



Published in final edited form as:

Crit Care Med. 2022 February 01; 50(2): e162–e172. doi:10.1097/CCM.0000000000005286.

Comparison of machine learning methods for predicting outcomes after in-hospital cardiac arrest

Anoop Mayampurath, PhD^{#1}, Raffi Hagopian, MD^{#2}, Laura Venable, MS², Kyle Carey, MPH², Dana Edelson, MD, MS², Matthew Churpek, MD, MPH, PhD^{3,†}

¹Department of Pediatrics, University of Chicago, Chicago, IL

²Department of Medicine, University of Chicago, Chicago, IL

³Department of Medicine, University of Wisconsin, Madison, WI

These authors contributed equally to this work.

Abstract

Objective: Prognostication of neurological status among survivors of in-hospital cardiac arrests (IHCA) remains a challenging task for physicians. While models such as the Cardiac Arrest Survival Post-Resuscitation In-hospital (CASPRI) score are useful for predicting neurological outcomes, they were developed using traditional statistical techniques. In this study, we derive and compare the performance of several machine learning models to each other and to the CASPRI score for predicting the likelihood of favorable neurological outcomes among survivors of resuscitation.

Design: Analysis of the Get With The Guidelines-Resuscitation (GWTG-R) registry

Setting: 755 hospitals participating in GWTG-R from January 1, 2001 until January 28, 2017

Patients: Adult IHCA survivors

Interventions: None

Measurements and Main Results: Out of 117,674 patients in our cohort, 28,409 (24%) had a favorable neurological outcome, as defined as survival with a Cerebral Performance Category (CPC) Score of 2 at discharge. Using patient characteristics, pre-existing conditions, pre-arrest interventions, and peri-arrest variables, we constructed logistic regression, support-vector machines, random forests, gradient boosted machines, and neural network machine learning models to predict favorable neurologic outcome. Events prior to October 20, 2009 were used for model derivation and all subsequent events were used for validation. The gradient boosted machine predicted favorable neurological status at discharge significantly better than the CASPRI score (c-statistic: 0.81 vs. 0.73, $P < 0.001$) and outperformed all other machine learning models in terms of discrimination, calibration, and accuracy measures. Variables that were consistently most important for prediction across all models were duration of arrest, initial cardiac arrest rhythm, admission CPC score, and age.

[†]Corresponding author: Matthew Churpek, MD, MPH, PhD, 8007 Excelsior Drive, Madison, WI 53717-1903, mchurpek@medicine.wisc.edu.

Conclusions: The gradient boosted machine algorithm was the most accurate for predicting favorable neurologic outcomes in IHCA survivors. Our results highlight the utility of machine learning for predicting neurological outcomes in resuscitated patients.

Keywords

cardiac arrest; prediction; neurological outcomes; machine learning

INTRODUCTION

In-hospital cardiac arrest (IHCA) patients that have been successfully resuscitated and discharged are likely to be afflicted by severe neurological deficits^{1,2}. Resuscitated patients are often intubated, sedated, and in a state of induced hypothermia making neurologic assessments particularly challenging. This is distressing to patient families who desire accurate patient-specific prognostic information to aid them in goals of care decision-making. Additionally, incorrect prognostication of neurological recovery may bias studies by impacting the rate of true outcomes.³ The Get-With-the-Guidelines-Resuscitation (GWTG-R) Registry is a large multicenter collection of cardiac arrest data.^{4,5} These data have been leveraged in previous studies for the development of tools, such as the Cardiac Arrest Survival Post-Resuscitation In-Hospital (CASPRI) and Good Outcome Following Attempted Resuscitation (GO-FAR) scores, that estimate the likelihood of IHCA patients surviving to discharge without neurological deficits.^{6,7} These tools are based on conventional statistical methods and have been validated in external studies.⁸

Alongside the growth of large datasets, there has been a recent surge in the application of machine learning algorithms in predicting a variety of patient outcomes.⁹⁻¹³ However, it remains to be explored whether combining these algorithms with large-scale data from the GWTG-R registry can better identify IHCA resuscitation survivors with favorable neurological outcome.

The aim of this study was to determine whether the application of machine learning methods improves detection of resuscitated IHCA patients who go on to be discharged from the hospital without severe neurological deficits in comparison to the CASPRI model. We hypothesize that advanced machine learning models will discriminate patients who survive to hospital discharge without neurological deficits better than the CASPRI score. We also hypothesize that these models will demonstrate better calibration than CASPRI.

METHODS

Data sources and study population

We accessed the GWTG-R Registry to build our study population. The GWTG-R contains IHCA medical data collected from 755 hospitals across the United States since January 1st, 2000 using the Utstein template.^{4,5} Hospitals participating in the registry submit clinical information regarding the medical history, hospital care, and outcomes of consecutive patients hospitalized for cardiac arrest using an online, interactive case report form and Patient Management Tool™ (IQVIA, Parsippany, New Jersey). Data entry is performed by

trained professionals through manual chart review. We identified 333,793 cardiac arrests within the GWTG-R registry. Each patient's first arrest during a hospitalization was utilized for analysis, thereby excluding 53,491 arrests. The following exclusion criteria were applied to be consistent with criteria used for the CASPRI model: removal of arrests that took place in locations other than general medicine ward or intensive care unit (ICU) settings (n=58,817), patients without recorded return of spontaneous circulation (n=91,762), missing discharge survival status (n=1,868), or missing Cerebral Performance Category (CPC) assessment on discharge (n=10,181, see Supplementary Figure 1). These exclusion criteria reflected the goal of including only patients hospitalized in the wards or ICU who had return of spontaneous circulation with a known outcome. The institutional review board at University of Chicago reviewed and approved the study with a waiver of informed consent (IRB 17-1342).

