PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Development and validation of a prehospital-stage prediction tool for traumatic brain injury: a multicentre retrospective cohort study in Korea
AUTHORS	Choi, Yeong Ho; Park, Jeong Ho; Hong, Ki Jeong; Ro, Young Sun; Song, Kyoung Jun; Shin, Sang Do

VERSION 1 – REVIEW

REVIEWER	Patrick Czorlich University Medical Center Hamburg-Eppendorf, Department of Neurosurgery
REVIEW RETURNED	15-Aug-2021

GENERAL COMMENTS	The manuscript deals with a globally clinically relevant question, namely which patients have suffered a traumatic brain injury so that these patients can be efficiently and promptly referred to a suitable clinic for further care. The authors used the data sets of 1,169 patients from three institutions in South Korea to compare different statistical models for predicting the presence of TBI and mortality, and came to the conclusion that machine learning, here elastic nets, has the best predictive value of the patient's death. The research question is not novel and has already been described in detail in the literature, and the results are not a little surprising either. In my opinion, the literature used is incomplete, especially the referencing of TBI studies from larger registries, which dealt with the question of the presence of TBI and the risk of death in patients. The results in these studies were not much worse than those Machine learning results of this study. Here the authors should once again carry out a detailed literature search and supplement the manuscript accordingly. The manuscript will certainly benefit from it. The use of ICD10 codes is generally associated with a limitation, as this is not necessarily linked to the severity of the trauma; in trauma research, this is generally done using the Abbreviated Injury Scale (AIS) or the Injury Severity Scale (AIS). This limitation should be mentioned in detail in the discussion. The sole use of the ICD10 code S06.7x is actually not permitted, as this code only codes for the duration of the unconscious and does not represent a meaningful diagnosis on its own. I recommend to check and revise this. The classification of the various groups is inconclusive and should be revised. Page 11, lines 189-192 makes no sense to me or are described in an incomprehensible manner.

	What does "included admission, transfer, or death" mean in this context in the TBI-ND group. Patients were transferred where? To another clinic? The authors' facilities are tertiary clinics or what is meant by this. From my point of view a transfer from a tertiary medical center is unusual or is this common in South Korea? And why are deceased patients included in both the TBI-ND and TBI-D groups? From my point of view, questions such as which patient was discharged home from the emergency room, which patient had to be admitted to the hospital, which patient had to receive ICU treatment, which patient had to undergo a neurosurgical intervention, etc. are more interesting. This paragraph should be revised and made clearer be formulated. What is the definition used for "Loss of consciousness"? This is a relevant aspect of the whole analysis and manuscript? This definition is of crucial importance for the assessment of the statistical analysis, since unconsciousness in general goes hand in hand with a change in the GCS and the factors are therefore not independent and can have a decisive influence on an analysis. The most interesting number of the study from my point of view is the number of false negative patients. If such an algorithm should be used in emergency medicine, then the crucial question is that no patient with TBI is underestimated and is admitted to a smaller hospital. This number can be calculated independently from the available numbers in the manuscript, but the authors should present these numbers to the readers and discuss the finding. The results of the study should also be interpreted with caution, as the preselection of patients for this study was determined on the basis of local criteria. An interesting question would have been whether machine learning can identify patients with a traumatic brain injury between all emergencies. This should be discussed or mentioned as a limitation. Minor points to address Authors talk about adults but included patients ≥ 15 years of age. A 15year old patient is not an adult it is an adolescent. I would recommend to revise the manuscript. From my personal point of view and AUROC value between 0.8-0.9 is not excellent, it is good as described by Nelson et al., Current Clinical Neurology 2020. Please provide the page number of reference #30 were the information was captured. Please provide data on the rate of patients with a GCS ≤ 8. Why do authors present data on SBP < 90mmHg but didn´t use this definition for the analysis, instead used cut-off values defined from the cohort. In general falls are distributed in less or more than 3 meters in the literature? Why is the cut-off in this study 6m? I would be happy if the authors would reconsider some of my

	suggestions and, if necessary, accept them, because I believe that the manuscript could benefit from them.

