

UCSF

UC San Francisco Previously Published Works

Title

IntraCranial pressure prediction AlgoRithm using machinE learning (I-CARE): Training and Validation Study.

Permalink

<https://escholarship.org/uc/item/19c3158v>

Journal

Critical Care Explorations, 6(1)

Authors

Fong, Nicholas

Feng, Jean

Hubbard, Alan

et al.

Publication Date

2024

DOI

10.1097/CCE.0000000000001024

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

IntraCranial pressure prediction AlgoRithm using machinE learning (I-CARE): Training and Validation Study

OBJECTIVES: Elevated intracranial pressure (ICP) is a potentially devastating complication of neurologic injury. Developing an ICP prediction algorithm to help the clinician adjust treatments and potentially prevent elevated ICP episodes.

DESIGN: Retrospective study.

SETTING: Three hundred thirty-five ICUs at 208 hospitals in the United States.

SUBJECTS: Adults patients from the electronic ICU (eICU) Collaborative Research Database was used to train an ensemble machine learning model to predict the ICP 30 minutes in the future. Predictive performance was evaluated using a left-out test dataset and externally evaluated on the Medical Information Mart for Intensive Care-III (MIMIC-III) Matched Waveform Database.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: Predictors included age, assigned sex, laboratories, medications and infusions, input/output, Glasgow Coma Scale (GCS) components, and time-series vitals (heart rate, ICP, mean arterial pressure, respiratory rate, and temperature). Each patient ICU stay was divided into successive 95-minute timeblocks. For each timeblock, the model was trained on nontime-varying covariates as well as on 12 observations of time-varying covariates at 5-minute intervals and asked to predict the 5-minute median ICP 30 minutes after the last observed ICP value. Data from 931 patients with ICP monitoring in the eICU dataset were extracted (46,207 timeblocks). The root mean squared error was 4.51 mm Hg in the eICU test set and 3.56 mm Hg in the MIMIC-III dataset. The most important variables driving ICP prediction were previous ICP history, patients' temperature, weight, serum creatinine, age, GCS, and hemodynamic parameters.

CONCLUSIONS: IntraCranial pressure prediction AlgoRithm using machinE learning, an ensemble machine learning model, trained to predict the ICP of a patient 30 minutes in the future based on baseline characteristics and vitals data from the past hour showed promising predictive performance including in an external validation dataset.

KEYWORDS: artificial intelligence; brain injury; intracranial pressure; machine learning; prediction

Nicholas Fong^{1,2}

Jean Feng, PhD³

Alan Hubbard, PhD⁴

Lauren Eyler Dang, MD⁴

Romain Pirracchio, MD, MPH,
PhD^{1,3,4}

Elevated intracranial pressure (ICP) is a potentially devastating complication of neurologic injury. The 2016 guidelines (1) for the management of patients with severe traumatic brain injury recommend using ICP monitoring to reduce in-hospital and 2-week post-injury mortality (level IIB). The same guidelines recommend treating ICP greater than 22 mm Hg because values above this level are associated with increased mortality (level IIB).

Unfortunately, the treatments available to decrease the ICP and optimize the cerebral perfusion pressure take time to be effective. The intensity and

Copyright © 2023 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/CCE.0000000000001024



KEY POINTS

Question: Given historical intracranial pressure (ICP) values, vitals, and laboratory values, can machine learning be used to predict the ICP value 30 minutes in the future?

Findings: The ICP value can be predicted 30 minutes in the future with encouraging predictive performance.

Meanings: Clinicians may be able to proactively adjust treatments and interventions to potentially prevent intracranial hypertension episodes.

duration of episodes of intracranial hypertension (intracranial hypertension dose) was found to be independently associated with mortality and long-term functional outcome in severe brain injuries from different origins such as traumatic brain injury (2) or subarachnoid hemorrhage (3). Hence, successful management of patients with elevated ICP requires early recognition and therapy directed at both reducing ICP and reversing its underlying cause. Furthermore, predicting the evolution of ICP could help the clinician to proactively adjust treatments and interventions to potentially prevent intracranial hypertension.

We have previously demonstrated that machine learning can be used to accurately predict the evolution of physiologic parameters (4, 5) using supervised ensemble machine learning methods that were proven to be superior to any single machine learning approaches in many situations (5). The goal of the present study is to use an ensemble learning approach to train and validate IntraCranial pressure prediction AlgoRithm using machinE learning (I-CARE), an ICP prediction algorithm to predict the ICP value 30 minutes in the future in patients hospitalized in the ICU with an acute brain injury and an ICP monitor.

METHODS

This study was based on retrospective data and is reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines (6).

Data Source

Two separate data sources were used for this study. The first one, electronic ICU (eICU) Collaborative Research Database, was used to train the model. The second one, the Medical Information Mart for Intensive Care-III (MIMIC-III) Matched Waveform Database, was used to externally validate the performance of the algorithm. There is no overlap between the two databases.

The eICU Collaborative Research Database is a multicenter publicly and freely accessible ICU database with high granularity data for over 200,000 admissions to ICUs monitored by eICU programs, a telehealth system developed by Philips Healthcare (Cambridge, MA) to support management of critically ill patients across the United States (7). Data in eICU are deidentified to meet the safe harbor provision of the U.S. HIPAA. Data in eICU were generated from over 130,000 unique patients admitted between 2014 and 2015 to one of 335 units at 208 hospitals in the United States. The deidentified data are publicly available after registration, including completion of a training course in research with human subjects and signing of a data use agreement mandating responsible handling of the data and adhering to the principle of collaborative research.

