BMJ Open

Development and validation of a prehospital-stage prediction tool for traumatic brain injury: a multicentre retrospective cohort study in Korea

Yeongho Choi,[1,2] Jeong Ho Park ,[1,2] Ki Jeong Hong,[1,2] Young Sun Ro ,[1,2] Kyoung Jun Song,[1,3] Sang Do Shin[1,2]

To cite: Choi Y, Park JH, Hong KJ, *et al*. Development and validation of a prehospital-stage prediction tool for traumatic brain injury: a multicentre retrospective cohort study in Korea. *BMJ Open* 2022;**12**:e055918. doi:10.1136/bmjopen-2021-055918

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-055918).

Received 02 August 2021
Accepted 20 December 2021

[1]Laboratory of Emergency Medical Services, Seoul National University Hospital Biomedical Research Institute, Seoul, South Korea
[2]Department of Emergency Medicine, Seoul National University Hospital, Seoul, South Korea
[3]Department of Emergency Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, South Korea

Correspondence to
Jeong Ho Park;
timthe@gmail.com

ABSTRACT

Objectives Predicting diagnosis and prognosis of traumatic brain injury (TBI) at the prehospital stage is challenging; however, using comprehensive prehospital information and machine learning may improve the performance of the predictive model. We developed and tested predictive models for TBI that use machine learning algorithms using information that can be obtained in the prehospital stage.

Design This was a multicentre retrospective study.

Setting and participants This study was conducted at three tertiary academic emergency departments (EDs) located in an urban area of South Korea. The data from adult patients with severe trauma who were assessed by emergency medical service providers and transported to three participating hospitals between 2014 to 2018 were analysed.

Results We developed and tested five machine learning algorithms—logistic regression analyses, extreme gradient boosting, support vector machine, random forest and elastic net (EN)—to predict TBI, TBI with intracranial haemorrhage or injury (TBI-I), TBI with ED or admission result of admission or transferred (TBI with non-discharge (TBI-ND)) and TBI with ED or admission result of death (TBI-D). A total of 1169 patients were included in the final analysis, and the proportions of TBI, TBI-I, TBI-ND and TBI-D were 24.0%, 21.5%, 21.3% and 3.7%, respectively. The EN model yielded an area under receiver–operator curve of 0.799 for TBI, 0.844 for TBI-I, 0.811 for TBI-ND and 0.871 for TBI-D. The EN model also yielded the highest specificity and significant reclassification improvement. Variables related to loss of consciousness, Glasgow Coma Scale and light reflex were the three most important variables to predict all outcomes.

Conclusion Our results inform the diagnosis and prognosis of TBI. Machine learning models resulted in significant performance improvement over that with logistic regression analyses, and the best performing model was EN.

Strengths and limitations of this study

- ► This study presented prehospital factors that could predict traumatic brain injury in patients with trauma chosen by model-specific metrics.
- ► We treated the missing variables as a different category, reflecting prehospital field uncertainties and increasing data utilisation.
- ► The retrospective observational study design could lead to certain types of bias (eg, selection bias, confounding bias).
- ► External validation for other areas should be conducted to generalise the developed prediction model.

INTRODUCTION

Traumatic brain injury (TBI) is a significant health burden worldwide.[1] It is the leading cause of mortality and disability among young individuals.[2] Patients with TBI are vulnerable to hypoxia and hypotension in the early period of their course, and these insults are associated with poor outcomes.[3 4] Prehospital assessment and management of patients with TBI are important,[5] as early prediction of TBI and correcting hypoxia and hypotension during the prehospital stage could be beneficial.[3] However, the identification of TBI can often be challenging in the prehospital area.[5] Vulnerable patients, including the elderly or patients who take medications like antiplatelet or anticoagulant drugs, often have TBI owing to low energy insults.[6] Prehospital clinical signs are also reported to have poor sensitivity for raised intracranial pressure following TBI.[7]

Several prediction models to target patients with TBI have been reported.[8–12] However, most incorporated information is available only in the hospital, such as laboratory results or image findings.[8 9 13] In addition, most previous prediction models focused on the outcomes of patients with TBI,[14–16] not the identification of TBI. Previously, predictors of older adult patients with TBI who required transport to a trauma centre were identified. However, this was consensus based; therefore, there is a lack of clinical data.[17]

Accurate prehospital prediction of TBI and its severity could prevent delays to definite care for patients with TBI. Most emergency medical service (EMS) providers collect various information including demographics, previous medical history, circumstances of the trauma and clinical signs including vital signs; but those variables have not been evaluated together as predictors of TBI and its severity. Using a variety of prehospital information and adapting newly emerging machine learning algorithms for predicting diagnosis, disposition and outcome of TBI might improve the accuracy of identification of TBI and its severity.

The aim of this study was to develop and test prediction models for the diagnosis and prognosis of TBI using prehospital information and machine learning algorithms among patients with severe trauma. We hypothesised that incorporating prehospital information could achieve acceptable performance in predicting TBI, and machine learning algorithms could contribute to performance improvement.

MATERIALS AND METHODS

Study design and settings

This was a multicentre retrospective study conducted at three tertiary academic emergency departments (EDs) located in an urban area (Seoul and Bundang) of South Korea. These EDs received 50 000–90 000 visits annually and are not designated trauma centres. We adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement on reporting predictive models.[18]

The EMS system in South Korea is operated by the National Fire Agency. The EMS level is considered intermediate, as EMS providers can perform bleeding control, spinal motion restriction, immobilisation and splintage, advanced airway management and administer fluid intravenously. As only physicians can declare death in South Korea, EMS providers cannot stop resuscitation and must transport all patients including those in cardiac arrest to the ED. For all EMS transport, EMS providers record an ambulance run sheet by law. Since 2012, the National Fire Agency adapted the US Centres for Disease Control and Prevention of the US field triage decision scheme to evaluate patients with trauma,[19] and they developed an EMS severe trauma in-depth registry. For said patients, EMS providers evaluate whether patients met trauma centre transport criteria in the field triage decision scheme. If they did, the in-depth registry should be recorded, and EMS transport protocol recommends that patients are transferred to a near-regional trauma centre; but it is not mandatory.

