ESC European Society of Cardiology

European Heart Journal: Acute Cardiovascular Care (2024) **13**, 472–480
https://doi.org/10.1093/ehjacc/zuae037

ORIGINAL SCIENTIFIC PAPER
Acute Coronary Syndromes

Development and external validation of a dynamic risk score for early prediction of cardiogenic shock in cardiac intensive care units using machine learning

Yuxuan Hu [1]*, Albert Lui[2], Mark Goldstein[3], Mukund Sudarshan[3], Andrea Tinsay[4], Cindy Tsui[4], Samuel D. Maidman[4], John Medamana[4], Neil Jethani[2,3], Aahlad Puli[3], Vuthy Nguy[5], Yindalon Aphinyanaphongs[5], Nicholas Kiefer[1], Nathaniel R. Smilowitz[1], James Horowitz[1], Tania Ahuja[6], Glenn I. Fishman[1], Judith Hochman[1], Stuart Katz[1], Samuel Bernard[1], and Rajesh Ranganath[3,5,7]

[1]Leon. H. Charney Division of Cardiology, NYU Langone Health, 550 1st Avenue, New York, NY 10016, USA; [2]NYU Grossman School of Medicine, New York, USA; [3]Courant Institute of Mathematics, New York University, New York, USA; [4]Department of Medicine, NYU Langone Health, New York, USA; [5]Department of Population Health, NYU Langone Health, New York, USA; [6]Department of Pharmacy, NYU Langone Health, New York, USA; and [7]Center for Data Science, New York University, New York, USA

Received 31 August 2023; revised 11 March 2024; accepted 19 March 2024; online publish-ahead-of-print 22 March 2024

Aims

Myocardial infarction and heart failure are major cardiovascular diseases that affect millions of people in the USA with morbidity and mortality being highest among patients who develop cardiogenic shock. Early recognition of cardiogenic shock allows prompt implementation of treatment measures. Our objective is to develop a new dynamic risk score, called CShock, to improve early detection of cardiogenic shock in the cardiac intensive care unit (ICU).

Methods and results

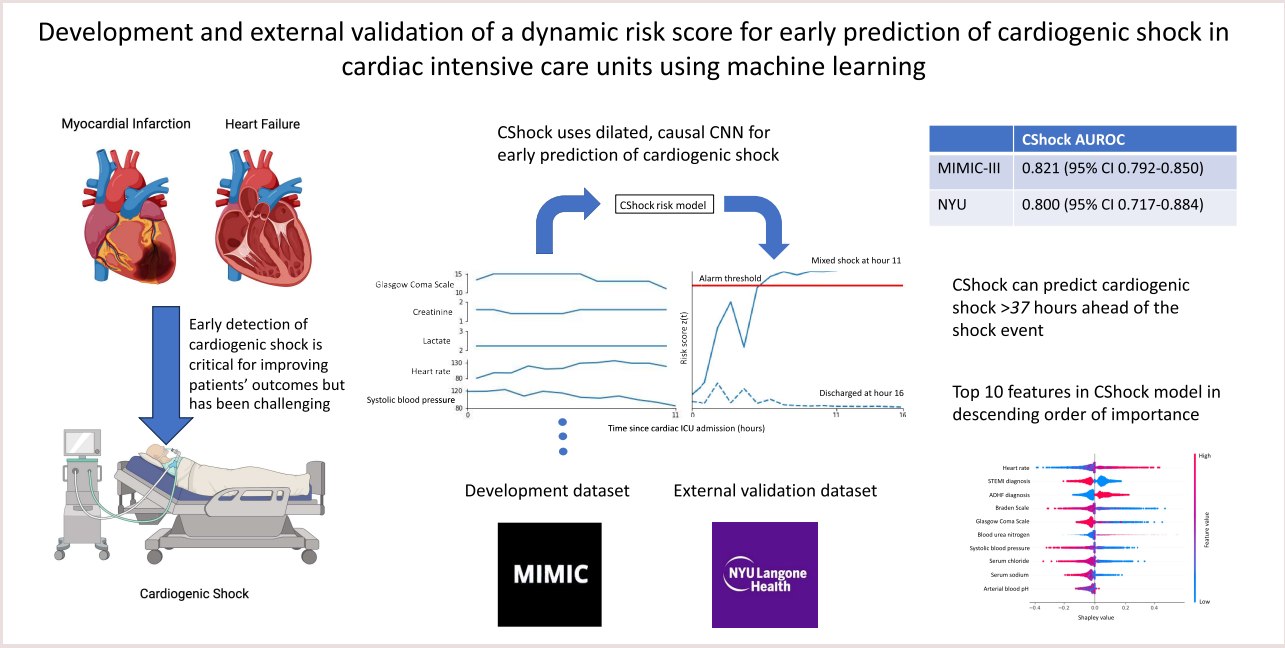
We developed and externally validated a deep learning-based risk stratification tool, called CShock, for patients admitted into the cardiac ICU with acute decompensated heart failure and/or myocardial infarction to predict the onset of cardiogenic shock. We prepared a cardiac ICU dataset using the Medical Information Mart for Intensive Care-III database by annotating with physician-adjudicated outcomes. This dataset which consisted of 1500 patients with 204 having cardiogenic/mixed shock was then used to train CShock. The features used to train the model for CShock included patient demographics, cardiac ICU admission diagnoses, routinely measured laboratory values and vital signs, and relevant features manually extracted from echocardiogram and left heart catheterization reports. We externally validated the risk model on the New York University (NYU) Langone Health cardiac ICU database which was also annotated with physician-adjudicated outcomes. The external validation cohort consisted of 131 patients with 25 patients experiencing cardiogenic/mixed shock. CShock achieved an area under the receiver operator characteristic curve (AUROC) of 0.821 (95% CI 0.792–0.850). CShock was externally validated in the more contemporary NYU cohort and achieved an AUROC of 0.800 (95% CI 0.717–0.884), demonstrating its generalizability in other cardiac ICUs. Having an elevated heart rate is most predictive of cardiogenic shock development based on Shapley values. The other top 10 predictors are having an admission diagnosis of myocardial infarction with ST-segment elevation, having an admission diagnosis of acute decompensated heart failure, Braden Scale, Glasgow Coma Scale, blood urea nitrogen, systolic blood pressure, serum chloride, serum sodium, and arterial blood pH.

Conclusion

The novel CShock score has the potential to provide automated detection and early warning for cardiogenic shock and improve the outcomes for millions of patients who suffer from myocardial infarction and heart failure.

