with eICU-CRD being the largest (> 130 K) and AmsterdamUMCDB the smallest (> 23 k). ICU mortality and intensity of treatment also varied, with 28-day mortality rates and frequency of ventilation being lowest in eICU-CRD. Frequency of lab values tended to be highest in MIMIC-IV, while frequency of vital signs was highest in AmsterdamUMCDB.
Conclusion. Several high-quality ICU databases are currently available. The research question, and thus required sample size, presence of covariates and frequency of measurements, should inform which database to use. Due to the underlying differences between the datasets, we suggest using at least two databases to ensure generalizability of findings.

Reference(s)

1. Johnson AEW, Pollard TJ, Shen L, et al. (2016) MIMIC-III, a freely accessible critical care database. Sci Data 3:1–9. https://doi.org/10.1038/sdata.2016.35
2. Pollard TJ, Johnson AEW, Raffa JD, et al. (2018) The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 5:1–13. https://doi.org/10.1038/sdata.2018.178
3. Thoral PJ, Peppink JM, Driessen RH, et al. (2021) Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example.
4. Faltys, M., Zimmermann, M., Lyu, X., et al. (2021). HiRID, a high time-resolution ICU dataset (version 1.1.1). PhysioNet. https://doi.org/10.13026/nkwc-js72.

000384

Machine learning to understand unmet needs for multiclass outcome prediction in the ICU after traumatic brain injury

S. Bhattacharyay[1]; E. Steyerberg[2]; A. Maas[3]; D. Menon[1]; A. Ercole[1]
[1]Division of Anaesthesia, University of Cambridge, Cambridge, United Kingdom; [2]Department of Biomedical Data Sciences, Leiden University Medical Center (LUMC), Leiden, Netherlands; [3]Department of Neurosurgery, Antwerp University Hospital, Edegem, Belgium
Correspondence: S. Bhattacharyay

Intensive Care Medicine Experimental 2021, **9(1)**: 000384

Introduction. Well validated models exist only for dichotomous outcome prognosis after traumatic brain injury (TBI). Dichotomisation, however, fails to capture the broad continuum of outcomes which may be crucial for shared decision-making and treatment stratification in the ICU. On the other hand, multiclass TBI outcome prediction performance is poor, especially for intermediate outcomes, but it is unclear whether this is due to inflexible modelling strategies (i.e., model complexity and outcome encoding) or inadequate disease characterisation by existing predictors.
Objectives. We assess multiclass prediction of the Glasgow Outcome Scale—Extended (GOSE) [1] at 6 months from standard TBI predictors as a function of endpoint encoding (multinomial vs ordinal) and model complexity (logistic regression vs deep learning).
Methods. We use data from a prospective cohort of 3,573 patients from the CENTER-TBI [2] study, extracted using Opal [3]. We use the IMPACT [4] extended predictor set (10 covariates collected within 24 h of ED/ICU admission) including age, clinical severity scores, secondary insult indicators, CT characteristics, and lab values. Missing values are multiply imputed ($m = 100$), and we train 4 multiclass prediction model types: multinomial logistic regression (MLR), ordinal logistic regression (OLR), neural network with a multinomial output layer (DeepM), and neural network with an ordinal output layer (DeepO). Model performance and calibration is assessed with repeated cross-validation (20 repeats, 5 folds). We also calculate predictor significance (Shapley values) [5].
Results. As shown in **Fig 1**, discrimination of GOSE does not vary significantly with model type based on 95% CIs. Even the most flexible models yield AUCs with only modest discrimination (AUC between 0.61 and 0.66) for intermediate outcomes. MLR, OLR, and DeepO predictions are well-calibrated to true GOSE distributions, but DeepM learns to negatively bias predictions of GOSE 8 to increase sensitivity to intermediate outcomes. Classification performance is consistently poor (mean accuracy ≤ 46%) due to excessive categorisation into GOSE 1 or GOSE 8 effectively reducing a multiclass problem to a dichotomous one. Based on paired Wilcoxon tests for predictor significance ($\alpha = 0.05$) age, motor Glasgow Coma Score, Marshall CT classification, and pupil reactivity are the most significant while the lab values (glucose and Hb) and presence of an epidural haematoma are the least.

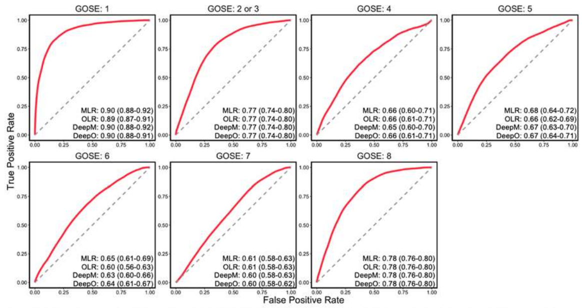


Fig. 1: Receiver operating characteristic (ROC) curves of the best-performing model type per each GOSE score at 6 months post-injury on the validation sets. The values in each box represent the mean area under the curve (with 95% confidence interval in parentheses) per model type.

