Development and Validation of Dynamic Predictive Models Using Vital Signs for Trauma-associated Severe Hemorrhage: a Comparative Study

Chengyu Guo

Nankai University

Minghui Gong

Tsinghua University

Qiaochu Shen

Tsinghua University

Ruolin Wang

Tsinghua University

Hongliang Zhang

General Hospital of the PLA

Junkang Wang

General Hospital of the PLA

Tanshi Li (✉ its301@163.com)

General Hospital of the PLA

Chunping Li

Tsinghua University

Research Article

Keywords: Trauma-associated severe hemorrhage, dynamic predictive models, vital signs, machine learning, deep learning

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1141709/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: This study aimed to develop and to validate dynamic predictive models for trauma-associated severe hemorrhage based on vital signs to register early warning and dynamic prediction of severe hemorrhage in trauma patients.

Methods: The MIMIC-IV cohort was collected retrospectively. The inclusion criteria were trauma patients aged ≥ 16 years with complete clinical data. Heart rate, respiratory rate, systolic blood pressure, diastolic blood pressure, and peripheral oxygen saturation were extracted as predictive variables. Based on logistic regression, support vector machine, random forest, adaptive boosting, gated recurrent unit, and gated recurrent unit-d, predictive models for trauma-associated severe hemorrhage were developed and validated to dynamically predict whether severe hemorrhage will occur in trauma patients in the next 1 h/2 h/3 h. This study was based on the Trauma database of the General Hospital of the People's Liberation Army for external validation. The models were developed and validated using Python 3.8.5 software. SPSS.21 software was used for statistical analysis.

Results: Of the 7522 trauma patients in the MIMIC-IV cohort, 283 (3.76%) had a severe hemorrhage. The area under the curve of the gated recurrent unit-d model was the best in the 1 h (0.946 ± 0.029), 2 h (0.940 ± 0.032), and 3 h groups (0.943 ± 0.034), and there was no significant difference among the three groups. In the Trauma cohort, the area under the curve of the gated recurrent unit-d model also achieved the best performance in the 1 h (0.779 ± 0.013), 2 h (0.780 ± 0.008), and 3 h groups (0.778 ± 0.009), and there was no significant difference among the three groups. When comparing the gated recurrent unit-d model with the traditional scoring systems, the gated recurrent unit-d model still has advantages. Moreover, we have developed a web-based predictive system to help clinicians use our models.

Conclusions: This study developed and validated dynamic predictive models for trauma-associated severe hemorrhage based on vital signs to assist pre-hospital or in-hospital emergency personnel to make decisions, and the gated recurrent unit-d model performed best.

Trial registration:

The MIMIC-IV database was previously de-identified and reviewed by the institutional review board (IRB) of its host organization and determined to be exempted from subsequent IRB. We obtained the administrative permissions to use the database (Certification Number: 27959316) for our research, after completing the National Institutes of Health web-based training course: Protecting Human Research Participants. We were reviewed and approved by the Ethics Committee of Chinese PLA General Hospital to use the Trauma database. The ethical batch number is S2021-466-01. Moreover, the informed consent of subjects was waived by the Ethics Committee of Chinese PLA General Hospital.

Background

Trauma is a major global public health issue, contributing to approximately 1 in 10 mortalities and resulting in the annual worldwide death of more than 5.8 million people. Severe hemorrhage is the main cause of preventable death due to trauma. Approximately 40% of trauma deaths are attributed to severe hemorrhage, and up to 50% of such cases are reported dead on arrival at the hospital [1–4].

Severe hemorrhage can lead to hemorrhagic shock, acute traumatic coagulopathy, and multiple organ dysfunction syndrome in trauma patients with compensatory shock, occult internal bleeding, atypical signs, or multiple injuries. If not found and treated on time, these may eventually lead to death [5]. Emergency personnel involved in pre-hospital and in-hospital first aid should assess the severity of trauma hemorrhage early and identify patients with severe hemorrhage to effectively make triage and evacuation decisions and to implement life-saving interventions, such as damage control surgery, damage control resuscitation, and massive transfusion programs early. Therefore, the construction of trauma-associated severe hemorrhage (TASH) predictive models and early identification of severe hemorrhage are very important to patients' outcomes [6, 7].

At present, the predictive models for severe hemorrhage are mostly scoring systems based on logistic regression, such as the TASH score [8], Assessment of Blood Consumption (ABC) score [9], and Prince of Wales (PWH) score [10]. These scores are time-consuming and complex, often requiring laboratory values or ultrasound evaluation upon arrival at the hospital to calculate the results. Furthermore, most of these scoring systems are static evaluations using a single measurement, and the long detection intervals and invasive operations of laboratory or ultrasound examinations make dynamic monitoring difficult to achieve.

In recent years, with the development of machine learning technology, especially sub-domain deep learning, as well as the emergence of electronic health records and explosive growth of the amount of information stored in it, current studies are increasingly applying machine learning algorithms to the medical field.

This study explores the combination of artificial intelligence fields, such as machine learning and deep learning, in the field of medical trauma to develop efficient and dynamic trauma-related clinical predictive models to support pre-hospital or in-hospital emergency personnel's decisions and to assist clinicians in diagnosis, to improve medical services, and to save more lives.

Methods

Data sources

Two large databases were selected for this study: the Medical Information Mart for Intensive Care (MIMIC)-IV database (version 0.4) and Trauma database of General Hospital of the People's Liberation Army (PLA) (hereafter referred to as the Trauma database). MIMIC-IV is a database containing real, de-identified medical information obtained during the hospitalization of patients in the Beth Israel Deaconess Medical Center from 2008 to 2019. The MIMIC-IV database includes demographic information, vital signs, laboratory values, imaging reports, treatment records, and death records [11]. The Trauma database is a large, comprehensive database established by the Medical Big Data Research Center of the General Hospital of the PLA. It contains the de-identified data of patients from 2015 to 2020, including vital signs, laboratory values, nursing records, and treatment records.

Data extraction

The same data extraction standards and preprocessing methods were used for the MIMIC-IV and Trauma databases.

