Intelligent Prediction of RBC Demand in Trauma Patients

Yannan Feng

Chinese PLA General Hospital <https://orcid.org/0000-0002-0596-4007>

Zhenhua Xu

Beijing Hexing Chuanglian Health Technology, Co., Ltd, China

Junting Liu

Chinese PLA General Hospital

Xiaolin Sun

Chinese PLA General Hospital

Deqing Wang

Chinese PLA General Hospital

Yang Yu (✉ yuyangpla301@163.com)

<https://orcid.org/0000-0002-9390-1080>

Research

Keywords: Mathematic model, Intelligent prediction, Trauma transfusion

Posted Date: July 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-41246/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Military Medical Research on May 24th, 2021. See the published version at <https://doi.org/10.1186/s40779-021-00326-3>.

Abstract

Background: The vital signs of trauma patients are complex and changeable, and the prediction of blood transfusion demand mainly depends on doctors' experience and trauma scoring system, therefore it can't be accurately predicted. In this study, the decision tree algorithm (Classification and Regression Tree, CRT and eXtreme Gradient Boosting, XGboost) of machine learning were proposed to the demand prediction of traumatic blood transfusion, hoping to provide technical support for doctors.

Methods: Total 1371 trauma patients who were diverted to the emergency department from 2014.1 to 2018.1 were collected from the emergency trauma database. The vital signs, laboratory examination parameters and blood transfusion volume were used as variables, and the non-invasive parameters and all (non-invasive + invasive) parameters were used to construct the intelligent prediction model of RBC demand by logistic regression (LR), CRT and Xgboost. The prediction accuracy of the model was compared with the Area Under Curve (AUC).

Results: The studies we have performed showed that non-invasive parameters were used to predict blood transfusion, LR method was the best, AUC 0.72 (95% confidence interval [CI] 0.657-0.775), highest than CRT AUC 0.69 (95%CI 0.633-0.751) and Xgboost AUC 0.71 (95%CI 0.654-0.756) ($P<0.05$). Trauma site and shock index are important prediction parameters. All the parameters to predict, Xgboost was the best with AUC 0.94 (95%CI 0.893-0.981), which was highest than LR AUC 0.80 (95%CI 0.744-0.850) and CRT AUC 0.82 (95%CI 0.779-0.853) ($P<0.05$). Hematocrit/Hemoglobin are important prediction parameters.

Conclusions: The prediction model of red blood cell transfusion in trauma patients constructed by decision tree algorithm can be used as a technical support to assist doctors to make rapid and accurate blood transfusion decisions in emergency rescue environment, so as to improve the success rate of patient treatment.

Background

Trauma accounts for about 9% of global deaths[1], and deaths mainly occur within the first 12 hours after trauma[2]. The first step in trauma treatment is to control the bleeding as soon as possible, advance to the start of the incident, and directly transfer the patient to a nearby trauma treatment institution[3]. Post-traumatic blood loss is a potential and preventable leading cause of death[4]. The core principle of treatment is to identify the risk of hemorrhagic shock as early as possible, meanwhile fluid resuscitation and blood transfusion are needed to maintain the stability of basic vital signs and hemodynamic[5]. The study found that blood transfusion products prehospital within 15 minutes or 15 minutes after injury was associated with 24-hour mortality (5.6% VS. 20.2%) and 30-day mortality (11.8% VS. 22.9%) compared with delayed or non-transfusion[6]. Delayed blood transfusion can lead to pulmonary complications and death[7]. Several studies have found that RBC transfusion in trauma patients is associated with increased morbidity and mortality[8, 9]. Kotwal RS et al. found that the death rate of massive blood transfusion group was significantly lower than non-massive blood transfusion group, especially in severe

and extremely severe trauma (ISS > 15). However, regardless of the severity of trauma, the mortality decreased gradually in massive blood transfusion group, non-massive blood transfusion group and non-transfusion group, and there was significant difference. With the increase of blood transfusion, the mortality rate gradually increased during hospitalization[10]. Therefore, blood products should be given early in the pre-hospital transfer to improve the patients' survival rate after trauma, and then other means should be intervened as soon as possible to strictly control the amount of blood transfusion.

At present, there are many studies on traumatic massive blood transfusion, including various trauma scoring systems on the battlefield and civilians[11–13], which are used to predict when to initiate massive blood transfusion programs. However, in recent years, with the improvement of early pre-hospital and hospitalization trauma management measures, the proportion of patients with massive blood transfusion has gradually decreased[14]. For traumatic patients who don't meet the standard of massive blood transfusion, there are few studies on the need for blood transfusion. The fifth edition of the European Trauma Guide recommends that the target Hb should be maintained at (70–90) g/L[5], which can be used as a reference for blood transfusion needs, but the guidelines also suggest that the normal initial test results of Hb may mask bleeding, and it is recommended to use the results of repeated Hb tests as laboratory indicators of bleeding. Therefore, only by the results of Hb to judge whether blood transfusion or not, reference value is limited. How to judge the best demand for blood transfusion according to the changing vital signs of trauma patients is a difficult problem for emergency doctors. At present, most of the blood transfusion decisions made by doctors based on their personal experience, but there is no feasible and recognized reference standard for different individuals. Transfusion too early will not only cause a waste of blood components, but also affect the prognosis of patients with excessive blood transfusion[10, 15]. Delayed blood transfusion will lead to hemorrhagic shock, aggravate complications such as hypothermia, acidosis and coagulation dysfunction, and seriously affect the survival rate of patients[16].

We believe that the application of machine learning methods, compared with traditional statistical methods, can help us to identify whether patients have the need for blood transfusion and reduce unnecessary complications caused by delayed transfusion, insufficient blood transfusion or excessive transfusion. Therefore, this study proposes a new method to establish a mathematical model of artificial intelligence by retrospective analysis of patients' vital signs, laboratory tests / tests and other data, to assist doctors to quickly make decisions on whether or not blood transfusion is needed after trauma, and to improve the success rate of patient treatment.