Primary outcome

Our primary outcome of interest was survival with favorable neurological status at discharge, as indicated by the discharge CPC score. The CPC score varies from one to five. A score of one represents minimal to no neurological deficits, two implies moderate disability, and three through five represent severe disability, comatose state, and brain death, respectively. For purposes of our analysis we used a threshold of a CPC score of ≥ 2 to define favorable neurological outcome, following the outcome definition of the CASPRI score.⁶

Predictor Variables

Our models were trained using variables that are expected to be knowable by providers after a successful IHCA resuscitation, similar to those considered for the development of the CASPRI score. We therefore obtained patient- and event-level variables from GWTG, including patient age, sex, neurologic status prior to arrest (pre-arrest CPC), comorbidities and conditions (sepsis, hypotension, heart failure, coronary artery disease, diabetes mellitus, renal dysfunction, liver dysfunction, respiratory failure, dementia, focal neurological deficits, history of stroke, trauma, metastatic malignancy, hematologic malignancy, or metabolic derangement), life support devices in place prior to event (vasopressors, dialysis, mechanical ventilation, intra-aortic balloon pumps), and cardiac arrest event descriptors (night vs. day, weekend vs. weekday, duration of CPR, initial cardiac arrest rhythm, time to defibrillation).

Model Development

We derived the following machine learning models: logistic regression (LR), random forests (RF), extreme gradient boosting (XGBoost), support vector machines (SVM), and a multi-layer perceptron (MLP) neural network model. We split data longitudinally to derive and validate all models wherein data including and prior to October 20, 2009 was considered as the derivation cohort and data from October 21, 2009 to January 28, 2017 formed the validation cohort. The choice of October 20, 2009 for the temporal split was to enable direct comparison of our models to the CASPRI score model that was created by using GWTG-R data prior to this date. We further considered a split-by-site validation study design, wherein models were derived using data from four groups of hospitals and validated

on an independent fifth group. The assignment of all hospital sites into one of five groups was random and performed prior to splitting. Finally, we considered a split-by-random study design, where 70% of data was randomly chosen as the derivation cohort while the remaining 30% was selected as the validation cohort.

Missing data for LR, RF, SVM, and MLP models was imputed using predictions from a classification or regression tree for categorical and numeric variables respectively. Briefly, we constructed decision trees using complete data from the derivation cohort. Missing values in both the derivation cohort and the validation cohort was imputed using predictions from these trees. Derivation data for the MLP and the SVM model were additionally encoded so that numeric variables were centered and scaled and categorical variable were dummy-coded. The validation cohort was encoded using center-scale parameters from the derivation cohort. No imputation or encoding was performed for the XGBoost model, as it can natively handle missing data.

In terms of model hyperparameters and structure, we considered a radial basis kernel for the SVM. The MLP was designed as two successive dense and dropout layers followed by a softmax classifier. Hyperparameters for all models except MLP were optimized using the derivation cohort through 5-fold cross-validation. The MLP was optimized using 20 epochs with early-stopping based on 80%-20% training-test split of the derivation data, with AUC used for hyperparameter optimization. A complete list of hyperparameters optimized for each model is available in Supplementary Table 1. Finally, we estimated variable importance using a permutation-based approach that computes changes in loss function based on resampling observations for each variable.¹⁴

Model Performance

Discrimination of all models was assessed using the area under the receiver operating characteristic curve (AUC) metric in the out-of-sample validation cohorts. Comparison of AUCs between models was performed using DeLong's method.¹⁵ We also compared model calibration curves that depict predicted vs. actual results for all models. We compared sensitivity, specificity, positive predictive value, and negative predictive value for the best model with that of CASPRI at various model output thresholds to evaluate model performance in a clinical context. A sensitivity analysis was conducted to compare model performance on the overall test data with a sub-population of patients with admission CPC score of less than or equal to 2, as patients with poor neurological status at baseline (CPC score > 2) are unlikely to experience favorable neurological outcomes at discharge. We followed recent recommendations made by the American Heart Association,³ and have reported the performance of the prediction models using the Transparent Reporting of multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) (checklist in Supplementary Table 2). Analyses were performed using R version 3.6.2 (R Project for Statistical Computing) and Python Version 3.6.2 with $P < 0.05$ indicating statistical significance.

RESULTS

Patient Characteristics

Our cohort consisted of 117,674 cardiac arrest patients who survived resuscitation and were neurologically assessed at discharge. Of these, 28,409 (24%) had favorable neurological outcomes as indicated by the CPC score 2 at the time of discharge. Table 1 compares patient and arrest characteristics between those with and without favorable outcomes at discharge. Comparison of pre-existing conditions and intervention prior to cardiac arrest between patients with and without the outcome is given in Supplementary Table 3.

Model Performance

Predictive performances of all models are illustrated in Table 2. On a test dataset of all patients post October 20, 2009 (n=65,572), the LR model showed improved performance in detecting patients with favorable neurological outcome at discharge over the CASPRI model (LR AUC 0.79 [0.79-0.80] vs. CASPRI AUC 0.73 [0.73-0.74], $P < 0.001$). Predictive performance did not improve when comparing the LR model to SVM (SVM AUC 0.78 [0.78-0.79] vs. LR AUC 0.79 [0.79-0.78], $P < 0.001$) and to RF (RF AUC 0.79 [0.79-0.79] vs. LR AUC 0.79 [0.79-0.80], $P = 0.29$). However, the MLP and XGBoost architectures had higher AUC in comparison to the LR model (MLP AUC 0.80 [0.80-0.80], $P < 0.001$; XGBoost AUC 0.81 [0.80-0.81] vs. LR AUC 0.79 [0.79-0.80], $P < 0.001$). The XGBoost model performance did not change when predicting favorable neurological outcomes in patients with admission CPC score 2 (AUC 0.81 [0.80-0.81]).