Extraction of the study population. The inclusion criteria of the study population were as follows: 1) patients aged ≥ 16 years admitted to the hospital due to trauma and 2) complete clinical data of at least one data record for any of the five vital signs, namely, heart rate (HR), respiratory rate (RR), systolic blood pressure (SBR), diastolic blood pressure (DBP), and peripheral oxygen saturation (SpO_2).

Patients in the experimental (that is, patients with TASH) and control groups (patients without TASH) were extracted from the study population. The extraction criteria were as follows [12–14]: 1) massive transfusion of three or more units of red blood cells within 1 h, anytime during the first 24 h after admission; 2) embolization or hemostatic surgery within 24 h after admission; and 3) death within 24 h after admission. If the patient met any of the above three conditions, it was considered that the patient met the outcome variable and was classified as the experimental group; otherwise, they were classified as the control group.

Extraction of vital signs data: 1) Study section time definition. For the experimental group, the study section time was defined as when the outcome variable was met for the first time. For the control group, the study section time was defined as the time the last vital sign was detected. 2) Data extraction of the five vital signs. First, the data in the three-time intervals of 1–13 h, 2–14 h, and 3–15 h before the study section time were extracted to construct models to predict the probability of severe hemorrhage 1 h/2 h/3 h after trauma (hereinafter referred to as the 1 h, 2 h, and 3 h groups). Second, no more than 50 recent data samples were extracted from each patient during this time interval. Finally, patients with complete deletion for any of the five recorded vital signs were removed after the extraction was completed.

Data preprocessing

We cleaned all vital signs data, including HR, RR, SBP, DBP, and SpO_2 , to circumvent any incorrect data entered by recording errors.

Filling in the missing data. For patient i , if data x_t^d , which is the vital sign d at the moment t , were missing, the last data before or after the moment t were selected to fill in the x_t^d .

Predictive algorithms and statistical methods

In this study, six machine learning and deep learning algorithms were used to develop and to validate the dynamic predictive models of TASH, which were: Logistic Regression (LR), Support Vector Machine (SVM) [15], Random Forests (RF) [16], Adaptive Boosting (AdaBoost) [17], Gated Recurrent Unit (GRU) [18], and Gated Recurrent Unit-D (GRU-D) [19]. The performance of each model was evaluated by accuracy, sensitivity, specificity, false positive rate (FPR), positive predictive value (PPV), negative predictive value (NPV), Youden index, and area under the curve (AUC).

In our study, big trauma data in the MIMIC-IV database were used to develop the predictive models. Ten-fold cross-validation was used to evaluate the model performance. The big trauma data in the Trauma database were used for external validation of each model. The models were developed and validated using Python 3.8.5 software. SPSS.21 software was used for statistical analysis. The Mann-Whitney U/ Wilcoxon rank-sum test was used for the comparison of two samples of quantitative data, and the Kruskal-Wallis H rank-sum test was used for the comparison of multiple groups of quantitative data. Chi-square test, continuous correction method, or Fisher exact probability test was used to compare the classified data. Differences were considered statistically significant at $p < 0.05$.

Results

Baseline characteristics

In this study, information on 13307 traumatic patients was extracted from the MIMIC-IV database. After screening according to the inclusion criteria, a total of 7522 patients were included, with a median age of 63 years (57.30% males). Among them, 283 met the outcome variable of severe hemorrhage, accounting for 3.76% of the total of the study (Figure 1a, Table 1). Information on a total of 25810 traumatic patients was initially extracted from the Trauma database. After screening according to the inclusion criteria, a total of 1686 patients were included, with a median age of 47 years (77.52% males). Among them, 306 met the outcome variable of severe hemorrhage, accounting for 18.15% of the total of the study (Figure 1b, Table 2).

Table 1
Baseline characteristics of the study population in the MIMIC-IV cohort.

	MIMIC-IV cohort	TASH group	Non-TASH group	p-value
N	7522	283	7239	
Male, n (%)	4310 (57.30)	165 (53.92)	4145 (57.26)	0.727
Age (years), median [Q1, Q3]	63.00 (48.00, 76.00)	63.00 (50.00, 77.50)	63.00 (48.00, 76.00)	0.368
BMI (kg/m ²), median [Q1, Q3]	27.05 (23.47, 31.88)	27.06 (23.55, 31.40)	27.05 (23.47, 31.89)	0.830
Admission Type, n (%)				0.001*
Elective	150 (1.99%)	4 (1.41%)	146 (2.02%)	
Emergency	4917 (65.37%)	210 (74.20%)	4707 (65.02%)	
Urgent	810 (10.77%)	34 (12.01%)	776 (10.72%)	
Other	1645 (21.87%)	35 (12.37%)	1610 (22.24%)	
Ethnicity, n (%)				<0.001*
Asian	184 (2.45%)	8 (2.83%)	176 (2.43%)	
Black	559 (7.43%)	16 (5.65%)	543 (7.50%)	
Hispanic	279 (3.71%)	11 (3.89%)	268 (3.71%)	
White	4981 (66.22%)	157 (55.48%)	4824 (66.64%)	
Other/Unknown	1519 (20.19%)	91 (32.16%)	1428 (19.73%)	
Comorbidities, n(%)				
Hypertension	2807 (37.32%)	103 (36.40%)	2704 (37.35%)	0.744
Diabetes	1492 (16.84%)	38 (13.42%)	1454 (20.09%)	0.006*
Pneumonia	1233 (16.39%)	41 (14.49%)	1192 (16.47%)	0.378
Congestive heart failure	500 (6.65%)	8 (2.83%)	492 (6.80%)	0.009*
Myocardial infarction	542 (7.21%)	17 (6.01%)	527 (7.25%)	0.417
Chronic liver disease	793 (10.54%)	46 (16.25%)	747 (10.32%)	0.001*
Chronic kidney disease	829 (11.02%)	19 (6.71%)	810 (11.19%)	0.018*
Vital signs, median [Q1, Q3]				
Heart rate (bpm)	87.00 (75.00, 100.00)	94.00 (81.00, 107.00)	87.00 (74.00, 100.00)	<0.001*
Respiratory rate (breaths/min)	19.00 (16.00, 24.00)	20.00 (16.00, 25.00)	19.00 (16.00, 24.00)	<0.001*
Systolic blood pressure (mm Hg)	120.00 (106.00, 136.00)	117.00 (104.00, 133.00)	120.00 (106.00, 136.00)	<0.001*
Diastolic blood pressure (mm Hg)	64.00 (55.00, 75.00)	62.00 (52.00, 72.00)	64.00 (55.00, 75.00)	<0.001*
Peripheral oxygen saturation (%)	98.00 (96.00, 100.00)	98.00 (96.00, 100.00)	98.00 (96.00, 100.00)	<0.001*
Fluid balance (1st 24 h), median [Q1,Q3]				
RBC transfusion (ml)	656.00 (375.00, 917.00)	1050.00 (700.00, 1868.00)	553.00 (350.00, 725.00)	<0.001*
FFP transfusion (ml)	598.00 (318.00, 840.00)	623.00 (508.88, 1218.25)	570.00 (316.00, 675.00)	<0.001*
PLT transfusion (ml)	300.00 (222.00, 550.50)	510.50 (286.50, 1059.75)	293.00 (206.00, 452.00)	<0.001*

