Anaesth Crit Care Pain Med 41 (2022) 101015

Contents lists available at ScienceDirect

Anaesthesia Critical Care & Pain Medicine

journal homepage: www.elsevier.com

Original Article

Computational signatures for post-cardiac arrest trajectory prediction: Importance of early physiological time series

Han B. Kim [a,c,g,1], Hieu T. Nguyen [a,g,1], Qingchu Jin [a,1], Sharmila Tamby [b], Tatiana Gelaf Romer [a], Eric Sung [a], Ran Liu [a], Joseph L. Greenstein [a], Jose I. Suarez [c,d,e], Christian Storm [f], Raimond L. Winslow [a], Robert D. Stevens [c,d,e,g,*]

[a] Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA
[b] Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA
[c] Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[d] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[e] Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[f] Department of Nephrology and Intensive Care Medicine, Charité-Universitätsmedizin, Berlin, Germany
[g] Laboratory of Computational Intensive Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Article history:
Available online

Keywords:
Cardiac arrest
Prognostication
Machine learning
Physiological time series
Precision medicine

ABSTRACT

Background: There is an unmet need for timely and reliable prediction of post-cardiac arrest (CA) clinical trajectories. We hypothesized that physiological time series (PTS) data recorded on the first day of intensive care would contribute significantly to discrimination of outcomes at discharge.
Patients and methods: Adult patients in the multicenter eICU database who were mechanically ventilated after resuscitation from out-of-hospital CA were included. Outcomes of interest were survival, neurological status based on Glasgow motor subscore (mGCS) and surrogate functional status based on discharge location (DL), at hospital discharge. Three machine learning predictive models were trained, one with features from the electronic health records (EHR), the second using features derived from PTS collected in the first 24 h after ICU admission (PTS_{24}), and the third combining PTS_{24} and EHR. Model performances were compared, and the best performing model was externally validated in the MIMIC-III dataset.
Results: Data from 2216 admissions were included in the analysis. Discrimination of prediction models combining EHR and PTS_{24} features was higher than models using either EHR or PTS_{24} for prediction of survival (AUROC 0.83, 0.82 and 0.79 respectively), neurological outcome (0.87, 0.86 and 0.79 respectively), and DL (0.80, 0.78 and 0.76 respectively). External validation in MIMIC-III (n = 86) produced similar model performance. Feature analysis suggested prognostic significance of previously unknown EHR and PTS_{24} variables.
Conclusion: These results indicate that physiological data recorded in the early phase after CA resuscitation contain signatures that are linked to post-CA outcome. Additionally, they attest to the effectiveness of ML for post-CA predictive modeling.

1. Introduction

Cardiac arrest (CA) is an abrupt cessation of myocardial function that affects more than half a million people in the United States annually. Patients resuscitated from cardiac arrest can experience a wide range of outcome trajectories, from complete recovery to death or severe neurologic disability [1]. A challenge in post-CA care is to accurately predict outcome, especially in the early phase when patients are treated in the intensive care unit (ICU). Physical examination findings and neurophysiological tests lack prognostic accuracy, especially when assessed less than 72 h after CA [2]. The recommended paradigm of multi-modality prognostication can be difficult to implement, and the predictive performance of its different elements, while studied individually, are unknown in aggregate [3]. Timely and accurate characterization of post-CA

* Corresponding author at: Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
E-mail address: rstevens@jhmi.edu (R. D. Stevens).
[1] Contributed equally

https://doi.org/10.1016/j.accpm.2021.101015

clinical severity and recovery probabilities could open the way for more effective, personalized therapeutic intervention.

Here, we propose a novel approach for post-CA clinical outcome prediction, based on the hypothesis that physiologic time series (PTS) signals, which are routinely recorded at the bedside, contain discriminative information and contribute to prognostic model performance. The hypothesis was tested using high-resolution data from the multicenter Philips eICU-CRD database [4], and externally validated on Medical Information Mart for Intensive Care (MIMIC) III database [5].

2. Patients and methods

2.1. Source of data

Data were extracted from the Philips eICU-Clinical Research Database (eICU), an open-source platform containing over 200,859 unique patient encounters in 208 hospitals across the US that use tele-ICU software [4]. This data consists of patients admitted to ICUs in 2014 and 2015 and contains granular EHR and PTS data. PTS data in eICU was recorded as a windowed median every 5 min and includes heart rate (HR), systolic and diastolic blood pressure (SBP, DBP), respiratory rate (RR), and O2 saturation by pulse oximeter (SpO_2). For the purpose of this research, we focused exclusively on PTS data collected in the first 24 h (PTS_{24}) after ICU admission for the five mentioned signals.

Data on post-CA patients within the Medical Information Mart for Intensive Care (MIMIC-III) database were used as the validation cohort [5]. This database contained 61,522 unique patient encounters at Beth Israel Deaconess Medical Center between 2001 and 2012.

2.2. Inclusion criteria

Patients were selected based on the following criteria: >18 years old, admitted to the ICU after CA, remained in the ICU for > 24 h, were mechanically ventilated, all three principal outcomes recorded, and had available PTS_{24} signals. We included only post-CA patients who were mechanically ventilated as they represent a subset with higher severity of illness in whom prognostication is most relevant. The same inclusion criteria were used to select patients in eICU and MIMIC-III.

2.3. Outcomes

The three outcomes of interest were survival, neurological status based on Glasgow Coma Score motor subscore (mGCS), and surrogate functional status based on discharge location (DL), all at hospital discharge. The neurological outcome indicator widely used in the CA population is the Cerebral Performance Category (CPC) score, [6] however, it was not recorded in eICU or MIMIC III. We defined a surrogate neurological outcome based on the motor subscore of the Glasgow Coma Score (mGCS) recorded at discharge, dichotomized as follows: mGCS of 6 (favorable outcome), mGCS $\leq$ 5 (unfavorable outcome). We defined a surrogate functional status based on hospital discharge location (DL), dichotomized as follows: discharge location of home or acute rehabilitation (favorable outcome), other location (unfavorable outcome). Availability of clinical outcome data varied in eICU and MIMIC (Table S1).

2.4. Variable extraction and selection

Based on prior studies, quantitative variables in the EHR relevant to post-CA patients were selected. EHR features were composed of 338 variables extracted from demographics, nurse examinations, laboratory results, medication, initial rhythm, admission diagnosis, and SOFA score/components, all limited to the first day of ICU admission. Precise documentation of targeted temperature management (TTM) was not available in the dataset. We developed a TTM identification algorithm based on temporal trends in body temperatures over the first 24 h; using this method, we classified 531 patients as having received TTM during the observation period.

Highly comparative time-series analysis (HTCSA), an automated feature extraction tool, was used to derive approximately 4000 derived features per PTS_{24} signal type [7]. HCTSA PTS_{24} derived features included distribution, correlation, trends, frequency, information theory features, and many others and has been successfully used across multiple fields of study [7,8].

The PTS_{24} and EHR features were pruned using a nested random forest feature importance ranking and variance inflation factor collinearity analysis for each clinical outcome. This resulted in three distinct feature spaces for each clinical outcome reducing the features to 410, 508, and 436 for survival outcome, mGCS neurologic outcome, and DL outcome respectively. A breakdown of HCTSA feature categories included in our model can be found in Table S2 and Table S3, and a more detailed explanation can be found in the supplemental text.

2.5. Model development

All models were developed exclusively using features extracted from the first 24 h following ICU admission. Four ML approaches were used to train the prediction models: generalized linear model (GLM), random forest (RF), gradient boost (XGBoost) and neural network (NN) to assess the performance of utilizing different feature subsets (EHR features, PTS_{24} features, and combined EHR and PTS_{24} features) for each principal outcome. Fig. S1 summarizes the model training and testing schema implemented. The implemented nested cross-validation contained an outer and an inner loop.

