

RESEARCH

Open Access



Pilot deployment of a machine-learning enhanced prediction of need for hemorrhage resuscitation after trauma – the ShockMatrix pilot study

Tobias Gauss^{1,2*}, Jean-Denis Moyer³, Clelia Colas⁴, Manuel Pichon⁵, Nathalie Delhayé⁶, Marie Werner^{7,8}, Veronique Ramonda⁹, Theophile Sempe⁴, Sofiane Medjkoune⁴, Julie Josse¹⁰, Arthur James¹¹, Anatole Harrois^{7,8} and the Traumabase Group¹²

Abstract

Importance Decision-making in trauma patients remains challenging and often results in deviation from guidelines. Machine-Learning (ML) enhanced decision-support could improve hemorrhage resuscitation.

Aim To develop a ML enhanced decision support tool to predict *Need for Hemorrhage Resuscitation (NHR)* (part I) and test the collection of the predictor variables in real time in a smartphone app (part II).

Design, setting, and participants Development of a ML model from a registry to predict *NHR* relying exclusively on prehospital predictors. Several models and imputation techniques were tested. Assess the feasibility to collect the predictors of the model in a customized smartphone app during prealert and generate a prediction in four level-1 trauma centers to compare the predictions to the *gestalt* of the trauma leader.

Main outcomes and measures Part 1: Model output was *NHR* defined by 1) at least one RBC transfusion in resuscitation, 2) transfusion ≥ 4 RBC within 6 h, 3) any hemorrhage control procedure within 6 h or 4) death from hemorrhage within 24 h. The performance metric was the F4-score and compared to reference scores (*RED FLAG*, *ABC*). In part 2, the model and clinician prediction were compared with Likelihood Ratios (LR).

Results From 36,325 eligible patients in the registry (Nov 2010—May 2022), 28,614 were included in the model development (Part 1). Median age was 36 [25–52], median ISS 13 [5–22], 3249/28614 (11%) corresponded to the definition of *NHR*. A XGBoost model with nine prehospital variables generated the best predictive performance for *NHR* according to the F4-score with a score of 0.76 [0.73–0.78]. Over a 3-month period (Aug—Oct 2022), 139 of 391 eligible patients were included in part II (38.5%), 22/139 with *NHR*. Clinician satisfaction was high, no workflow disruption observed and LRs comparable between the model and the clinicians.

Conclusions and relevance The ShockMatrix pilot study developed a simple ML-enhanced *NHR* prediction tool demonstrating a comparable performance to clinical reference scores and clinicians. Collecting the predictor variables in real-time on prealert was feasible and caused no workflow disruption.

Keywords Trauma, Shock, Prediction tool, Machine Learning, Decision Support

*Correspondence:

Tobias Gauss

tgauss@chu-grenoble.fr

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Key points

Question

Is it feasible to develop a machine-learning-enhanced decision support tool capable to predict the need for trauma hemorrhage resuscitation on hospital arrival relying exclusively on prehospital predictors, able to handle missing data and collect the predictor variables in real-time without workflow interruption?

Findings

This pilot study developed a machine learning enhanced need for hemorrhage resuscitation decision support tool and tested the collection of the predictor variables in a smartphone tool.

Meaning

The study demonstrates the feasibility to develop and deploy a user-friendly decision support tool to predict the need for hemorrhage resuscitation after trauma.

Introduction

Emergency situations, such as major trauma management, place considerable cognitive and emotional burdens on even the most experienced clinicians [1]. Under these constraints, clinical decision-making fluctuates between a heuristic and a rational cognitive mode [2]. This pattern contributes to inconsistent and non-reproducible decision-making, leading to deviations from evidence-based guidelines and compromises subsequently care quality and patient outcomes [3, 4].

Trauma-specific decision-support systems offer the potential to mitigate variable decision-making and guideline deviation. These systems range from simple checklists [5], flowcharts [6, 7] to digital tools [8, 9]. Recent advances in mathematical and technological paradigms have enabled the exploration of algorithm-based strategies for enhanced real-time decision-making and complex prediction in trauma. These developments led to numerous machine learning models with a heterogeneous set of predictor variables, often including intrahospital predictors, heterogeneous target outcomes (outputs) and performance levels [10, 11].

Most studies focus on model development, few perform external validation and rarely prospective, and even fewer attempt a prospective workflow integration or real-life validation [12, 13]. Although a recent guideline highlights the need to consider usability, ergonomics, explicability, and human-machine interaction when dealing with decision support tools [14], a knowledge gap persists about patient or workflow impacts and real-life feasibility.

The ShockMatrix pilot study aimed to bridge this gap and was conducted in two steps. The first part developed

a machine-learning prediction algorithm relying exclusively on routine prehospital variables and was capable of handling missing data. The primary hypothesis was that this algorithm would predict the *need for hemorrhage resuscitation* as reliably as established by prediction rules and clinical judgement (part 1). In part 2, the investigators prospectively assessed the ergonomics and usability of a smartphone app to collect the predictor variables required to feed the tool. They did so in a real-life setting during prealert by the trauma leader in a level-1 center and generated predictions. In the French EMS, prealert is usually given between 10 to 60 min before arrival of the patient (median 30 min). If proven to be deployable in a clinical setting, the tool could provide real-time decision support for a timelier mobilization of hemorrhage resuscitation resources and pave the way to large-scale prospective testing.

Methods

The study comes in two parts. Part 1 consisted in the development and selection of the best machine learning model to predict NHR. Part 2 consisted in a prospective observational study to evaluate the usability of a smartphone application to collect the information required to generate a prediction in a real-life setting. The description of the study followed the TRIPOD [15, 16] and DECIDE-AI [14] statements.

Machine learning model development

Sample cohort

This study was conducted on the Traumabase registry. The registry prospectively collects socio-demographic, clinical, biological, therapeutic, and in-hospital evolution data from the prehospital scene to hospital discharge for all severely injured patients admitted to a participating center [17]. The registry dataset contained information collected between November 2010 and May 2022 [18, 19].

All consecutive patients over 18 years directly admitted to one of the 26 participating trauma centers were considered eligible for inclusion. Patients who suffered from a cardio-respiratory arrest during the prehospital phase or with missing information to determine the outcome criterion NHR were excluded. The study obtained approval from the University Paris Nord Ethics committee (CER-2021-106 /project SHOCKMATRIX, 11th February 2022). The Traumabase registry has obtained approval from the Institutional Review Board (Comité de Protection des Personnes, Paris VI) from the Advisory Committee for Information Processing in Health Research (*Comite Consultatif Pour le Traitement de l'information en matière de recherche dans le domaine de la santé* CCTIRS, 11.305bis) and from the National

Data Protection Agency (*Commission Nationale de l'Informatique et des Libertés* CNIL, 911,461).

Model output

The primary outcome, *Need for Hemorrhage Resuscitation* (NHR), was defined as a composite outcome including any one of the following criteria: 1) administration of at least one packed red blood cell (RBC) in the resuscitation room or 2) transfusion of 4 or more RBC Cells within 6 h since admission or 3) the need for a hemorrhage control procedure (interventional radiology and/or surgical control in the operating room) within 6 h of admission or 4) death resulting from hemorrhagic shock within 24 h. These events were registered by trained research assistants into the registry, blinded and without knowledge of the study objective. All participating clinicians were blinded to the model output.

