

A Machine Learning–Based Model to Predict Acute Traumatic Coagulopathy in Trauma Patients Upon Emergency Hospitalization

Clinical and Applied
Thrombosis/Hemostasis
Volume 26: 1-10
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1076029619897827
journals.sagepub.com/home/cat



Kaiyuan Li, MD¹, Huitao Wu, MD², Fei Pan, PhD¹, Li Chen, PhD¹,
Cong Feng, PhD¹, Yihao Liu, MD¹, Hui Hui, MD¹, Xiaoyu Cai, MD³,
Hebin Che, MD², Yulong Ma, PhD⁴, and Tanshi Li, PhD¹

Abstract

Acute traumatic coagulopathy (ATC) is an extremely common but silent murderer; this condition presents early after trauma and impacts approximately 30% of severely injured patients who are admitted to emergency departments (EDs). Given that conventional coagulation indicators usually require more than 1 hour after admission to yield results—a limitation that frequently prevents the ability for clinicians to make appropriate interventions during the optimal therapeutic window—it is clearly of vital importance to develop prediction models that can rapidly identify ATC; such models would also facilitate ancillary resource management and clinical decision support. Using the critical care Emergency Rescue Database and further collected data in ED, a total of 1385 patients were analyzed and cases with initial international normalized ratio (INR) values >1.5 upon admission to the ED met the defined diagnostic criteria for ATC; nontraumatic conditions with potentially disordered coagulation systems were excluded. A total of 818 individuals were collected from Emergency Rescue Database as derivation cohorts, then were split 7:3 into training and test data sets. A Pearson correlation matrix was used to initially identify likely key clinical features associated with ATC, and analysis of data distributions was undertaken prior to the selection of suitable modeling tools. Both machine learning (random forest) and traditional logistic regression were deployed for prediction modeling of ATC. After the model was built, another 587 patients were further collected in ED as validation cohorts. The ATC prediction models incorporated red blood cell count, Shock Index, base excess, lactate, diastolic blood pressure, and potential of hydrogen. Of 818 trauma patients filtered from the database, 747 (91.3%) patients did not present ATC (INR ≤ 1.5) and 71 (8.7%) patients had ATC (INR > 1.5) upon admission to the ED. Compared to the logistic regression model, the model based on the random forest algorithm showed better accuracy (94.0%, 95% confidence interval [CI]: 0.922-0.954 to 93.5%, 95% CI: 0.916-0.95), precision (93.3%, 95% CI: 0.914-0.948 to 93.1%, 95% CI: 0.912-0.946), F1 score (93.4%, 95% CI: 0.915-0.949 to 92%, 95% CI: 0.9-0.937), and recall score (94.0%, 95% CI: 0.922-0.954 to 93.5%, 95% CI: 0.916-0.95) but yielded lower area under the receiver operating characteristic curve (AU-ROC) (0.810, 95% CI: 0.673-0.918 to 0.849, 95% CI: 0.732-0.944) for predicting ATC in the trauma patients. The result is similar in the validation cohort. The values for classification accuracy, precision, F1 score, and recall score of random forest model were 0.916, 0.907, 0.901, and 0.917, while the AU-ROC was 0.830. The values for classification accuracy, precision, F1 score, and recall score of logistic regression model were 0.905, 0.887, 0.883, and 0.905, while the AU-ROC was 0.858. We developed and validated a

¹ Department of Emergency, The First Medical Center to Chinese People's Liberation Army General Hospital, Beijing, China

² National Engineering Laboratory for Medical Big Data Application Technology, The First Medical Center to Chinese People's Liberation Army General Hospital, Beijing, China

³ Department of Blood Transfusion, The First Medical Center to Chinese People's Liberation Army General Hospital, Beijing, China

⁴ Anesthesia and Operation Center, The First Medical Center to Chinese PLA General Hospital, Beijing, China

Corresponding Author:

Tanshi Li, Department of Emergency, The First Medical Center to Chinese People's Liberation Army General Hospital, No. 28, Fu Xing Road, Beijing 100853, China.

Email: lts301@163.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

prediction model based on objective and rapidly accessible clinical data that very confidently identify trauma patients at risk for ATC upon their arrival to the ED. Beyond highlighting the value of ED initial laboratory tests and vital signs when used in combination with data analysis and modeling, our study illustrates a practical method that should greatly facilitate both warning and guided target intervention for ATC.

Keywords

acute traumatic coagulopathy, prediction modeling, random forest, logistic regression, hemorrhagic shock

Date received: 11 September 2019; revised: 28 November 2019; accepted: 3 December 2019.

Introduction

Annually, traumatic injury leads to more than 5 million deaths worldwide, comprising 10% of global mortality, and trauma care is thus a huge burden on medical resources and health services.^{1,2} Uncontrolled hemorrhage is associated with 40% of traumatic deaths,^{3,4} and among such cases, ~60% of hemorrhagic deaths occur within 3 hours of admission and a total of 94% occur within 24 hours.⁵ A variety of factors contribute to massive bleeding, yet acute traumatic coagulopathy (ATC) is widely considered to be a leading cause, affecting up to 30% of severely injured patients; indeed, ATC is understood to predict up to 4-fold increases in bleeding-related mortality.⁶⁻⁹

Acute traumatic coagulopathy is an endogenous coagulopathy multifactorial with many mechanisms, including tissue damage, shock, hemodilution, hypothermia, acidemia, inflammation, and hypoperfusion,^{8,10-12} and early detection and intervention of ATC is known to significantly reduce mortality and to ameliorate the outcomes of trauma patients.^{5,13} Therefore, it is of great importance to develop prediction models that can alert clinicians to potential ATC cases when trauma patients arrive at the emergency department (ED). Ideally, such models would include the following characteristics: (1) Simplicity: The model needs to be based on data that are easily accessible upon admission to the ED. (2) Universality: Indicators for building models can be widely applied to medical institutions around the world. (3) Timeliness: Observed indicators can be obtained quickly, helping doctors get prediction result rapidly and enabling them to take model-guided interventions in time.^{14,15}

Several previous studies have reported the development of predictive models for ATC,^{16,17} including, for example, the Prediction of Acute Coagulopathy of Trauma (PACT) score. The PACT score is primarily based on prehospital treatment data including age, injury mechanism, prehospital Shock Index, Glasgow Coma Score values, prehospital cardiopulmonary resuscitation, and endotracheal intubation. However, due to the distinct prehospital practice patterns and variation in the experience and technical backgrounds of ambulance crew in different countries, these subjective variables are ultimately limited in clinical application. In short, there is a great opportunity to use objective clinical features assessed upon admission to the ED to develop more powerful predictive models for ATC.