Class imbalance was handled by weighting the disproportionate classes to impose a heavier cost when errors were made in the minority class. The class imbalance of our eICU and MIMIC CA cohort per clinical outcome can be found in Table S1.

2.6. Model performance metrics

The sensitivity, specificity, and discrimination estimated by the area under the receiver operating characteristic curve (AUROC) of each model was computed across 25 outer validation loops (5 outer $\times$ 5 inner loop) and performances were compared between ML algorithms per clinical outcome. Additionally, precision, area under the precision recall curve (AUPRC), and F1 score were computed to better observe the effect of different clinical outcome prevalence. These performance metrics were evaluated after probability calibration to improve the distribution of predicted probabilities to better match the distribution of ground truth. Isotonic regression calibration was implemented to calibrate the predicted probabilities. Youden's index was used to compare performances of models using different feature subsets and ML classifiers. Youden's index (J statistic) is commonly used to express the performance of a binary classification model and is defined as the point on the ROC curve, which maximizes the sum of sensitivity and specificity minus feature ranking and Interpretation.

Feature ranking was performed to evaluate model interpretability. The motivation was to extract the most important features used in our final model and to provide insight into which HCTSA PTS_{24} derived features and/or potentially previously unknown EHR

features were most valuable for discriminative performance. We implemented a nested random forest feature importance ranking, which establishes hierarchical importance for each feature estimated by the frequency and placement (tree depth) of each feature in each decision tree [9,10]. This ranking was then normalized as a "relative importance" (RI) with a range of 0–1, with 1 being the most important feature.

To further enhance variable interpretation, the top 50 features from the nested random forest rankings were analyzed by their beta coefficients from each trained generalized linear model (GLM) per clinical outcome. This allowed for the analysis of the logistic regression beta coefficients to understand the positive and negative correlation to each of the clinical outcomes.

2.7. External validation

The resulting three eICU models were externally validated on 86 post-CA patients in MIMIC. This subset of patients was conservatively selected to contain those with an unquestionable certainty of an out of hospital cardiac arrest and resulted in a limited external validation sample size. This occurred due to missing diagnosis timing within the ICU, which left us only with admission diagnosis to rely on to determine out-of-hospital CA. From these 86 post-CA patients, different subsets had varying availability of clinical outcomes and variable availability, therefore, for each clinical outcome, there were fewer patients as shown in Table S1.

3. Results

We analyzed 2216 unique post-CA ICU admissions in the eICU database (Fig. 1) and 86 unique post-CA ICU admissions meeting the same criteria in MIMIC. The demographic summary and outcome distributions of eICU and MIMIC post-CA cohorts are provided in Table 1 and Table S1.

3.1. Model performance

Performance of all clinical outcome prediction models, evaluated for each feature subset (EHR-only, PTS_{24}-only, and EHR and PTS_{24} combined) is shown in Table 2, Fig. 2, and Fig. 3. This table also includes the discriminative performance of baseline logistic regression models for each outcome label created utilizing only variables used to compute the widely implemented APACHE IV in-hospital mortality. This is labeled as APACHE reference and serves as the true baseline comparison for only our equivalent CA survival outcome.

The AUROC of our in-hospital mortality (survival outcome) model outperformed the APACHE in-hospital mortality reference by 10% ($p < 0.01$), and provided significantly higher sensitivity, specificity, and precision. Given the mortality of 40.3% in our eICU cohort, compared to the APACHE reference, the positive (PPV) and negative predictive value (NPV) of our survival outcome model increased by 9.4% (0.66) and 8.0% (0.85) respectively (Table S5). It is important to note that all models were optimized for the AUROC at the Youden's index for ease of comparative evaluation, therefore, based on future implementation requirements, the PPV and NPV can be further optimized.

Across the three hospital discharge outcomes, the XGBoost model provided the highest discrimination with prediction of neurological outcome reaching an AUROC of 0.87 $\pm$ 0.1 and AUPRC of 0.86 $\pm$ 0.1. The PPV and NPV for the neurological outcome model shows significant implementation utility, highlighting that given the 52.8% prevalence of unfavorable neurological outcome, regardless of the predicted discharge neurological status, our model prediction was 80% correct (PPV: 0.80; NPV:0.79) (Table S5).

According to the DL outcome, 70.5% of the eICU CA cohort was found to have an unfavorable outcome. The EHR and PTS_{24} derived features showed promising discriminative performance and attained an AUROC of 0.80 $\pm$ 0.1 for the XGBoost model. The PPV of correctly identifying an unfavorable outcome was 0.88 with an NPV of 0.51. This indicates that the model, at the Youden's index, is optimized to be increasingly certain of patients predicted to have an unfavorable discharge location compared to patients predicted to be sent home or to acute rehabilitation.

3.2. Feature subset model performance

The ROC curves of each feature subset (EHR-only, PTS_{24}-only, and EHR + PTS_{24}) are shown with model performances in

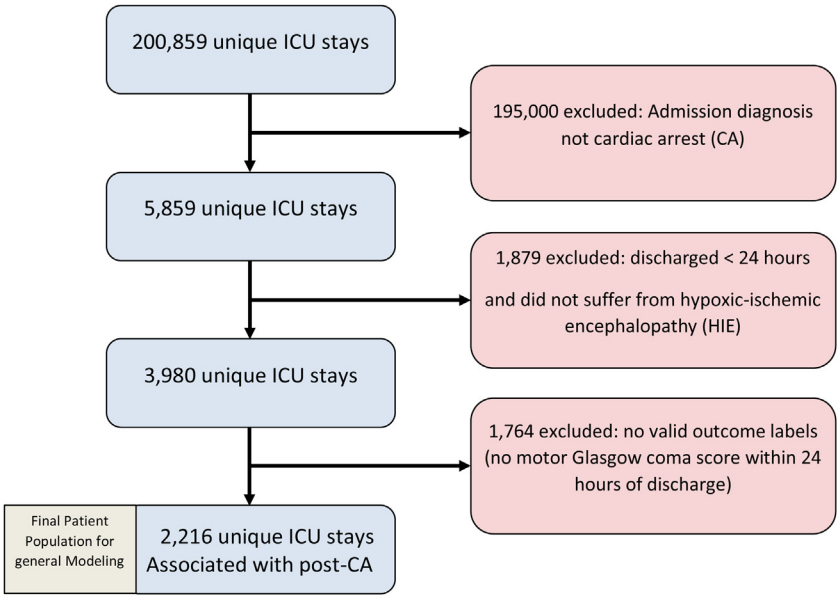

Fig. 1. Study flow diagram for eICU patients. MIMIC III patients were identified using the same criteria.

Table 1
Patient demographic summary.

	eICU-CRD	MIMIC III	*P*-value
n	2216	86	
Age (SD)	62.50 (15.86)	66.19 (14.65)	0.034
Body mass index (SD)	30.00 (8.30)	29.64 (4.06)	0.69
Ideal body weight (SD)	63.89 (10.94)	67.05 (6.37)	0.008
Gender male (%)	1280 (57.8)	53 (61.6)	0.548
Motor GCS on admission (IQR)	3 (4)	3 (2)	0.109
Total GCS on admission (IQR)	6 (6)	5 (6)	0.765
African American (%)	368 (16.6)	10 (11.6)	0.283
Caucasian (%)	1562 (70.5)	55 (64.0)	0.238
Other ethnicity (%)	286 (12.9)	21 (24.4)	0.004
Patients on ventilator (%)	1960 (88.4)	77 (89.5)	0.89
Patients with asystole (%)	180 (8.1)	Not available	
Patients with pulseless (%)	345 (15.6)	Not available	
Patients with ventricular fibrillation (%)	196 (8.8)	Not available	
Patients with ventricular tachycardia (%)	69 (3.1)	Not available	
Patients with unknown rhythm (%)	1426 (64.4)	Not available	
SOFA suspected sepsis (%)	680 (30.7)	11 (12.8)	0.001
SOFA septic shock (%)	328 (14.8)	7 (8.1)	0.118
SOFA score (SD)	6.63 (2.73)	5.29 (3.03)	< 0.001
qSOFA score (SD)	1.24 (0.66)	0.80 (0.72)	< 0.001
Neurological outcome (%)			< 0.001
Favorable	1170 (52.8)	22 (25.6)	
Unfavorable	1046 (47.2)	39 (45.3)	
Not available	0 (0.0)	25 (29.1)	
Survival (%)			0.548
Alive	1322 (59.7)	48 (55.8)	
Expired	894 (40.3)	38 (44.2)	
Discharge location outcome (%)			0.028
Favorable	646 (29.2)	36 (41.9)	
Unfavorable	1542 (69.6)	50 (58.1)	
Not available	28 (1.3)	0 (0.0)	

mGCS: motor Glasgow Coma Scale subscore; BMI: body mass index; DL: discharge location neurological; SOFA: sequential organ failure assessment; qSOFA: quick sequential organ failure assessment.