MIMIC-III (8) is a publicly and freely available database associating medico-administrative data, physiologic measurements and treatment administration prospectively and consecutively collected at the bedside between 2001 and 2012 from five ICUs in Boston's Beth Israel Deaconess Medical Center. Data were de-identified; data collection was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and Massachusetts Institute of Technology (Cambridge, MA). We used the subset of patients in the MIMIC-III dataset that also have ICP data in the MIMIC-III Waveform Database Matched Subset (9) for the validation set.

Participants

From the eICU database, we selected all adult (≥ 18 yr old) patients with recorded ICP monitoring for at least 4 consecutive hours.

Outcome

The outcome was defined as the median ICP measurement within a 5-minute interval 30 minutes after the last observed ICP value (**Fig. 1**).

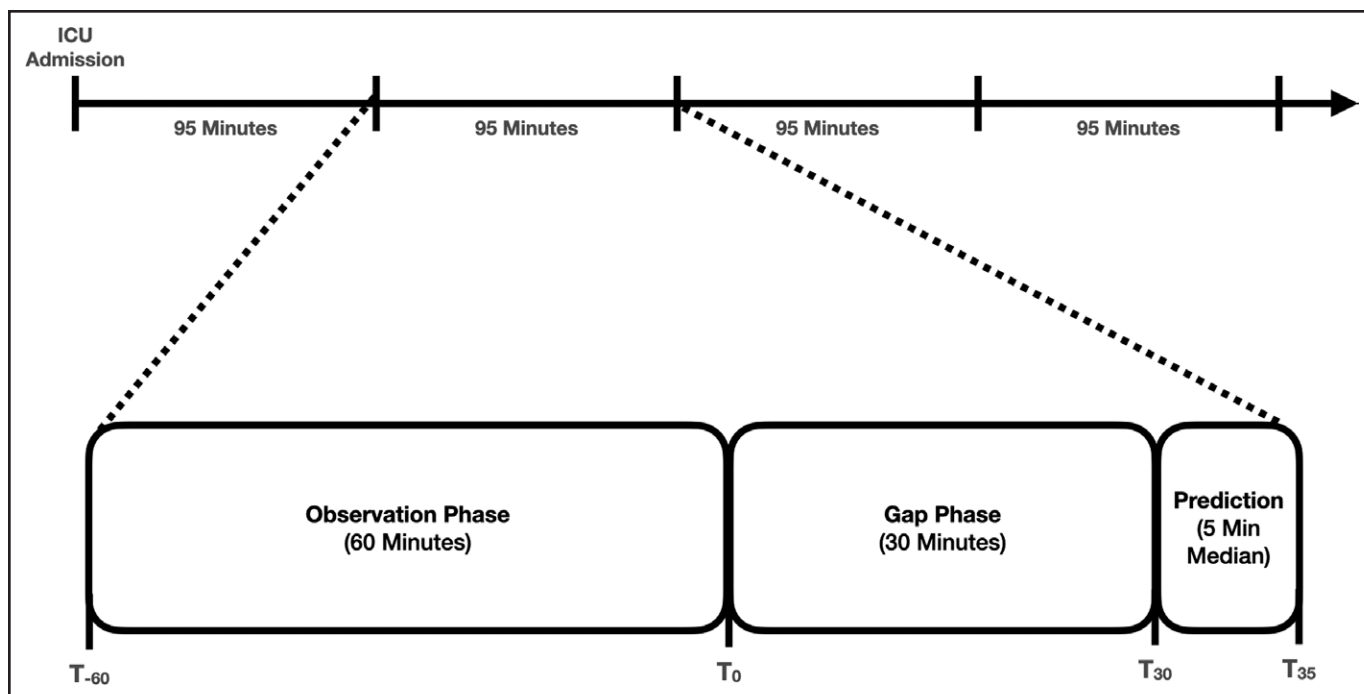


Figure 1. Illustration of the composition of the definition of a timeblock. For each patient, their time in the ICU is divided into 95-min blocks. Each 95-min block is subdivided into a 60-min observation period, a 30-min gap period, and a 5-min prediction period. The model is given data from the observation period and is asked to predict the median intracranial pressure during the 5-min prediction period.

Study Periods

Each included ICU admission in the eICU dataset was divided into successive 95-minute timeblocks. Each 95-minute period was divided into three consecutive time windows of 60 (“observation window” to define predictors), 30 (“gap window”), and 5 minutes (“prediction window”; Fig. 1). The 30-minute “gap window” was motivated by the fact that ICP prediction is helpful only if the predicted value is far enough into the future that there is sufficient time for therapeutic adjustment. In the eICU dataset, vital signs were internally collected at 1-minute interval, and 5-minute medians were archived in the dataset. Consequently, there are 12 time-varying measurements in each observation window, six measurements in the gap window, and one measurement in the observation window.

Predictors

Predictor selection was based on clinical expertise and data availability. Variables with more than 20% missing values were not considered as predictors. Predictors in the model included baseline demographics (age, assigned sex), reason for ICU admission, laboratories

(arterial blood gases, sodium, creatinine, hematocrit, hemoglobin, platelets, glucose, fibrinogen, and international normalized ratio), medications and infusions (sedatives, vasopressors, hypertonic solutions, benzodiazepines, neuromuscular blockers, and opioids), input/output, Glasgow Coma Scale (GCS) components, and time-series vitals (heart rate, ICP, mean arterial pressure [MAP], respiratory rate, and temperature) from the observation window only. For each timeblock, the model was trained on nontime-varying covariates as well as on the 12 values of the time-varying covariates at 5-minute intervals and asked to predict the 5-minute median ICP 30 minutes after the last observed ICP value. In the MIMIC-III Waveform Database Matched Subset dataset, vital signs were collected at 1-minute intervals with all data archived in the dataset, so data was post-processed to take the 5-minute median to match the eICU data format.

