

Predicting 30-day mortality in severely injured elderly patients with trauma in Korea using machine learning algorithms: a retrospective study

Jonghee Han, MD¹ , Su Young Yoon, MD¹ , Junepill Seok, MD² , Jin Young Lee, MD² , Jin Suk Lee, MD² , Jin Bong Ye, MD² , Younghoon Sul, MD^{2,3} , Se Heon Kim, MD² , Hong Rye Kim, MD⁴ 

¹Department of Cardiovascular and Thoracic Surgery, Trauma Center, Chungbuk National University Hospital, Cheongju, Korea

²Department of Trauma Surgery, Trauma Center, Chungbuk National University Hospital, Cheongju, Korea

³Department of Trauma Surgery, Chungbuk National University College of Medicine, Cheongju, Korea

⁴Department of Neurosurgery, Trauma Center, Chungbuk National University Hospital, Cheongju, Korea

Purpose: The number of elderly patients with trauma is increasing; therefore, precise models are necessary to estimate the mortality risk of elderly patients with trauma for informed clinical decision-making. This study aimed to develop machine learning based predictive models that predict 30-day mortality in severely injured elderly patients with trauma and to compare the predictive performance of various machine learning models.

Methods: This study targeted patients aged ≥ 65 years with an Injury Severity Score of ≥ 15 who visited the regional trauma center at Chungbuk National University Hospital between 2016 and 2022. Four machine learning models—logistic regression, decision tree, random forest, and eXtreme Gradient Boosting (XGBoost)—were developed to predict 30-day mortality. The models' performance was compared using metrics such as area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, specificity, F1 score, as well as Shapley Additive Explanations (SHAP) values and learning curves.

Results: The performance evaluation of the machine learning models for predicting mortality in severely injured elderly patients with trauma showed AUC values for logistic regression, decision tree, random forest, and XGBoost of 0.938, 0.863, 0.919, and 0.934, respectively. Among the four models, XGBoost demonstrated superior accuracy, precision, recall, specificity, and F1 score of 0.91, 0.72, 0.86, 0.92, and 0.78, respectively. Analysis of important features of XGBoost using SHAP revealed associations such as a high Glasgow Coma Scale negatively impacting mortality probability, while higher counts of transfused red blood cells were positively correlated with mortality probability. The learning curves indicated increased generalization and robustness as training examples increased.

Conclusions: We showed that machine learning models, especially XGBoost, can be used to predict 30-day mortality in severely injured elderly patients with trauma.

Prognostic tools utilizing these models are helpful for physicians to evaluate the risk of mortality in elderly patients with severe trauma.

Keywords: Wounds and injuries; Aged; Mortality; Prediction model; Machine learning

Received: April 22, 2024

Revised: May 23, 2024

Accepted: May 29, 2024

Correspondence to

Jonghee Han, MD
Department of Cardiovascular and
Thoracic Surgery, Trauma Center,
Chungbuk National University Hospital,
776 1sunhwan-ro, Seowon-gu,
Cheongju 28644, Korea
Tel: +82-43-269-7849
Email: medihan7@naver.com

© 2024 The Korean Society of Traumatology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Background

Trauma has been viewed as a common cause of disability and death in patients [1,2]. Approximately 500,000 elderly patients are admitted to trauma centers after injury annually in the United States [3]. In Korea, there are as many as 200,000 severe trauma patients annually, about 40% of whom aged ≥ 60 years. Recently, the proportion of elderly patients with trauma has been steadily increasing [4]. The care and management of critically ill elderly patients with severe trauma poses unique challenges for healthcare professionals, necessitating innovative approaches to improve outcomes and enhance patient care [5].

Many trauma scoring systems and predictors have been developed, such as the Injury Severity Score (ISS), Revised Trauma Score, Trauma and Injury Severity Score, and Geriatric Trauma Outcome Score (GTOS) [3,6,7]. The above scores provide clinical importance, but require the assumption of linear relationships between factors. However, in many cases, relationships exist beyond simple linear relationships, and multicollinearity problems often exist.