Conclusion. Since the poor performance of GOSE prediction models is independent of complexity and outcome encoding, it follows that features known to predict dichotomous outcomes are insufficient for multiclass prediction. This result suggests that either admission characterisation is incomplete for intermediate outcomes or these outcomes are better explained by events during ICU stay. This motivates the search for a better classification of TBI that is related to a more nuanced understanding of outcome.

Reference(s)

1. [2] Maas AIR, Menon DK, Steyerberg EW et al. (2014) Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): A Prospective Longitudinal Observational Study. Neurosurgery. https://doi.org/10.1227/NEU.0000000000000575
2. [1] McMillan T, Wilson L, Ponsford J, Levin H, Teasdale G, Bond M (2016) The Glasgow Outcome Scale—40 years of application and refinement. Nature Reviews Neurology. https://doi.org/10.1038/nrneurol.2016.89
3. [5] Lundberg SM, Lee S (2017) A Unified Approach to Interpreting Model Predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems 30:4765–4774. https://doi.org/10.5555/3295222.3295230
4. [4] Steyerberg EW, Mushkudiani N, Perel P et al. (2008) Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. PLOS Medicine. https://doi.org/10.1371/journal.pmed.0050165
5. [3] Doiron D, Marcon Y, Fortier I, Burton P, Ferretti V (2017) Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonization and dissemination. Int J Epidemiol. https://doi.org/10.1093/ije/dyx180

6. S.B. is a recipient of the Gates Cambridge Scholarship. S.B. would like to thank Abhishek Dixit (University of Cambridge) for helping him extract the data.

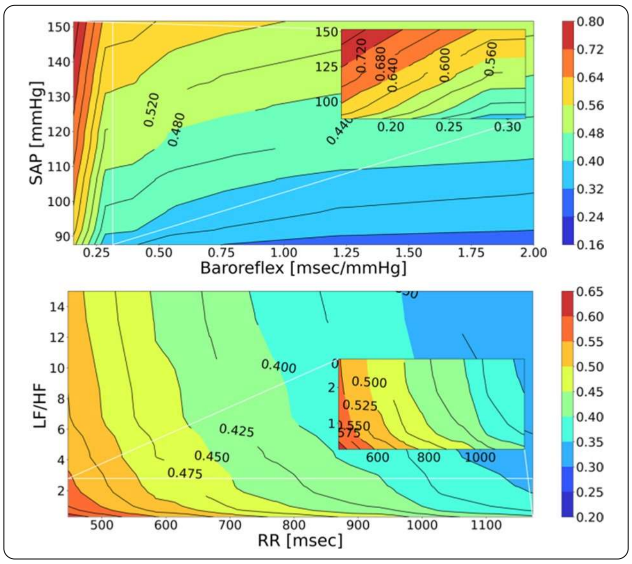
000433

An AI-based cardiovascular monitoring tool for sepsis identification in ICU

M. Mollura[1]; LW. Lehman[2]; R. Barbieri[3]
[1]DEIB, Politecnico di Milano, Milano, Italy; [2]Institute for Medical Engineering & Science (IMES), Massachusetts Institute of Technology, Cambridge, United States of America; [3]Dipartimento di elettronica, informazione e bioingegneria, Politecnico di Milano, Milan, Italy
Correspondence: M. Mollura

Intensive Care Medicine Experimental 2021, **9(1)**: 000433

Introduction. Sepsis is one of the major causes of mortality in ICU with an occurrence up to 40% worldwide [1]. It is considered a serious public health issue with 1 in 3 hospitalizations that end in death with sepsis [2]. Sepsis showed a strong influence on cardiovascular functioning in terms of both myocardial and cardiac autonomic dysfunction [3]. The major effects of sepsis on the cardiovascular system can be summarized as follows: a systolic and diastolic cardiac dysfunction, an increased heart rate despite an overall reduction in autonomic modulation of heart activity and an impairment in the baroreflex sensitivity.
Objectives. Our study is aimed at exploring the ability of ECG and arterial blood pressure (ABP) waveforms, recorded in the first hour of ICU stay, in recognizing patients with sepsis with an AI-based physiological and cardiovascular monitoring tool.
Methods. We extracted the first hour of ECG and ABP waveforms of patients admitted in the ICU, from the publicly available MIMIC-III database on PhysioNet [4,5]. The final population includes 142 patients, 50% of whom with sepsis.
The ECG and ABP signals were processed in order to extract the R-peak occurrences from the ECG and the systolic, diastolic and onset occurrences and values from the pressure signal. 68 features were extracted from the heart rate and blood pressure variability domain through mathematical modelling of the closed loop cardiovascular system, which allowed also for the extraction of baroreflex gain [6]. Finally, 7 confoundings comprising the undergoing sedative and vasoactive agent treatment and mechanical ventilation as well as age, gender, diabetes and hypertension were included.
A logistic regression model was then trained on a 80% training set and tested on the remaining 20% of data.
Results. Best results on the test set show an AUROC = 0.91 and AUPRC = 0.90, thus highlighting the ability of continuously recorded vital signs in recognizing septic patients.