REVIEWER	Kazuya Matsuo Kobe University Graduate School of Medicine School of Medicine, Neurosurgery
REVIEW RETURNED	03-Sep-2021

GENERAL COMMENTS	The authors attempted to predict the presence of TBI, as well as TBI with intracranial hemorrhage and TBI with death by using prehospital information. The AUROC for the prediction by machine learning models are generally above 0.8, which is not bad. Their hypothesis set as "incorporating prehospital information could achieve acceptable performance in predicting TBI, and machine learning algorithms could contribute to performance improvement" is appropriate and their results support the hypothesis. The results are clear and easy to understand. However, I would like to recommend some issues to re-consider before publication. Materials and Methods 1. Looking at Supplementary Table 1, there are 27 factors. Furthermore, the continuous variables are categorized and one-hot encoded. It seems to be too many learning parameters. Did you use all of them as learning parameters? The prediction performance might be better if you reduce the number of learning parameters by using additional feature selection. Have you tried it? 2. The continuous variables do not need to be categorized when train and test machine learning models. Is there a reason why you dared to categorize them? 3. The method of feature selection should be described in more detail. "Backward stepwise LR was selected for feature selection" is not enough to understand the method. 4. Which did you use regularized logistic regression or traditional logistic regression model? 5. Basically, representative hyperparameters of the best models should be provided. Result 6. Please explain the AVPU scale in Table 1, and consider using units other than mmHg for the RR. 7. Did you perform feature selection based on all items in Supplementary Table 2? In general, such as “High-risk auto crash” seems to be a valid predictor of TBI. Wasn't it determined to be valid during the feature selection process? 8. There appears to be a large discrepancy between the AUROC and NRI results in the TBI-D predictions. Do you have any opinion on the reason for this? Discussion 9. In the introduction, it is noted that "Prehospital clinical signs are also reported to have poor sensitivity for raised intracranial pressure following TBI." Did you examine the relationship with raised intracranial pressure and machine learning model in this study?

	10. Please discuss the potential clinical use of the model and implications for future research as required by the TRIPOD Checklist. 11. There are too many references. Reducing the number of references may help to focus the article.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Patrick Czorlich , University Medical Center Hamburg-Eppendorf

Comments to the Author:

The manuscript deals with a globally clinically relevant question, namely which patients have suffered a traumatic brain injury so that these patients can be efficiently and promptly referred to a suitable clinic for further care.

The authors used the data sets of 1,169 patients from three institutions in South Korea to compare different statistical models for predicting the presence of TBI and mortality, and came to the conclusion that machine learning, here elastic nets, has the best predictive value of the patient's death.

The research question is not novel and has already been described in detail in the literature, and the results are not a little surprising either.

In my opinion, the literature used is incomplete, especially the referencing of TBI studies from larger registries, which dealt with the question of the presence of TBI and the risk of death in patients. The results in these studies were not much worse than those Machine learning results of this study. Here the authors should once again carry out a detailed literature search and supplement the manuscript accordingly. The manuscript will certainly benefit from it.

(ANSWER) Thank you for the review and comments. We carried out a detailed literature search and added relevant and important studies including 1) two systematic reviews of prognostic models for TBI (Pablo Perel, et al, BMC Med Inform Decis Mak, 2006 and Samuel et al, Journal of Neurotrauma, 2020), 2) External validation study of IMACT and CRASH prediction models (Bob Roozenbeek et al, Critical Care Med. 2012), 3) TBI prediction study using IMPACT-II database and Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) database (Benjamin et al, J Clin Epidemiol. 2020), and 4) In-hospital mortality prediction study using NTDB dabatse (Ahmad et al, Scanadivian Journal of Trauma, Resuscitation, and Emergency medicine, 2020) in introduction.

(Revision) Introduction

Several prediction models to target patients with TBI have been reported.[13-17] However, most incorporated information that is available only in the hospital, such as laboratory results or image findings.[13 14 18] In addition, most previous prediction models focused on the outcomes of patients with TBI,[19-21] not the identification of TBI.