Selection of the predictors and interpretability

The initial phase of model development was designed to identify the most relevant predictors. This choice entailed the selection of a set of variables among all pre-hospital variables from the registry and available to the dispatch physician. Predictors were retained based on the *Shapley Value* approach. Shapley Value of each pre-hospital feature were computed to measure its weight on the prediction of NHR. To obtain and facilitate interpretability the SHAP framework (Shapley Additive exPlanations) was applied. The frameworks assigned each feature an importance value for a prediction based on the game theory optimal Shapley values. For every patient, the SHAP framework gave access to summary plots combining feature importance and effects [20]. The extensive list of initially explored and tested pre-hospital variables are listed in the supplementary material.

Model development

We divided the dataset at random into a *training* (50%), a *validation* (20%) and a *test* set (30%). Outliers and inappropriate observations, as defined by clinicians, were considered as missing values in continuous variables and a new category was created for categorical variables to be implemented subsequently in the model development (Supplementary Material Fig.1). Missing values were handled by mean imputation for continuous variables and augmented with missing-data mask concatenation for all predictors [21].

Being a rare event, the primary outcome, NHR, was considered as imbalanced with an asymmetric distribution. To avoid a bias in favor of the majority class in the prediction, the investigators adopted a random *undersampling* method (training and test set). A percentage of patients without the target feature NHR was randomly

removed from the training set [22]. Model performances were computed on an unbalanced test set to reproduce real-life conditions (Supplementary Material Fig. 2, *Distribution in training, validation and test set before/ after undersampling*). Training on the *undersampled* balanced dataset implied a miscalibration of the probabilities of the model. To correct that effect, the Bayes Minimum Risk theory was used to recalibrate the probabilities and adjust the classification threshold [23] (Supplementary material Fig. 3).

Model selection

In the training set, four tree-based algorithms were trained and compared CART tree, Random Forest, XGBoost and CatBoost. To reduce false negative results, training was performed and iteratively evaluated by F4-score through tenfold cross validation (Supplementary material, Fig. 3). The validation set allowed to determine the thresholds and adjust hyperparameters.

The models were evaluated and compared on the test set using a panel of both common metrics sensitivity (Se), specificity (Sp), accuracy, precision, recall, area under the Precision-Recall curve (AUC PR) and receiver operating curve (AUC ROC), positive and negative likelihood ratio and the more specific F β -measure metric chosen to evaluate at-once Sensitivity and Positive Predictive Value (PPV) with emphasis on false negative (FN) prediction error. Confidence intervals of each performance metric were computed by Bootstrap on the test set for each model through 2000 samples and 70% of resampling. The performance of the models was compared to the reference clinical rules *Red Flag* [24] and ABC score [25] with and without imputation. These rules were chosen since they rely exclusively on pre-hospital variables. The diagnostic threshold was set to 0.11 after consensus of the clinical advisory board [26] to reduce the risk of missing patients in shock.

Statistical analysis and sample size

Continuous data were described as median (quartiles 1–3) and categorical variables as count (percentages).

The sample size was determined by a bootstrap with 2000 iterations to provide a compromise between a confidence interval around the *bootstrap* mean of F4 with an inferior margin sufficiently lower than the clinician reference F4 of 0.63 and required inclusion of 1000 patients into the training set. All calculations performed on Python 3.11.0.

Part II: Pilot study

The smartphone application was developed by professional developers from the scientific partner Cap Gemini Invent (Issy-Les-Moulineaux, Paris, France) as part of a

non-for-profit contribution to for the research project. The application was available to participating clinicians free of charge on the on AppStore and Playstore server (see Supplementary Fig. 4 for the Android screenshot) and to be downloaded to their personal smartphone. A personal login and password granted access to the application. The participants received specific training and manuals to facilitate the use of the application.

Participants and data collection

This phase was conducted over a three-month period from June to August 2022 in five Traumabase centers. After each trauma call activation (*prealert*) and before patient arrival, trauma leaders in the resuscitation room had to connect to the application and collect the nine predictor variables with the possibility to indicate that the variable was not known or unavailable. After collecting the nine predictor variables, clinicians indicated their level of confidence in the information obtained from the dispatch center (“*Very low*”, “*Rather low*”, “*Rather high*”, “*High*”). At the end, each trauma leaders indicated their own subjective appreciation, *gestalt*, of the probability of the model output (NHR, see above) with “*Yes*” or “*No*” and their confidence about their own prediction (ranging from 0, “*I am absolutely not confident in my prediction*” to 100, “*I could not be more confident with my prediction*”). Finally, participants specified their level of experience as trauma leader in years (<3, 3 to 6, 6 to 9, more than 9).

In this pilot study, users did not have access to the results of the model prediction for two reasons. First, for medico-legal and regulatory reasons, and second because the investigators felt it was premature to provide a prediction before a prospective large-scale assessment of the model and to minimize patient risk.

Data processing

The smartphone application stored the data to a HIPAA (Health Insurance Portability and Accountability Act) compliant *Microsoft Azur* server. The data were fed into the prediction model to generate the output, NHR. Patient outcomes including the primary outcome, NHR, were retrieved from the Traumabase registry and compared to trauma leader predictions and to calculate sensitivity, specificity, positive, negative predictive value, and likelihood ratios. In this pilot study, this comparison was not intended to assess the model performance, since the investigators were aware of a lack of power.

Qualitative assessment of the usability

Every participating trauma leader responded to an online questionnaire (see Supplementary material). This questionnaire evaluated the following dimensions on a scale from 1 to 10:

- App ergonomics (touch screen, readability, ease of use, input time, workflow disruption, feasibility,...)
- Acceptability (clinical relevance, response rate, commitment to use, patient impact,
- Overall impression (interest raised, satisfaction, likelihood to participate in follow up study)
- Protocol compliance (level of concordance between server and clinical data, inclusion capacity, missing data,...)
- Project feasibility (recruitment capacity, sample size calculation,...)

Results

Part 1: Model development

Sample cohort

Between November 2010 and May 2022, 36 325 patients were registered in the Traumabase. Among them, 28 614 patients met the inclusion criteria and 3 249 (11%) were diagnosed with *Need for Hemorrhage Resuscitation* (NHR, flowchart Fig. 1). Median age was 36 years [IQR 25–52] and the cohort was mainly composed of men 79% (22 356 male, 6 062 female). Median ISS was 13 [5–22]. Table 1 provides a detailed description of the cohort.

Predictor selection

Based on the Shapley values, nine predictor variables were identified: type of trauma (blunt or penetrating), minimal diastolic blood pressure, minimal systolic blood pressure, maximal heart rate, capillary hemoglobin concentration, volume of crystalloid fluid expansion, intubation, catecholamine use, clinically obvious pelvic trauma. Those with the highest weight were the minimum systolic blood pressure, the capillary hemoglobin, the total volume of fluid expansion and the maximum heart rate. Low values of minimum systolic blood pressure and capillary hemoglobin with high values of heart rate and crystalloid fluid expansion volume implied a high risk of NHR according to the model (Fig. 2).

Model development and selection

The recalibration method applied to the predictions of the model indicated a decision threshold of 11%; patients with a probability above that threshold were considered at risk of a *NHR*. The best predictive performance for *NHR* based according to the F4-score was obtained with the XGBoost technique with a score of 0.76 [0.73–0.78] (Table 2). This threshold limited the rate of false negative cases (174 FN) (Fig. 3).

