trained in a manner that achieves high sensitivity and specificity despite severe class imbalance. To our knowledge, these techniques have not previously been combined with real-time predictors for CV events.

Models for risk prediction in hospitalized patients have typically focused on predicting mortality risk or length of stay for patients in the ICU. Traditional models incorporate variables thought to indicate physiologic instability or end-organ injury (eg, respiratory rate, serum bilirubin level, serum creatinine, etc).[27-29] While these models generally have good discriminative power,[30] they fail to provide specific, actionable information and simply notify health care teams that particular patients are at increased mortality risk at some point in their ICU stay. In most cases, predictive scores are calculated based on the most extreme variable values during the initial 24 hours of the ICU admission, with repeat calculations every 24 to 72 hours.

Newer models have higher predictive performance than traditional models; they are trained to predict the incidence of a particular outcome (eg, bleeding, renal failure, mortality, etc) at an indefinite future time. They are not designed to predict the time periods during which patients are at the highest risk. Furthermore, in terms of ML for risk prediction in COVID-19, prior studies have focused largely on initial diagnosis, mortality, or severity of illness, but none have specifically focused on CV events, including in-hospital AM/CA and TEs, both clinically important complications with implication for cardiac treatment and monitoring. Moreover, to our knowledge, our model is the first to utilize continuous time-series physiologic data as well as laboratory and electrocardiographic data to provide a continuously updating risk score for an outcome within a particular future time window (eg, risk of TE in the next 24 hours). By providing a risk score for a specific outcome window, our model provides timely, actionable information, allowing the health care team to allocate resources and initiate therapies when they are most needed.

With respect to TEs, we found that 40 of 41 events occurred in patients already ordered for high-intensity venous thromboembolism (VTE) prophylaxis, suggesting an even more aggressive anticoagulant regimen may be needed for those patients identified by the model. Additionally, VTE prophylaxis is 1 of the treatments most frequently omitted by nursing staff or declined by patients. An analysis of VTE events at our institution over a 72-day period during the Spring 2020 COVID-19 wave demonstrated

CENTRAL ILLUSTRATION Overview of the COVID-HEART Study



Shade JK, et al. JACC Adv. 2022;1(2):100043.

We developed and validated an interpretable predictor of all-cause mortality/cardiac arrest and thromboembolic events in COVID-19 using retrospective registry data from over 3,600 patients admitted to multiple hospitals. It achieved AUROCs of 0.917 (95% CI: 0.916-0.919) and 0.757 (95% CI: 0.751-0.763) for prediction of all-cause mortality/cardiac arrest and thromboembolic events, respectively. The predictor can facilitate triage and resource allocation by providing real-time risk scores for complications that commonly occur in COVID-19 patients. It could be retrained to predict additional cardiovascular events, or adverse cardiovascular events after COVID. The methodology could be extended to clinical scenarios that require screening or early detection. AUROC = area under the receiver operating characteristic curve; CI = confidence interval.

that 4 of 11 SARS-CoV-2 positive patients who experienced VTE events had at least 1 missed dose of VTE prophylaxis.[31] While care providers should ideally strive for 100% compliance with VTE prophylaxis in all eligible patients, the identification of patients at high risk for TEs may help target these interventions to the patients most in need.

With respect to interventions to address impending AM/CA, we found in our detailed chart review that a number of AM/CA events were not unprovoked but were a consequence of a precipitating event that altered the patients' hemodynamics, such as intubation, patient positioning (eg, supine to prone), or hemodialysis. Therefore, in addition to predicting unprovoked arrest (in approximately half of the cases), our model predicted an unstable physiologic state that resulted in arrest due to otherwise well-tolerated hemodynamic perturbations. Identification of patients as high risk for mortality would aid clinicians by imploring them to defer any treatments that may provoke an arrest until the patient's physiology recovers. For those treatments that cannot be deferred, identification of high-risk patients would prompt the primary team to assemble specialized staff and equipment, given the high risk of arrest (eg, calling the anesthesia team for intubation in a high-risk patient, having adequate nursing staff for a possible resuscitation, etc).