P-value is the result of data comparison between the TASH group and non-TASH group. The Mann-Whitney U/ Wilcoxon rank-sum test was used for the comparison of two samples of quantitative data. Chi-square test, continuous correction method, or Fisher exact probability test was used to compare the classified data. * represents the difference is statistically significant. MIMIC-IV, Medical Information Mart for Intensive Care database-IV; TASH, trauma-associated severe hemorrhage; BMI, body mass index; RBC, red blood cell; FFP, fresh frozen plasma; PLT, platelet; SOFA, sequential organ failure assessment score; SAPS II, simplified acute physiology score II; OASIS, Oxford acute severity of illness score; Hosp., hospital; LOS, length of stay; ICU, intensive care unit.

	MIMIC-IV cohort	TASH group	Non-TASH group	p-value
Urine output (ml)	1320.00	1170.50	1325.00	0.014*
	(790.00, 2045.00)	(475.50, 2122.75)	(800.00, 2043.50)	
Drug (1st 24 h), n (%)				
Epinephrine	848 (11.27%)	106 (37.46%)	742 (10.25%)	<0.001*
Norepinephrine	821 (10.91%)	114 (40.28%)	707 (9.77%)	<0.001*
Dopamine	57 (0.76%)	15 (5.30%)	42 (0.58%)	<0.001*
Dobutamine	25 (0.33%)	1 (0.35%)	24 (0.33%)	0.617
Tranexamic acid use	11 (0.15%)	9 (3.18%)	2 (0.03%)	<0.001*
Severity score, median [Q1, Q3]				
SOFA	4 (2, 7)	7 (4, 10)	4 (2, 7)	<0.001*
SAPS II	35 (26, 44)	37 (31, 47)	35 (26, 44)	<0.001*
OASIS	32 (26, 39)	36 (29, 42)	32 (26, 39)	<0.001*
Primary outcomes (1st 24 h), n (%)				
Massive transfusion	170 (2.26)	170 (60.07)		
Embolization or hemostatic surgery	15 (0.20)	15 (5.30)		
Death	106 (1.41)	106 (37.46)		
Total	283 (3.76)	283 (100.00)		
Secondary outcomes				
Hosp. LOS (d), median [Q1, Q3]	7.68 (4.34, 14.21)	5.53 (0.64, 12.93)	7.73 (4.46, 14.22)	<0.001*
ICU LOS (d), median [Q1, Q3]	2.23 (1.12, 5.80)	1.73 (0.72, 7.05)	2.26 (1.14, 5.78)	0.002*
In-hospital mortality, n (%)	753 (10.01%)	135 (47.70%)	618 (8.54%)	<0.001*
ICU mortality, n (%)	572 (7.50%)	124 (43.82%)	448 (6.19%)	<0.001*
P-value is the result of data comparison between the TASH group and non-TASH group. The Mann-Whitney U/ Wilcoxon rank-sum test was used for the comparison of two samples of quantitative data. Chi-square test, continuous correction method, or Fisher exact probability test was used to compare the classified data. * represents the difference is statistically significant. MIMIC-IV, Medical Information Mart for Intensive Care database-IV; TASH, trauma-associated severe hemorrhage; BMI, body mass index; RBC, red blood cell; FFP, fresh frozen plasma; PLT, platelet; SOFA, sequential organ failure assessment score; SAPS II, simplified acute physiology score II; OASIS, Oxford acute severity of illness score; Hosp., hospital; LOS, length of stay; ICU, intensive care unit.				

Table 2
Baseline characteristics of the study population in the Trauma cohort.

	Trauma cohort	TASH group	Non-TASH group	p-value
N	1686	306	1380	
Male, n (%)	1307 (77.52)	240 (78.43)	1067 (77.32)	0.673
Age (years), median [Q1, Q3]	47.00 (32.00, 60.00)	43.00 (30.00, 54.00)	48.00 (33.00, 61.00)	<0.001*
Vital signs, median [Q1, Q3]				
Heart rate (bpm)	90.00 (77.00, 104.00)	98.00 (82.00, 115.00)	89.00 (76.00, 103.00)	<0.001*
Respiratory rate (breaths/min)	20.00 (19.00, 20.00)	20.00 (18.00, 21.00)	20.00 (19.00, 20.00)	<0.001*
Systolic blood pressure (mm Hg)	121.00 (109.00, 135.00)	114.00 (102.00, 127.00)	123.00 (110.00, 137.00)	<0.001*
Diastolic blood pressure (mm Hg)	74.00 (65.00, 83.00)	70.00 (61.00, 79.00)	75.00 (66.00, 83.00)	<0.001*
Peripheral oxygen saturation (%)	98.00 (97.00, 99.00)	98.00 (97.00, 99.00)	98.00 (97.00, 99.00)	0.002*
Fluid balance (1st 24 h), median [Q1, Q3]				
RBC transfusion (u)	4.00 (2.50, 4.00)	4.00 (4.00, 5.25)	2.00 (2.00, 2.00)	<0.001*
FFP transfusion (u)	3.75 (2.40, 4.50)	4.00 (2.50, 4.50)	2.20 (2.08, 2.40)	<0.001*
Drug (1st 24 h), n (%)				
Adrenoceptor agonists	70 (4.15%)	34 (11.11%)	36 (2.6%)	<0.001*
Dopamine	86 (5.10%)	46 (15.03%)	40 (2.90%)	<0.001*
Hemostatic drugs	1064 (63.11%)	240 (78.43%)	824 (59.71%)	<0.001*
Primary outcomes (1st 24 h), n(%)				
Massive transfusion	149 (8.84)	149 (48.69)		
Embolization or hemostatic surgery	180 (10.68)	180 (58.82)		
Death	20 (1.19)	20 (6.54)		
Total	306 (18.15)	306 (100.00)		
Secondary outcomes, n (%)				
In-hospital mortality	34 (2.02%)	22 (7.19%)	12 (0.87%)	<0.001*
P-value is the result of data comparison between the TASH group and non-TASH group. The Mann-Whitney U/ Wilcoxon rank-sum test was used for the comparison of two samples of quantitative data. Chi-square test, continuous correction method, or Fisher exact probability test was used to compare the classified data. * represents the difference is statistically significant. TASH, trauma-associated severe hemorrhage; RBC, red blood cell; FFP, fresh frozen plasma.				