The Ministry of Health and Welfare designated three ED levels according to the resources and functional requirements; level 1 (n=36) and level 2 (n=118) EDs have more resources and better facilities for emergency care and must be staffed by emergency physicians 24 hours a day/365 days a year, whereas level 3 EDs (n=248) can be staffed by general physicians. In accordance with the EMS Act, all EDs participated annually in a nationwide functional performance evaluation programme, which was administered by the Ministry of Health and Welfare. The three participating hospitals in this study were all level 1 EDs that can perform acute trauma care for patients with TBI 24 hours a day/365 days a year—including emergency neurosurgical operation and angiographic interventions. The Ministry of Health and Welfare also designated trauma centres in Korea. Total 16 trauma centres were designated as trauma centres in 2018. Among them, 15 were level I EDs.

Data source

We used an EMS ambulance run sheet, EMS trauma in-depth registry and ED administrative database. The EMS database information, including ambulance run sheet and trauma in-depth registry, was collected electronically by EMS providers using tablets. The EMS record review for each severe trauma has been performed by EMS medical directors of each fire department since 2012. The ED administrative database contains patients' demographic characteristics, route of visit, time of visit and diagnosis and disposition. We merged the EMS database with the ED administrative database based on patients' arrival time, age and sex.

Study population

We included adult (age ≥15) EMS users who were transported to participating hospitals with severe trauma from 1 January 2014 to 31 December 2018. Severe trauma was assessed by EMS providers and defined as patients who fulfilled trauma centre transport criteria (physiologic criteria, anatomic criteria, mechanism of injury criteria or special patients or system consideration criteria) in the field triage decision scheme.[20] Patients were excluded if they had out-of-hospital cardiac arrest or their main cause of EMS call was medical or non-traumatic injury, including choking, drowning, fire, flame, heat, cold, poisoning, chemical, sexual assault, weather or natural disaster. Patients with an unknown outcome were also excluded.

Outcome measure

The primary outcome measure was the diagnosis of TBI. TBI diagnosis was defined as patients whose diagnostic code, according to the International Statistical Classification of Diseases and Related Health Problems (ICD-10), was between S06.0 and S06.9.[21 22] Although S06.7 is coded for the duration of unconscious, we included S06.7 in our study outcome according to the previous studies.[21–23] However, no patients only have S06.7 code for TBI diagnosis in our study. The ED administrative database has two types of primary diagnostic codes: the final diagnostic codes at ED discharge and at hospital discharge. We extracted up to 20 codes for each. We defined the diagnostic code as positive for TBI if a confirmative diagnostic code was found in any level of the discharge

record. Because ICD 10 code is not directly linked to the severity of TBI, we further included a variety of additional outcome measures to perform analysis that takes into account severity. A secondary outcome measure was TBI diagnosis with intracranial haemorrhage or injury (TBI-I), defined as patients with TBI excluding concussion (ICD 10 code with S06.0). A tertiary outcome was TBI with non-discharge (TBI-ND), defined as patients with TBI excluding ED discharged patients. Because patients with TBI-ND needed further management by hospitalisation or transfer, we thought that this group of patients had clinically significant severity. A quaternary outcome measure was TBI with death (TBI-D), defined as patients with TBI who died in ED or hospital. Because patients with TBI-D are most severe group, patients with TBI-D were also included in TBI-ND.

Variables and preprocessing

We collected patients' demographic data, circumstances of trauma, chief complaints, EMS vital sign assessment, EMS management and hospital outcomes. The detailed descriptions of each variable are described in online supplemental table 1. Categorical variables were preprocessed with the one-hot encoding (dummy variable encoding) method. Continuous variables were divided into four quantiles, and unknown or missing values were categorised as a fifth category. One hot encoding was also applied to discretised continuous variables. Preprocessing measures including discretisation results of continuous variables are presented in online supplemental table 1.

Model development

We developed prediction models for outcomes by using five machine learning algorithms: traditional logistic regression (LR) analyses, extreme gradient boost (XGB), random forest (RF), support vector machine (SVM) and elastic net (EN). The LR algorithm was chosen as baseline comparison algorithm because it is widely used in the medical field and has been used for previous prediction model development in TBI studies.[12] Backward stepwise LR was selected for feature selection, and we used the default parameter of stepAIC function from MASS package (V.7.3–53.1) in R for the selection. The other four algorithms were selected based on their ability to model non-linear associations, their relative ease of implementation and their general acceptance in the machine learning community.[24–26] All algorithms have a method to calculate the probability of the outcome occurring and algorithms other than LR need hyperparameter tuning for proper training and prediction.

The study population was split into training cohorts that included development, validation and test cohorts. The development cohort included a training cohort from which each of the machine learning prediction models were derived and a validation cohort in which the prediction models were applied to adjust the hyperparameters of the algorithm. The test cohort was used for the final evaluation of the performance of the prediction models. Chronological split was used for data split. Patients enrolled from 1 January 2014 to 31 December 2016 were used as the training cohort; patients from 1 January 2017 to 31 December 2017 were used as the validation cohort and patients from 1 January 2018 to 31 December 2018 were used as the test cohort. Hyperparameter tuning using validation data was conducted by, first, a random search within 10 000 randomly generated hyperparameters; then, grid search hyperparameters were chosen from random search with five candidates per each hyperparameter. Finally, hyperparameter with best area under receiver–operation curve (AUROC) in validation cohorts were selected. Test data were separated during training and tuning processes and used to measure algorithm performance.

Statistical analysis

The demographic findings and outcomes of the study population were described in this study. Additionally, the baseline characteristics of the training cohort and the validation cohort were compared. The continuous variables were compared by using Student's t test or the Wilcoxon rank sum test, and the categorical variables were compared by using the χ^2 test or the Fisher exact test, as appropriate.

