Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms

Wu, X.; Sun, Y.Y.; Xu, X.; Steyerberg, E.W.; Helmrich, I.R.A.R.; Lecky, F.; ... ; CTR-TBI Participants Investigator

Journal of Neurotrauma

Open camera or QR reader and
scan code to access this article
and other resources online.

ORIGINAL ARTICLE   **CLINICAL STUDIES**

Mortality Prediction in Severe Traumatic Brain Injury Using Traditional and Machine Learning Algorithms

Xiang Wu,[1,**] Yuyao Sun,[2,**] Xiao Xu,[2] Ewout W. Steyerberg,[3,4] Isabel R.A. Retel Helmrich,[3] Fiona Lecky,[5] Jianying Guo,[2] Xiang Li,[2] Junfeng Feng,[6,7] Qing Mao,[6,7] Guotong Xie,[2,8,9,*] Andrew I.R. Maas,[10,11] Guoyi Gao,[1,6,*] Jiyao Jiang[6,7]; and the CENTER-TBI Participants and Investigators

Abstract

Prognostic prediction of traumatic brain injury (TBI) in patients is crucial in clinical decision and health care policy making. This study aimed to develop and validate prediction models for in-hospital mortality after severe traumatic brain injury (sTBI). We developed and validated logistic regression (LR), LASSO regression, and machine learning (ML) algorithms including support vector machines (SVM) and XGBoost models. Fifty-four candidate predictors were included. Model performance was expressed in terms of discrimination (C-statistic) and calibration (intercept and slope). For model development, 2804 patients with sTBI in the Collaborative European NeuroTrauma Effectiveness Research in TBI (CENTER-TBI) China Registry study were included. External validation was performed in 1113 patients with sTBI in the CENTER-TBI European Registry study. XGBoost achieved high discrimination in mortality prediction, and it outperformed logistic and LASSO regression. The XGBoost model established in this study also outperformed prediction models currently available, including the International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) core and International Mission for Prognosis and Analysis of Clinical Trials (CRASH) basic models. When including 54 variables, XGBoost and SVM reached C-statistics of 0.87 (95% confidence interval [CI]: 0.81-0.92) and 0.85 (95% CI: 0.79-0.90) at internal validation, and 0.88 (95% CI: 0.87-0.88) and 0.86 (95% CI: 0.85-0.87) at external validation, respectively. A simplified version of XGBoost and SVM using 26 variables selected by recursive feature elimination (RFE) reached C-statistics of 0.87 (95% CI: 0.82-0.92) and 0.86 (95% CI: 0.80-0.91) at internal validation, and 0.87 (95% CI: 0.87-0.88) and 0.87 (95% CI: 0.86-0.87) at external validation, respectively. However, when the number of variables included decreased, the difference between ML and LR diminished. All the prediction models can be accessed via a web-based calculator. Glasgow Coma Scale (GCS) score, age, pupillary light reflex, Injury Severity Score (ISS) for brain region, and the presence of acute

[1]Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.
[2]Ping An Healthcare Technology, Beijing, China.
[3]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
[4]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
[5]Centre for Urgent and emergency care Research (CURE), Health Services Research Section, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, United Kingdom.
[6]Shanghai Institute of Head Trauma, Shanghai, China.
[7]Brain Injury Center, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.
[8]Ping An Health Cloud Company Ltd., Beijing, China.
[9]Ping An International Smart City Technology Co., Ltd., Beijing, China.
[10]Department of Neurosurgery, Antwerp University Hospital, Edegem, Belgium.
[11]University of Antwerp, Edegem, Belgium.
**The first two authors contributed equally.

*Address correspondence to: *Guoyi Gao, MD, PhD, Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, 85 Wujin Road, Shanghai, China* E-mail: gao3@sina.com; *Guotong Xie, PhD, Ping An Healthcare Technology, 23 Financial Street, Beijing, China* E-mail: xieguotong@pingan.com.cn

subdural hematoma were the five strongest predictors for mortality prediction. The study showed that ML techniques such as XGBoost may capture information hidden in demographic and clinical predictors of patients with sTBI and yield more precise predictions compared with LR approaches.

Keywords: extreme gradient boosting; logistic regression; machine learning; prognostic model; traumatic brain injury

Introduction

Traumatic brain injury (TBI) is the main cause of death and disability in young adults worldwide, and it is regarded as one of the conditions with the greatest health care and economic impact in society.[1] Prediction of outcome in patients after TBI is crucial in clinical decision-making and health care policy making. Patients with TBI differ in demographic characteristics, pre-injury health, cause of injury, injury severity, and clinical severity and treatments; their outcomes are highly variable. The high heterogeneity of TBI poses challenges to outcome prediction.

Much effort has been applied to prediction modeling in patients with TBI. The majority of previous models use traditional statistical analyses, such as logistic regression (LR). The two most widely validated prediction models in TBI are the International Mission for Prognosis and Analysis of Clinical Trials (CRASH) and the International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) models.[2] These models focus on modeling a limited set of key predictors. However, they only explain approximately 35% of variance in outcome.[3]

To improve the performance of the current models, machine learning (ML) algorithms may be useful. ML is a branch of artificial intelligence and is entering the realm of clinical research at an increasing pace because of a data explosion and increasing computational power.[4–7] It enables computer algorithms to learn from experience, without explicitly being guided by humans.[8] ML techniques provide new opportunities for better prediction.[9–12] However, when applied to patients with TBI, no improvements were noted.[13,14] Possible explanations include that a rather limited set of key predictors was studied, and ML methods require large numbers of potential predictors in large data sets to benefit from their greater flexibility over traditional methods.

In this study we aimed to develop and validate models to predict in-hospital mortality of patients with severe TBI (sTBI). We compared with traditional LR modeling the performance of two commonly used ML models: support vector machine (SVM) and extreme gradient boosting (XGBoost).

