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

Journal of Neurotrauma

DOI: 10.1089/neu.2022.0221

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

Running Title: Mortality prediction in sTBI

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[***]

[1] *Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; X. Wu MD; Prof G.Y. Gao MD;*

[2] *Ping An Healthcare Technology, Beijing, China; Prof Y.Y. Sun; Prof X. Xu; Prof J.Y. Guo; Prof X. Li; Prof G.T. Xie;*

[3] *Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; Prof E. Steyerberg; I.R.A. Retel Helmrich MSc;*

[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, UK; Prof. F. Lecky MD;*

[6] *Shanghai Institute of Head Trauma, Shanghai, China; Prof G.Y. Gao MD; Prof J.F. Feng MD; Prof Q. Mao MD; Prof J.Y. Jiang MD;*

[7] *Brain Injury Center, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; Prof J.F. Feng MD; Prof Q. Mao MD; Prof J.Y. Jiang MD;*

[8] *Ping An Health Cloud Company Limited, Beijing, China; Prof G.T. Xie;*

[9] *Ping An International Smart City Technology Co., Ltd., Beijing, China; Prof G.T. Xie;*

[10] *Department of Neurosurgery, Antwerp University Hospital, Edegem, Belgium; Prof. A.I.R. Maas MD;*

[11] *University of Antwerp, Edegem, Belgium; Prof. A.I.R. Maas MD*

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)
This paper has been peer-reviewed and accepted for publication, but has yet to undergo copyediting and proof correction. The final published version may differ from this proof.

* Joint First Authors

** Joint Corresponding Authors

***Listed in the Appendix; Study group monitored email address: center-tbi@uza.be

Xiang Wu, Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; drwuxiang@163.com; Tel: +86 158 0059 0792

Yuyao Sun, Ping An Healthcare Technology, Beijing, China; sunyuyao188@pingan.com.cn; Tel: +86 021 38637279

Xiao Xu, Ping An Healthcare Technology, Beijing, China; xuxiao780@pingan.com.cn; Tel: +86 021 38637279

Ewout W. Steyerberg, Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; E.W.Steyerberg@lumc.nl; Tel: +31 71 526 9700

Isabel R.A. Retel Helmrich, Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; i.retelhelmrich@erasmusmc.nl; Tel: +31 10 704 0 704

Fiona Lecky, Centre for Urgent and emergency care Research (CURE), Health Services Research Section, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK; f.e.lecky@sheffield.ac.uk; Tel: +44 114 222 4345 3023

Jianying Guo, Ping An Healthcare Technology, Beijing, China; guojyjane@outlook.com; Tel: +86 021 38637279

Xiang Li, Ping An Healthcare Technology, Beijing, China; lixiang453@pingan.com.cn; Tel: +86 021 38637279

Junfeng Feng, Shanghai Institute of Head Trauma, Shanghai, China; fengjfmail@163.com; Tel: +86 136 1186 0825

Qing Mao, Shanghai Institute of Head Trauma, Shanghai, China; neurojack@163.com; Tel: +86 136 7192 3990

Guotong Xie, Ping An Healthcare Technology, Beijing, China; xieguotong@pingan.com.cn; Tel: +86 021 38637279

Andrew I.R. Maas, Department of Neurosurgery, Antwerp University Hospital, Edegem, Belgium; andrew.maas@uza.be; Tel: +32 3 821 30 00

Guoyi Gao, Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; gao3@sina.com; Tel: +86 138 0187 4393

Jiyao Jiang, Shanghai Institute of Head Trauma, Shanghai, China; jiyaojiang@126.com; Tel: +86 139 0199 2806

Corresponding author:

Guoyi Gao

Department of Neurosurgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

E-mail: gao3@sina.com

Text Word Count: 3395 words; Abstract Word Count: 302 words

Figure Count: 4; Table Count: 1; Reference Count: 30

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Abstract

Prognostic prediction of traumatic brain injury (TBI) 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 sTBI patients in CENTER-TBI China Registry study were included. External validation was performed in 1113 sTBI patients in CENTER-TBI European registry study. XGBoost achieved high discrimination in mortality prediction, and outperformed Logistic and LASSO regression. The XGBoost model established in this study also outperformed prediction models currently available, including IMPACT core and CRASH basic models. When including 54 variables, XGBoost and SVM reached C-statistics of 0.87 (95% 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, respectively, 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. GCS, age, pupillary light reflex, ISS for brain region and the presence of acute subdural hematoma were the 5 strongest predictors for mortality prediction. The study showed that machine learning techniques such as XGBoost may capture information hidden in demographic and clinical predictors of patients with sTBI and yield more precise predictions compared to logistic regression approaches.