Cumulated distribution of probabilities over the negative class predicted by the XGBoost model on the validation set generated the results illustrated in Table 3. Setting the threshold for the positive cases lower than the default reduced misclassification of false negative cases

Table 1 Sample cohort clinical characteristics

	Cohort n = 28 614	Missing values n (%)	NHR n = 3 249	No-NHR n = 25 365
Predictors				
Demography				
Median age [Q1-Q3]	36 [25–52]	0 (0%)	40 [27–59]	36 [25–51]
Male, n (%)	22 356 (78.1%)	190 (0.7%)	2 305 (70.9%)	20 051 (79%)
ISS [Q1-Q3]	13 [5-22]	1 147 (4%)	26 [17–38]	10 [5-20]
Continuous				
Prehospital Median minimum SBP [Q1-Q3]	116 [100–130]	3 325 (11.6%)	87 [70–106]	119 [104–130]
Prehospital Median minimum DBP [Q1-Q3]	70 [59–80]	3 428 (12%)	50 [40–65]	70 [60–80]
Prehospital Median maximum HR [Q1-Q3]	94 [80–110]	3 404 (12%)	114 [93–130]	92 [80–107]
Prehospital Median capillary hemoglobin [Q1-Q3]	14 [12.8–15]	5 835 (20.4%)	12.4 [11-14]	14 [13–15.2]
Prehospital Median crystalloid fluid expansion volume [Q1-Q3]	500 [250–1000]	785 (2.7%)	1 000 [500–1 500]	500 [100–750]
Categorical				
Prehospital Cristalloid Fluid expansion, n (%)	21 560 (75.4%)	785 (2.7%)	2 842 (87.5%)	18 718 (73.8%)
Prehospital Intubation, n (%)	6 344 (22.2%)	472 (1.7%)	1 688 (52%)	4 656 (18.4%)
Prehospital Catecholamine use, n (%)	2 557 (8.9%)	813 (2.8%)	1 283 (39.5%)	1 274 (5%)
Prehospital suspected pelvic trauma, n (%)	1 376 (4.8%)	348 (1.2%)	496 (15.3%)	888 (3.5%)
Prehospital suspected penetrating trauma, n (%)	3 621 (12.7%)	126 (0.4%)	546 (16.8%)	3 075 (12.1%)
Output variables				
Median number of packed RBC transfused in the trauma room [Q1-Q3]	0 [0–0]	1 067 (3.7%)	2 [0–3]	0 [0–0]
Median number of packed RBC transfused within 6 h [Q1-Q3]	0 [0–0]	161 (0.6%)	4 [0–6]	0 [0–0]
Hemostatic interventional procedure, n (%)	2 074 (7.3%)	2 264 (7.9%)	2 068 (63.7%)	6 (0.02%)
Median duration before surgery in minutes [Q1-Q3]	174.5 [104–309.25]	1 329 (4.6%)	120 [70–189]	235 [142–450]
In hospital mortality, n (%)	2 201 (7.7%)	1 725 (6%)	916 (28.2%)	1 285 (5.1%)
Death of hemorrhagic shock, n (%)	1 242 (4.3%)	954 (3.3%)	688 (21.2%)	554 (2.2%)
Median duration of hospitalization in days [Q1-Q3]	8 [3-19]	3 624 (12.7%)	18 [4–42]	7 [3-17]

SBP Systolic blood pressure, DBP Diastolic blood pressure, HR Heart rate, RBC Red blood cells, CI Confidence interval

Minimum corresponds to the lowest recorded systolic or diastolic blood pressure during the prehospital phase; maximum corresponds to the highest recorded heart rate pressure during the prehospital phase

and accounting appropriately the true negative cases (Fig. 4).

Part 2: pilot study

The three-month pilot study included 139 of 361 eligible patients (38.5%). The input time was less than 2 min (SD 1 min). In total 22/139 (15.8%) patients presented with a NHR. Twenty-three out of 54 (42%) targeted clinicians participated, and all responded to the survey. In the sample cohort median ISS was 9 [4; 20], with road traffic accidents as predominant mechanism and a higher amount of penetrating trauma compared to the derivation sample (21% versus 11%). Participants indicated a global satisfaction of 87% with the tool and 91% of interest in the study and 91% found the smartphone application ergonomic and user-friendly. Figure 5 illustrates participant appreciation of the protocol feasibility and app ergonomics.

In the small sample of the pilot study, the F-4 score for the XGBoost model was 0.49 (with imputation of missing

data using MIA) against 0.69 for the clinician gestalt (no imputation). This corresponds for the XGBoost model to a sensitivity of 50% versus 64% for the clinician gestalt for the target output NHR, a specificity of 85%, versus 79%, a PPV of 39% versus 37% and a NPV of 90% versus 92%. These results indicate a positive LR of 3.44 and a negative LR of 0.59 for the model and positive LR of 3.1 and negative of 0.46 for the clinicians and corresponding accuracies of 0.8 and 0.77, respectively.

Discussion

The Shockmatrix pilot study demonstrated the feasibility of the development of a prediction model NHR from registry data based exclusively on a limited set of pre-hospital predictors with the capacity to handle missing predictors. This model shows comparable performance to established clinical decision rules and experienced clinicians. Furthermore, the study demonstrated the capacity to implement the model into an easy-to-use

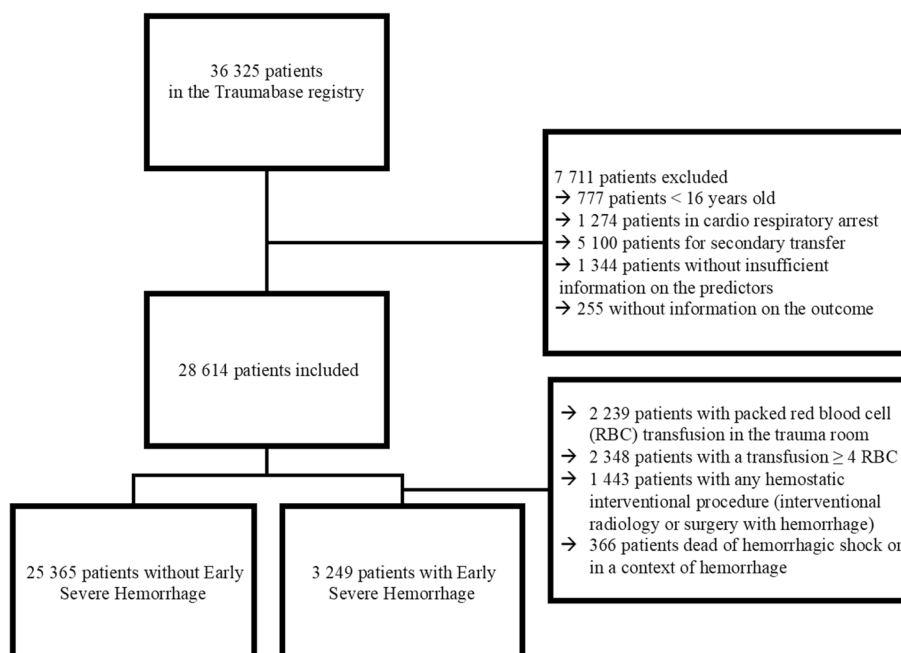


Fig. 1 Study flowchart

smartphone-based tool to capture the necessary predictors and make the prediction model available to clinicians without workflow interruption.

Two recent reviews provide an extensive overview of algorithm-based prediction of haemorrhage and trauma outcomes [10, 11]. These models can be classified according to the outcome, predictors, data source and methods applied. These models frequently predict mortality, transfusion, perform either risk quantification or detect hemorrhage and coagulopathy. Specific patient needs other than transfusion occur rarely as target and most model outputs are not actionable.

