Enhancing Performance of The National Field Triage Guidelines Using Machine Learning: Development of a Prehospital Triage Model to Predict Severe Trauma (pTEST)

Qi Chen, Yuchen Qin, Zhichao Jin, Xinxin Zhao, Jia He, Cheng Wu, Bihan Tang

Submitted to: Journal of Medical Internet Research
on: March 24, 2024

Table of Contents

Enhancing Performance of The National Field Triage Guidelines Using Machine Learning: Development of a Prehospital Triage Model to Predict Severe Trauma (pTEST)

Qi Chen[1*] PhD; Yuchen Qin[1*] PhD; Zhichao Jin[1*] PhD; Xinxin Zhao[2] PhD; Jia He[1] PhD; Cheng Wu[1*] PhD; Bihan Tang[3] PhD

[1]Department of Health Statistics Naval Medical University Shanghai CN
[2]School of Medicine Tongji University Shanghai CN
[3]Department of Health Management Naval Medical University Shanghai CN
[*]these authors contributed equally

Corresponding Author:
Bihan Tang PhD
Department of Health Management
Naval Medical University
No. 800 Xiangyin Road
Shanghai
CN

Abstract

Background: Prehospital trauma triage is essential to get the right patient to the right hospital. However, the national field triage guidelines proposed by the American College of Surgeons proved relatively insensitive when identifying severe traumas.

Objective: This study aimed to build a prehospital triage model to predict severe trauma and enhance the performance of the national field triage guidelines.

Methods: This is a multi-site prediction study, and the data were extracted from the National Trauma Data Bank between 2017 and 2019. All patients with injury, aged ?16 years, and transported by ambulance from the injury scene to any trauma center were potentially eligible. The data were divided into training, internal, and external validation sets of 672,309, 288,134, and 508,703 patients. As national field triage guidelines recommended, age, seven vital signs, and eight injury patterns at the pre-hospital stage were included as candidate variables for model development. Outcomes are severe trauma with Injured Severity Score ?16 (primary) and critical resource use within 24 h of emergency department arrival (secondary). The triage model was developed using an extreme gradient boosting model and Shapley additive explanation analysis. The model's accuracy regarding discrimination, calibration, and clinical utility was assessed.

Results: At a fixed specificity of 0.5, the model showed a sensitivity of 0.799(0.797–0.801), an undertriage rate of 0.080(0.079–0.081), and an overtriage rate of 0.743(0.742–0.743) for predicting severe trauma. The model showed a sensitivity of 0.774(0.772–0.776), an undertriage rate of 0.158(0.157–0.159), and an overtriage rate of 0.609(0.608–0.609) when predicting critical resource use, fixed at 0.5 specificity. The triage model's areas under the curve were 0.755(0.753–0.757) for severe trauma prediction and 0.736(0.734–0.737) for critical resource use prediction. The triage model's performance was better than those of the Glasgow Coma Score, Prehospital Index, revised trauma score, and the 2011 national field triage guidelines RED criteria. The model's performance was consistent in the two validation sets.

Conclusions: The prehospital triage model is promising for predicting severe trauma and achieving an undertriage rate of <10%. Moreover, machine learning enhances the performance of field triage guidelines.

(JMIR Preprints 24/03/2024:58740)
DOI: https://doi.org/10.2196/preprints.58740

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Enhancing Performance of The National Field Triage Guidelines Using Machine Learning: Development of a Prehospital Triage Model to Predict Severe Trauma (pTEST)

Qi Chen[3$] PhD, Yuchen Qin[3$] PhD, Zhichao Jin[3$] PhD, Xinxin Zhao[4] PhD, Jia He[3,4] PhD, Cheng Wu[3$] PhD, Bihan Tang[1,2*] PhD

Author Affiliations:

[1]Department of Health Management, Naval Medical University, Shanghai, China

[2]Department of Community Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA, USA

[3]Department of Health Statistics, Naval Medical University, Shanghai, China

[4]School of Medicine, Tongji University, Shanghai, China

[$]These authors contributed equally and are co-first authors of this article.

[*]**Corresponding author:**

Prof Bihan Tang, Department of Health Management, Naval Medical University, No. 800 Xiangyin Road, Shanghai 200433, China, Tel: +86-021-81871425, Fax: +86-021-81871425, E-mail: mangotangbihan@126.com.

Abstract

Background: Prehospital trauma triage is essential to get the right patient to the right hospital. However, the national field triage guidelines proposed by the American College of Surgeons proved relatively insensitive when identifying severe traumas.
Objective: This study aimed to build a prehospital triage model to predict severe trauma and enhance the performance of the national field triage guidelines.
Methods: This is a multi-site prediction study, and the data were extracted from the National Trauma Data Bank between 2017 and 2019. All patients with injury, aged ≥16 years, and transported by ambulance from the injury scene to any trauma center were potentially eligible. The data were divided into training, internal, and external validation sets of 672,309, 288,134, and 508,703 patients. As national field triage guidelines recommended, age, seven vital signs, and eight injury patterns at the pre-hospital stage were included as candidate variables for model development. Outcomes are severe trauma with Injured Severity Score ≥16 (primary) and critical resource use within 24 h of emergency department arrival (secondary). The triage model was developed using an extreme gradient boosting model and Shapley additive explanation analysis. The model's accuracy regarding discrimination, calibration, and clinical utility was assessed.
Results: At a fixed specificity of 0.5, the model showed a sensitivity of 0.799(0.797–0.801), an undertriage rate of 0.080(0.079–0.081), and an overtriage rate of 0.743(0.742–0.743) for predicting severe trauma. The model showed a sensitivity of 0.774(0.772–0.776), an undertriage rate of 0.158(0.157–0.159), and an overtriage rate of 0.609(0.608–0.609) when predicting critical resource use, fixed at 0.5 specificity. The triage model's areas under the curve were 0.755(0.753–0.757) for severe trauma prediction and 0.736(0.734–0.737) for critical resource use prediction. The triage model's performance was better than those of the Glasgow Coma Score, Prehospital Index, revised trauma score, and the 2011 national field triage guidelines RED criteria. The model's performance was consistent in the two validation sets.
Conclusions: The prehospital triage model is promising for predicting severe trauma and achieving an undertriage rate of <10%. Moreover, machine learning enhances the performance of field triage guidelines.

Keywords: Severe Trauma; Field Triage; Machine Learning; Prediction model

Introduction

Trauma is a universal health challenge that places a massive burden on national economies. It causes 4.4 million deaths annually, and an estimated 10% of all years lived with disability.[1,2] The American College of Surgeons Committee on Trauma (ACS-COT) recommends that severe trauma be treated at levels 1 and 2 trauma care facilities.[3] Patients with severe trauma have approximately 25% lower mortality rates when treated at levels 1 or 2 trauma centers than when treated at lower-level or non-trauma centers.[4] Prehospital estimation of injury severity is essential for prehospital triage. It is a critical step for emergency medical service (EMS) providers in making decisions regarding patient destination. Under- and overtriage are incorrect triages. A low-sensitivity triage tool results in many false-negative cases indicating undertriage and a possible failure in trauma first aid. Conversely, low specificity is associated with a high rate of false-positive cases, indicating overtriage.[5]

The national field triage guidelines were initially developed by ACS-COT in 1976 and revised in 2011 and 2021.[6,7] The national field triage guidelines have been widely implemented in the US and represent one of the few standardized national protocols for EMS. It was developed based on

peer-reviewed research, resulting in biased estimates and reduced generalizability.[3,8] A prospective national triage guidelines validation study for identifying high-risk trauma patients found that the guidelines were relatively insensitive in identifying severely injured patients and those requiring early critical resource use.[9] In addition, other triage tools, such as the Glasgow Coma Score (GCS), Prehospital Index (PHI), and revised trauma score (RTS), have not shown ideal predictive performance.[5,10-12] Therefore, it emphasizes the need to improve the prehospital triage tool.[13]

Machine learning (ML) development has advanced rapidly in the medical field, notably in trauma medicine, and has demonstrated that the ML model's predictive ability is significantly better than that of the conventional trauma triage tools for mortality outcomes, hospitalization, and critical care admission.[14,15] Therefore, this study aimed to build a prehospital triage model to predict severe trauma (pTEST) and enhance the performance of the national field triage guideline.

Methods

Recruitment

This multi-site prediction study was conducted to predict trauma severity during field triage. We developed, validated, and reported our triage model following the Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis statement,[16] as shown in Multimedia Appendix 1. The Naval Medical University Ethics Committee approved the study protocol. Our study was secondary analyses utilizing the National Trauma Data Bank (NTDB) with primary consent, and the data is anonymized.