No significant difference was found in sex, age, and body mass index between the experimental (TASH group) and control groups (non-TASH group) in the MIMIC-IV cohort (Table 1). However, the TASH group had a higher proportion of emergency hospitalization, higher HR and RR, lower blood pressure, and more red blood cell, platelet, and plasma transfusions. Patients in the TASH group received an increased infusion of vasoactive drugs, positive inotropic drugs, and hemostatic drugs, and their sequential organ failure assessment (SOFA), simplified acute physiology (SAPS II), and Oxford acute severity of illness (OASIS) scores were higher; hospital stay and hospital mortality were also increased. In the Trauma cohort (Table 2), patients in the TASH and non-TASH groups had a similar pattern.

Development of TASH dynamic predictive models based on the MIMIC-IV cohort

As shown in Table 3, compared with the other five models, the GRU-D model obtained the best results with respect to accuracy (0.885±0.042), sensitivity (0.894±0.080), specificity (0.883±0.043), FPR (0.117±0.043), PPV (0.573±0.100), NPV (0.980±0.015), Youden index (0.777±0.101), and AUC (0.943±0.034) in the 3 h group. It also had the best performance in the 1 h and 2 h groups. In the 3 h group, the ROC curves of the predictive models based on the six algorithms were compared and the AUCs' differences were analyzed. There were statistical differences between the GRU-D model and other five models, and the same results were obtained in the 1 h and 2 h groups (Figure 2a, 2b, 2c, Figure 3a, 3b, 3c). Moreover, the AUC of the GRU-D model showed no statistical difference among the three groups.

Table 3
Comparison of the effects of TASH predictive models based on the MIMIC-IV cohort.

		Accuracy	Sensitivity	Specificity	FPR	PPV	NPV	Youden	AUC
1 h group	LR	0.686±0.038	0.644±0.099	0.693±0.042	0.307±0.042	0.261±0.039	0.921±0.021	0.337±0.105	0.722±0.060
	SVM	0.786±0.035	0.683±0.098	0.803±0.039	0.197±0.039	0.370±0.057	0.939±0.018	0.486±0.101	0.816±0.044
	RF	0.792±0.034	0.655±0.117	0.815±0.038	0.185±0.038	0.374±0.061	0.935±0.021	0.470±0.117	0.804±0.055
	AdaBoost	0.794±0.032	0.680±0.112	0.813±0.036	0.187±0.036	0.380±0.057	0.939±0.020	0.493±0.111	0.811±0.052
	GRU	0.872±0.033	0.825±0.084	0.880±0.034	0.121±0.034	0.540±0.078	0.968±0.015	0.704±0.096	0.900±0.046
	GRU-D	0.892±0.040	0.900±0.069	0.891±0.043	0.109±0.043	0.592±0.101	0.982±0.013	0.790±0.089	0.946±0.029
2 h group	LR	0.604±0.045	0.735±0.107	0.582±0.048	0.418 ±0.048	0.228±0.033	0.930±0.027	0.317±0.119	0.703±0.073
	SVM	0.791±0.035	0.683±0.109	0.809±0.042	0.191±0.042	0.377±0.055	0.939±0.019	0.492±0.106	0.803±0.053
	RF	0.805±0.038	0.658±0.125	0.830±0.039	0.170±0.039	0.396±0.071	0.936±0.022	0.488±0.130	0.802±0.067
	AdaBoost	0.784±0.038	0.711±0.106	0.796±0.043	0.204±0.043	0.372±0.061	0.943±0.020	0.507±0.109	0.809±0.061
	GRU	0.874±0.042	0.844±0.087	0.879±0.045	0.121±0.045	0.551±0.099	0.971±0.016	0.723±0.105	0.903±0.046
	GRU-D	0.892±0.037	0.889±0.077	0.892±0.037	0.108±0.037	0.590±0.096	0.980±0.014	0.782±0.097	0.940±0.032
3 h group	LR	0.677±0.043	0.692±0.117	0.675±0.048	0.325±0.048	0.263±0.042	0.930±0.025	0.366±0.121	0.729±0.075
	SVM	0.747±0.045	0.673±0.127	0.759±0.053	0.241±0.053	0.321±0.059	0.934±0.024	0.432±0.127	0.788±0.067
	RF	0.781±0.036	0.621±0.119	0.808±0.043	0.192±0.043	0.354±0.060	0.928±0.021	0.429±0.114	0.796±0.063
	AdaBoost	0.777±0.042	0.630±0.134	0.802±0.046	0.198±0.046	0.350±0.073	0.929±0.024	0.432±0.137	0.787±0.067
	GRU	0.873±0.043	0.817±0.101	0.882±0.043	0.118±0.043	0.549±0.104	0.967±0.018	0.699±0.120	0.898±0.053
	GRU-D	0.885±0.042	0.894±0.080	0.883±0.043	0.117±0.043	0.573±0.100	0.980±0.015	0.777±0.101	0.943±0.034
Logistic Regression, (LR); Support Vector Machine, (SVM); Random Forests, (RF); Adaptive Boosting, (AdaBoost); Gated Recurrent Unit, (GRU); Gated Recurrent Unit-D, (GRU-D); false positive rate, (FPR); positive predictive value, (PPV); negative predictive value, (NPV); area under the curve, (AUC).									