Methods

Clinical data

The Emergency Trauma Database of the Chinese People's Liberation Army(PLA) General Hospital is a comprehensive, unidentified dataset containing medical information on 22, 491 critically ill patients during January 2014 to January 2018[17]. All patients were admitted to the emergency department. A

total of 1371 trauma patients' medical information were extracted who were triaged to a critical rescue room. The data related to blood transfusion are provided by the clinical blood transfusion intelligent management and evaluation system established by the Transfusion Department of the Chinese PLA General Hospital[18]. The patient's uniquely identifies the information of the two databases associated with the outpatient number. In the process of data extraction, the original data is completely consistent with the database data through quality control. The Medical Ethics Committee of the Chinese PLA General Hospital approved the study and waived the requirement for written informed consent.

Contains variables

Basic information (age, sex, height, weight); diagnosis; admission time; discharge time; after-department track; blood transfusion time; blood transfusion components; infusion volume of RBC.

Non-invasive detection parameters: vital signs [heart rate(HR), respiration(R), shock index(SI), systolic blood pressure(SBP), diastolic blood pressure(DBP), blood oxygen saturation(SpO₂), temperature(T)] and test time, trauma site;

Invasive detection parameters: routine blood test parameters [hemoglobin(Hb), hematocrit(Hct), platelet count(PLT), C-reactive protein(CRP), Interleukin(IL)-6] and test time; coagulation indicators [prothrombin time(PT), activated partial thromboplastin time(APTT), international standardized ratio(INR), prothrombin activity(PTA), fibrinogen(Fib)] and test time; blood gas test parameters [potential of hydrogen(PH), partial pressure of oxygen(PO₂), partial pressure of carbon dioxide(PCO₂), total carbon dioxide(TCO₂), lactate(Lac), actual bicarbonate(AB), Standard bicarbonate(SB), potassium(K)] and detection time; trauma severity classification (first level, second level and third level); endotracheal intubation; vasoactive drugs;

Construct new variables: For the trauma diagnostic classification, we divided them into the fields of trauma type (open trauma, blunt injury) and trauma site (head and neck, upper extremity, lower extremity, chest and abdomen, spine, trunk and pelvis).

Variable dimensionality reduction: In order to reduce the time and complexity of the model operation, only one variable with high correlation coefficient is retained, such as Hb and Hct, only the variable Hct is retained.

Inclusion and exclusion criteria: Inclusion criteria: (1) patients' diagnosis were matching or fuzzy matching with "injury", (2) patients were triaged from the emergency department to a critical rescue room. Exclusion criteria: (1) patients with non-external trauma, (2) age < 18 years old.

Acquisition of variables

The process of getting variables includes extracting and aggregating variables, cleaning variables, processing variables.

Extraction of variables: The numerical variables can be extracted directly, including vital sign parameters, laboratory test results, and information related to blood transfusion in the database. The results of the first examination when entering the emergency department are used as variables to predict the demand for blood transfusion. If multiple tests are performed before or after blood transfusion, the results closest to the time blood transfusion are included in the analysis. We use natural language processing to extract effective information from unstructured text variables in the database in advance, such as diagnosis, medical orders, etc. We extract the variable information from the emergency trauma database, and then take the patient unique identification as the center, associate with the blood transfusion information of the clinical blood transfusion database system, and aggregate into a record.

Variable cleaning: We needed clean up duplicate data and formulate retention principles, such as testing the changes of vital signs many times after entering the emergency department, and taking the results of the first test as the key variable to judge whether blood transfusion or not; check the invalid value and establish the criteria, such as height, weight with - 1, 0, etc.; check the logical relationship among the data, such as the time of admission, the time of laboratory examination, the start time of blood transfusion and so on.

Variable processing: (1) Classify variable processing, turn it into numerical vector and then use it to build models, such as gender and other variables; (2) Unstructured text variable processing, using the automatic counting word segmentation algorithm in natural language processing to transform words into numerical variables; (3) In order to construct new variables, the diagnostic information of patients such as diagnostic details and variable processing of trauma sites, were divided into phrases and fields, and then counted and scored in the target variables of different categories, and trained the model use the learned rules to construct new variables.

The establishment of the model

SPSS 22.0 software was used to establish LR model and CRT model.

CRT is the supervised analysis technology, which uses the binary syncopation method to cut the data into two pieces at a time and enter into the left and right two trees respectively. The root node of the tree is a dependent variable, and the child node is based on the classification variable (parent node). The minimum sample size on the CRT parent node established by the non-invasive parameter is 20 and the child node is 10. The minimum sample size of the CRT parent node for invasive parameters is 50 and the child node is 20. If the sample size on the node does not meet this requirement, the node is the terminal node and will no longer be segmented[19].

Xgboost is a gradient lifting decision tree algorithm that is executed by using the `train_test_plt()` function provided by the Python language. Xgboost belongs to supervised learning method, and is an integrated learning model which is used for classification analysis (processing discrete data) and regression tree analysis (processing continuous data). The Xgboost algorithm is composed of loss function and regular function. The loss function calculates the error between the prediction and the real result, and the loss

function is constrained based on the minimum error in the actual calculation. The regular function is used to detect the complexity of the model to avoid over-fitting. The loss function and the objective function are given according to the actual situation.

Statistic analysis

The counting data are described by frequency and percentage [n (%)], and the measurement data are expressed by mean and standard deviation [Mean (SD)] or median and quartile spacing [Median [Range]]. The measurement data of the two groups were compared by analysis of variance or Kruskal-Wallis nonparametric test, and the counting data of the two groups were compared by chi-square test. If the test $P < 0.05$ indicated the difference was statistically significant.