Machine learning (ML) learns models from past data to predict future data [8]. ML models show better predictive power when multiple variables show complicated and nonlinear relationships [9–13]. ML methods can extract various important factors that influence the prognosis of patients [14]. Using such technology can help improve the speed and accuracy of medical decisions, particularly in critically ill patients where swift decision-making is vital [15]. Many ML models, such as support vector machines, decision trees, logistic regression, random forests, eXtreme Gradient Boosting (XGBoost), and others, have been developed in various medical fields. In several studies, ML models have demonstrated high predictive accuracy in critically ill trauma patients [16–18].

Objectives

In this study, we developed a model to predict 30-day mortality in severely injured elderly patients with trauma, evaluated its performance, and analyzed risk factors.

METHODS

Ethics statement

This study was approved by the Institutional Review Board of Chungbuk National University Hospital (No. 2024-01-034-001). The requirement for informed consent was waived due to the

retrospective nature of the study.

Study population

This retrospective study was conducted at a single level I trauma center at Chungbuk National University Hospital (Cheongju, Korea) between January 2016 and December 2022. Data from the medical records of patients admitted to this hospital and registered in the Korean Trauma Database were collected and analyzed.

The inclusion criteria were patients aged ≥ 65 years, admitted to traumatic intensive care unit with ISS ≥ 15 . Those who were hopelessly discharged, died at the time of arrival, did not survive after cardiopulmonary resuscitation, died after 30 days after injury, or died unrelated to trauma were excluded from the study.

Study design

The patients were divided into two groups: survivor and nonsurvivor groups. We compared baseline characteristics, vital signs, and mental status upon arrival at the emergency room, Abbreviated Injury Score (AIS), ISS, Glasgow Coma Scale (GCS), need for emergency surgery, transfusion of red blood cells (RBCs) within 24 hours, number of transfused RBCs, and type of accident.

Finally, we selected 10 potential risk factors after the univariate analysis through least absolute shrinkage and selection operator (LASSO) regression [19]. We selected risk factors with nonzero LASSO coefficient as a feature (Fig. 1). The 10 risk factors were age, pulse rate, GCS score, respiratory rate, need for surgery, chest AIS, pelvis AIS, ISS, number of transfused RBCs, and type of accident. Potential risk factors were used as the same features to compare the models equally.

ML modeling establishment

The original dataset was divided into training (80%) and testing (20%). Four different ML models (logistic regression, decision tree, random forest, and XGBoost) were used to create and test the models for predicting mortality in elderly patients with severe trauma. Each model was trained on the training set and assessed using the test set. To obtain the best prediction model, the optimal hyperparameters of the models were selected and tuned using a grid search method with a 10-fold cross-validation procedure. The 10-fold cross-validation was used to evaluate the models by dividing the dataset into 10 subsamples, training and testing the models 10 times, and providing a more reliable estimate of their generalization performance [11,16]. The best estimator model was evaluated using the test set.