External validation of TASH dynamic predictive models based on the Trauma cohort

As shown in Table 4, in the Trauma cohort, the GRU-D model was superior to other models in sensitivity, NPV, Youden index, and AUC, while the RF model performed better in terms of accuracy, specificity, FPR, and PPV. In this study, the purpose was to identify patients with TASH as much as possible; thus, we paid more attention to sensitivity than accuracy. Besides, the AUC is the best index to measure the comprehensive performance of the model. Overall, the GRU-D model is still the best model in the Trauma cohort. Thus, it can be observed that the GRU-D model has better generalization ability than other models. The ROC curve comparison and AUC difference analysis of the six models showed statistical differences between the GRU-D model and other five models (Figure 3a, 3b, 3c, Figure 4a, 4b, 4c). Furthermore, there was no statistical difference in the AUC of the GRU-D model among the three groups.

Table 4
Comparison of the effects of TASH predictive models based on the Trauma cohort.

		Accuracy	Sensitivity	Specificity	FPR	PPV	NPV	Youden	AUC
1 h group	LR	0.678±0.008	0.621±0.007	0.692±0.011	0.309±0.011	0.322±0.006	0.886±0.001	0.312±0.007	0.701±0.002
	SVM	0.723±0.009	0.567±0.010	0.760±0.013	0.240±0.013	0.358±0.010	0.882±0.002	0.327±0.009	0.698±0.002
	RF	0.744±0.017	0.493±0.058	0.803±0.032	0.197±0.032	0.372±0.020	0.871±0.010	0.296±0.036	0.690±0.004
	AdaBoost	0.726±0.022	0.513±0.041	0.776±0.035	0.224±0.035	0.353±0.022	0.872±0.006	0.289±0.020	0.684±0.004
	GRU	0.561±0.032	0.824±0.034	0.499±0.047	0.501±0.047	0.280±0.012	0.924±0.008	0.323±0.022	0.744±0.011
	GRU-D	0.587±0.043	0.844±0.054	0.526±0.065	0.474±0.065	0.298±0.021	0.936±0.012	0.370±0.021	0.779±0.013
2 h group	LR	0.603±0.007	0.680±0.009	0.586±0.009	0.415±0.009	0.274±0.004	0.888±0.003	0.265±0.010	0.687±0.003
	SVM	0.726±0.007	0.558±0.010	0.765±0.011	0.235±0.011	0.353±0.008	0.883±0.002	0.323±0.007	0.690±0.003
	RF	0.757±0.006	0.470±0.024	0.823±0.012	0.177±0.012	0.380±0.009	0.871±0.004	0.293±0.016	0.664±0.006
	AdaBoost	0.726±0.007	0.525±0.011	0.772±0.010	0.228±0.010	0.347±0.009	0.876±0.002	0.298±0.010	0.679±0.005
	GRU	0.559±0.033	0.835±0.032	0.496±0.048	0.504±0.048	0.277±0.013	0.929±0.008	0.330±0.024	0.741±0.011
	GRU-D	0.582±0.039	0.871±0.041	0.516±0.057	0.485±0.057	0.294±0.016	0.947±0.011	0.387±0.020	0.780±0.008
3 h group	LR	0.636±0.010	0.606±0.012	0.643±0.013	0.357±0.013	0.261±0.005	0.887±0.002	0.249±0.009	0.684±0.004
	SVM	0.711±0.008	0.563±0.014	0.741±0.012	0.259±0.012	0.313±0.006	0.891±0.002	0.305±0.007	0.699±0.002
	RF	0.767±0.010	0.461±0.017	0.831±0.014	0.169±0.014	0.364±0.016	0.881±0.003	0.292±0.014	0.691±0.004
	AdaBoost	0.744±0.019	0.491±0.025	0.797±0.028	0.203±0.028	0.337±0.021	0.882±0.003	0.288±0.015	0.687±0.005
	GRU	0.577±0.041	0.818±0.060	0.527±0.061	0.473±0.061	0.267±0.017	0.934±0.011	0.345±0.022	0.741±0.011
	GRU-D	0.579±0.048	0.874±0.058	0.517±0.069	0.483±0.069	0.277±0.023	0.953±0.013	0.391±0.020	0.778±0.009
Logistic Regression, (LR); Support Vector Machine, (SVM); Random Forests, (RF); Adaptive Boosting, (AdaBoost); Gated Recurrent Unit, (GRU); Gated Recurrent Unit-D, (GRU-D); false positive rate, (FPR); positive predictive value, (PPV); negative predictive value, (NPV); area under the curve, (AUC).									

Comparison between the GRU-D dynamic predictive models and traditional scoring systems

Figures 5a, 5b, and 5c compare the GRU-D dynamic predictive model with shock index, Larson score, and Vandromme score in the three groups. As shown in Figure 5, the GRU-D model is significantly superior to the traditional severe hemorrhage predictive scores, with the highest AUC. Figures 6a, 6b, and 6c compare the GRU-D dynamic predictive model with the OASIS, SAPS II, and SOFA scores in the three groups. As shown in the Figure 6, the GRU-D model is significantly better than the previous severity scores and has the highest AUC.

Development of TASH predictive system

To facilitate doctors to use our predictive models, we have developed a website tool; the website is <http://82.156.217.249:5000/>. The predictive system integrated the GRU-D models with the best predictive effect, including the data input page and predictive result display page (as shown in Figure 7a and 7b). On the data input page, the input data template can be downloaded. After inputting or importing vital signs time series data according to the template format, you can click the "Submit" button to jump to the predictive result display page. This page shows the predictive probability of TASH 1 h/2 h/3 h after trauma and shows the input predictive variable data.