The use of ICD10 codes is generally associated with a limitation, as this is not necessarily linked to the severity of the trauma; in trauma research, this is generally done using the Abbreviated Injury Scale (AIS) or the Injury Severity Scale (AIS). This limitation should be mentioned in detail in the discussion. The sole use of the ICD10 code S06.7x is actually not permitted, as this code only codes for the duration of the unconscious and does not represent a meaningful diagnosis on its own. I recommend to check and revise this.

(ANSWER) Thank you for the review and comments. We added a limitation for the use of ICD 10 codes in our study in the discussion. Because ICD 10 codes is not directly linked to the severity of TBI, we further included a variety of additional outcome measure to perform analysis that takes into account severity. The TBI patients who had only S06.7x codes were not present in our study. We added description for our study population as follows.

(Revision) Methods-Outcome measure

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 take into account severity. A secondary outcome measure was TBI diagnosis with intracranial hemorrhage or injury (TBI-I), defined as TBI patients excluding concussion (ICD 10 code with S06.0). A tertiary outcome was TBI with non-discharge (TBI-ND), defined as TBI patients excluding ED discharged patients. Because TBI-ND patients needed further management by hospitalization 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 TBI patients who died in ED or hospital. Because TBI-D patients are most severe group, TBI-D patients were also included in TBI-ND.

(Revision) Discussion

Fifth, Abbreviated Injury Scale (AIS) codes were not used to identify our study outcome because of a lack of information. To compensate for this limitation, we further identified TBI-I, TBI-ND, and TBI-D patients to consider severity. However, different definitions of clinical severity, including ICU admission or emergency operation, might be possible.

(Revision) Methods-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. [22 23] Although S06.7 is codes for the duration of unconscious, we included S06.7 in our study outcome according to the previous studies. [27-29] However, no patients only have S06.7 code for TBI diagnosis in our study.

The classification of the various groups is inconclusive and should be revised. Page 11, lines 189-192 makes no sense to me or are described in an incomprehensible manner.

(ANSWER) Thank you for the review and comments. We revised the description of classification of the various groups in methods section

(Revision) Methods-Outcome measure
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 take into account severity. A secondary outcome measure was TBI diagnosis with intracranial hemorrhage or injury (TBI-I), defined as TBI patients excluding concussion (ICD 10 code with S06.0). A tertiary outcome was TBI with non-discharge (TBI-ND), defined as TBI patients excluding ED discharged patients. Because TBI-ND patients needed further management by hospitalization 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 TBI patients who died in ED or hospital. Because TBI-D patients are most severe group, TBI-D patients were also included in TBI-ND.

What does "included admission, transfer, or death" mean in this context in the TBI-ND group. Patients were transferred where? To another clinic? The authors' facilities are tertiary clinics or what is meant by this. From my point of view a transfer from a tertiary medical center is unusual or is this common in South Korea?

(ANSWER) Thank you for the review and comments. Due to lack of intensive care unit, lack of ward or unavailability of the operation room, transferring was occurred in tertiary clinics in Korea. The tertiary clinics are designated by Ministry of Health and Welfare for the general emergency care. The Ministry of Health and Welfare also designate a trauma center in Korea. In 2018, the Ministry of Health and Welfare designated 16 trauma centers, and 15 were existing tertiary clinics. The participating hospitals were not trauma center. We added description of setting in methods

(Revision) Methods-Study design and setting
This was a multi-center 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 centers.

(Revision) Methods-Study design and setting
The Ministry of Health and Welfare also designated trauma centers in Korea. Total 16 trauma centers were designated as trauma centers in 2018. Among them, 15 were Level I EDs.

And why are deceased patients included in both the TBI-ND and TBI-D groups?

(ANSWER) Thank you for the review and comments. Because our prediction models predict binary outcome, we had to divided patients into two groups: those with less severity and those with more severity. Because TBI-D patients were most severe patients in our study group, those patients were included in TBI-ND patients.