Methods

Study population

This study included clinical data of 2804 patients with sTBI (initial Glasgow Coma Scale [GCS] score ≤8) from the Collaborative European NeuroTrauma Effectiveness Research in TBI (CENTER-TBI) China Registry and 1113 patients with sTBI from the CENTER-TBI European Registry. In total, 13,138 patients with TBI were recruited from 52 centers across China between December 22, 2014 and August 1, 2017 in the China Registry, and 22,849 patients with TBI were recruited from 65 centers in 19 countries between December 19, 2014 and December 17, 2017 in the European Registry.[15,16] Both registries were prospective longitudinal observational studies. Data were collected for patients with a clinical diagnosis of TBI and an indication for computed tomography (CT).

Information was collected using a web-based electronic case report form (eCRF) and managed by the QuesGen data management platform. Data were coded in accordance with the Common Data Elements (CDE) scheme. During the data uploading process, the system ran data validation checks. All study data in the database were de-identified and stored securely under the supervision of the Karolinska Institutet International Neuroinformatics Coordinating Facility (KI-INCF).

Ethics approval and consent to participate

The China CENTER-TBI Registry has been conducted in accordance with all relevant laws of the People's Republic of China, including but not limited to, the relevant privacy and data protection laws and regulations (the "Privacy Law"), the relevant laws and regulations on the use of human materials, and all relevant guidance relating to clinical studies from time to time in force including, but not limited to, the ICH Harmonised Tripartite Guideline for Good Clinical Practice (CPMP/ICH/135/95) ("ICH GCP") and the World Medical Association Declaration of Helsinki entitled "Ethical Principles for Medical Research Involving Human Subjects." Ethical approval was obtained for all recruiting sites. The study protocol was approved by the ethics committees of participating centers, which waived the need for informed consent as only routinely collected clinical data were recorded. The CENTER-TBI study was registered with ClinicalTrials.gov (NCT02210221).

Outcome and predictors

The primary outcome was mortality before discharge. Fifty-four variables were available in the database, which were included to predict in-hospital mortality, including

baseline demographic characteristics, injury-related characteristics, clinical severity, radiological findings, and clinical interventions (Supplementary Table S1). Baseline, injury-related characteristics, clinical severity, and radiological findings were assessed at arrival, and clinical interventions, immediately performed as emergency procedures upon admission, were recorded at discharge. Missing data were imputed with mean value. The rate of missing data was 0.63% in the training and internal validation set, and 1.15% in the external validation set.

Model development

Regression techniques. Standard LR and LASSO regression (a logistic regression with LASSO penalization) were used. Standard LR is prone to overfitting, whereas LASSO is expected to improve the performance of LR models by shrinking some coefficients to zero.[17,18] No non-linear or interaction terms were included in the regression models.

Machine learning algorithms. Two ML tools were applied: XGBoost and SVM.[19,20] These are widely used in medical research.[5,6,9–11,20] To simplify the XGBoost model, recursive feature elimination (RFE) was applied for feature selection.[21] Briefly, this method removes the weakest features until the specified number of features is reached. Ten-fold cross-validation was used to find the optimal feature number, by scoring and selecting the best feature subsets, and to evaluate performance. Moreover, Bayesian optimization was used to fine-tune the parameters automatically for each of the ML models. Traditional tuning is often a ''black art'' requiring expert experience, rules of thumb, or sometimes brute force search. Instead, we consider this problem through the framework of Bayesian optimization, which therefore has great appeal for automatic approaches that can optimize the performance of any given learning algorithm to the problem. All participants with sTBI were randomly divided into 10 subsets. Models were trained in all but one subset (Fig. 1). The 10-fold cross validation was repeated 10 times with change in the randomization.
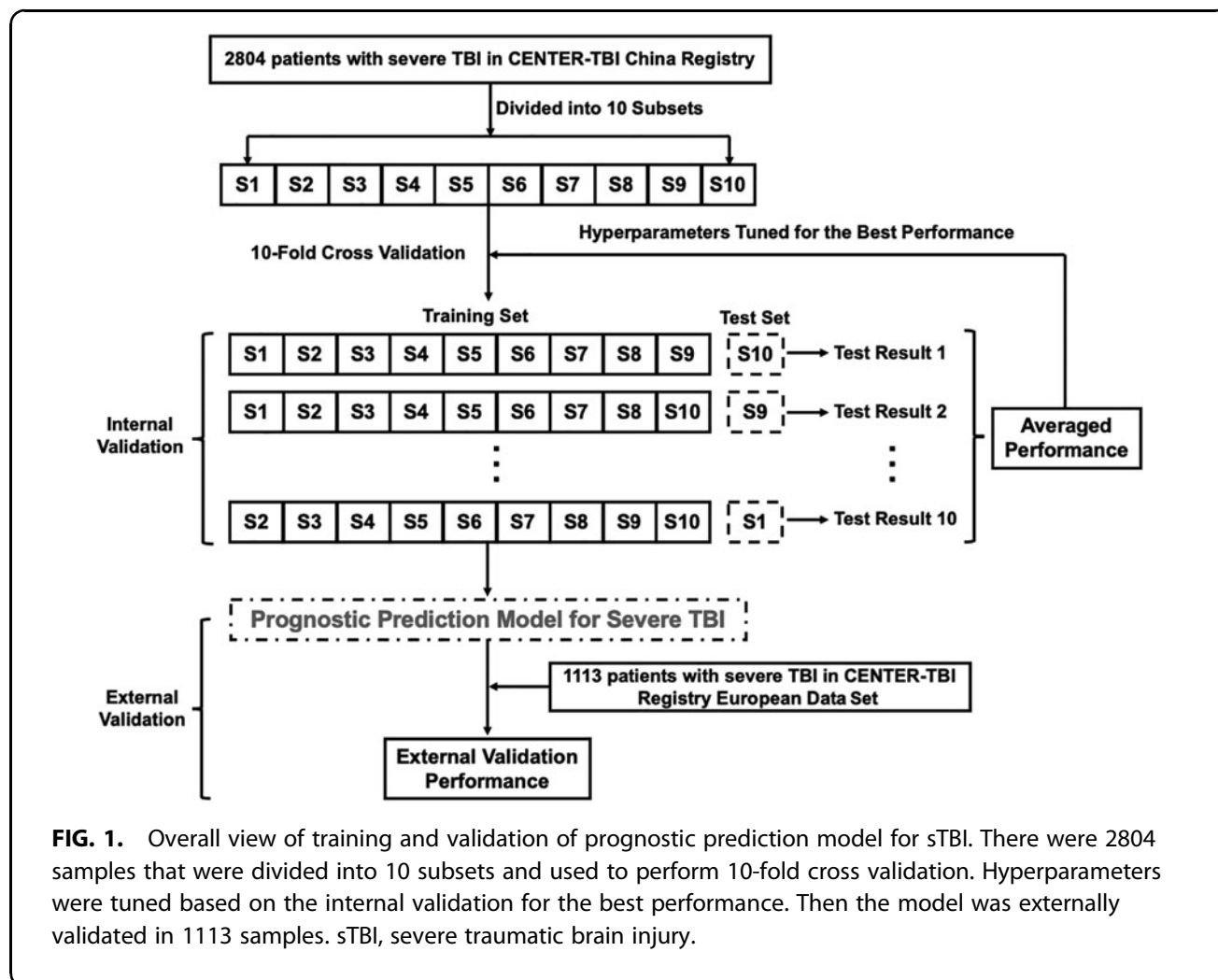