Sample Size

The eICU database includes 200,859 ICU encounters for 139,367 unique patients admitted between 2014 and 2015. From 931 ICU encounters that met inclusion criteria, 46,207 timeblocks were extracted from the database; 698 patients (75%, 35,128 timeblocks) were

randomized to the training set and 233 patients (25%, 11,079 timeblocks) were randomized to the test set. To avoid any risk of data leakage, the same patient could not contribute time periods to both training and testing sets. The MIMIC III Waveform Database Matched Subset Version 1.0 contains 22,317 high frequency waveform and 22,247 numeric records corresponding to ICU stays from 10,282 patients also included in the MIMIC III Clinical Database.

Missing Data and Outliers Filtering

ICP values that were less than 0 or greater than the MAP were considered invalid and consequently treated as missing. Timeblocks with three or more missing ICP measurements (i.e. three values missing from the source dataset, implying missing ICP data for at least three 5-min windows) during the observation window or missing ICP measurement during the prediction window were excluded from the analysis. For blocks with fewer than three missing ICP measurements during the observation window, the missing ICP values were imputed with the median ICP measurement during the observation window. Additionally, a missing variable indicator was included in the model for each ICP measurement to indicate if the ICP measurement was imputed. **eFigure 1** (<http://links.lww.com/CCX/B285>) depicts the proportion of eICU timeblocks contributed by each institution and the number of timeblocks with one or two missing ICP values during the observation period by institution. Missing non-ICP vital signs were imputed by forward filling data when previous data in the observation window was available. When a previous datapoint from a given time window was not available, missing non-ICP vital signs were imputed by taking the median of nonmissing data in the observation window. The other missing variables were imputed by taking the median value for the patient or the median value for the dataset if the patient had no valid data. The proportion of missing data in observation windows is available in **eTable 1** (<http://links.lww.com/CCX/B285>).

Statistical Analysis

Algorithm. The model used in this study is a supervised ensemble machine learning algorithm called Super Learner (10). The Super Learner is a method for selecting

via cross-validation the optimal regression algorithm among all weighted combinations of a set of given candidate algorithms, henceforth referred to as the library. Thus, the Super Learner algorithm requires the user to input a library. Theoretical results suggest that to optimize the performance of the resulting algorithm, the inputted library should include as many sensible algorithms as possible. In this study, the library included ten algorithms (**eTable 2**, <http://links.lww.com/CCX/B285>). Comparison of the algorithms relied on ten-fold cross-validation. In this process, the data are first split into ten mutually exclusive and exhaustive blocks of approximately equal size. One of these blocks, the validation set, is excluded, and all remaining data, referred to as the training set, are used to fit each of the algorithms. Each fitted algorithm is used to predict the ICP for all patients in the validation set and the squared errors between predicted and observed outcomes are averaged. The performance of each algorithm is evaluated in this manner. This procedure is repeated exactly 10 times, with a different block used as validation set every time. Performance measures are aggregated over all 10 iterations, yielding a cross-validated estimate of the mean-squared error (CV-MSE) for each algorithm. A crucial aspect of this approach is that for each iteration not a single patient appears in both the training and validation sets. The potential for overfitting, wherein the fit of an algorithm is overly tailored to the available data at the expense of performance on future data, is thereby mitigated, as overfitting is more likely to occur when training and validation sets intersect. Candidate algorithms are ranked according to their CV-MSE and the algorithm with least CV-MSE was identified. This algorithm was then refitted using all available data, leading to a prediction rule referred to as the Discrete Super Learner. Subsequently, the prediction rule consisting of the CV-MSE-minimizing weighted convex combination of all candidate algorithms was also computed and refitted on all data. The resulting algorithm is referred to as the SuperLearner algorithm.

Model Performance. Following recommendations from “Guidelines for developing and reporting machine learning predictive models in biomedical research” (11), we randomly divided the eICU patients into a training set (75% of patients) and an internal validation dataset (the remaining 25% of the patients), and used the MIMIC-III cohort for external validation.

Calibration. Model calibration was graphically assessed by plotting the predicted vs. observed ICP

for the internal and the external validation dataset. Calibration was also illustrated using Bland-Altman plots, accounting for repeated measures (12). The root mean squared error (RMSE), bias, and limits of agreement were computed and reported for the internal and the external validation dataset.

Thresholds. To assess the ability of I-CARE to detect significant changes in the ICP, performance was also specifically studied in a subset of the test and the validation sets where the actual ICP value during the prediction window increased by at least 20% of the mean ICP from the observation window. In addition to predicting ICP, we also want to detect clinically significant changes in the ICP. Therefore, we looked at the accuracy of the I-CARE algorithm in predicting significant increases in the ICP. Specifically, we assessed the area under the receiver operating characteristic curve, sensitivity, specificity, accuracy, positive and negative predictive values, and positive and negative likelihood ratios for: 1) the detection of an ICP increase in the next 30 minutes of more than 10%, 20%, and 30% compared with the mean ICP value during the observation window and 2) the detection of an episode of intracranial hypertension as defined by a median ICP greater than 15, 20, and 22 mm Hg during the prediction window.