Source of data and patients

Data from the NTDB, the largest aggregation of trauma registry data in the United States assembled by the ACS, was used in this study.[17] In 2017, the ACS Trauma Quality Program transitioned to a new technical vendor and redesigned the NTDB infrastructure. The 2017 and 2018 NTDB datasets comprising 2,041,706 patients were used for pTEST model development, hyperparameter tuning, and internal validation. The 2019 NTDB dataset comprising 1,097,190 patients was used for the external validation of the pTEST model.

Notably, all patients with injury, aged ≥16 years, and transported by ground or aerial ambulance from the injury scene to any trauma center were potentially eligible. Due to a lack of crucial information on outcomes and predictors, we excluded patients who died in the EMS, were discharged from the emergency department (ED) to another hospital, and those without any EMS records or an Injury Severity Score (ISS). In total, 960,443 and 508,703 participants were included in the development and validation sets, respectively. Figure 1 shows the patient selection process.

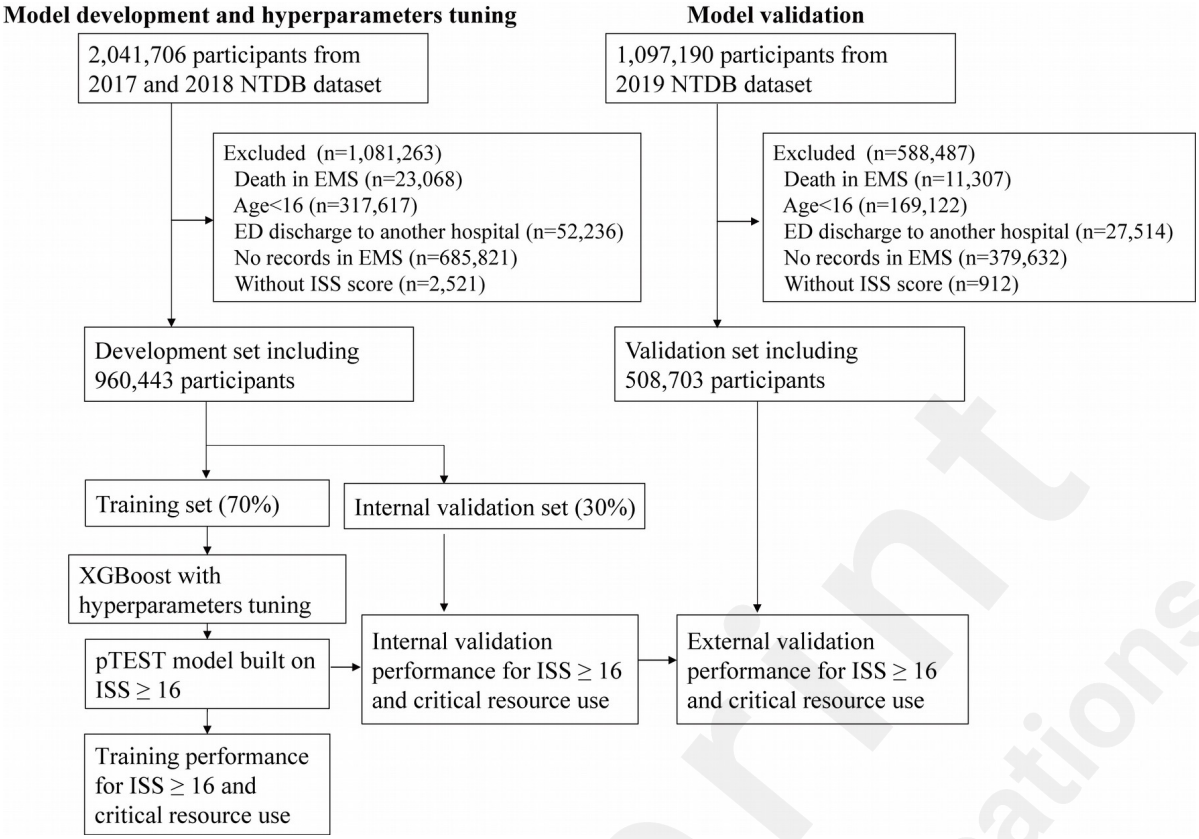


Figure 1. The flow chart of model development and validation.

Outcome and predictors

Defining “severe trauma” is a challenge in prehospital triage model development. It varies widely across studies. The reference standard (primary outcome) of “severe injury” was considered as an ISS ≥16 as the benchmark to evaluate triage accuracy recommended by the ACS.[7,18,19] The ISS calculated by anatomical criteria is assumed to be consistent with the patient status on-scene and is associated with high mortality.[4,19] However, it does not reflect resource utilization directly; therefore, we included a composite resource-based early critical resource use measure as the secondary outcome. According to similar studies,[9,18] early critical resource use included intubation in the EMS or ED, discharge to the intensive care unit from the ED, surgery for hemorrhage control, interventional radiology procedures, cerebral monitoring, and in-hospital death, all within 24 h. A detailed definition of severe trauma is provided in Multimedia Appendix 2.

According to a recent field triage protocol review in 2017,[10] the significant predictors of a severely injured patient were age, vital signs, injury patterns, and injury mechanism. In addition, in the US National Guidelines for the Field Triage of Injured Patients in 2011 and 2021,[6,7] age, vital signs, and injury pattern measurements were the field triage’s top priorities. In the 2011 national field triage guidelines, severely injured patients who should be transported preferentially to the highest-level trauma center were identified using the RED criteria, which included three vital signs (GCS, systolic blood pressure, and respiratory rate) and eight injury patterns. During field triage, time is essential, and the number and complexity of hand-collected variables must be limited. Therefore, as the US National Guideline for the Field Triage recommended and recorded in the NTDB, we incorporated 16 candidate variables in EMS for model development, including age at the time of injury (AGEYEARS), GCS Eye (EMSGCSEYE), GCS Motor (EMSGCSMOTOR), GCS Verbal (EMSGCSVERBAL), systolic blood pressure (EMSSBP), oxygen saturation (EMSPULSEOXIMETRY), respiratory rate (EMSRESPIRATORYRATE), pulse rate (EMSPULSERATE), penetrating injuries (TCCPEN), chest wall instability (TCCCHEST), long-

bone fractures (TCCLONGBONE), crushed extremity (TCCCRUSHED), amputation (TCCAMPUTATION), pelvic fracture (TCCPELVIC), skull fracture (TCCSKULLFRACTURE), and paralysis (TCCPARALYSIS). The detail definitions of candidate variables are listed in Multimedia Appendix 3.

Model development

The pTEST model was developed using the extreme gradient boosting model (XGBoost) and Shapley additive explanation analysis (SHAP). XGBoost is a novel boosting tree-based ensemble algorithm through which new models are created to predict residuals or errors of prior models and then combined to make a final prediction.[20] Recently, XGBoost has been widely used in ML due to its outstanding prediction performance, ability to employ continuous and categorical inputs, lack of data preprocessing, imbalanced data handling capacity, high internal optimization, and relatively modest computational costs.[21]

Patients in the development set from the 2017 and 2018 NTDB datasets were randomly grouped into training (70%) and internal validation (30%) sets for model development. Using a grid search, a 10-fold cross-validation process was used on the training set for hyperparameter tuning. The goal of hyperparameter tuning is to find the values that lead to the best-predicted performance. The optimal values of hyperparameters were as follows: learning rate = 0.04071151, minimum loss reduction required to make a further partition = 20.36485, maximum tree depth = 14, minimum sum of instance weights needed in a leaf node=39, maximum number of boosting iterations = 1051, subsample ratio of the training instance= 0.7763707, and the other hyperparameters were default. In addition to the training set, the model's reproducibility, transportability, and generalizability were evaluated using internal and external validation sets.

Missing values are an essential concern in trauma triage because there may not always be time to measure critical variables. The absent proportions of the training, internal validation, and external validation sets are shown in Multimedia Appendix 4, with all variables, except pulse oximetry, missing below 6%. XGBoost supports branch directions for predictors with missing values, creating an advantage in real-world situations where XGBoost can still achieve individual prediction without complete prehospital data.

To gain insight into the risk prediction model, we investigated different predictors' contributions based on Shapley values using SHAP, a game theory concept introduced in the 1950s.[22,23] A predictor's SHAP value can be positive or negative, suggesting an increased or decreased probability of severe trauma, respectively. In our study, the SHAP values were visualized at global (dataset level) and local (patient-specific) levels to investigate the predictors' impact. XGBoost and SHAP were implemented using the R packages tidy models and Shapviz.