We assessed discrimination performance by comparing the AUROC for each model in the test cohort. We considered an AUROC of 0.5 as no discrimination, 0.7–0.8 as acceptable, 0.8–0.9 as excellent and more than 0.9 is considered outstanding.[27] Area under the precision-recall curve (AUPRC) was assessed for each model in the test cohort. We assessed the calibration power by using the Hosmer–Lemeshow test, the scaled Brier score and a calibration plot in the test cohort. For the delineation of test characteristics, the sensitivity, specificity and positive and negative predictive values with 95% CIs were determined using a cut-off probability at a sensitivity of 80%. Given that poor sensitivity of clinical predictors for TBI in previous studies,[7] and almost 75% sensitivity level for other severe disease prediction in prehospital settings,[28 29] we thought that 80% sensitivity was an appropriate target for our prediction model. We calculated false-positive rate as 1—specificity. The added prognostic power of each prediction model compared with the LR model was also evaluated by continuous net reclassification index (NRI). NRI is a statistical method to quantify how well a new model correctly reclassifies the study population with the other models. Details of NRI are described elsewhere.[30]

By using a model-specific metric, the variable importance of each model was assessed, except for the SVM algorithm. The variable importance was determined by the coefficient effect sizes for the LR model. The XGB and RF models were ranked by variable importance on the selection frequency of the variable as a decision node. The absolute value of the coefficients corresponding to the tuned model were used for the measurement of variable importance in the EN algorithm. To compare the

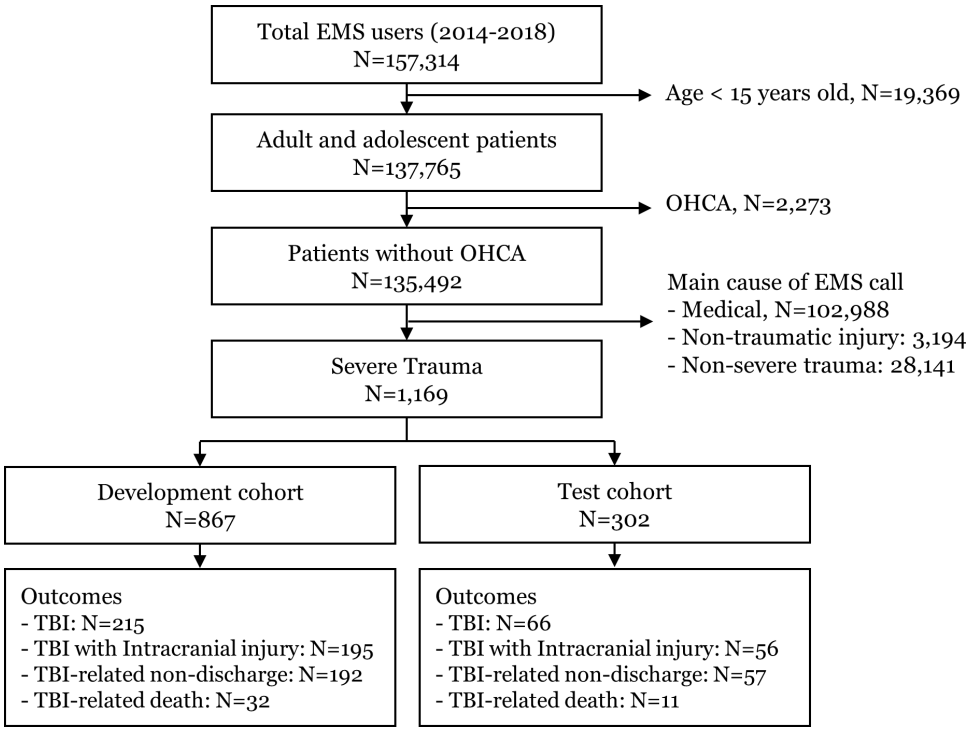

Figure 1 Population flow. EMS, emergency medical service; OHCA, out-of-hospital cardiac arrest; TBI, traumatic brain injury.

variable importance of each prediction model efficiently, top five variables of each model were presented.

All statistical analyses were performed with R Statistical Software (V.4.0.1; R Foundation for Statistical Computing, Vienna, Austria). Packages included caret, e1071, xgboost, randomForest and glmnet for the analysis of the machine learning algorithms.

Patient and public involvement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

RESULT

Demographic findings

Among the 157 134 EMS users transported to three hospitals from 2014 to 2018, 1169 patients were included in the final analysis (figure 1). Patients were split into two data sets: data from 2014 to 2017, consisting of 867 patients (74.2%) in the development cohort and the remaining data from 2018 consisting of 302 patients (25.8%) in the test cohort (figure 1). Among the development cohort, data from 2014 to 2016—consisting of 661 patients—were used as the training cohort, and 2017 data—consisting of 206 patients—were used as the validation cohort in the model.

Table 1 shows key demographic findings of the development and test cohorts. Median (IQR) age was 52 years (35–66) in the development cohort and 56 years (40–69) in the test cohort. Traffic accident was most common mechanism of trauma (43.3% for the development cohort and 41.4% for the test cohort). The proportion of patients with alert mental status was 58.1% for the development cohort and 69.5% in the test cohort. Overall, TBI, TBI-I, TBI-ND, TBI-D occurred in 215 (24.8%), 195 (22.5%), 192 (22.1%) and 32 (3.7%) in the development cohort; and 66 (21.9%), 56 (18.5%), 57 (18.9%) and 11 (3.6%) in the test cohort. All demographic characteristics of the development and test cohorts are described in online supplemental table 2.