Model Interpretability. The contribution of each predictor was quantified by computing the SHapley Additive exPlanations (SHAP) framework (13). SHAP is a game theoretic approach that quantifies the average expected marginal contribution of one predictor after all possible combinations have been considered. Using the Interpretable Machine Learning package; <https://CRAN.Rproject.org/package=iml> (14), Shapley values were generated for each prediction in the validation set using 100 Monte Carlo simulations. The Shapley values provide insights into the relative importance of each predictor variable. By quantifying the average expected marginal contribution of a specific predictor after considering all possible combinations, the Shapley values enable discernment of the exact influence of individual variables on the likelihood of intracranial hypertension according to the model.

All analyses were performed using statistical software R Version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). The I-CARE Super Learner algorithm was trained using SuperLearner R package, Version 2.0-24 (R Foundation for Statistical

Computing, Vienna, Austria). The I-CARE model will be made available upon reasonable request.

RESULTS

Participants

The distribution of ICP values in the eICU dataset before applying exclusion criteria is shown in **eFigure 2** (<http://links.lww.com/CCX/B285>). Nine hundred thirty-one ICU admissions from the eICU dataset met our inclusion criteria, for a total of 46,207 timeblocks: 35,128 timeblocks (corresponding to 698 patients) were included in the training set, 11,079 timeblocks (corresponding to 233 patients) were included in the test set (**Fig. 2**). Six thousand eight hundred thirty-five timeblocks from 127 patients from the MIMIC-III dataset were used for external validation. Patient characteristics are provided in **eTable 3** (<http://links.lww.com/CCX/B285>). The median (Q1–Q3) across timeblocks of the average ICP during the observation period was 9.83 mm Hg (6.42–14.0 mm Hg), 9.75 mm Hg (6.42–13.7 mm Hg), and 9.25 mm Hg (6.63–12.2 mm Hg) in the training set, test set, and external validation set, respectively (**eTable 4**, <http://links.lww.com/CCX/B285>). During the prediction phase, the observed median (Q1–Q3) ICP was 10.0 mm Hg (6.00–14.0 mm Hg), 10.0 mm Hg (6.00–14.0 mm Hg), and 9.00 mm Hg (6.00–12.0 mm Hg) in the training set, test set, and external validation set, respectively. The number of timeblocks where the actual ICP value during the prediction window increased by at least 20% of the mean ICP from the observation window was 8018 (22.8%), 2648 (23.9%), and 1348 (19.7%) in the training set, test set, and external validation set, respectively.

Model Performance

Model calibration is illustrated in **Figure 3**. The RMSE in the test set was 4.51 mm Hg. As illustrated in the Bland-Altman plots (**Fig. 3, B and D**), systematic bias in the test set was –1.15 mm Hg with limits of agreements of –9.92 and 7.63 mm Hg. When the model was evaluated on 6835 timeblocks from the external dataset, the RMSE was 3.56 mm Hg, the systematic bias was 0.657 mm Hg with limits of agreements of –6.52 and 7.83 mm Hg. A subset of the test and validation sets where the actual ICP value during the prediction

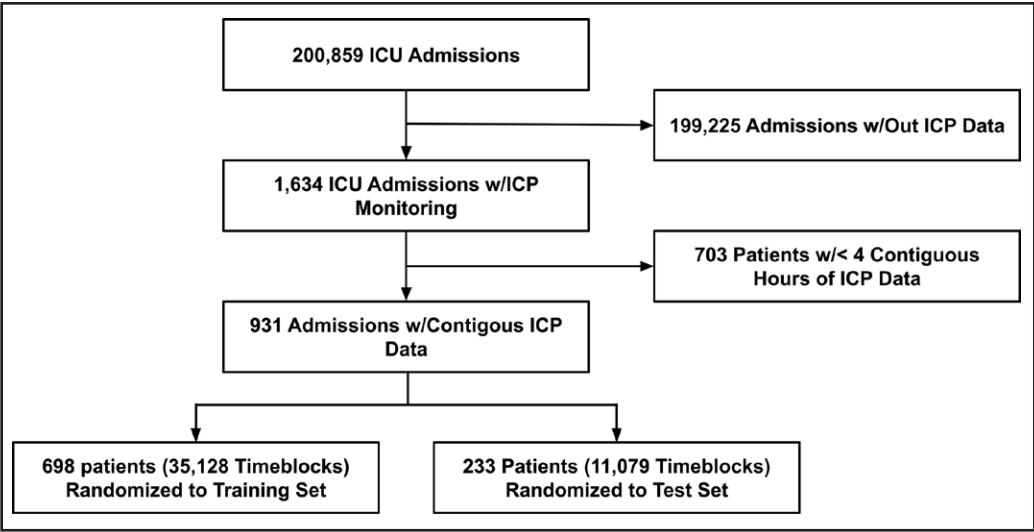


Figure 2. Flowchart depicting how timeblocks from the electronic ICU dataset were generated from the database. ICP = intracranial pressure, W = with.

window increased by at least 20% of the mean ICP from the observation window was additionally examined (eFigs. 3 and 4, <http://links.lww.com/CCX/B285>). The RMSE was 7.38 and 5.60 mm Hg in this subset of the test and the validation set, respectively. I-CARE’s performance to detect an episode of intracranial hypertension as defined by a median ICP greater than 15, 20, and 22 mm Hg during the prediction window is provided in **Table 1**. In the external validation dataset, I-CARE was able to predict these episodes with an accuracy of 92%, 97%, and 98%, and a positive likelihood ratio of 22.06, 69.85, and 104.48, respectively. We also evaluated the performance of I-CARE to detect an ICP increase in the next 30 minutes of more than 10%, 20%, and 30% of baseline, defined as the mean ICP during the observation window. In the external validation dataset, I-CARE was able to predict these increases with 64%, 73%, and 80% accuracy, with a positive likelihood ratio of 2.49, 3.84, and 6.30, respectively (Table 1).