Statistical analysis methods

For the sample size calculation, the prevalence of events was set at 17.8%, and the number of predictors was 16 based on the development set. The area under the curve (AUC) of the optimal prehospital triage model in a previous study was 0.68,[11] and our pTEST model was expected to achieve an AUC of 0.7. At least 1871 patients were required for model development or validation using the R package pmsampsize.[24]

Continuous data are presented as mean and standard deviation (SD), and categorical data are presented as frequencies and percentages (%). The t-test was used to evaluate the differences in continuous data, which followed a normal distribution and variance homogeneity; otherwise, the Wilcoxon rank-sum test was used. The differences in categorical data were evaluated using Pearson's chi-square test. The area under the receiver operating characteristic curve was calculated to assess model discrimination. The AUCs between models were compared using the DeLong test. The best thresholds of the models were determined by maximizing the Youden index, and performance metrics, including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative

predictive value (NPV), were calculated. Performance metrics 95% confidence intervals were calculated using 500 bootstrap replicates. In addition, the pTEST model is intended to identify severe trauma that requires high sensitivity and NPV to rule out it. Sensitivity and specificity are inversely proportional and a tradeoff needs to be made between sensitivity and specificity. Therefore, the sensitivities and NPV of the different models were compared using a 0.5 specificity as previous studies [25,26].

Our study defined the over- and undertriage rates.

Overtriage rate (1-PPV) = number of patients with negative outcomes (ISS <16 or no critical resource use) predicted as positive outcomes/total number of predicted positive outcomes.

Undertriage rate (1-NPV) = number of patients with positive outcomes predicted as negative outcomes/total number of predicted negative outcomes.

In addition, a calibration plot and Brier score were generated to assess how closely the predicted probability approximated the actual probability. The clinical utility of the models was evaluated using a decision curve analysis method. The discrimination, calibration, and clinical utility of the pTEST were compared with the GCS, PHI, RTS, and RED criteria of the 2011 National Field Triage Guidelines. Statistical significance was set at $P < .05$. All statistical analyses were performed using the R software (version 4.3.1).

Results

Patient characteristics

Table 1 shows the patients' baseline characteristics in the training, internal validation, and external validation sets. In these three sets, severe trauma proportions were 17.80%, 17.80%, and 17.08%, respectively, and critical resource use proportions were 29.36%, 29.56%, and 28.17%, respectively. Notably, most variables showed statistically significant differences among the three sets for large sample sizes, but the differences were minimal. The demographic characteristics, vital signs, and injury patterns of non-severe and severe trauma are listed in Multimedia Appendix 5–7, and the attributes of non-critical and critical resource users are listed in Multimedia Appendix 8–10. Severe trauma and critical resource users are usually male, air-transported, taken to higher-level trauma centers, and have extreme trauma patterns.

Table 1. Baseline characteristics of the patients from the training set, internal validation set and external validation set

Characteristics	Training set (n=672309)	Internal validation set (n=288134)	External validation set (n=508703)	*P* value
Sex, Male	403434(60.01)	172821(59.98)	300711(59.12)	<.001
Transport mode				<.001
Ground	622489(92.59)	266622(92.53)	474645(93.30)	
Helicopter	48464(7.21)	20911(7.26)	33346(6.56)	
Fixed-wing	1356(0.20)	601(0.21)	712(0.14)	
Trauma center level				<.001
Level 1	275723(55.75)	118199(55.72)	206563(54.39)	
Level 2	179419(36.28)	77005(36.30)	142433(37.50)	
Level 3	39397(7.97)	16944(7.99)	30819(8.11)	
TCCPEN, yes	26987(4.01)	11319(3.93)	17658(3.47)	<.001
TCCCHEST, yes	4295(0.64)	1825(0.63)	3216(0.63)	.89
TCCLONGBONE, yes	4598(0.68)	1867(0.65)	3213(0.63)	.002
TCCCRUSHED, yes	3188(0.47)	1374(0.48)	2712(0.53)	<.001
TCCAMPUTATION, yes	832(0.12)	372(0.13)	613(0.12)	.58
TCCPELVIC, yes	7495(1.11)	3346(1.16)	5852(1.15)	.07
TCCSKULLFRACTURE, yes	5127(0.76)	2237(0.78)	4087(0.80)	.043

TCCPARALYSIS, yes	4379(0.65)	1887(0.65)	3077(0.60)	.003
ISS score, ≥16	119690(17.80)	51296(17.80)	86902(17.08)	<.001
Surgery for hemorrhage control, yes	14714(2.45)	6441(2.50)	10886(2.37)	.002
Cerebral monitor, yes	8527(1.42)	3563(1.38)	5819(1.27)	<.001
Interventional radiology procedures, yes	5154(0.86)	2182(0.85)	3848(0.84)	.59
Discharge to the ICU from ED, yes	142422(21.46)	61384(21.58)	105670(21.06)	<.001
In-hospital death within 24 h, yes	9303(1.38)	3987(1.38)	6371(1.25)	<.001
Intubation in the EMS or ED, yes	72755(10.82)	31080(10.79)	49779(9.79)	<.001
Critical resource use, yes	177570(29.36)	76604(29.56)	129551(28.17)	<.001
RED criteria, yes	87577(13.03)	37448(13.00)	62194(12.23)	<.001
Age	53.12(21.86)	53.22(21.84)	54.40(21.77)	<.001
EMSSBP, mmHg	139.89(28.37)	139.92(28.42)	140.74(28.62)	<.001
EMSPULSERATE, n/minute	90.58(20.31)	90.58(20.29)	90.34(20.41)	<.001
EMSRESPIRATORYRATE, n/minute	18.42(4.73)	18.44(4.79)	18.46(4.82)	.08
EMSPULSEOXIMETRY, %	96.26(5.50)	96.26(5.48)	96.19(5.44)	<.001
EMSGCSEYE	3.81(0.65)	3.81(0.65)	3.82(0.63)	<.001
EMSGCSVERBAL	4.60(0.95)	4.60(0.95)	4.61(0.93)	.02
EMSGCSMOTOR	5.73(0.97)	5.73(0.97)	5.75(0.94)	<.001
EMSTOTALGCS	14.12(2.44)	14.12(2.43)	14.17(2.36)	<.001
Total time spent in ED, minutes	189.94(150.35)	207.75(400.72)	207.68(1355.03)	<.001
Length of stay, days	6.20(9.67)	6.21(8.51)	6.21(8.38)	.001
ISS score	9.69(8.37)	9.69(8.32)	9.55(8.20)	<.001
PHI score	1.32(2.23)	1.32(2.23)	1.31(2.22)	.26
RTS score	11.73(0.88)	11.73(0.89)	11.74(0.86)	.045

Note: The continuous data were described by mean (SD), and categorical data were described by n (%).

Model performance

For predicting severe trauma, we compared the performance metrics of the other models at the same specificity fixed at a moderate number of 0.5. The pTEST model showed a higher sensitivity of 0.799 (0.797–0.801), a lower undertriage rate of 0.080 (0.079–0.081), and a lower overtriage rate of 0.743 (0.742–0.743) in the training set (Table 2). In addition, for critical resource use prediction fixed at a specificity of 0.5, the pTEST model showed a higher sensitivity of 0.774 (0.772–0.776), lower undertriage rate of 0.158 (0.157–0.159) and lower overtriage rate of 0.609 (0.608–0.609) than the other models in the training set (Multimedia Appendix 11). We validated the pTEST model performance using two validation sets and obtained consistent results (Table 2 and Multimedia Appendix 11). The model performance metrics for predicting severe trauma and critical resource use at the best thresholds with the maximum Youden index are listed in Multimedia Appendix 12 and 13, demonstrating a higher pTEST Youden index than other models.