Figure 1 depicts calibration plots for all models for predicting favorable neurological outcomes. A calibration fit with a slope of 1 and an intercept of 0, indicated by the dashed-line, signifies perfect calibration of the model to the outcome. While the prevalence estimated by Chan et al. at specific thresholds generally matched the true prevalence, the CASPRI score overestimates the true prevalence of patients with favorable neurological outcomes. The machine learning models demonstrate varying levels of calibration – while the XGBoost, RF and, the LR models show excellent calibration with slopes and intercepts approximating 1 and 0 respectively, the MLP and SVM show comparatively weaker calibration. Supplementary Figure 2 depicts rate of survival with a favorable neurological status compared to the XGBoost predicted probability of favorable neurological outcomes.

Table 2 also depicts model performances when conducting the split-by-site and the random-split study designs. Because the CASPRI model was trained on data prior to 2009, there is potential for data leakage for the split-by-site and split-by-random study designs, wherein the derivation and validation cohort may overlap. As can be seen, while there was a marginal improvement in performance for all models in the random-split design, there was no discernible difference between the longitudinal split and the split-by-site designs. Similar trends were also observed in comparing calibration metrics for all models across different derivation-validation study designs (see Supplementary Table 4).

Table 3 compares the sensitivity, specificity, positive and negative predictive values for the XGBoost model for predicting favorable neurological outcomes at discharge, with that of CASPRI scores (inverted to associate higher values with likelihood of favorable neurological

outcomes) and the LR model for the split-by-time study design. At a sensitivity of 54%, the XGBoost model had a higher specificity (87% vs 79%; 87% vs 85%), positive predictive value (57% vs. 44%; 57% vs 54%), and negative predictive value (86% vs. 84%; 86% vs. 84%) than both the CASPRI and the LR models. At a specificity of 90%, the XGBoost model had a higher sensitivity (48% vs. 35%; 48% vs. 44%), positive predictive value (61% vs. 54%; 61% vs. 59%), and negative predictive value (85% vs. 82%; 85% vs 84%) than both CASPRI and the LR model. At the best performing threshold, as identified using Youden's J-Statistic, the XGBoost model had a sensitivity of 72% (70%-77%), a specificity of 74% (69%-76%), a positive predictive value of 90% (89%-90%), and a negative predictive value of 45% (45%-48%) for predicting patients with favorable neurological outcomes. At the threshold corresponding to a 5% false-positive rate (i.e., 5% of patients identified as favorable neurological outcome will experience poor neurological outcome), the XGBoost model had a sensitivity of 34% (33%-34%), a positive predictive value of 95% (95%-96%), and a negative predictive value of 31% (31%-31%). In contrast, at the threshold corresponding to 5% false-positive rate, the CASPRI model had a sensitivity of 31% (30%-32%), a positive predictive value of 91% (91%-91%), and a negative predictive value of 29% (29%-30%). At a 10% or lower likelihood of surviving to discharge without neurological deficits (at XGBoost thresholds of 23 and Inverted CASPRI scores 30), the XGBoost model had a higher sensitivity than CASPRI (27% [95% CI: 26%-27%] vs. 14% [95% CI: 13%-14%]). Similarly, at a 5% or lower likelihood of surviving to discharge with favorable neurological outcomes (at XGBoost thresholds of 10 and Inverted CASPRI score 23), the XGBoost model had a higher sensitivity than CASPRI (7% [95% CI: 7%-8%] vs. 2% [95% CI: 1%-2%]).

Figure 2 illustrates the variables that were important for predicting favorable neurological outcomes within each model, ordered by median normalized importance across all models. Variables most important for prediction include duration of resuscitation, initial cardiac arrest rhythm, admission CPC score, and age. The most important pre-existing conditions were sepsis and cancer, while the top interventions in place were mechanical ventilation and administration of vasoactive agents. Supplementary Table 5 depicts the odds-ratios for each variable within the logistic regression model. Supplementary Figure 3 depicts partial dependence plots from the XGBoost model for age, duration of resuscitation, and admission CPC score. As shown, patients 60 or more have a lower probability of surviving to discharge without neurological deficits as compared to younger patients. The probability of favorable neurological outcome plateaus for patients whose cardiac arrest lasted for at least 15 minutes. Finally, admission CPC scores < 3 were associated with higher probability of favorable neurological outcomes in the XGBoost model.

DISCUSSION

In this study, we developed machine learning models to predict favorable neurological survival in patients who experience in-hospital cardiac arrests using the large, multicenter GWTG registry. In a test set of approximately 65,000 patients admitted in a time period after development of CASPRI in 2009, we found that a model based gradient boosting techniques outperformed the CASPRI as well as other machine learning models in terms of

discrimination, accuracy measures, and calibration. Our results indicate the potential of this model as a useful tool among resuscitated patients for assessing neurological outcomes.