The methods cover a large spectrum from regression (e.g. stepwise, Poisson, logistic, Cox) over neural, artificial, or deep networks to tree- or kernel-based methods. No method seems superior to another regarding predictive performance. Most studies are retrospective and do not test feasibility in a real-life setting. Studies differ in the time points to collect the required predictors, rely on late or intrahospital predictors. In consequence, results become available late in the patient pathway to become actionable. Reports do not follow a stringent reporting structure and use different metrics. A combination of systematic reporting of a complete panel of indicators from sensitivity to $F4$ - and $F\beta$ -Score seems appropriate [9, 11, 14], including the specification of diagnostic thresholds. Imputation is not always performed or not mentioned systematically. Numerous imputation techniques are available, their explanation is beyond the scope of this

paper. The fact that an information is missing represents itself an information, since information is rarely missing at random. For this reason, the input of some crucial missing information has been provided in the SHAP diagram in Fig. 2.

The following examples illustrates these observations. Maurer et al. developed a smartphone-based predictor tool, including age, systolic blood pressure, respiratory rate, mechanism, temperature, SpO_2 , comorbidities, GCS and AIS, based on the large reliable Trauma Improvement Quality database [27]. The predictive performance achieves an AUROC of 0.92 for penetrating and 0.83 for blunt trauma. Yet the model predicts mortality, a non-actionable outcome and relies on AIS data, that require a complete body-scan, usually obtained after the resuscitation phase. The need for complete injury description also applies Lee et al. and Nederpelt et al. and Follin et al. [28–30]. Despite excellent performance, the need for detailed injury data makes these models less applicable. Liu et al. developed a model based primarily on prehospital data, including heart rate variability to predict life-saving interventions (transfusion, angioembolization, intubation, thoracotomy, needle decompression...) with variable performance between AUROC 0.9 and correlation coefficient of 0.77 [31]. Yet, the predicted outcome is not specific enough to allow targeted anticipation and preparation. Perkins et al. demonstrated a very elegant tool based on an innovative Bayes network [32]. The model shows excellent diagnostic performance for

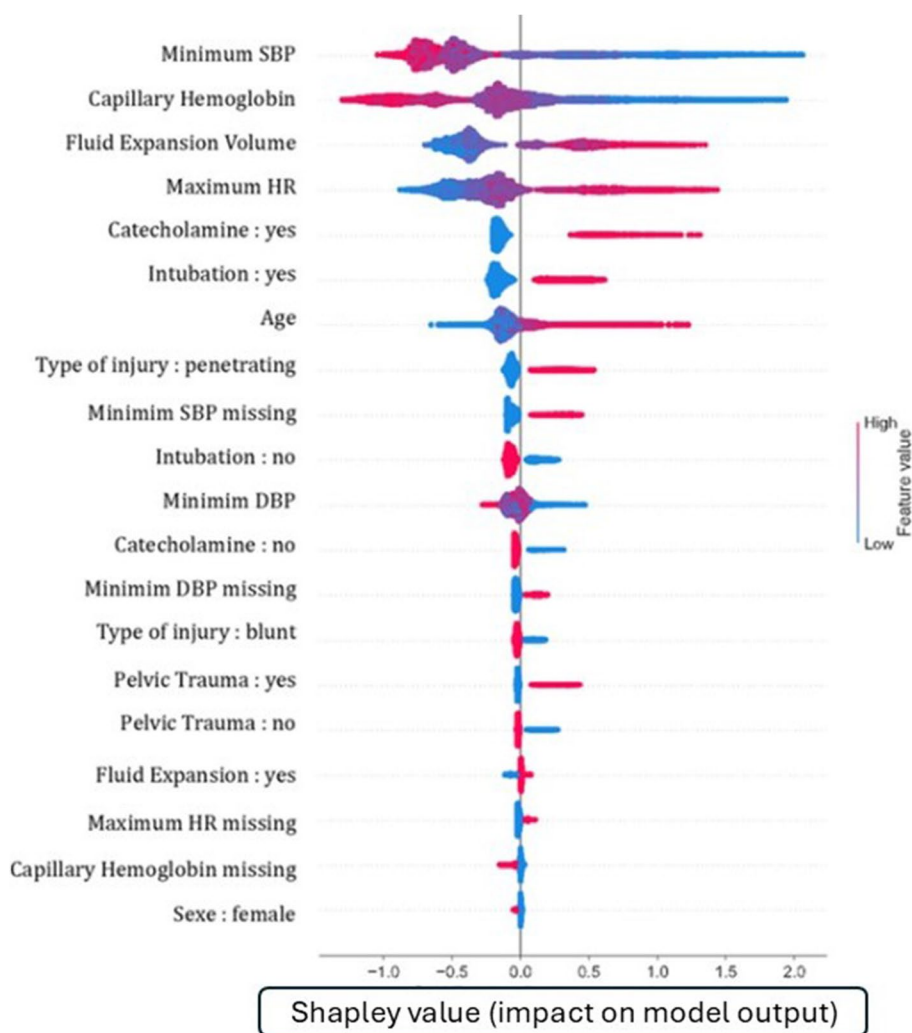


Fig. 2 SHAP Diagram (XGBoost model). The SHAP value is used to report the weight of the variables in the model. Variables are ranked from the most important at the top (minimal SBP) to the less important at the bottom (sex). Variable influence on hemorrhagic shock prediction is represented from the right (positive) to left (negative) and the value of the observation is colored from red (highest for continuous variables and "yes" pour binomial variables) to blue (lowest for continuous variables and "No" pour binomial variables). As an example, "The minimal SBP is the better predictor of hemorrhagic shock (on the top) and the lower minimal SBP value (in blue) have the most positive impact on hemorrhagic shock prediction (right)"

coagulopathy and need for critical resource, but requires an extensive dataset, available only after resuscitation room work up, which provides limited use before the admission to the hospital.

In this context, the Shockmatrix model compares favorably. The model relies exclusively on prehospital routine variables, avoids dichotomization of continuous variables, such as Shock Index and clinical practice patterns as potential source of bias. The diagnostic performance is comparable and acceptable with an AUC of 0.89 for the XGBoost, including the iteration with imputation. The investigators considered the capacity of the model to impute missing data and still perform adequately, since

missing features are a reality in any clinical setting. The model strikes an equilibrium between diagnostic performance and a minimal set of easily available exclusive prehospital predictors to predict a set of composite actionable outcomes [33]. These features provide the model an operational character based on an actionable output that anticipates patient need and resource mobilization. Numerous prehospital.

Few studies attempt a prospective workflow integration and real-life validation of their existing tools to assess usability, ergonomics, explicability, and human-machine and clinician uptake and compliance [14]. To close this gap, this pilot study represented a pivotal step towards

Table 2 Model performances metrics with 95% confidence intervals and a diagnostic threshold of 0,11