A major barrier to clinical adoption of prognostic ML models is the lack of appropriate validation on a representative test cohort. The temporally divided test sets in this study demonstrated the performance of the predictor in a set of patients in the development set admitted after the end of data collection. A prospective cohort would not be expected to have the same composition as the development set; indeed, there were several statistically significant differences in demographics, clinical characteristics, and prevalence of adverse CV events between the development and test sets in this study. However, the strong test results show that the predictor is robust to changes in clinical treatment guidelines and evolving demographics. We hypothesize that it maintains its accuracy because it considers data which describe the patient's physiologic state, not variables that are directly influenced by physician input such as ventilator settings or medication use. Furthermore, the predictor maintained strong performance in leave-hospital-out validation, which demonstrated its robustness when trained and tested with data of patients from different populations.

STUDY LIMITATIONS. A limitation in this study is the requirement for imaging confirmation of TEs. All TE diagnoses were adjudicated by a clinician to ensure they were clinically relevant. If the radiologist made an incorrect diagnosis and the adjudicating clinician incorrectly agreed that the event was supported by clinical evidence, this would unfortunately constitute an error in our data set. Similarly, it is likely that patients in the study experienced TEs that were either the precipitating cause of death or that were not identified on imaging and were therefore not counted as events. There were only 35 patients in the development set with imaging-confirmed TEs, and these outcomes could only be identified per-day, not at the exact time they occurred, as with AM/CA. As a result, only a few features could be selected; it is possible that a larger feature set would lead to more accurate prediction of the patients' risk of TEs since more details of the patients' clinical states could be considered. In addition, complete details about the primary causes of death were not known for all patients, and therefore, it was not possible to distinguish if AM was secondary to CV, respiratory, or other causes.

Additional limitations stem from the use of the JH-CROWN registry.[32] These include the potential for measurement error, inaccurate patient-reported history (eg, smoking), and missing data. Another potential limitation is confounding by indication, which means that treatments were selected based on clinical indication. While our model did not include treatments or other variables that were directly influenced by clinical indication, some variables in the model were likely indirectly influenced by clinical indication. For example, the pulse oxygen saturation may have been affected by changes in ventilator settings for patients who were receiving mechanical ventilation. There is also a subgroup of patients who had pre-existing do not resuscitate/do not intubate/comfort care orders. These patients would have received no interventions leading up to an adverse event, which means that the sequalae of physiologic changes for these patients may be different from those for patients who received interventions prior to an adverse event. Finally, there is selection bias inherent to including only patients who sought care at a hospital; patients without insurance, undocumented patients, and patients with other barriers to seeking care may be less likely to be included.

CONCLUSIONS

In this study, we demonstrated highly accurate prediction of AM/CA and TE in hospitalized COVID-19 patients using the continuously updating COVID-HEART predictor. In its current implementation, the predictor can facilitate practical, meaningful changes

in patient triage and the allocation of resources by providing real-time risk scores for complications occurring commonly in COVID-19 patients. The COVID-HEART can be retrained to predict additional adverse CV events including myocardial infarction and arrhythmia. The potential utility of the predictor extends well beyond hospitalized COVID-19 patients, as COVID-HEART could be applied to the prediction of CV adverse events after hospital discharge or in patients with chronic COVID syndrome ("long COVID"). Additionally, the ML methodology utilized here could be expanded to use in other clinical scenarios that require screening or early detection, such as risk of hospital readmission, with the goal of improved clinical outcomes through early warnings and resultant opportunity for timely intervention.