Table 2. model performance metrics for predicting severe trauma fixed at a specificity of 0.5

Prediction Tool	Specificity	Sensitivity	Accuracy	Undertriage rate (1-NPV)	Overtriage rate (1-PPV)	Youden index
Training set						
pTEST	0.500	0.799(0.797-0.801)	0.553(0.553-0.554)	0.080(0.079-0.081)	0.743(0.742-0.743)	1.299
GCS	0.500	0.682(0.680-0.684)	0.532(0.532-0.532)	0.119(0.119-0.120)	0.774(0.774-0.775)	1.182
PHI	0.500	0.711(0.709-	0.536(0.536-	0.107(0.107-	0.772(0.771-	1.211

		0.713)	0.537)	0.108)	0.772)	
RTS	0.500	0.634(0.632-0.636)	0.523(0.523-0.523)	0.132(0.132-0.133)	0.791(0.791-0.792)	1.134
RED criteria	0.500	0.620(0.619-0.621)	0.521(0.521-0.522)	0.141(0.141-0.142)	0.788(0.788-0.789)	1.120
Internal validation set						
pTEST	0.500	0.794(0.791-0.798)	0.552(0.552-0.553)	0.082(0.081-0.083)	0.744(0.743-0.745)	1.294
GCS	0.500	0.682(0.680-0.685)	0.532(0.532-0.533)	0.120(0.119-0.120)	0.774(0.773-0.775)	1.182
PHI	0.500	0.710(0.707-0.714)	0.536(0.536-0.537)	0.108(0.106-0.109)	0.772(0.771-0.773)	1.210
RTS	0.500	0.633(0.631-0.636)	0.523(0.523-0.523)	0.132(0.132-0.133)	0.791(0.791-0.792)	1.133
RED criteria	0.500	0.620(0.619-0.622)	0.521(0.521-0.522)	0.141(0.141-0.142)	0.788(0.788-0.789)	1.120
External validation set						
pTEST	0.500	0.794(0.792-0.797)	0.550(0.550-0.551)	0.078(0.077-0.079)	0.753(0.753-0.754)	1.294
GCS	0.500	0.681(0.679-0.683)	0.531(0.530-0.531)	0.115(0.115-0.116)	0.783(0.782-0.783)	1.181
PHI	0.500	0.709(0.706-0.711)	0.535(0.534-0.535)	0.103(0.103-0.104)	0.781(0.780-0.782)	1.209
RTS	0.500	0.633(0.631-0.635)	0.522(0.522-0.522)	0.127(0.126-0.128)	0.799(0.799-0.800)	1.133
RED criteria	0.500	0.616(0.615-0.618)	0.520(0.520-0.520)	0.137(0.136-0.137)	0.798(0.797-0.798)	1.116

In Figure 2, pTEST AUCs for severe trauma prediction were 0.755 (0.753–0.757), 0.751 (0.749–0.754), and 0.750 (0.749–0.752) in training, internal validation, and external validation sets, respectively, and the AUCs for predicting critical resource use were 0.736 (0.734–0.737), 0.732 (0.730–0.734), and 0.733 (0.732–0.735), respectively, demonstrating better discrimination ability than GCS, PHI, RTS, and RED criteria. Multimedia Appendix 14 depicts the pTEST model predicted outcome probability as a waterfall plot. The calibration curves in Multimedia Appendix 15 show that the severe trauma predicted probability and pTEST critical resource use agreed with the proportion observed using the smallest Brier score. In Multimedia Appendix 16, pTEST provides a consistently higher net benefit across a broad range of risk thresholds (10–100%) than the two default strategies and other models.

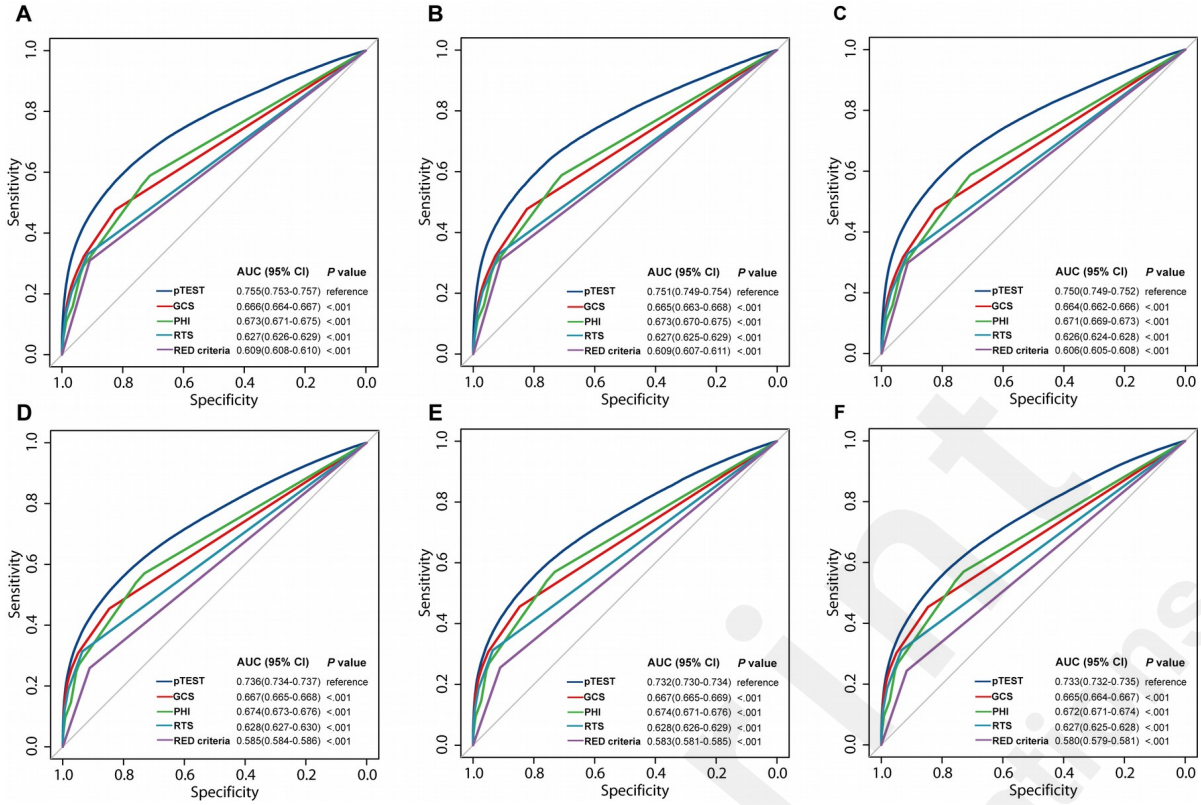


Figure 2. The ROC curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.

Model interpretation

As shown in the SHAP summary plots (Figure 3A), the contributions of the variables to the pTEST model for severe trauma prediction were evaluated using the average absolute SHAP values; the top five important variables were EMSGCSVERBAL, EMSSBP, EMSRESPIRATORYRATE, EMSGCSMOTOR, and EMSPULSEOXIMETRY. Figure 3B lists the impact of the different variables illustrated by the SHAP values for severe trauma prediction. Figure 3C shows each variable's SHAP values versus measured values. Figures 3B and C show that the higher the EMSGCSVERBAL score, the lower the probability of severe trauma ("negative" impact). Similarly, AGEYEARS, EMSGCSEYE, EMSGCSMOTOR, and EMSPULSEOXIMETRY negatively contributed to the predicted probability. In contrast, eight injury patterns contributed positively to the predicted probability. The SHAP summary and dependence plots of the pTEST model for critical resource use prediction are shown in Multimedia Appendix 17. The personalized feature attributes for two representative patients with and without severe trauma in the training set are provided in Multimedia Appendix 18.