Table 2. The AUROC of the PTS_{24}-only model for all clinical outcomes was 4%–5% higher than the APACHE references utilizing validated clinical variables. Additionally, feature analysis identified which HCTSA PTS features contribute most for each of the post-CA clinical outcome prognostications and provides a ranked list for use in other studies (Table S3).

For all clinical outcomes, the best performing model combined EHR and PTS_{24} features. Comparing the GLM, RF, and XGBoost models across different feature subsets, the AUROC increase of the combined EHR and PTS_{24} features was 1%–2% higher than the AUROC of EHR-only model across all clinical outcomes.

3.3. Other performance metrics

The area under the precision-recall curve (AUPRC) of the best model for each outcome was 0.74 ± 0.1 for mortality, 0.86 ± 0.1 for mGCS, and 0.89 ± 0.1 for DL outcomes. The F1 score, a harmonic mean of precision (PPV) and recall (sensitivity), provides a class imbalance adjusted accuracy and shows that for both neurological outcomes with higher unfavorable outcome prevalence, the F1 score has been adjusted to be higher than the accuracy. The opposite is true for the survival outcome where the unfavorable outcome (expired) accounts for only 40.3% of the CA cohort. Lastly, we analyzed the Brier score of the isotonic regression calibrated predicted probabilities. While most models had good agreement between actual and predicted probabilities prior to isotonic calibration, the mortality calibration plot in Fig. 2 shows that the calibrated probabilities had improved agreement. Overall, the best performing models for all clinical outcomes had a Brier score ranging from 0.15 to 0.17. All these metrics are detailed in Table 2.

3.4. Model interpretability

Features were ranked utilizing the minimum depth of a nested random forest and the beta coefficients of a GLM. Fig. 4 shows the ranking results for both methods applied to survival outcome features. For clarity, in the random forest relative importance plot we used abbreviated feature designations (Fig. 4A); a dictionary can be found in Table S3. The top 50 features from the RF ranking were further analyzed using the GLM beta coefficients to better understand the relationship between feature value and its correlation and contribution to the binary outcomes (Fig. 4B). The ranking results for the neurological outcome labels are in Figs. S3 and S4. It was notable that TTM was not among the ranked predictive features.

For all clinical outcomes, the GCS total and subscores taken from the first day of ICU admission ranked as the most important features. The positive and negative beta coefficients for the clinical variables were verified as plausibly contributing to the favorable and unfavorable outcomes for all three outcomes. For survival outcome, of the top 50 features, 24 were HCTSA PTS_{24} derived features mainly originating from heart rate, SpO_2, and respiratory rate signals. Similarly, the top 50 features for the neurological outcome models were mostly HCTSA PTS derived features (33 for mGCS and 30 for DL). This strengthens our hypothesis that although there may be similar predictive performance between EHR and PTS_{24} models, the HCTSA PTS_{24} features make an independent contribution to the final model performance of the combined models.

These rankings also help identify which signals have more utility. While the top 50 features for survival and neurological outcome were lab results and heart rate derived features, the SpO_2

Table 2
Model performance summary for all models for each clinical outcome and feature subset.