Key words: Traumatic brain injury; Prognostic model; Logistic regression; Machine learning; Extreme gradient boosting

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Introduction

Traumatic brain injury (TBI) is the main cause of death and disability in young adults worldwide, and is regarded as one of the conditions with the greatest healthcare 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, clinical severity and treatments, and their outcomes are highly variable. The high heterogeneity of TBI poses challenges to outcome prediction.

Much efforts have been made in prediction modeling in patients with TBI. A majority of previous models use traditional statistical analyses, such as logistic regression (LR). The two most widely validated prediction models in TBI are the CRASH and IMPACT models[2]. These models focused 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 the 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]. Explanations may include that a rather limited set of key predictors was studied, while ML methods require large numbers of potential predictors in large data sets to benefit from their greater flexibility than traditional methods.

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

Materials and Methods

Study population

This study included clinical data of 2804 patients with severe TBI (sTBI, initial GCS≤8) from the CENTER-TBI China Registry and 1113 patients with severe TBI from CENTER-TBI Europe Registry. In total, 13138 TBI patients were recruited from 52 centers across China between 22nd, Dec, 2014, and 1st, Aug, 2017 in the China Registry, and 22849 TBI patients were recruited from 65 centers in 19 countries between 19th Dec, 2014, and 17th Dec, 2017 in the European Registry[15,16]. Both registries were prospective longitudinal observational studies. Data was collected for patients with a clinical diagnosis of TBI and an indication for Computed Tomography (CT). The study protocol was approved by the ethics committees of participating centres, who waived the need for informed consent as only routinely collected clinical data were recorded. The CENTER-TBI study was registered with ClinicalTrials.gov (NCT02210221).

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 (https://www.commondataelements.ninds.nih.gov). 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 Karolinska Institutet International Neuroinformatics Coordinating Facility (KI-INCF).

Outcome and predictors

The primary outcome was mortality before discharge. A total of 54 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 (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 training and internal validation set, and 1.15% in 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, while LASSO is expected to improve the performance of logistic regression 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 machine learning 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 use to finetune the parameters automatically for each of the machine learning 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 is therefore 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 (Figure 1). The 10-fold cross validation was repeated 10 times with change in the randomization. Sample weighting was added to solve label imbalances. The codes of model training and hyperparameters of final models were available in Github.

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

nonlinear 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 1,113 patients with sTBI from 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 prediction model was tested via the C-statistic, calibration slope and intercept.

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

Statistical analysis

Continuous variables were reported as median and IQRs, and categorical data as numbers and percentages. A two-tailed p value of 0.05 or less was used to define statistical significance. DeLong method was used to compare C-statistics between models. A total of 5 comparison were made in multiple comparison among XGBoost vs. SVM, XGBoost vs. LASSO, XGBoost vs. naïve LR, SVM vs. LASSO, and SVM vs. naïve LR, and the p value was adjusted to 0.01 according to Bonferroni correction.

All the model training and validation were performed using “scikit-learn” module, and XGboost package in Python (version 3.5). The hyperparameters and coding of model training and testing are available at GitHub repository (https://github.com/Yuyoo/Mortality-prediction-in-sTBI). The statistical analyses (including statistical description and performance comparison) were performed using R (version

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

3.5.0) statistical software, with RStudio (version 1.1.447) used as the implementation IDE. Delong test was performed using "roc.test" function of pROC package (version 1.18.0). Modeling results were reported in accordance with the TRIPOD guidelines (Supplement). To allow further validation, the XGBoost and SVM can be accessed using a web-based calculator at http://101.89.95.81:8654/.

Results

Study population

In total, 2,804 patients with severe TBI (sTBI) (GCS ≤ 8) were included for model development and internal validation, of whom 552 (20%) had died at discharge (figure S1). Among them, 79% were male. The median age was 49 (IQR: 36-61) years. Most of the sTBI occurred on the streets or highways (n=1731; 62%), and 18% (n=511) got injured at home. The median GCS was 6 (IQR: 4-7) and the median Injury Severity Score (ISS) was 25 (IQR: 17-32), respectively. 39% of patients had at least one-side pupillary light reflex absent and 2,785 (99%) showed abnormal CT results. A Total of 1,113 sTBI patients were included in the external validation dataset, of whom 372 (33%) had died at discharge. Compared to the training set, patients in the external validation set were slightly older, had lower GCS, and more patients got injured at home (Table 1).