LR method was used to screen the significant variables with $p < 0.05$ as independent variables and whether blood transfusion was used as dependent variables to establish the model. After the regression coefficient was standardized, the risk factors OR and 95%CI were used to express the relationship between variables and the occurrence of blood transfusion.

CRT and Xgboost models, using the original variables, combined variables or constructed new variables of historical data sets for model training. The historical data set is randomly divided into 80% training set and 20% test set. The model is trained on the training set and the effect of the model is evaluated on the test set.

LR, CRT and Xgboost models were compared with whether blood transfusion was used as the target variable, method 1-basic information + non-invasive parameters as analysis variables, and method 2-basic information + non-invasive parameters + invasive parameters as analysis variables to establish models, and AUC were drawn and analyzed respectively. The AUC results of two methods and three models were compared by python software, if $P < 0.05$, the difference was statistically significant.

According to the node level (root node, child node) of each variable in the decision tree, the CRT model reflects the importance of each variable. The Xgboost model is represented by the weight of the factors in the tree model of the gradient lifting decision tree algorithm.

Results

Patient characteristics

The emergency trauma database of the General Hospital of the Chinese PLA contains the medical information of 22491 critically ill patients. We included 1371 patients who met the study criteria for analysis. Among them, there were 324 females (23.6%) and 1047 males (76.4%). 1183 patients (86.3%) did not receive blood transfusion, and 188 patients (13.7%) received blood transfusion. There was significant difference between transfusion group and non-transfusion group in age, HR, SBP, DBP, SI, HB, Hct, PLT, PT, APTT, PTA, Fib, PH, PO₂, TC0₂, Lac, AB, SB, K, endotracheal intubation, vasoactive drugs, trauma site, RBC volume, 24-hour RBC and emergency department time ($P < 0.05$). There was no

significant difference between transfusion group and non-transfusion group in sex, height, weight, R, SpO₂, T, CRP, IL-6, INR, PCO₂, trauma severity classification and trauma type($P > 0.05$) (Table 1).

Table 1
Univariate analysis on predicting factors for transfusion

Variable		N(%)	No-transfusion Mean(SD)/ Median[Range] N = 1183	Transfusion Mean(SD)/ Median[Range] N = 188	P-value
Age*		1371(100%)	44.00 [29.00,56.50]	42.00 [28.00,54.25]	0.049
Sex	Female	324(23.6%)	284 (24.01)	40 (21.28)	0.468
	Male	1047(76.4%)	899 (75.99)	148 (78.72)	
Height(cm)		488(35.6%)	170.00 [164.00,175.00]	170.00 [163.50,173.25]	0.242
Weight(kg)		477(34.8%)	68.00 [60.00,75.00]	67.75 [60.00,76.00]	0.541
Non-invasive parameters	HR**	785(57.3%)	96.95 (24.31)	103.55 (25.87)	0
	R	786(57.3%)	21.00 [19.00,23.00]	21.00 [19.00,26.00]	0.071
	SBP(mmHg) **	787(57.4%)	124.32 (25.28)	117.65 (27.46)	0
	DBP(mmHg) **	787(57.4%)	77.58 (15.96)	74.62 (17.83)	0
	SpO2(%)	783(57.1%)	98.00 [96.00,99.00]	98.00 [96.00,99.00]	0.113
	SI**	785(57.3%)	0.82 (0.29)	0.95 (0.43)	0
	T(°C)	637(46.5%)	37.00 [36.80,37.30]	37.00 [36.70,37.30]	0.389
Invasive detection parameters	HB(g/L) **	1287(93.9%)	126.00 [107.00,143.00]	107.00 [82.00,135.00]	0
	Hct(L/L) **	1300(94.8%)	3.80 [0.50,22.00]	0.46 [0.25,4.00]	0
Remarks: HR: heart rate, R: respiration, SBP: systolic blood pressure, DBP: diastolic blood pressure, SpO2: blood oxygen saturation, T: temperature, SI: Shock index, Hb: hemoglobin, Hct: hematocrit, PLT: platelet count, CRP: C-reactive protein, IL-6: Interleukin-6, PT: prothrombin time, APTT: activated partial thromboplastin time, INR: international standardized ratio, PTA: prothrombin activity, Fib: fibrinogen, PH: potential of hydrogen, PO2: partial pressure of oxygen, PCO2: partial pressure of carbon dioxide, TCO2: total carbon dioxide, SPO2: oxygen saturation, Lac: lactate, AB: actual bicarbonate, SB: Standard bicarbonate, K: potassium, RBC: volume of red blood cell transfusion, 24 h RBC: the volume of 24-hour red blood cell transfusion. N: number.					
P < 0.05,there are statistical differences. “*”: P < 0.05, “**”: P < 0.01.					

Variable	N(%)	No-transfusion	Transfusion	P-value
		Mean(SD)/	Mean(SD)/	
		Median[Range]	Median[Range]	
		N = 1183	N = 188	
PLT($10^9/L$) **	1225(98.4%)	216.81 (94.70)	201.76 (99.53)	0
CRP(mg/L)	1075(78.4%)	0.95 [0.10,4.87]	0.41 [0.10,3.41]	0.806
IL-6(pg/ml)	581(42.4%)	182.37 (380.08)	219.26 (388.41)	0.412
PT(s) **	1230(89.7%)	14.70 [14.00,16.00]	15.40 [14.20,17.08]	0
APTT(s) **	1227(89.5%)	37.03 (10.39)	38.50 (12.86)	0
INR	1224(89.3%)	15.40 [14.60,16.40]	15.60 [14.60,16.60]	0.698
PTA(%) **	1230(89.7%)	80.00 [68.00,89.00]	73.50 [61.25,85.00]	0
Fib(g/L) **	1216(88.7%)	3.15 (1.76)	2.72 (1.51)	0
PH**	1230(89.7%)	1.16 [1.08,1.28]	1.22 [1.11,1.39]	0
PO2(mmHg) **	1123(81.9%)	120.73 (62.16)	134.12 (73.88)	0.001
PCO2(mmHg)	1124(82.0%)	37.00 [33.00,41.00]	37.00 [32.00,41.00]	0.116
TCO2(mmol/L) **	1121(81.8%)	24.01 (4.38)	22.66 (4.85)	0
Lac(mmol/L) **	1122(81.8%)	7.41 [7.37,7.45]	7.39 [7.35,7.43]	0
AB(mmol/L) **	1124(82.0%)	22.88 (4.24)	21.54 (4.73)	0
SB(mmol/L) **	1122(81.8%)	23.69 (3.56)	22.39 (4.31)	0