To date, the diagnosis of ATC has been based on conventional coagulation indicators such as activated partial thromboplastin time, prothrombin time, and international normalized ratio (INR), among others,⁷ and these typically require a minimum of 1 to 2 hours of processing time after admission to the ED. This time lag can result in missing of therapeutic windows for treating ATC. Notably, a previous study with ATC cases showed that an INR >1.5 during admission to the ED is associated with multiple adverse prognoses.⁹ So, it is of vital importance to develop a prediction model that can identify ATC rapidly because it could inform ancillary resource management and clinical decision support to help saving the lives of trauma patients.

Here, we used machine learning methods (eg, random forest algorithm) to develop and validate a prediction model for ATC that is based on objective indicators which are already routinely obtained as patients are admitted at the hospital. Our study also developed a predictive model based on a more traditional logistic regression method, enabling us to rigorously compare the performance of the 2 models.

Methods

Source of Data

Known as Emergency Rescue Database used for modeling, this critical care database is an integrated and de-identified data set generated by the Chinese People's Liberation Army General Hospital and contains medical information for 22 941 critically ill patients from May 2015 to July 2017.¹⁸ The medical data represented in the database include human demographic information, triage target, medical records, vital signs, laboratory tests, and imaging examinations. For validation, medical information for another 924 severe trauma patients was collected in ED from April 2018 to May 2019. All the patients were admitted via the ED and were triaged to a critical rescue room.

Study Cohort

Patients whose triage target was classified as "trauma" were selected. Some of the patients had more than 1 admission to the ED; only the data for the first admission were extracted and analyzed. In order to rule out the nontraumatic conditions that may have confounded interpretations about the coagulation

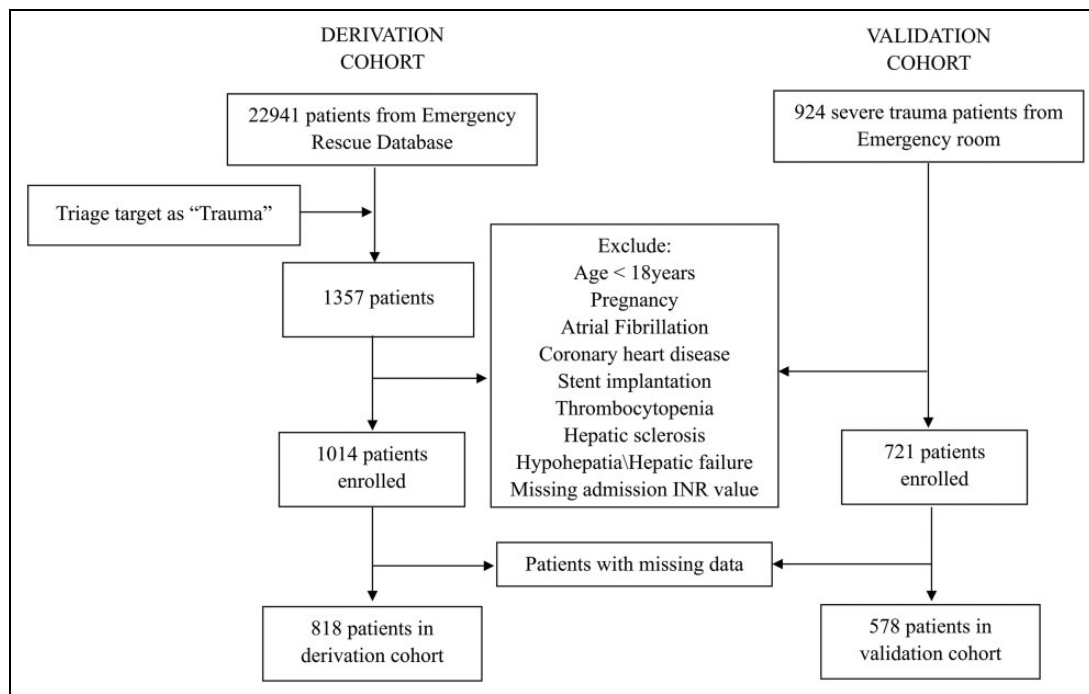


Figure 1. Flowchart of patient selection.

system, we performed keyword in the diagnosis charts, and patients with the following comorbidity or medical conditions were excluded from the cohort: (1) age <18 years, (2) pregnancy, (3) coronary heart disease, (4) atrial fibrillation, (5) history of stent implantation, (6) thrombocytopenia, (7) hepatic sclerosis, (8) hypohepatia/hepatic failure, and (9) missing admission INR value. Finally, patients with incomplete initial data of admission to the ED were also excluded. A total of 818 patients from Emergency Rescue Database were included as derivation cohort, and a total of 567 patients further collected from ED were included in the study as validation cohort (Figure 1).

Predictor and Outcome Definitions

The initial INR value >1.5 after admission to the ED was defined as the diagnostic criteria for ATC.⁹ The initial indicators for predictive modeling included:

1. Demographic variables: age and gender.

Triage level: for the Emergency Rescue Database, a 4-level triage scale for prioritization (wherein level 1 indicates the most severe) was used, and prior to final data entry for each patient, the triage level was curated by a highly trained and experienced registered nurse.

Initial ED admission vital signs: Temperature, heart rate (HR), respiratory rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), and SpO₂.

2. Shock Index: Admission HR divided by SBP.

Initial ED quick laboratory data: potential of hydrogen (PH), PO₂, PCO₂, calcium, base excess (BE), lactate (Lac), red blood cell (RBC) count, hemoglobin (HB), and platelet count.