Determining risk of neurological deficits is challenging in cardiac arrest survivors since the presence of intubation precludes proper neurological assessment. Recently, prior studies have successfully developed scores that allow accurate prognostication after IHCA. For example, the CASPRI score was developed using data from the GWTG registry to predict survival to hospital discharge without neurological deficits, as indicated by a discharge CPC score ≥ 2 .⁶ A CPC score at discharge of 1 or 2 indicate moderate to mild neurological deficit that still allow for the patient to live independently.¹⁶ Using the CASPRI score, clinicians could estimate where a patient falls in terms of deciles of risk and consider appropriate adjustments regarding goals of care. The CASPRI score has also been validated in an independent, external setting.⁸ Another model derived using the GWTG-R data is the GO-FAR score, that utilized variables similar to the CASPRI model to identify resuscitated patients with likelihood of surviving to discharge with minimal neurological deficits (defined as a discharge CPC Score of 1). Here we considered favorable neurological outcomes as defined in the CASPRI model as our primary outcome.

Recently, machine learning algorithms have been applied to large-scale, multi-site, national registries to predict outcomes such as mortality in rapid response events or risk of 30-day readmissions in IHCA patients.^{17,18} However, their application to predicting patients that survive to discharge without neurological deficits has not been explored. Other studies have used machine learning to predict neurological outcomes 12-24 hours post cardiac arrest using electroencephalogram (EEG) data,¹⁹ and for predicting neurological status at discharge for out-of-hospital cardiac arrests.²⁰ We derived and compared the performance of modern machine learning models in discriminating patients with favorable neurological outcomes from those without in a large national registry of in-hospital cardiac arrest patients. Among the models explored, the gradient-boosted model was overall the best in terms of improvements in discrimination over the CASPRI score, improvements in accuracy metrics such as sensitivity, specificity, positive, and negative predictive values at various thresholds, and had excellent calibration for the outcome. Our model could be used as a baseline risk adjustment or comparator algorithms when considering the association between other data modalities, such as EEG, and outcomes.³ Our model offers flexibility in setting score thresholds that are specific to the needs of a hospital or additional applications such as quality initiatives or risk adjustment for prospective studies. Our study is another example of the utility of combining machine learning methods with large datasets for development of better risk stratification methods.

The modest improvements of the more advanced methods compared to logistic regression linear model may be attributed to the highly structured predictor variables, where all but two (duration and age) were categorical variables. The improvement in discrimination observed for the gradient boosted technique may be attributed to the model's capacity to be more flexible than a logistic regression model by utilizing possible non-linearity and interactions between variables to improve performance, as observed in a recent review article.³ The logistic regression and gradient boosted model also demonstrated better calibration to the actual rate of the outcome than the other models. We additionally illustrate how partial

dependence plots can aid in interpreting direction of risk for the gradient boosted model. It is possible that the observed statistical differences in the machine learning model performances may not translate to significant change in impact from a clinical standpoint. We note that the logistic regression model is more interpretable as a bedside score. However, it does require imputation of missing values. On the other hand, the data itself does not need any pre-processing or imputation before ingesting into the gradient boosted model. Thus, the choice behind implementation may be based on accuracy, interpretability, and implementation considerations. If the motivation is to acquire accurate prediction of favorable neurological survival without need for preprocessing data, then the gradient boosted model is preferred. However, the logistic regression model is preferable in cases where data preprocessing is not an issue, and there is a need to more directly assess the association between each variable and neurological survival for each new patient.

By testing three validation study-designs, our study further informs implementation decisions regarding the models. The temporal validation allowed us to assess model generalizability to new patient cases, which is of most relevant to clinicians already participating in the GWTG database seeking to apply these models to future patients. The external validation strategy allows for assessing model performance in a new hospital setting. Our objective in conducting random validation, which typically ensures maximal performance, was to ensure there was no bias in the temporal or external validation. The equivalent performance observed for all validation study designs suggests that these models are stable across time and across hospital site.

Our work also highlights the importance of duration of resuscitation, initial cardiac arrest rhythm, admission CPC, and patient age in determining likelihood of neurological survival in resuscitation survivors. These variables are included in the CASPRI score as well as considered to be crucial for outcomes in IHCA patients.^{6,21–23} Pre-existing conditions such as cancer, hepatic insufficiency, and sepsis that were associated with poor neurological outcomes in this study are also supported by other studies.^{21,24} Similarly, the association between poor outcomes and incidence of mechanical ventilation and administration of vasoactive agents in IHCA patients have also been documented previously.^{25,26}

Our study has the following limitations. First, our models utilize retrospective data elements within the GWTG registry, and thus may miss important features. Second, data from the GWTG registry may not be generalizable externally. Further, our exclusion criteria involved variables, such as ROSC, that may not be missing at random and could introduce bias in our cohort selection. In particular, our data might contain patients with poor neurological outcomes due to early withdrawal of life-sustaining treatment who might have otherwise survived with a favorable neurologic status.²⁷ Finally, similar to the CASPRI model, our models cannot be extended to out-of-hospital survivors nor utilized directly for treatment or prognostication. As with any prediction model, individual patient prognosis and care decisions need to be made using additional data and context. Previous studies have highlighted the limitations of predictive models in neurological prognostication.²⁸ Further, prospective studies have demonstrated that a significant proportion of comatose cardiac arrest survivors achieved good neurological outcomes even with awakening after 6 days.²⁹

Our models are best suited for estimating likelihood of neurological survival for IHCA survivors for the purpose of quality initiatives and risk-adjustment.