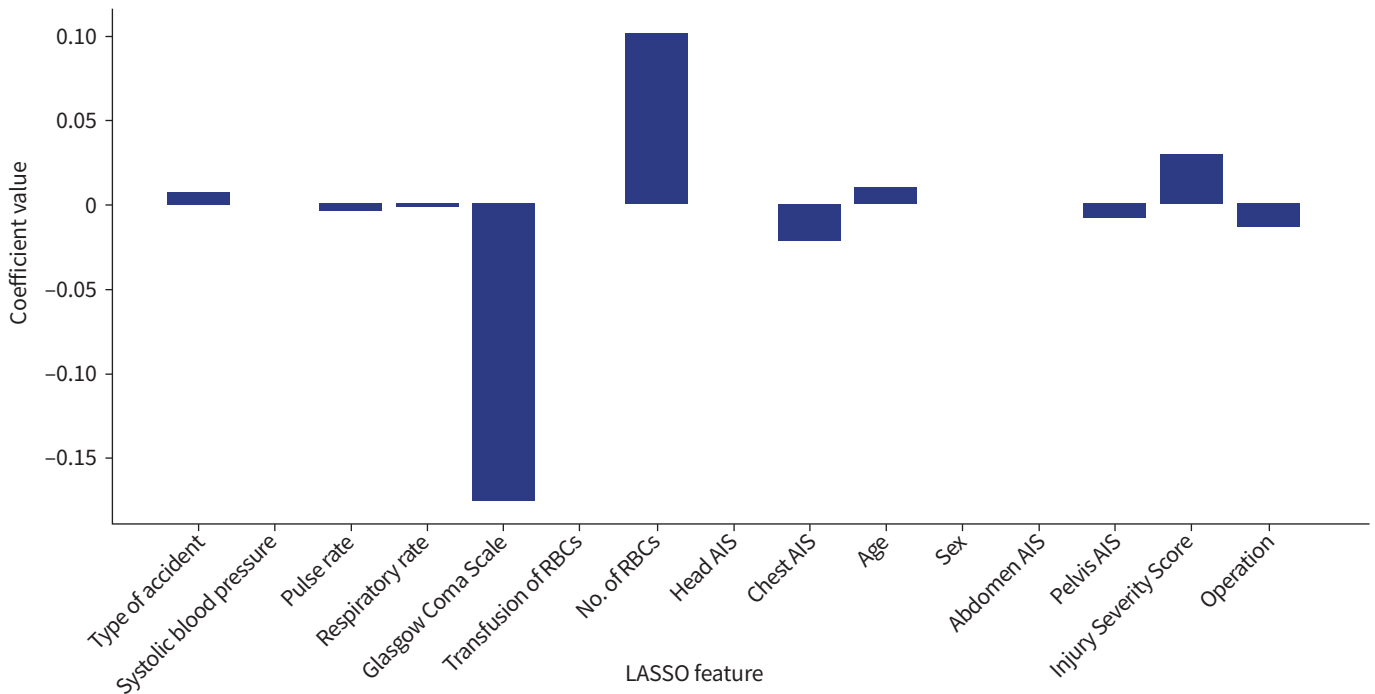


Fig. 1. Feature selection using the least absolute shrinkage and selection operator (LASSO) method with coefficients. RBC, red blood cell; AIS, Abbreviated Injury Severity.

Model performance and comparison

The model was evaluated using a confusion matrix, the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall (sensitivity), specificity, and F1 score. Shapley Additive Explanations (SHAP) is a popular technique in the field of ML for explaining the output of a model. This provided a way to understand the contribution of each feature to the prediction of the model [20]. To demonstrate the importance of features, we use the SHAP method. We used a learning curve to represent of the model's performance on a task as a function of the amount of training data or training iterations [21,22].

Statistical analysis

Statistical analyses were performed using Python ver. 3.10.5 (Python Software Foundation) and R ver. 4.2.3 (R Foundation for Statistical Computing). For categorical variables, we used the chi-square test or Fisher exact test to determine significance. For continuous variables, we used the t-test or Mann-Whitney U-test. A P-value of <0.05 was considered statistically significant. ML models were constructed using Python and several Python modules (pandas, numpy, sklearn, xgboost, matplotlib, and shap).

RESULTS

Clinical characteristics of the study population

A total of 555 patients were included in the study. Of the 555 patients, 430 (77.5%) survived, and 125 (22.5%) died. The differences in baseline characteristics between survivors and non-survivors are shown in Table 1. The nonsurvivor group was significantly older than the survivor group (79 years vs. 75 years, $P=0.021$). The GCS score was significantly lower in the nonsurvivor group than in the survivor group (6 vs. 14, $P<0.001$). The nonsurvivor group had a greater proportion of patients with hypotension (20.8% vs. 11.4%, $P=0.011$), need for RBC transfusion (66.4% vs. 44.7%, $P<0.001$), and number of transfused RBCs (5 units vs. 3 units, $P<0.001$) than the survivor group. ISS (26 vs. 22, $P<0.001$) and GTOS (162 vs. 141, $P<0.001$) were also higher in the nonsurvivor group than in the survivor group. The number of patients exhibiting higher head and chest AIS scores was significantly higher in the nonsurvivor group (Table 2).