Feature Selection and Model Evaluation

Our initial explorations analyzed the baseline levels for the data of the cohort. Clinical characteristics between the ATC group and the non-ATC group were compared using either Student *t* tests or rank-sum tests, as appropriate. The χ^2 tests or Fisher exact tests were used to compare differences in categorical variables. We further used Pearson correlation coefficients to distinguish the variables that may have impacted our target INR, which identified the top 6 variables; these were chosen for incorporation into the final model. Prior to algorithm selection, we analyzed the distribution of the data. Whether INR > 1.5 was used to separate the ATC group and non-ATC group. In a 3-D coordinate system, every combination of 3 factors was chosen for distinguishing ATC patients and non-ATC patients. The 2 groups present in comparatively 2 zone ("ATC zone" and "relative safety zone") in these coordinate systems. The most typical one is shown in Figure 2. So it was reasonable to select an algorithm from among classification algorithms (Figure 2). Two algorithms, either random forest or logistic regression, were developed to predict ATC. Accordingly, the clinical meanings could be learned through the coefficients of the derived model. Random forest is an ensemble learning method that grows many classification trees. To classify an object from

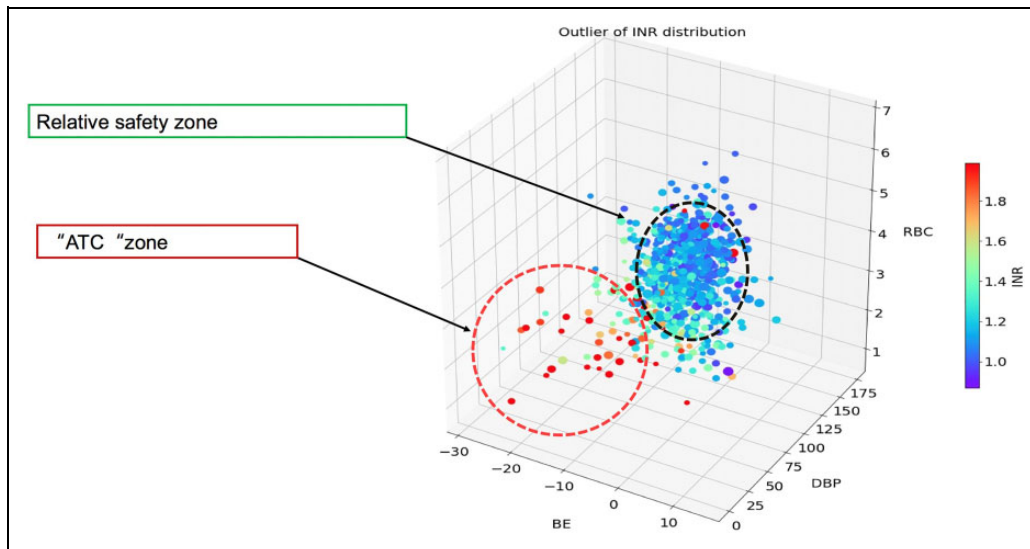


Figure 2. Distribution of the data. A 3-D coordinate system was used to distinguish the ATC group and non-ATC group. The admission criteria of ATC patients (in “ATC zone”) and non-ATC patients (in “relative safety zone”) are clearly distinguishable. ATC indicates acute traumatic coagulopathy.

an input vector, each tree gives a classification. The forest selects the classification that has most votes.

The derivation cohort was split into 2 data sets with the same proportion of ATC patients at a ratio of 7:3 training set:testing set, while the validation cohort was used as validating set. The model was trained by training set, while the prediction result of the model in testing and validating set was used to test and validate the performance of model. The area under the receiver operating characteristic curve (AU-ROC), classification accuracy, precision, F1 score, and recall score were used as metrics to evaluate the models. The classification accuracy is the proportion of the correct prediction in all prediction results. The precision is the proportion of true positive samples in all predicted positive samples. Recall score is the proportion of predicted positive samples in all true positive samples. The F1 score is an integrated metrics combining precision and recall. In the random forest algorithm, we used the Gini Index as the optimization criterion, with 1000 estimators used in calculation. In the logistic regression, the l2 penalty was used to prevent overfitting with $C = 100$. The baseline statistical analysis and hypothesis test were done in R studio 3.5.1. The whole process of data mining and model fitting were done in python 3.6.

Results

Cohort Characteristics

Of 1357 trauma patients initially identified in the derivation cohort, 1014 patients met the inclusion criteria for the current study, among which 196 (19.3%) patients had missing data for one or more covariates and were, therefore, excluded; finally, 818 patients with complete data were used for modeling (Figure 1). Demographic and other injury

Table 1. Characteristics Between the ATC and Non-ATC Groups in Derivation Cohort.^a

	INR ≤ 1.5 (n = 747)	INR > 1.5 (n = 71)	P
Age, mean (SD)	46.9 (17.3)	44.5 (16.7)	.259
Male sex, n (%)	589 (91)	56 (9)	.996
Triage level, n (%)			.044
Level 1	22 (76)	7 (24)	
Level 2	624 (92)	55 (8)	
Level 3	92 (92)	8 (8)	
Level 4	9 (90)	1 (10)	
Initial ED vital signs, mean (SD)			
Temperature (°C)	36.92 (0.75)	36.97 (1.08)	.743
HR	94 (22)	111 (27)	<.001
RR	21 (5)	23 (9)	.041
SBP (mm Hg)	127 (25)	107 (39)	<.001
DBP (mm Hg)	78 (15)	63 (23)	<.001
SpO ₂ (%)	96 (6)	92 (13)	.008
SI, mean (SD)	0.77 (0.24)	1.12 (0.50)	<.001
Initial ED blood gas analysis, mean (SD)			
PH	7.41 (0.07)	7.33 (0.14)	<.001
Po ₂ (mm Hg)	116 (56)	131 (95)	.232
Pco ₂ (mm Hg)	37 (7)	34 (12)	.066
Calcium	1.23 (0.30)	1.13 (0.18)	<.001
BE (mmol/L)	-0.97 (4.12)	-7.13 (6.59)	<.001
Lac (mmol/L)	2.3 (1.9)	5.6 (3.7)	<.001
Initial ED blood test, mean (SD)			
RBC (10 ¹² /L)	4.06 (0.80)	3.09 (0.94)	<.001
HB (g/L)	125 (26)	93 (29)	<.001
Platelet (10 ⁹ /L)	210 (82)	178 (135)	.054

Abbreviations: ATC, acute traumatic coagulopathy; BE, base excess; DBP, diastolic blood pressure; ED, emergency department; Hb, hemoglobin; HR, heart rate; INR, international normalized ratio; Lac, lactate; RR, respiratory rate; SBP, systolic blood pressure; SD, standard deviation; SI, Shock Index; RBC, red blood cell.

