Development and Validation of a Physiology-Driven Machine Learning Model for Post-Cardiac Arrest Outcome Prediction Using a Large Multi-Center Database

Han Kim MSE[1,2,3], Hieu Nguyen[1], Qingchu Jin[1], Jose Suarez MD[2], Christian Storm MD[2], Raimond Winslow PhD[1], Robert D. Stevens MD[2,3]

[1] Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA, [2] Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [3] Laboratory of Computational Intensive Care Medicine, Johns Hopkins University, Baltimore, MD, USA

Introduction

Cardiac arrest (CA) is an abrupt cessation of myocardial function which affects more than half a million people in the United States annually. An estimated 80% of patients are unconscious after resuscitation from CA, and these patients can experience a wide range of outcomes, from complete recovery to death or severe neurologic disability.[1] A major 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 in the early phase of care.[2] The recommended paradigm of multi-modality prognostication to be implemented >72 hours after CA can be a challenge to implement, and the predictive performance of its different elements, while studied individually, are unknown in aggregate.[3]

Objectives/Aims

Here, we propose a novel approach for post-CA clinical outcome prediction, based on two hypotheses. **First**, that variables derived from early (first 24h in ICU) physiologic time series (PTS) signals widely available at the bedside contain discriminative information and contribute to prognostic model performance; and **second,** that a combination of electronic health record (EHR) clinical and derived PTS variables will yield the best short-term prognostic capabilities for post-CA data-driven prognostication. Three short-term hospital discharge endpoints were evaluated and modeled: 1. in-hospital mortality, 2. neurological outcome based on dichotomized motor Glasgow Coma score, and 3. post-hospital discharge location. Hypotheses were tested using data from the multicenter Philips eICU-CRD database,[4] and externally validated on Medical Information Mart for Intensive Care (MIMIC) III database.[5]

Methods

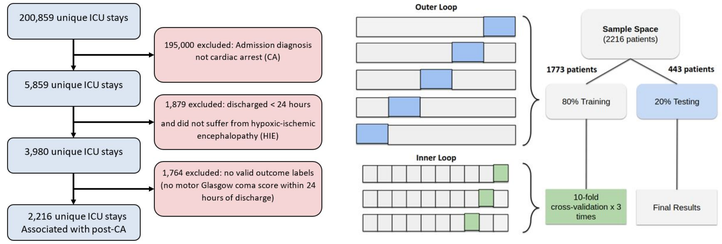

Figure 1. Study inclusion and exclusion criteria identifying the 2,216 eICU post-CA cohort. 86 MIMIC III post-CA cohort were identified following the same criteria.

Figure 2. Supervised machine learning pipeline showing the nested cross validation implementation in which the inner loop is used for hyperparameter tuning and the outer loop is used to evaluate the generalized model performance.

Machine learning models were created to predict hospital discharge endpoints within 24 hours of ICU admission. We evaluated four different statistical and machine learning methods: 1. generalized linear model (glm), random forest (rf), XGBoost (xg), and neural networks (nn). EHR and PTS derived features were evaluated separately and together to assess the added predictive capabilities of PTS derived features. The principal outcomes were survival and neurological function recorded at the time near discharge from the hospital *(longer term outcomes are not available in eICU)*. The neurological outcome indicator widely used in the CA population is the Cerebral Performance Category (CPC) score,[6] however, CPC was not recorded in eICU or MIMIC III. Therefore, to best account for this limitation, we defined two surrogate outcomes, one based on the motor subscore of the Glasgow Coma Score (mGCS) at discharge, dichotomized as follows: mGCS of 6 (favorable outcome), mGCS ≤5 (unfavorable outcome) and the second based on hospital discharge location (DL), dichotomized as follows: discharge location of home and rehabilitation (favorable outcome), other location (unfavorable outcome).

Results

	eICU-CRD	MIMIC III	P-value
n	2,216	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 (%)	1,280 (57.8)	53 (61.6)	0.548
Motor GCS on Admission (SD)	3.19 (1.98)	2.84 (1.50)	0.109
Total GCS on Admission (SD)	6.55 (3.54)	6.67 (4.14)	0.765
African American (%)	368 (16.6)	10 (11.6)	0.283
Caucasian (%)	1,562 (70.5)	55 (64.0)	0.238
Other Ethnicity (%)	286 (12.9)	21 (24.4)	0.004
Patients on Ventilator (%)	1,960 (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 (%)	1,426 (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	1,170 (52.8)	22 (25.6)	
Unfavorable	1,046 (47.2)	39 (45.3)	
Not Available	0 (0.0)	25 (29.1)	
Survival (%)			0.548
Alive	1,322 (59.7)	48 (55.8)	
Expired	894 (40.3)	38 (44.2)	
Discharge Location Dichotomized (%)			0.028
Favorable	646 (29.2)	36 (41.9)	
Unfavorable	1,542 (69.6)	50 (58.1)	
Not Available	28 (1.3)	0 (0.0)	

Table 1. Demographic Summary of the Post-Cardiac arrest population in the eICU-CRD and MIMIC-III database.

Table 1 provides the demographic summary of the eICU and MIMIC-III post-CA cohorts. As our model required stringent data availability to assess the efficacy of incorporating high frequency time series physiologic signal (PTS) features, our external validation post-CA population was reduced to 86 ICU admissions. Figures 3 demonstrates that the incorporation of PTS derived features increased predictive performance and the PTS derived features alone may provide significant information that have previously been overlooked, both in other models and in the clinical setting.

Figure 4 shows model performances across different clinical endpoints for each of our four evaluated statistical and machine learning algorithms. Table 2 shows the MIMIC-III external validation results on our best performing eICU post-CA prognostication models. The degree of external validation performance reduction varied between 3 - 7% AUROC and suggested overall good generalizability.

