

Identifying risk of adverse outcomes in COVID-19 patients via artificial intelligence–powered analysis of 12-lead intake electrocardiogram

Arun R. Sridhar, MBBS, MPH,^{*1} Zih-Hua Chen (Amber), MS,[†] Jacob J. Mayfield, MD,^{*} Alison E. Fohner, PhD,[‡] Panagiotis Arvanitis, MD,[§] Sarah Atkinson, BS,^{*} Frieder Braunschweig, MD, PhD,^{||¶} Neal A. Chatterjee, MD, MSc,^{*} Alessio Falasca Zamponi, MD,[§] Gregory Johnson, PhD,[#] Sanika A. Joshi,[†] Mats C.H. Lassen, MD,^{**} Jeanne E. Poole, MD, FHRS,^{*} Christopher Rumer, BS,^{*} Kristoffer G. Skaarup, MD,^{**} Tor Biering-Sørensen, MD, PhD, MPH,^{**1} Carina Blomstrom-Lundqvist, MD, PhD,^{§1} Cecilia M. Linde, MD, PhD,^{||¶1} Mary M. Maleckar, PhD,^{††1} Patrick M. Boyle, PhD, FHRS^{††§§1}

From the ^{*}Division of Cardiology, University of Washington, Seattle, Washington, [†]Department of Bioengineering, University of Washington, Seattle, Washington, [‡]Department of Epidemiology, University of Washington, Seattle, Washington, [§]Department of Medical Science, Uppsala University, Uppsala, Sweden, ^{||}Department of Medicine, Karolinska Institutet, Stockholm, Sweden, [¶]Department of Cardiology, Karolinska University Hospital, Stockholm, Sweden, [#]Unaffiliated independent researcher, Seattle, Washington, ^{**}Department of Cardiology, Herlev & Gentofte University Hospital, Copenhagen University, Copenhagen, Denmark, ^{††}Department of Computational Physiology, Simula Research Laboratory, Oslo, Norway, ^{‡‡}Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, Washington, and ^{§§}Center for Cardiovascular Biology, University of Washington, Seattle, Washington.

BACKGROUND Adverse events in COVID-19 are difficult to predict. Risk stratification is encumbered by the need to protect healthcare workers. We hypothesize that artificial intelligence (AI) can help identify subtle signs of myocardial involvement in the 12-lead electrocardiogram (ECG), which could help predict complications.

OBJECTIVE Use intake ECGs from COVID-19 patients to train AI models to predict risk of mortality or major adverse cardiovascular events (MACE).

METHODS We studied intake ECGs from 1448 COVID-19 patients (60.5% male, aged 63.4 ± 16.9 years). Records were labeled by mortality (death vs discharge) or MACE (no events vs arrhythmic, heart failure [HF], or thromboembolic [TE] events), then used to train AI models; these were compared to conventional regression models developed using demographic and comorbidity data.

RESULTS A total of 245 (17.7%) patients died (67.3% male, aged 74.5 ± 14.4 years); 352 (24.4%) experienced at least 1 MACE (119 arrhythmic, 107 HF, 130 TE). AI models predicted mortality and MACE with area under the curve (AUC) values of 0.60 ± 0.05 and

0.55 ± 0.07 , respectively; these were comparable to AUC values for conventional models (0.73 ± 0.07 and 0.65 ± 0.10). There were no prominent temporal trends in mortality rate or MACE incidence in our cohort; holdout testing with data from after a cutoff date (June 9, 2020) did not degrade model performance.

CONCLUSION Using intake ECGs alone, our AI models had limited ability to predict hospitalized COVID-19 patients' risk of mortality or MACE. Our models' accuracy was comparable to that of conventional models built using more in-depth information, but translation to clinical use would require higher sensitivity and positive predictive value. In the future, we hope that mixed-input AI models utilizing both ECG and clinical data may be developed to enhance predictive accuracy.

KEYWORDS 12-lead ECG; COVID-19; Artificial intelligence; Deep learning; Arrhythmia; Heart failure prognosis; Mortality; Risk factors

(Cardiovascular Digital Health Journal 2022; ■:1–13) © 2021 Heart Rhythm Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

¹Drs Sridhar, Biering-Sørensen, Blomstrom-Lundqvist, Linde, Maleckar, and Boyle are site principal investigators. **Address reprint requests and correspondence:** Dr Arun R. Sridhar, Division of Cardiology, University of Washington, 1959 Pacific Street NE, Box 356422, Seattle, WA 98195, USA. E-mail address: arun11@cardiology.washington.edu; or Dr Patrick M. Boyle, Department of Bioengineering, University of Washington, Foege Hall N310H, 3720 15th Ave NE, Box 355061, Seattle, WA 98195, USA. E-mail address: pmjboyle@uw.edu.

KEY FINDINGS

- We demonstrate the feasibility and examine the potential effectiveness of a rapid, inexpensive triage tool for COVID-19 patients using a 12-lead electrocardiogram (ECG) at admission augmented by machine learning.
- When compared with the traditional statistical model incorporating extensive clinical data, our results using only the deep neural network-augmented ECG did not yield improved discrimination.
- Our model's low sensitivity and positive predictive value would be a barrier to successful deployment in clinical settings in its current form; our findings suggest that mixed-input models using both ECG data and clinical variables might be a viable path towards better predictive capabilities.

Introduction

Coronavirus disease 2019 (COVID-19) has now been documented in at least 230 million people worldwide, resulting in at least 4.7 million deaths.^{1–6} This affliction has been linked with a range of cardiovascular complications including arrhythmias, myocarditis, acute coronary syndrome, and thromboembolism.⁷ These *de novo* cardiovascular events, as well as pre-existing cardiovascular comorbidity, are linked to adverse outcomes in patients with COVID-19. Work in cardiomyocytes derived from human pluripotent stem cells showed that SARS-CoV-2 infection impaired electrophysiological and contractile function and led to widespread cell death, suggesting cardiovascular symptoms may be a direct consequence of cardiotoxicity.⁸ Atrial fibrillation (AF)⁹ is associated with increased mortality in infected patients, especially when it newly presents during hospitalization for COVID-19.⁵ Biomarker-detected myocardial injury,¹⁰ thromboembolism, and abnormal laboratory coagulopathy studies are frequent complications of SARS-CoV-2 infection and are also associated with worse outcomes.^{11–13} Early prediction of mortality and morbidity risk may improve patient management.

One of the greatest challenges during the pandemic, especially in resource-strained settings, has been early identification of individual patients at higher risk for adverse outcomes. Machine learning (ML) has the potential to facilitate triage by providing clinicians with useful risk stratification data to inform decisions regarding level of care and follow-up monitoring. The electrocardiogram (ECG), a rapid, inexpensive, and noninvasive diagnostic test suitable for repetitive recordings, is an ideal target for ML augmentation, especially in light of associations observed between ECG abnormalities and adverse COVID-19 outcomes.^{14,15} ECG ML applications have been shown to recognize subtle patterns in the electrical signals, imperceptible to human readers, that can be leveraged to predict and classify different arrhythmic conditions, including AF,¹⁶ and to screen for other cardiovascular conditions, including hypertrophic

cardiomyopathy,¹⁷ left ventricular systolic dysfunction,¹⁸ and aortic stenosis.¹⁹ ML networks have also demonstrated promise identifying clinically meaningful markers of prognosis, predicting both 1-year mortality²⁰ and incident cardiac arrest within 24 hours.²¹

In this study, we developed a deep neural network^{22,23} (DNN)-based ECG analysis system to predict the risk of adverse events in patients with COVID-19 solely using 12-lead intake ECGs recorded at the time of COVID-19 hospital admissions as input. We trained 2 DNNs, 1 to predict mortality and 1 to predict major adverse cardiovascular events (MACE, including arrhythmic events, new-onset heart failure, or thromboembolic complications). We report the performance of these DNN systems, which use the ECGs alone to predict these outcomes of interest, and compare these to a conventional statistical model, which uses a broad set of demographics, comorbidities, and clinical variables to make the same predictions.

Methods

Collaboration setup

We enrolled patients from the University of Washington Healthcare System (UW; Seattle, WA), the Karolinska Institutet University Hospital (KI; Stockholm, Sweden), the Uppsala University Hospital (UU; Uppsala, Sweden), and the Copenhagen University Hospital (UC; Copenhagen, Denmark). Data collection was retrospective for all centers except UC, where collection was prospective, but data were analyzed retrospectively. This study was approved by the Institutional Review Board of the UW, the Swedish Ethical Review Authority, and the regional Danish Ethical Committee. Associated reference numbers are as follows: IRB6878 (UW), 2020-02627 (KI), 2020-02662 (UU), and H-20021500 (UC).