Prediction model construction using logistic and LASSO regression

We considered 54 candidate predictors for model development, including, age, gender, pre-injury status, ISS, GCS, injury causes, injury places, pupillary reflex, SpO_2, blood pressure, CT results, ICU admission and emergency interventions (table S1). The LR model demonstrated overfitting when including a total of 54 variables, with an average C-statistic of 0.88 (95% 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 (figure 2a). A simplified LR model included 36 variables which showed significance ($p<0.05$). This model had a higher C-statistic at external validation (0.84, 95% CI: 0.81-0.86, table S2). Further simplification using 8 variables and 6 variables (table 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 (figure 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 support vector machine (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 (figure 2c). XGBoost performed slightly better compared to SVM, 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 (figure 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 (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 internal and 0.87 (95% CI: 0.87-0.88) in external validation set for XGBoost, and 0.86 (95% CI: 0.80-0.91) in internal and 0.87 (95% CI: 0.86-0.87) in external validation set for SVM. Calibration intercept for external validation was -0.33 for XGBoost and -0.52 for SVM. Calibration slope for external validation was 1.22 for XGBoost and 1.06 for SVM (figure S2).

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

SHAP analysis for the XGBoost model revealed that the 5 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 AIS, with a max 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, needs for intensive care, too low or too high systolic blood pressure, low GCS motor score. Secondary referral and CSF drainage was associated with a lower mortality rate (Figure 3 and Figure S3).

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

Besides, SHAP model can better interpret XGBoost model, which, unlike logistic regression, is difficult to explain due to its non-linearity. Its application in explaining outcome prediction of two individuals was demonstrated in figure S5. In the first case, the predicted mortality was above average because severe comorbidity, severe injury with ISS of 75 and GCS of 3, low oxygen saturation at 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, the initial CT only showed 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 in c-statistic (P<0.0001 and P<0.001, respectively, Figure 4 and Figure 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 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 6 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 was presented in Table S6 and Table S7, and the detailed performance of 10 randomization repetition were shown in 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-statistics of 0.82 (95%CI: 0.79-0.84) and IMPACT core model reached 0.80 (95%: 0.78-0.83). Calibration slopes were 0.92 and 1.17 for CRASH and IMPACT models respectively, and calibration intercepts were -0.49 and -0.02 for CRASH and IMPACT models respectively. Due to limitation of database, variables required for IMPACT core+CT and CRASH-CT model were not available.

Model presentation

To facilitate external validation by independent researches, all models including XGBoost, SVM and LR can be accessed using a web-based calculator at http://101.89.95.81:8654/. Both the 54-variable model and the simplified versions are available online by clicking corresponding labels (figure 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 after sTBI based on commonly available demographic and clinical data. A total of 2804 patients after 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 outperformed established prediction models for outcome prediction in TBI.

Compared to 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. [13,24,25]. Thus, the model might be used for early prediction in the emergency room.

Since 20% of patients with sTBI died before discharge, early determination of prognosis is a priority for both the physicians and relatives involved[15,26]. 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, e.g., the International Mission for Prognosis and Clinical Trials in Traumatic Brain Injury (IMPACT) prognostic models and the Corticosteroid Randomization after Significant Head Injury (CRASH) prognostic models[27-29]. Most predictors identified in the current model were in line with established models including IMPACT and 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 non-linear

association between some variables (e.g. GCS, age and ISS) and outcome. In addition, some predictors including head ISS, secondary referral and emergency interventions were found 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 investigated the application of ML in outcome prediction of TBI, and achieved quite contradictory conclusions. Gravesteijn and colleagues found that ML may not outperform logistic regression for outcome prediction after moderate or sTBI[13]. While studies by Matsuo, Lu, Feng and their colleagues indicated a relatively good predictive performance of modern ML for TBI outcome compared with regression approach[24,25,30]. The current study indicated 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 result of our study revealed 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 to 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 reached higher discrimination and calibration performance than regression models, because more information, including signals and noises, were 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, while 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. A limitation of this study included 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. Besides, there is no fixed time for outcome evaluation (death).