		Model type	AUROC	Sensitivity	Specificity	AUPRC	Precision	F1 score	Brier score	Accuracy
Survival outcome	EHR + PTS	GLM	0.82 (0.82, 0.81)	0.80 (0.82, 0.79)	0.67 (0.69, 0.65)	0.72 (0.73, 0.70)	0.63 (0.64, 0.61)	0.70 (0.71, 0.69)	0.17 (0.18, 0.17)	0.72 (0.73, 0.72)
		RF	0.81 (0.82, 0.80)	0.79 (0.80, 0.77)	0.69 (0.71, 0.67)	0.70 (0.72, 0.69)	0.63 (0.65, 0.62)	0.70 (0.71, 0.69)	0.17 (0.18, 0.17)	0.73 (0.74, 0.72)
		*XGBoost	0.83 (0.84, 0.82)	0.79 (0.81, 0.77)	0.71 (0.73, 0.70)	0.74 (0.75, 0.73)	0.65 (0.67, 0.64)	0.71 (0.72, 0.70)	0.17 (0.17, 0.16)	0.74 (0.75, 0.74)
		NN	0.81 (0.82, 0.80)	0.64 (0.68, 0.59)	0.81 (0.84, 0.79)	0.63 (0.66, 0.60)	0.69 (0.71, 0.68)	0.66 (0.68, 0.63)	0.18 (0.18, 0.17)	0.74 (0.75, 0.73)
	EHR	GLM	0.81 (0.82, 0.81)	0.78 (0.80, 0.77)	0.69 (0.71, 0.68)	0.72 (0.73, 0.71)	0.63 (0.65, 0.62)	0.70 (0.71, 0.69)	0.17 (0.18, 0.17)	0.73 (0.74, 0.72)
		*RF	0.82 (0.82, 0.81)	0.75 (0.76, 0.73)	0.73 (0.74, 0.72)	0.71 (0.73, 0.70)	0.65 (0.66, 0.64)	0.70 (0.70, 0.69)	0.18 (0.18, 0.17)	0.74 (0.74, 0.73)
		XGBoost	0.81 (0.82, 0.80)	0.78 (0.80, 0.76)	0.70 (0.71, 0.68)	0.72 (0.74, 0.71)	0.64 (0.65, 0.63)	0.70 (0.71, 0.69)	0.17 (0.18, 0.17)	0.73 (0.74, 0.72)
	PTS	GLM	0.76 (0.77, 0.75)	0.69 (0.71, 0.67)	0.69 (0.71, 0.67)	0.65 (0.66, 0.64)	0.60 (0.62, 0.59)	0.64 (0.65, 0.63)	0.19 (0.20, 0.19)	0.69 (0.70, 0.68)
		RF	0.75 (0.76, 0.74)	0.74 (0.76, 0.72)	0.64 (0.65, 0.62)	0.64 (0.66, 0.62)	0.58 (0.59, 0.57)	0.65 (0.66, 0.64)	0.20 (0.20, 0.20)	0.68 (0.69, 0.67)
		*XGBoost	0.79 (0.80, 0.78)	0.74 (0.76, 0.72)	0.69 (0.71, 0.66)	0.68 (0.70, 0.67)	0.62 (0.63, 0.60)	0.67 (0.68, 0.66)	0.18 (0.19, 0.18)	0.71 (0.72, 0.70)
	APACHE reference		0.74 (0.75, 0.73)	0.72 (0.75, 0.70)	0.63 (0.65, 0.60)	0.62 (0.63, 0.61)	0.57 (0.58, 0.56)	0.64 (0.64, 0.63)	0.20 (0.21, 0.20)	0.67 (0.68, 0.66)
Neurological outcome (mGCS)	EHR + PTS	GLM	0.86 (0.86, 0.85)	0.81 (0.83, 0.78)	0.73 (0.75, 0.71)	0.85 (0.86, 0.84)	0.77 (0.78, 0.76)	0.79 (0.80, 0.78)	0.15 (0.16, 0.15)	0.77 (0.78, 0.76)
	L	RF	0.86 (0.86, 0.85)	0.82 (0.83, 0.80)	0.73 (0.74, 0.71)	0.84 (0.85, 0.84)	0.77 (0.78, 0.76)	0.79 (0.80, 0.79)	0.15 (0.16, 0.15)	0.77 (0.78, 0.77)
	L	*XGBoost	0.87 (0.88, 0.86)	0.81 (0.82, 0.79)	0.76 (0.78, 0.75)	0.86 (0.87, 0.85)	0.79 (0.80, 0.78)	0.80 (0.81, 0.80)	0.15 (0.15, 0.14)	0.79 (0.79, 0.78)
		NN	0.86 (0.86, 0.85)	0.80 (0.83, 0.78)	0.76 (0.78, 0.74)	0.68 (0.72, 0.63)	0.79 (0.80, 0.78)	0.80 (0.81, 0.79)	0.16 (0.16, 0.15)	0.78 (0.79, 0.78)
	EHR	GLM	0.85 (0.86, 0.85)	0.79 (0.80, 0.77)	0.77 (0.78, 0.75)	0.84 (0.85, 0.83)	0.79 (0.80, 0.78)	0.79 (0.80, 0.78)	0.15 (0.16, 0.15)	0.78 (0.78, 0.77)
		RF	0.86 (0.87, 0.85)	0.79 (0.80, 0.77)	0.77 (0.79, 0.75)	0.84 (0.85, 0.83)	0.80 (0.81, 0.78)	0.79 (0.80, 0.78)	0.15 (0.16, 0.15)	0.78 (0.79, 0.77)
		*XGBoost	0.86 (0.87, 0.85)	0.80 (0.81, 0.79)	0.76 (0.77, 0.74)	0.85 (0.87, 0.84)	0.79 (0.80, 0.78)	0.79 (0.80, 0.79)	0.15 (0.16, 0.15)	0.78 (0.79, 0.78)
	PTS	GLM	0.74 (0.75, 0.74)	0.70 (0.71, 0.68)	0.68 (0.70, 0.66)	0.74 (0.75, 0.73)	0.71 (0.72, 0.70)	0.70 (0.71, 0.69)	0.21 (0.21, 0.20)	0.69 (0.70, 0.68)
		RF	0.76 (0.77, 0.75)	0.71 (0.73, 0.69)	0.66 (0.68, 0.64)	0.76 (0.77, 0.75)	0.70 (0.71, 0.69)	0.71 (0.72, 0.70)	0.20 (0.21, 0.20)	0.69 (0.70, 0.68)
		*XGBoost	0.79 (0.80, 0.79)	0.73 (0.74, 0.71)	0.71 (0.73, 0.70)	0.79 (0.80, 0.78)	0.74 (0.75, 0.73)	0.73 (0.74, 0.72)	0.19 (0.19, 0.18)	0.72 (0.73, 0.71)
	APACHE reference		0.75 (0.76, 0.74)	0.76 (0.78, 0.75)	0.63 (0.64, 0.62)	0.72 (0.74, 0.71)	0.70 (0.70, 0.69)	0.73 (0.74, 0.72)	0.20 (0.20, 0.20)	0.70 (0.71, 0.69)
Discharge location outcome (DL)	EHR + PTS	GLM	0.78 (0.79, 0.78)	0.72 (0.74, 0.70)	0.71 (0.73, 0.69)	0.88 (0.89, 0.87)	0.86 (0.87, 0.85)	0.78 (0.79, 0.77)	0.16 (0.17, 0.16)	0.72 (0.73, 0.71)
		RF	0.77 (0.78, 0.76)	0.65 (0.67, 0.63)	0.76 (0.78, 0.74)	0.87 (0.88, 0.86)	0.87 (0.87, 0.86)	0.74 (0.75, 0.73)	0.17 (0.17, 0.16)	0.68 (0.69, 0.67)
		*XGBoost	0.80 (0.81, 0.79)	0.70 (0.73, 0.68)	0.75 (0.76, 0.73)	0.89 (0.90, 0.89)	0.87 (0.88, 0.86)	0.78 (0.79, 0.76)	0.16 (0.16, 0.16)	0.72 (0.73, 0.70)
		NN	0.77 (0.78, 0.76)	0.82 (0.85, 0.80)	0.54 (0.58, 0.50)	0.72 (0.75, 0.69)	0.81 (0.82, 0.79)	0.81 (0.82, 0.80)	0.18 (0.18, 0.17)	0.74 (0.75, 0.73)
	EHR	*GLM	0.78 (0.79, 0.77)	0.70 (0.72, 0.68)	0.72 (0.74, 0.69)	0.87 (0.88, 0.87)	0.86 (0.86, 0.85)	0.77 (0.78, 0.76)	0.17 (0.17, 0.16)	0.71 (0.72, 0.70)
		RF	0.77 (0.78, 0.76)	0.68 (0.70, 0.66)	0.72 (0.75, 0.69)	0.86 (0.87, 0.85)	0.85 (0.86, 0.84)	0.76 (0.77, 0.74)	0.17 (0.17, 0.17)	0.69 (0.71, 0.68)
		XGBoost	0.76 (0.77, 0.75)	0.69 (0.70, 0.67)	0.71 (0.73, 0.69)	0.86 (0.87, 0.85)	0.85 (0.86, 0.84)	0.76 (0.77, 0.75)	0.17 (0.17, 0.17)	0.69 (0.70, 0.68)
	PTS	GLM	0.74 (0.75, 0.73)	0.69 (0.70, 0.67)	0.68 (0.70, 0.66)	0.85 (0.86, 0.84)	0.84 (0.84, 0.83)	0.75 (0.76, 0.74)	0.18 (0.18, 0.17)	0.68 (0.69, 0.67)
		RF	0.73 (0.74, 0.72)	0.65 (0.67, 0.63)	0.69 (0.71, 0.67)	0.84 (0.86, 0.82)	0.83 (0.84, 0.83)	0.73 (0.74, 0.72)	0.18 (0.18, 0.18)	0.66 (0.67, 0.65)
		*XGBoost	0.76 (0.76, 0.75)	0.70 (0.72, 0.68)	0.67 (0.69, 0.65)	0.86 (0.87, 0.86)	0.84 (0.84, 0.83)	0.76 (0.77, 0.75)	0.17 (0.17, 0.17)	0.69 (0.70, 0.68)
	APACHE reference		0.71 (0.72, 0.70)	0.70 (0.73, 0.67)	0.60 (0.63, 0.57)	0.83 (0.83, 0.82)	0.81 (0.82, 0.80)	0.75 (0.76, 0.73)	0.18 (0.19, 0.18)	0.67 (0.68, 0.66)

AUROC: area under the receiver operating characteristic curve; AUPRC: area under the precision recall curve; RF: random forest; XGBoost: extreme gradient boosting; mGCS: motor Glasgow Coma Score; DL: discharge location; PTS: physiologic time series; EHR: electronic health record; Reference APACHE (Acute Physiology and Chronic Health Evaluation) refers to a logistic regression model created utilizing variables used to compute the APACHE IV score to predict our three clinical outcomes**. APACHE reference is only used as an approximate reference/baseline point.** The only direct APACHE reference comparison would be for post-CA survival (in-hospital mortality) outcome.
* The best performing models in each feature subset according to AUROC are marked with an asterisk.