FIG. 1. Overall view of training and validation of prognostic prediction model for sTBI. There were 2804 samples that were divided into 10 subsets and used to perform 10-fold cross validation. Hyperparameters were tuned based on the internal validation for the best performance. Then the model was externally validated in 1113 samples. sTBI, severe traumatic brain injury.

Sample weighting was added to solve label imbalances. The codes of model training and hyperparameters of final models were available in Github.

The Shapley Additive exPlanations (SHAP) method was applied for better interpretability of XGBoost prediction results. SHAP is a method to explain individual predictions. The effect of each feature on outcome prediction is summed in each patient according to the non-linear XGBoost model. The impact of each feature on the outcome can hence be interpreted from the SHAP values.

Internal and external validation procedures. During 10-fold cross validation, the one subset that was not included in the model training served as the internal validation set. This process was repeated 10 times until each subset was used to test the accuracy of the model, and the performance was averaged. To capture the distributional performance of trained models, the 10-fold cross validation was repeated 10 times with change in the randomization. The hyperparameters were tuned for the best discriminating power in internal validation sets.

Data of 1113 patients with sTBI from the CENTER-TBI European Registry were used for external validation. The two studies used for data development and external validation included the same variables. The performance of the prediction model was tested via the C-statistic, calibration slope, and intercept.

Model performance of the external validation set was also compared with that of the CRASH basic model and IMPACT core modeles.[22,23]

Statistical analysis

Continuous variables were reported as median and interquartile ranges (IQRs), and categorical data as numbers and percentages. A two-tailed p-value of 0.05 or less was used to define statistical significance. The DeLong method was used to compare C-statistics between models. A total of five comparisons were made with multiple comparisons among XGBoost versus SVM, XGBoost versus LASSO, XGBoost versus naïve LR, SVM versus LASSO, and SVM versus naïve LR, and the p-value was adjusted to 0.01 according to Bonferroni correction.

All the model training and validation was performed using the "scikit-learn" module, and the XGboost package in Python (version 3.5). The hyperparameters and coding of model training and testing are available at the GitHub repository.[24] The statistical analyses (including statistical description and performance comparison) were performed using R statistical software (version 3.5.0), with RStudio (version 1.1.447) used as the implementation Integrated Development Environment (IDE). The Delong test was performed using the "roc.test" function of the pROC package (version 1.18.0). Modeling results were reported in accordance with the TRIPOD guidelines (Supplementary Appendix SA1). To allow further validation, the XGBoost and SVM can be accessed using a web-based calculator.[25]

Results

Study population

In total, 2804 patients with sTBI (GCS score ≤8) were included for model development and internal validation, of whom 552 (20%) had died in the hospital (Supplementary Fig. S1). Among them, 79% were male. The median age was 49 (IQR: 36–61) years. Most of the sTBIs occurred on the streets or highways ($n = 1731$; 62%), and 18% ($n = 511$) occurred at home. The median GCS score was 6 (IQR: 4–7) and the median Injury Severity Score (ISS) was 25 (IQR: 17–32), respectively. Thirty-nine percent of patients had at least one-side pupillary light reflex absent and 99% (2785) showed abnormal CT results. A total of 1113 patients with sTBI were included in the external validation data set, of whom 372 (33%) had died in the hospital. Compared with the training set, patients in the external validation set were slightly older, had lower GCS scores, and more were injured at home (Table 1).