(Revision) Methods-Outcome measure
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 take into account severity. A secondary outcome measure was TBI diagnosis with intracranial hemorrhage or injury (TBI-I), defined as TBI patients excluding concussion (ICD 10 code with S06.0). A tertiary outcome was TBI with non-discharge (TBI-ND), defined as TBI patients excluding ED discharged patients. Because TBI-ND patients needed further management by hospitalization 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 TBI patients who died in ED or hospital. Because TBI-D patients are most severe group, TBI-D patients were also included in TBI-ND.

From my point of view, questions such as which patient was discharged home from the emergency room, which patient had to be admitted to the hospital, which patient had to receive ICU treatment, which patient had to undergo a neurosurgical intervention, etc. are more interesting. This paragraph should be revised and made clearer be formulated.

(ANSWER) Thank you for the review and comments. We also think the questions you suggested are important and interesting. However, because our prediction models predict a binary outcome, we had to divide patients into two groups: those with less severity and those with more severity. Because we already adapted four outcome measures, we did not further conduct analysis for ICU admission or emergency operation. We revised the paragraph, and we also added limitation for non-conducting ICU admission or emergency operation in discussion.

(Revision) Methods-Outcome measure

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 take into account severity. A secondary outcome measure was TBI diagnosis with intracranial hemorrhage or injury (TBI-I), defined as TBI patients excluding concussion (ICD 10 code with S06.0). A tertiary outcome was TBI with non-discharge (TBI-ND), defined as TBI patients excluding ED discharged patients. Because TBI-ND patients needed further management by hospitalization 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 TBI patients who died in ED or hospital. Because TBI-D patients are most severe group, TBI-D patients were also included in TBI-ND.

(Revision) Discussion

Fifth, Abbreviated Injury Scale (AIS) codes were not used to identify our study outcome because of a lack of information. To compensate for this limitation, we further identified TBI-I, TBI-ND, and TBI-D patients to consider severity. However, different definitions of clinical severity, including ICU admission or emergency operation, might be possible.

What is the definition used for "Loss of consciousness"? This is a relevant aspect of the whole analysis and manuscript? This definition is of crucial importance for the assessment of the statistical analysis, since unconsciousness in general goes hand in hand with a change in the GCS and the factors are therefore not independent and can have a decisive influence on an analysis.

(ANSWER) Thank you for the review and comments. Loss of consciousness is information collected by EMS providers whether patients had a “loss of consciousness” between injury and EMS provider’s assessment. This information is different from level of consciousness, which is evaluated by EMS providers using GCS or AVPU scales. We added further description in Supplementary Table 1.

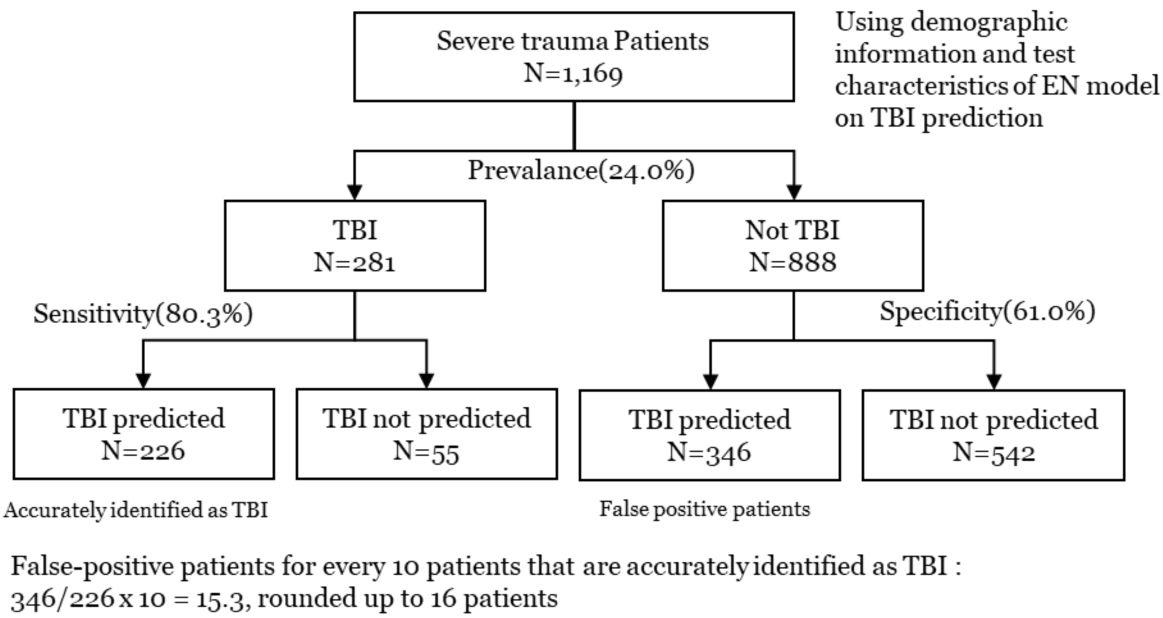
(Revision) Supplementary table 1.