Dataset design

We first identified 1448 patients who were admitted to 1 of the 4 study centers between March 2, 2020, and February 28, 2021 for COVID-19; those who underwent an ECG recording during hospital intake were included. All included patients had laboratory-confirmed COVID-19 diagnosis (ICD-10-CM U07.1). Each center identified records from March 2, 2020, to June 8, 2020. UW, which was the main center for the study, continued extracting records through February 28, 2021. ECGs were adjudicated to exclude cases with pacing artifacts caused by cardiac implantable electronic devices or unacceptable quality (owing to missing leads or significant noise); in total, 62 records were excluded owing to these criteria (24 pacing artifacts; 38 poor quality). The remaining 1386 hospitalized COVID-19 patients (60.5% male, 63.4 ± 16.9 years old) comprised our final dataset. Standard 10-second 12-lead ECGs and inpatient outcomes (see next subsection) were collected for the study cohort, along with demographics, comorbidities, hospitalization variables, and laboratory values (Table 1 includes details of variables collected). ECGs were acquired using GE ECG systems

Table 1 Demographics, comorbidities, and outcomes, overall and by enrolling hospital system

	Overall	UW	KI	UU	UC	P value	Missing (%)
N	1386	420	481	308	177		
Demographics							
Age (years), mean (SD)	63.43 (16.90)	63.26 (16.41)	59.91 (17.04)	65.88 (17.96)	69.15 (13.24)	<.001	0.0
Sex at birth = female, n (%)	547 (39.5)	169 (40.2)	161 (33.5)	134 (43.5)	83 (46.9)	.004	0.0
BMI (kg/m ²), mean (SD)	28.38 (6.58)	29.70 (7.81)	28.63 (5.96)	27.85 (6.90)	27.06 (5.79)	.001	20.8
Ethnicity, n (%)						-	13.1
Hispanic/Latinx	85 (7.1)	84 (20.2)	0 (0.0)	1 (0.3)	-		
Non-Hispanic/Latinx	609 (50.5)	327 (78.6)	13 (2.7)	269 (87.3)	-		
Unknown / unavailable	511 (42.4)	5 (1.2)	468 (97.3)	38 (12.3)	-		
Race, n (%)							
FN/AK Native	9 (0.6)	9 (2.1)	0 (0.0)	0 (0.0)	-	<.001	0.0
Asian	118 (8.5)	71 (16.9)	7 (1.5)	40 (13.0)	-	<.001	0.0
Black/AA	74 (5.3)	57 (13.6)	11 (2.3)	6 (1.9)	-	<.001	0.0
HI FN/Pac Isl	7 (0.5)	7 (1.7)	0 (0.0)	0 (0.0)	-	.001	0.0
White	694 (50.1)	265 (63.1)	228 (47.4)	201 (65.3)	-	<.001	0.0
Other	57 (4.1)	3 (0.7)	50 (10.4)	4 (1.3)	-	<.001	0.0
Unknown / unavailable	252 (18.2)	8 (1.9)	187 (38.9)	57 (18.5)	-	<.001	0.0
Comorbidity							
Hypertension (%)	737 (54.0)	214 (53.2)	236 (49.3)	184 (59.7)	103 (58.2)	.021	1.4
CAD (%)	187 (13.8)	59 (14.8)	65 (13.7)	48 (15.6)	15 (8.5)	.144	2.1
CIED (%)						-	13.3
Pacemaker	22 (1.8)	9 (2.2)	8 (1.7)	5 (1.6)	-		
ICD	5 (0.4)	4 (1.0)	1 (0.2)	0 (0.0)	-		
Outcomes							
Arrhythmic event (%)	125 (9.0)	28 (6.7)	48 (10.0)	36 (11.7)	13 (7.3)	.084	0.0
TE (%)	132 (9.5)	36 (8.6)	52 (10.8)	29 (9.4)	15 (8.5)	.660	0.0
HF (%)	109 (7.9)	19 (4.5)	58 (12.1)	23 (7.5)	9 (5.1)	<.001	0.0
Mortality (%)	245 (17.7)	88 (21.0)	73 (15.2)	69 (22.4)	15 (8.5)	<.001	0.0

P values are for tests of differences between centers (continuous variables: ANOVA; categorical variables: χ^2).

AA = African American; BMI = body mass index; CAD = coronary artery disease; CIED = cardiac implanted electronic device; FN/AK Native = First Nations or Alaskan Native; HF = heart failure; HI FN/Pac Isl = Hawaiian First Nations / Pacific Islander; ICD = implanted cardioverter-defibrillator; KI = Karolinska Institutet; TE = thromboembolic event; UC = University of Copenhagen; UU = Uppsala University; UW = University of Washington.

(General Electric Company, Boston, MA) and raw data were managed using the GE MUSE Cardiology Information System. Of the 1386 intake ECGs, 1207 (95.0%) were acquired within 24 hours of the hospital admission date; for the remaining 179 patients, the median number of days between hospital admission and intake ECG acquisition was 4 (interquartile range: 3, 8). For patients who had multiple ECGs in the MUSE system from date of intake, the latest one was used in the dataset. All data were collected and managed using Research Electronic Data Capture (REDCap) electronic data capture tools hosted at the UW Institute of Translational Health Sciences.^{24,25} Data entry was overseen or performed by experienced research nurses or doctors at all 4 sites. A PDF copy of the REDCap data collection instrument used to populate our database is provided in the [Supplemental Material](#).

Clinical outcomes

Records were labeled by clinician and/or trained study coordinators with clinician supervision for (1) patient all-cause mortality vs survival and (2) incidence of major adverse cardiovascular events (MACE) during COVID-19 hospitalization (controls without events vs cases with arrhythmic, heart failure [HF], or thromboembolic events). Co-primary

endpoints were mortality and composite of MACE, which included thromboembolic, arrhythmic, or HF events. Thromboembolic events were defined as acute myocardial infarction, ischemic stroke, or pulmonary embolism. Arrhythmic events were defined as new-onset AF, high burden of premature ventricular complexes, sustained ventricular tachycardia, ventricular fibrillation, and cardiac arrests owing to bradyarrhythmias and tachyarrhythmias. HF events included new-onset HF and cardiogenic shock. Patients with previously known or recurrent AF were noted, but these were not classified as events under the arrhythmic endpoint, since our aim was to predict complications that arose *de novo* because of COVID-19. The non-event cases formed the background class, in which patients developed no MACE during their COVID-19 hospitalization. The study period only comprised the hospitalization for COVID-19 and no long-term follow-up was performed after discharge. Thus, all included patients either died or were discharged.

Artificial intelligence model development

Essential details about the artificial intelligence (AI) models are provided in this section; [Supplemental Methods](#) provides further information. [Figure 1](#) presents a dataflow diagram for our study. In addition to the general exclusion criteria described

above, 6 records were excluded from the MACE network analysis owing to the presence of AF on the intake ECG in a patient with new-onset AF as an outcome. ECGs and associated labeling information were split into training and testing sets using a 10-fold stratified cross-validation scheme.²⁶ The ratio of records in the training, validation, and testing sets was 8:1:1. The deep learning model input data was a stack of standard 10-second 12-lead ECG records sampled at either 500 ($n = 1217$) or 250 Hz ($n = 169$). For each clinical record used the model input was an 8×5000 matrix.

We formulated 2 distinct convolutional neural networks with long short-term memory (CNN-LSTMs); these were designed to predict the incidence during COVID-19 hospitalization of, respectively, mortality (CNN-LSTM1) and life-threatening cardiovascular events as described above (CNN-LSTM2). Figure 2 shows schematic representations for the model architectures, both of which had 3 main sections. In section I, a series of 1D convolution layers was used to extract temporal features from each channel separately. Every convolution layer was followed by a batch normalization layer to centralize data distribution and a rectified linear unit to weight the output of past layers.²⁷ Section II was a recurrent neural network, consisting of 2 LSTM layers²³ to process spatial features across all 8 channels; each LSTM layer contained up to 4 units with feedback connections. Section III was a fully connected layer and activation function, used to output a value describing the model's confidence that a particular ECG belonged to each class. CNN-LSTM1 was a binary model (ie, predicted likelihood of survival vs death), whereas CNN-LSTM2 was a multilabel model (independent likelihoods for all 4 event types: arrhythmic, heart failure, thromboembolic, or none).

In addition to the primary analysis via 10-fold cross-validation, a secondary analysis was conducted to evaluate whether the timing of patient intake might affect the predictive power of our AI models. In this case, we trained the models with only ECGs of patients who were admitted between March 2, 2020, and June 8, 2020. Model performance was then evaluated with a holdout test set that contained only ECGs of patients who were admitted on or after June 9, 2020. Lastly, to explore the possibility that changing the cohort size might affect model accuracy, we reran the entire 10-fold train/validate/test process for CNN-LSTM1 for subsets with 20%, 40%, 60%, or 80% of the population. These were used to construct a learning curve (ie, model accuracy vs cohort size), as in prior work.²⁸

Training and testing of CNN-LSTM models was carried out using advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system of the University of Washington. All jobs were run on 1 standard compute node (32 cores, Intel® Xeon® Gold 6130 CPU @ 2.10 GHz, 128 GB RAM).