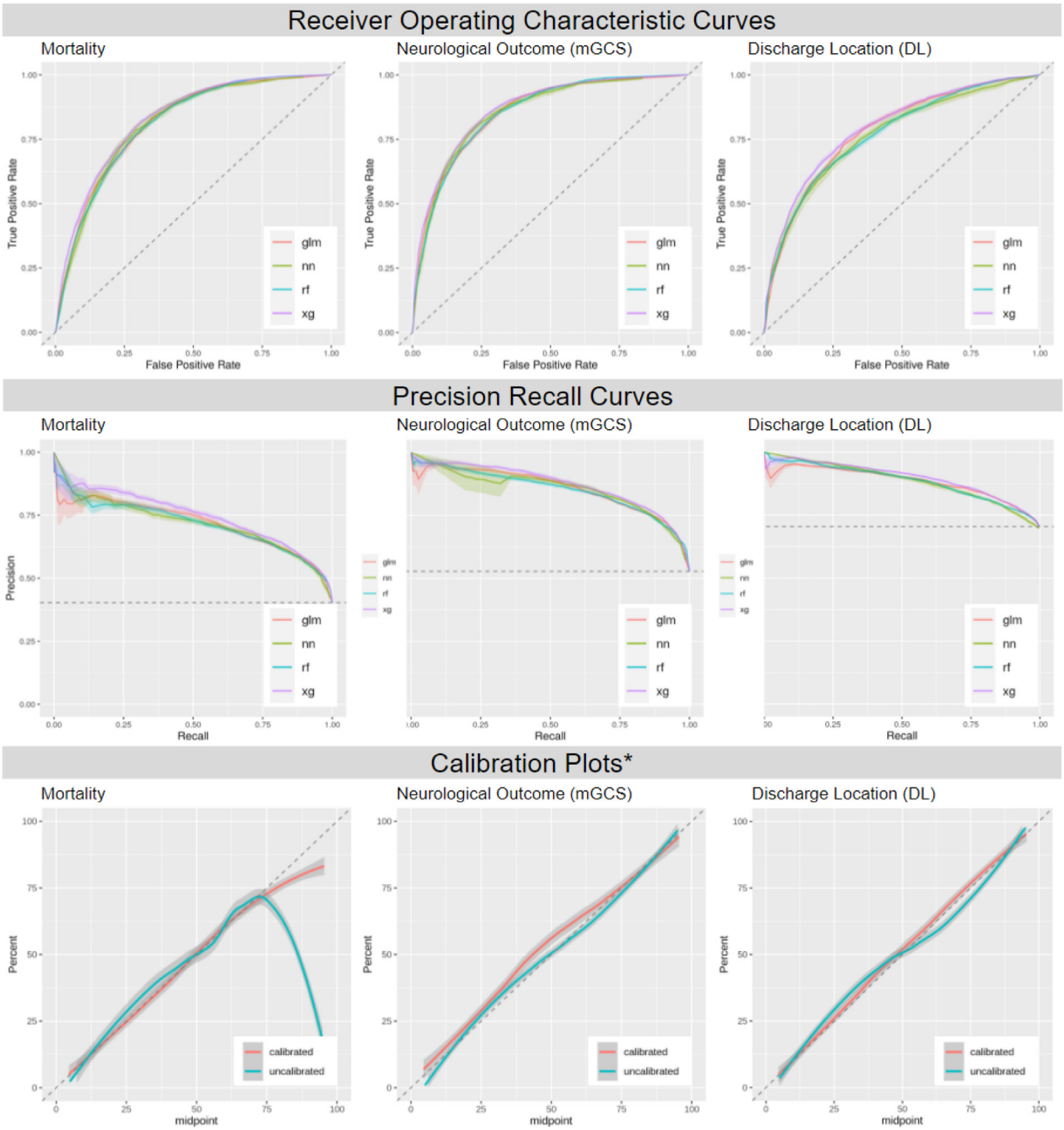

Fig. 2. Performance of computational models for all three labels.
*The following calibration plots are visualized only for the best performing model. AUROC: Area under the receiver operating curve; AUPRC: Area under the precision recall curve. GLM: generalized linear model. NN, neural network. Rf, random forest. Xg, gradient boost. mGCS, motor subscore of the Glasgow Coma Scale. The dotted line in the AUPRC plots represent the class imbalance.

and respiratory rate derived features had the largest impact on the DL outcome. Additionally, the DL outcome was associated with different lab results and SOFA component features compared to the top 50 features of the other two outcomes.

The HCTSA PTS derived categories can be found in Table S2. In agreement with the frequency count of each category of HCTSA variables selected for each clinical outcome, many of the top PTS features described correlation, time series model fitting, symbolic transformation, information theory, and wavelet coefficient features. A detailed breakdown of each HCTSA variable broken down by category and feature number can be found in Table S3.

3.5. External validation

Based on our inclusion and exclusion criteria, we found 86 matching MIMIC III post-CA ICU admission with available clinical outcomes of interest and physiologic time series data (Table 1). MIMIC III validation indicated reduced performance when compared to eICU: The AUROC of the best MIMIC III models was reduced by 7% for survival outcome, 2% for neurological outcome, and 4% drop for DL outcome. A comprehensive overview of external validation results is in Table S4 and Fig. S6.

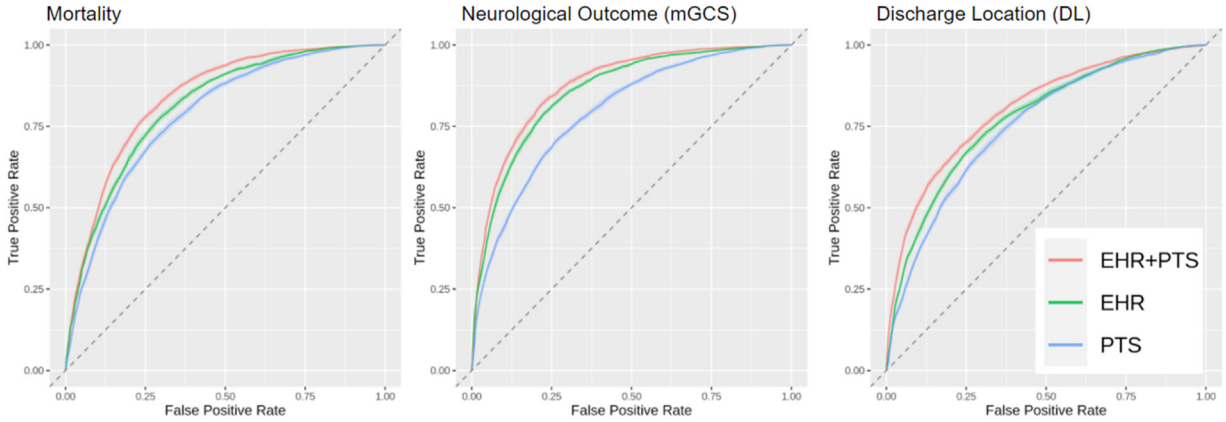

Fig. 3. Receiver operating characteristic curves (ROC) of EHR-only *vs.* PTS_{24}-only *vs.* EHR & PTS_{24} feature subset models. ROCs of the best performing models are plotted for each feature subset. Algorithms with an asterisk in Table 2 identifies which models are visualized. mGCS, motor subscore of the Glasgow Coma Scale. EHR, model established with electronic health record data. PTS, model established with physiological time series data. EHR + PTS, model combining features from EHR and PTS.

4. Discussion

Results demonstrate the value of a computational modeling approach to predict short-term post-CA clinical trajectories in the ICU setting. These models were established exclusively with data available in the first 24 h after ICU admission and document the importance of PTS as predictive variables. The findings show that time-series feature engineering is a promising method to decode clinically meaningful information from high-frequency physiological data.

Feature exploration demonstrated that time series physiological signals contain important predictive information. The majority of the top 50 predictive features across all clinical outcomes were PTS_{24} derived features. Our analysis makes clear that interpretable statistical and mathematical features can be extracted from clinical time series data, distinguishing this approach from "black box" neural network models.

Our analyses were performed using data, which are routinely acquired in ICUs, suggesting the method could be scaled for broader validation and use. The results suggest a pragmatic and efficient computational approach to post-CA outcome prediction that might complement existing prognostication systems. Additionally, modeling was based exclusively on data collected in the first 24 h after ICU admission, demonstrating that early physiological signatures are linked to outcome and need further investigation. These early indicators could, if further validated, represent a different paradigm from the current approach of delaying prognostication until at least 72 h after CA [3].