Main analysis

The final hyperparameters of prediction models are described in online supplemental table 3. The discrimination and NRI of the prediction models on the test cohort are presented in table 2. The AUROC for outcomes was 0.770–0.809 for TBI, 0.812–0.844 for TBI-I, 0.767–0.811 for TBI-ND and 0.664–0.889 for TBI-D (table 2 and online supplemental figure 1). Compared with LR, XGB performed significantly well in predicting TBI, and RF and EN performed well in predicting TBI-ND and TBI-D. EN model generally performed well on all outcomes. The AUROC of the EN model for outcomes was 0.799 (95% CI 0.732 to 0.867), 0.844 (95% CI 0.779 to 0.910), 0.811 (95% CI 0.741 to 0.882) and 0.871 (95% CI 0.764 to 0.978) for TBI, TBI-I, TBI-ND and TBI-D, respectively. Machine learning models generally resulted in significant reclassification improvement compared with LR for TBI, TBI-I and TBI-ND. Prediction of TBI-D, AUROC difference and reclassification improvement compared with LR was non-significant in all machine learning models. The

Table 1 Key characteristics of the development and test cohorts

	n (%) or median (IQR)			
	Total	Development cohort	Test cohort	P
Total	N=1169	n=867	n=302	
Demographics				
Age, years	53 (36–66)	52 (35–66)	56 (40–69)	<0.01
Male	809 (69.2)	592 (68.3)	217 (71.9)	0.25
Job, unemployed	299 (25.6)	197 (22.7)	102 (33.8)	<0.01
Diabetes	62 (5.3)	35 (4.0)	27 (8.9)	<0.01
Hypertension	105 (9.0)	61 (7.0)	44 (14.6)	<0.01
Circumstances of trauma				
Location, road/highway	444 (38.0)	326 (37.6)	118 (39.1)	0.65
Season, summer	336 (28.7)	253 (29.2)	83 (27.5)	0.57
Weekday, weekend	811 (69.4)	599 (69.1)	212 (70.2)	0.72
Time, 18:00 to midnight	361 (30.9)	265 (30.6)	96 (31.8)	0.69
Mechanism of injury, TA	500 (42.8)	375 (43.3)	125 (41.4)	0.57
Chief complaint				
Fracture/abrasion/laceration	302 (25.8)	204 (23.5)	98 (32.5)	<0.01
EMS vital sign assessment				
SBP, mm Hg	130 (109–150)	130 (104–146)	131 (115–150)	<0.01
DBP, mm Hg	80 (70–91)	80 (69–90)	80 (70–92)	0.21
RR, /min	18 (16–20)	18 (16–20)	18 (16–20)	0.33
HR, /min	86 (75–99)	86 (74–99)	86 (76–100)	0.40
SpO_2, %	98 (95–99)	98 (95–99)	98 (96–99)	0.67
AVPU scale, alert	714 (61.1)	504 (58.1)	210 (69.5)	<0.01
EMS management				
Intravenous route	176 (15.1)	129 (14.9)	47 (15.6)	0.77
Haemorrhage control	586 (50.1)	426 (49.1)	160 (53.0)	0.25
Spinal motion restriction	811 (69.4)	606 (69.9)	205 (67.9)	0.51
Oxygen supply	233 (19.9)	176 (20.3)	57 (18.9)	0.59
In-hospital mortality	90 (7.7)	74 (8.5)	16 (5.3)	0.07
Outcomes				
TBI	281 (24.0)	215 (24.8)	66 (21.9)	0.30
TBI with intracranial injury	251 (21.5)	195 (22.5)	56 (18.5)	0.15
TBI-related non-discharge	249 (21.3)	192 (22.1)	57 (18.9)	0.23
TBI-related death	43 (3.7)	32 (3.7)	11 (3.6)	0.95

AVPU, mental status in alert, verbal, pain, and unresponsive scale; DBP, diastolic blood pressure; EMS, emergency medical service; RR, respiratory rate; SBP, systolic blood pressure; TA, traffic accident; TBI, traumatic brain injury.

precision-recall curve is shown in online supplemental figure 2. AUPRC was 0.479–0.564 for TBI, 0.469–0.606 for TBI-I, 0.477–0.551 for TBI-ND and 0.094–0.293 for TBI-D. EN model showed highest AUPRC among all prediction models. Online supplemental figure 3 shows the calibration plot of prediction models according to outcomes. All prediction models generally showed poor calibration. Given the high AUROC and AUPRC among prediction models, and reclassification improvement compared with LR, we determined EN as a best-performing prediction model in our analysis.

Using cut-off of 80% sensitivity, specificity was 47.5%–72.5% for TBI, 71.1%–81.3% for TBI-I, 46.1%–74.3% for TBI-ND and 42.6%–79.0% for TBI-D. EN showed the highest specificity and PPV among all outcomes. False-positive rate (1—specificity) was almost 19.7%–39.0% according to outcomes in the EN model. The 95% CI of specificity of the EN model was not overlapped with LR

Table 2 Discrimination and reclassification of prediction models for outcomes on test cohort

Outcome	Model	AUROC (95% CI)	p*	NRI (95% CI)	p†	AUPRC
TBI						
	LR	0.770 (0.698 to 0.841)	NA	NA	NA	0.492
	XGB	0.809 (0.743 to 0.876)	0.04	0.689 (0.427 to 0.951)	<0.01	0.552
	SVM	0.776 (0.708 to 0.844)	0.77	0.339 (0.072 to 0.607)	0.01	0.479
	RF	0.800 (0.735 to 0.865)	0.13	0.308 (0.047 to 0.569)	0.02	0.532
	EN	0.799 (0.732 to 0.867)	0.06	0.698 (0.441 to 0.954)	<0.01	0.564
TBI-I						
	LR	0.820 (0.751 to 0.890)	NA	NA	NA	0.551
	XGB	0.838 (0.775 to 0.901)	0.28	0.539 (0.258 to 0.821)	<0.01	0.554
	SVM	0.812 (0.748 to 0.875)	0.66	0.729 (0.464 to 0.994)	<0.01	0.469
	RF	0.836 (0.772 to 0.899)	0.38	0.333 (0.058 to 0.607)	0.02	0.552
	EN	0.844 (0.779 to 0.910)	0.15	1.093 (0.845 to 1.342)	<0.01	0.606
TBI-ND						
	LR	0.767 (0.690 to 0.844)	NA	NA	NA	0.482
	XGB	0.800 (0.727 to 0.873)	0.07	0.605 (0.326 to 0.884)	<0.01	0.496
	SVM	0.778 (0.704 to 0.852)	0.56	0.285 (–0.001 to 0.572)	0.05	0.477
	RF	0.809 (0.739 to 0.880)	0.03	0.194 (–0.059 to 0.448)	0.13	0.535
	EN	0.811 (0.741 to 0.882)	0.02	0.768 (0.496 to 1.039)	<0.01	0.551
TBI-D						
	LR	0.664 (0.490 to 0.838)	NA	NA	NA	0.138
	XGB	0.714 (0.512 to 0.917)	0.64	–0.026 (–0.605 to 0.553)	0.93	0.094
	SVM	0.814 (0.718 to 0.910)	0.09	0.209 (–0.325 to 0.742)	0.44	0.140
	RF	0.889 (0.801 to 0.976)	<0.01	–0.204 (–0.742 to 0.334)	0.46	0.196
	EN	0.871 (0.764 to 0.978)	0.01	0.119 (–0.415 to 0.654)	0.66	0.293