Remarks: HR: heart rate, R: respiration, SBP: systolic blood pressure, DBP: diastolic blood pressure, SpO2: blood oxygen saturation, T: temperature, SI: Shock index, Hb: hemoglobin, Hct: hematocrit, PLT: platelet count, CRP: C-reactive protein, IL-6: Interleukin-6, PT: prothrombin time, APTT: activated partial thromboplastin time, INR: international standardized ratio, PTA: prothrombin activity, Fib: fibrinogen, PH: potential of hydrogen, PO2: partial pressure of oxygen, PCO2: partial pressure of carbon dioxide, TCO2: total carbon dioxide, SPO2: oxygen saturation, Lac: lactate, AB: actual bicarbonate, SB: Standard bicarbonate, K: potassium, RBC: volume of red blood cell transfusion, 24 h RBC: the volume of 24-hour red blood cell transfusion. N: number.

P < 0.05, there are statistical differences. “*”: P < 0.05, “**”: P < 0.01.

Variable		N(%)	No-transfusion Mean(SD)/ Median[Range] N = 1183	Transfusion Mean(SD)/ Median[Range] N = 188	P-value
	K(mmol/L) **	1278(93.2%)	3.88 [3.56,4.10]	3.90 [3.60,4.24]	0.008
Endotracheal intubation*	No	1148(83.7%)	1003 (84.78)	145 (77.13)	0.011
	Yes	223(16.3%)	180 (15.22)	43 (22.87)	
Vasoactive drugs**	No	1264(92.2%)	1113 (94.08)	151 (80.32)	0
	Yes	107(7.8%)	70 (5.92)	37 (19.68)	
Trauma site**	Upper extremity	24(1.8%)	21 (1.87)	3 (1.66)	0
	Lower extremity	49(3.6%)	42 (3.73)	7 (3.87)	
	Head and neck	414(30.2%)	386 (34.28)	28 (15.47)	
	Chest and abdomen	639(46.7%)	538 (47.78)	101 (55.8)	
	Spine	75(5.5%)	62 (5.51)	13 (7.18)	
	Trunk	27(2.0%)	20 (1.78)	7 (3.87)	
	Pelvis	79(5.8%)	57 (5.06)	22 (12.15)	
Trauma severity classification	First level	1160(84.6%)	997 (84.42)	163 (86.7)	0.529
	Second level	204(14.9%)	179 (15.16)	25 (13.3)	
	Third level	5(0.4%)	5 (0.42)	0	
Trauma type	Open trauma	867(63.2%)	747 (63.14)	120 (63.83)	0.921
	Blunt injury	504(36.8%)	436 (36.86)	68 (36.17)	
RBC(U) **		1371(100%)	0.00	2.00 [0.00,4.00]	0
Remarks: HR: heart rate, R: respiration, SBP: systolic blood pressure, DBP: diastolic blood pressure, SpO2: blood oxygen saturation, T: temperature, SI: Shock index, Hb: hemoglobin, Hct: hematocrit, PLT: platelet count, CRP: C-reactive protein, IL-6: Interleukin-6, PT: prothrombin time, APTT: activated partial thromboplastin time, INR: international standardized ratio, PTA: prothrombin activity, Fib: fibrinogen, PH: potential of hydrogen, PO2: partial pressure of oxygen, PCO2: partial pressure of carbon dioxide, TCO2: total carbon dioxide, SPO2: oxygen saturation, Lac: lactate, AB: actual bicarbonate, SB: Standard bicarbonate, K: potassium, RBC: volume of red blood cell transfusion, 24 h RBC: the volume of 24-hour red blood cell transfusion. N: number.					
P < 0.05,there are statistical differences. “*”: P < 0.05, “**”: P < 0.01.					

Variable	N(%)	No-transfusion	Transfusion	P-value
		Mean(SD)/	Mean(SD)/	
		Median[Range]	Median[Range]	
		N = 1183	N = 188	
24 h RBC(U) **	1371(100%)	0.00	2.00 [0.00,4.00]	0
Emergency department time(h) **	1371(100%)	23.58 (33.69)	25.61 (37.80)	0.001
Remarks: HR: heart rate, R: respiration, SBP: systolic blood pressure, DBP: diastolic blood pressure, SpO2: blood oxygen saturation, T: temperature, SI: Shock index, Hb: hemoglobin, Hct: hematocrit, PLT: platelet count, CRP: C-reactive protein, IL-6: Interleukin-6, PT: prothrombin time, APTT: activated partial thromboplastin time, INR: international standardized ratio, PTA: prothrombin activity, Fib: fibrinogen, PH: potential of hydrogen, PO2: partial pressure of oxygen, PCO2: partial pressure of carbon dioxide, TCO2: total carbon dioxide, SPO2: oxygen saturation, Lac: lactate, AB: actual bicarbonate, SB: Standard bicarbonate, K: potassium, RBC: volume of red blood cell transfusion, 24 h RBC: the volume of 24-hour red blood cell transfusion. N: number.				
P < 0.05,there are statistical differences. “*”: P < 0.05, “**”: P < 0.01.				