^aValues reported as median (SD) or number (%).

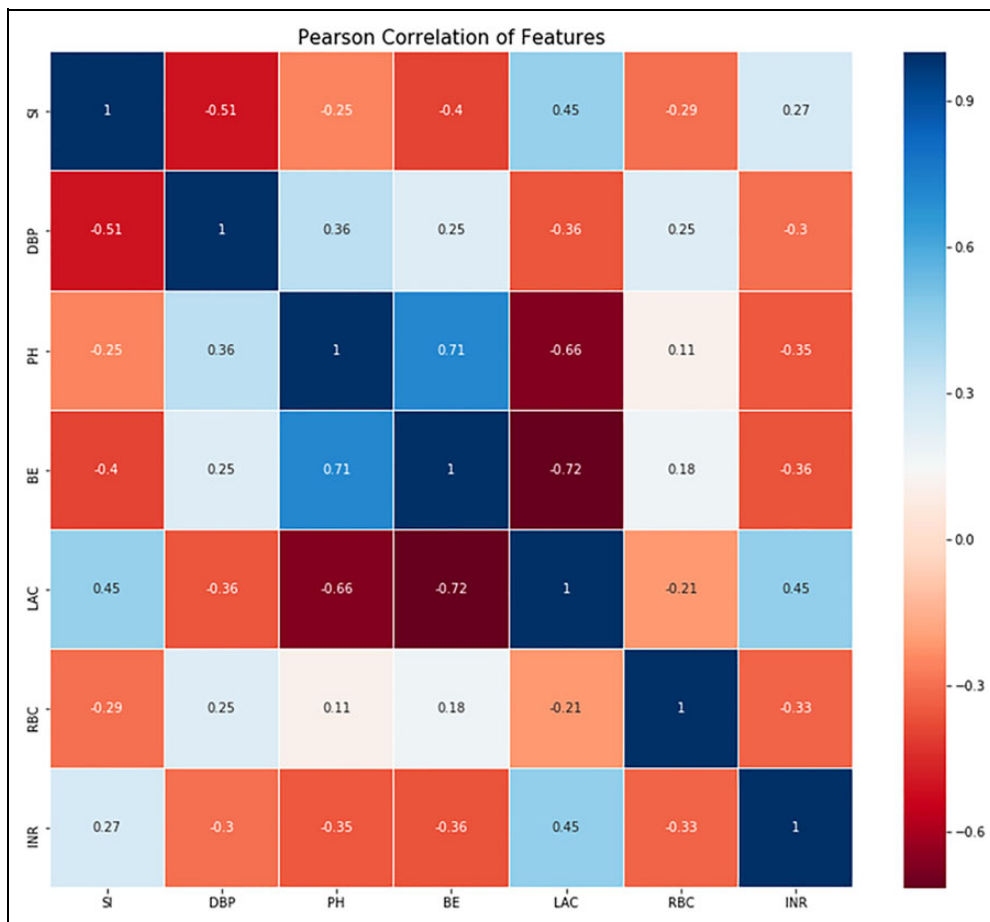


Figure 3. Pearson correlation of clinical features. Pearson correlation coefficient was used to distinguish the clinical features that may affect INR. The top 6 ranking variables were Lac (0.45), BE (−0.36), PH (−0.35), RBC (−0.33), DBP (−0.3), and SI (0.27). Positive values indicate positive correlation and negative values indicate negative correlation. BE indicates base excess; DBP, lactate, diastolic blood pressure; INR, international normalized ratio; Lac, lactate; RBC, red blood cell; SI, Shock Index.

Table 2. Results From the Logistic Regression Model.

	Intercept	SI	DBP	PH	BE	Lac	RBC
Coef	−7.172	1.367	−0.008	0.930	−0.099	0.118	−0.949
Exp (coef.)	−	3.923	0.992	2.535	0.905	1.125	0.387

Abbreviations: BE, base excess; DBP, diastolic blood pressure; Lac, lactate; RBC, red blood cell; SI: Shock Index; PH: potential of hydrogen.

characteristics for the 818 trauma patients for the primary analysis are shown in Table 1. Seven hundred forty-seven (91.3%) patients did not present ATC, and 71 (8.7%) patients had ATC at their ED admission. We noted that patients with ATC expressed significantly higher values for clinical features including HR, SI, BE, and Lac but had lower values for SBP, DBP, SpO₂, PH, calcium, RBC, and HB. The details are shown in Table 1. We generated a Pearson correlation coefficient matrix (Figure 3) which revealed that the top 6 variables associated with ATC were Lac (0.45), BE (−0.36), PH (−0.35), RBC (−0.33), DBP

(−0.3), and SI (0.27). Both logistic regression and random forest algorithm were used to fit the model based on these 6 variables.

Logistic Regression Model

The hyperparameters used in our analysis were as follows: $C = 100.0$, $dual = False$, $fit_intercept = True$, $intercept_scaling = 1$, $max_iter = 100$, $n_jobs = 1$, $penalty = "l2"$, $random\ state = 1$, $solver = "liblinear"$, $tol = 0.0001$, $verbose = 0$, and $warm\ start = False$. In the logistic regression model, the coefficient and odds ratio (OR) for the 6 selected variables are shown in Table 2. Notably, DBP (coefficient = −0.008, OR for each 1 mm Hg increase = 0.992), BE (coefficient = −0.099, OR for each 1 mmol/L increase = 0.905), RBC (coefficient = −0.949, OR for each 10¹²/L increase = 0.387) were associated with decreased probability of ATC. In contrast, SI (coefficient = 1.367, OR for each 1-unit increase = 3.923), PH (coefficient = 0.930, OR for each 1-unit increase = 2.535), and Lac (coefficient = 0.118, OR for each 1 mmol/L increase, 1.125) were associated with increased probability of ATC (Table 2).