ACKNOWLEDGMENTS The data utilized for this publication were from the Johns Hopkins COVID-19 Precision Medicine Center of Excellence, officially designated by the Precision Medicine Analytics Platform (PMAP) leadership, and based on the contribution of many patients and clinicians. This work would not be possible without the assistance of Dr Paul Nagy, Bonnie Woods, Michael Cook, Brandon Smith, Philip Gianuzzi, Brian Garibaldi, and all other clinicians and data scientists at Johns Hopkins University who contributed to the COVID data resource design and collation. The authors are grateful to Dr Vadim Zipunnikov for his helpful suggestions.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This study was funded by the National Science Foundation (NSF) through a RAPID grant (2029603) to Dr Trayanova and a Graduate Research Fellowship (DGE-1746891) to Dr Shade. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Natalia A. Trayanova, Johns Hopkins University, Hackerman 216, 3400 North Charles Street, Baltimore, Maryland 21218, USA. E-mail: ntrayanova@jhu.edu. Twitter: @ntrayanova.

PERSPECTIVES

COMPETENCY IN PRACTICE-BASED LEARNING

COMPETENCY IN PRACTICE-BASED LEARNING: The COVID-HEART predictor can identify patient at-risk AM/CA and TEs by quantitatively evaluating changes in dozens of clinical variables, enhancing clinical practice by providing data-driven clinical decision support.

TRANSLATION OUTLOOK: Clinical implementation of the algorithm would require a one-time engineering investment to convert the model and preprocessing algorithms into predictive model markup language. The model could then be fully integrated with an electronic health record system and would require no manual input or time investment by a clinician to calculate or view a patient's risk score and the clinical variables that most influenced the score. Prospective validation would be required to increase clinical confidence in the predictor, and a larger training data set would likely improve accuracy of TE prediction.

REFERENCES

1. Zheng Y-Y, Ma Y-T, Zhang J-Y, Xie X. COVID-19 and the cardiovascular system. *Nat Rev Cardiol*. 2020;17:259–260. https://doi.org/10.1038/s41569-020-0360-5

2. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. 2020;5:802–810. https://doi.org/10.1001/jamacardio.2020.0950

3. Bin SY, Heo JY, Song MS, et al. Environmental contamination and viral shedding in MERS patients during MERS-CoV outbreak in South Korea. *Clin Infect Dis*. 2015;62:755–760. https://doi.org/10.1093/cid/civ1020

4. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395:1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3

5. Puntmann VO, Carerj ML, Wieters I, et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol*. 2020;5:1265–1273. https://doi.org/10.1001/jamacardio.2020.3557

6. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost*. 2020;18:844–847. https://doi.org/10.1111/jth.14768

7. Malas MB, Naazie IN, Elsayed N, Mathlouthi A, Marmor R, Clary B. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: a systematic review and meta-analysis. *Lancet*. 2020;29-30:100639. https://doi.org/10.1016/j.eclinm.2020.100639

8. Aboughdir M, Kirwin T, Khader AA, Wang B. Prognostic value of cardiovascular biomarkers in COVID-19: a review. *Viruses*. 2020;12:527. https://doi.org/10.3390/v12050527

9. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Solitons Fractals*. 2020;139:110059.

10. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight Covid-19. *Physiol Genomics*. 2020;52:200–202. https://doi.org/10.1152/physiolgenomics.00029.2020

11. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med*. 2020;46:846–848. https://doi.org/10.1007/s00134-020-05991-x

12. Yan L, Zhang H-T, Goncalves J, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv*. Posted March 17, 2020. https://doi.org/10.1101/2020.02.27.20028027

13. Vaid A, Somani S, Russak AJ, et al. Machine learning to predict mortality and critical events in

COVID-19 positive New York city patients: a cohort study (preprint). *J Med Internet Res*. 2020;22:e24018. https://doi.org/10.2196/24018

14. Li X, Ge P, Zhu J, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ*. 2020;8:e10337. https://doi.org/10.7717/peerj.10337

15. Aktar S, Talukder A, Ahamad MM, et al. Machine learning and meta-analysis approach to identify patient comorbidities and symptoms that increased risk of mortality in COVID-19. *arXiv*. Submitted on August 21, 2020. https://doi.org/10.48550/arXiv.2008.12683

