A Physiology-Driven Machine Learning Model for Traumatic Brain Injury Outcome Prediction Using a Large Multi-Center Database

LCICM

JOHNS HOPKINS UNIVERSITY & MEDICINE

Han Kim MSE[1,2], Robert D Stevens MD[1,2]

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

Introduction

Traumatic Brain Injury (TBI) is a global health concern associated not only with significant mortality but also cognitive impairment and reduced quality of life in a significant proportion of survivors.[1] In the United States alone, TBI accounts for about 30% of all injury related deaths with about 300,000 hospitalizations per year and 57,000 deaths.[2] The estimated economic burden of TBI (direct and indirect medical costs) in 2010 was $76.5 billion.[3,4] Existing approaches to classify TBI and predict its outcomes, which include the widely validated International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomization After Significant Head injury (CRASH) models, do not capture the complexity of TBI which encompasses a broad array of clinical and biological features.[5,6]

Objectives/Aims

This goal of the study was to identify physiologic signatures in TBI patients recorded in the first 24 hours of ICU admission, testing the hypothesis that these signatures are associated with short term clinical outcomes. **We hypothesize that extending the physiologic signatures beyond traditional electronic health record (EHR) features and IMPACT and CRASH risk factors to include time dependent aperiodic clinical measures and physiologic time series (PTS) signals will significantly improve clinical outcome prognostication of ICU stratum TBI patients.** Two short-term hospital discharge endpoints were evaluated and modeled: 1. in-hospital mortality, and 2. neurological outcome based on dichotomized motor Glasgow Coma score. The study was conducted using data from the multicenter Philips eICU-CRD database,[7] and externally validated on Medical Information Mart for Intensive Care (MIMIC) III database.[8] Published CRASH and IMPACT model performances along with our TBI cohort's CRASH and IMPACT predictive performance were used as benchmark prediction systems.