Statistical assessment of AI model performance

Model performance was evaluated by calculating area under the receiver operator characteristic curves (AUROC),

sensitivity and specificity values, and confusion matrices for holdout testing sets in a 10-fold cross-validation scheme, as described above. Optimal probability thresholds were determined based on the ROC curves of internal validation sets for each output class:

$$D = fpr_{val}^2 + 2 \times (1 - tpr_{val})^2$$

where D is a weighted distance from the origin of the coordinate system to a specific point in the ROC curve of validation set, fpr is false-positive rate, and tpr is true-positive rate. This formulation intentionally weighted sensitivity (ie, $1 - tpr$) over specificity (ie, fpr), since the importance of minimizing false-negative rate (fnr) was deemed a much higher priority than reducing fpr . Each fpr , tpr pair maps to a unique threshold determined by minimizing D .

The ROC curves for each class in CNN-LSTM2 independently consider the comparison of model output to ground truth label. Thus, assessing the overall performance of this model necessitated the use of an ROC aggregation technique. We thus calculated the *macro-average* ROC curve (ie, each class's ROC curve contributed equally to the average), which considered each prediction problem separately (eg, survival vs non-survival AND death vs non-death, independently) then averaged them together.

Finally, when aggregating results of different model performances from all different train/test iterations carried out as part of the 10-fold cross-validation process, 2-sided 95% confidence intervals were used to estimate the overall performance matrix for the system.

Conventional statistical model development

Clinical model comparisons were developed using multivariable logistic regression. The cohort was split into the same 10 training/testing sets used for the AI models. The endpoints of mortality and *composite* cardiac event (experience of at least 1: HF, arrhythmia, or thromboembolism) were modeled separately. The following steps were performed for each of the 10 cohort splits: Missing data were imputed in the training set using mean imputation and the same mean values were applied to missing data in the test and validation sets. Logistic regression was used to predict the outcome. The models adjusted for age, sex, race/ethnicity, body mass index (BMI), history of hypertension, history of coronary artery disease (CAD), and presence of an intracardiac device. Four metrics extracted from ECGs were also included: ventricular rate, PR interval, QRS duration, and QT interval. ECG metrics were treated as continuous variables. Race categories were Asian, Black, White Hispanic, White non-Hispanic, and Other/Unknown. BMI was either entered by clinical staff or calculated from height and weight. Age and BMI were treated as continuous variables. History of hypertension, history of CAD, and presence of intracardiac device were treated as binary variables, with the reference being not present. The reference category for sex assigned at birth was male. Race/ethnicity was included as a factor, with non-Hispanic White as the reference. The logistic regression

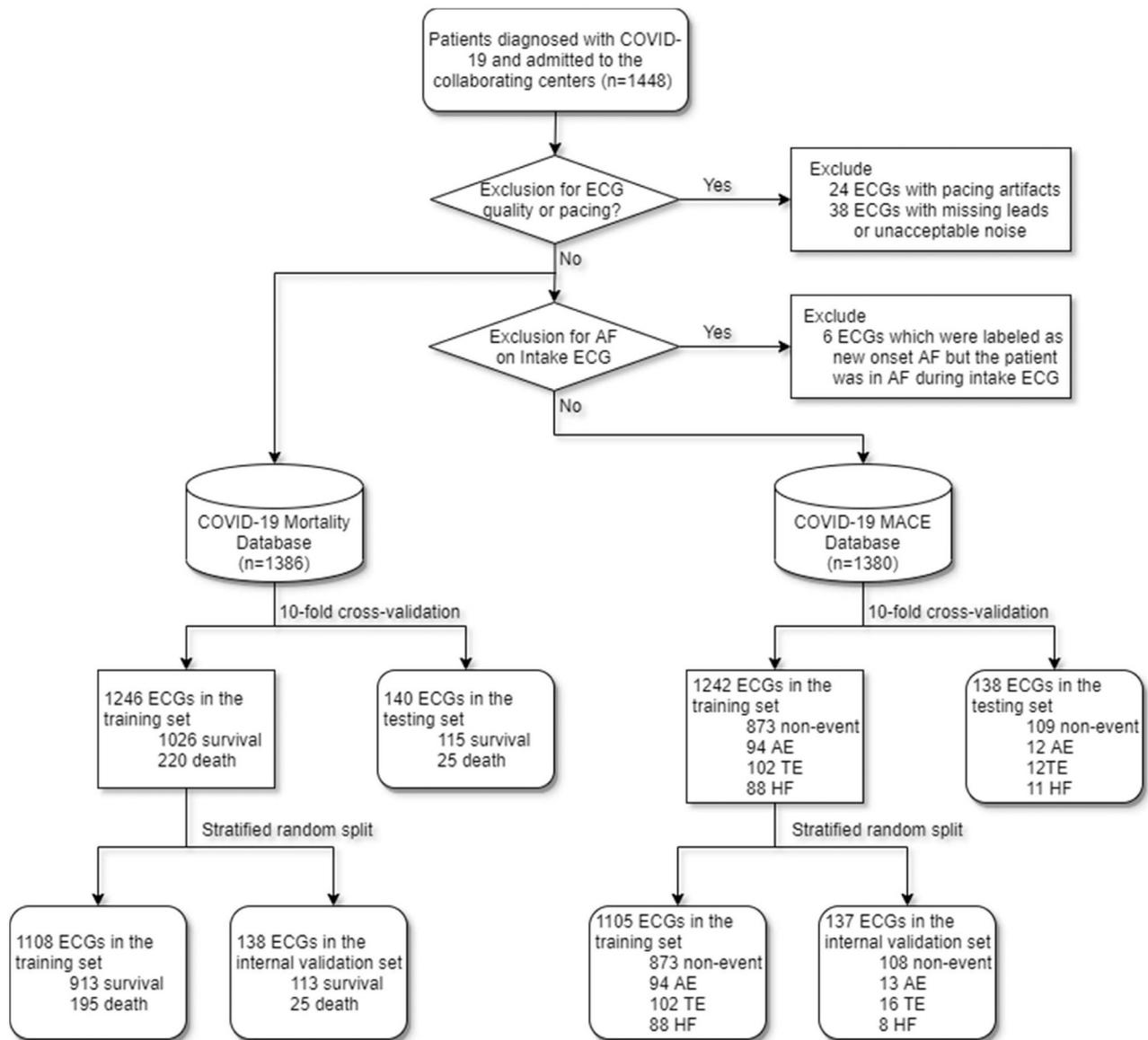


Figure 1 Flowchart showing inclusion of patients into databases for both classification problems to be addressed via artificial intelligence–based predictive modeling. The left branch of the tree concerns the first convolutional neural network with long short-term memory (CNN-LSTM1), concerned with differentiating between electrocardiograms (ECGs) from patients who survived vs died. The right branch describes the database used for CNN-LSTM2, which independently predicts the likelihood that each ECG belongs to a patient from 4 groups (no event vs major adverse cardiovascular events [MACE], as shown in legend). AE = arrhythmic event; HF = heart failure; TE = thromboembolic event.

models used the same optimal threshold identification process as outlined above for the AI models in the test set and performance metrics were evaluated in the validation cohort. Results of the model performances from all different train/validation/test iterations carried out as part of the 10-fold cross-validation process were aggregated to determine 2-sided 95% confidence intervals. Analyses were performed using the R programming language, version 3.6.1, and used packages readxl,²⁹ tidy,³⁰ dplyr,³¹ caret,³² and pROC.³³

Results

We used intake ECGs and outcomes during COVID-19 hospitalization for 1386 patients, after all exclusion criteria

had been applied (see Methods). All 4 centers were university hospitals (Table 1). The mean age was 63.43 ± 16.9 years; 60.5% ($n = 839$) were assigned male sex at birth, 54.0% ($n = 737$) had hypertension, and 1.9% ($n = 27$) had a cardiac implantable electronic device (22 [1.6%] pacemakers; 5 [0.4%] implantable cardioverter-defibrillators). A total of 116 intake ECGs included atrial arrhythmia, which accounted for 8.4% of the total database. Ninety-seven were AF and 12 were atrial flutter, with various other arrhythmias accounting for the remaining cases (Supplemental Table 1); of note, 6 records with new-onset AF were excluded from the development of CNN-LSTM2 to avoid confounding, as outlined in Methods.