Conclusions

In conclusion, 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. Besides, XGBoost was promising as a machine learning tool, which revealed superior performance by capturing information hidden in demographic and clinical predictors in large datasets of patients after sTBI.

Acknowledgements

The CENTER-TBI project was supported by the European Commission 7th Framework program (EC grant 602150). We are immensely grateful to all patients and investigators for helping us in our efforts to improve care and outcome for TBI.

Competing interests

All authors declare no competing interests.

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

No specific funding was provided for the China TBI registry. The coordinating center received support from the European Commission 7th Framework program (602150), in the context of CENTER-TBI.

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.

Consent for publication

All authors have seen and approved the final manuscript.

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. Proposals may be submitted online https://www.center-tbi.eu/data. A Data Access Agreement is required, and all access must comply with regulatory restrictions imposed on the original study.

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, Murray GD, Maas AI. Early prognosis in traumatic brain injury: from prophecies to predictions. Lancet Neurol. 2010;9(5):543-54;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;109:476-86 e1;doi:10.1016/j.wneu.2017.09.149

5. Burbidge R, Trotter M, Buxton B, Holden S. 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):444-66;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, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev. 2012;36(4):1140-52;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):2497-505;doi:10.1056/NEJMoa1812757

11. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. 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):1216-9;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;78:83-9;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):670-7;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):923-34;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 T. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B. 1996;58:267–88

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):1565-7;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):1234-42;doi:10.1136/annrheumdis-2019-216599

22. Collaborators MCT, Perel P, Arango M, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ. 2008;336(7641):425-9;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 e;doi:10.1371/journal.pmed.0050165

24. 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

25. 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):202-10;doi:10.1089/neu.2018.6276

26. 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):19881-5

27. 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):1598-606;doi:10.1089/neu.2015.4182

28. 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

29. 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):1306-12;doi:10.1089/neu.2011.1988

30. 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;54:110-6;doi:10.1016/j.jcrc.2019.08.010

Table 1. Baseline of patients included in the CENTER TBI China Registry and 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%)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Abnormal	2785 (99%)	961 (86%)
Systolic BP (mmHg)	131 (115-150)	128 (110-147)
≤90 mmHg	161 (5.7%)	124 (11%)
Diastolic BP (mmHg)	80 (70-89)	75 (60-90)
SPO_2 (%)	98 (95-99)	99 (96-100)
≤95%	829 (30%)	225 (20%)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Figure legend