*Comparing the AUROC and the logistic regression model.
†Comparing the NRI and the logistic regression model.
AUPRC, area under precision-recall curve; AUROC, area under the receiver operating characteristic curve; EN, elastic net; LR, logistic regression analysis; NRI, net reclassification index; RF, random forest; SVM, support vector machine; TBI, traumatic brain injury; TBI-D, traumatic brain injury with death; TBI-I, traumatic brain injury with intracranial injury; TBI-ND, traumatic brain injury with non-discharge; XGB, extreme gradient boosting.

in TBI, TBI-ND and TBI-D predictions. NPV was almost 89%–99% for all outcomes in the prediction models (table 3).

Table 4 shows the top five variable importance of prediction models according to outcomes. Variables related to patients' symptom of loss of consciousness, Glasgow Coma Scale component and light reflex were the three most important variables to predict all outcomes. Compared with other outcomes, the difference between variable importance for TBI-D was prominent, and the mechanism of injury, heart rate and age showed the highest importance for predicting TBI-D.

DISCUSSION

By using prehospital data from EMS users visiting three teaching hospitals, we developed and validated prediction models for the diagnosis and prognosis of TBI using machine learning algorithms among patients with severe trauma, identified by EMS providers in South Korea. We found that 24% of patients were diagnosed with TBI, 22% showed intracranial injury, 21% could not be discharged from the ED with a TBI diagnosis and 4% showed TBI-related death. Machine learning models showed acceptable-to-excellent discrimination performance (AUROCs were 0.799–0.871 according to outcomes in the best-performing EN model). When identifying 80% of target patients with TBI, the false-positive rate was almost 19.7%–39.0%. Consciousness status-related variables ranging from patients' symptom to EMS providers' assessment showed the highest importance for predicting all outcomes. This study adds considerably to the understanding of prehospital prediction performance of TBI among patients with severe trauma. Use of comprehensive prehospital information and certain machine learning approaches led to increased performance with a diminished false-positive rate compared with those of the traditional statistical model.

Table 3 Test characteristics of prediction models for outcomes on test cohort

Outcome	Model	Specificity (95% CI)	Sensitivity (95% CI)	PPV (95% CI)	NPV (95% CI)	Cut-off
TBI						
	LR	47.5 (40.9 to 54.0)	80.3 (68.7 to 89.1)	29.9 (23.3 to 37.3)	89.6 (82.9 to 94.3)	0.136
	XGB	72.5 (66.3 to 78.1)	80.3 (68.7 to 89.1)	44.9 (35.7 to 54.3)	92.9 (88.2 to 96.2)	0.268
	SVM	64.8 (58.4 to 70.9)	80.3 (68.7 to 89.1)	39.0 (30.7 to 47.7)	92.2 (87.0 to 95.8)	0.191
	RF	68.2 (61.9 to 74.1)	80.3 (68.7 to 89.1)	41.4 (32.8 to 50.4)	92.5 (87.6 to 96.0)	0.185
	EN	61.0 (54.5 to 67.3)	80.3 (68.7 to 89.1)	36.6 (28.7 to 44.9)	91.7 (86.3 to 95.5)	0.205
TBI-I						
	LR	71.1 (65.0 to 76.7)	80.4 (67.6 to 89.8)	38.8 (29.9 to 48.3)	94.1 (89.7 to 97.0)	0.164
	XGB	74.0 (68.0 to 79.4)	80.4 (67.6 to 89.8)	41.3 (31.9 to 51.1)	94.3 (90.0 to 97.1)	0.143
	SVM	71.1 (65.0 to 76.7)	80.4 (67.6 to 89.8)	38.8 (29.9 to 48.3)	94.1 (89.7 to 97.0)	0.172
	RF	76.0 (70.2 to 81.2)	80.4 (67.6 to 89.8)	43.3 (33.6 to 53.3)	94.4 (90.3 to 97.2)	0.205
	EN	81.3 (75.9 to 86.0)	80.4 (67.6 to 89.8)	49.5 (38.8 to 60.1)	94.8 (90.9 to 97.4)	0.204
TBI-ND						
	LR	46.1 (39.8 to 52.6)	80.7 (68.1 to 90.0)	25.8 (19.6 to 32.9)	91.1 (84.7 to 95.5)	0.090
	XGB	66.5 (60.2 to 72.4)	80.7 (68.1 to 90.0)	35.9 (27.7 to 44.9)	93.7 (89.0 to 96.8)	0.242
	SVM	59.2 (52.7 to 65.4)	80.7 (68.1 to 90.0)	31.5 (24.1 to 39.7)	92.9 (87.7 to 96.4)	0.147
	RF	60.4 (54.0 to 66.6)	80.7 (68.1 to 90.0)	32.2 (24.6 to 40.5)	93.1 (88.0 to 96.5)	0.138
	EN	74.3 (68.3 to 79.6)	80.7 (68.1 to 90.0)	42.2 (32.8 to 52.0)	94.3 (90.0 to 97.1)	0.201
TBI-D						
	LR	42.6 (36.9 to 48.5)	81.8 (48.2 to 97.7)	5.1 (2.4 to 9.5)	98.4 (94.4 to 99.8)	0.005
	XGB	57.7 (51.8 to 63.5)	81.8 (48.2 to 97.7)	6.8 (3.2 to 12.5)	98.8 (95.8 to 99.9)	0.002
	SVM	74.2 (68.8 to 79.2)	81.8 (48.2 to 97.7)	10.7 (5.0 to 19.4)	99.1 (96.7 to 99.9)	0.039
	RF	74.9 (69.5 to 79.8)	81.8 (48.2 to 97.7)	11.0 (5.1 to 19.8)	99.1 (96.8 to 99.9)	0.005
	EN	79.0 (73.9 to 83.6)	81.8 (48.2 to 97.7)	12.9 (6.1 to 23.0)	99.1 (96.9 to 99.9)	0.033