Feature Importance

Plots depicting Shapley values are provided in **Figure 4**; and **eFigures 5 and 6** (<http://links.lww.com/CCX/B285>). As illustrated in Figure 4, the most important variable driving I-CARE’s prediction of the ICP 30 minutes in the future is previous ICP history. Additionally, patients’ temperature, weight, serum creatinine, age, GCS, and hemodynamic parameters were identified as important predictors. eFigure 6 (<http://links.lww.com/CCX/B285>) illustrates the breakdown of the top time-varying

predictors broken down by relative time in the observation window.

DISCUSSION

In this study, the I-CARE algorithm was trained to predict the ICP 30 minutes in the future. In an independent external dataset, I-CARE was able to predict the ICP with a RMSE of less than 4 mm Hg. In addition, I-CARE was shown to have encouraging predictive performance for

the detection of acute changes in the ICP in the next 30 minutes.

There is a growing interest in the use of machine learning techniques to predict the evolution of important physiologic parameters, such as the MAP in critically ill patients (4, 5, 15). However, only few studies have described the use of machine learning to predict the ICP in patients with a severe brain injury. Most studies thus far were performed in pediatric patients (16). Available studies in adult patients present some limitations (17–20). The algorithm developed by Güiza et al (20) was trained in a relatively small population of 178 neurocritical care patients, only 61% of which presenting an episode of elevated ICP. The model by Güiza et al (20) was later externally validated (21), confirming the model’s ability to detect episodes of increased ICP in traumatic brain injury patients (22). However, in this study, as in many others, the algorithm detects intracranial hypertension as defined based on a single ICP threshold. The definition and the clinical meaningfulness of intracranial hypertension depend on the clinical context and the patient. Training an algorithm to predict the actual ICP value rather than intracranial hypertension as a binary outcome gives the clinician the opportunity to tailor the threshold for concern and treatment plan instead of applying a one-size-fits-all treatment strategy. The same team has recently published a new algorithm that predicts the intracranial hypertension dose (23). Although interesting, this approach suffers the same limitation in terms of relying on a somewhat arbitrary binary definition of the outcome of interest.