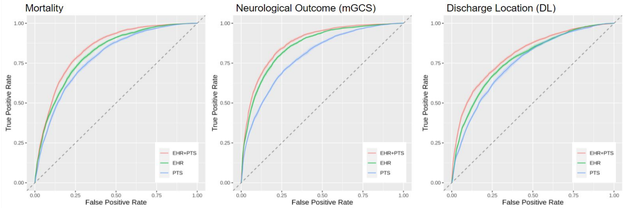

Figure 3. ROC curves for each clinical endpoint [Mortality (left), Neurological outcome (middle) and Discharge location (right)] evaluating the model performance differences between three different feature spaces: 1. EHR and PTS derived features, 2. EHR derived features only, and 3. Our findings show that not only do PTS derived features alone provide almost as much information as EHR derived features, but also there is an additive performance improvement when PTS and EHR derived features are combined.

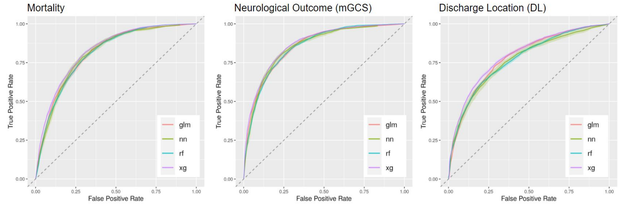

Figure 4. ROC curves of glm, nn, rf, and xg boost models utilizing both EHR and PTS derived features to predict Mortality (left), Neurological outcome (middle) and Discharge location (right).

		AUROC	Sensitivity	Specificity
In-hospital Survival Outcome	eICU model development	0.83 (0.84, 0.82)	0.79 (0.81, 0.77)	0.71 (0.73, 0.70)
	MIMIC III External Validation	0.76 (0.77, 0.75)	0.79 (0.82, 0.76)	0.59 (0.61, 0.56)
Neurological Outcome (motor GCS surrogate)	eICU model development	0.87 (0.88, 0.86)	0.81 (0.82, 0.79)	0.76 (0.78, 0.75)
	MIMIC III External Validation	0.84 (0.85, 0.83)	0.92 (0.93, 0.90)	0.58 (0.60, 0.56)
Dichotomized Discharge Location	eICU model development	0.80 (0.81, 0.79)	0.70 (0.73, 0.68)	0.75 (0.76, 0.73)
	MIMIC III External Validation	0.76 (0.76, 0.75)	0.75 (0.77, 0.72)	0.59 (0.61, 0.56)

Table 2. Performance metric summary of the eICU-CRD developed model and MIMIC III external validation for each evaluated clinical endpoint. Results show a loss of performance during external validation but that may be due to the limited samples that met our inclusion and exclusion criteria (86 samples).

Results (cont.)

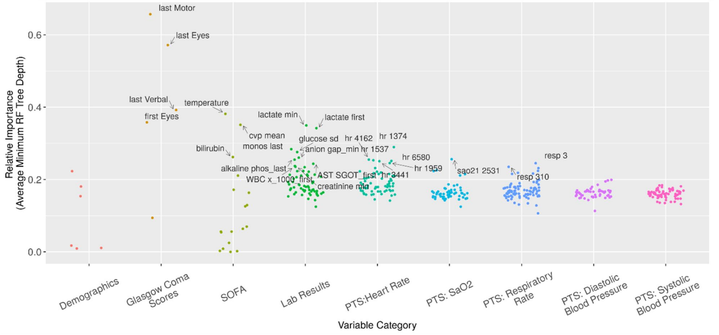

Figure 5. Categorized random forest feature ranking for our in-hospital survival prediction model. Each dot represents an individual feature, and is grouped into feature categories. A relative importance of 1.00 signifies the most important feature. The top 20 features are labeled and for the sake of simplicity, PTS derived features are identified by numbers.

The feature rankings for our in-hospital survival endpoint are shown in Figure 5. It shows the relative importance based on the average minimum random forest tree depth. The more a feature contributes to the prediction, the higher the relative importance. Overall, rankings for each endpoint show that majority of the top 50 features across all clinical outcomes were PTS derived features. Each models' features were pruned and selected from a list of 19,691 features collected across all five PTS signal types (heart rate, SpO2, respiratory rate, diastolic, and systolic blood pressure) from the first 24 hours of ICU admissions.

Conclusions & Future Work

Taken together, findings demonstrate that computational models trained with high-resolution ICU time series data recorded in the first 24 hours after ICU admission can successfully discriminate discharge neurological outcome and survival of patients resuscitated from CA. We found that physiological signals contain valuable prognostic information and that features derived from the first 24 hours 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.

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. This study warrants further exploration of the predictor variables, to gain insights on the clinical correlation and interpretability provided by those variables. In addition, 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.

References

1. Cariou, A., et al. Targeted temperature management in the ICU: Guidelines from a French expert panel. Anaesth Crit Care Pain Med 37, 481-491 (2018).
2. Sandroni, C., D'Arrigo, S. & Nolan, J.P. Prognostication after cardiac arrest. Crit Care 22, 150 (2018).
3. Callaway, C.W., et al. Part 8: Post-Cardiac Arrest Care: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 132, S465-S482 (2015).
4. Pollard, T.J., et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 5, 180178 (2018).
5. Johnson, A.E., et al. MIMIC-III, a freely accessible critical care database. Sci Data 3, 160035 (2016).
6. Balouris, S.A., et al. Development and validation of the Cerebral Performance Categories-Extended (CPC-E). Resuscitation 94, 98-105 (2015).