Random Forest

The hyperparameters used in our analysis were as follows: bootstrap = True, criterion = “gini,” max_depth = 10, max_features = “auto,” min_samples_split = 2, n_estimators = 1000, n_jobs = 2, random_state = 1, verbose = 0, and warm_start = False. We classified all the patients according to their ATC status and then used the random forest algorithm to assess the feature importance for each of the 6 clinical features on the disease (Figure 4). Specifically, feature importance was calculated as the sum of the decrease in error when split by a variable, which reflects the contribution that each variable makes in classifying a patient’s status in the ATC or non-ATC group; this revealed that indicators for traumatic coagulopathy were ranked as follows: RBC (0.254), SI (0.180), BE (0.175), Lac (0.146), DBP (0.141), and PH (0.104; Figure 4).

Performance Comparison of the Models

Model discrimination was assessed using the AU-ROC, besides which we also applied standard machine learning evaluation metrics including accuracy, precision, recall, and F1 score. The details of the results are presented in Table 3. Notably, the random forest model offers superior accuracy (0.940) and precision (0.933), as well as higher F1 (0.934) and recall scores

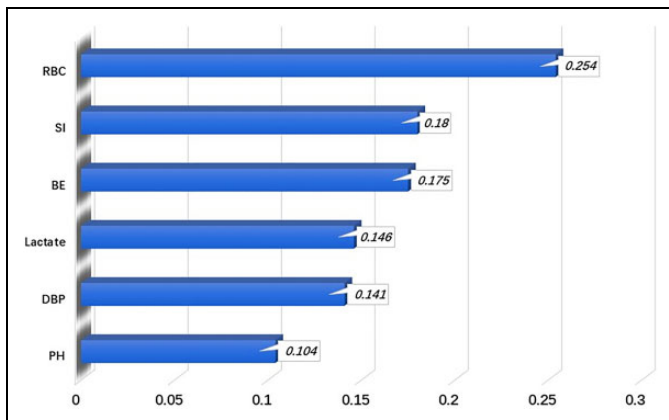


Figure 4. Feature importance derived from the random forest. The most important factors on the disease by the random forest algorithm. Feature importance reflects the contribution of each variable makes in classifying ATC groups and non-ATC groups. The clinical features affecting ATC were ranked thusly: RBC (0.254), SI (0.180), BE (0.175), Lac (0.146), DBP (0.141), and PH (0.104). ATC indicates acute traumatic coagulopathy; BE, base excess; DBP, lactate, diastolic blood pressure; Lac, lactate; RBC, red blood cell.

(0.940). In contrast, the logistic regression–based model had a higher AU-ROC (0.849; Figure 5).

Performance Comparison of the Validation Cohort

For the random forest algorithm, the values for classification accuracy, precision, F1 score, and recall score of our prediction model were 0.916 (95% CI: 0.891-0.936), 0.907 (95% CI: 0.881-0.928), 0.901 (95% CI: 0.874-0.922), and 0.917 (95% CI: 0.892-0.937), while the AU-ROC was 0.830 (95% CI: 0.770-0.887; (Table 4). For the logistic regression, the values for classification accuracy, precision, F1 score, and recall score of our prediction model were 0.905 (95% CI: 0.879-0.926), 0.887 (95% CI: 0.859-0.910), 0.883 (95% CI: 0.855-0.906), and 0.905 (95% CI: 0.879-0.926), while the AU-ROC was 0.858 (95% CI: 0.808-0.903; Figure 6).

Discussion

In this study, by establishing 2 predictive models for ATC, we screened for admission indicators that are informative for the onset of ATC. With our newly developed models—both of which incorporated only 6 objective and readily obtainable indicators that are already universally available during admission to the ED—we can very confidently predict patients at risk for ATC (random forest: accuracy 94%, precision 93.3%, recall 94%, and F1 score 93.4% vs logistic regression: accuracy 93.5%, precision 93.1%, recall 93.5%, and F1 score 92%). Preliminary inferential statistical and correlation-based analyses suggested that the most informative clinical features for ATC were likely RBC, SI, BE, Lac, DBP, and PH (Figure 4); findings are in line with other studies.^{7,11,19}

It is important to note that both models performed very well in both test and validation samples. The discrepancy between the 2 modeling strategies is that the random forest model is better in terms of accuracy, precision, F1 score, and recall scores, while the logistic regression model offered superior predicted performance in AU-ROC. Considering a lower rate of missed diagnosis is preferred in medical model, the logistic model may provide more information in this given topic. Further, there are many reports in the literature which highlight that machine learning–based models can often outperform logistic regression models for predicting various medical outcomes,²⁰⁻²³ yet our study emphasizes that traditional regression-based models also often perform well. It may be because our study cohort is not large enough to show critical difference in every performance metrics. Further study could be done if possible. Ultimately, our comparison of the

Table 3. Performance Comparison Between the Random Forest and Logistic Regression Prediction Models for ATC.

	AU-ROC (CI) (%)	Accuracy (CI) (%)	Precision (CI) (%)	F1 score (CI) (%)	Recall (CI) (%)
Random forest	0.810 (0.673-0.918)	0.940 (0.922-0.954)	0.933 (0.914-0.948)	0.934 (0.915-0.949)	0.940 (0.922-0.954)
Logistic regression	0.849 (0.732-0.944)	0.935 (0.916-0.95)	0.931 (0.912-0.946)	0.920 (0.9-0.937)	0.935 (0.916-0.95)

Abbreviations: ATC, acute traumatic coagulopathy; AU-ROC, area under receiver operating characteristic curve; CI, confidence interval.

performance of the 2 models suggests that clinical computer-aided workflows should probably combine a variety of models to capitalize on their individual strengths while overcoming any particular limitations.