	F4-Score	Sensitivity	Precision	Specificity	Accuracy	AUC PR	AUC ROC	LR+	LR-
Red Flag (1337 patients with missing values deleted 38 in NHR)	0.78 [0.77–0.80]	0.97 [0.95–0.98]	0.19 [0.18–0.21]	0.41 [0.39–0.42]	0.48 [0.46–0.49]	0.19 [0.18–0.21]	0.69 [0.68–0.70]	1.63 [1.58–1.67]	0.08 [0.05–0.12]
Red Flag with imputation	0.76 [0.74–0.78]	0.93 [0.91–0.95]	0.19 [0.18–0.21]	0.51 [0.49–0.52]	0.56 [0.54–0.57]	0.19 [0.18–0.20]	0.72 [0.71–0.73]	1.88 [1.82–1.95]	0.14 [0.10–0.18]
ABC (930 patients with missing values deleted 82 in NHR)	0.70 [0.67–0.73]	0.79 [0.75–0.82]	0.25 [0.23–0.27]	0.68 [0.67–0.70]	0.69 [0.68–0.71]	0.22 [0.20–0.24]	0.73 [0.72–0.75]	2.48 [2.33–2.63]	0.31 [0.26–0.36]
ABC with imputation	0.65 [0.62–0.68]	0.72 [0.69–0.76]	0.25 [0.23–0.27]	0.72 [0.71–0.73]	0.72 [0.71–0.73]	0.21 [0.19–0.23]	0.72 [0.70–0.74]	2.55 [2.39–2.73]	0.39 [0.34–0.44]
CART	0.72 [0.69–0.74]	0.82 [0.79–0.85]	0.24 [0.22–0.25]	0.66 [0.65–0.68]	0.68 [0.67–0.69]	0.40 [0.36–0.44]	0.81 [0.79–0.83]	2.43 [2.31–2.56]	0.27 [0.23–0.32]
Random Forest	0.73 [0.71–0.76]	0.79 [0.76–0.82]	0.34 [0.32–0.37]	0.81 [0.80–0.82]	0.80 [0.79–0.81]	0.55 [0.51–0.59]	0.88 [0.86–0.89]	4.08 [3.81–4.36]	0.26 [0.22–0.30]
XGBoost	0.76 [0.73–0.79]	0.82 [0.79–0.85]	0.35 [0.32–0.37]	0.80 [0.79–0.81]	0.80 [0.79–0.81]	0.56 [0.52–0.60]	0.89 [0.87–0.90]	4.12 [3.89–4.40]	0.22 [0.19–0.26]
CatBoost	0.75 [0.72–0.77]	0.81 [0.78–0.84]	0.34 [0.32–0.36]	0.80 [0.79–0.81]	0.80 [0.79–0.81]	0.51 [0.47–0.56]	0.88 [0.86–0.89]	4.00 [3.77–4.26]	0.24 [0.20–0.28]

Red Flag Score: Shock Index > 1, MAP < 70 mmHg, point of care haemoglobin ≤ 13 g/dl, unstable pelvis, pre-hospital intubation

ABC Score: penetrating trauma, heart rate > 120 b/min, systolic blood pressure < 90 mmHg, positive abdominal FAST

AUC Area under the curve, ROC Receiver operating curve, LR Likelihood ratio

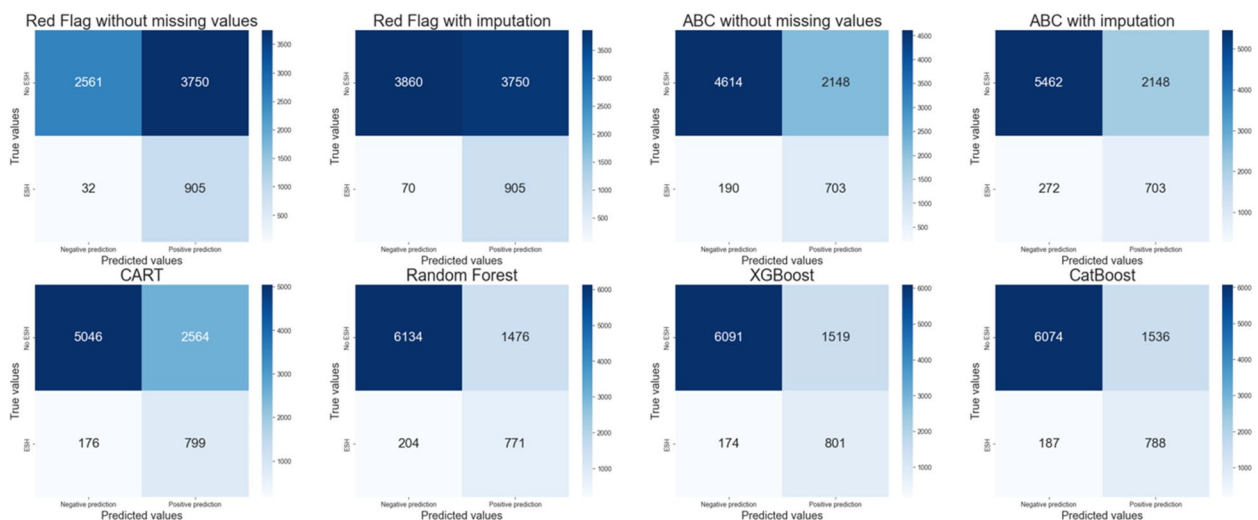


Fig. 3 Confusion matrices

the establishment of an operational Clinical Decision Support Tool (CDST). It demonstrated the feasibility to create a tool to collect the predictor data quickly and feed these into the model. Clinician compliance, satisfaction and interest was high, suggesting potential future uptake and tool compliance. The predictors selected (Shapley value diagram) remain in the realms of a clinical and physiological representation which is important for

clinician confidence and involvement. Despite not being the declared objective of the study, the performance of the model in comparison to clinician gestalt in the pilot phase underscored the need to retrain and improve the initial model; a step rarely performed in the current literature. Retraining resulted in the inclusion of the trauma mechanism “penetrating”; improved the model, and integrated a calibration method. A prospective study will be

Table 3 Cumulated distribution of probabilities over the negative class predicted by the XGBoost model on the validation set

Probability of ESH predicted by the XGBoost model	Cumulated count of True Negative	Cumulated percentage of True Negative	Cumulated count of False Negative	Cumulated percentage of False Negative
0%	231	5.3%	1	0.7%
1%	1 954	45.2%	13	9.2%
2%	2 653	61.4%	27	19.0%
3%	3 126	72.3%	45	31.7%
4%	3 431	79.4%	55	38.7%
5%	3 623	83.8%	71	50.0%
6%	3 799	87.9%	86	60.6%
7%	3 934	91.0%	99	69.7%
8%	4 046	93.6%	108	76.1%
9%	4 149	96.0%	120	84.5%
10%	4 239	98.0%	129	90.8%
11%	4 324	100%	142	100%

Adjusting the threshold for the positive cases from 11 to 5% reduces the number of False Negative predictions by half while keeping 84% of the True Negative predictions