Table 1. Baseline of Patients Included in the CENTER-TBI China Registry and the CENTER-TBI EU Registry

	China Registry sTBI, n = 2804	EU Registry sTBI, n = 1113
Male gender	2223 (79%)	810 (73%)
Age	49 (36-61)	50 (30-68)
Pre-injury ASA-PS		
ASA I	2190 (78%)	461 (41%)
ASA II	439 (16%)	290 (26%)
ASA III	111 (4%)	234 (21%)
ASA IV	36 (1%)	23 (2%)
Unknown	28 (1%)	105 (9%)
Injury Place		
Street/Highway	1731 (62%)	466 (42%)
Home	511 (18%)	381 (34%)
Public location (eg., bar, station, nightclub)	264 (9%)	48 (4%)
Work/School	18 (1%)	16 (1%)
Sport/Recreation	266 (9%)	141 (13%)
Other	12 (0%)	34 (3%)
Unknown	2 (0%)	27 (2%)
GCS score	6 (4-7)	3 (3-6)
Total ISS	25 (17-32)	29 (24-50)
Pupillary reflex		
Both exist	1703 (61%)	681 (61%)
One absent	370 (13%)	144 (13%)
Both absent	731 (26%)	288 (26%)
CT Result		
Normal	19 (1%)	152 (14%)
Abnormal	2785 (99%)	961 (86%)
Systolic BP (mm Hg)	131 (115-150)	128 (110-147)
≤ 90 mm Hg	161 (5.7%)	124 (11%)
Diastolic BP (mm Hg)	80 (70-89)	75 (60-90)
SPo$_2$ (%)	98 (95-99)	99 (96-100)
≤ 95%	829 (30%)	225 (20%)

ASA-PS, American Society of Anesthesiologists Physical Status; BP, blood pressure; CT, computed tomography; GCS, Glasgow Coma Scale; SPo$_2$, oxygen saturation as measured by pulse oximetry; sTBI, severe traumatic brain injury.

Prediction model construction using logistic and LASSO regression

We considered 54 candidate predictors for model development, including, age, gender, pre-injury status, ISS, GCS score, injury causes, injury places, pupillary reflex, oxygen saturation as measured by pulse oximetry (SpO_2), blood pressure, CT results, intensive care unit (ICU) admission, and emergency interventions (Supplementary Table S1). The LR model demonstrated overfitting when it included a total of 54 variables, with an average C-statistic of 0.88 (95% confidence interval [CI]: 0.86-0.90) in training sets, 0.83 (95% CI: 0.79-0.86) in internal validation sets, and 0.79 (95% CI: 0.76-0.82) in the external validation set. The calibration intercept and slope were −0.05 and 0.15, respectively, at external validation (Fig. 2A). A simplified LR model included 36 variables that showed significance ($p < 0.05$). This model had a higher C-statistic at external validation (0.84, 95% CI: 0.81-0.86, Supplementary Table S2). Further simplification using eight variables and six variables (Supplementary Tables S3 and S4) showing significance in the previous models demonstrated better performance with a C-statistic of 0.85 (95% CI: 0.82-0.87) and 0.84 (95% CI: 0.82-0.87) at external validation, respectively.

With 54 candidate variables, LASSO regression performed better than LR without penalization. The LASSO model shrunk variables to zero, leaving 36 predictors in the model. It reached a C-statistic of 0.85 (95%CI: 0.81-0.88) at internal validation and 0.86 (95%CI: 0.83-0.88) at external validation. The calibration intercept and slope were −0.48 and 1.03, respectively, at external validation (Fig. 2B). A simplified model with 36, 8, and 6, candidate variables showed similar C-statistics of 0.86 (95%CI: 0.83-0.88), 0.85 (95%: 0.83-0.88), and 0.85 (95%: 0.83-0.88) at external validation, respectively.

Prediction model construction using ML algorithms

When including all 54 predictors, the SVM model reached an average C-statistic of 0.85 (95% CI: 0.79-0.90) in internal validation sets, and 0.86 (95% CI: 0.85-0.87) in the external validation set. Both SVM and XGBoost achieved better calibration performance than regression models. The calibration intercept and slope were −0.21 and 1.19, respectively, at external validation (Fig. 2C). XGBoost performed slightly better compared with SVM, it and achieved 0.87 (95% CI: 0.81-0.92) in internal validation sets, and 0.88 (95% CI: 0.87-0.88) in the external validation set. Calibration intercept and slope were −0.10 and 1.34, respectively, at external validation (Fig. 2D). The emphasis on sensitivity and specificity will be determined by the users. At a cutoff value of 0.27, the XGBoost model had a sensitivity of 90% and a specificity of 62%; at a cutoff value of 0.57, the model had a sensitivity of 64% and a specificity of 90%.

After RFE, which removed the weakest features until the optimal number was reached, a simplified ML model was built using the 26 variables selected by RFE (Supplementary Table S5), which reached similar performance compared with all 54 variables. The average C-statistic was 0.87 (95% CI: 0.82-0.92) in the internal and 0.87 (95% CI: 0.87-0.88) in the external validation set for XGBoost, and 0.86 (95% CI: 0.80-0.91) in the internal and 0.87 (95% CI: 0.86-0.87) in the external validation set for SVM. The calibration intercept for external validation was −0.33 for XGBoost and −0.52 for SVM. The calibration slope for external validation was 1.22 for XGBoost and 1.06 for SVM (Supplementary Fig. S2).

SHAP analysis for the XGBoost model revealed that the five strongest predictors for mortality were: low GCS score, elder age, absent pupillary light reflex, high ISS for brain region (which is the quadratic of brain abbreviated injury scale (AIS), with a maximum of 75 assigned when brain AIS was 6) and presence of acute subdural hematoma. Other important features included low oxygen saturation, high total ISS, midline shift over 5 mm, presence of contusions, need for intensive care, too low or too high systolic blood pressure, and low GCS motor score. Secondary referral and cerebrospinal fluid (CSF) drainage was associated with a lower mortality rate (Fig. 3 and Supplementary Fig. S3).