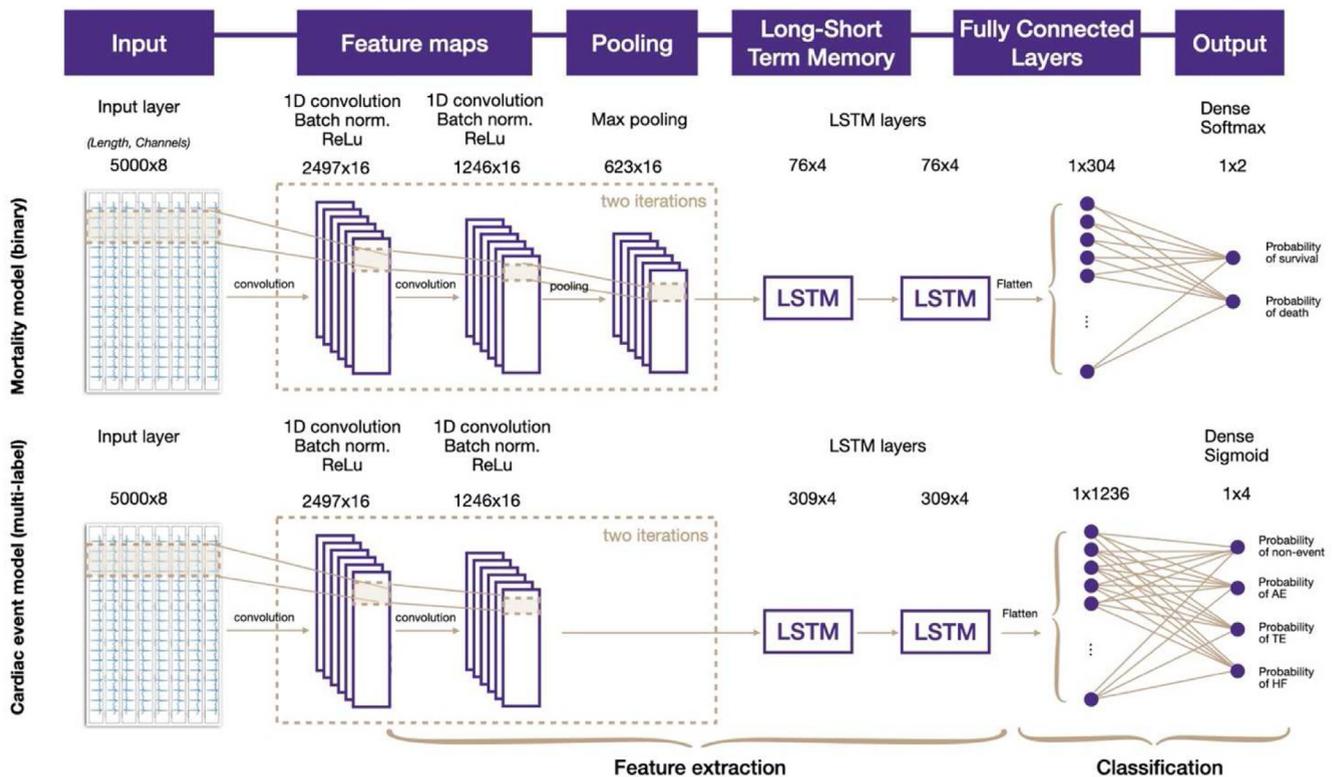


Figure 2 Schematics illustrating machine learning network architectures for convolutional neural network long short-term memories 1 and 2 (LSTM1/2). Each network consists of 3 sections: (I) convolution layers, shown here as “feature maps” and “pooling” subsections; (II) recurrent neural network layers (labeled “Long-Short Term Memory”); and (III) fully connected layers that produce outputs (ie, predicted probabilities for each class). AE = arrhythmic event; HF = heart failure; ReLu = rectified linear unit; TE = thromboembolic event.

The all-cause mortality rate was 17.7% ($n = 245$); the survival group included patients who were discharged either to home ($n = 776$; 63.7%) or to other care units after stabilization ($n = 235$; 19.3%). As shown in [Table 1](#), subcohorts from different regions were distinct, with statistical tests showing a lack of homogeneity in all demographic and comorbidity categories except hypertension. In contrast, outcomes were more uniformly distributed across the centers, with exceptions being a higher incidence of HF events at KI and a lower mortality rate at UC. We also tabulated these data, along with ECG metrics, according to mortality vs survival outcome ([Supplemental Table 2](#)) and event type (none vs arrhythmic vs HF vs thromboembolic event; [Supplemental Table 3](#)). Compared to those who survived, patients who died were older (average age: 74 vs 61 years) and had higher prevalence of hypertension (70% vs 50%) and CAD (23% vs 12%); for comparison between event types, significant intergroup differences existed for several demographic variables as well as hypertension and CAD (both of which were higher in patients who experienced arrhythmia or HF events). In 2 subgroups (patients who died vs those who survived and patients with HF vs no MACE), ventricular rate was faster, QRS duration was greater, and corrected QT interval was prolonged ($P < .01$ in all cases). Given the high rate of atrial arrhythmia in this cohort, it is notable that there were no significant intergroup differences whatsoever in PR interval or P-wave duration.

The ability of the fully trained CNN-LSTM1 model to successfully predict death during hospitalization vs survival from intake ECG is illustrated by the ROC curve in [Figure 3A](#) (green line; AUROC: 0.60 ± 0.05); the ROC curve for the conventional statistical model is also shown for comparison (red line; AUROC: 0.73 ± 0.07). Additional detail regarding the conventional model (odds ratio and P values for each covariate in the model) are presented in [Supplemental Table 4](#). In both cases, the line shown is the average across all 10 cross-validation data sets and the shaded region shows the ± 1 standard deviation range. The predictive power of CNN-LSTM1 was not approved by either of the transfer learning (TL) approaches described in the Supplemental Methods (TL1 AUROC: 0.42 ± 0.06 ; TL2 AUROC: 0.54 ± 0.06). At the optimal threshold value, positive and negative predictive values (PPV/NPV) [95% confidence intervals] for CNN-LSTM1 were 0.22 [0.20–0.23] and 0.87 [0.85–0.89]; the overall sensitivity and specificity were 0.66 [0.60–0.73] and 0.47 [0.41–0.53]. These values were distinct from those observed for the conventional statistical model, in which positive and negative predictive powers were inverted (PPV: 0.90 [0.88–0.92]; NPV: 0.35 [0.32–0.38]; sensitivity: 0.75 [0.69–0.80]; specificity: 0.61 [0.50–0.72]). To further examine the tradeoff between highly undesirable *fpr* (ie, survival predicted for a patient at risk of dying) and more tolerable *fpr* (ie, death predicted for a patient tracking towards survival), we plotted both quantities as a