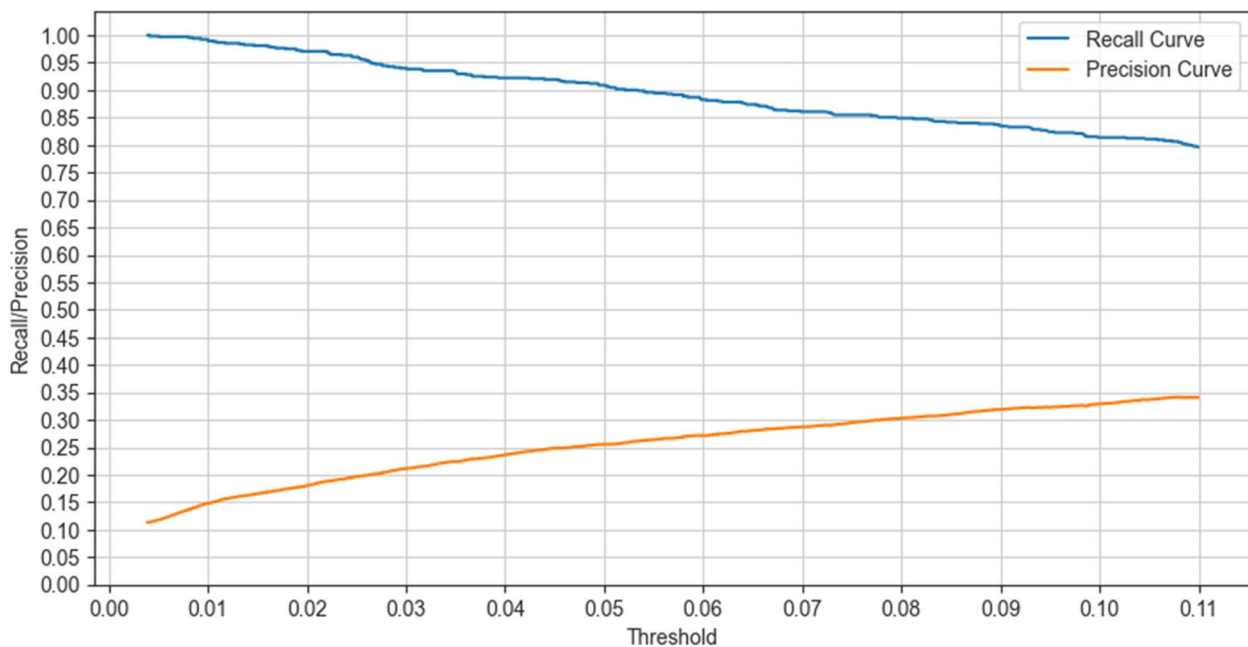


Fig. 4 Recall / Precision curve over threshold of the XGBoost model on the validation set. Adjusting the threshold for the positive cases from 11 to 5% improves the Sensitivity (Recall) to 0.90 while keeping the Precision over 0.25

completed in June 2024 (ClinicalTrials.gov Identifier: NCT06270615) in seven French centers to assess the predictive performance of the model and compare against clinician gestalt using the same information by a real-life prealert with clinicians blinded to the prediction. A randomized cluster trial is in preparation with a deployment of three machine learning algorithms in 16 dispatch centers across France with a planned start in 2025. The algorithms will predict need for hemorrhage control and

resuscitation, need for neurosurgery and need for trauma intervention.

Limitations

Limitations of the study relate to the development without validation in an independent dataset. The investigators share the belief that prospective validation in a real-life workflow with real patient data and assessment of patient impact is crucial. For this reason, a prospective

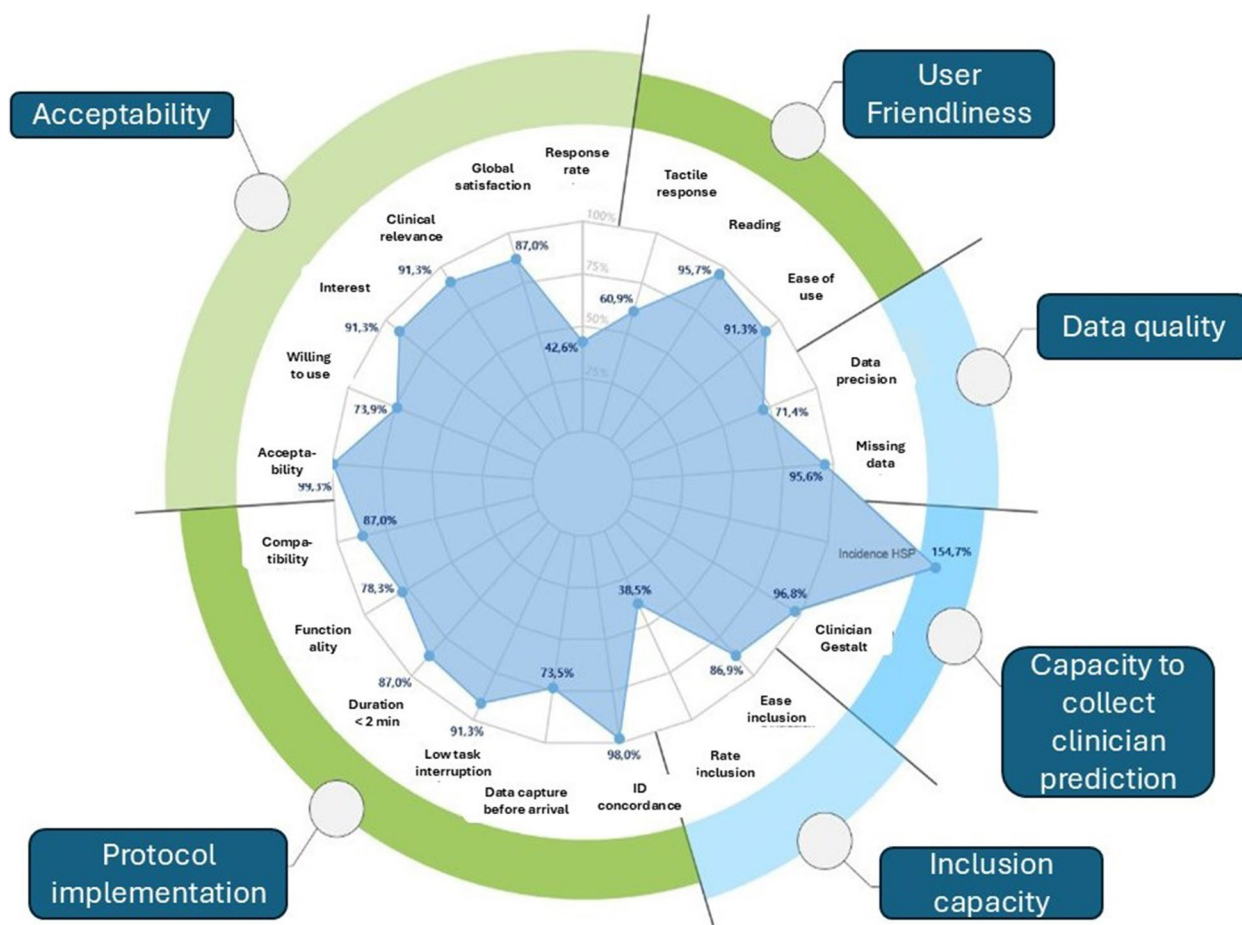


Fig. 5 App assessment by participating clinicians in six dimensions, acceptability, ergonomics, data quality, protocol implementation, sample size, recruitment

real-life validation study is underway in seven centers comparing the performance of the retrained model with the prediction by clinicians. Some predictors such as catecholamine and capillary hemoglobin use are more infrequent in some systems; a prospective validation will assess the external validity. The model did not outperform clinicians in the pilot study. First this pilot was not powered to assess this question. Second, the sample contained an overrepresentation of penetrating trauma linked to the case load in one center, an observation that allowed retraining of the model. Third, the investigators feel participating clinicians were eager to demonstrate their performance compared to the model, equivalent to a Hawthorne effect. The observed performance might thus overestimate a “real-life” performance. Confrontation with the algorithm output might reduce decisional uncertainty. Despite lack of power and a considerable selection bias, the model performed as well as the clinicians in this pilot study. The ML tool could become an objective reference against which the clinician could

benchmark his own appreciation. Finally, the investigators acknowledge that only 45% of eligible patients were included and a 42% response rate among the targeted clinicians may appear low; a selection bias cannot be excluded. Both rates are high compared to other pilot studies. In consequence, the pilot study helped to devise steps to increase clinician involvement and participation to reduce selection bias.

Conclusion

The ShockMatrix pilot study bridged the gap between model development and prospective field test to explore clinician AI interaction. The study developed a robust machine-learning prediction capable of imputing missing data based exclusively on routine prehospital variables to predict need for haemorrhage resuscitation. In a second step, the investigators evaluated prospectively the ergonomics and usability of a smartphone-based application to collect the information required to feed the prediction tool in a real-life setting and generate predictions with the algorithm.