Since the data for our CA patients was collected in multiple institutions, we expected that the developed models might be generalizable to a broader population of post-CA patients admitted to hospitals in the USA. Notwithstanding clinical guidelines, there is considerable variability between institutions (and even between ICUs within the same institutions) in the acute care of patients resuscitated from CA [11]. Our external validation in MIMIC III found a decrease in model discrimination of 3%–8%. The reduction in discrimination is not unexpected [12,13].

An important goal will be to validate our results on other post-CA populations, in particular in prospective cohorts that would allow the efficacy, generalizability, and practicability of this approach to be tested in a real-world and real-time setting. Additionally, we plan to validate our models in post-CA populations outside North America to determine model fit and performance within epidemiologically disparate CA populations. This study warrants further exploration of the predictor variables, to gain insights on the clinical correlation and interpretability provided by those variables. Furthermore, it will be important to examine the value of computational approaches in integrating prognostic data from different modalities, including neurophysiology and brain MRI [29]. We plan to assess the impact of these multimodal biomarkers in future studies. Last, work is needed to understand the relevance of such models to long-term outcomes, and to determine if comparable predictions are possible even earlier in the ICU stay.

Several limitations in this work need to be noted. This was a retrospective analysis suggesting potential errors due to bias, confounders, and unrecorded or missing data. Our analysis was centered on a subset of patients who remained in the ICU for $>$ 24 h and for whom adequate data were available; these inclusion criteria almost certainly introduced bias, since patients in very unstable condition and/or dying in the first 24 h would have been excluded, and documentation of PTS and/or outcomes may have been less complete in this group. As a result, survival observed in this sample was higher than in contemporary cohorts [30]. No information was available about many variables, which are commonly used for post-CA prognostication including bystander resuscitation, time to return of spontaneous circulation, end-tidal CO2, and results of tests such as EEG, evoked potentials or imaging, among others. It is possible that inclusion of such variables would have enhanced the predictive value of our EHR model.

In addition, documentation was lacking on the use of TTM, an important treatment variable, which can significantly influence functional outcome in CA patients, or on withdrawal of life sustaining therapy (WLST), which may occur in a significant proportion of CA patients and can signal self-fulfilling prophecies [14,31]. Another major limitation is that validated outcome measures such as the cerebral performance category (CPC) score or cognitive tests were not available in the eICU database. We devised a surrogate outcome based on mGCS and discharge location, however neither of these captures the functional state information that is encompassed in scores like the CPC [15–17]. Finally, it should be recognized that the performance of our models, while rivaling those of scores derived at later time points ($>$ 72 h), will need to be significantly increased in order for them to be considered in clinical decision-making.

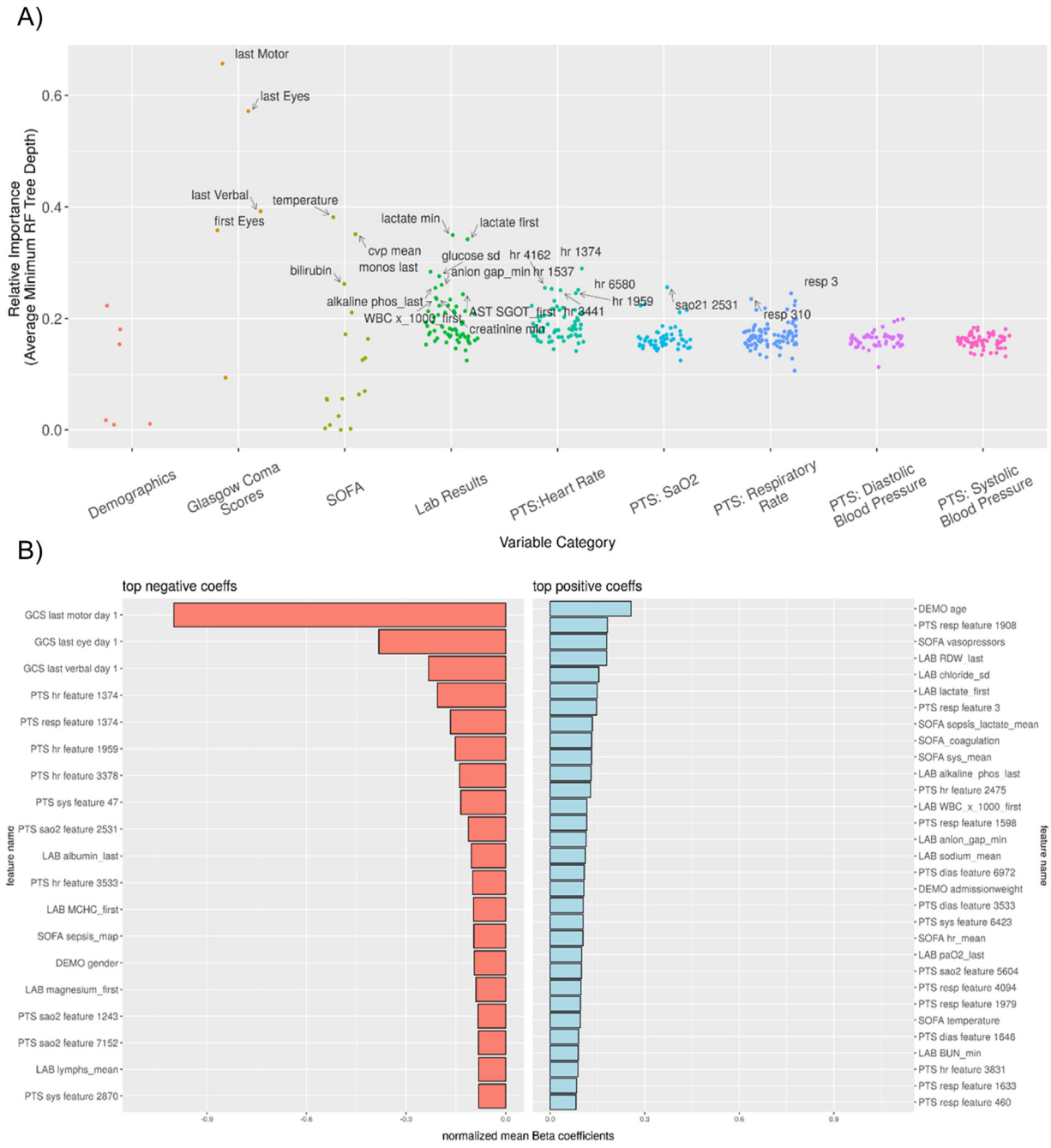

Fig. 4. Feature importance visualized for the 410 survival outcome features using A) Random Forest relative importance visualized using feature categories with the top 25 features labeled, and B) Normalized GLM beta coefficients of the top 50 features. Both figures show simplified naming conventions for the HCTSA PTS derived features (full dictionary available in the supplements). Similar plots for mGCS and DL outcomes are available in Figs. S3 and S4.

5. Conclusions

Taken together, these findings demonstrate that computational models trained with high-resolution ICU time series data can successfully discriminate discharge neurological outcome, discharge location, and survival of patients resuscitated from CA. We found that physiological signals contain valuable prognostic information and that features derived from the first 24 h of ICU admission are associated with early post-CA recovery trajectories. Our models are interpretable and indicate a number of predictive features, which warrant exploration in future studies. Early and accurate characterization of post-CA severity and clinical trajectories could in the future provide a window for personalizing therapeutic interventions with the goal of achieving better outcomes.

Declarations

None.

Funding

None.

Conflicts of interest/competing interests

The authors have no conflicts of interest to declare.

Availability of data and material

The eICU and MIMIC databases analyzed in this study are publicly available from PhysioNet (https://physionet.org/).