* Corresponding author. Tel: 646-929-7875, Email: yuxuanh@gmail.com

Graphical Abstract



Keywords Heart failure • Myocardial infarction • Data science • Machine learning • Cardiac critical care

Introduction

Cardiogenic shock is a life-threatening condition that is characterized by reduced cardiac output in the presence of adequate intravascular volume, resulting in tissue hypoxia. Myocardial infarctions[1,2] and acute decompensated heart failure[3] are the two most common aetiologies for cardiogenic shock.[4] Mortality for cardiogenic shock remains high (30–50%).[1,4,5] Early identification of cardiogenic shock can facilitate interventions that may mitigate the consequences of prolonged end-organ insult,[6,7] such as rapid employment of haemodynamic support with pharmacologic and nonpharmacologic agents, engagement of a dedicated shock team, and prompt transfer of patients at lower-acuity hospitals to a tertiary high-volume shock hub.[1] In the setting of acute myocardial infarction, early recognition of patients at high risk for cardiogenic shock allows providers to proceed with urgent revascularization of the infarct-related coronary artery.[5] Therefore, timely recognition of patients at high risk for cardiogenic shock is crucial for improving the care of patients and their outcomes.

Despite the potential benefits, early identification of cardiogenic shock has been challenging. Intensive care unit (ICU) providers are presented with tremendous amounts of data generated from multiple sources including laboratory measurements, vital signs, haemodynamics, and cardiac function studies. The limited ability of human providers to process, interpret, and act upon all the data stored in the electronic health record (EHR) in a timely fashion can lead to poor patient outcomes. Moreover, it is difficult to provide around-the-clock monitoring and risk assessment of cardiac ICU patients by human caregivers, especially in settings with low provider-to-patient ratios.

Furthermore, there lacks an effective risk stratification tool for the development of cardiogenic shock. Current risk stratification strategies such as the Society for Cardiovascular Angiography and Intervention classification schema[8,9] lack specific quantitative criteria, making it challenging to apply clinically. Well-established scores such as CardShock,[10] intraaortic balloon pump (IABP)-shock II,[11] and Acute Physiology and Chronic Health Evaluation II[12] were developed to predict outcomes such as in-hospital or 30-day mortality rather than the development of cardiogenic shock. The Observatoire Régional Breton sur l'Infarctus (ORBI) risk score[13] was developed for predicting cardiogenic shock and is transparent and can be applied at the bedside. However, the study cohort for the ORBI risk score was myocardial infarction with ST-elevation (STEMI) patients only.

Early identification and treatment of cardiogenic shock could lead to better outcomes, shorter cardiac ICU length of stay, and fewer complications in cardiac ICUs.[6] To improve risk stratification of cardiac critical care patients and provide early warning of cardiogenic shock for patients admitted into cardiac ICU with (i) acute decompensated heart failure and/or (ii) STEMI or myocardial infarction without ST-elevation (NSTEMI), we develop a new dynamic risk score, called CShock, by training a machine learning model with a novel loss function designed specifically for risk scoring. Our goal is to predict the development of late cardiogenic shock after cardiac ICU admission for at-risk patients. We prepared a cardiac ICU dataset using Medical Information Mart for Intensive Care-III (MIMIC-III) database[14] by annotating with physician-adjudicated outcomes and relevant features manually extracted from echocardiogram and catheterization reports. This dataset was then used to train CShock. We externally validated the risk model on the New York University (NYU)

Langone Health cardiac ICU database which was also annotated with physician-adjudicated outcomes.

Methods

Preparation of a cardiac intensive care unit dataset with physician-adjudicated outcomes and cardiac features extracted from unstructured data

We used a publicly available critical care database MIMIC-III,[14] which has 8188 admissions involving at least one cardiac ICU stay during the hospitalization. The inclusion criterion for chart review was that patients needed to have a possible diagnosis of acute decompensated heart failure and/or myocardial infarction with or without ST-elevation. The exclusion criteria were patients with age <18 years old or >89 years old, total hospital stay <24 h, surgery ICU admission prior to cardiac ICU admission during the same hospitalization, and shock on arrival to cardiac ICU (details in Supplementary material online, *Methods*, section Chart review). The discharge summary and echocardiogram reports of the admissions that were eligible for chart review were subsequently split among four physicians to review, which provided information regarding additional exclusion criteria, admission diagnoses, outcomes (no shock vs. non-cardiogenic shock only vs. cardiogenic shock/mixed shock), left heart catheterization, and echocardiogram data (details in Supplementary material online, *Methods*, section Chart review). Patients who developed shock within 4 h of cardiac ICU admission were excluded due to the assumption that the patient was likely in a peri-shock state at the time of admission. Patients' outcomes were retrospectively adjudicated from chart review. Time of shock onset was determined as the earliest time after cardiac ICU admission when (i) systolic blood pressure was <90mmHg for at least 30 min and there was evidence of systemic hypoperfusion or (ii) pharmacologic agents/mechanical circulatory support was initiated to maintain systolic blood pressure >90mmHg,[7] consistent with the clinical criteria used in prior landmark trials on cardiogenic shock.[5,15] If there was a discrepancy between physician-adjudicated outcomes and objective data, i.e. physician-adjudicated outcome revealed no shock but there was shock based on structured data, a second physician reviewed both the discharge summary and the structured data to reconcile the discrepancy (details in Supplementary material online, *Methods*, section Chart review). Primary physician assessment of outcomes from discharge summary was blinded to the predictors. The validation of the primary assessment and the time of the event was done in an unblinded manner with respect to blood pressure, lactate, creatinine, and urine output.

Development of a dynamic risk score for prediction of cardiogenic shock

We developed CShock to determine the risk a patient has for developing cardiogenic shock after admission to the cardiac ICU. CShock is updated every hour after cardiac ICU admission. A patient was considered positive if the patient developed cardiogenic or mixed shock during their cardiac ICU stay; the patients who developed non-cardiogenic shock only or no shock in the study cohort were treated as negative cases. The features used to train the model for CShock included patient demographics, cardiac ICU admission diagnoses, routinely measured laboratory values and vital signs, and relevant features manually extracted from echocardiogram and left heart catheterization reports (see Supplementary material online, *Methods*, List of Features in the models). The physiological time series of the time-varying features from cardiac ICU admission to event (discharge if no shock; shock onset otherwise) were inputted into the CShock model. The model that underlies CShock was based on a dilated causal convolutional neural network (CNN) architecture, often used for time series modelling (See Supplementary material online, *Methods*, Model development for more details on model architecture).