CONCLUSION

We developed machine learning models using large-scale multicenter data that improve detection of resuscitation survivors who are likely to be discharged without poor neurological deficits. These models may provide clinicians with important information for quality initiatives as well as could be used in risk adjustment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to thank Mary Akel for administrative assistance. We also thank the members of the American Heart Association's Get With the Guidelines Adult Research Task Force: Anne Grossestreuer PhD; Ari Moskowitz MD; Joseph Ornato MD; Mary Ann Peberdy MD; Michael Kurz MD MS-HES; Monique Anderson Starks MD MHS; Paul Chan MD MSc; Saket Girotra MBBS SM; Sarah Perman MD MSCE; Zachary Goldberger MD MS. IQVIA serves as the data collection (through their Patient Management Tool – PMT™) and coordination center for GWTG. The University of Pennsylvania serves as the data analytic center and has an agreement to prepare the data for research purposes.

Conflicts & Disclosures:

Dr. Mayampurath is supported by a career development award from the National Heart, Lung, and Blood Institute (K01HL148390). Dr. Churpek is supported by an R01 from NIGMS (R01 GM123193). Dr. Mayampurath has performed consulting services for Litmus Health (Austin, TX). Drs. Churpek and Edelson have a patent pending (ARCD. P0535US.P2) for risk stratification algorithms for hospitalized patients and has received research support from EarlySense (Tel Aviv, Israel). Dr. Edelson has received research support and honoraria from Philips Healthcare (Andover, MA). Dr. Edelson has ownership interest in AgileMD (San Francisco, CA), which licenses eCART, a patient risk analytic.

Copyright Form Disclosure:

Dr. Mayampurath received funding from the National Institutes of Health (NIH) and the National Heart, Lung, and Blood Institute and Litmus Health. Dr. Edelson received funding from the Department of Defense (E01W81XWH2110009), the University of Chicago through a patent pending (ARCD.P0535US.P2), and AgileMD. She received research support and honoraria from Philips Healthcare. Dr. Churpek received funding from the NIH and Early Sense. Drs. Mayampurath and Churpek received support for article research from the NIH. The remaining authors have disclosed that they do not have any potential conflicts of interest.

REFERENCES

1. Ebell MH, Becker LA, Barry HC, Hagen M. Survival After In-Hospital Cardiopulmonary Resuscitation. *J Gen Intern Med.* 1998;13(12):805–816. doi:10.1046/j.1525-1497.1998.00244.x [PubMed: 9844078]
2. Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS. Trends in Survival after In-Hospital Cardiac Arrest. *New England Journal of Medicine.* 2012;367(20):1912–1920. doi:10.1056/NEJMoa1109148
3. Geocadin Romergryko G, Callaway Clifton W, Fink Ericka L, et al. Standards for Studies of Neurological Prognostication in Comatose Survivors of Cardiac Arrest: A Scientific Statement From the American Heart Association. *Circulation.* 2019;140(9):e517–e542. doi:10.1161/CIR.0000000000000702 [PubMed: 31291775]

4. Goldberger ZD, Chan PS, Berg RA, et al. Duration of resuscitation efforts and survival after in-hospital cardiac arrest: an observational study. *Lancet*. 2012;380(9852):1473–1481. doi:10.1016/S0140-6736(12)60862-9 [PubMed: 22958912]
5. Peberdy MA, Kaye W, Ornato JP, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation*. 2003;58(3):297–308. doi:10.1016/s0300-9572(03)00215-6 [PubMed: 12969608]
6. Chan PS, Spertus JA, Krumholz HM, et al. A Validated Prediction Tool for Initial Survivors of In-Hospital Cardiac Arrest. *Arch Intern Med*. 2012;172(12):947–953. doi:10.1001/archinternmed.2012.2050 [PubMed: 22641228]
7. Ebell MH, Jang W, Shen Y, Geocadin RG, Get With the Guidelines–Resuscitation Investigators. Development and validation of the Good Outcome Following Attempted Resuscitation (GO-FAR) score to predict neurologically intact survival after in-hospital cardiopulmonary resuscitation. *JAMA Intern Med*. 2013;173(20):1872–1878. doi:10.1001/jamainternmed.2013.10037 [PubMed: 24018585]
8. Wang C-H, Chang W-T, Huang C-H, et al. Validation of the Cardiac Arrest Survival Postresuscitation In-hospital (CASPRI) score in an East Asian population. *PLoS One*. 2018;13(8). doi:10.1371/journal.pone.0202938
9. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Critical care medicine*. 2016;44(2):368–374. doi:10.1097/CCM.0000000000001571 [PubMed: 26771782]
10. Giannini HM, Ginestra JC, Chivers C, et al. A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice*. *Critical Care Medicine*. 2019;47(11):1485–1492. doi:10.1097/CCM.0000000000003891 [PubMed: 31389839]
11. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. *Crit Care Med*. 2018;46(7):1070–1077. doi:10.1097/CCM.0000000000003123 [PubMed: 29596073]
12. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572(7767):116–119. doi:10.1038/s41586-019-1390-1 [PubMed: 31367026]
13. Mayampurath A, Sanchez-Pinto LN, Carey KA, Venable L-R, Churpek M. Combining patient visual timelines with deep learning to predict mortality. *PLOS ONE*. 2019;14(7):e0220640. doi:10.1371/journal.pone.0220640 [PubMed: 31365580]
14. Biecek P DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*. 2018;19(84):1–5.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845. [PubMed: 3203132]
16. Stiell IG, Nesbitt LP, Nichol G, et al. Comparison of the Cerebral Performance Category score and the Health Utilities Index for survivors of cardiac arrest. *Ann Emerg Med*. 2009;53(2):241–248. doi:10.1016/j.annemergmed.2008.03.018 [PubMed: 18450329]
17. Shappell C, Snyder A, Edelson DP, Churpek MM. Predictors of In-hospital Mortality after Rapid Response Team Calls in a 274 Hospital Nationwide Sample. *Crit Care Med*. 2018;46(7):1041–1048. doi:10.1097/CCM.0000000000002926 [PubMed: 29293147]
18. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. *JAMA Cardiol*. 2017;2(2):204–209. doi:10.1001/jamacardio.2016.3956 [PubMed: 27784047]
19. Tjepkema-Cloostermans MC, da Silva Lourenço C, Ruijter BJ, et al. Outcome Prediction in Postanoxic Coma With Deep Learning. *Crit Care Med*. 2019;47(10):1424–1432. doi:10.1097/CCM.0000000000003854 [PubMed: 31162190]