Methods

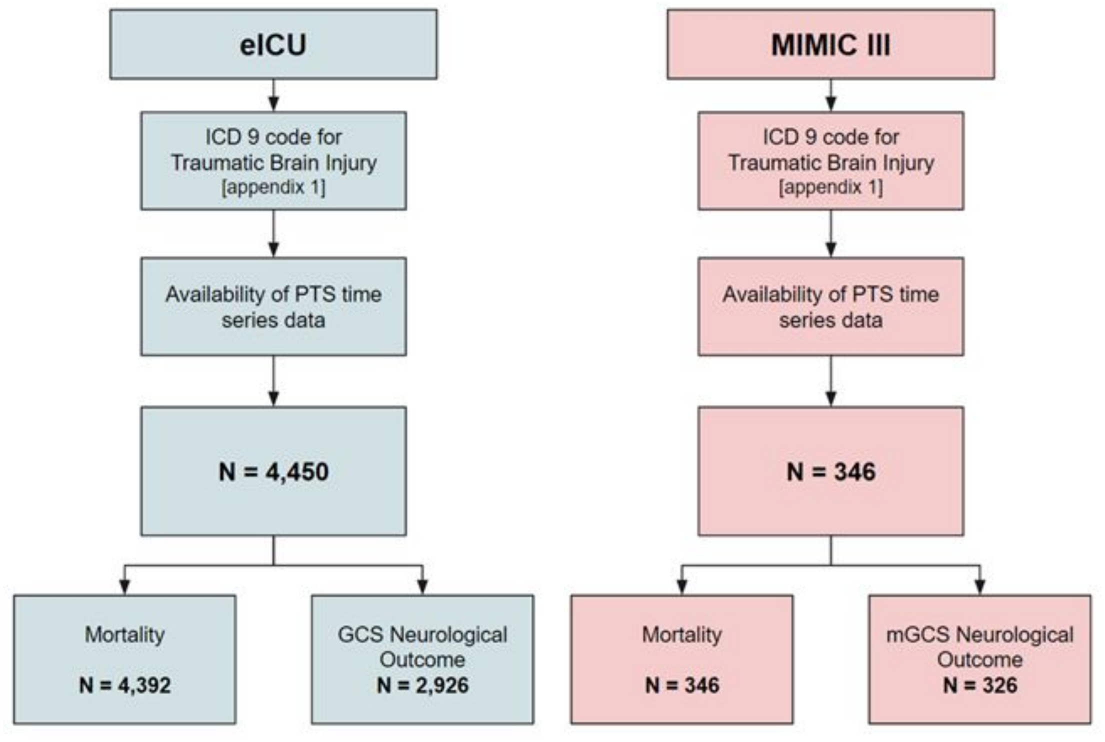

Figure 1. Study inclusion and exclusion criteria identifying the 4,450 eICU TBI patients based on admission diagnosis (ICD 9 code) and data availability. 346 MIMIC III TBI patients were identified following the same criteria. Based on the availability of clinical endpoints, the final modeled sample size differed.

In a multisite clinical database of 208 institutions in the US (eICU), we identified patients admitted to the ICU with a diagnosis of TBI (n=4450). Predictive features of interest were clinical variables, laboratory results, and physiologic time series data (PTS, i.e., high frequency monitoring data including heart rate, SaO2, respiratory rate, and blood pressure). Three different machine learning (ML) algorithms generalized linear models (GLM), random forest (RF), and XGBoost (XG), were trained on a statistically pruned feature space of 147 derived variables using mortality and neurological outcome at hospital discharge as clinical endpoints.

Results

	eICU (n = 4,450)	MIMIC (n = 346)
Age (sd)	61.57 (21.9)	61.70 (21.77)
Gender = male (%)	2,706 (60.8)	207 (59.8)
Mild TBI (%)	2,340 (52.6)	158 (45.7)
Moderate TBI (%)	1,081 (24.3)	71 (20.5)
Severe TBI (%)	1,029 (23.1)	117 (33.8)
Neurological outcome (mGCS) = unfavorable outcome (%)	657 (22.5)	55 (16.9)
In hospital mortality = expired (%)	546 (12.4)	40 (11.6)

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

ML models performed well for both neurological outcome prediction and for mortality prediction (Figure 2, Table 2). ML model performance was significantly higher than IMPACT and CRASH for neurological outcome and mortality at discharge. Additionally, published AUROC for IMPACT and CRASH models ranged from 0.79 to 0.82 for mortality and 0.77 and 0.78 for neurological outcome prediction.[9] Our IMPACT and CRASH AUROCs were well within the literature review bounds and provides confidence in the consistency of our eICU and MIMIC III TBI population to prior works.

External validation utilizing MIMIC III corroborated the results from eICU for both neurological outcome and mortality. Our eICU developed model was generalizable to the MIMIC III TBI cohort as observed by an increase in performance metrics for both clinical outcomes.

The value of integrating PTS derived features was clearly observed in features ranked by beta coefficients of our trained GLM model. Figure 4 shows the rankings for neurological outcome prediction. We see that of the top 50 features, 22 features were PTS derived features. Similar importance of PTS features was seen for mortality prediction (not shown). 24 of the 50 top beta coefficients were PTS derived features. We were able to observe the clear discriminative benefits of PTS derived variables in combination to EHR derived variables.