Description of loss of consciousness: Symptom of loss of consciousness (whether patients had loss of consciousness between injury and EMS provider’s assessment)

The most interesting number of the study from my point of view is the number of false negative patients. If such an algorithm should be used in emergency medicine, then the crucial question is that no patient with TBI is underestimated and is admitted to a smaller hospital. This number can be calculated independently from the available numbers in the manuscript, but the authors should present these numbers to the readers and discuss the finding.

(ANSWER) Thank you for the review and comments. We added example of false positive rate calculation process in supplementary Figure 2.

(Revision) Supplementary Figure 4. Example of calculating false-positive patients for accurately identified patients. TBI, traumatic brain injury; EN, elastic net.



The results of the study should also be interpreted with caution, as the preselection of patients for this study was determined on the basis of local criteria. An interesting question would have been whether machine learning can identify patients with a traumatic brain injury between all emergencies. This should be discussed or mentioned as a limitation.

(ANSWER) Thank you for the review and comments. We added limitation in our discussion.

(Revision) Discussion

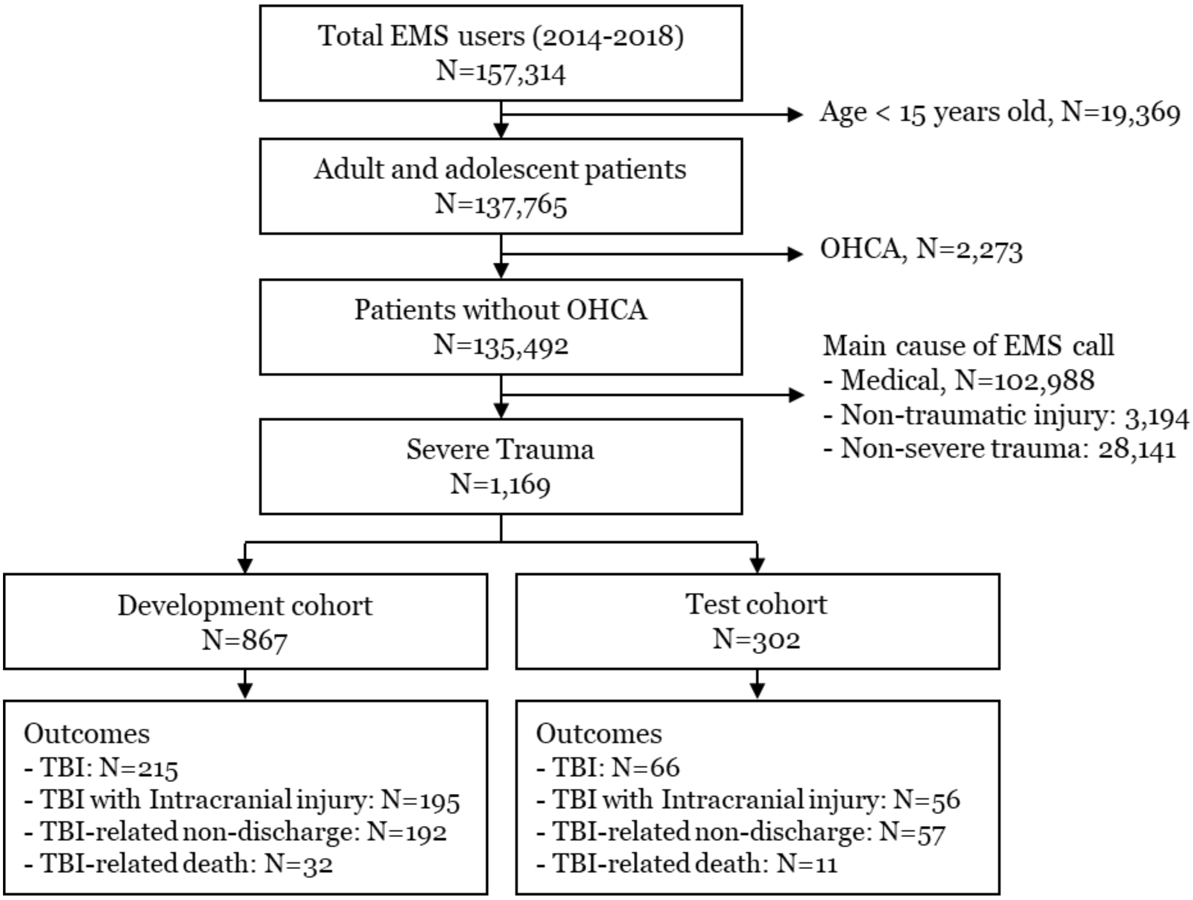
Fourth, we selected our study population using trauma center transport criteria for EMS providers in Korea. Although those criteria are based on the field triage decision scheme which is the most widely used prehospital trauma triage protocol [11], extrapolation to another EMS setting or general trauma patients would be limited.