The study cohort was split 50, 25, and 25% into training, validation, and testing in a four-fold cross-validation. Each fold had the same proportion of patients who developed cardiogenic shock vs. did not develop cardiogenic shock. To improve performance, the model was pretrained in an auxiliary task of predicting in-hospital mortality for ICU patients in MIMIC-III database. The receiving operator characteristic (ROC) curve and its corresponding area under the curve (AUC) were obtained by varying the alarm threshold that determined which patients were identified by the model as at risk for cardiogenic shock. We considered a comparison of CShock against the ORBI risk score[13] but could not accurately calculate the ORBI risk score, as many of the variables involved in the calculation of the ORBI score were not available such as history of previous stroke/transient ischemic attack (TIA), post-primary percutaneous coronary intervention (pPCI) thrombolysis in myocardial infarction (TIMI) flow <3, Killip class, and first medical contact-to-pPCI delay >90 min. In addition, the ORBI risk score was developed for patients admitted for STEMI only, whereas our cohort also included patients admitted for acute decompensated heart failure and NSTEMI. We followed TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis)[16] in reporting the results (see Supplementary material online, *Appendix* for details).

External validation

The model was externally validated using a dataset that comprised patients admitted to the cardiac ICU at NYU Langone Health who met the same inclusion and exclusion criteria as the study cohort. Patients' outcomes in the external validation cohort were retrospectively adjudicated by a physician. The NYU Langone Health patient cohort is more contemporary than the MIMIC-III cohort (MIMIC-III data were between 2001 and 2012 and NYU data were between 2018 and 2022). A subset of the features in the full model were available for external validation. We trained the model on the MIMIC-III cohort using the subset of features. The time-dependent variables also had missing indicators that took a value of 1 if the corresponding feature value was missing at each unique hour after cardiac ICU admission; 0 otherwise. Missing values were forward filled with the last known values and if the last known values were not available, imputed using population mean, as they were done for the development dataset. Primary physician assessment of outcomes from discharge summary was blinded to the predictors. The validation of the primary assessment and the time of the event was done in an unblinded manner with respect to blood pressure, lactate, creatinine, and urine output. The input features of the external validation cohort were fed into the portable model to produce the time-varying CShock risk score for cardiogenic shock. Receiving operator characteristic curve and its corresponding AUC were obtained by varying the alarm threshold that determined which patients were identified by the model as at risk for cardiogenic shock.

Model interpretation

We used Shapley values[17] to provide explanations for how each feature influenced predictions. We employed FastSHAP,[18] an efficient algorithm for calculating Shapley values.[17] The Shapley values were based on a FastSHAP's surrogate model that was trained to produce CShock risk scores given different feature subsets by masked prediction[19,20] rather than model the covariates[21] (see Supplementary material online, *Methods* for more details on model interpretation). Shapley values were determined for the features in the portable model using the training dataset.

Results

Cardiogenic shock early warning system

In total, 3220 admissions met the inclusion and exclusion criteria for chart review (*Figure 1*). After applying the additional criteria (details in Supplementary material online, *Methods*, section Study cohort) obtained from chart review, the final study cohort comprised 1500 patients (*Figure 1*), comparable in size to the cohorts used to derive other risk scores for cardiogenic shock.[11,22] A total of 1264 patients had no shock event before cardiac ICU discharge; 204 had cardiogenic/mixed shock; and 32 had non-cardiogenic shock only (*Figure 1*). The 204 patients who developed cardiogenic/mixed shock were considered positive; 1296 patients who developed non-cardiogenic shock only or no shock in the study cohort were treated as negative cases. The causes of non-cardiogenic shock were found to be septic or haemorrhagic shock during chart review. The average age of the study cohort in MIMIC-III was 68.1 ± 13.5. There were 918 male and 582 female

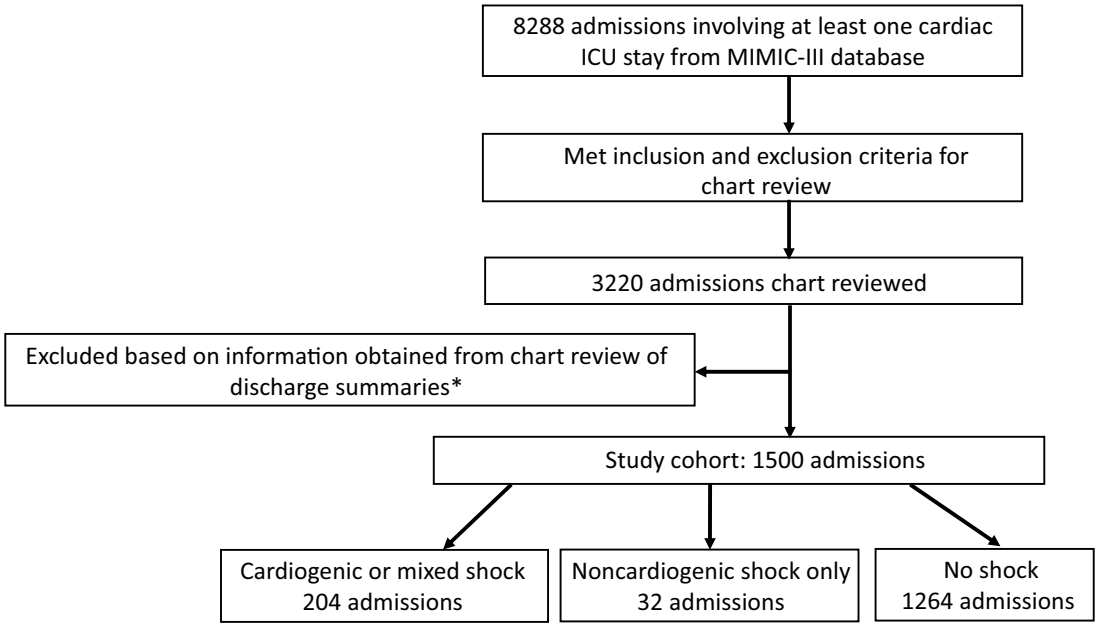


Figure 1 Study cohort. *The admissions that met the following criteria were excluded: development of acute respiratory distress syndrome, massive pulmonary embolism, cardiac tamponade and mechanical complications of myocardial infarction, urgent coronary artery bypass surgery, cardiac surgery in the 7 days prior to first cardiac intensive care unit admission, and cardiac arrest in the 7 days prior to first cardiac intensive care unit admission and admission to cardiac intensive care unit for diagnosis other than myocardial infarction with ST-elevation/myocardial infarction without ST-elevation/ADHF. Patients who developed shock within 4 h of cardiac intensive care unit admission were also excluded due to the assumption that the patient was likely in a peri-shock state at the time of admission.

patients in the MIMIC-III study cohort. Race, ethnicity, and socioeconomic status were not provided for the study cohort. Baseline characteristics of the study cohort are shown in *Table 1*.

There were 194 input variables, including 182 time-varying (physiological) and 12 static (e.g. demographic) features. The left panel of *Figure 2* demonstrates the physiological time series of five example features that were entered into the dilated CNN-based model, which then outputted the CShock score as shown by the blue lines in the right panel. The dashed blue line is the CShock score for a patient who develops no shock and gets discharged from the cardiac ICU at Hour 16; the solid blue line shows the CShock score for a patient who goes into mixed cardiogenic/non-cardiogenic shock at Hour 11. Cardiogenic shock detection occurs if a patient's CShock score exceeds the alarm threshold value.