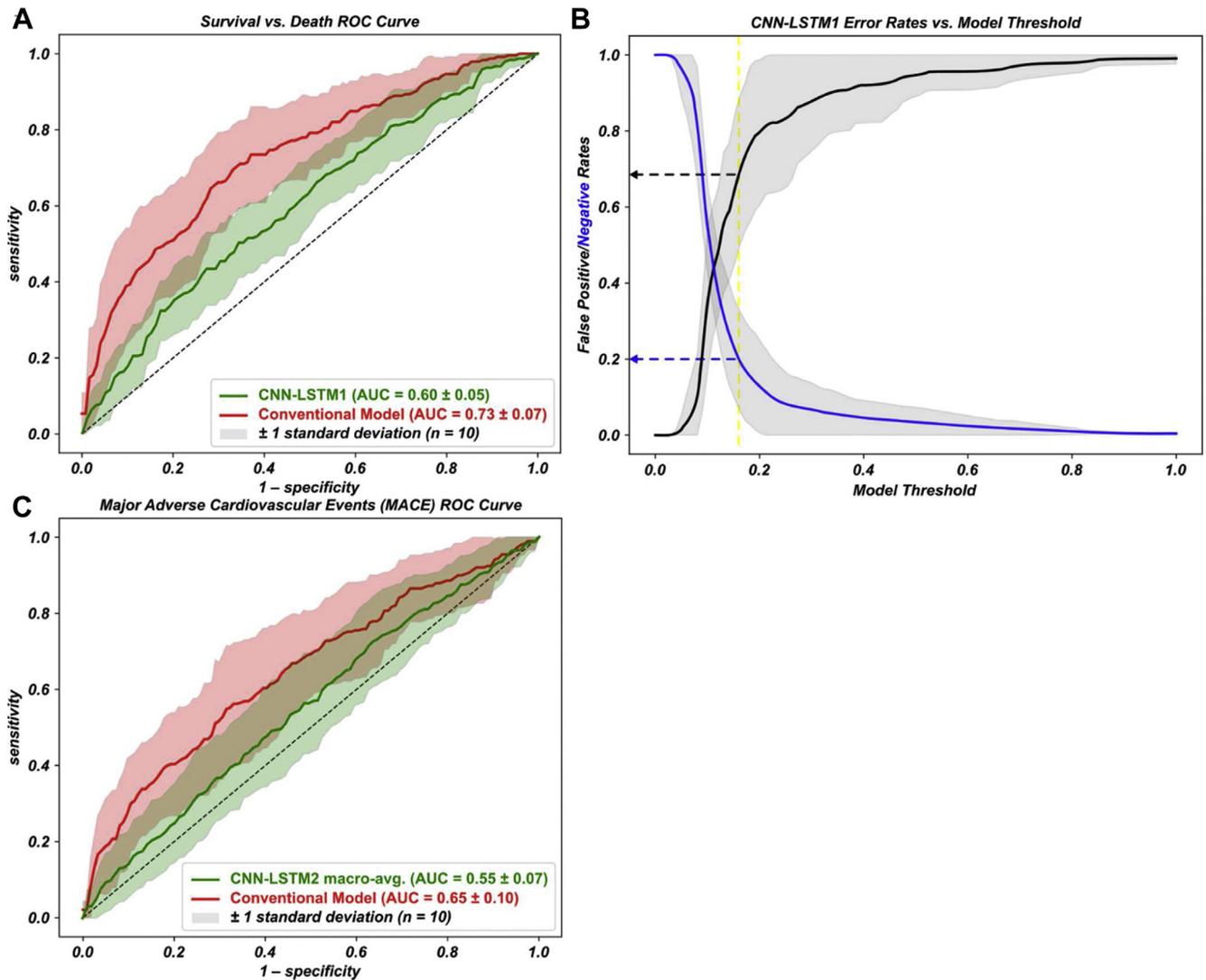


Figure 3 Data summarizing predictive power of artificial intelligence–based and conventional models. **A:** Receiver operator characteristic (ROC) curves for convolutional neural network long short-term memory 1 (CNN-LSTM1) and the corresponding conventional model, which attempted to differentiate between electrocardiograms of patients who survived vs died. **B:** False-positive and false-negative rates (*fpr* and *fnr*, respectively) for CNN-LSTM1 as a function of model threshold; at a nominal threshold (dashed yellow line) associated with a 20% *fnr*, the corresponding *fpr* ($\sim 70\%$) is shown by a dashed black arrow. **C:** ROC curves for CNN-LSTM2 (multilabel model independent prediction of different major adverse cardiovascular event types; single curve derived via macro-averaging) and the corresponding conventional model (binary prediction: any event vs no event). See Figure 4 and Supplemental Figure 2 for additional plots. AUC = area under the curve; MACE = major adverse cardiovascular event.

function of the raw CNN-LSTM1 output threshold for distinguishing between predicted death and survival (Figure 3B). This analysis shows that to maintain a reasonable *fnr* (eg, 20% for the nominal threshold shown by the dashed yellow line), it was necessary to accept a high *fpr* ($\sim 70\%$ for the example shown); this is consistent with the high NPV and low PPV values reported above. Construction of a learning curve (Supplemental Figure 1) for CNN-LSTM1 showed that there was no trend towards increasing model accuracy as a function of cohort size. Rather, there was a trend towards reduced accuracy, although differences in accuracy for successive cohort sizes were not statistically significant.

Summary data (ie, macro-average ROC curve; AUROC: 0.55 ± 0.07) for prediction of cardiovascular events via CNN-LSTM2 are shown in Figure 3C (green line). These

results are contrasted with those from the logistic regression model shown in Figure 3C (red line), which tackled a distinct classification problem that was binary (ie, MACE vs no MACE; AUROC: 0.65 ± 0.10); see Supplemental Table 4 for additional covariate information from this model. Both the ML-based and conventional approach had inferior power compared to the respective death vs survival models, indicating that MACE prediction (multiclass using CNN-LSTM2 or binary using the conventional model) was more challenging. Quantitative metrics of the CNN-LSTM2 aggregate performance were as follows: PPV = 0.28 [0.27–0.29]; NPV = 0.76 [0.74–0.78]; sensitivity = 0.67 [0.61–0.72]; specificity = 0.42 [0.36–0.47]. The predictive power of the logistic regression model was comparatively higher: PPV: 0.82 [0.81–0.84]; NPV: 0.32 [0.26–0.38]; sensitivity: 0.75

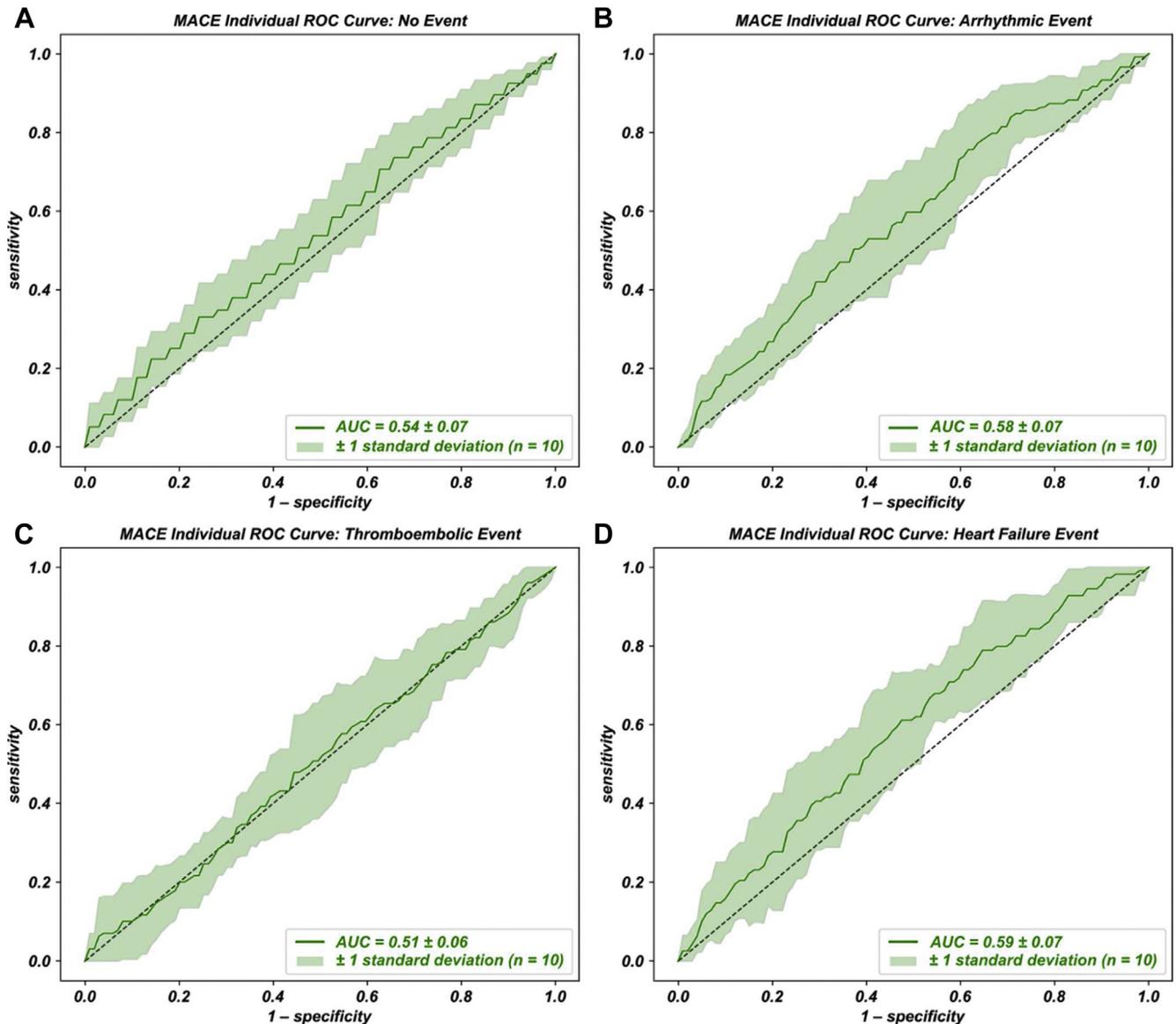


Figure 4 Individual receiver operator characteristic (ROC) curves for the 4 independent classification tasks performed by convolutional neural network long short-term memory 2 (CNN-LSTM2). **A:** Prediction of no event during hospitalization from intake electrocardiogram. **B:** Prediction of arrhythmic event. **C:** Prediction of thromboembolic event. **D:** Prediction of heart failure event. AUC = area under the curve; MACE = major adverse cardiovascular event.

[0.69–0.82]; specificity: 0.39 [0.29–0.49]. Since TL was not beneficial in the context of the mortality vs survival network, we elected not to attempt it for CNN-LSTM2.