20. Park JH, Shin SD, Song KJ, et al. Prediction of good neurological recovery after out-of-hospital cardiac arrest: A machine learning analysis. *Resuscitation*. 2019;142:127–135. doi:10.1016/j.resuscitation.2019.07.020 [PubMed: 31362082]
21. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-Hospital Cardiac Arrest: A Review. *JAMA*. 2019;321(12):1200–1210. doi:10.1001/jama.2019.1696 [PubMed: 30912843]
22. Coppler PJ, Elmer J, Calderon L, et al. Validation of the Pittsburgh Cardiac Arrest Category illness severity score. *Resuscitation*. 2015;89:86–92. doi:10.1016/j.resuscitation.2015.01.020 [PubMed: 25636896]
23. Rohlin O, Taeri T, Netzereab S, Ullemark E, Djärv T. Duration of CPR and impact on 30-day survival after ROSC for in-hospital cardiac arrest-A Swedish cohort study. *Resuscitation*. 2018;132:1–5. doi:10.1016/j.resuscitation.2018.08.017 [PubMed: 30138650]
24. Ebell MH, Afonso AM. Pre-arrest predictors of failure to survive after in-hospital cardiopulmonary resuscitation: a meta-analysis. *Fam Pract*. 2011;28(5):505–515. doi:10.1093/fampra/cmr023 [PubMed: 21596693]
25. Grmec S, Mally S. Vasopressin improves outcome in out-of-hospital cardiopulmonary resuscitation of ventricular fibrillation and pulseless ventricular tachycardia: a observational cohort study. *Crit Care*. 2006;10(1):R13. doi:10.1186/cc3967 [PubMed: 16420660]
26. Sutherasan Y, Peñuelas O, Muriel A, et al. Management and outcome of mechanically ventilated patients after cardiac arrest. *Crit Care*. 2015;19(1). doi:10.1186/s13054-015-0922-9
27. Elmer J, Torres C, Tp A, et al. Association of early withdrawal of life-sustaining therapy for perceived neurological prognosis with mortality after cardiac arrest. *Resuscitation*. 2016;102:127–135. doi:10.1016/j.resuscitation.2016.01.016 [PubMed: 26836944]
28. Carter EL, Hutchinson PJA, Koliass AG, Menon DK. Predicting the outcome for individual patients with traumatic brain injury: a case-based review. *British Journal of Neurosurgery*. 2016;30(2):227–232. doi:10.3109/02688697.2016.1139048 [PubMed: 26853860]
29. Nakstad ER, Stær-Jensen H, Wimmer H, et al. Late awakening, prognostic factors and long-term outcome in out-of-hospital cardiac arrest – results of the prospective Norwegian Cardio-Respiratory Arrest Study (NORCAST). *Resuscitation*. 2020;149:170–179. doi:10.1016/j.resuscitation.2019.12.031 [PubMed: 31926258]