Potentially pathogenic roles of RBCs in coagulopathy has been little explored until recently, yet there are now multiple studies that propose multiple functions for RBCs in hemostasis and in thrombosis in trauma patients. The hemorheological effects of RBCs can be an essential prothrombotic factor because impaired blood flow is a known pathophysiological mechanism of thrombosis.²⁴ It is also known that many bleeding disorders can be treated by increasing the RBC count, regardless of the platelet level.²⁵ Furthermore, RBC count is interconnected with clot contraction, fibrinolysis, and endothelium homeostasis.²⁶ Our random forest model ranked RBC first in terms of feature importance, a finding which affirms previously reported findings.

Shock Index and DBP are sensitive indices of shock.^{27,28} Corresponding to our observation, hemorrhagic shock (HS) has been widely illustrated to be significantly correlated with the occurrence of ATC in both clinical and animal experiments.²⁹⁻³² Shock Index was interpreted to be a favorable predictor of massive hemorrhage in both prehospital and ED settings, when $SI \geq 0.9$ postinjury can recognize patients with critical bleeding.²⁷ As a primary indicator of hemodynamic monitoring,

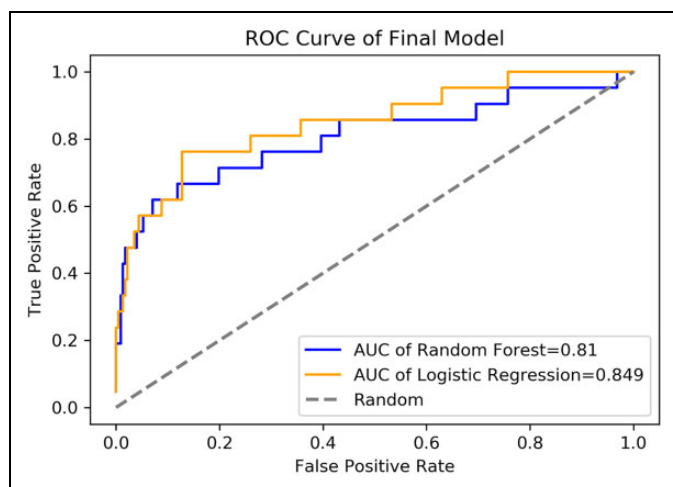


Figure 5. Discrimination of the 2 acute traumatic coagulopathy prediction models. Model discrimination was assessed using AU-ROC. In the logistic regression, AU-ROC is 0.849 (95% CI: 0.732-0.944), while in the random forest model, the AU-ROC is 0.810 (95% CI: 0.673-0.918). AU-ROC indicates area under receiver operating characteristic curve; CI, confidence interval.

DBP may be predictive for patient mortality in cardiogenic shock and/or septic shock.^{28,33} For HS, the HR undergoes a compensatory increase due to hypovolemia that leads to a short diastolic phase and results in the elevation of DBP. Furthermore, ATC is associated with metabolic acidosis that originates from systemic hypoperfusion as a result of hypovolemia.^{7,34} As excellent indicators of hypoperfusion, BE, Lac, and PH predict higher incidence of shock-related complications such as acute respiratory distress syndrome, multi-organ failure, and the need for transfusion.³⁵⁻³⁸

Similar to past prediction models like PACT score,¹⁷ we intended to incorporate risk stratification into our model, so we selected patients with severe injuries with triage target reported as “trauma” and narrowed down to patients who were assigned to a critical rescue room. In contrast to previous models, our model is based on the routinely obtained data that are taken during admission to the ED, rather than data acquired in a prehospital stage. On the one hand, this means our model inherently lags behind in prediction window compared to the PACT score. On the other hand, however, given that we applied objective indicators that showed strong linear relationships with the occurrence of ATC, we anticipate that our model will be extremely helpful in the ED setting, especially as computer-aided prediction and triage technologies become more and more common. Our study has several limitations. First, we have used admission INR >1.5 to define ATC, a value that has been well verified in both civil and military settings.^{9,39} However, this selected value cannot fully account for ATC pathogenesis; consider, for example, the known role of fibrinolysis and insufficient platelet function in ATC.⁴⁰⁻⁴² Multiple studies have suggested that viscoelastic tests could be effective supplements for the detection of ATC, and these have been increasingly applied in trauma situations^{43,44}; they provide rapid information about the underlying mechanism of ATC and allow clinicians to focus on particular aspects of clotting, facilitating targeted coagulation supervision and intervention(s) in accordance with the particular needs of a given individual. Unfortunately, the Emergency Rescue Database had relatively little information from viscoelastic test for subjects. As such tests become increasingly common, there should be adequate data to conduct follow-on analyses from our modeling work which will likely add yet better predictive performance for ATC. Second, injury severity score (ISS) has been independently associated with increased coagulopathy,^{34,45,46} yet in practice—owing to the need for staff training as well as enormous workloads and ED urgency—it is difficult to routinely calculate ISS. We did anticipate this in our study design and had intended to enroll triage level as a potential proxy indicator

Table 4. Validation Comparison Between the Random Forest and Logistic Regression Prediction Models.

	AU-ROC	Accuracy	Precision	F1 score	Recall
Random forest	0.830 (0.770-0.887)	0.916 (0.891-0.936)	0.907 (0.881-0.928)	0.901 (0.874-0.922)	0.917 (0.892-0.937)
Logistic regression	0.858 (0.808-0.903)	0.905 (0.879-0.926)	0.887 (0.859-0.910)	0.883 (0.855-0.906)	0.905 (0.879-0.926)

Abbreviation: AU-ROC, area under receiver operating characteristic curve.