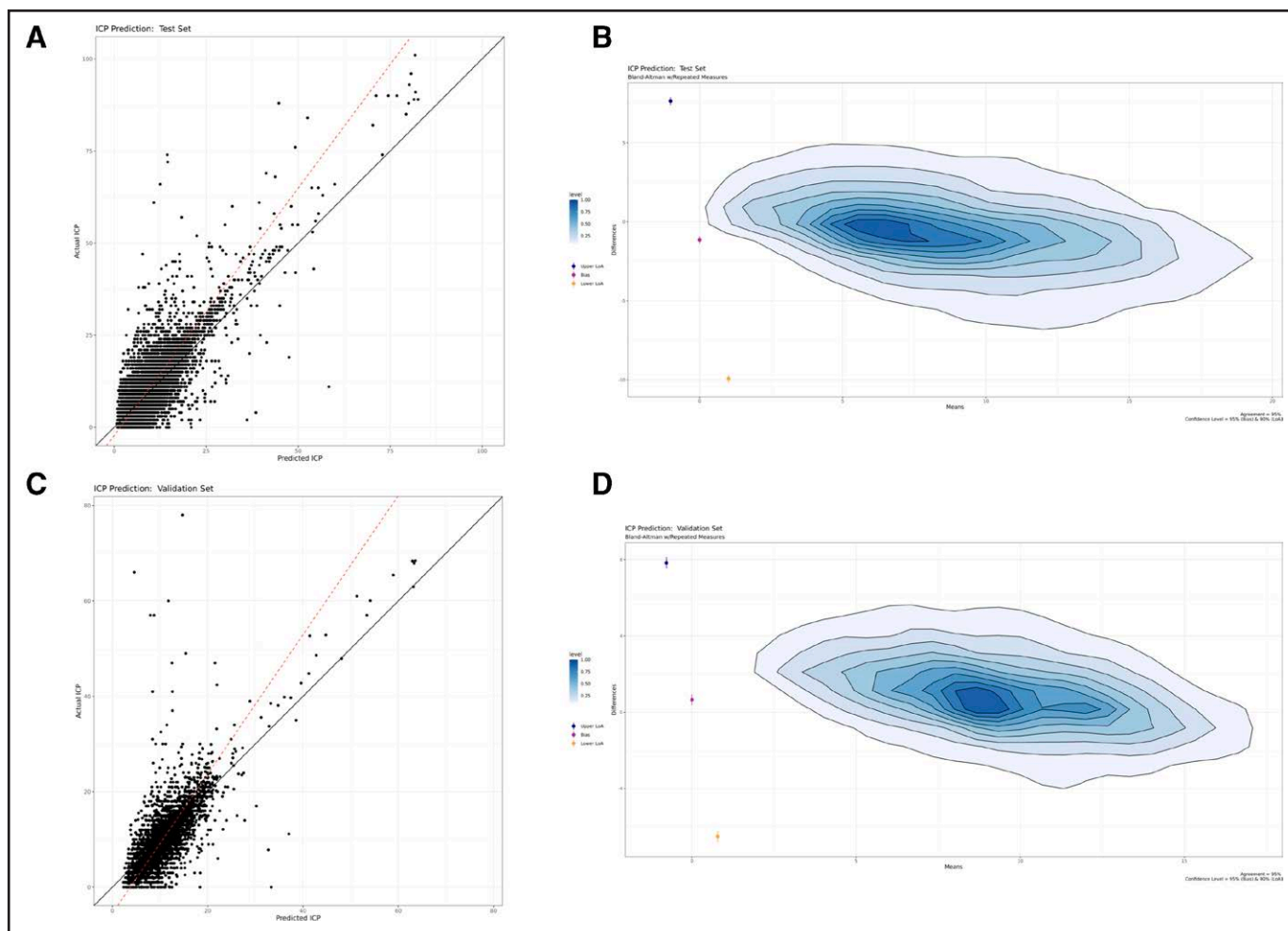


Figure 3. Calibration and Bland-Altman plots for test and external validation datasets. **A**, Test set calibration plot; **B** test set Bland-Altman plot; **C** external validation set calibration plot; **D** external validation set Bland-Altman plot. **A** and **C**, The x-axis represents the predicted intracranial pressure (ICP), the y-axis the actual ICP. Color represents the patient for which a prediction was made. **B** and **D**, The x-axis represents the median ICP, the y-axis the difference between the predicted and the observed ICP. LoA = limits of agreement, W = with.

Several studies have used machine learning to predict the ICP in patients with no ICP monitor (24–26). This is potentially useful at the early stage of brain injury management before the insertion of the ICP monitor or in settings where ICP monitoring is not available. I-CARE focuses on ICP prediction 30 minutes in the future in patients already equipped with an ICP monitoring device. This time gap of 30 minutes was chosen to give enough time to the clinician to adjust interventions in response to the predicted ICP and potentially avoid intracranial hypertension episodes. By doing so, clinicians may be able to decrease the intracranial hypertension dose, which has been shown to be associated with increased mortality and poor long-term functional outcomes (2, 3).

Unsurprisingly, previous ICP values were found to be the most important predictors of future ICP values

(eFig. 6, <http://links.lww.com/CCX/B285>). This is consistent with other ICP prediction models (23). This finding reflects clinical practice in that without knowledge of baseline ICP information, prediction of future ICP is improbable. Interestingly though, other clinical parameters such as age, GCS, weight, temperature, hemodynamic status (MAP and heart rate), and serum creatinine were found to also play some role in the prediction.

This study has some limitations. First, patients with different types of brain injuries were pooled together. The physiology driving the evolution of the ICP may differ between injuries, and even better performance would be expected if the algorithm were trained on a more homogenous patient population. However, training I-CARE on a variety of brain

TABLE 1.
Model Performance for Intracranial Hypertension Prediction Using Different Intracranial Pressure Thresholds in the Validation Set

Measurement	Relative Increase			Specific Intracranial Pressure Threshold		
	10% Increase From Baseline	20% Increase From Baseline	30% Increase From Baseline	15 mm Hg Hypertension	20 mm Hg Hypertension	22 mm Hg Hypertension
Accuracy	0.64 (0.63–0.65)	0.73 (0.72–0.74)	0.80 (0.79–0.81)	0.92 (0.91–0.92)	0.97 (0.97–0.98)	0.98 (0.98–0.98)
Prevalence	0.30	0.20	0.14	0.13	0.034	0.022
Positive likelihood ratio	1.74 (1.66–1.84)	2.38 (2.22–2.55)	3.22 (2.95–3.53)	20.63 (17.50–24.31)	94.37 (61.56–144.65)	104.48 (59.48–183.54)
Negative likelihood ratio	0.60 (0.57–0.64)	0.60 (0.57–0.64)	0.61 (0.58–0.65)	0.46 (0.43–0.50)	0.64 (0.59–0.71)	0.75 (0.69–0.83)
Sensitivity	0.61 (0.59–0.63)	0.53 (0.50–0.56)	0.48 (0.44–0.51)	0.55 (0.52–0.58)	0.36 (0.30–0.42)	0.25 (0.18–0.33)
Specificity	0.65 (0.64–0.67)	0.78 (0.77–0.79)	0.85 (0.84–0.86)	0.97 (0.97–0.98)	1.00 (0.99–1.00)	1.00 (1.00–1.00)
Positive predictive value	0.43 (0.41–0.45)	0.37 (0.35–0.39)	0.35 (0.32–0.38)	0.76 (0.72–0.79)	0.77 (0.68–0.85)	0.70 (0.56–0.82)
Negative predictive value	0.79 (0.78–0.81)	0.87 (0.86–0.88)	0.91 (0.90–0.91)	0.94 (0.93–0.94)	0.98 (0.97–0.98)	0.98 (0.98–0.99)
False positive rate	0.35 (0.33–0.36)	0.22 (0.21–0.23)	0.15 (0.14–0.16)	0.03 (0.02–0.03)	0.00 (0.00–0.01)	0.00 (0.00–0.00)
False negative rate ^a	0.39 (0.37–0.41)	0.47 (0.44–0.50)	0.52 (0.49–0.56)	0.45 (0.42–0.48)	0.64 (0.58–0.70)	0.75 (0.67–0.82)
Area under the receiver operating characteristic curve	0.63 (0.62–0.64)	0.65 (0.64–0.67)	0.66 (0.65–0.68)	0.76 (0.75–0.78)	0.68 (0.65–0.71)	0.62 (0.59–0.66)

^aWe additionally conducted a sensitivity analysis evaluating the intracranial hypertension episode prediction performance of IntraCranial pressure prediction Algorithm using machineE learning by examining the model's ability to predict an intracranial pressure (ICP) that is within 2 mm Hg of the threshold (e.g., predicting an ICP ≥ 18 mm Hg in the 20 mm Hg threshold). We found that the false negative rate drops by around 20%. Specifically, the 15 mm Hg false negative rate is 0.22 in the sensitivity analysis vs. 0.45 in this table, 20 mm Hg is 0.43 vs. 0.64, and 22 mm Hg is 0.57 vs. 0.75.