Performance evaluation of the models

The accuracy, precision, recall, specificity, F1 score, and AUC of the ML models for predicting mortality in severely injured elderly patients with trauma were validated using the test dataset. In

Table 1. Basic characteristics of the study population

Characteristic	Total (n=555)	Survivor (n=430)	Nonsurvivor (n=125)	P-value
Age (yr)	75 (70–81)	75 (70–80)	79 (71–82)	0.021
Male sex	357 (64.3)	273 (63.5)	84 (67.2)	0.499
Injury mechanism				0.095
Pedestrian TA	124 (22.3)	89 (20.7)	35 (28.0)	
Bicycle TA	35 (6.3)	24 (5.6)	11 (8.8)	
Motorcycle TA	85 (15.3)	63 (14.7)	22 (17.6)	
Car TA	75 (13.5)	67 (15.6)	8 (6.4)	
Slip down	84 (15.1)	63 (14.7)	21 (16.8)	
Fall from height	88 (15.9)	74 (17.2)	14 (11.2)	
Stuck by object	25 (4.5)	20 (4.7)	5 (4.0)	
Other	39 (7.0)	30 (6.9)	9 (7.2)	
Pulse rate (beats/min)	85 (72–101)	84 (72–99)	92 (78–112)	0.004
Respiratory rate (breaths/min)	20 (20–24)	20 (20–24)	22 (18–24)	0.913
Systolic blood pressure (mmHg)	130 (102–158)	130 (104–154)	134 (100–170)	0.463
Glasgow Coma Scale	14 (8–15)	14 (11–15)	6 (3–10)	<0.001
Hypotension	75 (13.5)	49 (11.4)	26 (20.8)	0.011
RBC transfusion	275 (49.5)	192 (44.7)	83 (66.4)	<0.001
No. of transfused RBCs (U)	4 (2–6)	3 (2–5)	5 (4–11)	<0.001
Injury Severity Score	25 (18–27)	22 (17–26)	26 (25–35)	<0.001
Trauma and Injury Severity Score	0.881 (0.500–0.881)	0.881 (0.731–0.953)	0.500 (0.269–0.731)	<0.001
Geriatric Trauma Outcome Score	144 (129–164)	141 (125–158)	162 (144–186)	<0.001
Emergency operation	187 (33.7)	139 (32.3)	48 (38.4)	0.247

Values are presented as median (interquartile range) or number (%). Percentages may not total 100 due to rounding. TA, traffic accident; RBC, red blood cell.

the test-set analysis, the AUCs of the logistic regression, decision tree, random forest, and XGBoost models were 0.938, 0.863, 0.919, and 0.934, respectively (Fig. 2). The AUC was slightly higher in the logistic regression than in XGBoost; however, XGBoost exhibited better results in terms of accuracy, precision, recall, specificity, and F1 score (0.91, 0.72, 0.86, 0.92, and 0.78, respectively) (Table 3).

To analyze the impact of important features on the model, a SHAP summary figure was used to show how features affected the probability of mortality. The colors and locations of the points indicate whether a certain feature has a higher or lower value than the other features. Using the XGBoost model as an example, a high GCS score is negatively associated with the probability of mortality, whereas a high number of transfused RBCs positively correlates with the probability of mortality. Lower GCS score, higher number of transfused RBCs, higher ISS, and old age positively impacted on 30-day mortality prediction (Fig. 3).

The learning curve of XGBoost showed that the training score decreased, and the validation score increased as the number of training examples increased (Fig. 4). With repeated training and

validation sets, the difference between the training and validation sets decreased, indicating that the model became more generalized and robust.

DISCUSSION

Predicting a patient's prognosis with insufficient patient information in an emergency plays a vital role in patient evaluation and the utilization of limited medical resources in trauma. However, owing to information and time limitations in emergencies, emergency physicians and trauma surgeons are often unable to obtain all features for predicting mortality. Traditional trauma scoring systems, although clinically important, often assume linear relationships between explanatory variables. ML models, which use robust mathematical methodologies, exhibit better predictive capabilities, especially for handling complicated, nonlinear, and covariate factors. This provides an opportunity in the field of trauma to improve predictive accuracy.