Minor points to address

Authors talk about adults but included patients ≥ 15 years of age. A 15 year old patient is not an adult it is an adolescent. I would recommend to revise the manuscript.

(ANSWER) Thank you for the review and comments. We changed term to “adult and adolescent” in Figure 1.

(Revision) Figure 1. Population flow



From my personal point of view and AUROC value between 0.8-0.9 is not excellent, it is good as described by Nelson et al., Current Clinical Neurology 2020. Please provide the page number of reference #30 were the information was captured.

(ANSWER) Thank you for the review and comments. “Applied Logistic Regression, Second edition, Chapter Assessing the fit of the model in page 162 described following general rule. We changed s reference to include the page.

As a general rule:

If ROC = 0.5:	this suggests no discrimination (i.e., we might as well flip a coin)
If $0.7 \leq ROC < 0.8$:	this is considered acceptable discrimination
If $0.8 \leq ROC < 0.9$:	this is considered excellent discrimination
If $ROC \geq 0.9$:	this is considered outstanding discrimination.

Please provide data on the rate of patients with a GCS ≤ 8. Why do authors present data on SBP < 90mmHg but didn´t use this definition for the analysis, instead used cut-off values defined from the cohort.

(ANSWER) Thank you for the review and comments. There was 56.4% (489/867) of patients with GCS ≤ 8 in train cohort, with 17.9% (155/867) of unknown GCS. We presented SBP < 90mmHg data to show patients who met physiologic criteria for the field triage decision scheme. All continuous vital sign data were preprocessed into quantile in our analysis. We categorized continuous vital sign data to incorporate missing as another category.

In general falls are distributed in less or more than 3 meters in the literature? Why is the cut-off in this study 6m?

(ANSWER) Thank you for the review and comments. Our EMS providers only collected information of fall > 6 meters or not, because the information is one of the criteria for the field triage decision scheme. In the field triage decision scheme, cutoff of 6 meters was used for adult and cutoff of 3 meters was used for children.