Code availability

All code used to extract, process, and analyze data will be made available on Github with the publication and will provide tools to reproduce the article results. External R libraries used for data extraction, processing, or analysis are available via CRAN (https://cran.r-project.org/). Deep learning analysis was accomplished using Pytorch (https://pytorch.org/). The HCTSA matlab feature extraction tool is available (https://github.com/benfulcher/hctsa) and requires Matlab versions R2018b or later with access to Statistics, Signal Processing, Curve Fitting, System Identification, Wavelet, and Econometrics Matlab toolboxes.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

Conception or design of the work: RDS, RLW, CS, JIS. Data Collection: HBK, HTN, QJ, ST, TGR. Data analysis and interpretation: all authors. Drafting the article: HBK, HTN, QJ, ST, TGR, ES. Critical revision of the article: HBK, HTN, QJ, RL, JLG, JIS, CS, RLW, RDS. Final approval of the version to be published: all authors.

Ethical statement

Research in this report was carried out on fully deidentified publicly available datasets made available via the Massachusetts Institute of Technology (MIT) PhysioNet repository (https://physionet.org/). Data in the MIMIC-III database have been deidentified, and the institutional review boards of MIT (number 0403000206) and Beth Israel Deaconess Medical Center (number 2001-P-001699/14) both approved the use of the database for research. Because the database does not contain protected health information, a waiver of the requirement for informed consent was included in the IRB approval. Data in eICU are also deidentified, and research using eICU data is exempt from institutional review board (IRB) approval due to the retrospective design, lack of direct patient intervention, and the security schema, for which the re-identification risk was certified as meeting safe harbor standards by an independent privacy expert (Privacert, Cambridge, MA, USA; Health Insurance Portability and Accountability Act Certification number 1031219–2).

Appendices

PTS preprocessing

Given medical data irregularities, specifically PTS data, preprocessing steps were needed to standardize and impute the data. Our preprocessing can be separated into two steps: (1) clinically implausible data (outlier) detection and (2) missing value imputation.

This process is illustrated in Fig. S5. We reasoned that outlier could be categorized as either an anomaly (artifact) or an accurate reading (real clinical events). To identify outliers, the sliding window median and median absolute deviation for PTS data were calculated. Within the window, any single point or continuous interval of points that falls outside of 3 absolute deviations from the sliding window median are considered potential outliers. Next, two rejecting bounds (lower and upper bounds deemed clinically implausible by physicians) are set and can be found in Table S6. For each potential outlier interval, the entire interval is removed if at least one point meets the above outlier criteria. The rationale behind these criteria is that outliers in the same interval are triggered by either a clinical event or a machine malfunction event. In any given temporal interval, if it was shown that one outlier was an artifact, then all other outliers in this outlier interval were considered an artifact, and the whole interval was removed.

For PTS data imputation, the first step was to capture data in the nurse charting records. In many ICUs, nurses manually record various features such as heart rate, blood pressure, and temperature. These data are then archived in the EHR. Inconsistencies can sometimes be observed between manually recorded data and PTS data; this may be due to offsets between the time vital signs are recorded and the physiological occurrence. However, when EHR and PTS data are strongly correlated, EHR data is a valuable resource for imputation of missing PTS data. We binned nurse charting data into 5 min intervals, mirroring the format of the PTS data. Next, the Pearson correlation for all overlapping time points between the EHR and PTS data was calculated. EHR data were used for imputation if the degree of correlation was larger than 0.8 for more than 15 common time points. The remaining missing data were imputed by linear interpolation.

Variable extraction and selection

Based on prior studies, quantitative variables in the EHR relevant to post-CA patients were selected. EHR features were composed of 338 variables extracted from demographics, nurse examinations, laboratory results, medication, initial rhythm, admission diagnosis, and SOFA score/components, all limited to the first day of ICU admission. Highly comparative time-series analysis (HTCSA), an automated feature extraction tool, was used to derive approximately 4000 derived features per PTS_{24} signal (heart rate, SpO2, respiratory rate, diastolic blood pressure, and systolic blood pressure). HCTSA PTS_{24} derived features included distribution, correlation, trends, frequency, information theory features, and many others. A simplified list of features by variable category utilized for subsequent modeling can be found in Table S2. HCTSA has been successfully used across multiple fields of study, and therefore was selected to identify signal derived features that be applicable in this use case [7,8].

The PTS_{24} (19,691 variables) and EHR (338 variables) features were pruned in two steps to ensure selected features were tailored for each clinical outcome. First, PTS_{24} signal derived features were down selected based on a nested random forest feature importance ranking. The average number of features needed to reach the maximal AUROC across 10 cross validation folds were selected per signal [18]. Then variance inflation factor (VIF) analysis was used to evaluate the collinearity between the combined feature space of EHR and derived features from each PTS_{24} signal to further reduce the number of features and ensure minimal cross correlation [19]. This resulted in three distinct feature spaces for each clinical outcome reducing the 20,029 features to 410, 508, and 436 features for survival outcome, mGCS neurologic outcome, and DL outcome respectively. A breakdown of HCTSA feature categories included in our model can be found in Table S2 and Table S3.

Missing data

EHR variables for which > 40% of data were missing were excluded from the analysis. For the remaining variables, a random forest unsupervised imputation was used to fill missing values. This multiple imputation method takes the non-linearity and interaction among variables into account and has been shown to provide robust imputations [20]. This multiple imputation method takes the non-linearity and interaction among variables into account and has been shown to provide robust imputations. Five imputation iterations were performed each creating 50,000 unsupervised decision trees using a subset of both the samples and features. The resulting imputation is averaged initially within the 50,000 decision trees then averaged over the five iterations to provide a robust missing data imputation of EHR variables.

PTS_{24} signals with missing data were imputed using two different criteria after the removal of implausible data points based on clinician-defined cut-offs. First, any missing signal gap less than 1 hour was linearly interpolated. Then, as the HCTSA tool required all 5-minute time points to have a corresponding value, the remaining missing values were filled in by carrying forward the previous value.

Model development

All models were developed exclusively using features extracted from the first 24 h following ICU admission. Models were trained using different feature subsets (EHR features, PTS_{24} features, and combined EHR and PTS_{24} features) following feature selection, to predict each of the three principal outcomes. Four ML approaches were used to train the prediction models: generalized linear model (GLM), random forest (RF), gradient boost (XGBoost) and neural network (NN). The NN architecture contained fully connected (FC) layers for static EHR features and recurrent layers (RNN) for the five raw PTS_{24} signals (Fig. S2). Fig. S1 summarizes the model training and testing schema implemented. The implemented nested cross-validation contained an outer and an inner loop. The outer loop resampled the 80% training and 20% testing five times, ensuring all samples were included in the test set once while keeping each outer loop's testing and training samples independent. The inner loop refers to the traditional k-fold cross-validation loop and was repeated three times per outer loop; and was used to evaluate the training performance and hyper parameter tune each ML approach. The five outer loops enabled the estimation of generalized model performances across more combinations of training and testing data compared to the traditional single training and single testing set. Therefore, the nested cross-validation approach reduced the risk of overestimating the final model performance due to chance.

Class imbalance was handled by weighting the disproportionate classes to impose a heavier cost when errors were made in the minority class. Weights were objectively determined by imposing the proportion of class 1 as weights to class-2 samples and the proportion of class 2 as weights to class-1 samples. Class weighting, therefore, provided a significant benefit to reduce over fitting to the majority class. The class imbalance of our eICU and MIMIC CA cohort per clinical outcome can be found in Table S1.