Interaction analysis suggested that the impact of age on outcome decreased at low GCS score. Besides, whether GCS is low or high, the younger age (<48 years) tended to decreased mortality, and the elder age (>48 years) tended to increased mortality. It was also found that the impact of brain injury ISS increased at low GCS score. In other words, when the GCS score is low, the brain ISS can give us extra information about mortality (Supplementary Fig. S4).

Besides, the SHAP model can better interpret the XGBoost model, which, unlike LR, is difficult to explain due to its non-linearity. Its application in explaining outcome prediction of two individuals is demonstrated in Supplementary Figure S5. In the first case, the predicted mortality was above average because severe comorbidity, severe injury with an ISS of 75 and GCS score of 3, low oxygen saturation at the scene, and mass subdural hematoma increased the mortality, although normal pupillary light reflex and CSF drainage lowered the mortality. In the second case, the predicted mortality was below average because this patient needed no ICU treatment, the brain ISS was relatively low, and the initial CT showed only the minor contusion without midline shift or subdural hematoma, although the age was high and the oxygen saturation was relatively low.

Comparison between Linear regression, LASSO regression, and machine learning algorithms

When including a total of 54 candidate variables, XGBoost outperformed naïve LR and LASSO regression
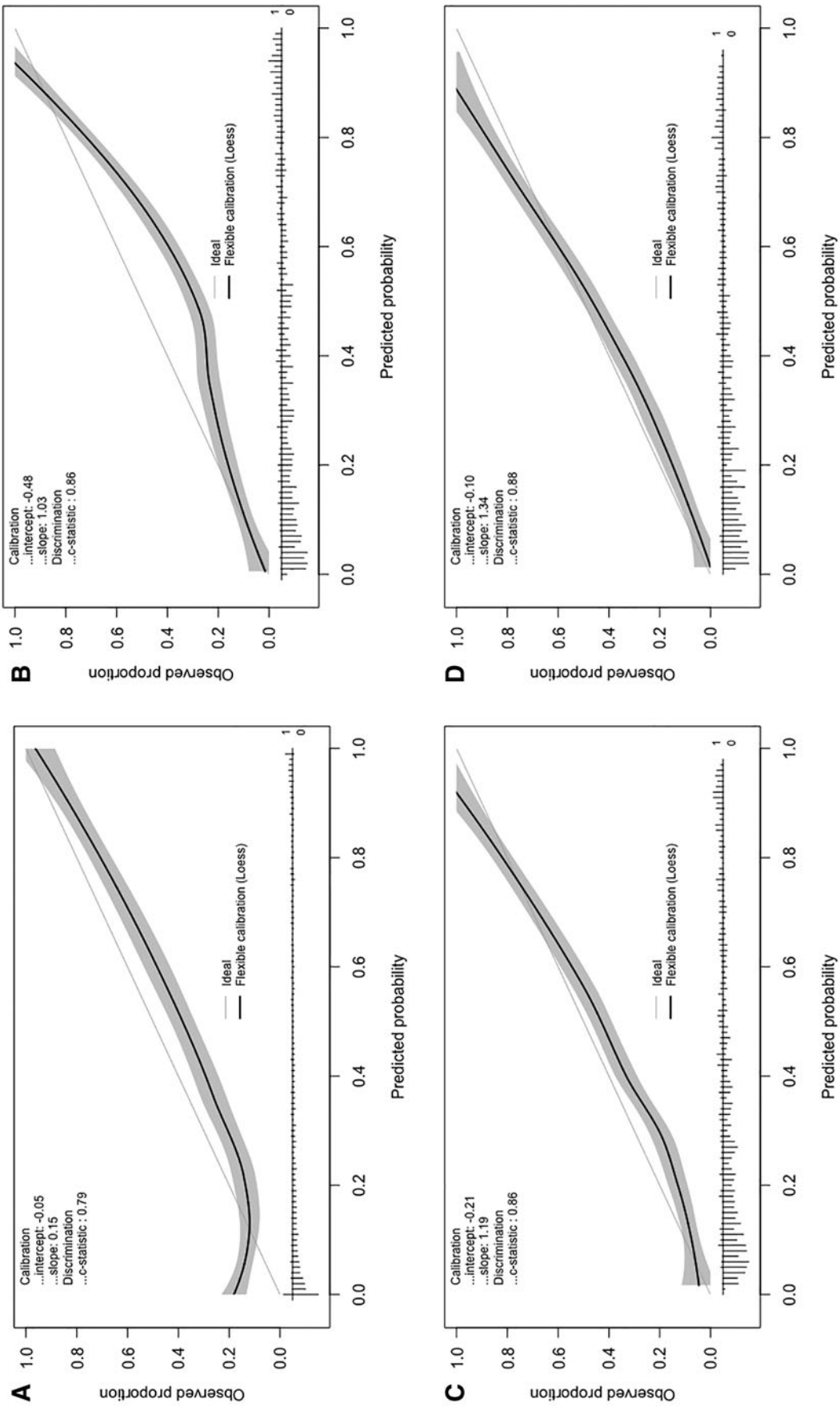
FIG. 2. Calibration plot of external validation in Logistic regression (**A**), LASSO regression (**B**), SVM (**C**), and XGBoost (**D**) for 54 variables. XGBoost outperformed the other three models, and it achieved a C-statistic of 0.88 (95% confidence interval: 0.87–0.88) in the external validation set. SVM, support vector machines.
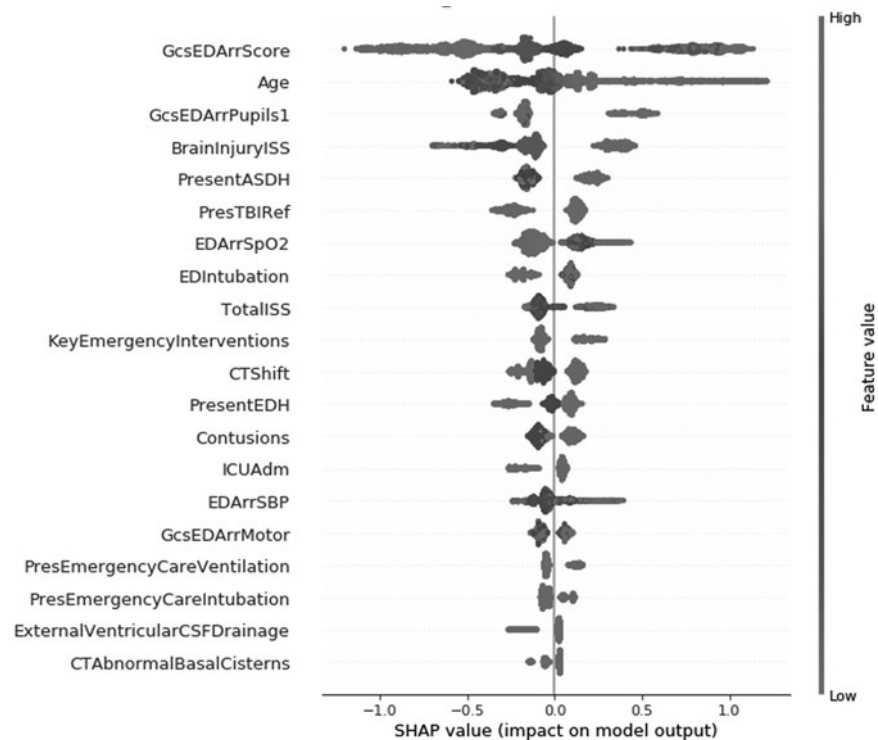
FIG. 3. Impact of features in the XGBoost model for sTBI mortality prediction using SHAP. The SHAP values were derived from the results of internal validation. The five strongest predictors for mortality were: low GCS score, elder age, absent pupillary light reflex, high ISS for brain region, and presence of acute subdural hematoma. GCS, Glasgow Coma Scale; ISS, Injury Severity Index; SHAP, Shapley Additive exPlanations; sTBI, severe traumatic brain injury.