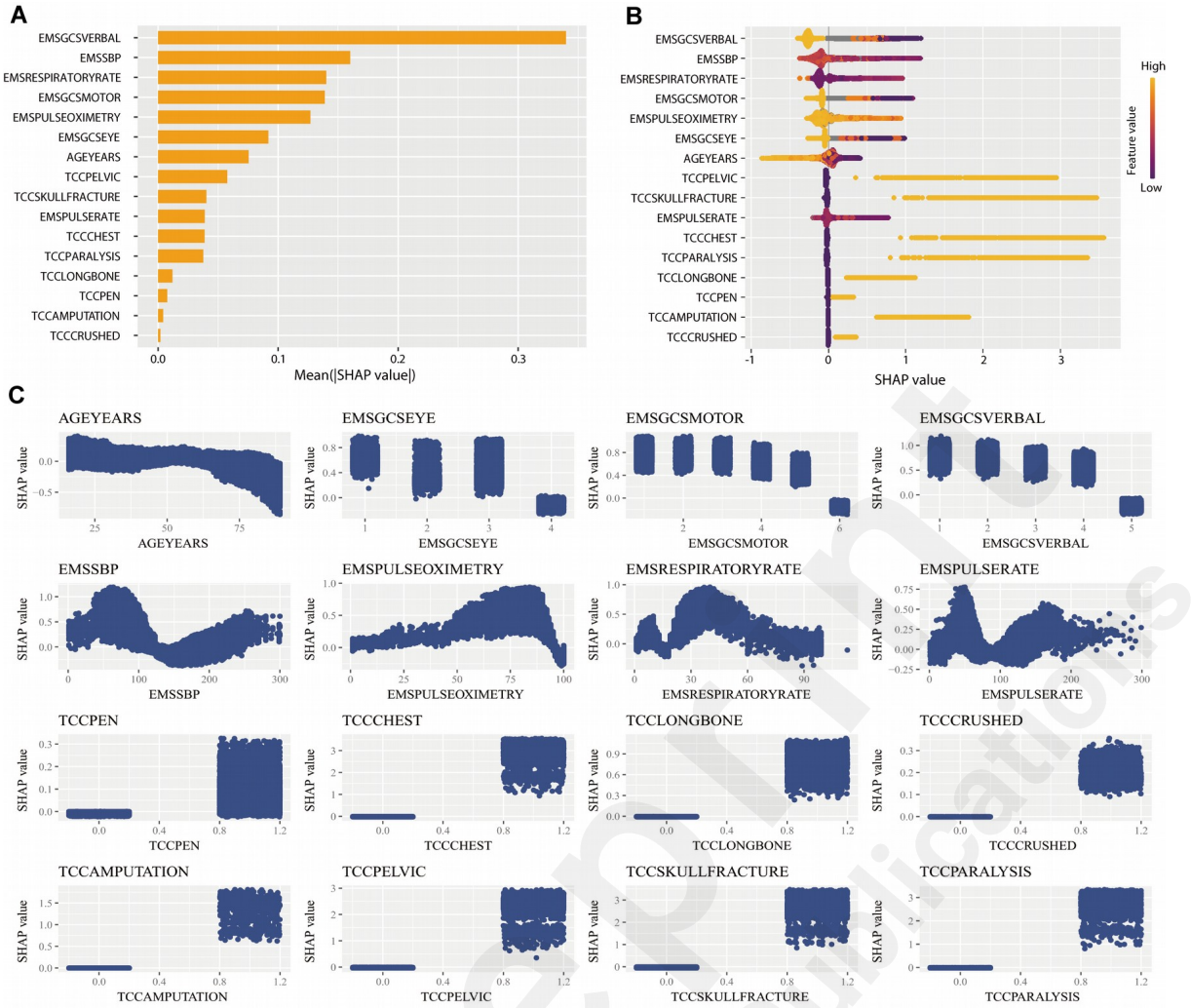


Figure 3. Global model explanation for predicting severe trauma by the SHAP method in training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. In each variable, a dot is made for each single patient, representing the SHAP value of this variable. The colors of the dots demonstrate the actual values of the features, and the dots are stacked vertically to show density. (C) SHAP dependence plot. Each dependence plot shows the association between the actual value and the SHAP value of the variable, and each dot represents a single patient.

Subgroup analysis

In Multimedia Appendix 19, subgroup analyses were performed according to age, sex, transport mode, trauma type, and prehospital time. The pTEST model AUCs for severe trauma prediction in patients ≥60 years old were relatively low at 0.717(0.714–0.720), 0.712(0.708–0.717), and 0.710(0.707–0.714) in the three sets, respectively. The AUCs in patients with penetrating injuries were relatively high at 0.815(0.810–0.820), 0.810(0.802–0.817), and 0.810(0.805–0.816) in the three sets, respectively. The high proportion of severe trauma in helicopter-transported patients (approximately 41%) led to a high undertriage rate (>0.2) and a low overtriage rate (<0.5). Multimedia Appendix 20 illustrates the pTEST model performance in critical resource use prediction in different subgroups.

Discussion

In this multi-site, large-sample study, we present a prehospital trauma triage tool, pTEST, for severe trauma prediction in EMS. To our knowledge, this is the first study combining ML with national triage guidelines. The pTEST performed optimally in internal and external validations. In addition, its diagnostic accuracy was evaluated using anatomical and resource-based outcomes, and the resource-based outcome is a better alternative to determine the need for specialized trauma care.[27] Furthermore, the pTEST was developed based on national triage guidelines, a globally adopted standard in many organizations. Therefore, the pTEST model can be conveniently applied in EMS practice and may have global relevance.

Consistent with previous studies[8], we also demonstrated the poor performance of the RED criteria from the national triage guidelines. Globally, all triage guidelines are based on a criteria checklist, including vital signs, injury type, and mechanism of injury.[7,28,29] These guidelines are simplistic: patients meeting any one of the criteria should be transported to the highest-level trauma center. In reality, there was an interaction among variables and non-linear effects of continuous variables (Figure 3C). XGBoost, as a non-linear ensemble method, can train a more accurate classifier from several weak classifiers and has other benefits, such as dealing with missing values and interaction, avoiding overfitting, and accelerating the training speed by parallel calculation.[30] Our study incorporated XGBoost into national field triage guidelines and developed the pTEST model to enhance the performance. The pTEST model included age, seven vital signs, and eight injury patterns. We did not perform further variable selection because these variables are the most important in the national field triage guidelines, and the overall number is moderate for field triage. The pTEST model did not achieve an undertriage rate of <5% or an overtriage rate of <35%, as the ACS Committee on Trauma targeted. However, two aspects must be noted. First, the definition of the undertriage rate by the ACS is different from that used in our study. For example, the undertriage rate in ACS = number of patients with ISS ≥16 transported to a low-level trauma center or non–trauma center / total number of patients transported to a low-level trauma center or non–trauma center.[19] In contrast, the undertriage rate in our study = number of patients with ISS ≥16 predicted as ISS <16 / total number of patients predicted ISS <16. In the national field triage guideline, the patients predicted as ISS <16 can be transported to a low-level or non-trauma center. If the patients with predicted ISS <16 and ≥16 are transferred to low- and high-level trauma centers, respectively, then the undertriage rate in our study is the same as that in ACS, but this is unlikely in practice. Second, the under- and overtriage rates are affected by severe trauma proportion. Since most (>90%) of our study population came from level 1 and 2 trauma centers, the severe trauma proportion was approximately 18%, resulting in undertriage rate overestimation and overtriage rate underestimation. Based on previous studies with good sample representativeness,[9] we assumed that the proportion of severe trauma was 3%. Keeping the sensitivity (0.794) and specificity (0.5) of the pTEST model and the sample size (n=508,703) in the external validation set unchanged, the under- and overtriage rates were 1.26% and 95.3%, respectively, in the external validation set (Multimedia Appendix 20), meeting the ACS undertriage rate target.

The pTEST model and the national field triage guidelines in subgroup analyses were particularly insensitive among older adults.[9] Possible explanations include different physiological responses to injury,[31] medication use that potentially worsens injury,[32] high prevalence of frailty, and comorbidities.[33] Previous studies have explored elderly-specific triage criteria.[15,34,35] The pTEST model performed better for penetrating traumas. Notably, several studies have found that penetration is a strong severe trauma predictor, and severely penetrated injured patients are more easily recognized.[10] The undertriage rate was high in patients transported using helicopters. Patients with a high proportion of severe trauma, such as those experiencing large-scale casualties, should be transferred to high-level trauma centers without field triage to reduce the undertriage rate.

Previous studies have reported controversial results regarding the utility of ML in medical

prediction issues.[36] Overall, in studies with a limited number of predictors, ML does not demonstrate advantages over traditional models, such as logistic regression,[37] whereas, for studies with many predictors, advanced ML may have an advantage.[38] A recent review, including 14 studies, demonstrated that the predictive ability of ML-based models was significantly better than that of conventional trauma triage tools for outcomes of mortality, hospitalization, and critical care admission, and XGBoost was the most commonly used ML algorithm.[14] In the present study, the relatively large number of predictors and sufficient amount of data tended to favor ML applications. We built the pTEST model using XGBoost but did not evaluate other ML methods. We believe an excellent model can be created using a large sample size, an advanced ML method, and robust hyperparameter tuning. In addition, we minimized the risk of chance findings and overfitting by avoiding exploring other modeling strategies.

This study had some limitations. First, most of our study population were from level 1 and 2 trauma centers, and the proportion of patients with severe trauma (approximately 18%) was significantly higher than that of all prehospital trauma patients (approximately 3%).[9] However, unlike PPV and NPV, the sensitivity, specificity, and AUC of the pTEST model were not affected by the proportion of severe trauma, and the high sensitivity, specificity, and AUC objectively reflect the pTEST model's good performance. In addition, some emergency resources may be unavailable in low-level trauma and non-trauma centers,[39] and samples from high-level trauma centers make it possible to evaluate the pTEST model with critical resource use as the endpoint. Second, the pTEST model was not developed into a software application, as in other studies,[40] because software development requires adaptation to existing information systems in EMS, which is a complex project. However, in the future, EMS providers can develop software based on available data and programs. Third, the 2017–2019 NTDB data followed the 2011 national field triage guidelines, and the latest guidelines have been revised in 2021. An additional "active bleeding" has been added to high-risk trauma types.[7] The new guidelines will take several years to be implemented, and our model must be further validated and updated as necessary.

Conclusions

We constructed a prehospital triage model, pTEST, to predict severe trauma and achieved an undertriage rate of <10%. Moreover, our study demonstrated that ML is a promising method for enhancing field triage guidelines performance. In the future, we will validate our pTEST model using populations from different countries and casualty backgrounds. In addition, software must be developed to increase user convenience of the pTEST model in the EMS.