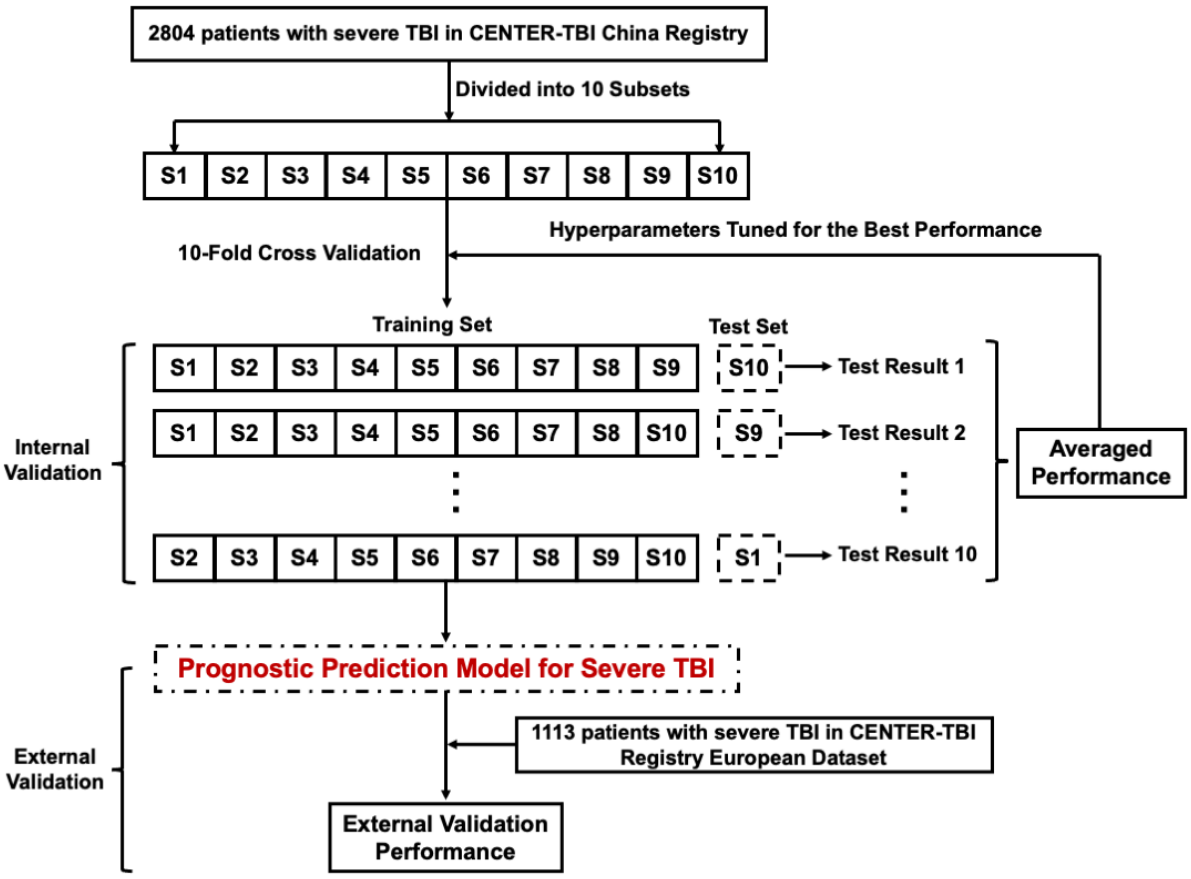

Figure 1. Overall view of training and validation of prognonstic prediction medel for sTBI.

2804 samples were divided into 10 subsets and used to do 10-fold cross validation, hyperparameters were tuned base on the internal validation for the best performance. Then the model was externally validated in 1113 samples.