in C-statistic ($p < 0.0001$ and $p < 0.001$, respectively; Fig. 4 and Supplementary Fig. S6), and SVM outperformed naïve LR ($p < 0.0001$). As the selected features reduced to 26, the performance of LR increased, but XGBoost still performed better than naïve LR and LASSO regression ($p < 0.0001$ and $p = 0.0016$, respectively), and SVM still outperformed naïve LR ($p = 0.00019$). However, when the number of features was further reduced, the performance of both SVM and XGBoost reduced significantly and showed a similar discriminating power with naïve LR ($p = 0.23$ and 0.20, respectively) and LASSO regression ($p = 0.24$ and 0.22, respectively) when it only included six variables. XGBoost showed high robustness and the best performance in discriminating hospital mortality throughout different numbers of variables included. The comparison of performance between each model is presented in Supplementary Tables S6 and S7, and the detailed performance of 10 randomization repetition is shown in Supplementary Figure S7.

Comparison with IMPACT and CRASH models

The XGBoost model (both original and simplified version) outperformed the currently widely accepted

IMPACT core and CRASH basic prognostic models. In the external validation set, the CRASH basic model achieved C-statistic of 0.82 (95%CI: 0.79-0.84) and the IMPACT core model reached 0.80 (95%: 0.78-0.83). Calibration slopes were 0.92 and 1.17 for the CRASH and the IMPACT models, respectively, and calibration intercepts were -0.49 and -0.02 for the CRASH and the IMPACT models, respectively. Due to limitations of the database, variables required for the IMPACT core+CT and the CRASH-CT model were not available.

Model presentation

To facilitate external validation by independent research, all models including XGBoost, SVM, and LR can be accessed using a web-based calculator.[25] Both the 54-variable model and the simplified versions are available online by clicking corresponding labels (Supplementary Fig. S8). The risk percentage calculated implies the predicted mortality rate at discharge.

Discussion

The current study developed and compared strategies for prediction modeling of in-hospital mortality in patients
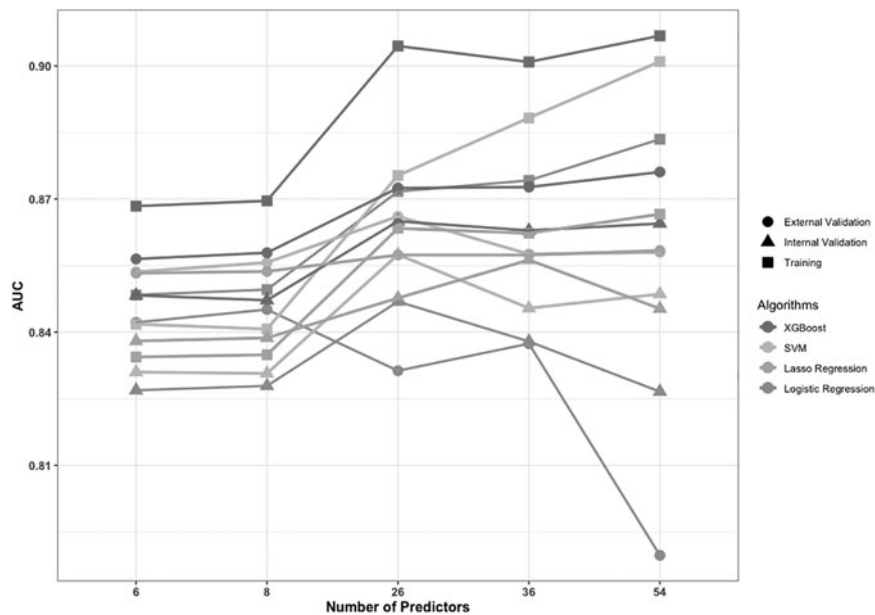
FIG. 4. Performance comparison based on the area under the curve of different algorithms when including a different amount of predictors. XGBoost showed the best performance in discriminating hospital mortality in a training set, internal validation set, and external validation set throughout different numbers of variables included. SVM, support vector machines.

after sTBI based on commonly available demographic and clinical data. A total of 2804 patients with sTBI in the CENTER-TBI China Registry were included in model development and 1113 in the CENTER-TBI European Registry were used for external validation. The XGBoost model achieved high discrimination and calibration performance in predicting in-hospital mortality, and it outperformed established prediction models for outcome prediction in TBI.