ROC curves illustrating model performance for independent prediction of the 4 individual MACE types are provided in Figure 4A–4D. Here, we see that poor performance of CNN-LSTM2 was disproportionately driven by the model’s difficulty in identifying ECGs from patients who had no event or experienced a thromboembolic event (AUROC = 0.54 ± 0.07 and 0.51 ± 0.06 , respectively); in contrast, the model’s ability to predict arrhythmic or HF events was on par with the AI-based ability to predict death (AUROC = 0.58 ± 0.07 and 0.59 ± 0.07 , respectively). Nevertheless, as shown by graphs of *fpr* and *fnr* as a function of event-specific model thresholds for all 4 outcome types in Supplemental Figure 2A–2D, the practical consequence of

this increased predictive power is modest (ie, dashed black lines showing the *fpr* that must be tolerated to achieve a 20% *tpr* all range between $\sim 75\%$ and $\sim 80\%$, with little difference between event types).

CNN-LSTM1 outputs for predicted death and survival probabilities for intake ECGs in a representative test set are shown in Figure 5A. As indicated by the box-and-whisker diagrams superimposed on the violin plots, the optimal threshold value (dashed line) resulted in 72% of the ECGs from patients who died being assigned a “death” prediction (ie, *tpr*; Figure 5A, right). The corresponding rate of correct classification for ECGs from patients who survived was lower (Figure 5A, left: 54% *tnr*). Similar plots for CNN-LSTM2 (Figure 5B and 5C and Supplemental Figure 3A and 3B) indicate that the multilabel classification problem was more challenging.

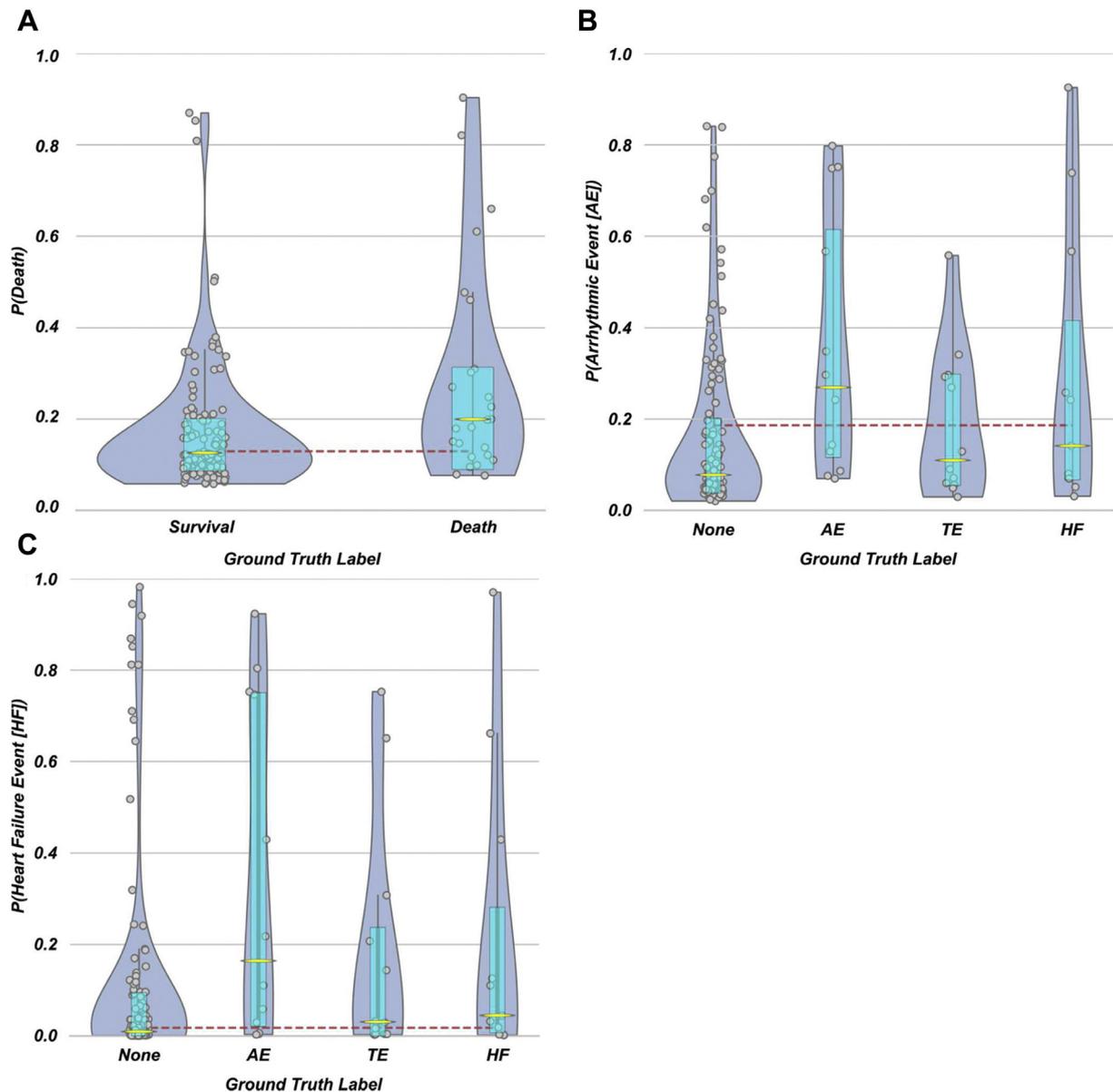


Figure 5 Violin plots with box-and-whisker annotations showing raw network outputs. **A:** Results for classification task (survival vs death) in convolutional neural network long short-term memory 1 (CNN-LSTM1). **B, C:** Results for 2 classification tasks in CNN-LSTM2 (**B:** arrhythmic events; **C:** heart failure events). X-axis labels show ground truth labels; dashed lines in each panel show optimal classification thresholds, as explained in Methods. See [Supplemental Figure 3](#) for additional plots. AE = arrhythmic event; HF = heart failure; TE = thromboembolic event.

Finally, as shown in [Figure 6](#), using a test set comprising only ECGs and outcomes from patients recruited after a cutoff date (June 9, 2020), the predictive power of both the survival vs death model ([Figure 6A](#)) and the multilabel MACE model ([Figure 6B](#); macro-average and individual predictive ROCs) were within the ± 1 standard deviation intervals for the corresponding models shown in [Figure 3A](#) and [3C](#), respectively. The numbers of patients enrolled at each center over the study period are shown in [Supplemental Figure 4A](#) and [4B](#). There were no discernable trends in mortality rate or MACE incidence on a month-by-month basis (see [Supplemental Figure 4C](#) and [4D](#), respectively).

Discussion

In this manuscript, we demonstrate the feasibility and examine the potential effectiveness of a rapid, inexpensive triage tool for COVID-19 patients using a 12-lead ECG at admission augmented by ML. To our knowledge, this is the first such attempt to triage any viral disease using a single ECG. Destabilization of healthcare infrastructure during the pandemic sharply highlighted the need for rapid and readily available triage tools to guide resource allocation during times of crisis. When compared with the traditional statistical model incorporating extensive clinical data, our results using only the DNN-augmented ECG did not yield improved discrimination; however, this approach has the advantage

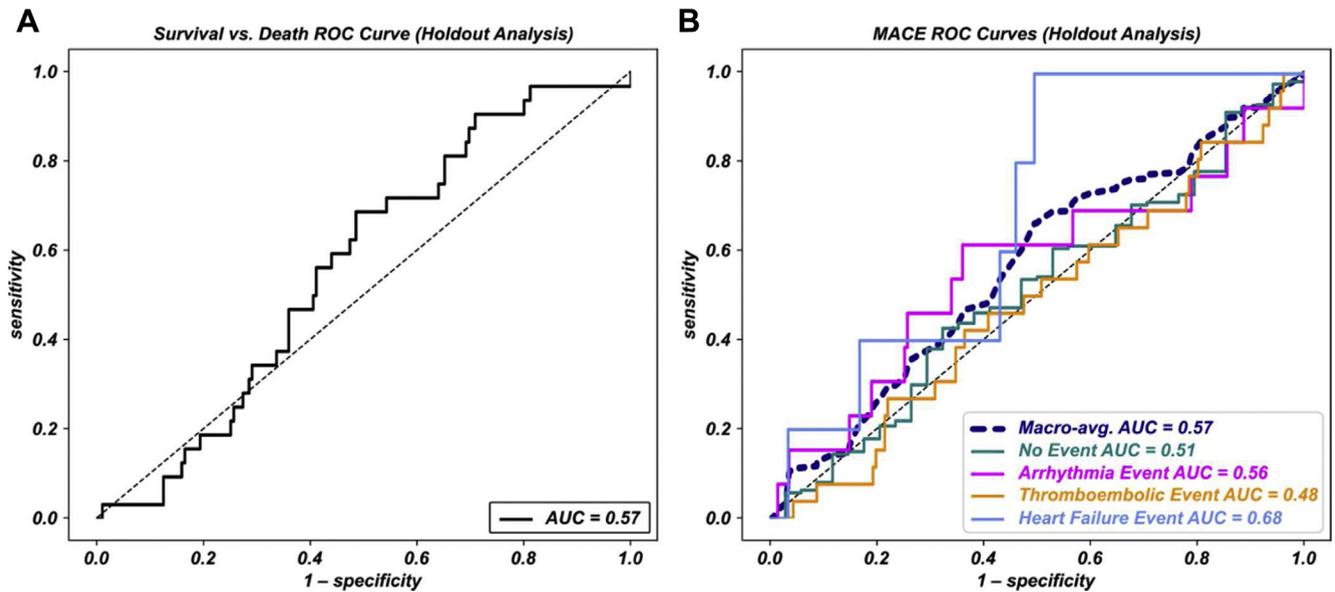


Figure 6 Summary data for holdout testing of both artificial intelligence-based models trained with electrocardiograms (ECGs) from before a cutoff date (June 9, 2020) and tested with ECGs from after that date. **A:** Receiver operator characteristic (ROC) curve for convolutional neural network long short-term memory 1 (CNN-LSTM1) trained and tested using the holdout protocol defined above. **B:** Same as panel A but for CNN-LSTM2. Both macro-averaged and individual event type ROCs are shown superimposed on this plot. AUC = area under the curve; MACE = major adverse cardiovascular event.