Acknowledgments

Authors' contributions

Q. Chen, C. Wu, and B. Tang discussed and developed the study question for this report. Q. Chen, Y. Qin, and Z. Jin conducted the data extraction and statistical analysis, which was validated by B. Tang. All authors were involved in the interpretation of the data and discussed the results. Q. Chen wrote the first draft of this paper. All authors agreed on the final draft of this study. The corresponding author has the right to grant on behalf of all authors.

Conflict of interest

None declared.

Funding/Support

The study was funded by grants 2023YFC3107201 and 2023YFC3107202 from the National Key

Research and Development Program of China, grant 2022-JCJQ-QT-012 from Young Elite Scientists Sponsorship Program by China Association for Science and Technology, and grant 2022JC011 from Shanghai Emerging Cross Disciplinary Research Project.

Funding support or role of Funder/sponsor

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data access, responsibility, and analysis

B. Tang had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Data sharing statement

The NTDB datasets analyzed during our study are available in ACS Trauma Quality Programs [https://www.facs.org/quality-programs/trauma/quality/national-trauma-data-bank/]. The final data used during the current study is available from the corresponding author upon reasonable request.

Abbreviations

ACS-COT: American College of Surgeons Committee on Trauma
AUC: area under the curve
ED: emergency department
EMS: emergency medical service
GCS: Glasgow Coma Score
ISS: Injury Severity Score
ML: machine learning
NPV: negative predictive value
NTDB: National Trauma Data Bank
PHI: Prehospital Index
PPV: positive predictive value
pTEST: prehospital triage model to predict severe trauma
RTS: revised trauma score
SD: standard deviation
SHAP: Shapley additive explanation analysis
XGBoost: extreme gradient boosting model

Multimedia Appendix 1

TRIPOD Checklist in the manuscript.

Multimedia Appendix 2

Detailed definition of severe trauma and critical resource use.

Multimedia Appendix 3

The candidate variables for model development.

Multimedia Appendix 4

The missing proportion of the candidate variables for model development. (A) The missing proportion in training set. (B) The missing proportion in internal validation set. (C) The missing proportion in external validation set.

Multimedia Appendix 5
The demographic characteristics, vital signs and injury patterns between non- severe traumas and severe traumas in training set.

Multimedia Appendix 6
The demographic characteristics, vital signs and injury patterns between non-severe traumas and severe traumas in internal validation set.

Multimedia Appendix 7
The demographic characteristics, vital signs and injury patterns between non-severe traumas and severe traumas in external validation set.

Multimedia Appendix 8
The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in training set.

Multimedia Appendix 9
The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in internal validation set.

Multimedia Appendix 10
The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in external validation set.

Multimedia Appendix 11
Model performance metrics for predicting critical resource use fixed at a specificity of 0.5.

Multimedia Appendix 12
Model performance metrics for predicting severe trauma at the best thresholds with maximum Youden index.

Multimedia Appendix 13
Model performance metrics for predicting critical resource use at the best thresholds with maximum Youden index.

Multimedia Appendix 14
Waterfall plot for the predicted probability of pTEST model related to the outcome. (A) Predicted probability of severe trauma in training set. (B) Predicted probability of severe trauma in internal validation set. (C) Predicted probability of severe trauma in external validation set. (D) Predicted probability of critical resource use in training set. (E) Predicted probability of critical resource use in internal validation set. (F) Predicted probability of critical resource use in external validation set.

Multimedia Appendix 15
Calibration curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.

Multimedia Appendix 16
DCA curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.

Multimedia Appendix 17
Global model explanation for predicting critical resource use by the SHAP method in training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. (C) SHAP dependence plot.

Multimedia Appendix 18

Local model explanation for predicting severe trauma by the SHAP method in training set. (A) Waterfall plot of risks contributed by each variable for a patient at high risk of severe trauma. (B) Waterfall plot for a patient at low risk of severe trauma.

Multimedia Appendix 19

The performance of the pTEST model for predicting severe trauma in different subgroups.

Multimedia Appendix 20

(A) The performance of the pTEST model for predicting critical resource use in different subgroups. (B) The undertriage and overtriage rates of the pTEST in external validation set with the fixed sensitivity, specificity and sample size.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396(10258):1204-1222. doi: 10.1016/S0140-6736(20)30925-9.

2. World Health Organization. Injuries and violence. 2021 [cited 2024 2-22]; Available from: https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence.

3. Morris RS, Karam BS, Murphy PB, Jenkins P, Milia DJ, Hemmila MR, et al. Field-triage, hospital-triage and triage-assessment: a literature review of the current phases of adult trauma triage. J Trauma Acute Care Surg 2021;90(6):e138-e145. doi: 10.1097/TA.0000000000003125.

4. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. A national evaluation of the effect of trauma-center care on mortality. N Engl J Med 2006;354(4):366-378. doi: 10.1056/NEJMsa052049.

5. Gianola S, Castellini G, Biffi A, Porcu G, Fabbri A, Ruggieri MP, et al. Accuracy of pre-hospital triage tools for major trauma: a systematic review with meta-analysis and net clinical benefit. World journal of emergency surgery: World J Emerg Surg 2021;16(1): 31. doi: 10.1186/s13017-021-00372-1.

6. Sasser SM, Hunt RC, Faul M, Sugerman D, Pearson WS, Dulski T, et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. MMWR Recomm Rep 2012;61(RR-1):1-20. PMID: 22237112

7. Newgard CD, Fischer PE, Gestring M, Michaels HN, Jurkovich GJ, Lerner EB, et al. National guideline for the field triage of injured patients: Recommendations of the National Expert Panel on Field Triage, 2021. J Trauma Acute Care Surg 2022;93(2):e49-e60. doi: 10.1097/TA.0000000000003627.

8. Lupton JR, Davis-O'Reilly C, Jungbauer RM, Newgard CD, Fallat ME, Brown JB, et al. Under-triage and over-triage using the field triage guidelines for injured patients: a systematic review. Prehosp Emerg Care 2023;27(1):38-45. doi: 10.1080/10903127.2022.2043963.

9. Newgard CD, Fu R, Zive D, Rea T, Malveau S, Daya M, et al. Prospective validation of the National Field Triage Guidelines for identifying seriously injured persons. J Am Coll Surg 2016;222(2):146-158.e2. doi: 10.1016/j.jamcollsurg.2015.10.016.

10. van Rein EAJ, Houwert RM, Gunning AC, Lichtveld RA, Leenen LPH, van Heijl M. Accuracy of prehospital triage protocols in selecting severely injured patients: A systematic review. J Trauma Acute Care Surg 2017;83(2):328-339. doi: 10.1097/TA.0000000000001516.

11. Malik NS, Chernbumroong S, Xu Y, Vassallo J, Lee J, Bowley DM, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. EClinicalMedicine 2021;36:100888. doi: 10.1016/j.eclinm.2021.100888.

12. Bhalla MC, Frey J, Rider C, Nord M, Hegerhorst M. Simple Triage Algorithm and Rapid Treatment and Sort, Assess, Lifesaving, Interventions, Treatment, and Transportation mass casualty triage methods for sensitivity, specificity, and predictive values. Am J Emerg Med 2015;33(11):1687-1691. doi: 10.1016/j.ajem.2015.08.021.

13. Voskens FJ, van Rein EAJ, van der Sluijs R, Houwert RM, Lichtveld RA, Verleisdonk EJ, et al. Accuracy of prehospital triage in selecting severely injured trauma patients. JAMA Surg 2018;153(4):322-327. doi: 10.1001/jamasurg.2017.4472.

14. Adebayo O, Bhuiyan ZA, Ahmed Z. Exploring the effectiveness of artificial intelligence, machine learning and deep learning in trauma triage: A systematic review and meta-analysis. Digit Health 2023;9:20552076231205736. doi: 10.1177/20552076231205736.

15. Garwe T, Newgard CD, Stewart K, Wan Y, Cody P, Cutler J, et al. Enhancing utility of interfacility triage guidelines using machine learning: Development of the Geriatric Interfacility Trauma Triage score. J Trauma Acute Care Surg 2023;94(4):546-553. doi: 10.1097/TA.0000000000003846.

16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. doi: 10.1136/bmj.g7594.

17. Hashmi ZG, Kaji AH, Nathens AB. Practical guide to surgical data sets: National Trauma Data Bank (NTDB). JAMA Surg 2018;153(9):852-853. doi: 10.1001/jamasurg.2018.0483.

18. van der Sluijs R, Lokerman RD, Waalwijk JF, de Jongh MAC, Edwards MJR, den Hartog D, et al. Accuracy of pre-hospital trauma triage and field triage decision rules in children (P2-T2 study): an observational study. Lancet Child Adolesc Health 2020;4(4):290-298. doi: 10.1016/S2352-4642(19)30431-6.