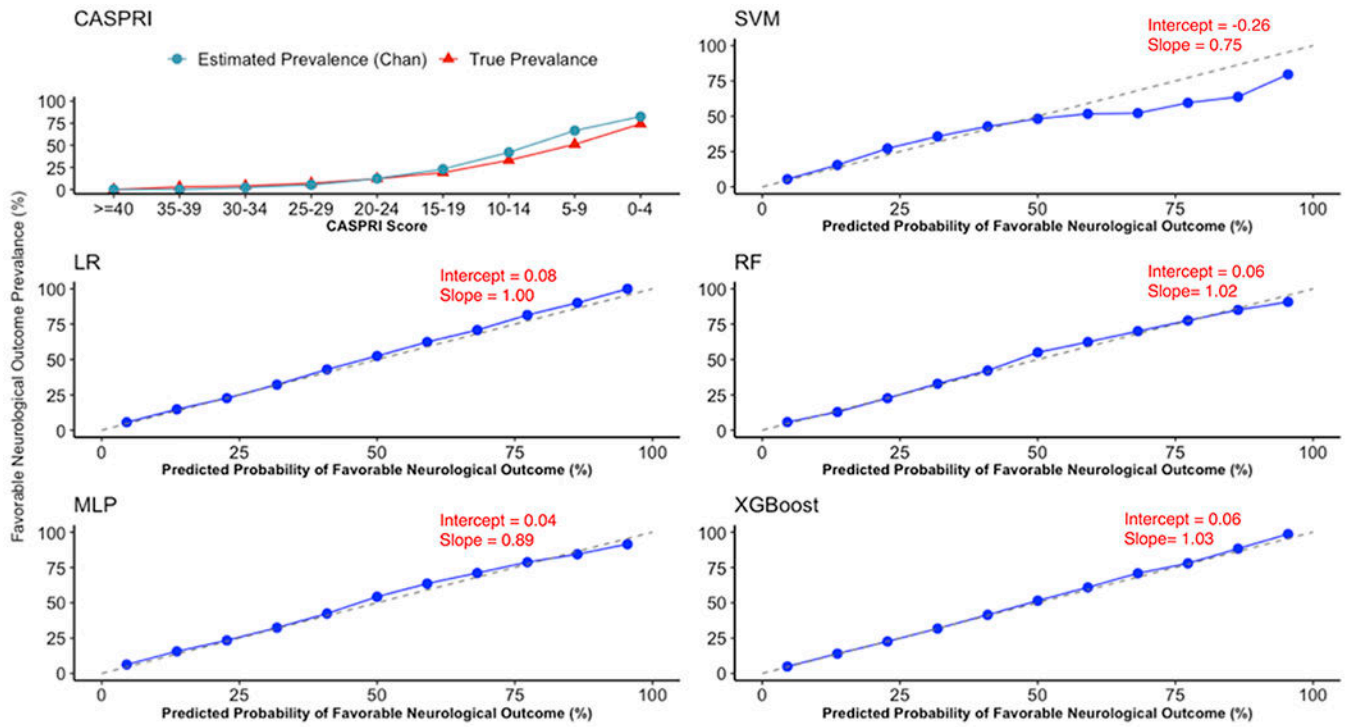


Figure 1: Calibration plot for CASPRI (score in depicted in reserve) and all machine learning models demonstrating alignment between predicted probability of favorable neurological outcome at discharge against true outcome rate.

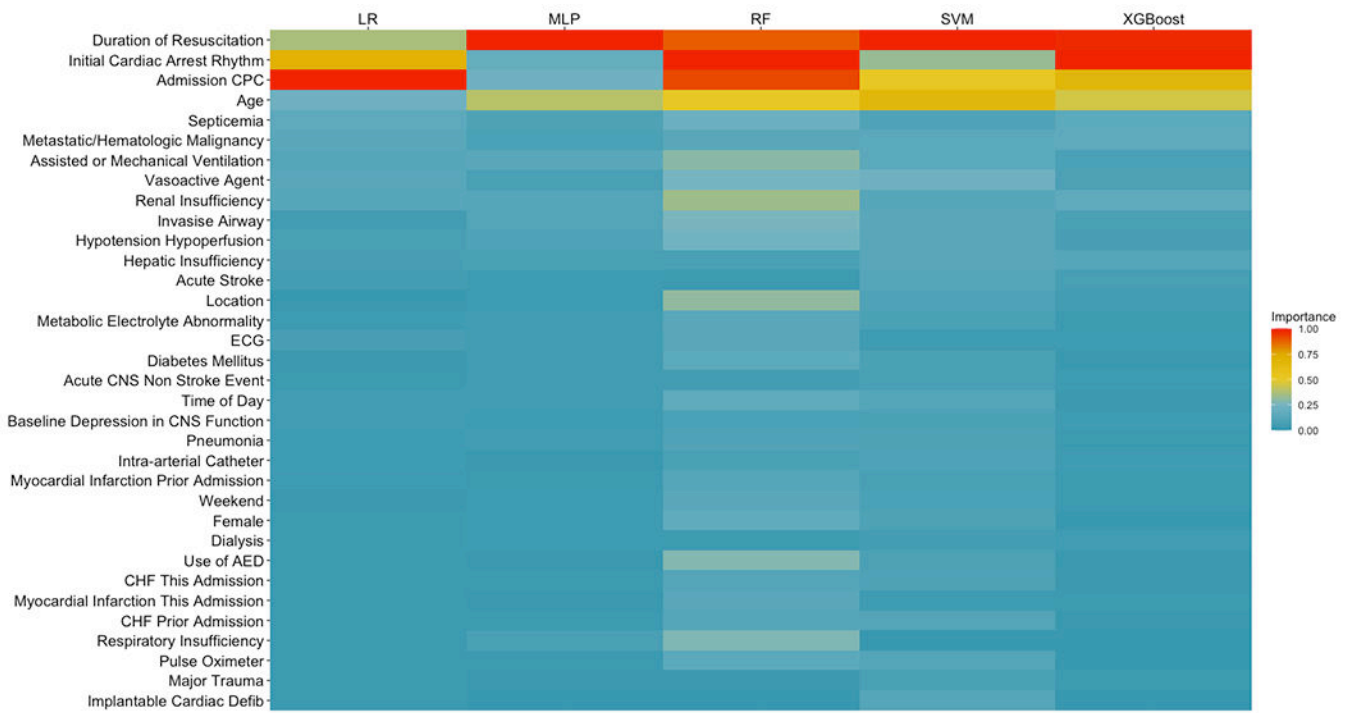


Figure 2: Variable importance across all models.

Table 1:

Clinical and Arrest Characteristics of Patients With and Without Favorable Neurological Outcome at Discharge