The CShock score was able to achieve an area under the receiver operator characteristic curve (AUROC) of 0.821 (95% CI 0.792–0.850) (*Figure 3*). *Table 2A* shows positive predictive value (PPV)/negative predictive value (NPV), sensitivity/specificity, overall accuracy of the model, and how early the model can predict cardiogenic shock at different thresholds. Each row in the table corresponds to the CShock model having a PPV of one-fourth, one-third, and 0.4.

Model performance in different subgroups

We performed subgroup analysis based on age, sex, and aetiology (*Table 3*). The model performed better for the younger group than the older group. We also found that the model performed better for male patients [0.844 (95% CI 0.807–0.881)] compared with female patients [0.781 (95% CI 0.733–0.828)]. The model performed best in the STEMI subgroup compared with NSTEMI and acute decompensated heart failure subgroups. The NSTEMI and acute decompensated heart failure subgroups showed weaker prediction compared with the entire study cohort. The STEMI subgroup had better performance than the entire study cohort.

External validation

The model was externally validated using a dataset that comprised all patients admitted to the cardiac ICU at NYU Langone Health who met the same inclusion and exclusion criteria as the study cohort (see Supplementary material online, *Figure S1*). The external validation cohort consisted of 131 patients with 25 patients experiencing cardiogenic/mixed shock. The average age of the external validation cohort was 64.5 ± 14.9. There were 85 male and 46 female patients. Race, ethnicity, and socioeconomic status were not provided for the external validation cohort. Of the 182 time-varying input features in the original model, 58 input features are routinely measured in the EHR at NYU Langone Health. Therefore, we developed the portable model by training with 70 input features (58 were time-varying and 12 were static) using the MIMIC-III database (list of features is available in the Supplementary material online, *Methods*, List of Features in the models, denoted as included in the portable model). The portable model was able to achieve similar AUROC at 0.806 (95% CI 0.777–0.835) on the study cohort. It was then externally validated using the NYU Langone Health cardiac ICU dataset and obtained an AUROC of 0.800 (95% CI 0.717–0.884).

Model interpretation

Figure 4A presents the 10 most important features, i.e. with the largest Shapley values; each dot represents the feature values averaged over time of an individual patient from the training dataset with a higher value being more red and a lower value being more blue. Positive and negative Shapley values are associated with an increase or decrease in the CShock score, i.e. risk of cardiogenic shock development, respectively.

Table 1 Baseline characteristics of the study cohort

	Did not have cardiogenic shock	Cardiogenic/mixed shock	*P*-value
N	1296	204	
Age, mean (SD)	67.8 (13.5)	69.9 (13.9)	0.045
Female, *n* (%)	491 (37.9)	91 (44.6)	0.079
STEMI, *n* (%)	536 (41.4)	30 (14.7)	<0.001
NSTEMI, *n* (%)	411 (31.7)	52 (25.5)	0.088
ADHF, *n* (%)	568 (43.8)	161 (78.9)	<0.001
Creatinine, mean (SD)	1.6 (1.6)	1.9 (1.3)	0.002
Glucose, mean (SD)	158.8 (78.4)	140.5 (66.5)	<0.001
Heart rate, mean (SD)	81.0 (16.8)	87.0 (18.4)	<0.001
SBP, mean (SD)	130.1 (22.2)	114.0 (19.6)	<0.001
DBP, mean (SD)	68.6 (15.6)	61.2 (13.1)	<0.001
EF, mean (SD)	41.6 (13.9)	37.1 (15.0)	0.004
Previous CABG, *n* (%)	132 (14.6)	20 (24.7)	0.024
Time to event, mean (SD)	54.2 (41.1)	41.0 (42.5)	<0.001

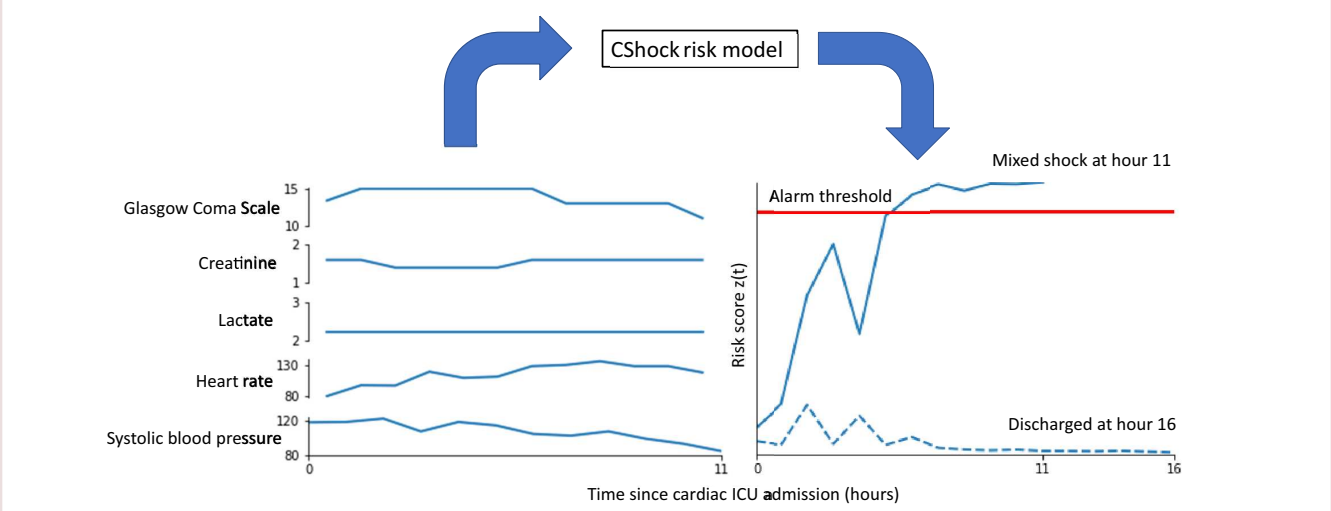


Figure 2 The physiological time series of five example features from cardiac intensive care unit admission to event (discharge if no shock; shock onset otherwise) are shown for a patient developing mixed cardiogenic/non-cardiogenic shock (left panel). Features displayed are the Glasgow Coma Scale (a score for coma severity, see Supplementary material online, *Methods*, List of Features in the models for the three components of the score), creatinine, lactate, heart rate, and systolic blood pressure without support. The physiological time series were fed into the convolutional neural network-based model, which outputted CShock scores. CShock scores (in the right panel) were calculated for each hour from cardiac intensive care unit admission until the event. Two patients' CShock scores are shown here: one patient develops no shock and gets discharged from cardiac intensive care unit at Hour 16 (dashed line); the other patient goes into mixed shock at Hour 11 (solid line). Cardiogenic shock detection occurs if a patient's CShock score exceeds the alarm threshold value. The horizontal line indicates the detection threshold corresponding to a sensitivity of 0.8.