EN, elastic net; LR, logistic regression analysis; RF, random forest; SVM, support vector machine; TBI, traumatic brain injury; TBI-D, traumatic brain injury with death; TBI-I, traumatic brain injury with intracranial injury; TBI-ND, traumatic brain injury with non-discharge; XGB, extreme gradient boosting.

Several studies reported that EMS providers' assessment using prehospital information is effective for the identification of patients with severe trauma who require direct transport to a trauma centre.[31–33] Because TBI accounts for a significant portion of patients with severe trauma,[32] and the majority of patients have poor access to trauma centres,[34] identification of TBI among patients with severe trauma by EMS providers could contribute to proper prehospital management and destination hospital decisions.[3] However, prehospital identification of TBI is challenging.[35] Prehospital clinical signs showed poor predictive performance for differentiating patients with TBI,[7] and previous prediction models related to TBI mostly focused on TBI outcomes.[8 9 13] One study reported the predictors for mild TBI with persistent symptoms, but a single-centre case–control study design and ED-based model development lack applicability to prehospital settings.[36] In this study, we developed and tested TBI prediction models that used prehospital information, and we found acceptable discrimination power for the prediction of diagnosis and prognosis of TBI. Uniquely, we incorporated various demographic variables, trauma circumstances, patients' complaints and EMS assessment information in the prediction models, and we adapted the machine learning algorithms.

When using a cut-off for 80% sensitivity for TBI detection, the false-positive rate was 19.7%–39.0% (table 2). Those false-positive rate levels are plausible for detecting severe diseases in EMS settings. A previous study reported a 26% of false-positive rate of EMS triage for myocardial infarction with a sensitivity of 74% and 50% of false-positive rate of EMS recognition of stroke in sensitivity of 74%.[28 29] Considering the prevalence of outcomes (24% in TBI, 22% in TBI-I, 21% in TBI-ND and 4% in TBI-D; table 1), there would be 16, 9, 12 and 67 false-positive patients for every 10 patients who are accurately identified as TBI, TBI-I, TBI-ND and TBI-D, respectively (online supplemental table 4). Because of the low prevalence of TBI-D, a similar specificity of the prediction model for outcomes resulted in a very low positive predictive value and a high proportion of false-positive cases, which suggested the limited applicability of prediction models for TBI-D in prehospital settings.

Consciousness-status-related variables ranging from patients' complaints to EMS assessment showed the highest importance regardless of models and outcomes

Table 4 Top five important variables for outcomes in descending order using model-specific metrics

Outcome	Rank	LR	XGB	RF	EN
TBI					
	1	Loss of consciousness	Loss of consciousness	Loss of consciousness	Loss of consciousness
	2	GCS, eye, 1	GCS, eye, 1	GCS, eye, 1	GCS, motor, 1
	3	GCS, verbal, 1	GCS, verbal, 1	GCS, verbal, 1	GCS, motor, 2
	4	Light reflex	Other mechanism	Light reflex	GCS, eye, 1
	5	GCS, motor, 1	GCS, verbal, 2	GCS, motor, 1	GCS, verbal, 1
TBI-I					
	1	Loss of consciousness	Loss of consciousness	Loss of consciousness	GCS, eye, 1
	2	GCS, eye, 1	GCS, eye, 1	GCS, eye, 1	Loss of consciousness
	3	GCS, verbal, 1	GCS, verbal, 1	GCS, verbal, 1	GCS, motor, 1
	4	Light reflex	GCS, verbal, 2	Light reflex	GCS, verbal, 1
	5	GCS, motor, 1	Other mechanism	GCS, motor, 1	Light reflex
TBI-ND					
	1	Loss of consciousness	Loss of consciousness	Loss of consciousness	Loss of consciousness
	2	GCS, eye, 1	GCS, eye, 1	GCS, eye, 1	GCS, eye, 1
	3	GCS, verbal, 1	GCS, verbal, 1	GCS, verbal, 1	GCS, motor, 1
	4	Light reflex	GCS, verbal, 2	GCS, verbal, 2	GCS, verbal, 1
	5	GCS, motor, 1	GCS, motor, 1	GCS, motor, 4	Light reflex
TBI-D					
	1	Loss of consciousness	GCS, verbal, 1	GCS, verbal, 1	GCS, motor, 2
	2	GCS, verbal, 1	Oxygen saturation<96%	Light reflex	GCS, verbal, 1
	3	GCS, eye, 1	Fall mechanism	Loss of consciousness	Loss of consciousness
	4	Light reflex	Afternoon	GCS, eye, 1	Age over 80
	5	GCS, motor, 1	Light reflex	GCS, motor, 1	HR 87–99

EN, elastic net; GCS, Glasgow coma scale; HR, heart rate; LR, logistic regression; RF, random forest; TBI, traumatic brain injury; TBI-D, traumatic brain injury with death; TBI-I, traumatic brain injury with intracranial injury; TBI-ND, traumatic brain injury with non-discharge; XGB, extreme gradient boosting.

in our study. Consciousness status is closely associated with head trauma. Head trauma can result in structural brain injury or physiological disruption of brain function, which could result in altered mental status.[37] Mental status is also associated with TBI severity[38] and its association with TBI outcomes has been reported.[8 9 13] History taking and physical examination for altered mental status are key to early diagnosis and proper management of TBI in prehospital settings.[39]