of not requiring clinical expertise to gather medical history.³⁴ Nevertheless, our models' low sensitivity and PPV would be a barrier to successful deployment in clinical settings in its current form. Mixed-input AI models that analyze both ECG data and clinical variables might enhance the accuracy, as discussed below.

Our network yielded AUROC values of 0.60 for predicting death and 0.55 for predicting MACE. The overall performance of our models can be compared to other commonly used risk stratification tools such as the CHADS₂ and CHA₂DS₂-VASc scores for stroke, which are estimated by the U.S. Agency for Healthcare Research and Quality to have AUROC values of 0.66–0.75.³⁵ Owing to the unpredictable nature of the disease, especially in hospitalized patients, and the high risk of devastating complications,³⁶ we optimized our DNN to maximize NPV. This prioritized correct identification of patients who had favorable outcomes as low risk (ie, true negatives), while minimizing false negatives. Our model achieved an NPV of 0.87 for mortality and 0.76 for MACE.

Of the 2 networks we developed, CNN-LSTM1 had distinctly better performance than CNN-LSTM2. Although the specifics of the classification tasks involved are distinct (multiclass vs binary), the general implication is that intake ECG data may be less suitable for predicting MACE during COVID-19 hospitalization compared to demographic data and comorbidity data. Nevertheless, results from CNN-LSTM2 must be carefully interpreted, since its ability to predict different MACE types was divergent. For the best-performing individual prediction (HF event classifier), at the optimal threshold the network had relatively low *fpr* (eg, 44% for ECGs from patients with no event; Figure 5C,

column 1) but this came at the cost of a relatively high *fnr* (27%; Figure 5C, column 4). In the case of independently predicting the likelihood that a patient would experience no MACE (Supplemental Figure 3A), while 73% of ECGs from patients who truly lacked adverse events were correctly classified (column 1), >50% of patients in all 3 MACE groups were incorrectly classified (columns 2–4).

The strategic holdout analysis presented in Figure 6 also warrants further discussion. The aim here was to determine if the model's predictive power was degraded by attempting to train and test the model with ECGs from patients hospitalized earlier and later in the pandemic, respectively. This might be expected, owing to the changing attributes of COVID-19 disease and treatment over time, although our analysis of temporal distribution of mortality rates and MACE incidence in this cohort (Supplemental Figure 4C and 4D) showed no apparent trends. In practice, holdout model performance was not obviously inferior to the main results presented in Figure 3. Notably, the holdout MACE model's ability to predict HF from intake ECG was among the highest observed in the study (AUROC = 0.68).

Even though our enthusiasm for the potential translational value of these particular models is low, our study has significant strengths. We initiated an international collaboration of 4 university hospitals, which were heavily affected by the pandemic across 2 continents. This yielded a diverse set of patients from different geographic regions and institutions with significant heterogeneity in presentation as well as therapies. Various conventional ML techniques have been proposed for ECG-based algorithms,³⁷ but most use ECG features, as opposed to the raw ECG signal itself as employed here, as input (eg, R-R interval, QRS complex duration, etc),

which requires labor-intensive preprocessing. Moreover, we opted for minimal data sanitization by enforcing only the most essential exclusion criteria and applying minimal preprocessing to model inputs. Thus, the data used to train and test our ML networks closely resemble “real-world” signals that could be found in any clinical setting. We highlight that atrial arrhythmia is common and likely to be encountered in many patients presenting with COVID-19, so we opted to include as many of these ECGs as possible in our study. The lone exception was that ECGs demonstrating new-onset AF were excluded from the MACE outcome, since new-onset AF was a component of the MACE outcome.

It is also worth noting that CNN-LSTM2 carries out a multilabel classification task in which the likelihood of a particular ECG belonging to each group (no event vs 3 types of MACE) is independently predicted via 1 single set of calculations. This is a distinct and more difficult classification task from the conventional statistical model against which the performance of CNN-LSTM2 was evaluated (ie, binary prediction of no MACE vs *any* MACE). Finally, we carried out a rigorous and comprehensive hyperparameter tuning exercise for both CNN-LSTM1 and CNN-LSTM2. In each case, we considered 128 unique network permutations and carried out a complete set of train/test cycles using our 10-fold cross-validation schemes (ie, 1280 unique runs per model), meaning that the predictive capabilities of the models shown in [Figure 2](#), [Supplemental Tables 4, 5](#) are the best possible outcome for this set of DNN parameters. As such, we are confident that the modest AUROC values for our 2 models are not the consequence of an inadequate exploration of what DNN technology might potentially offer.

The imperfect performance of our network demonstrates an important aspect of ML applications: explainability. While many DNN architectures allow for identification of specific features predictive of adverse outcomes, we were unable to do so because of the LSTM components included in our DNNs. The CNN models in our study were trained to identify adverse outcomes regardless of the exact changes occurring on the ECG. We believe it is entirely possible that the CNNs could be using any changes encoded in that signal (eg, heart rate, heart rate variability, QRS width, ST changes, or combinations of these elements) to make predictions about outcomes. In the case of COVID-19, it is possible that the most salient ECG markers are those that are indicative of baseline comorbidity, such as right atrial abnormality or right ventricular strain pattern, which suggest pulmonary disease. Alternatively, the network may have preferentially identified novel features associated with severe COVID-19 disease. To examine exactly what features the CNN models are examining to make these predictions, we would have to use certain explainable AI models, which would be more interpretable. This was beyond the scope of the current project, but future studies may be able to harness new computational techniques³⁸ to improve approaches in this regard. It is also difficult to speculate what steps may be taken to improve model performance. We hypothesize that the heterogeneity of both the components of MACE and varied causes of mortality

contributed to poor performance. Future efforts may benefit from identification of more narrowly defined outcomes with linear and biologically plausible mechanistic connections between cardiac electrical activity and the outcome.

Some interesting observations can be gleaned from the odds ratios for covariates in the 2 conventional models used as a basis of comparison for the AI-based models. Advanced age, male sex, and higher ventricular rate were associated with increased risk of both mortality and MACE; for the second model, there were additional associations between MACE risk and elevated BMI or increased PR interval duration, albeit with weaker *P* values. The clinical significance of these features is not entirely clear. In the context of the current study, it is noteworthy that at least 1 ECG feature (ventricular rate) was strongly associated with both death and MACE. While it is possible that formulation of alternative statistical models using exclusively ECG metrics, demographic data, or comorbidity information might elucidate the key drivers of predictive capability, these analyses were deemed beyond the scope of this AI-focused study. Nevertheless, the fact that the predictive accuracy of both conventional models outstripped that of our AI-based methods suggests that future development of mixed-input CNN-LSTMs that analyze both raw ECG and demographic/comorbidity data might prove useful.

Limitations

Our study has several additional limitations. First, the modest performance of our networks limits clinical applicability at the current stage. Although the accuracy we observed was not superior to the traditional model based on clinical data, we note that triage tools for potentially fatal diseases should prioritize high NPV and our models did achieve that design goal. A deep learning model based on both clinical and ECG data might be able to achieve better accuracy; however, building such a network with heterogeneous inputs is complicated and, to the best of our knowledge, has not been attempted in this field. Notably, while we note that the size of the cohort considered in this study is small compared to many prior AI-based ECG studies,^{39,40} our analysis ([Supplemental Figure 1](#)) suggests this may not be the key factor constraining model performance. While this finding does not guarantee that increasing the cohort size would not increase the accuracy of our model, it suggests a more likely explanation is that the deep learning approach is relatively ill suited for the chosen problem.