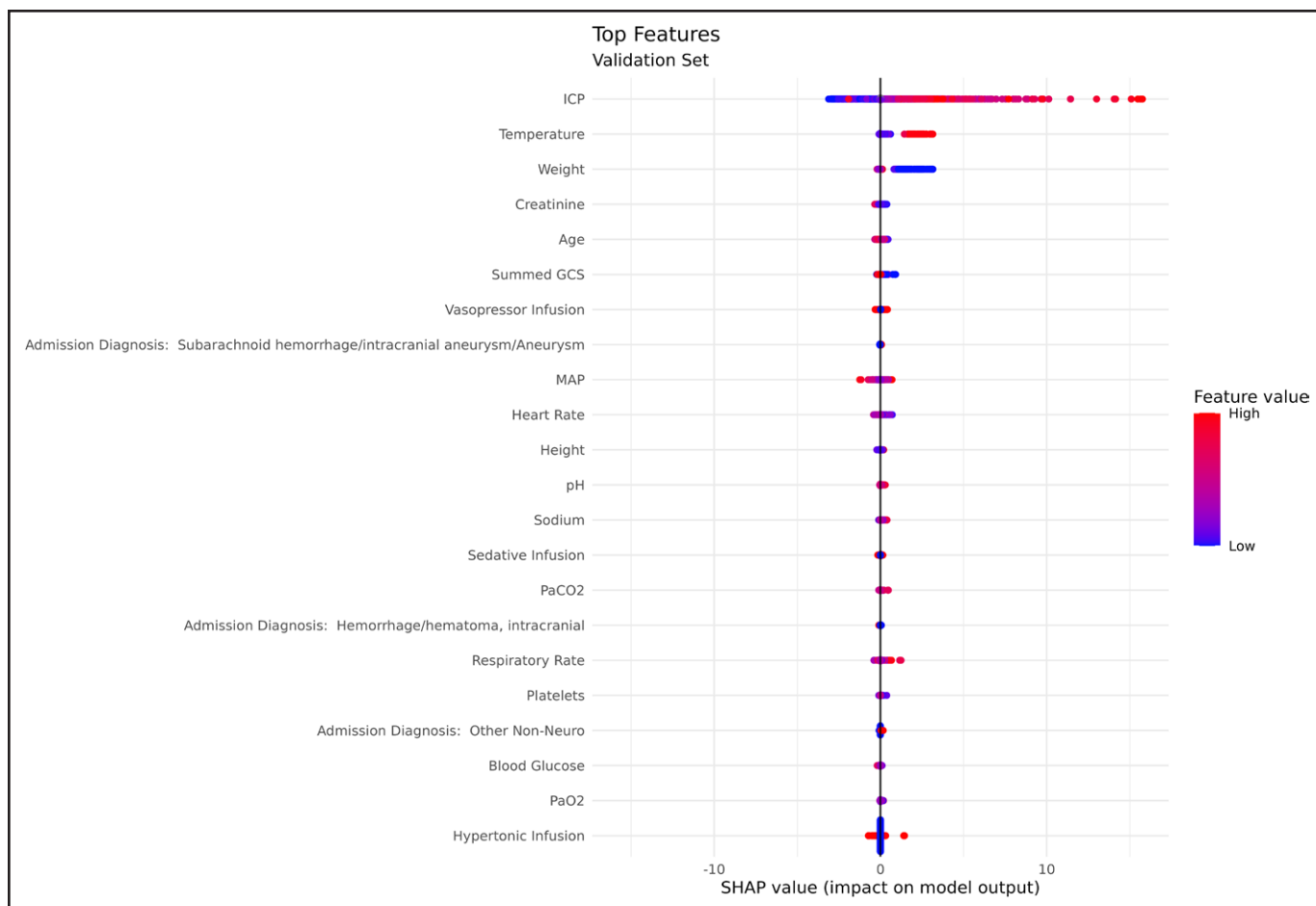


Figure 4. Feature importance for the top predictors based on Shapley values. Color of each SHapley Additive exPlanations (SHAP) value corresponds to the relative value of the feature; high values are indicated by red and low values are indicated by blue. Time-varying predictors are grouped in this figure; predictors are stratified by time in the observation window in eFigure 6 (<http://links.lww.com/CCX/B285>). GCS = Glasgow Coma Scale, ICP = intracranial pressure, MAP = mean arterial pressure.

injured patients increases the generalizability of our results; additionally, our feature importance analysis identified that the etiology of brain injury could have a minimal but non-null effect on the predicted ICP value. Second, although external validation was performed using a completely independent dataset, a prospective validation in real-life conditions is still yet to be performed. This prospective validation in real-life conditions is planned for a follow-up study. Third, the ICP of patients included in the training, test, and external validation sets were relatively low. But more than 20% of the analyzed timeblocks had a greater than 20% increase in the ICP between the observation and the prediction window. Fourth, although I-CARE uses several time-evolving variables as predictors, including vital signs, we did not use high-fidelity waveform signals in this first version of the algorithm. An updated version of the algorithm,

trained with vital signs of higher granularity, will be released in the future. Fifth, while I-CARE's false positive rate was minimal (Table 1), limiting the risk of overtreatment, the false negative rate may appear substantial, leading to a potential risk of undertreatment. However, the primary goal of I-CARE is to predict the continuous ICP and not whether an ICP threshold will be reached. Hence, the algorithm was not training to optimize classification, but rather to minimize the error in predicting the actual continuous ICP value. Thus, I-CARE should be used to predict a trend in the ICP and not as classifier for elevated ICP.