19. Committee on Trauma American College of Surgeons. Resources for optimal care of the injured patient 2014. 2014, 633 Chicago, IL: N. Saint Clair St.; 2014.p. 60611-3211.

20. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. in proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016.

21. Sharma V, Kulkarni V, Jess E, Gilani F, Eurich D, Simpson SH, et al. Development and

validation of a machine learning model to estimate risk of adverse outcomes within 30 days of opioid dispensation. JAMA Netw Open 2022;5(12):e2248559. doi: 10.1001/jamanetworkopen.2022.48559.

22. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020;2(1):56-67. doi: 10.1038/s42256-019-0138-9.

23. Fahmy AS, Csecs I, Arafati A, Assana S, Yankama TT, Al-Otaibi T, et al. An explainable machine learning approach reveals prognostic significance of right ventricular dysfunction in nonischemic cardiomyopathy. JACC Cardiovasc Imaging 2022;15(5):766-779. doi: 10.1016/j.jcmg.2021.11.029.

24. Riley RD, Ensor J, Snell KIE, Harrell Jr FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441. doi: 10.1136/bmj.m441.

25. Jianxing H, Bo W, Jinsheng T, Liu Q, Peng M, Xiong S, et al. Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study. Lancet Digit Health 2023,5:e647-e656. doi: 10.1016/S2589-7500(23)00125-5.

26. L Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. Inhal Toxicol 2014;26(13):811-828. doi:10.3109/08958378.2014.955932.

27. Lerner EB, Cushman JT, Drendel AL, Badawy M, Shah MN, Guse CE, et al. Effect of the 2011 Revisions to the field triage guidelines on under- and over-triage rates for pediatric trauma patients. Prehosp Emerg Care 2017;21(4):456-460. doi: 10.1080/10903127.2017.1300717.

28. Hamada SR, Gauss T, Duchateau FX, Truchot J, Harrois A, Raux M, et al. Evaluation of the performance of French physician-staffed emergency medical service in the triage of major trauma patients. J Trauma Acute Care Surg 2014;76(6):1476-1483. doi: 10.1097/TA.0000000000000239.

29. Dinh MM, Bein KJ, Oliver M, Veillard A-S, Ivers R. Refining the trauma triage algorithm at an Australian major trauma centre: derivation and internal validation of a triage risk score. Eur J Trauma Emerg Surg 2014;40(1):67-74. doi: 10.1007/s00068-013-0315-1.

30. Dupont T, Kentish-Barnes N, Pochard F, Duchesnay E, Azoulay E. Prediction of post-traumatic stress disorder in family members of ICU patients: a machine learning approach. Intensive Care Med 2024;50(1):114-124. doi: 10.1007/s00134-023-07288-1.

31. Heffernan DS, Thakkar RK, Monaghan SF, Ravindran R, Adams Jr CA, Kozloff MS, et al. Normal presenting vital signs are unreliable in geriatric blunt trauma victims. J Trauma 2010;69(4):813-820. doi: 10.1097/TA.0b013e3181f41af8.

32. Nishijima DK, Shahlaie K, Sarkar K, Rudisill N, Holmes JF. Risk of unfavorable long-term outcome in older adults with traumatic intracranial hemorrhage and anticoagulant or antiplatelet use. Am J Emerg Med 2013;31(8):1244-1247. doi: 10.1016/j.ajem.2013.04.035.

33. Galimberti S, Graziano F, Maas AIR, Isernia G, Lecky F, Jain S, et al. Effect of frailty on 6-month outcome after traumatic brain injury: a multicentre cohort study with external validation. Lancet Neurol 2022;21(2):153-162. doi: 10.1016/S1474-4422(21)00374-4.

34. Ichwan B, Darbha S, Shah MN, Thompson L, Evans DC, Boulger CT, et al. Geriatric-specific triage criteria are more sensitive than standard adult criteria in identifying need for trauma center care in injured older adults. Ann Emerg Med 2015;65(1):92-100.e3. doi: 10.1016/j.annemergmed.2014.04.019.

35. Newgard CD, Holmes JF, Haukoos JS, Bulger EM, Staudenmayer K, Wittwer L, et al. Improving early identification of the high-risk elderly trauma patient by emergency medical services. Injury 2016;47(1):19-25. doi: 10.1016/j.injury.2015.09.010.

36. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22. doi: 10.1016/j.jclinepi.2019.02.004.

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

38. Larsson A, Berg J, Gellerfors M, Gerdin Wärnberg M. The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study. BMC Med Inform Decis Mak 2021;21(1):192. doi: 10.1186/s12911-021-01558-y.

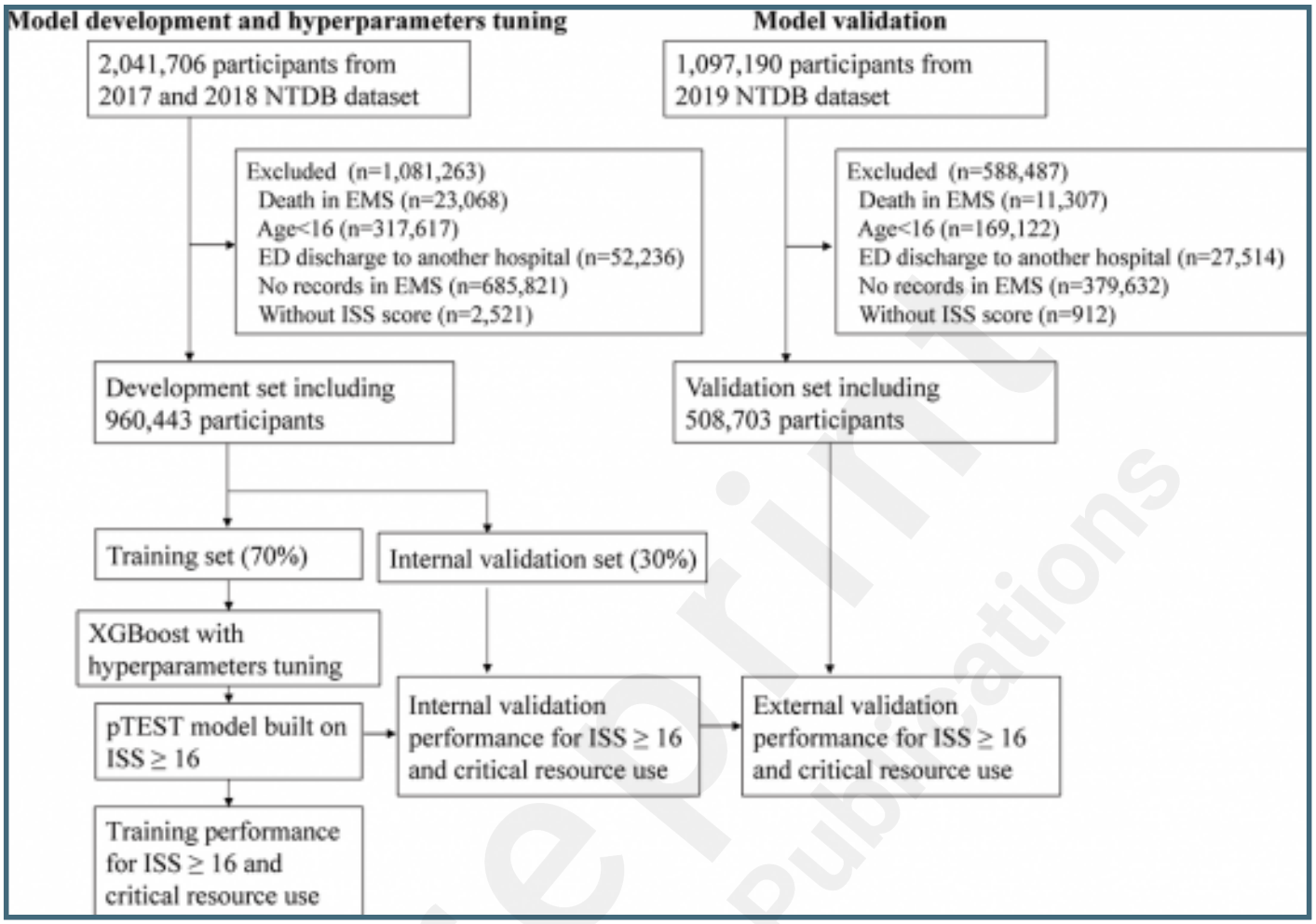
39. Soto JM, Zhang Y, Huang JH, Feng DX. An overview of the American trauma system. Chin J Traumatol 2018;21(2):77-79. doi: 10.1016/j.cjtee.2018.01.003.