Second, this study included only hospitalized patients with COVID-19. Whether this test can be used to prognosticate in settings where patients have few or no symptoms requires a separate study. Third, certain agents used to treat COVID-19, which we did not aim to capture in our database, affect the ECG (eg, hydroxychloroquine, azithromycin, and other critical care medications) and may have impacted network performance. Regarding the latter point, the geographic and management diversity of the included patients who were treated in different institutions may have enhanced the robustness and generalizability of our analysis.

Lastly, MACE definitions were based on clinicians' determinations as adjudicated by chart review and notes. While we collected data on certain laboratory values, we opted to rely on clinician judgment for diagnosis of MACE elements such as acute myocardial infarction. These definitions carry an important implication of subjectivity and interobserver bias, but we determined that this would be the more ideal adjudication method compared to the alternative of having an arbitrary threshold lab value or premature ventricular complex burden when these parameters were not being measured systematically in the pandemic setting.

Conclusion

Our analysis shows that AI-based algorithms could be helpful for rapid prognostication of adverse events in COVID-19 patients using a single ECG taken at the time of admission. While these models do not have perfect accuracy by any means, their performance is comparable to that of conventional statistical models relying on a multitude of clinical and demographic variables.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are grateful to research nurse Helena Karlsson (KI) for assistance with data entry for this project.

Funding Sources

This work was supported by a COVID-19 Rapid Response Grant from the University of Washington Population Health Initiative (awarded to ARS, AEF, NAC, JEP, and PMB). The work was further supported by AFA Insurance in Sweden, Dnr 200223, Novo Nordisk Foundation in Denmark (NFF20SA0062835), and Norwegian Center for Research-Based Innovation Precision Health Center for optimized cardiac care (ProCardio), Project No.: 309762. The REDCap database system, which was essential to the completion of this project, was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1 TR002319.

Disclosures

ARS and PMB report that during the initial data collection phase at UW, GE provided access to the MUSE plugin necessary for exporting ECG data in machine-readable format free of charge; otherwise, GE was not involved in this project. CBL reports: consulting fees from Medtronic, Boston Sci, Cathprint; speaker honoraria from Bayer, Medtronic, Boston Sci, MSD, BMS, and Philips; and funding from the Swedish Heart Lung Foundation and the Swedish Research Council. CML reports: consulting fees from AstraZeneca, Roche Diagnostics, and Bayer; speaker honoraria from Novartis, AstraZeneca, Bayer, Medtronic, Impulse Dynamics, and Vi-for; and funding from the Swedish Heart Lung Foundation,

Stockholm County Council, and the Royal Society of Science. TBS reports receiving research grants from Sanofi Pasteur and GE Healthcare, is a Steering Committee member of the Amgen-financed GALACTIC-HF trial, is on advisory boards for Sanofi Pasteur and Amgen, and has received speaker honorariums from Novartis and Sanofi Pasteur. The remaining authors have nothing to disclose.

Authorship

All authors attest they meet the ICMJE criteria for authorship.

Patient Consent

Patient consent was waived for this research as all data were collected through retrospective chart review. Only deidentified data was shared between institutions.

Ethics Statement

This study was approved by the Institutional Review Board (IRB) of the UW, the Swedish Ethical Review Authority, and the regional Danish Ethical Committee. Associated reference numbers: IRB6878 (UW), 2020-02627 (KI), 2020-02662 (UU), and H-20021500 (UC).

Appendix Supplementary data

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.cvdhj.2021.12.003>.

References

1. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020;324:782–793.
2. World Health Organization. WHO Coronavirus (COVID-19) Dashboard. 2021. <https://covid19.who.int>.
3. Garcia S, Albaghdadi MS, Meraj PM, et al. Reduction in ST-segment elevation cardiac catheterization laboratory activations in the United States during COVID-19 pandemic. *J Am Coll Cardiol* 2020;75:2871–2872.
4. Sharma R, Kuohn LR, Weinberger DM, et al. Excess cerebrovascular mortality in the United States during the COVID-19 pandemic. *Stroke* 2021;52:563–572.
5. Kaufman HW, Chen Z, Niles J, Fesko Y. Changes in the number of US patients with newly identified cancer before and during the coronavirus disease 2019 (COVID-19) pandemic. *JAMA Netw Open* 2020;3:e2017267.
6. Blumenthal D, Fowler EJ, Abrams M, Collins SR. Covid-19-implications for the health care system. *N Engl J Med* 2020;383:1483–1488.
7. Nishiga M, Wang DW, Han Y, Lewis DB, Wu JC. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nat Rev Cardiol* 2020;17:543–558.
8. Marchiano S, Hsiang TY, Khanna A, et al. SARS-CoV-2 infects human pluripotent stem cell-derived cardiomyocytes, impairing electrical and mechanical function. *Stem Cell Reports* 2021;16:478–492.
9. Michaud GF, Stevenson WG. Atrial fibrillation. *N Engl J Med* 2021;384:353–361.
10. 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.
11. Castro RA, Frishman WH. Thrombotic complications of COVID-19 infection: a review. *Cardiol Rev* 2021;29:43–47.
12. Modin D, Claggett B, Sindet-Pedersen C, et al. Acute COVID-19 and the incidence of ischemic stroke and acute myocardial infarction. *Circulation* 2020;142:2080–2082.
13. Katsoularis I, Fonseca-Rodriguez O, Farrington P, Lindmark K, Fors Connolly AM; Risk of acute myocardial infarction and ischaemic stroke

- following COVID-19 in Sweden: a self-controlled case series and matched cohort study. *Lancet* 2021;398:599–607.
14. Fye WB. A history of the origin, evolution, and impact of electrocardiography. *Am J Cardiol* 1994;73:937–949.
 15. Lanza GA, De Vita A, Ravenna SE, et al. Electrocardiographic findings at presentation and clinical outcome in patients with SARS-CoV-2 infection. *Europace* 2021;23:123–129.
 16. Christopoulos G, Graff-Radford J, Lopez CL, et al. Artificial intelligence-electrocardiography to predict incident atrial fibrillation: a population-based study. *Circ Arrhythm Electrophysiol* 2020;13:e009355.
 17. Ko WY, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol* 2020;75:722–733.
 18. Adedinsowo D, Carter RE, Attia Z, et al. Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circ Arrhythm Electrophysiol* 2020;13:e008437.
 19. Kwon JM, Lee SY, Jeon KH, et al. deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc* 2020;9:e014717.
 20. Raghunath S, Ulloa Cerna AE, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med* 2020;26:886–891.
 21. Kwon JM, Kim KH, Jeon KH, Lee SY, Park J, Oh BH, et al. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. *Scand J Trauma Resusc Emerg Med* 2020;28:98.
 22. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–444.
 23. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–1780.
 24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–381.
 25. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
 26. Poldrack RA, Huckings G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* 2020;77:534–540.
 27. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning*. Haifa, Israel: Omnipress 2010;807–814.
 28. Perlich C, Provost F, Simonoff J. Tree induction vs. logistic regression: a learning-curve analysis. *J Mach Learn Res* 2003;4:211–255.
 29. Wickham H, Bryan J. readxl: Read Excel Files. R package version 1.3.1, 2021. Available at: <https://CRAN.R-project.org/package=readxl>.
 30. Wickham H. tidy: Tidy Messy Data. R package version 1.1.3, 2021. Available at: <https://CRAN.R-project.org/package=tidy>.
 31. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.7, 2021. Available at: <https://CRAN.R-project.org/package=dplyr>.
 32. Kuhn M. caret: Classification and Regression Training. R package version 6.0-86, 2021. Available at: <https://CRAN.R-project.org/package=caret>.
 33. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
 34. De Vita A, Ravenna SE, Covino M, et al. Electrocardiographic findings and clinical outcome in patients with COVID-19 or other acute infectious respiratory diseases. *J Clin Med* 2020;9:3647.
 35. Sanders GD, Lowenstern A, Borre E, et al. Stroke Prevention in Patients With Atrial Fibrillation: A Systematic Review Update. In: *Services USDoHaH*, ed. Rockville, MD: Agency for Healthcare Research and Quality; 2018.
 36. Wiley Z, Ross-Driscoll K, Wang Z, Smothers L, Mehta AK, Patzer RE. Racial and ethnic differences and clinical outcomes of COVID-19 patients presenting to the emergency department. *Clin Infect Dis* 2021.04.02; <https://doi.org/10.1093/cid/ciab290>.
 37. Jambukia SH, Dabhi VK, Prajapati HB. Classification of ECG signals using machine learning techniques: A survey. Paper presented at: 2015 International Conference on Advances in Computer Engineering and Applications; 19-20 March 2015, 2015.
 38. Hicks SA, Isaksen JL, Thambawita V, et al. Explaining deep neural networks for knowledge discovery in electrocardiogram analysis. *Sci Rep* 2021;11:10949.
 39. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;394:861–867.
 40. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;25:70–74.