Model prediction

Methods 1-the model established with non-invasive parameters predicted the need for blood transfusion after trauma. The AUC of LR 0.72 (95%CI 0.657–0.775) was highest than the AUC of Xgboost model 0.71 (95%CI 0.654–0.756) and AUC of CRT model 0.69 (95%CI 0.633–0.751) (Fig. 1A). There was significant difference in AUC among the three models (P < 0.05). The accuracy of Xgboost model was 0.75, which was highest than that of LR 0.55 and CRT 0.48. Methods 2-the model established with all parameters were used to predict the need for blood transfusion after trauma. The AUC of Xgboost model was 0.94 (95%CI 0.893–0.981), which was highest than that of AUC of CRT model 0.82 (95%CI 0.779–0.853) and AUC of LR model 0.80 (95%CI 0.744–0.850)(Fig. 1B). There was significant difference in AUC among the three models (P < 0.05). The accuracy of CRT model is 0.89, which is highest than that of Xgboost 0.83 and LR 0.72 (Table 2).

Table 2
Comparison between LR, CRT and Xgboost model in predicting Blood transfusion

Comparison between LR, CRT and Xgboost model in predicting blood transfusion							
Parameters type	Methods	AUC	Sensitivity	Specificity	Accuracy	Youden Index	P-value
Non-invasive parameters	Xgboost	0.705	0.66	0.77	0.75	0.19	< 0.001
	Logistic	0.716	0.86	0.5	0.55	0.12	
	CRT	0.692	0.89	0.42	0.48	0.16	
All parameters	Xgboost	0.937	0.94	0.82	0.83	0.10	< 0.001
	Logistic	0.797	0.8	0.7	0.72	0.12	
	CRT	0.816	0.69	0.92	0.89	0.09	
Remarks: AUC: area under the curve; CRT: classification and regression tree.							

Variable importance analysis

Method 1 Non-invasive detection parameters were used to predict blood transfusion. LR analysis showed that, trauma site (OR = 18.371, 95% CI 4.019–83.931, $P < 0.05$) and SI (OR = 3.463, 95% CI 1.763–6.801, $P < 0.05$) were risk factors for predicting blood transfusion (An additional table shows this in more detail [see Additional Table 1]). The results of CRT model analysis show that the order of importance of variables were SI, trauma site, age and SpO2 (Fig. 2A). The top five variables in the Xgboost model were: trauma site, SBP, SI, DBP and HR (Fig. 3A).

Method 2 All test parameters were used to predict blood transfusion. LR analysis showed that trauma site (OR = 7.961, 95% CI 1.422–44.567), vasoactive drugs (OR = 2.039, 95% CI 1.092–3.808), PLT (OR = 0.995, 95% CI 0.992–0.998), PTA (OR = 0.975, 95% CI 0.964–0.988), Hct/Hb (OR = 0.923, 95% CI 0.899–0.948), SB and Fib were risk factors for blood transfusion ($P < 0.05$)(Table 3). The results of CRT model analysis show that the order of importance of variables were Hct/HB, Fib and CRP (Fig. 2B). The top five variables in the Xgboost model were in turn: Hct/HB, TC02, PH, PCO2 and CRP(Fig. 3B).

Table 3
Binary Logistic regression analysis for predicting transfusion with all (non-invasive + invasive) parameters

Variable	OR	2.5%	97.5%	P-value
SB	0.898	0.844	0.957	0.001
Hct	0.923	0.899	0.948	0
VD	2.039	1.092	3.808	0.025
Trauma site	7.961	1.422	44.567	0.018
PTA	0.975	0.964	0.988	0
SpO2	1.023	0.977	1.071	0.323
PLT	0.995	0.992	0.998	0.001
Fib	0.789	0.674	0.924	0.003
Variable	OR	2.5%	97.5%	P-value
T	0.843	0.626	1.137	0.262
Trauma site	18.371	4.019	83.931	0
SI	3.463	1.763	6.801	0
Remarks: T: temperature, SI: Shock index. OR: odd ratio.				
Remarks: SB: Standard bicarbonate, Hct: hematocrit, VD: Vasoactive drugs, PTA: prothrombin activity, SpO2: blood oxygen saturation, PLT: platelet count, Fib: fibrinogen, OR: odd ratio.				

Discussion

In our study, non-invasive detection parameters and all parameters were established to predict blood transfusion in trauma patients, and the decision tree algorithm (CRT and Xgboost) was compared with the traditional statistical method (LR). The results showed that the LR model with basic information and non-invasive parameters was the best, but the sensitivity of CRT model was the highest, and the specificity and accuracy of Xgboost model was the highest. The AUC of basic information + non-invasive parameter + invasive parameter model was higher than that of non-invasive parameter model. The Xgboost model was the best and the sensitivity was the highest, but the CRT model was the specificity with the highest accuracy.

AUC embodies the classification ability of the model. LR had the best classification ability in non-invasive parameter prediction, but it was suitable for data analysis and couldn't be used in clinical application. The decision tree algorithm had its own advantages, and the CRT model had the highest sensitivity and the best ability to identify patients who needed blood transfusion. The specificity and accuracy of Xgboost model was the highest, and the ability to identify blood transfusion / non-transfusion was the

best. When predicting all the parameters, the Xgboost model was the best, and the ability to identify blood transfusion was the best. CRT model had the best ability to identify transfusion / non-transfusion. The results showed that the more parameters, the more prominent the advantages of the decision tree model. The non-invasive parameters can be quickly obtained after the trauma patients have obtained medical resources, and the input data can be used to quickly feedback the results of whether the patients need blood transfusion or not by using the decision tree model. Although the prediction efficiency is slightly lower than all parameters, its time advantage is incomparable. Moreover, the trauma patients are accompanied by the changes of blood loss and fluid volume, and the vital signs are complex and changeable. The detection time of invasive parameters is about 1 hour. When the results are obtained, they can no longer reflect the current physiological parameters of the patients. Therefore, the non-invasive parameters obtained at any time can reflect the vital signs of patients at that time, and the model can be used to predict at any time, which is convenient for clinical application. When predicting all the parameters, the blood transfusion decisions made by clinicians based on experience are often not accurate. In the case of covering as many data and variables as possible, through a part of the data as a training set, on the basis of learning the experience of clinicians, the method of machine learning can more accurately and digitally assist doctors in the decision support of blood transfusion for trauma patients.