16. Di Castelnuovo A, Bonaccio M, Costanzo S, et al. Common cardiovascular risk factors and in-hospital mortality in 3,894 patients with COVID-19: survival analysis and machine learning-based findings from the multicentre Italian CORIST Study. *Nutr Metab Cardiovasc Dis*. 2020;30:1899–1913. https://doi.org/10.1016/j.numecd.2020.07.031

17. Cui W, Robins D, Finkelstein J. Unsupervised machine learning for the discovery of latent clusters in COVID-19 patients using electronic health records. *Stud Health Technol Inform*. 2020;272:1–4. https://doi.org/10.3233/SHTI200478

18. Wu H, Ruan W, Wang J, et al. Interpretable machine learning for COVID-19: an empirical study on severity prediction task. *arXiv*. Submitted September 30, 2020. https://doi.org/10.48550/arXiv.2010.02006

19. Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J*. 2020;56:2000775. https://doi.org/10.1183/13993003.00775-2020

20. Haimovich AD, Ravindra NG, Stoytchev S, et al. Development and validation of the quick COVID-19 severity index: a prognostic tool for early clinical decompensation. *Ann Emerg Med*. 2020;76:442–453.

21. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. https://doi.org/10.1136/bmj.m1328

22. Richards DM, Tweardy MKJ, Steinhubl SR, et al. Wearable sensor derived decompensation index for continuous remote monitoring of COVID-19 diagnosed patients. *NPJ Digit Med*. 2021;4:155. https://doi.org/10.1038/s41746-021-00527-z

23. van de Sande D, van Genderen ME, Rosman B, et al. Predicting thromboembolic complications in COVID-19 ICU patients using machine learning. *J Clin Transl Res*. 2020;6:179–186. https://doi.org/10.18053/jctres.06.202005.003

24. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–W73. https://doi.org/10.7326/M14-0698

25. Ribes A, Vardon-Bounes F, Mémier V, et al. Thromboembolic events and Covid-19. *Adv Biol Regul*. 2020;77:100735.

26. Maier CL, Truong AD, Auld SC, Polly DM, Tanksley C, Duncan A. COVID-19-associated hyperviscosity: a link between inflammation and thrombophilia? Caution against COVID-19 treatment. *Lancet*. 2020;395:1758–1759. https://doi.org/10.1016/S0140-6736(20)31209-5

27. Zimmerman A, Kalra D. Usefulness of machine learning in COVID-19 for the detection and prognosis of cardiovascular complications. *Rev Cardiovasc Med*. 2020;21:345–352. https://doi.org/10.31083/j.rcm.2020.03.120

28. Moreno RP, Metnitz PGH, Almeida E, et al. Saps 3 - from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med*. 2005;31:1345–1355. https://doi.org/10.1007/s00134-005-2763-5

29. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPMO-III). *Crit Care Med*. 2007;35:827–835. https://doi.org/10.1097/01.CCM.0000257337.63529.9F

30. Nassar AP, Mocelin AO, Nunes ALB, et al. Caution when using prognostic models: a prospective comparison of 3 recent prognostic models. *J Crit Care*. 2012;27:423.e1–423.e7. https://doi.org/10.1016/j.jcrc.2011.08.016

31. Varasteh M, Brandyn K, Owodunni OP, Kraus PS, Holzmueller CG. Nonadministration of pharmacologic venous thromboembolism prophylaxis is less common in hospitalized patients with COVID-19. *J Thromb Thrombolysis*. 2021;52:471–475. https://doi.org/10.1007/s11239-021-02384-9

32. Johns Hopkins Institute for Clinical and Translational Research. *JH-CROWN COVID-19 PMAP Registry*. 2020. https://ictr.johnshopkins.edu/covid-research-center/registry-dashboard/jh-crown/

KEY WORDS **big data, cardiac arrest, machine learning, SARS-CoV-2, thromboembolism**

APPENDIX **For supplemental Methods, tables, figures, and references, please see the online version of this paper.**