Discussion

In this study, we developed and validated three deep learning models to dynamically predict the probability of severe hemorrhage occurring at three points in time following trauma, based on vital signs data of trauma patients from a large-scale public database. It was further validated in the Trauma database of the University Teaching Hospital. Moreover, we provide an open and accessible data interface for the public to use and to validate our model. Our predictive models can help pre-hospital or in-hospital clinicians in the early identification, dynamic prediction, and decision making regarding patients with severe hemorrhage from trauma, thus saving more lives.

There are already some scoring systems for TASH. For example, the TASH score [8], PWH score [10], traumatic bleeding severity score (TBSS) [20], and modified TBSS [21] require clinical assessment, laboratory values, and ultrasound assessment. Scores such as the Hsu [22], Larson [23], McLaughlin [24], and Vandromme scores [25] require clinical assessment and laboratory values. The ABC score [9] requires clinical and ultrasound assessment. The above scoring systems, which require results of laboratory values or ultrasound assessment, are more complex and often require patients to arrive at the hospital to calculate the score results; thus, it is time-consuming for in-hospital evaluation and not suitable for pre-hospital evaluation [26].

Furthermore, most of these scoring systems are static evaluations using a single measurement, and the long detection intervals and invasive operations of laboratory or ultrasound examinations make dynamic monitoring difficult to achieve.

The TASH dynamic predictive models developed in our study only depend on vital signs, which can be easily obtained in pre-hospital or in-hospital environments, and medical staff can easily record the data regularly. Simple feature selection also ensures that the predictive models can be continuously and automatically recalculated before or during hospitalization, providing valuable information on whether patients are responding to treatment, thus making it easier for medical professionals to modify their treatment plans. In addition, the simplicity of the input and output improves the interpretability of the predictive models, thus increasing the possibility that health care providers trust their predictions [27–32].

Comparing the evaluation indexes of each model based on the MIMIC-IV database, in general, the GRU-D model is better than the GRU model; the GRU model is better than the AdaBoost, RF, and SVM models; there is no obvious difference among the AdaBoost, RF, and SVM models; and these three models are better than the LR model. The reason for the difference between the above six models may be that the LR, SVM, RF, and AdaBoost models are traditional machine learning algorithms, and the input of the models is a five-dimensional vector. The GRU and GRU-D models belong to the deep learning algorithm, and the input data is a time series of five-dimensional vectors comprising five vital signs. Moreover, the GRU model is a variant of the traditional recurrent neural network, which solves the problem of gradient disappearance. The GRU-D model is a variant model based on GRU proposed by Che et al. in 2018, which can deal with irregular sampling time series data with missing values. Its input includes the time series data, the mask, and time interval information. Then, in the process of training, it processes the time interval information between the two recorded data before and after, captures the relationship between the time series data, fills in the missing data, and makes predictions at the same time. In GRU-D, data filling and prediction of results are both conducted in the process of neural network training; thus, the parameters related to data filling will be continuously optimized in the process of training and make the predictive result better [19].

The GRU-D model, which has the best performance in our study, is compared with the traditional scoring systems. The shock index has been recommended to predict massive blood transfusion and emergency operation after trauma and is widely used in pre-hospital and battlefield environments [33, 34]. The Vandromme score was put forward by Vandromme and his colleagues in 2012, which was used to identify patients with massive blood transfusion risk [25]. The Larson's score was put forward by Larson and his colleagues in 2010 based on a combat database, which was used to predict the massive blood transfusion needs of combat casualties [23]. The OASIS, SAPS II, and SOFA scores are commonly used severe illness scoring systems at present, which are often used to predict the severity of patients' illness or hospitalization mortality. Based on the MIMIC-IV cohort, our study compares the GRU-D model with the above scoring systems. The GRU-D model has the highest AUC, which reflects the advantages of the GRU-D model in the dynamic prediction of TASH.

The results of our study were compared with those of other studies of the same type. For example, Brockamp et al. used the TraumaRegister DGU database to compare six scoring systems and algorithms for predicting persistent hemorrhage and blood transfusion needs after trauma, namely the TASH, PWH, Vandromme, Larson, Schreiber, and ABC scores. The corresponding AUCs were 0.889, 0.860, 0.840, 0.823, 0.800, 0.763, respectively. The performance of the TASH score in that study was better than those of the other scoring systems [35]. Compared with our study, the performance of the above TASH score was not as good as that of the GRU-D model in the MIMIC-IV cohort. In addition, Mitra et al. used data from the Alfred Trauma Registry to compare three scoring systems for predicting massive blood transfusion after trauma: the TASH score (AUC 0.899), PWH score (AUC 0.842), and ABC score (AUC 0.782). The performance of the TASH score was better than that of the other two scoring systems [36]. Compared with our study, the performance of the TASH score was not as good as that of the GRU-D model in the MIMIC-IV cohort. By comparing the AUC of predictive models among different studies, the GRU-D model based on vital signs still has a good predictive effect, which confirms the advantage of the GRU-D model in predicting TASH.

In addition to internal validation, we validate the model externally based on the Trauma database of the General Hospital of the PLA. As shown in Figures 4a, 4b, and 4c, the AUC of the GRU-D model is larger than that of the other models, indicating that our models have significant generalization ability and clinical value. To help clinicians use our models, we have developed a web-based predictive system, which provides a user-friendly interface. After entering the variables, the probability of severe hemorrhage occurring at three time points after trauma is shown. These results will help clinical decision-makers understand the condition of patients and prepare appropriate treatment strategies.

This study had certain limitations. First, the study population in this study only included adult patients, and further study population division based on age was not considered. However, age plays an important role in predicting the risk of severe hemorrhage. The age of the experimental group was significantly younger than that of the control group in the Trauma dataset in our study. Some studies have shown that elderly patients are more likely to have severe hemorrhage [20]. In future studies, we will divide the patients into different subgroups based on age for further discussion [37]. Second, the severe hemorrhage predictive models can only guide the doctor's clinical decision-making process and cannot replace the doctor's clinical judgment and other diagnostic tests. Finally, this is a retrospective observational study. Although the quality of the MIMIC-IV and Trauma databases is very high, there are still data losses and input errors. Therefore, prospective validation is still needed in the future. In future studies, it is also necessary to determine whether the use of dynamic predictive models for TASH reduces the waiting time before massive blood transfusion or damage control surgery and its impact on the prognosis of trauma patients.