Reviewer: 2

Dr. Kazuya Matsuo, Kobe University Graduate School of Medicine School of Medicine

Comments to the Author:

The authors attempted to predict the presence of TBI, as well as TBI with intracranial hemorrhage and TBI with death by using prehospital information. The AUROC for the prediction by machine learning models are generally above 0.8, which is not bad. Their hypothesis set as "incorporating prehospital information could achieve acceptable performance in predicting TBI, and machine learning algorithms could contribute to performance improvement" is appropriate and their results support the hypothesis. The results are clear and easy to understand. However, I would like to recommend some issues to re-consider before publication.

Materials and Methods

1. Looking at Supplementary Table 1, there are 27 factors. Furthermore, the continuous variables are categorized and one-hot encoded. It seems to be too many learning parameters. Did you use all of them as learning parameters? The prediction performance might be better if you reduce the number of learning parameters by using additional feature selection. Have you tried it?

(ANSWER) Thank you for the review and comments. Excessive variables could lead to overfitting, which worsens model performance on test data. We used all variables in supplementary table 1 in our analysis. We used backward stepwise selection in logistic regression to diminish variables. The elastic net also performs variable selection and regularization simultaneously. Otherwise, we did not remove variables manually from predictors. We added limitation in our discussion section

(Revision) Discussion

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 minimize this issue, we rigorously searched hyperparameters and carefully chose hyperparameters according to the performance in independent validation cohorts.

2. The continuous variables do not need to be categorized when train and test machine learning models. Is there a reason why you dared to categorize them?

(ANSWER) Thank you for the review and comments. We used categorization to incorporate cases with any missing continuous variables to the dataset.

3. The method of feature selection should be described in more detail. "Backward stepwise LR was selected for feature selection" is not enough to understand the method.

(ANSWER) Thank you for the review and comments. We added further description in methods section.

(Revision) Methods-Model development
We used the default parameter of stepAIC function from MASS package (version 7.3-53.1) in R for variable selection.

4. Which did you use regularized logistic regression or traditional logistic regression model?
(ANSWER) Thank you for the review and comments. We used traditional logistic regression model as base model. Elastic net has regularized logistic regression features among our models. We change sentences in model development as follows.

(Revision) Methods-Model development
We developed prediction models for outcomes by using five machine learning algorithms: traditional logistic regression analyses (LR), extreme gradient boost (XGB), random forest (RF), support vector machine (SVM), and elastic net (EN).

5. Basically, representative hyperparameters of the best models should be provided.

(ANSWER) Thank you for the review and comments. We added Supplementary table 3 for hyperparameters of the best model for XGB, SVM, RF, EN, respectively.

(Revision) Supplementary Table 3. Hyperparameters of the final prediction models*

Model	Outcome	Hyperparameters
Elastic net	TBI	alpha: 0.325, lambda: 0.07506346
	TBI-I	alpha: 0.325, lambda: 0.07506346
	TBI-ND	alpha: 0.325, lambda: 0.07017153
	TBI-D	alpha: 0.325, lambda: 0.01565599
Random forest	TBI	ntree:500, mtry: 18
	TBI-I	ntree:500, mtry: 18
	TBI-ND	ntree:500, mtry: 18
	TBI-D	ntree:500, mtry: 15
Support vector machine	TBI	sigma: 0.008047; C: 4
	TBI-I	sigma: 0.008047; C: 4
	TBI-ND	sigma: 0.008047; C: 4
	TBI-D	sigma: 0.008047; C: 4
Extreme gradient boosting	TBI	nrounds: 299; max_depth: 1; eta: 0.4807096; gamma: 2.336623; colsample_bytree: 0.3657893; min_child_weight: 8; subsample: 0.8182623

	TBI-I	nrounds: 299; max_depth: 1; eta: 0.4807096; gamma: 2.336623; colsample_bytree: 0.3657893; min_child_weight: 8; subsample: 0.8182623
	TBI-ND	nrounds: 301; max_depth: 1; eta: 0.02154674; gamma: 4.696105; colsample_bytree: 0.590754; min_child_weight: 1; subsample: 0.5070866
	TBI-D	nrounds: 50; max_depth: 0.3; eta: 0.3; gamma: 0; colsample_bytree: 0.8; min_child_weight: 1; subsample: 0.5510204

*Aside from the hyperparameters mentioned, all other hyperparameters are used as the default value. TBI, traumatic brain injury; TBI-I, TBI with intracranial hemorrhage or injury; TBI-ND, TBI non-discharge; TBI-D, TBI with death.