The treatment of trauma should take into account the mechanism of the trauma (open trauma or blunt injury), the location of the trauma (head, chest, etc.), pre-hospital resources, hospital emergency room settings (I, II, etc.) and trauma center facilities (immediate detection equipment and resources)[20]. Similarly, this study found that when predicting non-invasive parameters, the location of trauma and SI had the greatest impact on blood transfusion. The model established by combining age, sex, pre-hospital SI, admission HR, Hb and SpO₂ can better predict blood transfusion 3 hours before admission[21]. The post-traumatic SI is important in assessing the need for blood transfusion and can predict the demand for massive blood transfusion, laparotomy and mortality[22]. Shock index is more sensitive than ABC score in predicting traumatic massive blood transfusion[23].

Among the predictive variables of all parameters, Hct/Hb had great influence on blood transfusion in the three models. Consistent with our study, many models or scoring systems use Hct/Hb as the main parameter for the prediction of traumatic massive blood transfusion[12, 13, 24]. It is also consistent with the recommendation of the guidelines that Hb repeat test results should be used as a laboratory indicator of bleeding[5]. Different models have different parameters that affect whether or not blood transfusion is carried out. LR model is to judge the influence of variables on blood transfusion by risk factor (OR), and the results are generally recognized clinically. Except for trauma site and Hct/Hb, vasoactive drugs, PLT, PTA, and Fib were risk factors for blood transfusion demand. The study found that the use of vasoactive drugs can improve vital signs[25], and early routine medication can improve the effective rate of treatment of patients with severe trauma. Traumatic coagulation is easy to occur in the early stage of trauma, and the coagulation index (PLT, PTA, Fib) affects the demand for blood transfusion[16, 26]. In the process of building the CRT model, the variables corresponding to the root nodes are the most important, followed by the leaf nodes which split in turn[19]. In addition to Hct/Hb, Fib, CRP is an important variable

for predicting blood transfusion. Because CRP is an indicator of body stress, CRP stress increases after trauma, which can reflect the severity of trauma[27]. In the process of establishing the Xgboost model, the more times the nodes are traversed, the more important the variables corresponding to the nodes are. The importance of variables is mathematical relevance, and whether it has clinical guiding value needs to be comprehensively analyzed in combination with clinical experience.

With the progress of science and technology, artificial intelligence methods have been widely used in the field of medicine[28–31]. There are a lot of research on machine learning methods in trauma[32–34]. There have been a lot of research on the prediction of massive blood transfusion, and the prediction accuracy of decision tree algorithm is 0.695–0.814[35, 36]. Machine learning (mostly neural network) in a large number of studies to predict the prognosis of trauma, most studies have proved the benefits of machine learning methods, the sensitivity-specificity difference ranges from 0.035 to 0.927[37]. The neural network algorithm accuracy (98.7%), specificity (51.5%) was the highest in predicting the survival rate of trauma patients[38].

Our research compares the traditional statistical methods with the decision tree algorithm of machine learning, and the decision tree algorithm has outstanding advantages: (1) Most of the data in the real world are incomplete (missing key indicators) and noisy (numerical errors / anomalies). Artificial intelligence can allow cases with missing or outliers to be retained by interpolation and other methods. The larger the number of cases, the more meaningful the statistical results. (2) Xgboost algorithm is widely used in medicine, and the prediction performance is good[39, 40]. (3) The model can reconstruct more effective features from the training process of blood transfusion big data, which can be used to predict the blood transfusion volume of patients, so as to make the model have stronger generalization ability and reduce overfitting. (4) Using the difference between the prediction results and the training data for training, with the gradual increase of the amount of data, the accuracy is improved in the iterative process, which ensures the incremental learning characteristics of the model. (5) At present, doctors are widely used to make the decision of blood transfusion by combining various physiological parameters, symptoms and clinical experience. Our research uses a large number of historical data as a reference, on the basis of doctors' rich clinical experience, establishes a mathematical model, and adjusts the output of multiple experiments to get the best results. It has more practical value for primary hospitals or inexperienced doctors.

Limitations of the study

In this study, the method of artificial intelligence is used to construct the mathematical model, which is limited to the fact that the amount of data is not large enough, and the accuracy of the model needs to be improved, but with the increase of the amount of data and the continuous optimization of the model, the prediction accuracy of the model will gradually improve. The variables extracted from unstructured text information are limited, which don't improve the performance of the model, so how to use the effective information to improve the prediction efficiency of the model is the direction of our further research. Some of the patients in our trauma database are transferred to our hospital from primary hospitals after

emergency treatment (including blood transfusion), so the number of patients requiring emergency massive blood transfusion is relatively small, but it does not affect the establishment and application of the model. Because our model can make decisions on whether or not to transfuse based on changing, real-time vital signs and laboratory data in the process of trauma development. With a large amount of blood loss after trauma, complications such as hypothermia, acidosis and coagulation dysfunction are easy to occur, and the amount of plasma and platelet transfusion has an effect on the demand of red blood cells. However, our model includes indicators that reflect these symptoms, so the effects of these complications and blood components on erythrocyte demand have been taken into account.

Conclusions

The intelligent evaluation of the demand for post-traumatic blood transfusion shows that the prediction of non-invasive parameters is rapid and fast, and the prediction of all parameters is accurate and convenient. The model established in our study can be connected with ambulances and doctors' work computers, and is widely used in clinic as an auxiliary tool to provide support for clinicians.