Conclusions

This study developed and validated the dynamic predictive models of TASH based on vital signs and the GRU-D algorithm, which can dynamically predict whether severe hemorrhage will occur in trauma patients in the next 1 h/2 h/3 h and assist pre-hospital or in-hospital emergency personnel to make decisions, and our models are superior to other predictive methods.

Abbreviations

Medical Information Mart for Intensive Care database, (MIMIC)-IV; Trauma database of General Hospital of the People's Liberation Army, (PLA); institutional review board, (IRB); trauma-associated severe hemorrhage, (TASH); assessment of Blood consumption score, (ABC) score; Prince of Wales score (PWH); traumatic bleeding severity score (TBSS); sequential organ failure assessment score (SOFA); simplified acute physiology score (SAPS II); Oxford acute severity of illness score (OASIS); heart rate (HR), respiratory rate (RR), systolic blood pressure (SBR), diastolic blood pressure (DBP), and peripheral oxygen saturation (SpO₂); Logistic Regression, (LR); Support Vector Machine, (SVM); Random Forests, (RF); Adaptive Boosting, (AdaBoost); Gated Recurrent Unit, (GRU); Gated Recurrent Unit-D, (GRU-D); false positive rate, (FPR); positive predictive value, (PPV); negative predictive value, (NPV); area under the curve, (AUC)

Declarations

Ethics approval and consent to participate

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki. The MIMIC-IV database was previously de-identified and reviewed by the institutional review board (IRB) of its host organization and determined to be exempted from subsequent IRB. We obtained the administrative permissions to use the database (Certification Number: 27959316) for our research, after completing the National Institutes of Health web-based training course: Protecting Human Research Participants. We were reviewed and approved by the Ethics Committee of Chinese PLA General Hospital to use the Trauma database. The ethical batch number is S2021-466-01. Moreover, the informed consent of subjects was waived by the Ethics Committee of Chinese PLA General Hospital.

Consent for publication

Not applicable.

Availability of data and materials

The MIMIC-IV cohort are available in the MIMIC database, <https://mimic.mit.edu/>. The Trauma cohort are available from the Trauma database of the General Hospital of the PLA, although 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 the General Hospital of the PLA.

Competing interests

The authors have no conflicts of interest to declare.

Funding

This work was supported by The National Key Research and Development Program of China [2019YFF0302300].

Authors contributions

TL provided research design ideas and data extraction access to the Trauma database. CL provided administrative support. CG provided research design ideas, was responsible for article writing and modification, and participated in data extraction and data analysis. MG and QS were responsible for model construction and participated in data extraction. RW, HZ, and JW participated in data extraction and data analysis. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Editage (www.editage.cn) for English language editing.

References

1. Sauaia A, Moore FA, Moore EE, Moser KS, Brennan R, Read RA, et al. Epidemiology of trauma deaths: a reassessment. *J Trauma*. 1995;38:185–93.
2. Zhu CS, Cobb D, Jonas RB, Pokorny D, Rani M, Cotner-Pouncy T, et al. Shock index and pulse pressure as triggers for massive transfusion. *J Trauma Acute Care Surg*. 2019;87:S159–64.
3. Stephens CT, Gumbert S, Holcomb JB. Trauma-associated bleeding: management of massive transfusion. *Curr Opin Anaesthesiol*. 2016;29:250–5.

4. Kauvar DS, Lefering R, Wade CE. Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *J Trauma*. 2006;60;Suppl:S3–11.
5. Parimi N, Hu PF, Mackenzie CF, Yang S, Bartlett ST, Scalea TM, et al. Automated continuous vital signs predict use of uncrossed matched blood and massive transfusion following trauma. *J Trauma Acute Care Surg*. 2016;80:897–906.
6. Henriksen HH, Rahbar E, Baer LA, Holcomb JB, Cotton BA, Steinmetz J, et al. Pre-hospital transfusion of plasma in hemorrhaging trauma patients independently improves hemostatic competence and acidosis. *Scand J Trauma Resusc Emerg Med*. 2016;24:145.
7. Pommerening MJ, Goodman MD, Holcomb JB, Wade CE, Fox EE, Del Junco DJ, et al. Clinical gestalt and the prediction of massive transfusion after trauma. *Injury*. 2015;46:807–13.
8. Yücel N, Lefering R, Maegele M, Vorweg M, Tjardes T, Ruchholtz S, et al. Trauma associated severe hemorrhage (TASH)-score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma. *J Trauma*. 2006;60:1228–36; discussion 1236.
9. Nunez TC, Voskresensky IV, Dossett LA, Shinall R, Dutton WD, Cotton BA. Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)? *J Trauma*. 2009;66:346–52.
10. Rainer TH, Ho AM, Yeung JH, Cheung NK, Wong RS, Tang N, et al. Early risk stratification of patients with major trauma requiring massive blood transfusion. *Resuscitation*. 2011;82:724–9.
11. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Roger M. MIMIC-IV (version 0.4). PhysioNet 2020.
12. Cattle PM, Cotton BA. Prediction of massive transfusion in trauma. *Crit Care Clin*. 2017;33:71–84.
13. Wilson BR, Bruno J, Duckwitz M, Akers N, Jeanmonod D, Jeanmonod R. Prehospital end-tidal CO₂ as an early marker for transfusion requirement in trauma patients. *Am J Emerg Med*. 2021;45:254–7.
14. Tran A, Matar M, Steyerberg EW, Lampron J, Taljaard M, Vaillancourt C. Early identification of patients requiring massive transfusion, embolization, or hemostatic surgery for traumatic hemorrhage: a systematic review protocol. *Syst Rev*. 2017;6:80.
15. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24:1565–7.
16. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
17. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55:119–39.
18. Cho K, Merrienboer BV, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Comput Sci*. 2014.
19. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values [Sci. rep.:6085]. *Sci Rep*. 2018;8:6085.
20. Ogura T, Nakamura Y, Nakano M, Izawa Y, Nakamura M, Fujizuka K, et al. Predicting the need for massive transfusion in trauma patients: the Traumatic Bleeding Severity Score. *J Trauma Acute Care Surg*. 2014;76:1243–50.
21. Ogura T, Lefor AK, Masuda M, Kushimoto S. Modified traumatic bleeding severity score: early determination of the need for massive transfusion. *Am J Emerg Med*. 2016;34:1097–101.
22. Hsu JM, Hitos K, Fletcher JP. Identifying the bleeding trauma patient: predictive factors for massive transfusion in an Australasian trauma population. *J Trauma Acute Care Surg*. 2013;75:359–64.
23. Larson CR, White CE, Spinella PC, Jones JA, Holcomb JB, Blackburn LH, et al. Association of shock, coagulopathy, and initial vital signs with massive transfusion in combat casualties. *J Trauma*. 2010;69;Suppl 1:S26–32.
24. McLaughlin DF, Niles SE, Salinas J, Perkins JG, Cox ED, Wade CE, et al. A predictive model for massive transfusion in combat casualty patients. *J Trauma*. 2008;64;Suppl:S57–63; discussion S63.
25. Vandromme MJ, Griffin RL, McGwin G, Weinberg JA, Rue LW, Kerby JD. Prospective identification of patients at risk for massive transfusion: an imprecise endeavor. *Am Surg*. 2011;77:155–61.
26. Shih AW, Al Khan S, Wang AY, Dawe P, Young PY, Greene A, et al. Systematic reviews of scores and predictors to trigger activation of massive transfusion protocols. *J Trauma Acute Care Surg*. 2019;87:717–29.
27. Orphanidou C. A review of big data applications of physiological signal data. *Biophys Rev*. 2019;11:83–7.
28. Convertino VA, Moulton SL, Grudic GZ, Rickards CA, Hinojosa-Laborde C, Gerhardt RT, et al. Use of advanced machine-learning techniques for noninvasive monitoring of hemorrhage. *J Trauma*. 2011;71;Suppl:S25–32.
29. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*. 2016;102:1–5.
30. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med*. 2019;109:79–84.
31. Baker S, Xiang W, Atkinson I. Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach [Sci. rep.:21282]. *Sci Rep*. 2020;10:21282.
32. Liu J, Khitrov MY, Gates JD, Odom SR, Havens JM, de Moya MA, et al. Automated analysis of vital signs to identify patients with substantial bleeding before hospital arrival: a feasibility study. *Shock*. 2015;43:429–36.
33. Marengo CW, Lammers DT, Morte KR, Bingham JR, Martin MJ, Eckert MJ. Shock index as a predictor of massive transfusion and emergency surgery on the modern battlefield. *J Surg Res*. 2020;256:112–8.