This study focused on the application of ML models in predicting 30-day mortality in severely injured elderly patients with trauma, presenting an innovative approach for handling complex

Table 2. Comparison of AIS of patients who survived or not

AIS	No. of patients (%)			P-value
	Total (n=555)	Survivor (n=430)	Nonsurvivor (n=125)	
Head				
0	142 (25.6)	130 (30.2)	12 (9.6)	<0.001
1	0	0	0	
2	36 (6.5)	31 (7.2)	5 (4.0)	
3	89 (16.0)	76 (17.7)	13 (10.4)	
4	118 (21.3)	98 (22.8)	20 (16.0)	
5	170 (30.6)	95 (22.1)	75 (60.0)	
6	0	0	0	
Chest				
0	258 (46.5)	186 (43.3)	72 (57.6)	0.006
1	7 (1.3)	4 (0.9)	3 (2.4)	
2	32 (5.8)	25 (5.8)	7 (5.6)	
3	209 (37.7)	176 (40.9)	33 (26.4)	
4	41 (7.4)	35 (8.1)	6 (4.8)	
5	7 (1.3)	4 (0.9)	3 (2.4)	
6	1 (0.2)	0	1 (0.8)	
Abdomen				
0	374 (67.4)	285 (66.3)	89 (71.2)	0.297
1	0	0	0	
2	91 (16.4)	74 (17.2)	17 (13.6)	
3	62 (11.2)	52 (12.1)	10 (8.0)	
4	24 (4.3)	17 (4.0)	7 (5.6)	
5	4 (0.7)	2 (0.5)	2 (1.6)	
6	0	0	0	
Pelvis				
0	308 (55.5)	231 (53.7)	77 (61.6)	0.081
1	4 (0.7)	2 (0.5)	2 (1.6)	
2	121 (21.8)	97 (22.6)	24 (19.2)	
3	65 (11.7)	57 (13.3)	8 (6.4)	
4	38 (6.8)	31 (7.2)	7 (5.6)	
5	19 (3.4)	12 (2.8)	7 (5.6)	
6	0	0	0	

Percentages may not total 100 due to rounding. AIS, Abbreviated Injury Scale.

multiple variables. Mortality prediction models for elderly trauma patients with ISS ≥ 15 were established based on four ML algorithms.

AUC is a metric used to evaluate the performance of a ML model in binary classification tasks, such as predicting mortality. An AUC value over 0.9 is generally considered indicative of excellent discriminatory ability in a predictive model [23]. Three of the four models (logistic regression, random forest, and XGBoost) showed outstanding predictive performance in our test dataset. AUC is a valuable metric for assessing the overall performance of a predictive model; however, it has some limitations.

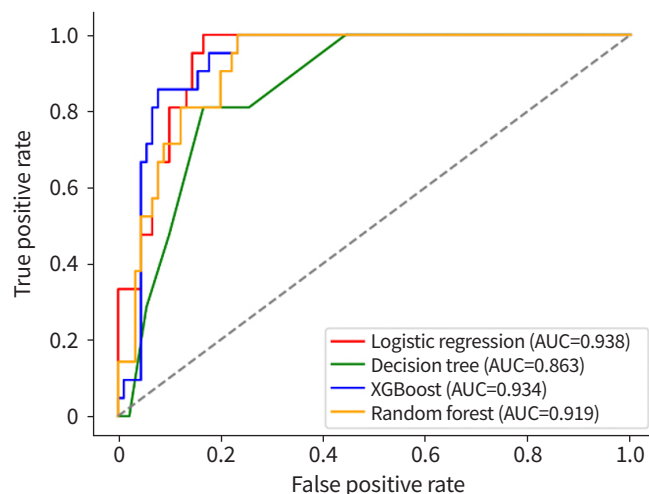


Fig. 2. Receiver operating characteristic curves of various machine learning models. AUC, area under the receiver operating curve; XGBoost, eXtreme Gradient Boosting.

These limitations include a lack of interpretability of specific contributors, inability to identify feature importance, and limited insight into the prediction impact. Various metrics can be used to evaluate binary classification models. Using only a subset can provide a false impression of a model's actual performance, leading to unexpected results when applied in clinical settings. Therefore, it is crucial to interpret the performance holistically by combining multiple metrics [24]. Therefore, we used the accuracy, precision, recall, specificity, and F1 score with the AUC to evaluate the models. The F1 score is a metric commonly used in binary classification problems, such as evaluating the performance of a ML model. The F1 score is the harmonic mean of precision and recall, and provide a balanced measure of a model's accuracy, particularly when dealing with imbalanced datasets [25,26].