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

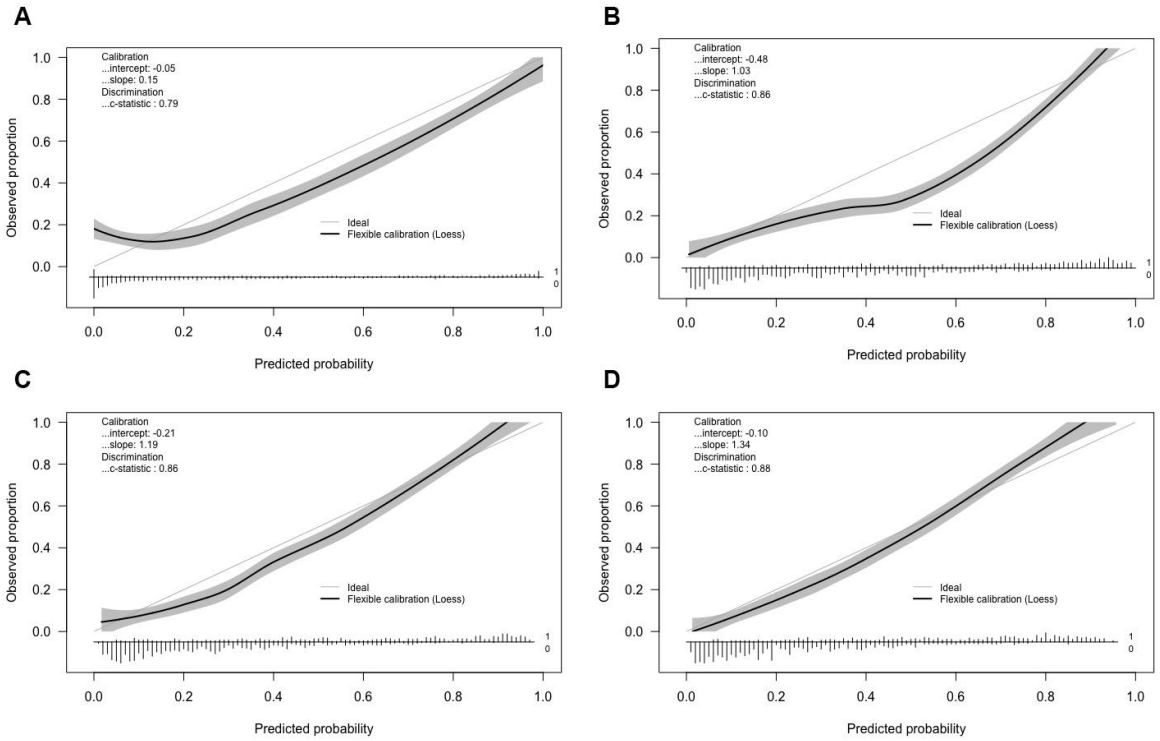

Figure 2. Calibration plot of external validation in Logistic regression (A), LASSO regression (B), SVM (C) and XGBoost (D) for 54 variables. XGBoost outperformed the other three models, and achieved a c-statistic of 0.88 (95% CI: 0.87-0.88) in the external validation set.