34. Sharma A, Naga Satish U, Tevatia MS, Singh SK. Prehospital shock index, modified shock index, and pulse pressure heart rate ratio as predictors of massive blood transfusions in modern warfare injuries: a retrospective analysis. *Med J Armed Forces India*. 2019;75:171–5.
35. Brockamp T, Nienaber U, Mutschler M, Wafaisade A, Peiniger S, Lefering R, et al. Predicting on-going hemorrhage and transfusion requirement after severe trauma: a validation of six scoring systems and algorithms on the TraumaRegister DGU. *Crit Care*. 2012;16:R129.
36. Mitra B, Rainer TH, Cameron PA. Predicting massive blood transfusion using clinical scores post-trauma. *Vox Sang*. 2012;102:324–30.
37. Sammy I, Lecky F, Sutton A, Leaviss J, O’Cathain A. Factors affecting mortality in older trauma patients-a systematic review and meta-analysis. *Injury*. 2016;47:1170–83.

Figures

Figure 1

a: Flow chart for patients’ enrollment and study design in the MIMIC-IV database. b: Flow chart for patients’ enrollment and study design in the Trauma database.

Figure 2

ROC curve comparison of TASH predictive models based on the MIMIC-IV cohort. a: 1 h group, b: 2 h group, c: 3 h group.

Figure 3

Comparison of model effects between the MIMIC-IV and Trauma cohorts. a: 1 h group, b: 2 h group, c: 3 h group.

Figure 4

ROC curve comparison of TASH predictive models based on the Trauma cohort. a: 1 h group, b: 2 h group, c: 3 h group.

Figure 5

Comparison of the GRU-D model with the SI, Larson score, and Vandromme score.

Figure 6

Comparison of the GRU-D model with the SAPS II score, OASIS score, and SOFA score.

Predictive Models for Trauma-associated Severe Hemorrhage

Please download the template: [click to download](#)

Please upload vital signs data (xlsx format file)

Select file Submit

a

Predictive Models for Trauma-associated Severe Hemorrhage

The probability of severe hemorrhage 1 h after trauma: **0.87**

The probability of severe hemorrhage 2 h after trauma: **0.93**

The probability of severe hemorrhage 3 h after trauma: **0.98**

Record Time	Heart Rate	Respiratory Rate	Systolic Blood Pressure	Diastolic Blood Pressure	SpO ₂
2146-10-02 16:45:00	82.00	17.00	148.00	86.00	100.00
2146-10-02 16:53:00	75.00	15.00	-	-	100.00
2146-10-02 17:00:00	75.00	14.00	135.00	44.00	100.00
2146-10-02 18:00:00	75.00	14.00	133.00	49.00	100.00
2146-10-02 19:00:00	73.00	14.00	116.00	54.00	100.00
2146-10-02 20:00:00	68.00	13.00	100.00	-	100.00
2146-10-02 21:00:00	87.00	16.00	94.00	-	80.00
2146-10-02 22:00:00	79.00	16.00	92.00	-	88.00
2146-10-02 23:00:00	71.00	18.00	91.00	-	91.00
2146-10-03 00:00:00	69.00	20.00	95.00	-	93.00
2146-10-03 01:00:00	69.00	21.00	93.00	-	91.00
2146-10-03 02:00:00	70.00	22.00	92.00	-	91.00
2146-10-03 03:00:00	71.00	21.00	78.00	-	91.00

b

Figure 7

Predictive system display based on GRU-D dynamic predictive models; a: Data input page, b: Predictive result display page.