Variable type	Variable	Patients with favorable neurological outcome (n=28409)	Patients without favorable neurological outcome (n=89265)	P-value
Demographics	Age, mean (sd)	64.1 (14.8)	66.1 (15.5)	<0.001
	Female, n (%)	11838 (41.7%)	38931 (43.6%)	<0.001
	Race, n (%)			
	Black	4764 (16.8%)	20572 (23.0%)	<0.001
	White	21371 (75.2%)	60035 (67.3%)	
	Other	2226 (7.8%)	8429 (9.4%)	
	Missing	48 (0.2%)	229 (0.3%)	
Characteristics of Arrest, n (%)	Initial Cardiac Arrest Rhythm			
	Asystole	5712 (20.1%)	25246 (28.3%)	<0.001
	Pulseless Electrical Activity	9557 (33.6%)	42797 (47.9%)	
	VT/VF T2FS<2min	5950 (20.9%)	7446 (8.3%)	
	VT/VF T2FS 2-3	1776 (6.3%)	2268 (2.5%)	
	VT/VF T2FS 3-4	677 (2.4%)	787 (0.9%)	
	VT/VF T2FS 4-5	347 (1.2%)	543 (0.6%)	
	VT/VF T2FS >5min	920 (3.2%)	1992 (2.2%)	
	Unknown	3470 (12.2%)	8186 (9.2%)	
	Duration of Resuscitation, minutes, median (IQR)	7 (3-14)	11 (6-21)	<0.001
	Hospital Location			
	Telemetry	7706 (27.1%)	15408 (17.3%)	<0.001
	Intensive Care Unit	15063 (53.0%)	55856 (62.6%)	
	Inpatient	5640 (19.8%)	18001 (20.2%)	
	Time and Day of Arrest			
	Night	8255 (29.1%)	29980 (33.6%)	<0.001
	Weekend	8220 (28.9%)	28476 (31.9%)	<0.001
	Use of AED			
	Yes	6450 (22.7%)	21063 (23.6%)	0.002
	No	13795 (48.6%)	43232 (48.4%)	
	Not used-by-facility/NA	8164 (28.7%)	24970(28.0%)	
	CPC Score prior to arrest			<0.001
	1	19261 (67.8%)	35609 (39.9%)	
	2	5250 (18.5%)	16291 (18.3%)	
	3	1017 (3.6%)	9559 (10.7%)	
	4 or 5	720 (2.5%)	5585 (6.3%)	
	Missing	2161 (7.6%)	22221 (24.9%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Comparison of Model Performances for Predicting Favorable Neurological Outcome at Discharge Based on Various Strategies for Constructing Derivation and Validation Data

Model	Split-by-time study design (AUC, 95%CI)	Split-by-site study design (AUC, 95%CI)	Split-by-random study design (AUC, 95%CI)
CASPRI	0.73 (0.73-0.74)	0.74 (0.74-0.75) *	0.74 (0.74-0.75) *
SVM	0.78 (0.78-0.79)	0.79 (0.79-0.79)	0.79 (0.78-0.79)
LR	0.79 (0.79-0.80)	0.79 (0.79-0.80)	0.80 (0.79-0.80)
RF	0.79 (0.79-0.79)	0.79 (0.79-0.80)	0.80 (0.79-0.81)
MLP	0.80 (0.80-0.80)	0.80 (0.80-0.80)	0.81 (0.80-0.81)
XGBoost	0.81 (0.80-0.81)	0.81 (0.81-0.81)	0.81 (0.80-0.81)

* Potential data leakage: the test data in the split-by-site and the split-by-random study design may include observations before 2009, which was used to derive the CASPRI model

AUC: Area Under the receiver operating characteristic Curve

CI: Confidence Interval

CASPRI: Cardiac Arrest Survival Post-Resuscitation In-hospital score

SVM: Support Vector Machine

LR: Logistic Regression

RF: Random Forest

MLP: Multi-Layer Perceptron

XGBoost: eXtreme Gradient Boosted machine

Table 3:

Sensitivity, specificity, positive and negative predictive values at various thresholds for the XGBoost, LR, and CASPRI models for predicting favorable neurological outcomes. Both XGBoost and LR model outputs were multiplied by 100. CASPRI scores were inverted (by subtracting from maximum score of 50) in order to associate higher scores with favorable neurological outcomes.

Model Cutoff	Sensitivity(% , 95%CI)	Specificity (% , 95% CI)	Positive Predictive Value (% , 95% CI)	Negative Predictive Value (% , 95% CI)
XGBoost (Predicted Probability x100)				
4	99 (99-99)	14 (14-14)	27 (27-27)	98 (97-98)
10	93 (92-93)	41 (41-42)	33 (33-33)	95 (94-95)
20	78 (77-79)	67 (67-67)	43 (42-43)	91 (90-91)
30	63 (62-63)	82 (81-82)	52 (51-52)	87 (87-88)
36	54 (53-55)	87 (87-88)	57 (57-58)	86 (86-86)
40	48 (47-49)	90 (90-90)	61 (60-61)	85 (85-85)
50	36 (35-36)	95 (95-95)	68 (67-69)	82 (82-82)
60	24 (24-25)	98 (97-98)	75 (74-76)	80 (80-81)
70	15 (15-16)	99 (99-99)	81 (80-82)	79 (79-79)
LR (Predicted Probability x100)				
4	98 (98-99)	16 (15-16)	27 (27-27)	97 (96-97)
10	92 (91-92)	41 (41-42)	33 (33-33)	94 (94-94)
20	77 (76-78)	66 (65-67)	41 (41-42)	90 (90-90)
30	61 (60-62)	81(80-81)	50 (50-51)	87 (87-87)
34	54 (54-55)	85 (85-85)	54 (53-54)	85 (86-86)
40	44 (44-45)	90 (90-91)	59 (58-60)	84 (83-84)
50	29 (29-30)	95 (95-96)	67 (66-68)	81 (81-81)
60	18 (18-19)	98 (98-98)	74 (73-75)	79 (79-79)
70	10 (10-10)	99 (99-99)	82 (80-83)	78 (78-78)
Inverted CASPRI score				
26	97 (96-97)	15 (15-15)	26 (26-27)	94 (93-94)
30	89 (89-90)	33 (32-33)	30 (29-30)	91 (90-91)
32	83 (82-83)	46 (46-46)	32 (32-33)	89 (89-90)
37	54 (54-55)	79 (78-79)	44 (44-45)	84 (84-85)
40	35 (35-36)	90 (90-91)	54 (53-55)	82 (81-82)
43	20 (19-21)	96 (96-97)	63 (62-65)	79 (79-79)