I-CARE is the first intracranial prediction algorithm allowing to accurately predict the ICP value 30 minutes in the future using advanced machine learning, trained on a large sample of neurocritical care patients, with external validation. More work is still needed to prospectively

validate the use of I-CARE in practice and determine the impact of treatment strategies to prevent the occurrence of ICP in patients with severe brain injury.

- 1 Department of Anesthesia and Perioperative Medicine, Zuckerberg San Francisco General Hospital and Trauma Center, University of California San Francisco, San Francisco, CA.
- 2 School of Medicine, University of California San Francisco, San Francisco, CA.
- 3 Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA.
- 4 Division of Biostatistics, School of Public Health, University of California Berkeley, Berkeley, CA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: Romain.Pirracchio@ucsf.edu

REFERENCES

1. Carney N, Totten AM, O'Reilly C, et al: Guidelines for the management of severe traumatic brain injury, fourth edition. *Neurosurgery* 2017; 80:6–15
2. Sheth KN, Stein DM, Aarabi B, et al: Intracranial pressure dose and outcome in traumatic brain injury. *Neurocrit Care* 2013; 18:26–32
3. Carra G, Elli F, Ianosi B, et al: Association of dose of intracranial hypertension with outcome in subarachnoid hemorrhage. *Neurocrit Care* 2021; 34:722–730
4. Cherifa M, Interian Y, Blet A, et al: The physiological deep learner: First application of multitask deep learning to predict hypotension in critically ill patients. *Artif Intell Med* 2021; 118:102118
5. Cherifa M, Blet A, Chambaz A, et al: Prediction of an acute hypotensive episode during an ICU hospitalization with a super learner machine-learning algorithm. *Anesth Analg* 2020; 130:1157–1166
6. Collins GS, Reitsma JB, Altman DG, et al: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med* 2015. doi: 10.1186/s12916-014-0241-z.
7. Pollard TJ, Johnson AEW, Raffa JD, et al: The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data* 2018; 5:180178
8. Johnson AEW, Pollard TJ, Shen L, et al: MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3:160035
9. Moody B, Moody G, Villarroel M, Clifford GD, Silva I. MIMIC-III Waveform Database Matched Subset (version 1.0). *Physio Net* 2020. <https://doi.org/10.13026/c2294b>. Accessed October 29, 2021
10. van der Laan MJ, Polley EC, Hubbard AE: Super learner. *Stat Appl Genet Mol Biol* 2007; 6:Article25
11. Luo W, Phung D, Tran T, et al: Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res* 2016; 18:e323
12. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–310
13. Lundberg S, Lee S-I: A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017
14. Molnar C, Schratz P: iml: Interpretable Machine Learning. R package version 0.11.1. 2022. Available at: <https://CRAN.R-project.org/package=iml>. Accessed January 18, 2022
15. Hatib F, Jian Z, Buddi S, et al: Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology* 2018; 129:663–674
16. Imaduddin SM, Fanelli A, Vonberg FW, et al: Pseudo-Bayesian model-based noninvasive intracranial pressure estimation and tracking. *IEEE Trans Biomed Eng* 2020; 67:1604–1615
17. Shaw M, Hawthorne C, Moss L, et al: Time series analysis and prediction of intracranial pressure using time-varying dynamic linear models. *Acta Neurochir Suppl* 2021; 131:225–229
18. Ye G, Balasubramanian V, Li JK-J, et al: Machine learning-based continuous intracranial pressure prediction for traumatic injury patients. *IEEE J Transl Eng Health Med* 2022; 10:4901008
19. Wijayatunga P, Koskinen L-OD, Sundström N: Probabilistic prediction of increased intracranial pressure in patients with severe traumatic brain injury. *Sci Rep* 2022; 12:9600
20. Güiza F, Depreitere B, Piper I, et al: Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury: Development and validation in a multicenter dataset. *Crit Care Med* 2013; 41:554–564
21. Carra G, Güiza F, Depreitere B, et al: CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators: Prediction model for intracranial hypertension demonstrates robust performance during external validation on the CENTER-TBI dataset. *Intensive Care Med* 2021; 47:124–126
22. Güiza F, Depreitere B, Piper I, et al: Early detection of increased intracranial pressure episodes in traumatic brain injury: External validation in an adult and in a pediatric cohort. *Crit Care Med* 2017; 45:e316–e320
23. Carra G, Güiza F, Piper I, et al: CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators: Development and external validation of a machine learning model for the early prediction of doses of harmful intracranial pressure in patients with severe traumatic brain injury. *J Neurotrauma* 2023; 40:514–522
24. Schmidt B, Czosnyka M, Smielewski P, et al: Noninvasive assessment of ICP: Evaluation of new TBI data. *Acta Neurochir Suppl* 2016; 122:69–73
25. Miyagawa T, Sasaki M, Yamaura A: Intracranial pressure based decision making: Prediction of suspected increased intracranial pressure with machine learning. *PLoS One* 2020; 15:e0240845
26. Cardim D, Robba C, Czosnyka M, et al: Noninvasive intracranial pressure estimation with transcranial Doppler: A prospective observational study. *J Neurosurg Anesthesiol* 2020; 32:349–353