For instance, heart rate is the most predictive feature of cardiogenic shock development and having an elevated heart rate is associated with an increased risk of cardiogenic shock development. *Figure 4B* shows how the AUROC changes as more features are added in descending order of importance into the FastSHAP surrogate model. AUROC improved as more features were included in the model and the best AUROCs were similar to that obtained using the portable model. With only the 10 most important features (as displayed in *Figure 4A*), the FastSHAP surrogate model was able to achieve AUROC (vertical line in *Figure 4B*) similar to that when more features were included. We used the top 10 variables in *Figure 4A* at the time of cardiac ICU admission to predict cardiogenic shock with a logistic regression model and found the AUROC of the reduced model to be 0.758 (95% CI 0.725–0.792), which is lower than the AUROC for the full (0.821, 95% CI 0.792–0.850) and portable (0.806, 95% CI 0.777–0.835) models. We reported the averaged beta coefficients and odds ratio of the reduced model in *Table 4*.

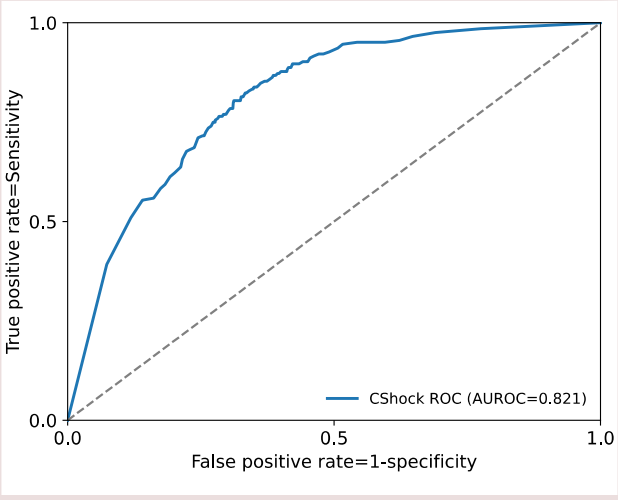


Figure 3 Cross-validated receiving operator characteristic for predicting cardiogenic shock using the developed CShock score with the Medical Information Mart for Intensive Care-III dataset.

Table 2 (A) Positive predictive value/negative predictive value, sensitivity/specificity, and overall accuracy of the model at different thresholds. Each row in the table corresponds to the CShock model having a positive predictive value of one-fourth, one-third, and 0.4. (B) Area under the receiver operator characteristic curve and area under precision recall curve (AUPRC) of the CShock model (in comparison, the prevalence of cardiogenic shock/mixed shock in the study cohort was 13.6%)

(A)

PPV	NPV	Sensitivity	Specificity	Accuracy	How early it can predict cardiogenic shock (h)
1/4	0.970	0.878	0.584	0.624	38.9
1/3	0.954	0.765	0.759	0.759	37.3
0.4	0.920	0.500	0.883	0.831	37.8

(B)

AUROC	AUPRC
0.821	0.387

Discussion

Myocardial infarction and heart failure affect nine and six million people in the US, respectively.[23] Cardiogenic shock is a common complication of myocardial infarction and heart failure and its occurrence is associated with substantial health and economic costs.[1] Early recognition of cardiogenic shock is of paramount importance as it facilitates timely treatments that could potentially prevent the vicious spiral of cardiogenic shock and reduce the health and financial burden. However, early identification of cardiogenic shock has remained challenging.[6] We have demonstrated that the CShock score can predict cardiogenic shock with excellent AUROC. The early warning system was tested in an independent and more contemporary patient cohort from a different hospital and showed comparable performance as in the development data.

At several different PPVs, CShock was able to achieve good NPV (>0.92) and predict cardiogenic shock, on average, >37 h ahead of the shock event. With increasing PPV, sensitivity decreased and specificity increased as expected. ORBI is an easy-to-use score with excellent performance for its study cohorts.[13] Unfortunately, some variables in the ORBI score calculation were not available for our study cohort such as history of previous stroke/TIA, post-pPCI TIMI flow <3, Killip class, and first medical contact-to-pPCI delay >90 min. Some variables such as post-pPCI TIMI flow <3 and first medical contact-to-pPCI delay >90 min are not relevant for acute decompensated heart failure patients subgroup in our study cohort, as ORBI was developed for STEMI patients only and our study cohort also had NSTEMI and acute decompensated heart failure patients.

The CShock risk score performed better for the younger subgroup compared with the older subgroup. This could be due to the fact that the elderly tend to have more comorbidities that were embedded only in the unstructured data of the EHR (e.g. clinical notes) and thus unaccounted for in the model. The CShock risk score also performed better for the male cohort compared with the female cohort. This was possibly a result of the study cohort having more male patients (918) as compared with female patients (582) and thus the model was better trained for the male cohort. The CShock risk score performed best for the STEMI cohort followed by the NSTEMI and acute decompensated heart failure (ADHF) subgroups, which is inversely associated with the prevalence of cardiogenic shock in the three subgroups.

The portable model is readily deployable. The portable model with 70 features that are routinely captured in the EHR database at NYU Langone Health achieved an AUROC of 0.806 (95% CI 0.777–0.835). This model could be easily implemented at other facilities with modern EHR systems, as most of the variables in the portable model are available in modern EHR systems. Using the portable model instead of the full model for EHR deployment can make implementation easier. It is interesting to note that the portable model performed on par with the full model despite having fewer features. This could be because (i) it is difficult to manually extract features from unstructured data (e.g. echocardiogram and catheterization data) and the lack of data can jeopardize the performance of the machine learning model; it is possible if those features were available as structured data, the performance of the full model would be even better. (ii) The structured data captured most of the predictive information that would be embedded in unstructured data; incorporating unstructured data will not improve the performance of the machine learning model beyond what has been achieved with structured data only. We were also able to achieve lower but still predictive AUROC using the top 10 variables in *Figure 4A* at the time of cardiac ICU admission to predict cardiogenic shock in a logistic regression model; this reduced model can be a good point-of-care tool at the time of admission.

Previous machine learning studies in broader medicine have relied on International Classification of Diseases (ICD) codes and natural language processing of clinical notes[24] to determine outcomes such as sepsis and septic shock. These types of approaches for determining outcomes are often inaccurate.[25] In contrast, the outcome events for the cardiac ICU patients were carefully adjudicated by physicians in this study. This will allow the CShock score developed by this project to translate more easily into clinical implementation.