40. van Rein EAJ, van der Sluijs R, Voskens FJ, Lansink KWW, Houwert RM, Lichtveld RA, et al. Development and validation of a prediction model for prehospital triage of trauma patients. JAMA Surg 2019;154(5):421-429. doi: 10.1001/jamasurg.2018.4752.

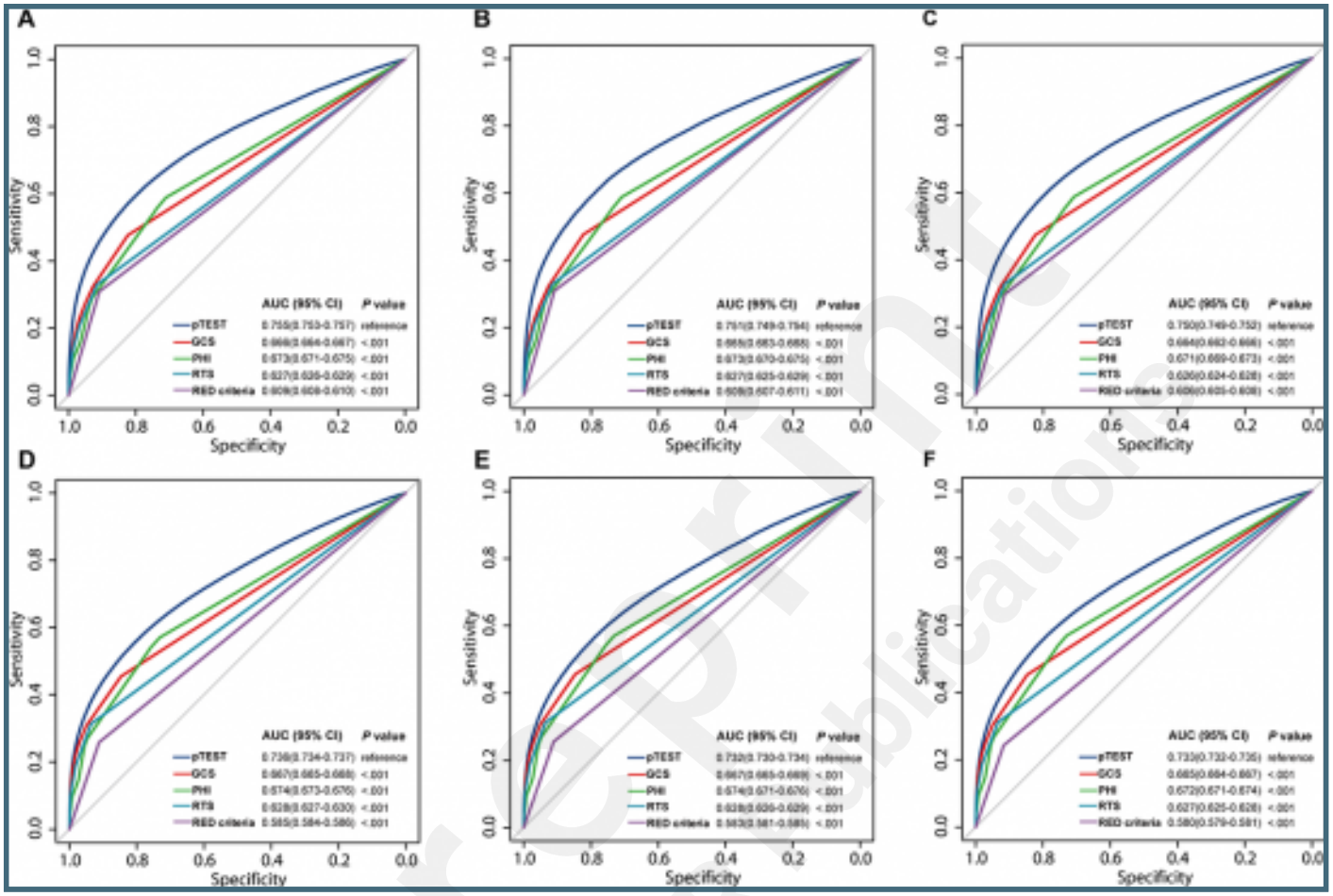
Supplementary Files

Figures

The flow chart of model development and validation.



The ROC curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.



Global model explanation for predicting severe trauma by the SHAP method in training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. In each variable, a dot is made for each single patient, representing the SHAP value of this variable. The colors of the dots demonstrate the actual values of the features, and the dots are stacked vertically to show density. (C) SHAP dependence plot. Each dependence plot shows the association between the actual value and the SHAP value of the variable, and each dot represents a single patient.

Multimedia Appendixes

TRIPOD Checklist in the manuscript.
URL: http://asset.jmir.pub/assets/6897de5ca6875d931d5c0a761218910f.docx

Detailed definition of severe trauma and critical resource use.
URL: http://asset.jmir.pub/assets/08d5e93e048abd567123b1f3a1c26295.docx

The candidate variables for model development.
URL: http://asset.jmir.pub/assets/3d5856803d00e7fa9addcea77be56043.docx

The missing proportion of the candidate variables for model development. (A) The missing proportion in training set. (B) The missing proportion in internal validation set. (C) The missing proportion in external validation set.
URL: http://asset.jmir.pub/assets/1c1d297669b71201421fb443dffa2528.docx

The demographic characteristics, vital signs and injury patterns between non- severe traumas and severe traumas in training set.
URL: http://asset.jmir.pub/assets/c6c7dd10d7054d5967daf6fb09ec462f.docx

The demographic characteristics, vital signs and injury patterns between non-severe traumas and severe traumas in internal validation set.
URL: http://asset.jmir.pub/assets/11831d32268ff2193db98ee05382e252.docx

The demographic characteristics, vital signs and injury patterns between non-severe traumas and severe traumas in external validation set.
URL: http://asset.jmir.pub/assets/1728be0af72ac1d1d7339866af764e35.docx

The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in training set.
URL: http://asset.jmir.pub/assets/d03e403a92a4de72be7db4bd59cd8b78.docx

The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in internal validation set.
URL: http://asset.jmir.pub/assets/e7c1cbe7baba4c4e976e75dbb02eed98.docx

The demographic characteristics, vital signs and injury patterns between no critical resource users and critical resource users in external validation set.
URL: http://asset.jmir.pub/assets/e44ab9f24f7973851e73217d0816528b.docx

Model performance metrics for predicting critical resource use fixed at a specificity of 0.5.
URL: http://asset.jmir.pub/assets/8a57390df70bb088c31ba53c8e6ac12c.docx

Model performance metrics for predicting severe trauma at the best thresholds with maximum Youden index.
URL: http://asset.jmir.pub/assets/b652edbe34743c7ca7bd72704702657f.docx

Model performance metrics for predicting critical resource use at the best thresholds with maximum Youden index.
URL: http://asset.jmir.pub/assets/e9244562d933042e906cc632724bb722.docx

Waterfall plot for the predicted probability of pTEST model related to the outcome. (A) Predicted probability of severe trauma in training set. (B) Predicted probability of severe trauma in internal validation set. (C) Predicted probability of severe trauma in external validation set. (D) Predicted probability of critical resource use in training set. (E) Predicted probability of critical resource use in internal validation set. (F) Predicted probability of critical resource use in external validation set.
URL: http://asset.jmir.pub/assets/9add51c9a53fad5d9a5faeed72df1fa4.docx

Calibration curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.
URL: http://asset.jmir.pub/assets/885d71edf264b3df987e69c3f5e9b28f.docx

DCA curves of five models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.

URL: http://asset.jmir.pub/assets/d02d5f8f973a49d82602cbc2d53d694b.docx

Global model explanation for predicting critical resource use by the SHAP method in training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. (C) SHAP dependence plot.
URL: http://asset.jmir.pub/assets/f5f42ec66e6a960a49fce793017775d8.docx

Local model explanation for predicting severe trauma by the SHAP method in training set. (A) Waterfall plot of risks contributed by each variable for a patient at high risk of severe trauma. (B) Waterfall plot for a patient at low risk of severe trauma.
URL: http://asset.jmir.pub/assets/365c84f24e6a415c7e7af0b629296a7d.docx

The performance of the pTEST model for predicting severe trauma in different subgroups.
URL: http://asset.jmir.pub/assets/7fb3e5653f5632827c97e0ae66322c58.docx

(A) The performance of the pTEST model for predicting critical resource use in different subgroups. (B) The undertriage and overtriage rates of the pTEST in external validation set with the fixed sensitivity, specificity and sample size.
URL: http://asset.jmir.pub/assets/5ec1f535767e136d92a71af9487420e7.docx