Supervised learning pipeline

Fig. S1 illustrates training and evaluating of our models using a nested cross-validation method containing two cross-validation loops. The inner loop is for hyper parameter tuning, while the outer, 5-fold × 5 times loop is used to estimate the generalized performance and to compare performances across different models. We used this nested cross-validation method instead of the traditional non-nested cross-validation method to avoid overestimation of the true performance. With the traditional k-fold cross-validation, the same data are often used both for hyper parameter tuning and for estimating the generalization error, which can lead to over fitting. Studies have demonstrated that the non-nested, cross-validated error estimate for the classifier with the optimal parameters is a substantially biased estimate of the true error that the classifier would incur on another, independent dataset [21–23]. The nested cross-validation resampling strategy has been shown to be an unbiased estimator of the true error [22,24].

Model optimization

For each model, hyper parameters were tuned in the inner loop of nested cross-validation (10-fold × 3 times, see Fig. S1). Hyper parameters for models except NN were tuned using the grid-search method with default hyper parameter space in the "caret" package. In addition, for the best performing first-level models (such as XGBoost, GLM-elastic net, RF, and NN), Bayesian (model-based) optimization was implemented using the "mlrMBO" package in R [25,26].

Neural network architecture

The neural network consists of two structures: fully connected (FC) network for the static data and recurrent neural network (RNN) for the dynamic data, where, for each patient, static data are constant and dynamic data varies over time. Fig. S2 shows the neural network pipeline for the input combing EHR and PTS_{24}. First, EHR data are separated into categorical data and continuous data. Continuous EHR and HCTSA package derived PTS features are normalized by the batch norm layer and then pass with the categorical EHR data through two fully connected layers. A gated recurrent unit (GRU) with the attention layer is implemented for the dynamic data. The attention layer has been widely used on medical data such as healthcare image data for lesion detection and PTS data [27,28]. The outputs from FC network and RNN network then pass through a fully connected layer to obtain a final prediction. The neural net schema described above was implemented in "Python 3.7" with "pytorch 0.4.1" package.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.accpm.2021.101015.

References

[1] Cariou A, Payen J-F, Asehnoune K, Audibert G, Botte A, Brissaud O, et al. Targeted temperature management in the ICU: guidelines from a French expert panel. Anaesth Crit Care Pain Med 2018;37:481–91.

[2] Sandroni C, D'Arrigo S, Nolan JP. Prognostication after cardiac arrest. Crit Care 2018;22:150.

[3] Callaway CW, Donnino MW, Fink EL, Geocadin RG, Golan E, Kern KB, et al. Part 8: Post-Cardiac Arrest Care: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015;132:S465–82.

[4] Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018;5180178.

[5] Johnson AE, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3160035.

[6] Balouris SA, Raina KD, Rittenberger JC, Callaway CW, Roger JC, Holm MB. Development and validation of the cerebral performance categories-extended (CPC-E). Resuscitation 2015;94:98–105. http://dx.doi.org/10.1016/j.resuscitation.2015.05.013.

[7] Fulcher BD, Jones NS. hctsa: a computational framework for automated time-series phenotyping using massive feature extraction. Cell Syst 2017;5:527–531.e3. http://dx.doi.org/10.1016/j.cels.2017.10.001.

[8] Fulcher B, Little M, Jones N. Highly comparative time-series analysis: the empirical structure of time series and their methods. J R Soc Interface 2013;1020130048. http://dx.doi.org/10.1098/rsif.2013.0048.

[9] Breiman L. Random forests. Mach Learn 2001;45:5–32. http://dx.doi.org/10.1023/A:1010933404324.

[10] Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. Pattern Recognit Lett 2010;31:2225–36. http://dx.doi.org/10.1016/j.patrec.2010.03.014.

[11] Balian S, Buckler DG, Blewer AL, Bhardwaj A, Abella BS, CARES Surveillance Group.. Variability in survival and post-cardiac arrest care following successful

resuscitation from out-of-hospital cardiac arrest. Resuscitation 2019;137:78–86. http://dx.doi.org/10.1016/j.resuscitation.2019.02.004.
[12] Steyerberg EW, Mushkhudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008;5e165. http://dx.doi.org/10.1371/journal.pmed.0050165. discussion e165.
[13] Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, et al. Prediction of outcome after moderate and severe traumatic brain injury: external validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation after Significant Head injury (CRASH) prognostic models. Crit Care Med 2012;40:1609–17. http://dx.doi.org/10.1097/CCM.0b013e31824519ce.
[14] May TL, Ruthazer R, Riker RR, Friberg H, Patel N, Soreide E, et al. Early withdrawal of life support after resuscitation from cardiac arrest is common and may result in additional deaths. Resuscitation 2019;139:308–13. http://dx.doi.org/10.1016/j.resuscitation.2019.02.031.
[15] Schefold JC, Storm C, Kruger A, Ploner CJ, Hasper D. The Glasgow Coma Score is a predictor of good outcome in cardiac arrest patients treated with therapeutic hypothermia. Resuscitation 2009;80:658–61. http://dx.doi.org/10.1016/j.resuscitation.2009.03.006.
[16] Dragancea I, Horn J, Kuiper M, Friberg H, Ullén S, Wetterslev J, et al. Neurological prognostication after cardiac arrest and targeted temperature management 33°C versus 36°C: results from a randomised controlled clinical trial. Resuscitation 2015;93:164–70. http://dx.doi.org/10.1016/j.resuscitation.2015.04.013.
[17] Roger C, Palmier C, Louart B, Molinari N, Claret P-G, de la Coussaye J-E, et al. Neuron specific enolase and Glasgow motor score remain useful tools for assessing neurological prognosis after out-of-hospital cardiac arrest treated with therapeutic hypothermia. Anaesth Crit Care Pain Med 2015;34:231–7. http://dx.doi.org/10.1016/j.accpm.2015.05.004.
[18] Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ Res 2017;121:1092–101. http://dx.doi.org/10.1161/CIRCRESAHA.117.311312.
[19] Miles J. Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd;; 2014. http://dx.doi.org/10.1002/9781118445112.stat06593.
[20] Tang F, Ishwaran H. Random forest missing data algorithms. Stat Anal Data Min: the ASA Data Sci J 2017;10:363–77. http://dx.doi.org/10.1002/sam.11348.
[21] Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics 2005;21:3301–7. http://dx.doi.org/10.1093/bioinformatics/bti499.
[22] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinf 2006;7:91. http://dx.doi.org/10.1186/1471-2105-7-91.
[23] Kim J-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal 2009;53:3735–45.
[24] Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. ArXiv 2018.
[25] Bischl B., Richter J., Bossek J., Horn D., Thomas J., Lang M. mlrMBO: a modular framework for model-based optimization of expensive black-box functions. ArXiv:170303373 [Stat] 2018.
[26] Snoek J., Larochelle H., Adams R.P. Practical bayesian optimization of machine learning algorithms. ArXiv:12062944 [Cs, Stat] 2012.
[27] Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019;3:173–82. http://dx.doi.org/10.1038/s41551-018-0324-9.
[28] Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bilhorac A, Rashidi P. DeepSOFA: a continuous acuity score for critically ill patients using clinically interpretable deep learning. Sci Rep 2019;9:1879. http://dx.doi.org/10.1038/s41598-019-38491-0.
[29] Quintard H, Velly L, Boussen S, Chiosi X, Amoretti M-E, Cervantes E, et al. Value of assessment of multivoxel proton chemical shift imaging to predict long term outcome in patients after out-of-hospital cardiac arrest: a preliminary prospective observational study. Resuscitation 2019;134:136–44. http://dx.doi.org/10.1016/j.resuscitation.2018.09.007.
[30] Gräsner JT, Herlitz J, Tjelmand IBM, Wnent J, Masterson S, Lilja G, et al. European resuscitation council guidelines 2021: epidemiology of cardiac arrest in Europe. Resuscitation 2021;161:61–79. http://dx.doi.org/10.1016/j.resuscitation.2021.02.007.
[31] Nielsen N, Wetterslev J, Cronberg T, Erlinge D, Gasche Y, Hassager C, et al. Targeted temperature management at 33°C versus 36°C after cardiac arrest. N Engl J Med 2013;369:2197–206. http://dx.doi.org/10.1056/NEJMoa1310519.