eICU Development Model

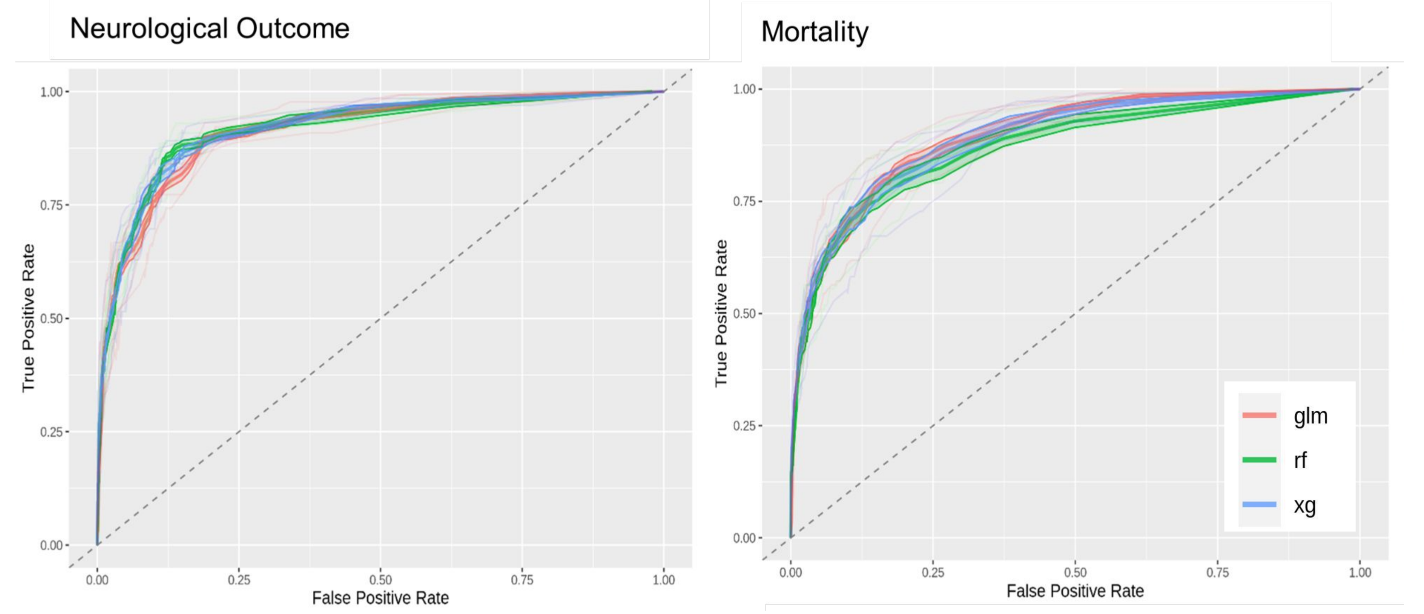

Figure 2. ROC curves of our glm, rf, and xgboost eICU TBI cohort developed models utilizing the 147 pruned EHR and PTS derived features to predict Neurological outcome (left) and mortality (right).

Short-term Neurological Outcome			Reference Models	
	eICU model	MIMIC III validation	IMPACT	CRASH
AUROC	0.923 ± 0.015	0.924 ± 0.005	0.790	0.776
Sensitivity	0.866 ± 0.048	0.844 ± 0.044	0.773	0.790
Specificity	0.866 ± 0.017	0.849 ± 0.027	0.646	0.613
Short-term Mortality Outcome			***Reference Models***	
	eICU model	MIMIC III validation	IMPACT	CRASH
AUROC	0.900 ± 0.027	0.931 ± 0.011	0.879	0.867
Sensitivity	0.796 ± 0.058	0.880 ± 0.040	0.852	0.810
Specificity	0.835 ± 0.034	0.841 ± 0.062	0.727	0.755

Table 2. Performance metric summary of the eICU-CRD development model and MIMIC III external validation for each evaluated clinical endpoint (neurological outcome and mortality). The MIMIC-III cohort external validation results provide evidence that our eICU developed TBI prognostication model is generalizable. Our models, both the eICU development and MIMIC-III validation performs significantly better than IMPACT and CRASH ($p<0.05$).