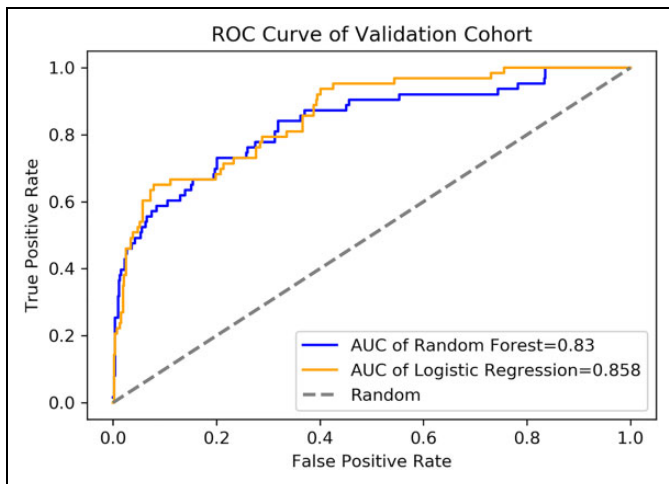


Figure 6. The AUC-ROC of validation cohort. Validation result was assessed using AU-ROC. In the logistic regression, AU-ROC is 0.858 (95% CI: 0.808-0.903), while in the random forest model, the AU-ROC is 0.830 (95% CI: 0.770-0.887). AU-ROC indicates area under receiver operating characteristic curve; CI, confidence interval.

for severity instead of ISS. However, our data analysis revealed that triage level did not show an obvious correlation or importance for ATC. Perhaps triage modes could be optimized in the future to better characterize the severity of trauma. Third, our single-center study will obviously need external validation with data from other trauma centers to enhance its application.

Conclusion

We developed a prediction model based on objective, rapidly accessible data that can powerfully predict which trauma patients are at risk for ATC upon admission to the ED. An important purpose of our study is that routinely acquired objective clinical features offer a rich source of predictive power as more and more computer-aided workflows are incorporated into modern medicine. Moreover, we ultimately found that a machine learning algorithm did offer certain benefits as we developed predictive models, but our impressive predictive power from a more traditional logistic regression model also highlighted the fact that computer-aided clinical guidance tools can profit from both emerging and traditional methods from computer science and biostatistics.

Authors' Note

Kaiyuan Li conceived the study and drafted the manuscript. Huitao Wu participated in study design and performed the analysis. Fei Pan, Li Chen, Cong Feng, and Yulong Ma helped to revise the manuscript. Hebin Che, Yihao Liu, Hui Hui, and Xiaoyu Cai participated in data extraction. Tanshi Li interpreted the results. All authors read and approved the final manuscript. Kaiyuan Li and Huitao Wu contributed equally to this study. The data that support the findings of this study are available from Chinese National Engineering Laboratory for Medical Big Data Application Technology but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available.

Data are however available from the authors upon reasonable request and with permission of Chinese National Engineering Laboratory for Medical Big Data Application Technology. Institutional review board of Chinese PLA General Hospital approved use of the Emergency Rescue Database and granted a waiver of informed consent for retrospective data analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by National Natural Science Foundation of China [grant number 81701961], Military Medical Innovation Project [grant numbers 14CXZ005, BWS14J042], The Big Data R&D Project of the PLA General Hospital [grant number 2017MBD-30]; and National Key R&D Program of China [grant number 2019YFF0302300], which had no role in study design, data collection, analysis or interpretation, or manuscript preparation.

ORCID iD

Tanshi Li  <https://orcid.org/0000-0001-6528-9595>

References

1. Injuries. <https://www.who.int/topics/injuries/about/en/>. Accessed Apr 10, 2019.
2. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the global burden of disease study 2010. *Lancet*. 2012;380(9859):2163-2196.
3. Cannon JW. Hemorrhagic shock. *N Engl J Med*. 2018;378(4):370-379.
4. Cothren CC, Moore EE, Hedegaard HB, Meng K. Epidemiology of urban trauma deaths: a comprehensive reassessment 10 years later. *World J Surg*. 2007;31(7):1507-1511.
5. Holcomb JB, Del Junco DJ, Fox EE, et al. The prospective, observational, multicenter, major trauma transfusion (PROMTTT) study: comparative effectiveness of a time-varying treatment with competing risks. *JAMA Surg*. 2013;148(2):127-136.
6. MacLeod JB, Lynn M, McKenney MG, Cohn SM, Murtha M. Early coagulopathy predicts mortality in trauma. *J Trauma*. 2003;55(1):39-44.
7. Frith D, Goslings JC, Gaarder C, et al. Definition and drivers of acute traumatic coagulopathy: clinical and experimental investigations. *J Thromb Haemost*. 2010;8(9):1919-1925.
8. Chang R, Cardenas JC, Wade CE, Holcomb JB. Advances in the understanding of trauma-induced coagulopathy. *Blood*. 2016;128(8):1043-1049.
9. Peltan ID, Vande Vusse LK, Maier RV, Watkins TR. An international normalized ratio-based definition of acute traumatic coagulopathy is associated with mortality, venous thromboembolism, and multiple organ failure after injury. *Crit Care Med*. 2015;43(7):1429-1438.