We adapted machine learning algorithms for the prediction of TBI-related outcomes and found an improvement in discrimination and an increase in specificity with the same sensitivity thresholds. However, the LR model also showed acceptable or similar performance compared with machine learning models, according to the outcomes. In clinical prediction models, a previous systematic review reported no performance benefit of the machine learning model over LR.[40] The previous study stated that machine learning models tend to show high performance with a strong signal-to-noise ratio problem like gaming, image recognition. However, clinical prediction problems often result in a poor signal-to-noise ratio.[40] If we could use unstructured data, which have strong signal-to-noise ratio like continuous vital sign monitoring data or audiovisual data for patients' appearance, machine learning models might perform better than LR models. In addition, if we analysed more patient data, the performance improvement of machine models might be elucidated.

Precise assessment in prehospital field could contribute to improved patient-related outcomes. High demand of EMS call and response, disparity in accessibility to definitive care capable hospitals according to regions[34] and the importance of timely management in acute disease care are the chief reasons behind the necessity for the accurate assessment of EMS providers. Although information acquisition and processing are quite difficult in prehospital areas, various instruments and information systems could attribute to diminish those problems. Complex data acquisition like mobile CT or other unstructured data,[41] information sharing through telemedicine[42] and decision support tools in prehospital environments[43] could contribute to the accurate assessment of EMS providers.

More information acquisition and real-time processing of those data could improve the clinical prediction models in prehospital areas, which could lead to the improvement of patients' safety and outcomes.

Our study had several limitations. First, our data were collected at three teaching hospitals in urban areas of South Korea. Therefore, external validation for other areas should be conducted to generalise the developed prediction model. Second, we used retrospective analysis of electronically collected prehospital and hospital data. There might be various information loss and missing data. We treated missing status as a separate category for our analysis;[44] however, there could be different reasons for missing data. Third, there is a possibility that the prediction model was overfitted or underfitted. The use of large number of predictors also can contribute to overfitting. To minimise this issue, we rigorously searched hyperparameters and carefully chose hyperparameters according to the performance in independent validation cohorts. Fourth, we selected our study population using trauma centre transport criteria for EMS providers in Korea. Although those criteria are based on the field triage decision scheme, which is the most widely used prehospital trauma triage protocol,[6] extrapolation to another EMS setting or general trauma patients would be limited. Fifth, Abbreviated Injury Scale codes were not used to identify our study outcome because of a lack of information. To compensate for this limitation, we further identified patients with TBI-I, TBI-ND and TBI-D to consider severity. However, different definitions of clinical severity, including ICU admission or emergency operation, might be possible.Finally, this study was performed in an intermediate-service-level EMS system. The generalisation of our study findings to different EMS settings should be made with caution.

In conclusion, we presented data on TBI among patients with severe trauma assessed by EMS providers, and our results inform the development of prediction models for the diagnosis and prognosis of TBI in our population. We used various information that can be obtained in prehospital settings and showed acceptable outcome performance. The consistent importance of consciousness-status-related variables emphasises the importance of assessment and monitoring of consciousness status in prehospital areas. Although prospective, and implementation studies are needed for TBI prediction in prehospital areas, our study outlined a novel method for the precise assessment of EMS providers using a machine learning-based prediction model. Further collection of various types of patient-related data would contribute to the enhanced performance of the clinical prediction model in prehospital settings.

Contributors YC and JHP designed and developed the study, analysed and interpreted the data, and drafted the initial manuscript. KJH, YSR, KJS and SDS were involved in the acquisition of data, the development of the research question and assisted with analysis and interpretation of data. JHP is responsible for the overall content as guarantor. All authors revised the drafts for intellectual content and edited the manuscript. All authors reviewed and approved the final draft.

Funding This study was supported by grant No. '04-2019-0680' from the Seoul National University Hospital Research Fund.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study complied with the Declaration of Helsinki, and its protocol was approved by the Institutional Review Board of the Seoul National University Hospital with a waiver of informed consent (IRB Number: E-2006-004-1128).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Not applicable.

ORCID iDs
Jeong Ho Park http://orcid.org/0000-0001-6330-9820
Young Sun Ro http://orcid.org/0000-0003-3634-9573

REFERENCES

1 Hsia RY, Markowitz AJ, Lin F, *et al*. Ten-year trends in traumatic brain injury: a retrospective cohort study of California emergency department and hospital revisits and readmissions. *BMJ Open* 2018;8:e022297.
2 Finfer SR, Cohen J. Severe traumatic brain injury. *Resuscitation* 2001;48:77–90.
3 Spaite DW, Bobrow BJ, Keim SM, *et al*. Association of statewide implementation of the prehospital traumatic brain injury treatment guidelines with patient survival following traumatic brain injury: the excellence in prehospital injury care (EPIC) study. *JAMA Surg* 2019;154:e191152.
4 McHugh GS, Engel DC, Butcher I, *et al*. Prognostic value of secondary insults in traumatic brain injury: results from the impact study. *J Neurotrauma* 2007;24:287–93.
5 Pélieu I, Kull C, Walder B. Prehospital and emergency care in adult patients with acute traumatic brain injury. *Med Sci* 2019;7. 10.3390/medsci7010012. [Epub ahead of print: 21 Jan 2019].
6 Sasser SM, Hunt RC, Faul M. Guidelines for field triage of injured patients: recommendations of the National expert panel on field triage, 2011. *Morbidity and Mortality Weekly Report: Recommendations and Reports* 2012;61:1–20.
7 Ter Avest E, Taylor S, Wilson M, *et al*. Prehospital clinical signs are a poor predictor of raised intracranial pressure following traumatic brain injury. *Emerg Med J* 2021;38:21–6.
8 MRC CRASH Trial Collaborators, Perel P, Arango M, *et al*. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ* 2008;336:425–9.
9 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:e165.
10 Gozt AK, Hellewell SC, Thorne J, *et al*. Predicting outcome following mild traumatic brain injury: protocol for the longitudinal, prospective, observational Concussion Recovery (*CREST*) cohort study. *BMJ Open* 2021;11:e046460.
11 Huth SF, Slater A, Waak M, *et al*. Predicting neurological recovery after traumatic brain injury in children: a systematic review of prognostic models. *J Neurotrauma* 2020;37:2141–9.
12 Perel P, Edwards P, Wentz R, *et al*. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak* 2006;6:38.