The figure depicts the results of the explainability analysis on model decision rules. The model's strategy results to be consistent with previously summarized clinical considerations about sepsis. Indeed, a lower (impaired) baroreflex gain despite the high systolic pressure values is associated with a higher probability of being septic. Similarly, the second plot shows that an increased heart rate (low RR-interval) is associated with higher sepsis probability and this probability increases when a reduced sympatho-vagal balance (LF/HF) is observed, corresponding to a reduced sympathetic activity or to an overall reduction of autonomic activity.
Conclusion. Our results show that continuously recorded patients' vital signs may help in automatic early identification of sepsis in the early hours of ICU stay. The proposed AI-based physiological monitoring also allows for the extraction of indices able to describe the state of the cardiovascular control system and providing key insights on cardiovascular functioning.

Reference(s)

1. [1]: Sakr, Y., Jaschinski, U., Wittebole, X., Szakmany, T., Lipman, J., Ñamendys-Silva, S. A., Martin-Loeches, I., Leone, M., Lupu, M.-N., & Vincent, J.-L. (2018). Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit. Open Forum Infectious Diseases, 5(12). https://doi.org/10.1093/ofid/ofy31
2. [2]: Fay K, Sapiano MRP, Gokhale R, et al. Assessment of Health Care Exposures and Outcomes in Adult Patients With Sepsis and Septic Shock. JAMA Netw Open. 2020; 3(7): e206004. https://doi.org/10.1001/jamanetworkopen.2020.6004
3. [4]: Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3(1). https://doi.org/10.1038/sdata.2016.35
4. [5]: Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation, 101(23), E215–E220. https://doi.org/10.1161/01.cir.101.23.e215
5. [6]: Chen, Z., Purdon, P., Pierce, E., Harrell, G., Brown, E., & Barbieri, R. (2009). Assessment of Baroreflex Control of Heart Rate During General Anesthesia Using a Point Process Method. Proceedings of the … IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP (Conference), 2009, 333–336. https://doi.org/10.1109/ICASSP.2009.4959588
6. [3]: Wang H. (2019). Cardiac Autonomic Nervous System and Sepsis-Induced Cardiac Dysfunction. In: Fu X., Liu L. (eds.) Severe Trauma and Sepsis. Springer, Singapore. https://doi.org/10.1007/978-981-13-3353-8_6

000450

Decoding digital health signatures for prediction of delirium in the intensive care unit

K. Gong,[1]; R. Lu[1]; T. Bergamaschi[1]; A. Sanyal[1]; J. Guo[1]; H. Kim[2]; R. Stevens[2]
[1]Biomedical Engineering, Johns Hopkins University, Baltimore, United States of America; [2]Anesthesiology and Critical Care Medicine, Johns Hopkins University, Baltimore, United States of America
Correspondence: R. Stevens
Intensive Care Medicine Experimental 2021, **9(1)**: 000450

Introduction. ICU delirium is frequent, associated with unfavorable outcomes, increased costs, and may be preventable. There is an unmet need for accurate methods to predict risk of delirium. Here, we trained machine learning models to predict ICU delirium using physiological and clinical features in large clinical database.
Objectives. The aim of this study was to leverage machine learning applied to early physiological and clinical data to predict delirium onset using physiological and clinical features in large clinical database.
Methods. The first model (A) was designed to predict delirium onset at any time during the ICU stay. Data from the first 24 h following ICU admission were extracted from the multi-center eICU database. A second "dynamic" model (B) was built to predict onset of delirium in the next hour to twelve hours. Features extracted from data included demographics, medical history, labs, medications, nurse charting, comorbidities, treatments, and physiological time series. Similar features were extracted from the MIMIC-III database for external validation. The top features were analyzed using logistic regression (LR), random forest (RF), or Shapley values, and then a statistically pruned feature space for each model was obtained. These pruned feature spaces included about 150 variables, with slight variations depending on the model, lead time, and data window used. For both models, three algorithms were evaluated: LR, RF, and gradient boosting [CatBoost]). Model performance was evaluated using nested cross-validation and area under the receiver operating characteristic curve