In our study, XGBoost exhibited the best performance in terms of accuracy, precision, recall, specificity, and F1 score. The F1 score of XGBoost was 0.78, the highest among the four models. An F1 score of 0.78 generally indicate a model is performing well, striking a balance between precision and recall. In the field of medicine, datasets often exhibit an imbalance, where certain classes or outcomes are underrepresented compared to others [27]. Gradient boosting, including algorithms such as XGBoost, can be strong for imbalanced datasets, even without explicitly oversampling the minority class. Gradient boosting inherently has mechanisms that make it suitable for handling imbalanced datasets [28].

We used the SHAP method to analyze the feature importance in the ML models. The SHAP method is crucial in ML for pro-

Table 3. Model performances of the test dataset

Prediction model	Accuracy	Precision	Recall	Specificity	F1 score	AUC
Logistic regression	0.86	0.58	0.86	0.86	0.69	0.938
Decision tree	0.83	0.53	0.81	0.83	0.64	0.863
Random forest	0.86	0.61	0.81	0.88	0.69	0.919
XGBoost	0.91	0.72	0.86	0.92	0.78	0.934

AUC, area under the receiver operating curve; XGBoost, eXtreme Gradient Boosting.



Fig. 3. Shapley Additive Explanations (SHAP) value of the eXtreme Gradient Boosting (XGBoost) model output. RBC, red blood cell; AIS, Abbreviated Injury Severity.

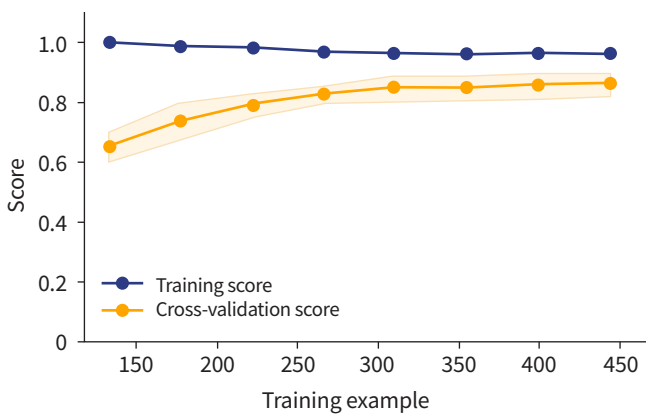


Fig. 4. Learning curves of eXtreme Gradient Boosting (XGBoost) model.

viding insightful explanations of model outputs by quantifying the contribution of each feature to predictions [29,30]. In this study, GCS was the most important feature, followed by the

number of transfused RBCs within 24 hours, ISS, pulse rate, and age in the XGBoost model.

Learning curves provide valuable insights into the performance of ML models during training by visualizing how a model's performance changes with the amount of training data or iterations [22]. Our results demonstrated a tendency toward generalization as the number of iterations increased across all models.

Limitations

Our study has some limitations. First, it was a retrospective study. Our cohort was specific to elderly patients. Therefore, there may have been selection bias. Second, the sample size was relatively small, and there may be problems such as overfitting, requiring additional data and external validation. Lastly, the unbalanced nature of the dataset may affect model generalization and reliability.

Conclusions

This study demonstrated that ML models, particularly XGBoost, perform well in predicting 30-day mortality among severely injured elderly trauma patients. ML models for mortality prediction can aid in the early identification of mortality risk factors and facilitate early intervention, potentially reducing mortality rates. Further research with larger datasets and external validation is recommended to enhance the robustness and generalizability of these models.

ARTICLE INFORMATION

Author contributions

Conceptualization: JH, SY; Data curation: JH, JS, JYL, JSL, JBY, YS, SHK, HRK; Formal analysis: JH; Methodology: JH; Project administration: YS; Supervision: SY; Writing—original draft: JH; Writing—review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are available from the corresponding author and the first author upon reasonable request.