We constructed a novel loss function specifically for risk scoring. The dilated causal CNN architecture along with the novel loss function are of broad applicability and can be employed for risk stratification in multiple clinical settings such as sepsis and pulmonary embolism. Pretraining is a relatively new technique for improving the performance

Table 3 Area under the receiver operator characteristic curve, number of patients, and number of patients with cardiogenic/mixed shock (percentage) for each age group, sex, and aetiology

	AUROC	Number of patients	Number of patients with cardiogenic/mixed shock (percentage)
Age (years)			
<25 percentile (<59)	0.856 (95% CI 0.788–0.924)	368	40 (10.9%)
Between 25 and 50 percentile (at least 59, <70)	0.816 (95% CI 0.757–0.875)	361	44 (12.2%)
Between 50 and 75 percentile (at least 70, <80)	0.812 (95% CI 0.758–0.867)	388	58 (14.9%)
>75 percentile (at least 80)	0.772 (95% CI 0.712–0.832)	383	62 (16.2%)
Sex			
Female	0.781 (95% CI 0.733–0.828)	582	91 (15.6%)
Male	0.844 (95% CI 0.807–0.881)	918	113 (12.3%)
Aetiology			
STEMI	0.880 (95% CI 0.808–0.953)	566	75 (13.3%)
NSTEMI	0.780 (95% CI 0.716–0.845)	463	66 (14.3)
ADHF	0.756 (95% CI 0.716–0.796)	729	110 (15.1%)

of physiological time series. It has been widely used for improving the performance of imaging analysis[26] and also for text via large language models.[27] Recently, there have been a few studies employing pretraining with labels that are related but not identical to the primary outcome of interest to improve the performance of physiological time series models.[28] We therefore employed a similar technique to improve the performance of our model in predicting cardiogenic shock by pretraining our model with a label such as mortality. Cardiogenic shock and mortality share certain physiological characteristics such as deranged metabolism. In this case, we showed that pretraining with an auxiliary task of predicting mortality for ICU patients in the MIMIC-III database improved the AUROC of the cardiogenic shock model from 0.750 (95% CI 0.715–0.786) to 0.821 (95% CI 0.792–0.850). We demonstrated that pretraining with a different cohort and a label that is related but not identical to the primary outcome of interest such as mortality can be a useful strategy to improve physiological time series model performance when there is a limited number of patients who meet all the inclusion and exclusion criteria of the study.

Shapley values are widely used to explain deep learning models and we used Shapley values to explain our model. We implemented a state-of-the-art algorithm for computing Shapley values[18] to reduce the amount of time it would take to calculate Shapley values relative to the traditional way that relies on optimizing the loss given by the least squares characterization for each new data point.[17] This model interpretation analysis using Shapley values indicated that an admission diagnosis of STEMI is associated with a lower risk of development of cardiogenic shock, whereas having an admission diagnosis of acute decompensated heart failure is associated with a higher risk of development of cardiogenic shock. This likely reflects advancements in acute coronary syndrome (ACS) management strategies over the years, which makes ACS patients less likely to develop cardiogenic shock. It could also relate to the fact that patients who were admitted with STEMI and acute decompensated heart failure have a higher risk of developing cardiogenic shock (12.7% in MIMIC-III study cohort), while patients admitted with STEMI but no acute decompensated heart failure have much lower risk of developing cardiogenic shock (4.5% in MIMIC-III study cohort). Among the 10 most important features, low Braden Scale (a risk score for identifying patients at risk for pressure ulcers,[29] see Supplementary material online, *Methods*, List of Features in the models for the six variables in Braden Scale), Glasgow Coma Scale, systolic blood pressure without support, and serum sodium are predictive of cardiogenic shock development, as expected clinically.[9,30,31] Lactate, a commonly used lab measurement to assess for severity of cardiogenic shock, is not among the 10 most important features; conceivably other top features captured most of the predictive information that would be embedded in lactate.

There have been other recent efforts to predict cardiogenic shock with machine learning.[32,33] However, they only considered patients who required inotropes/mechanical circulatory support as developing cardiogenic shock and used the time to initiate supportive measures as shock onset, which are not in alignment with the clinical criteria used in previous landmark trials;[5,15] the early presentation of cardiogenic shock with low blood pressure and end-organ hypoperfusion would be missed by their algorithms. In addition, both studies lacked external validation to demonstrate generalizability in other populations. 1% of the study cohort developed cardiogenic shock in the study by Rahman *et al.*,[33] which resulted in the best PPV being only 11% (with recall/sensitivity 27%) and would lead to the algorithm sounding many false alarms secondary to extreme class imbalance and missing the majority of cardiogenic shock patients. However, the comparison is limited as their score was developed using data from floors/emergency rooms and thus the population would have a lower prevalence; their reported PPVs of cardiogenic shock would be expected to be lower. Despite these challenges, like developing a dynamic risk score for early prediction of cardiogenic shock in cardiac ICU, developing risk scores for the general floor is an important clinical problem. It can also be interesting to extend the CShock model to non-ICU settings in future studies. The study by Chang *et al.*[32] relied on ICD codes to determine outcomes, which would be inaccurate as discussed above; they excluded mixed cardiogenic/non-cardiogenic shock patients from the study cohort, which could limit the algorithm's applicability. Lastly, machine learning techniques have been improving for better prediction of cardiogenic shock. The recent publication by Jajcay *et al.* developed a data processing pipeline for cardiogenic shock prediction using machine learning and achieved good classification performance.[34]