Glossary

AUC	Area under the Curve
CatBoost	open data framework providing a gradient boosting framework which among other features attempts to solve for Categorical features using a permutation driven alternative compared to the classical algorithm
F4-Score	$F4 - Score : F\beta = \left(1 + \beta^2\right) \frac{Pre \cdot Se}{(\beta^2 \cdot Pre) + Se}$ with $\beta = 4$
ML	Machine Learning
NHR	Need for Haemorrhage Resuscitation: 1) administration of at least one packed red blood cell (RBC) in the resuscitation room or 2) transfusion of 4 or more RBC Cells within 6 hours since admission or 3) the need for a hemorrhage control procedure (interventional radiology and/or surgical control in the operating room) within 6 hours of admission or 4) death resulting from hemorrhagic shock within 24 hours
Oversampling and Undersampling	techniques used to adjust the class distribution of a data set
Precisions-Recall Curve	A precision-recall curve (or PR Curve) is a plot of the precision (y-axis) and the recall (x-axis) for different probability thresholds
RBC	Red Blood Cell Concentrate
SHAP framework	stands for SHapley Additive exPlanations — a way to express explainability in Machine Learning. SHAP values are used when a complex model (could be a gradient boosting, a neural network, or anything that takes some features as input and produces some predictions as output) and you want to understand what decisions the model is making
XG Boost	open-source software library which provides a regularizing gradient boosting framework

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-024-02723-9>.

Supplementary Material 1.

The Traumabase Group

¹⁵ Clichy-APHP—Beaujon, Clichy, France	JEANTRELLE	Caroline
⁷ Le Kremlin-Bicêtre APHP—Bicêtre, Le Kremlin Bicêtre, France	WERNER HARROIS	Marie Anatole
¹³ Paris-APHP—Pitié Salpêtrière, Paris, France	RAUX	Mathieu
¹⁶ Créteil-APHP—Henri Mondor, Créteil, France	PASQUERON QUESNEL	Jean Christophe
⁶ Paris-APHP—HEGP, Paris, France	DELHAYE GODIER	Nathalie Anne
¹⁷ Clamart-HIA Percy, Clamart, France	BOUTONNET	Mathieu
¹⁸ Lille-CHRU de Lille—Roger Salengro, Lille, France	GARRIGUE BOURGEOIS BIJOK	Delphine Alexandre Benjamin
¹⁹ Strasbourg-CHRU de Strasbourg—Hautepierre, Strasbourg, France	POTTECHER MEYER	Julien Alain
¹ CHU Grenoble Alpes, Grenoble, France	GAUSS BANCO	Tobias Pierluigi

²⁰ Toulon-HIA Sainte Anne, Toulon, France	MONTALESCAU MEAUDRE	Etienne Eric
³ Caen-CHU de Caen—Côte de Nacre, Caen, France	HANOUS LEFRANCOIS	Jean-Luc Valentin
²¹ Nancy-CHU de Nancy—Central, Nancy, France	AUDIBERT	Gérard
²² Marseille-APHM—Nord, Marseille, France	LEONE HAMMAD	Marc Emmanuelle
²³ CHU Reims, Reims, France	DUCLOS FLOCH	Gary Thierry
⁵ CHU Toulouse—Purpan, Toulouse, France	GEERAERTS	Thomas
⁵ CHU Toulouse—Rangueil, Toulouse, France	BOUNES	Fanny
²⁴ CHU Clermont Ferrand, Clermont Ferrand, France	BOUILLON RIEU	Jean Baptiste Benjamin
²⁵ CHR Metz—Thionville, France	GETTES MELLATI	Sébastien Nouchan
²⁶ Hôpitaux Civils de Colmar, Colmar, France	DUSSAU GAERTNER	Leslie Elisabeth
²⁷ CHU Rouen, Rouen, France	POPOFF CLAVIER LEPÊTRE	Benjamin Thomas Perrine
²⁸ CHU Bordeaux, Bordeaux, France	SCOTTO	Marion
⁵ CHU Toulouse—Purpan URGENCE, Toulouse, France	ROTIVAL	Julie
²⁹ CH Valenciennes, Valenciennes, France	MALEC JAILLETTE	Loan Claire
³⁰ Amiens- CHU D'Amiens Sud, Amiens, France	GOSSET	Pierre
³¹ Dunkerque—CH De Dunkerque, France	COLLARD	Clément
³² Centre Hospitalier de Cayenne, France	PUJO KALLEL FREMERY HIGEL	Jean Hatem Alexis Nicolas
³³ Centre Hospitalier Universitaire DIJON—BOURGOGNE, Dijon, France	WILLIG	Mathieu
³⁴ Centre hospitalier de Tours, Tours, France	COHEN ABBACK	Benjamin Paer Selim
³⁵ CHU Annecy, Annecy, France	GAY ESCUDIER MERMILLOD BLONDIN	Samuel Etienne Romain

Authors' contributions

Design, data acquisition, analysis, writing of manuscript: TG. Design, conception, data analysis, methodological supervision, writing of manuscript: JJ. Data analysis, model construction, writing of manuscript: CC, MP, SM. Design, data acquisition, analysis, critical review: JDM, AH, VR, ND, AJ, TS, MW.

Funding

The study received no specific funding.

Data availability

All data and scripts are available upon request to the lead author tgauss@chu-grenoble.fr.

Declarations

Ethics approval and consent to participate

The study obtained approval from the University Paris Nord Ethics committee (CER-2021–106 /project SHOCKMATRIX, 11th February 2022) and waived the need for informed consent. The Traumabase registry has obtained approval from the Institutional Review Board (Comité de Protection des Personnes, Paris VI) from the Advisory Committee for Information Processing in Health Research (Comite Consultatif Pour le Traitement de l'information en matière de recherche dans le domaine de la santé CCTIRS, 11.305bis) and from the National Data Protection Agency (Commission Nationale de l'Informatique et des Libertés CNIL, 911461).

Consent for publication

Not applicable.

Competing interests

TG reports honoraria from Laboratoire du Biomédicament Français. JDM reports honoraria from Octapharma. MW reports honoraria from Edwards. AH reports honoraria from Laboratoire du Biomédicament Français, Edwards and Octapharma. VR reports honoraria from Pfizer.

Author details

¹Service Anesthésie-Réanimation, CHU Grenoble Alpes, Grenoble, France. ²Université Grenoble Alpes, Inserm, Grenoble Institute Neurosciences, Grenoble U1216, France. ³Service Anesthésie-Réanimation, CHU Caen, Caen, France. ⁴Cap Gemini Invent, Issy-Les-Moulineaux, Paris, France. ⁵Service Anesthésie-Réanimation, CHU Toulouse, Toulouse III – Université Paul Sabatier, Toulouse, France. ⁶Service Anesthésie-Réanimation, Hôpital Européen Georges Pompidou, AP-HP Paris, France. ⁷Service d'Anesthésie Réanimation Chirurgicale, DMU 12 Anesthésie Réanimation Chirurgicale Médecine Péri-Opératoire et Douleur Hôpital Bicêtre, AP-HP, Université Paris-Saclay, Le Kremlin-Bicêtre, Paris, France. ⁸Équipe DYNAMIC, Inserm UMR_S999, Le Kremlin-Bicêtre, Paris, France. ⁹Pôle Anesthésie, Service de Réanimation Polyvalente URM Purpan, CHU Toulouse, Médecine Péri-Opératoire, Toulouse, France. ¹⁰Institut National de Recherche en Sciences Et Technologies du Numérique, Premedical Team, Université de Montpellier, Montpellier, France. ¹¹DMU DREAM, Service Anesthésie-Réanimation, Hôpital Pitié-Salpêtrière, Sorbonne Université, GRC 29, AP-HP, Paris, France. ¹²The Traumabase Group, Hôpital Beaujon, AP-HP, Clichy, France.