10. Maegele M. Frequency, risk stratification and therapeutic management of acute post-traumatic coagulopathy. *Vox Sang.* 2009; 97(1):39-49.
11. Cap A, Hunt B. Acute traumatic coagulopathy. *Curr Opin Crit Care.* 2014;20(6):638-645.
12. Lier H, Bottiger BW, Hinkelbein J, Krep H, Bernhard M. Coagulation management in multiple trauma: a systematic review. *Intensive Care Med.* 2011;37(4):572-582.
13. CRASH-2 Trial Collaborators, Shakur H, Roberts I, et al. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet.* 2010;376(9734):23-32.
14. Wu TB, Wu S, Buoni M, et al. Computational model for hyperfibrinolytic onset of acute traumatic coagulopathy. *Ann Biomed Eng.* 2018;46(8):1173-1182.
15. Reisner AT, Khitrov MY, Chen L, et al. Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. *Appl Clin Inform.* 2013;4(3): 392-402.
16. Mitra B, Cameron PA, Mori A, et al. Early prediction of acute traumatic coagulopathy. *Resuscitation.* 2011;82(9):1208-1213.
17. Peltan ID, Rowhani-Rahbar A, Vande Vusse LK, et al. Development and validation of a prehospital prediction model for acute traumatic coagulopathy. *Crit Care.* 2016;20(1):371.
18. Zhao Y, Wang J, Pan F, et al. Pilot research: construction of emergency rescue database [in Chinese]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue.* 2018;30(6):609-612.
19. Giordano S, Spiezia L, Campello E, Simioni P. The current understanding of trauma-induced coagulopathy (TIC): a focused review on pathophysiology. *Intern Emerg Med.* 2017;12(7): 981-991.
20. Hernesniemi JA, Mahdiani S, Tynkkynen JA, et al. Extensive phenotype data and machine learning in prediction of mortality in acute coronary syndrome—the MADDEC study. *Ann Med.* 2019:1-8.
21. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* 2019;23(1):112.
22. Lee S, Choe EK, Park B. Exploration of machine learning for hyperuricemia prediction models based on basic health checkup tests. *J Clin Med.* 2019;8(2):172.
23. Lee HC, Yoon SB, Yang SM, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. *J Clin Med.* 2018; 7(11):428.
24. Wolberg AS, Aleman MM, Leiderman K, Machlus KR. Procoagulant activity in hemostasis and thrombosis: Virchow's triad revisited. *Anesth Analg.* 2012;114(2):275-285.
25. Weiss HJ, Lages B, Hoffmann T, Turitto VT. Correction of the platelet adhesion defect in delta-storage pool deficiency at elevated hematocrit—possible role of adenosine diphosphate. *Blood.* 1996;87(10):4214-4222.
26. Litvinov RI, Weisel JW. Role of red blood cells in haemostasis and thrombosis. *ISBT Sci Ser.* 2017;12(1):176-183.
27. Olausson A, Blackburn T, Mitra B, Fitzgerald M. Review article: shock index for prediction of critical bleeding post-trauma: a systematic review. *Emerg Med Australas.* 2014;26(3):223-228.
28. Axler O. Low diastolic blood pressure as best predictor of mortality in cardiogenic shock. *Crit Care Med.* 2013;41(11): 2644-2647.
29. Torres Filho I. Hemorrhagic shock and the microvasculature. *Compr Physiol.* 2017;8(1):61-101.
30. Gando S, Sawamura A, Hayakawa M. Trauma, shock, and disseminated intravascular coagulation: lessons from the classical literature. *Ann Surg.* 2011;254(1):10-19.
31. White NJ, Martin EJ, Brophy DF, Ward KR. Coagulopathy and traumatic shock: characterizing hemostatic function during the critical period prior to fluid resuscitation. *Resuscitation.* 2010; 81(1):111-116.
32. Iwamoto S, Takasu A, Sakamoto T. Therapeutic mild hypothermia: effects on coagulopathy and survival in a rat hemorrhagic shock model. *J Trauma.* 2010;68(3):669-675.
33. Benchekroune S, Karpati PC, Berton C, et al. Diastolic arterial blood pressure: a reliable early predictor of survival in human septic shock. *J Trauma.* 2008;64(5):1188-1195.
34. Floccard B, Rugeri L, Faure A, et al. Early coagulopathy in trauma patients: an on-scene and hospital admission study. *Injury.* 2012;43(1):26-32.
35. Rixen D, Raum M, Bouillon B, Lefering R, Neugebauer E. Arbeitsgemeinschaft "polytrauma" of the deutsche gesellschaft fur U: base deficit development and its prognostic significance in posttrauma critical illness: an analysis by the trauma registry of the deutsche gesellschaft fur unfallchirurgie. *Shock.* 2001;15(2): 83-89.
36. Kiyatkin ME, Bakker J. Lactate and microcirculation as suitable targets for hemodynamic optimization in resuscitation of circulatory shock. *Curr Opin Crit Care.* 2017;23(4):348-354.
37. Gale SC, Kocik JF, Creath R, Crystal JS, Dombrovskiy VY. A comparison of initial lactate and initial base deficit as predictors of mortality after severe blunt trauma. *J Surg Res.* 2016;205(2): 446-455.
38. Brohi K, Cohen MJ, Ganter MT, Matthay MA, Mackersie RC, Pittet JF. Acute traumatic coagulopathy: initiated by hypoperfusion: modulated through the protein C pathway? *Ann Surg.* 2007; 245(5):812-818.
39. Simmons JW, White CE, Ritchie JD, Hardin MO, Dubick MA, Blackbourne LH. Mechanism of injury affects acute coagulopathy of trauma in combat casualties. *J Trauma.* 2011;71(1 suppl): S74-S77.
40. Davenport RA, Guerreiro M, Frith D, et al. Activated protein C drives the hyperfibrinolysis of acute traumatic coagulopathy. *Anesthesiology.* 2017;126(1):115-127.
41. Madurska MJ, Sachse KA, Jansen JO, Rasmussen TE, Morrison JJ. Fibrinolysis in trauma: a review. *Eur J Trauma Emerg Surg.* 2018;44(1):35-44.
42. White NJ, Newton JC, Martin EJ, et al. Clot formation is associated with fibrinogen and platelet forces in a cohort of severely injured emergency department trauma patients. *Shock.* 2015; 44(suppl 1):39-44.

43. Maegele M, Nardi G, Schochl H. Hemotherapy algorithm for the management of trauma-induced coagulopathy: the German and European perspective. *Curr Opin Anaesthesiol.* 2017;30(2):257-264.
44. Hagemo JS, Christiaans SC, Stanworth SJ, et al. Detection of acute traumatic coagulopathy and massive transfusion requirements by means of rotational thromboelastometry: an international prospective validation study. *Crit Care.* 2015;19:97.
45. Cohen MJ, Kutcher M, Redick B, et al. Clinical and mechanistic drivers of acute traumatic coagulopathy. *J Trauma Acute Care Surg.* 2013;75(1 suppl 1):S40-S47.
46. Wafaisade A, Wutzler S, Lefering R, et al. Trauma registry of DGU: drivers of acute coagulopathy after severe trauma: a multivariate analysis of 1987 patients. *Emerg Med J.* 2010;27(12):934-939.