13 Miller PR, Chang MC, Hoth JJ, *et al*. Predicting mortality and independence at discharge in the aging traumatic brain injury population using data available at admission. *J Am Coll Surg* 2017;224:680–5.
14 Abujaber A, Fadlalla A, Gammoh D, *et al*. Prediction of in-hospital mortality in patients with post traumatic brain injury using national trauma registry and machine learning approach. *Scand J Trauma Resusc Emerg Med* 2020;28:44.
15 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.
16 Roozenbeek B, Lingsma HF, Lecky FE, *et al*. Prediction of outcome after moderate and severe traumatic brain injury: external validation of the International mission on prognosis and analysis of clinical trials (impact) and corticoid randomisation after significant head injury (crash) prognostic models. *Crit Care Med* 2012;40:1609–17.
17 Wasserman EB, Shah MN, Jones CMC, *et al*. Identification of a neurologic scale that optimizes EMS detection of older adult traumatic brain injury patients who require transport to a trauma center. *Prehosp Emerg Care* 2015;19:202–12.
18 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med* 2015;162:735–6.
19 Sasser SM, Hunt RC, Sullivent EE. *Guidelines for field triage of injured patients: recommendations of the National expert panel on field triage*, 2009.
20 Sasser SM, Hunt RC, Faul M, *et al*. Guidelines for field triage of injured patients: recommendations of the National expert panel on field triage, 2011. *MMWR Recomm Rep* 2012;61:1–20.
21 Andelic N, Anke A, Skandsen T, *et al*. Incidence of hospital-admitted severe traumatic brain injury and in-hospital fatality in Norway: a national cohort study. *Neuroepidemiology* 2012;38:259–67.
22 Ro YS, Shin SD, Holmes JF, *et al*. Comparison of clinical performance of cranial computed tomography rules in patients with minor head injury: a multicenter prospective study. *Acad Emerg Med* 2011;18:597–604.
23 Chan V, Thurairajah P, Colantonio A. Defining pediatric traumatic brain injury using International classification of diseases version 10 codes: a systematic review. *BMC Neurol* 2015;15:7.
24 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B* 2005;67:301–20.
25 Hearst MA, Dumais ST, Osuna E, *et al*. Support vector machines. *IEEE Intelligent Systems and their Applications* 1998;13:18–28.
26 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining. San Francisco, California, USA: association for computing machinery 2016:785–94.
27 Menard S. *Applied logistic regression analysis*. Sage, 2002: 162.
28 Oostema JA, Konen J, Chassee T, *et al*. Clinical predictors of accurate prehospital stroke recognition. *Stroke* 2015;46:1513–7.
29 Swan PY, Nighswonger B, Boswell GL, *et al*. Factors associated with false-positive emergency medical services triage for percutaneous coronary intervention. *West J Emerg Med* 2009;10:208–12.
30 Park JH, Shin SD, Song KJ, *et al*. Prediction of good neurological recovery after out-of-hospital cardiac arrest: a machine learning analysis. *Resuscitation* 2019;142:127–35.
31 Esposito TJ, Offner PJ, Jurkovich GJ, *et al*. Do prehospital trauma center triage criteria identify major trauma victims? *Arch Surg* 1995;130:171–6.
32 Ocak G, Sturms LM, Hoogeveen JM, *et al*. Prehospital identification of major trauma patients. *Langenbecks Arch Surg* 2009;394:285–92.
33 Fries GR, McCalla G, Levitt MA, *et al*. A prospective comparison of paramedic judgment and the trauma triage rule in the prehospital setting. *Ann Emerg Med* 1994;24:885–9.
34 Branas CC, MacKenzie EJ, Williams JC, *et al*. Access to trauma centers in the United States. *JAMA* 2005;293:2626–33.
35 Whiting MD, Dengler BA, Rodriguez CL, *et al*. Prehospital detection of life-threatening intracranial pathology: an unmet need for severe TBI in Austere, rural, and remote areas. *Front Neurol* 2020;11:599268.
36 Wojcik SM. Predicting mild traumatic brain injury patients at risk of persistent symptoms in the emergency department. *Brain Inj* 2014;28:422–30.
37 Widerström-Noga EG, Finnerup NB, Siddall PJ. Management of Concussion/m TBIWG. VA/DoD clinical practice guideline for management of Concussion/Mild traumatic brain injury. *J Rehabil Res Dev* 2009;46:01–68.
38 Grote S, Böcker W, Mutschler W, *et al*. Diagnostic value of the Glasgow coma scale for traumatic brain injury in 18,002 patients with severe multiple injuries. *J Neurotrauma* 2011;28:527–34.
39 Badjatia N, Carney N, Crocco TJ, *et al*. Guidelines for prehospital management of traumatic brain injury 2nd edition. *Prehosp Emerg Care* 2008;12:S1–52.
40 Christodoulou E, Ma J, Collins GS, *et al*. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22.
41 Nakada TA, Masunaga N, Nakao S, *et al*. Development of a prehospital vital signs chart sharing system. *Am J Emerg Med* 2016;34:88–92.
42 Kim Y, Groombridge C, Romero L, *et al*. Decision support capabilities of telemedicine in emergency prehospital care: systematic review. *J Med Internet Res* 2020;22:e18959.
43 Reisner AT, Khitrov MY, Chen L, *et al*. Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. *Appl Clin Inform* 2013;4:392–402.
44 Maslove DM, Podchiyska T, Lowe HJ. Discretization of continuous features in clinical datasets. *J Am Med Inform Assoc* 2013;20:544–53.