REFERENCES

- DiMaggio C, Ayoung-Chee P, Shinseki M, et al. Traumatic injury in the United States: in-patient epidemiology 2000-2011. *Injury* 2016;47:1393–403.
- The American Association for the Surgery of Trauma. Trauma facts [Internet]. The American Association for the Surgery of Trauma [cited 2024 Jan 2]. Available from: <https://www.aast.org/resources/trauma-facts>
- Zhao FZ, Wolf SE, Nakonezny PA, et al. Estimating geriatric mortality after injury using age, injury severity, and performance of a transfusion: the Geriatric Trauma Outcome Score. *J Palliat Med* 2015;18:677–81.
- Korean Statistical Information Service (KOSIS). Emergency medical status statistics [Internet]. KOSIS; 2021 [cited 2023 Dec 22]. Available from: https://kosis.kr/statHtml/statHtml.do?orgId=411&tblId=DT_41104_244&conn_path=I2
- Sawada Y, Isshiki Y, Ichikawa Y, et al. The significance of the treatment for elderly severe trauma patients who required intensive care. *Cureus* 2023;15:e39110.
- Cook AC, Joseph B, Inaba K, et al. Multicenter external validation of the Geriatric Trauma Outcome Score: a study by the Prognostic Assessment of Life and Limitations After Trauma in the Elderly (PALLIATE) consortium. *J Trauma Acute Care Surg* 2016;80:204–9.
- Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method: Trauma Score and the Injury Severity Score. *J Trauma* 1987;27:370–8.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007;2:59–77.
- Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med* 2020;18:462.
- Xu Y, Han D, Huang T, et al. Predicting ICU mortality in rheumatic heart disease: comparison of XGBoost and logistic regression. *Front Cardiovasc Med* 2022;9:847206.
- Gao W, Wang J, Zhou L, et al. Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. *Comput Biol Med* 2022;140:105097.
- Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. *Shock* 2017;48:504–10.
- Hanko M, Grendar M, Snopko P, et al. Random forest-based prediction of outcome and mortality in patients with traumatic brain injury undergoing primary decompressive craniectomy. *World Neurosurg* 2021;148:e450–8.
- Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data* 2018;5:180178.
- Guo C, Pan J, Tian S, Gao Y. Using machine learning algorithms to predict 28-day mortality in critically ill elderly patients with colorectal cancer. *J Int Med Res* 2023;51:300060 5231198725.
- Choi H, Lee JY, Sul Y, et al. Comparing machine learning and logistic regression for acute kidney injury prediction in trauma patients: a retrospective observational study at a single tertiary medical center. *Medicine (Baltimore)* 2023;102:e34847.
- Rau CS, Kuo PJ, Chien PC, Huang CY, Hsieh HY, Hsieh CH. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PLoS One* 2018;13:e0207192.

18. Yang S, Cao L, Zhou Y, Hu C. A retrospective cohort study: predicting 90-day mortality for ICU trauma patients with a machine learning algorithm using XGBoost using MIM-IC-III database. *J Multidiscip Healthc* 2023;16:2625–40.
19. Muthukrishnan R, Rohini R. LASSO: a feature selection technique in predictive modeling for machine learning. 2016 IEEE International Conference on Advances in Computer Applications (ICACA); 2016 Oct 24; Coimbatore, India. IEEE; 2016. p. 18–20.
20. Li Z. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. *Comput Environ Urban Syst* 2022;96:101845.
21. Allyn J, Allou N, Augustin P, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS One* 2017;12:e0169772.
22. Perlich C. Learning curves in machine learning. In: Sammut C, Webb GI, editors. *Encyclopedia of machine learning*. Springer; 2011. p. 577–80.
23. Corbacioglu SK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: a guide to interpreting the area under the curve value. *Turk J Emerg Med* 2023;23:195–8.
24. Hicks SA, Strumke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* 2022;12:5979.
25. Muller AC, Guido S. *Introduction to machine learning with Python*. O'Reilly Media; 2016.
26. Miao J, Zhu W. Precision–recall curve (PRC) classification trees. *Evol Intell* 2022;15:1545–69.
27. Li DC, Liu CW, Hu SC. A learning method for the class imbalance problem with medical data sets. *Comput Biol Med* 2010;40:509–18.
28. Ma Y, Gan M. Gradient boosting based prediction method for patient death in hospital treatment. In: Chen H, Zeng D, Yan X, Xing C, editors. *Smart health. International Conference on Smart Health (ICSH) 2019*; 2019 Jul 1–2; Shenzhen, China. Springer; 2019. p. 283–93.
29. Wei C, Zhang L, Feng Y, Ma A, Kang Y. Machine learning model for predicting acute kidney injury progression in critically ill patients. *BMC Med Inform Decis Mak* 2022;22:17.
30. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed* 2022;214:106584.