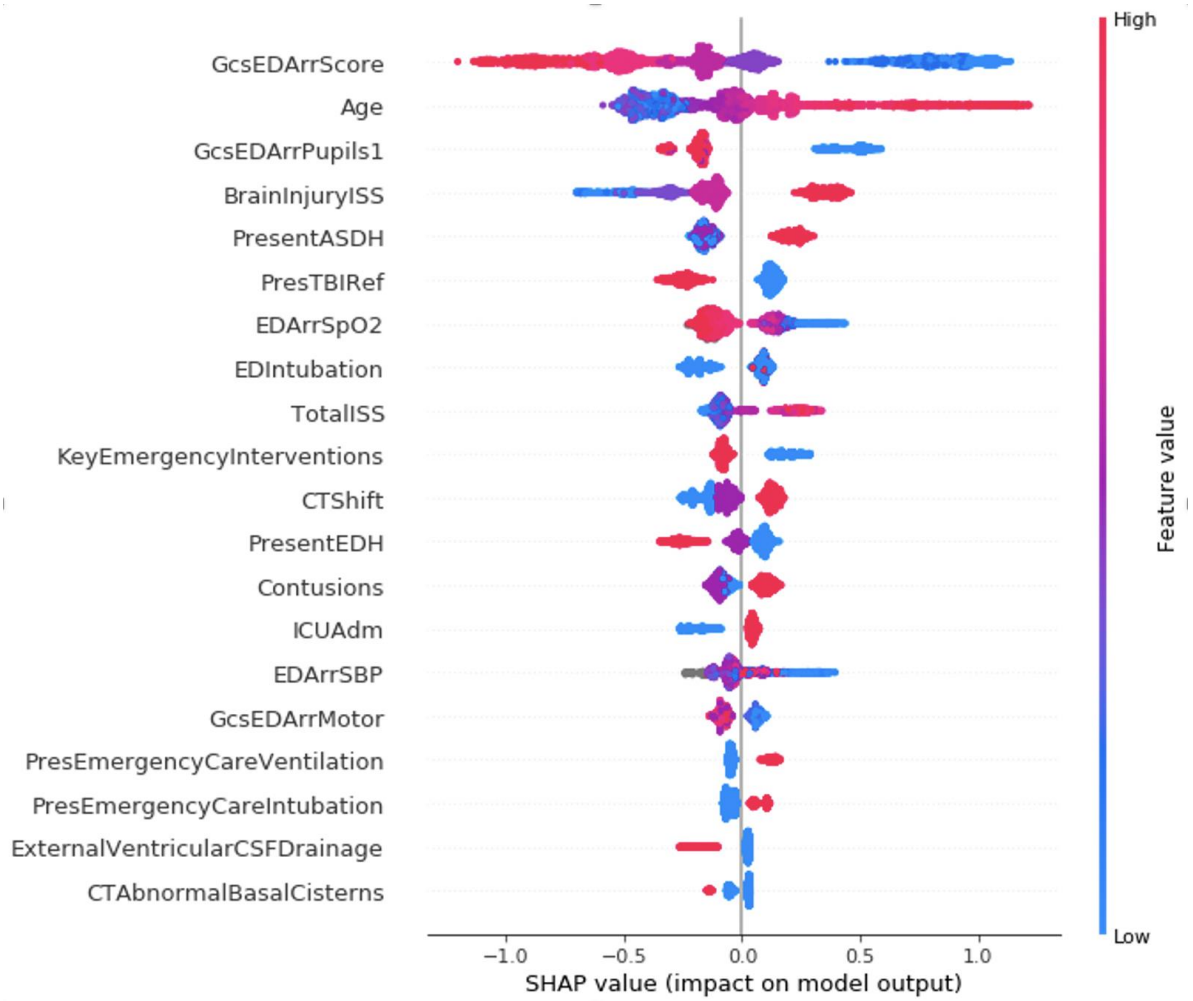

Figure 3. Impact of features in the XGBoost model for sTBI mortality prediction using Shapley Additive exPlanations (SHAP). The SHAP values were derived from the results of internal validation. The 5 strongest predictors for mortality were: low GCS Score, elder age, absent pupillary light reflex, high ISS for brain region, and presence of acute subdural hematoma.

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

Journal of Neurotrauma
Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms (DOI: 10.1089/neu.2022.0221)

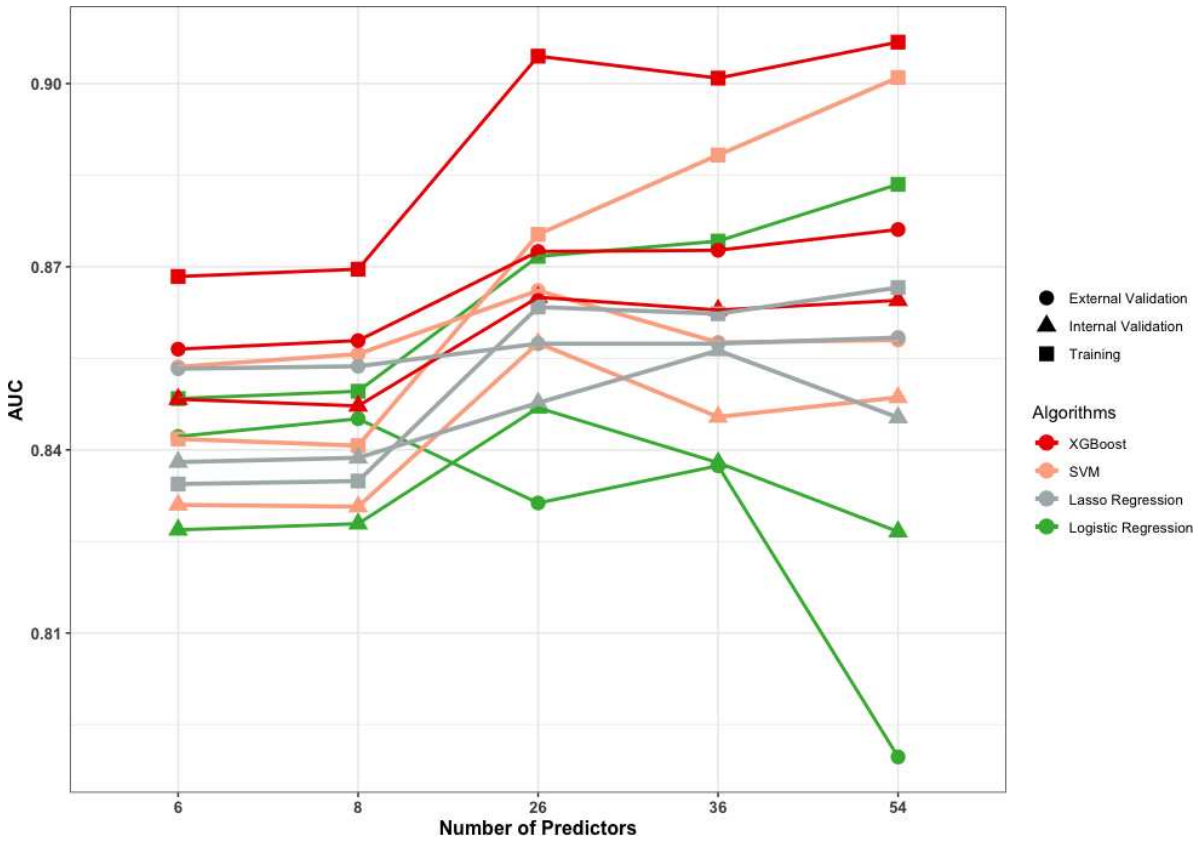

Figure 4. Performance comparison based on AUC of different algorithms when including different amount of predictors. XGBoost showed the best performance in discriminating hospital mortality in training set, internal validation set, and external validation set throughout different numbers of variables included.