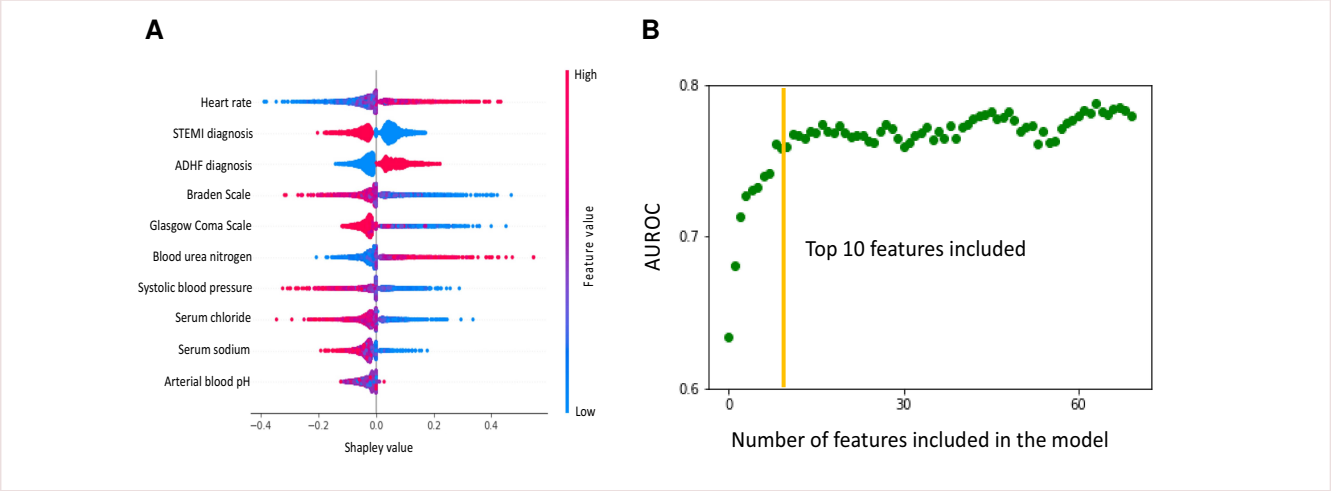


Figure 4 (A) The 10 most important features based on Shapley values in descending order of importance (1 heart rate, 2 having an admission diagnosis of myocardial infarction with ST-elevation, 3 having an admission diagnosis of acute decompensated heart failure, 4 Braden Scale, 5 Glasgow Coma Scale, 6 blood urea nitrogen, 7 systolic blood pressure, 8 serum chloride, 9 serum sodium, 10 arterial blood pH). Each dot represents the feature values averaged over time of an individual patient from the training dataset with a higher value being more red and a lower value being more blue. Positive and negative Shapley values are associated with an increase or decrease in the risk of cardiogenic shock development, respectively. (B) Area under the receiver operator characteristic curve obtained with an increasing number of features included in the FastSHAP surrogate model using the evaluation dataset. Vertical line indicates the area under the receiver operator characteristic curve when only the 10 most important features (as displayed in *Figure 4A*) were included in the FastSHAP surrogate model that was trained to predict on subsets of features.

Table 4 Beta coefficients averaged across folds and the corresponding odds ratio for the reduced model

Averaged beta coefficient	0.265	−0.233	0.480	−0.377	−0.262	0.341	−0.843	−0.274	−0.049	−0.598
Odds ratio	1.303	0.792	1.617	0.686	0.769	1.407	0.431	0.760	0.952	0.550

Their pre-processing pipeline could prove helpful for improving the performance of future machine learning-based risk models for cardiogenic shock including our CShock risk model.

Given the performance of our CShock score, we hypothesize that this CNN-based early warning model can help cardiac ICU teams clinically integrate complex data and more rapidly identify patients at risk for cardiogenic shock. A recent study using the Critical Care Cardiology Trials Network data showed that a multidisciplinary shock team approach improved outcomes in cardiogenic shock[6] and this machine learning-based model could build upon this improvement, alerting shock teams to impending cardiogenic shock. Each cardiac ICU team can choose a PPV value that is best for their team in terms of sensitivity so that they can achieve a balance in capturing most of the cardiogenic shock patients but also minimizing alert fatigue. The exact way to implement the algorithm can take on different forms: it can be an alert notifying the team of the possible impending cardiogenic shock; alternatively, the CShock score of the patient developing cardiogenic shock can be displayed. The best way to implement this algorithm will depend on each cardiac ICU team's preference and requires further studies with principles from human–computer interaction. The CShock model is able to predict cardiogenic shock at least 37 h in advance at several PPV thresholds. It can, therefore, help alert the shock team in advance to give the team ample time to prepare for impending cardiogenic shock. In summary, we demonstrate that the CShock score we developed has the potential to provide automated detection and early warning for cardiogenic shock and improve the outcomes for millions of patients who suffer from myocardial infarction and heart failure.

Limitations

The major limitation of the study is that the CShock model was developed and externally validated using cardiac ICU data from two academic medical centres in metropolitan areas and that the MIMIC-III database was collected from 2001 to 2012, which is not contemporary. Future studies should be conducted to evaluate the performance of the CShock model in other contemporary cardiac ICU databases of different hospital practices and different patient populations, especially in populations that are different from the typical demographics seen in academic centres in the Northeast USA. In addition, even though we treated the patients who developed non-cardiogenic shock as negative cases in this manuscript, this could present a bias as septicaemia can lead to cardiac suppression. Lastly, the availability of some input features such as pulmonary artery catheter data, lactate, and arterial blood gas data could be that the team is worried of impending shock. Using those features in training the model could introduce a bias to the model.

Supplementary material

Supplementary material is available at *European Heart Journal: Acute Cardiovascular Care* online.

Author contribution

Y.H., M.G., and R.R. designed the experiments. Y.H. and A.L. preprocessed and cleaned the MIMIC data. Y.H., A.T., C.T., S.M., and J.M. reviewed the MIMIC charts. V.N., Y.A., N.K., Ja.H., and T.A. selected and provided the NYU clinical data and context. Y.H. preprocessed, cleaned, and chart reviewed the NYU data. Y.H., M.G., and R.R. developed the pipeline for the CShock risk model. Y.H. and M.G. implemented the CShock model. Y.H. implemented pretraining, analysed subgroup performance, and externally validated the model with NYU data. Y.H., M.G., N.J., and R.R. devised and implemented the model interpretation strategy. M.S., A.P., N.S., and S.K. contributed to various analyses of the data. Y.H., A.L., M.G., and R.R. conceived and directed the project. Y.H. and M.G. made the figures. Y.H. and M.G. drafted the manuscript. G.F., Ju.H., S.K., S.B., and R.R. provided critical review.

Funding

Y.H. was supported by an National Institute of Health T32 grant (T32HL098129). R.R. was supported by National Institute of Health/National Heart, Lung, and Blood Institute Award R01HL148248, National Science Foundation Award 1922658 NRT-HDR: FUTURE Foundations, Translation, and Responsibility for Data Science, and National Science Foundation CAREER Award 2145542. N.J. was supported by an National Institute of Health T32 grant (T32GM007308). The funder of the study (National Institute of Health and National Science Foundation) had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication.

Conflict of interest: N.S. reports consulting for Abbott Vascular as a member of an advisory board.

Data availability

The full MIMIC dataset can be downloaded from http://physionet.org. The labels for cardiogenic shock will be made available on request based on the MIMIC policies. The computer code used in this research will be available at www.github.com/yuxuanhu12/cshock under an open-source license.