MIMIC-III External Validation

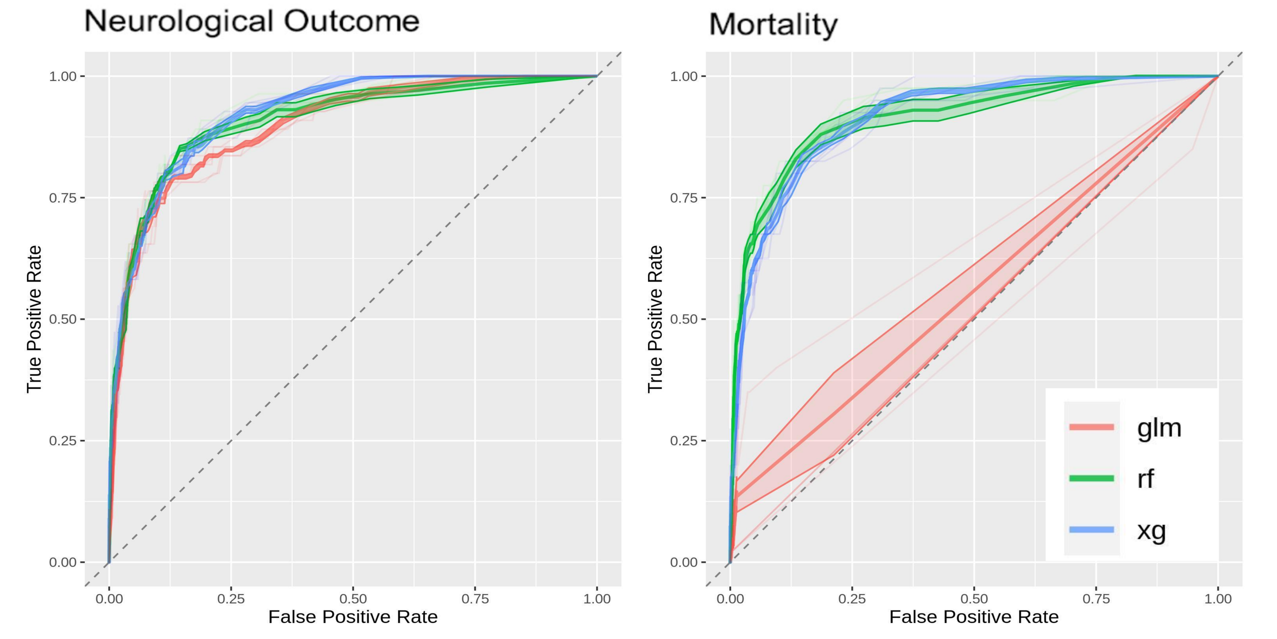

Figure 3. ROC curves for our MIMIC III external validation of the glm, rf, and xgboost eICU TBI cohort developed models to predict Neurological outcome (left) and mortality (right). Aside from the overfitting experience by our at-discharge mortality prediction GLM model, the MIMIC III validation cohort showed similar if not better model performances (Table 2).

Results (cont.)