Result

6. Please explain the AVPU scale in Table 1, and consider using units other than mmHg for the RR.

(ANSWER) Thank you for the review and comments. We added the explanation about AVPU, and corrected the unit for RR to "/min"

(Revision) Table 1 foot note

IQR, interquartile range; TA, traffic accident; SBP, systolic blood pressure; DBP, diastolic blood pressure; RR, respiratory rate; AVPU, mental status in alert, verbal, pain, and unresponsive scale; ED, emergency department; TBI, traumatic brain injury.

7. Did you perform feature selection based on all items in Supplementary Table 2? In general, such as "High-risk auto crash" seems to be a valid predictor of TBI. Wasn't it determined to be valid during the feature selection process?

(ANSWER) Thank you for the review and comments. We used predictors in Supplementary Table 1, which is different from Supplementary Table 2. In the trauma patients screening process by EMS providers in Korea, they screened criteria orderly from physiologic, anatomical, mechanism of injury. If patients had physiologic abnormality or anatomical abnormality, the mechanism of injury was not recorded in the registry. Therefore, "high-risk auto crash" was not collected for all patients and was not included in predictors in our prediction model. We added a footnote in Supplemntary Table 2.

(Revision) Supplementary Table 2 foot note

Field triage decision scheme criteria*

*EMS providers check specific criteria orderly from physiologic, anatomical, and mechanism of injury. If the preceding criteria are satisfied, the information of the latter criteria is not collected.

8. There appears to be a large discrepancy between the AUROC and NRI results in the TBI-D predictions. Do you have any opinion on the reason for this?

(ANSWER) Thank you for the review and comments. In the case of predicting TBI-D, the outcome proportion was very low (3.7%). In addition, because in-hospital management after ED arrival have large effects on mortality, we also thought that it would be difficult to predict death from the prehospital stage. Our study only used prehospital variables as the predictor, so poor performance of death prediction might be plausible.

Discussion

9. In the introduction, it is noted that "Prehospital clinical signs are also reported to have poor sensitivity for raised intracranial pressure following TBI." Did you examine the relationship with raised intracranial pressure and machine learning model in this study?

(ANSWER) Thank you for the review and comments. Intracranial pressure was not collected in our EMS setting. We alternatively used pupil reflex as a predictor.

10. Please discuss the potential clinical use of the model and implications as required by the TRIPOD Checklist.

(ANSWER) Thank you for the review and comments. We missed completing the potential clinical use of the model and implications part in the TRIPOD checklist. We reviewed our manuscript again and completed the TRIPOD checklist.

11. There are too many references. Reducing the number of references may help to focus the article. Thank you for your comment

(ANSWER) Thank you for the review and comments. We reviewed our references and refined them. We deleted several redundant references and added references from large cohort study about TBI as Review 1 recommended.

VERSION 2 – REVIEW

REVIEWER	Patrick Czorlich University Medical Center Hamburg-Eppendorf, Department of Neurosurgery
REVIEW RETURNED	26-Oct-2021

GENERAL COMMENTS	The authors have substantially revised and improved the manuscript.

	EMS are usually very country-specific and the work will be very interesting, especially for South-Korea. If such work is also published from other countries, a comparison of the relevant parameters could be of interest in order to adapt the respective emergency documentation protocols if necessary. Overall, the authors are to be congratulated for this overall fine work.

REVIEWER	Kazuya Matsuo Kobe University Graduate School of Medicine School of Medicine, Neurosurgery
REVIEW RETURNED	26-Oct-2021

GENERAL COMMENTS	The submitted manuscript was thoroughly revised in accordance with my comments. I have no comments any more.