References

1. van Diepen S, Katz JN, Albert NM, Henry TD, Jacobs AK, Kapur NK *et al.* Contemporary management of cardiogenic shock: a scientific statement from the American Heart Association. *Circulation* 2017;**136**:e232–e268.
2. Wayangankar SA, Bangalore S, McCoy LA, Jneid H, Latif F, Karrowni W *et al.* Temporal trends and outcomes of patients undergoing percutaneous coronary interventions for cardiogenic shock in the setting of acute myocardial infarction: a report from the CathPCI registry. *JACC Cardiovasc Interv* 2016;**9**:341–351.
3. Chioncel O, Parissis J, Mebazaa A, Thiele H, Desch S, Bauersachs J *et al.* Epidemiology, pathophysiology and contemporary management of cardiogenic shock—a position statement from the Heart Failure Association of the European Society of Cardiology. *Eur J Heart Fail* 2020;**22**:1315–1341.
4. Berg DD, Bohula EA, van Diepen S, Katz JN, Alviar CL, Baird-Zars VM *et al.* Epidemiology of shock in contemporary cardiac intensive care units. *Circ Cardiovasc Qual Outcomes* 2019;**12**:e005618.
5. Hochman JS, Sleeper LA, Webb JG, Sanborn TA, White HD, Talley JD *et al.* Early revascularization in acute myocardial infarction complicated by cardiogenic shock. SHOCK investigators. Should we emergently revascularize occluded coronaries for cardiogenic shock. *N Engl J Med* 1999;**341**:625–634.
6. Papolos AI, Kenigsberg BB, Berg DD, Alviar CL, Bohula E, Burke JA *et al.* Management and outcomes of cardiogenic shock in cardiac ICUs with versus without shock teams. *J Am Coll Cardiol* 2021;**78**:1309–1317.
7. Reyentovich A, Barghash MH, Hochman JS. Management of refractory cardiogenic shock. *Nat Rev Cardiol* 2016;**13**:481–492.
8. Baran DA, Grines CL, Bailey S, Burkhoff D, Hall SA, Henry TD *et al.* SCAI clinical expert consensus statement on the classification of cardiogenic shock: this document was endorsed by the American College of Cardiology (ACC), the American Heart Association (AHA), the Society of Critical Care Medicine (SCCM), and the Society of Thoracic Surgeons (STS) in April 2019. *Catheter Cardiovasc Interv* 2019;**94**:29–37.
9. Naidu SS, Baran DA, Jentzer JC, Hollenberg SM, van Diepen S, Basir MB *et al.* SCAI SHOCK stage classification expert consensus update: a review and incorporation of validation studies: this statement was endorsed by the American College of Cardiology (ACC), American College of Emergency Physicians (ACEP), American Heart Association (AHA), European Society of Cardiology (ESC) Association for Acute Cardiovascular Care (ACVC), International Society for Heart and Lung Transplantation (ISHLT), Society of Critical Care Medicine (SCCM), and Society of Thoracic Surgeons (STS) in December 2021. *J Am Coll Cardiol* 2022;**79**:933–946.
10. Harjola VP, Lassus J, Sionis A, Kober L, Tarvasmaki T, Spinar J *et al.* Clinical picture and risk prediction of short-term mortality in cardiogenic shock. *Eur J Heart Fail* 2015;**17**:501–509.
11. Poss J, Koster J, Fuernau G, Eitel I, de Waha S, Ouarrak T *et al.* Risk stratification for patients in cardiogenic shock after acute myocardial infarction. *J Am Coll Cardiol* 2017;**69**:1913–1920.
12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;**13**:818–829.
13. Auffret V, Cottin Y, Leurent G, Gilard M, Beer JC, Zabalawi A *et al.* Predicting the development of in-hospital cardiogenic shock in patients with ST-segment elevation myocardial infarction treated by primary percutaneous coronary intervention: the ORBI risk score. *Eur Heart J* 2018;**39**:2090–2102.
14. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;**3**:160035.
15. Thiele H, Zeymer U, Neumann FJ, Ferenc M, Olbrich HG, Hausleiter J *et al.* Intraaortic balloon support for myocardial infarction with cardiogenic shock. *N Engl J Med* 2012;**367**:1287–1296.
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 2015;**102**:148–158.
17. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;**30**.
18. Jethani N, Sudarshan M, Covert IC, Lee S-I, Ranganath R. Fastshap: real-time shapley value estimation. In: *International Conference on Learning Representations*, 2021.
19. Covert IC, Lundberg S, Lee S-I. Explaining by removing: a unified framework for model explanation. *J Mach Learn Res* 2021;**22**:9477–9566.
20. Jethani N, Sudarshan M, Aphinyanaphongs Y, Ranganath R. Have we learned to explain?: how interpretability methods can learn to encode predictions in their interpretations. In: *International Conference on Artificial Intelligence and Statistics*, 2021. PMLR.
21. Sudarshan M, Tansey W, Ranganath R. Deep direct likelihood knockoffs. *Adv Neural Inf Process Syst* 2020;**33**:5036–5046.
22. Beer BN, Jentzer JC, Weimann J, Dabboura S, Yan I, Sundermeyer J *et al.* Early risk stratification in patients with cardiogenic shock irrespective of the underlying cause—the cardiogenic shock score. *Eur J Heart Fail* 2022;**24**:657–667.
23. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW *et al.* Heart disease and stroke statistics-2021 update: a report from the American Heart Association. *Circulation* 2021;**143**:e254–e743.
24. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (trewscore) for septic shock. *Sci Transl Med* 2015;**7**:299ra122.
25. Misset B, Nakache D, Vesin A, Darmon M, Garrouste-Orgeas M, Mourvillier B *et al.* Reliability of diagnostic coding in intensive care patients. *Crit Care* 2008;**12**:R95.
26. Huang S-C, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digit Med* 2023;**6**:74.
27. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P *et al.* Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020;**33**:1877–1901.
28. McDermott M, Nestor B, Kim E, Zhang W, Goldenberg A, Szolovits P *et al.* A comprehensive EHR timeseries pre-training benchmark. In: *Proceedings of the Conference on Health, Inference, and Learning*, 2021.
29. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden scale for predicting pressure sore risk. *Nurs Res* 1987;**36**:205–210.
30. Kataja A, Tarvasmaki T, Lassus J, Kober L, Sionis A, Spinar J *et al.* Altered mental status predicts mortality in cardiogenic shock—results from the CardShock study. *Eur Heart J Acute Cardiovasc Care* 2018;**7**:38–44.
31. Leier CV, Dei Cas L, Metra M. Clinical relevance and management of the major electrolyte abnormalities in congestive heart failure: hyponatremia, hypokalemia, and hypomagnesemia. *Am Heart J* 1994;**128**:564–574.
32. Chang Y, Antonescu C, Ravindranath S, Dong J, Lu M, Vicario F *et al.* Early prediction of cardiogenic shock using machine learning. *Front Cardiovasc Med* 2022;**9**:862424.
33. Rahman F, Finkelstein N, Alyakin A, Gilotra NA, Trost J, Schulman SP *et al.* Using machine learning for early prediction of cardiogenic shock in patients with acute heart failure. *J Soc Cardiovasc Angiogr Interv* 2022;**1**:100308.
34. Jajcay N, Bezak B, Segev A, Matetzky S, Jankova J, Spartalis M *et al.* Data processing pipeline for cardiogenic shock prediction using machine learning. *Front Cardiovasc Med* 2023;**10**:1132680.