Compared with other ML algorithms, the current model included more clinical scales and medical interventions.[13] It did not require laboratory indicators, including serum glucose level, C-reactive protein, sodium level, etc.[13,26,27] Thus, the model might be used for early prediction in the emergency room.

Because 20% of patients with sTBI died before discharge, early determination of prognosis is a priority for both the physicians and relatives involved.[15,28] Reliable assessment of prognosis in patients with TBI is critical for clinical decision-making, health care policy making, family counseling, allocation of resources, research, and assessment of the quality of health care.[3] Of note, this model included emergency clinical interventions, so it is specific to current practice and indications for starting these interventions. The effectiveness of the treatments, however, cannot be derived from the current modeling and requires further study.

To predict the outcome of patients with TBI, many prediction models have been developed. Some of the

prediction models have been validated and showed high accuracy, for example, the IMPACT prognostic models and the CRASH prognostic models.[29–31] Most predictors identified in the current model are in line with established models including the IMPACT and the CRASH models. The IMPACT model includes age, GCS motor score, pupil reactivity hypoxia, hypotension, and CT findings to predict mortality or unfavorable outcome at 6 months.[23] The CRASH model includes age, GCS score, pupil reactivity, major extracranial injury, and CT findings to predict mortality at 14 days or unfavorable outcome at 6 months.[22]

Compared with these established models, the current XGBoost model achieved higher discriminative accuracy. Consistent with previous studies, the current XGBoost model revealed that low GCS score, elder age, absent pupillary light reflex, and presence of acute subdural hematoma were among the most important features for mortality. However, this study found a nonlinear association between some variables (e.g., GCS score, age, and ISS) and outcome. In addition, some predictors including head ISS, secondary referral, and emergency interventions were found to be relevant for prediction of in-hospital mortality by XGBoost and were rarely explored in previous models. This may underlie the performance improvement of the XGBoost model.

Currently, ML is ubiquitous and indispensable for solving complex problems of unstructured data in most sciences, due to its ability to handle large numbers of

predictors.[12] However, only a few studies have investigated the application of ML in outcome prediction of TBI, and they have achieved quite contradictory conclusions. Gravesteijn and colleagues found that ML may not outperform LR for outcome prediction after moderate or sTBI.[13] Whereas studies by Lu, Matsuo, Feng, and their colleagues indicated a relatively good predictive performance of modern ML for TBI outcome compared with the regression approach.[26,27,32] The current study indicates that one source of contradiction may be the numbers and types of predictors included in each model, and the balance of large numbers of predictors to large sample size.

The results of our study reveal that the number of variables affected the performance of ML and LR conversely in TBI prognostic prediction. When including only a small number of predictors, ML didn't show better performance compared with LR, and some ML algorithms even perform more poorly than LR. As the included number of variables increased, the performance of XGBoost and SVM improved, and they reached higher discrimination and calibration performance than regression models, because more information, including signals and noises, was contained in the predictors, and ML can eliminate redundant noise and better capture features of the patient before making predictions. The LR performance decreased when including more predictors, indicating low robustness for high-dimensional settings. LR is more suitable for low-dimensional data, whereas ML shows more potential in large-scale, multi-modality settings. Currently, continuous long-term multi-modality monitoring is commonly applied in critical care patients, and together with increasing biomarkers and radiological images, it may promote the use of ML for TBI outcome prediction.

The main strengths of this study are the large scale of the cohort, the prospective recording of patient data, and the external validation of models in the CENTER-TBI EU Registry study with an identical data collection protocol to CENTER-TBI China. Limitations of this study include a lack of lab and detailed radiological findings. The limited number of features included may hamper the performance of ML algorithms and led to a minimal increase in discrimination power compared with traditional regression algorithms. Further studies are needed to provide any clinically meaningful decision. In addition, there is no fixed time for outcome evaluation (death).

Conclusions

We developed and compared prediction models for in-hospital mortality in patients after sTBI based on demographic and clinical data in the CENTER-TBI China and EU Registries. The result demonstrated that the simplified XGBoost model achieved both accuracy

and clinical usability. In addition, XGBoost was promising as a ML tool, which revealed superior performance by capturing information hidden in demographic and clinical predictors in large data sets of patients after sTBI.

Availability of Supporting Data

Researchers who submit a methodologically sound study proposal that is approved by the management committee can have access to the study protocol, individual participant data, data dictionary, analytic code, and analysis scripts. A data access agreement is required, and all access must comply with regulatory restrictions imposed on the original study. The corresponding authors can provide the online address for proposals and the study group monitored email address.

Transparency, Rigor, and Reproducibility Statement

The study was pre-registered at clinicaltrials.gov (NCT02210221). A total of 2804 patients with sTBI in China Registry and 1113 patients with sTBI in European Registry were included. Final clinical outcome assessments and adjudications were performed by team members blinded to relevant characteristics of the participants. All surveys and questionnaires used to develop prognostic models are available from the authors. The key inclusion criteria and outcome evaluations are established standards. Replication by the study group was performed as part of this study. De-identified data from this study can be available after approval of submitted study proposal at https://www.center-tbi.eu/data. Analytic code used to conduct the prognostic analyses is available in (https://github.com/Yuyoo/Mortality-prediction-in-sTBI) and a website for performing batch prognostic predictions is available at (http://101.89.95.81:8654/). This paper will be published under a Creative Commons Open Access license, and upon publication will be freely available at https://www.liebertpub.com/loi/neu.