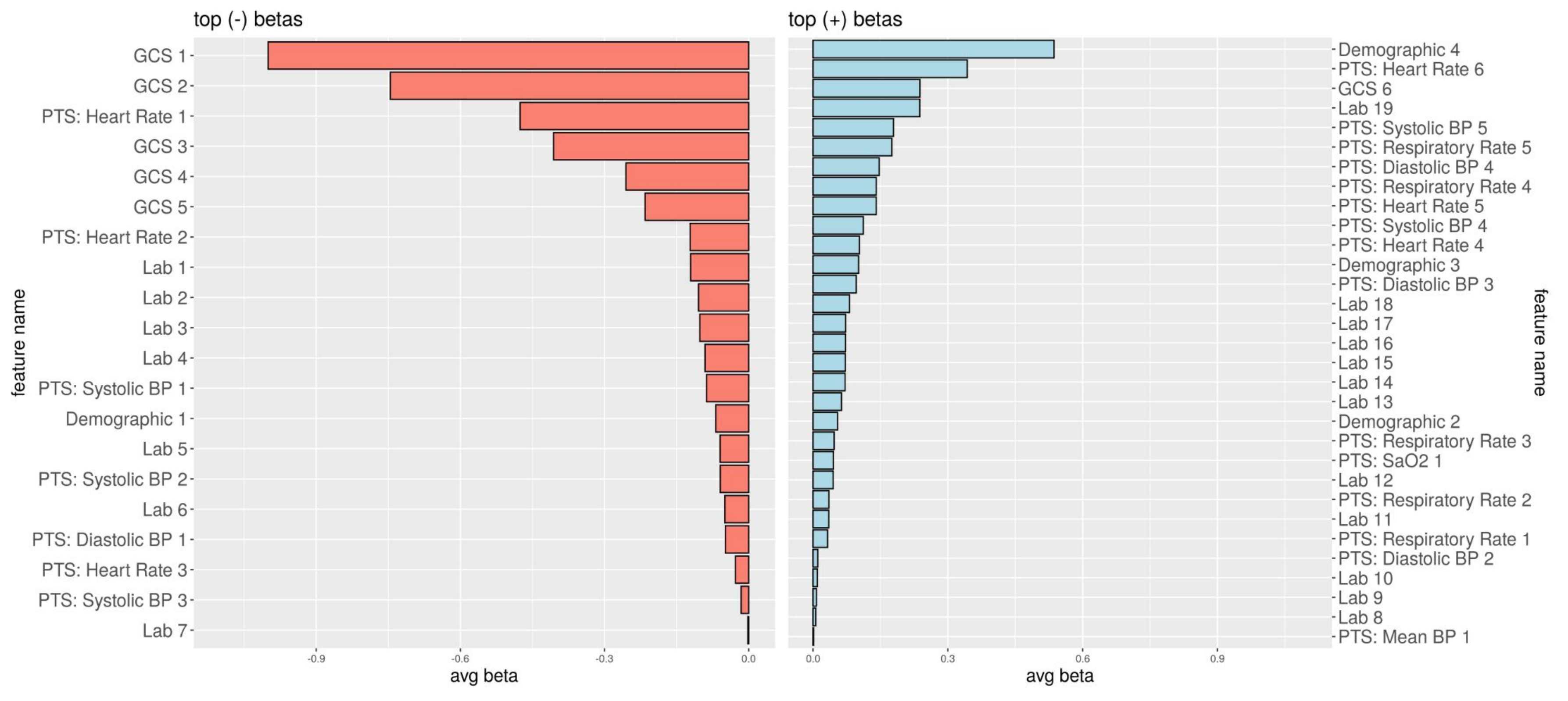

Figure 4. Generalized Linear Model (GLM) average beta coefficient feature ranking for neurological outcome prediction. Left side (red) are the negative beta coefficients, with higher individual feature values increasing the likelihood for favorable neurological outcome. Larger values of features on the right side (blue) increases the likelihood of unfavorable neurological outcome. Of the top 50 features shown, 22 were PTS signal derived features. Feature names were simplified based on categories to simplify the interpretability.

Conclusions & Future Work

Our results demonstrate that physiology-driven ML approaches significantly outperform IMPACT and CRASH logistic regression models for both neurological outcome and mortality prediction for ICU stratum TBI patients. These models, established using the multi-center eICU-CRD cohort, underwent successful external validation in the MIMIC-III single center TBI cohort and provides increasing confidence in the multicenter model's generalizability.

Results suggest that a data-driven approach incorporating PTS derived features captures prognostically relevant information on TBI patients which may be overlooked in existing TBI prediction systems.

Further work is being conducted to extend the TBI external validation analysis to MIMIC-IV, Amsterdam University Medical Center (AUMC), and Johns Hopkins Medicine patient datasets. This will help provide further evidence that our eICU developed model can be generalized to multiple adult TBI populations, especially two European cohorts and another large single center education medical institute.

References

1. Stocchetti, N., et al., Traumatic brain injury in an aging population. J Neurotrauma, 2012. 29(6): p. 1119-25.
2. Dutton, R.P. and M. McCunn, Traumatic brain injury. Curr Opin Crit Care, 2003. 9(6): p. 503-9.
3. Perel, P., et al., Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak, 2006. 6: p. 38.
4. Maas, A.I.R., et al., Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol, 2017. 16(12): p. 987-1048.
5. Steyerberg, E.W., et al., Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med, 2008. 5(8): p. e165; discussion e165.
6. Collaborators, M.C.T., et al., Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ, 2008. 336(7641): p. 425-9.
7. Pollard, T.J., et al., The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data, 2018. 5: p. 180178.
8. Johnson, A.E., et al., MIMIC-III, a freely accessible critical care database. Sci Data, 2016. 3: p. 160035.
9. Dijkland, S.A., et al., Prognosis in Moderate and Severe Traumatic Brain Injury: A Systematic Review of Contemporary Models and Validation Studies. J Neurotrauma, 2019.