Abbreviations

CRT

Classification and Regression Tree, XGboost:eXtreme Gradient Boosting, LR:logistic regression, AUC:Area Under Curve, PLA:People's Liberation Army, HR:heart rate, R:respiration, SBP:systolic blood pressure, DBP:diastolic blood pressure, SpO2:blood oxygen saturation, T:temperature, SI:Shock index, Hb:hemoglobin, Hct:hematocrit, PLT:platelet count, CRP:C-reactive protein, IL-6:Interleukin-6, PT:prothrombin time, APTT:activated partial thromboplastin time, INR:international standardized ratio, PTA:prothrombin activity, Fib:fibrinogen, PH:potential of hydrogen, PO2:partial pressure of oxygen, PCO2:partial pressure of carbon dioxide, TCO2:total carbon dioxide, SPO2:oxygen saturation, Lac:lactate, AB:actual bicarbonate, SB:Standard bicarbonate, K:potassium, RBC:volume of red blood cell transfusion, 24 h RBC:the volume of 24-hour red blood cell transfusion. VD:Vasoactive drugs, Trauma_loc:trauma site.

Declarations

Acknowledgements

The authors thank Tanshi Li's team in the emergency department and Wanguo Xue's team in big data Center for their help and support. The authors thank Chunya Ma for helpful discussions or comments.

Authors' contributions

Yannan Feng performed data analysis, wrote, and submitted the manuscript. Zhenhua Xu analyzed the data and visualized the results. Junting Liu and Xiaolin Sun collected the data. Deqing Wang designed the study, and contributed to preparation of the manuscript. Yang Yu contributed to design the study, data analysis and revise the manuscript. All authors read and approved the final manuscript.

Funding

The authors wish to thank the Key Project-subtopic of "13th five-year plan" Military Logistics Service Research of China (Grant No. BWS16J006), China, for financially supporting this study.

Availability of data and materials

All authors had full access to all the data in the study.

Ethics approval and consent to participate

The Medical Ethics Committee of the Chinese PLA General Hospital approved the study and waived the requirement for written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflicts of interest.

Author details

¹Department of Blood Transfusion, Chinese PLA General Hospital, Beijing, 100853, China;

²Beijing hexing chuanglian Health Technology Co., Ltd, Beijing, 100000, China.

References

- [1] GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1151-210.
- [2] Demetriades D, Murray J, Charalambides K, Alo K, Velmahos G, Rhee P, et al. Trauma fatalities: time and location of hospital deaths. *J Am Coll Surg*. 2004;198(1):20-6.
- [3] Scerbo MH, Holcomb JB, Taub E, Gates K, Love JD, Wade CE, et al. The trauma center is too late: Major limb trauma without a pre-hospital tourniquet has increased death from hemorrhagic shock. *J Trauma Acute Care Surg*. 2017;83a(6):1165-72.
- [4] Eastridge BJ, Mabry RL, Seguin P, Cantrell J, Tops T, Uribe P, et al. Death on the battlefield (2001-2011): implications for the future of combat casualty care. *J Trauma Acute Care Surg*. 2012;73(6 Suppl 5):S431-7.
- [5] Spahn DR, Bouillon B, Cerny V, Duranteau J, Filipescu D, Hunt BJ, et al. The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition. *Crit Care*. 2019;23(1):98.
- [6] Shackelford SA, Del Junco DJ, Powell-Dunford N, Mazuchowski EL, Howard JT, Kotwal RS, et al. Association of Prehospital Blood Product Transfusion During Medical Evacuation of Combat Casualties in Afghanistan With Acute and 30-Day Survival. *JAMA*. 2017;318(16):1581-91.

- [7] Croce MA, Tolley EA, Claridge JA, Fabian TC. Transfusions result in pulmonary morbidity and death after a moderate degree of injury. *J Trauma*. 2005;59(1):19-23; discussion 23-4.
- [8] Marik PE, Corwin HL. Efficacy of red blood cell transfusion in the critically ill: a systematic review of the literature. *Crit Care Med*. 2008;36(9):2667-74.
- [9] Malone DL, Dunne J, Tracy JK, Putnam AT, Scalea TM, Napolitano LM. Blood transfusion, independent of shock severity, is associated with worse outcome in trauma. *J Trauma*. 2003;54(5):898-905; discussion 905-7.
- [10] Kotwal RS, Howard JT, Orman JA, Tarpey BW, Bailey JA, Champion HR, et al. The Effect of a Golden Hour Policy on the Morbidity and Mortality of Combat Casualties. *JAMA Surg*. 2016;151(1):15-24.
- [11] 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(2):346-52.
- [12] 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(6):1228-36; discussion 1236-7.
- [13] Schreiber MA, Perkins J, Kiraly L, Underwood S, Wade C, Holcomb JB. Early predictors of massive transfusion in combat casualties. *J Am Coll Surg*. 2007;205(4):541-5.
- [14] Maegele M. Challenges to improving patient outcome following massive transfusion in severe trauma. *Expert Rev Hematol*. 2020;13(4):323-30.
- [15] Leal-Noval SR, Rincón-Ferrari MD, Muñoz-Gómez M. Red blood cell transfusion may be more detrimental than anemia for the clinical outcome of patients with severe traumatic brain injury. *Crit Care*. 2019;23(1):189.
- [16] Chang R, Cardenas JC, Wade CE, Holcomb JB. Advances in the understanding of trauma-induced coagulopathy. *Blood*. 2016;128(8):1043-9.
- [17] Zhao Y, Wang J, Pan F, Li P, Jia L, Li K, et al. [Pilot research: construction of emergency rescue database]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*. 2018;30(6):609-12.
- [18] Chen LF, Li H, Zhuang Y, Luo Q, Yu Y, Pan JC, et al. Application and construction of a clinical transfusion intelligent management and evaluation system. *Chinese Journal of Blood Transfusion*. 2015;28(09): 1167-73.
- [19] Zhang S, Hu Z, Ye L, Zheng Y. [Application of Logistic regression and decision tree analysis in prediction of acute myocardial infarction events]. *Zhejiang Da Xue Xue Bao Yi Xue Ban*. 2019;48(6):594-602.