Authors' Contributions

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated in the concept, design, analysis, writing, or revision of the manuscript. All authors participated in the reported analyses and interpretation of results relevant to their domain

of interest. XW and YS prepared the draft manuscript and coordinated its finalization. XW, YS, and XX performed statistical analyses and drafting of tables and figures. ES, IH, FL, JG, and XL revised the manuscript and gave support in statistical analyses and figures drafting. All authors have seen and approved the final manuscript.

Funding Information

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Figure S7
Supplementary Figure S8
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5
Supplementary Table S6
Supplementary Table S7
Supplementary Appendix SA1

References

1. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 2017;16(12):987–1048; doi: 10.1016/S1474-4422(17)30371-X

2. Dijkland SA, Foks KA, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J Neurotrauma 2020;37(1):1–13; doi: 10.1089/neu.2019.6401

3. Lingsma HF, Roozenbeek B, Steyerberg EW, et al. Early prognosis in traumatic brain injury: from prophecies to predictions. Lancet Neurol 2010;9(5):543–554; doi: 10.1016/S1474-4422(10)70065-X

4. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg 2018;109:476.e1–486.e1; doi: 10.1016/j.wneu.2017.09.149

5. Burbidge R, Trotter M, Buxton B, et al. Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput Chem 2001;26(1):5–14.

6. Magnin B, Mesrob L, Kinkingnehun S, et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 2009;51(2):73–83; doi: 10.1007/s00234-008-0463-x

7. Johnson AE, Ghassemi MM, Nemati S, et al. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016;104(2):444–466; doi: 10.1109/JPROC.2015.2501978

8. Senders JT, Zaki MM, Karhade AV, et al. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien) 2018; 160(1):29–38; doi: 10.1007/s00701-017-3385-8

9. Orru G, Pettersson-Yeo W, Marquand AF, et al. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 2012;36(4):1140–1152; doi: 10.1016/j.neubiorev.2012.01.004

10. Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med 2019; 380(26):2497–2505; doi: 10.1056/NEJMoa1812757

11. Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8–17; doi: 10.1016/j.csbj.2014.11.005

12. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med 2016;375(13):1216–1219; doi: 10.1056/NEJMp1606181

13. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95–107; doi: 10.1016/j.jclinepi.2020.03.005

14. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. J Clin Epidemiol. 2016;78:83–89; doi: 10.1016/j.jclinepi.2016.03.002

15. Gao G, Wu X, Feng J, et al. Clinical characteristics and outcomes in patients with traumatic brain injury in China: a prospective, multicentre, longitudinal, observational study. Lancet Neurol 2020;19(8):670–677; doi: 10.1016/S1474-4422(20)30182-4

16. Steyerberg EW, Wiegers E, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol 2019;18(10):923–934; doi: 10.1016/S1474-4422(19)30232-7

17. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2019.

18. R Tibshirani. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 1996;58:267–288.

19. Bennett KP, Campbell C. Support Vector machines: hype or hallelujah? SIGKDD Explorations 2000;2(2):1–13.

20. Noble WS. What is a support vector machine? Nat Biotechnol 2006;24(12): 1565–1567; doi: 10.1038/nbt1206-1565

21. Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, et al. Machine learning algorithms reveal unique gene expression profiles in muscle biopsies from patients with different types of myositis. Ann Rheum Dis 2020; 79(9):1234–1242; doi: 10.1136/annrheumdis-2019-216599

22. MRC CRASH Trial Collaborators; Perel P, Arango M, Clayton, T et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ 2008; 336(7641):425–429; doi: 10.1136/bmj.39461.643438.25

23. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008;5(8):e165; discussion e165; doi: 10.1371/journal.pmed.0050165

24. https://github.com/Yuyoo/Mortalityprediction-in-sTBI

25. http://101.89.95.81:8654/

26. Lu HY, Li TC, Tu YK, et al. Predicting long-term outcome after traumatic brain injury using repeated measurements of Glasgow Coma Scale and data mining methods. J Med Syst 2015;39(2):14; doi: 10.1007/s10916-014-0187-x

27. Matsuo K, Aihara H, Nakai T, et al. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. J Neurotrauma 2020;37(1):202–210; doi: 10.1089/neu.2018.6276

28. Gao J, Zheng Z. Development of prognostic models for patients with traumatic brain injury: a systematic review. Int J Clin Exp Med 2015; 8(11):19881–19885.

29. Castano-Leon AM, Lora D, Munarriz PM, et al. Predicting outcomes after severe and moderate traumatic brain injury: an external validation of impact and crash prognostic models in a large spanish cohort. J Neurotrauma 2016;33(17):1598–1606; doi: 10.1089/neu.2015.4182

30. Lingsma HF, Yue JK, Maas AI, et al. Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study. J Neurotrauma 2015;32(2):83–94; doi: 10.1089/neu.2014.3384

31. Roozenbeek B, Chiu YL, Lingsma HF, et al. Predicting 14-day mortality after severe traumatic brain injury: application of the IMPACT models in the brain trauma foundation TBI-trac(R) New York State database. J Neurotrauma 2012;29(7):1306–1312; doi: 10.1089/neu.2011.1988

32. Feng JZ, Wang Y, Peng J, et al. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. J Crit Care 2019;54:110–116; doi: 10.1016/j.jcrc.2019.08.010