- [20] El-Menyar A, Mekkodathil A, Abdelrahman H, Latifi R, Galwankar S, Al-Thani H, et al. Review of Existing Scoring Systems For Massive Blood Transfusion in Trauma Patients: Where Do We Stand. *Shock*. 2019.
- [21] Galvagno SM Jr, Hu P, Yang S, Gao C, Hanna D, Shackelford S, et al. Accuracy of continuous noninvasive hemoglobin monitoring for the prediction of blood transfusions in trauma patients. *J Clin Monit Comput*. 2015;29(6):815-21.
- [22] El-Menyar A, Goyal P, Tilley E, Latifi R. The clinical utility of shock index to predict the need for blood transfusion and outcomes in trauma. *J Surg Res*. 2018;22752-9.
- [23] Schroll R, Swift D, Tatum D, Couch S, Heaney JB, Llado-Farrulla M, et al. Accuracy of shock index versus ABC score to predict need for massive transfusion in trauma patients. *Injury*. 2018;49(1):15-9.
- [24] 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(6):724-9.
- [25] Thorup L, Koch KU, Upton RN, Østergaard L, Rasmussen M. Effects of Vasopressors on Cerebral Circulation and Oxygenation: A Narrative Review of Pharmacodynamics in Health and Traumatic Brain Injury. *J Neurosurg Anesthesiol*. 2020;32(1):18-28.
- [26] Levy JH, Welsby I, Goodnough LT. Fibrinogen as a therapeutic target for bleeding: a review of critical levels and replacement therapy. *Transfusion*. 2014;54(5):1389-405; quiz 1388.
- [27] Halldorsdottir HD, Eriksson J, Persson BP, Herwald H, Lindbom L, Weitzberg E, et al. Heparin-binding protein as a biomarker of post-injury sepsis in trauma patients. *Acta Anaesthesiol Scand*. 2018;62(7):962-73.
- [28] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8.
- [29] Liang H, Tsui BY, Ni H, Valentim C, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. 2019.
- [30] Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388-96.
- [31] Al'Aref SJ, Maliakal G, Singh G, van Rosendael AR, Ma X, Xu Z, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *Eur Heart J*. 2019.
- [32] Bertsimas D, Dunn J, Steele DW, Trikalinos TA, Wang Y. Comparison of Machine Learning Optimal Classification Trees With the Pediatric Emergency Care Applied Research Network Head Trauma Decision

Rules. JAMA Pediatr. 2019;173(7):648-56.

[33] Ruffle JK, Farmer AD, Aziz Q. Artificial Intelligence-Assisted Gastroenterology- Promises and Pitfalls. Am J Gastroenterol. 2019;114(3):422-8.

[34] Afshar M, Phillips A, Karnik N, Mueller J, To D, Gonzalez R, et al. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. J Am Med Inform Assoc. 2019;26(3):254-61.

[35] Hodgman EI, Cripps MW, Mina MJ, Bulger EM, Schreiber MA, Brasel KJ, et al. External validation of a smartphone app model to predict the need for massive transfusion using five different definitions. J Trauma Acute Care Surg. 2018;84(2):397-402.

[36] Seheult JN, Anto VP, Farhat N, Stram MN, Spinella PC, Alarcon L, et al. Application of a recursive partitioning decision tree algorithm for the prediction of massive transfusion in civilian trauma: the MTPitt prediction tool. Transfusion. 2019;59(3):953-64.

[37] Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. Shock. 2017;48(5):504-10.

[38] Rau CS, Wu SC, Chuang JF, Huang CY, Liu HT, Chien PC, et al. Machine Learning Models of Survival Prediction in Trauma Patients. J Clin Med. 2019;8(6).

[39] Xu Y, Ju L, Tong J, Zhou CM, Yang JJ. Machine Learning Algorithms for Predicting the Recurrence of Stage IV Colorectal Cancer After Tumor Resection. Sci Rep. 2020;10(1):2519.

[40] Lu Y, Fu X, Chen F, Wong K. Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning. Artif Intell Med. 2020;102101748.

Figures

Figure 1 Comparison of AUC between LR, CRT and Xgboost models in predicting Blood transfusion. A. Non-invasive parameters to predict.B. All parameters to predict

Comparison of AUC between LR, CRT and Xgboost models in predicting Blood transfusion Remarks: AUC: Area Under Curve, LR: Logistic regression, CRT: Classification and Regression Tree, Xgboost: eXtreme Gradient Boosting.

CRT model analysis for predicting transfusion. A. Non-invasive parameters to predict. B. All parameters to predict. Remarks: RBC: red blood cell, SI: Shock index, Trauma_loc: trauma site, SpO2: blood oxygen saturation. Hct: hematocrit, Fib: fibrinogen, CRP: C-reactive protein. CRT: classification and regression tree.

Figure 3 Feature variables ranked by weight in the prediction model of transfusion. A. Non-invasive parameters to predict. B. All parameters to predict

Figure 3

Feature variables ranked by weight in the prediction model of transfusion. A. Non-invasive parameters to predict. B. All parameters to predict. Remarks: Trauma_loc: trauma site, SBP: systolic blood pressure, SI: Shock index, DBP: diastolic blood pressure, HR: heart rate, R: respiration, T: temperature, SpO2: blood oxygen saturation. Hct: hematocrit, TCO2: total carbon dioxide, PH: potential of hydrogen, PCO2: partial pressure of carbon dioxide, CRP: C-reactive protein, VD: Vasoactive drugs, Fib: fibrinogen, SB: Standard bicarbonate, SCFL: severity classification in first level, Lac: lactate, PO2: partial pressure of oxygen, AB: actual bicarbonate, PLT: platelet count, EDT: Emergency department time, INR: international standardized ratio.