Received: 4 February 2024 Accepted: 14 October 2024

Published online: 28 October 2024

References

1. Wohlgenut JM, Kyrimi E, Stoner RS, Pisirir E, Marsh W, Perkins ZB, et al. The outcome of a prediction algorithm should be a true patient state rather than an available surrogate. *J Vasc Surg.* 2022;75:1495–6. <https://doi.org/10.1016/j.jvas.2021.10.059>.
2. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16. <https://doi.org/10.3402/meo.v16i0.5890>.
3. Rice TW, Morris S, Tortella BJ, Wheeler AP, Christensen MC. Deviations from evidence-based clinical management guidelines increase mortality in critically injured trauma patients*. *Crit Care Med.* 2012;40:778–86. <https://doi.org/10.1097/CCM.0b013e318236f168>.
4. Lang E, Neuschwander A, Favé G, Abback P-S, Esnault P, Geeraerts T, et al. Clinical decision support for severe trauma patients: Machine learning based definition of a bundle of care for hemorrhagic shock and traumatic brain injury. *J Trauma Acute Care Surg.* 2022;92:135–43. <https://doi.org/10.1097/TA.0000000000003401>.
5. van Maarseveen OEC, Ham WHW, van de Ven NLM, Saris TFF, Leenen LPH. Effects of the application of a checklist during trauma resuscitations on ATLS adherence, team performance, and patient-related outcomes: a systematic review. *Eur J Trauma Emerg Surg.* 2020;46:65–72. <https://doi.org/10.1007/s00068-019-01181-7>.
6. Mercer SJ, Kingston EV, Jones CPL. The trauma call. *BMJ* 2018;k2272. <https://doi.org/10.1136/bmj.k2272>.
7. Gauss T, Quintard H, Bijok B, Bouhours G, Clavier T, Cook F, et al. Intra-hospital Trauma Flowcharts - cognitive aids for intrahospital trauma management from the French Society of Anaesthesia and Intensive Care Medicine and the French Society of Emergency Medicine. *Anaesthesia Critical Care & Pain Medicine* 2022:101069. <https://doi.org/10.1016/j.accpm.2022.101069>.
8. Fitzgerald M, Cameron P, Mackenzie C, Farrow N, Scicluna P, Gocentas R, et al. Trauma resuscitation errors and computer-assisted decision support. *Arch Surg.* 2011;146:218–25. <https://doi.org/10.1001/archsurg.2010.333>.
9. Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. *Shock.* 2017;48:504–10. <https://doi.org/10.1097/SHK.0000000000000898>.
10. Hunter OF, Perry F, Salehi M, Bandurski H, Hubbard A, Ball CG, et al. Science fiction or clinical reality: a review of the applications of artificial intelligence along the continuum of trauma care. *World J Emerg Surg.* 2023;18:16. <https://doi.org/10.1186/s13017-022-00469-1>.
11. Peng HT, Siddiqui MM, Rhind SG, Zhang J, Teodoro da Luz L, Beckett A. Artificial intelligence and machine learning for hemorrhagic trauma care. *Mil Med Res* 2023;10:6. <https://doi.org/10.1186/s40779-023-00444-0>.
12. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med.* 2021;47:750–60. <https://doi.org/10.1007/s00134-021-06446-7>.
13. Gauss T, Perkins Z, Tjardes T. Current knowledge and availability of machine learning across the spectrum of trauma science. *Curr Opin Crit Care.* 2023;29:713–21. <https://doi.org/10.1097/MCC.0000000000001104>.
14. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ.* 2022;377:e070904. <https://doi.org/10.1136/bmj-2022-070904>.
15. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med.* 2015;162:W1–73. <https://doi.org/10.7326/M14-0698>.
16. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open.* 2021;11:e048008. <https://doi.org/10.1136/bmjopen-2020-048008>.
17. Hamada SR, Gauss T, Duchateau F-X, Truchot J, Harrois A, Raux M, et al. Evaluation of the performance of French physician-staffed emergency medical service in the triage of major trauma patients. *J Trauma Acute Care Surg.* 2014;76:1476–83. <https://doi.org/10.1097/TA.0000000000000239>.
18. Gauss T, Ageron F-X, Devaud M-L, Debaty G, Travers S, Garrigue D, et al. Association of Prehospital Time to In-Hospital Trauma Mortality in a Physician-Staffed Emergency Medicine System. *JAMA Surg.* 2019;154:1117–24. <https://doi.org/10.1001/jamasurg.2019.3475>.
19. Gauss T, Richards JE, Tortù C, Ageron F-X, Hamada S, Josse J, et al. Association of early norepinephrine administration with 24-hour mortality among patients with blunt trauma and hemorrhagic shock. *JAMA Netw Open.* 2022;5:e2234258. <https://doi.org/10.1001/jamanetworkopen.2022.34258>.
20. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. Explainable AI for Trees: From Local Explanations to Global Understanding 2019. <https://doi.org/10.48550/ARXIV.1905.04610>.
21. Josse J, Prost N, Scornet E, Varoquaux G. On the consistency of supervised learning with missing values 2019. <https://doi.org/10.48550/ARXIV.1902.06931>.
22. Josse J, Reiter JP. Introduction to the Special Section on Missing Data. *Statist Sci.* 2018;33:139–41. <https://doi.org/10.1214/18-STS332IN>.
23. Pozzolo AD, Caelen O, Johnson RA, Calibrating BG. Probability with Undersampling for Unbalanced Classification. *IEEE Symposium Series on Computational Intelligence.* Cape Town, South Africa: IEEE. 2015;2015:159–66. <https://doi.org/10.1109/SSCI.2015.33>.
24. Hamada SR, Rosa A, Gauss T, Desclèfs J-P, Raux M, Harrois A, et al. Development and validation of a pre-hospital "Red Flag" alert for activation of intra-hospital haemorrhage control response in blunt trauma. *Crit Care.* 2018;22:113. <https://doi.org/10.1186/s13054-018-2026-9>.
25. Nunez TC, Voskresensky IV, Dossett LA, Shinnall 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. <https://doi.org/10.1097/TA.0b013e3181961c35>.
26. Rousson V, Zumbo T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak*. 2011;11:45. <https://doi.org/10.1186/1472-6947-11-45>.
 27. Maurer LR, Bertsimas D, Bouardi HT, El Hechi M, El Moheb M, Giannoutsou K, et al. Trauma outcome predictor: An artificial intelligence interactive smartphone tool to predict outcomes in trauma patients. *J Trauma Acute Care Surg*. 2021;91:93–9. <https://doi.org/10.1097/TA.0000000000003158>.
 28. Lee K-C, Lin T-C, Chiang H-F, Horng G-J, Hsu C-C, Wu N-C, et al. Predicting outcomes after trauma: Prognostic model development based on admission features through machine learning. *Medicine (Baltimore)*. 2021;100:e27753. <https://doi.org/10.1097/MD.00000000000027753>.
 29. Nederpelt CJ, Mokhtari AK, Alser O, Tsiligkaridis T, Roberts J, Cha M, et al. Development of a field artificial intelligence triage tool: Confidence in the prediction of shock, transfusion, and definitive surgical therapy in patients with truncal gunshot wounds. *J Trauma Acute Care Surg*. 2021;90:1054–60. <https://doi.org/10.1097/TA.0000000000003155>.
 30. Follin A, Jacquemin S, Chhor V, Bellenfant F, Robin S, Guinvarc'h A, et al. Tree-based algorithm for prehospital triage of polytrauma patients. *Injury*. 2016;47:1555–61. <https://doi.org/10.1016/j.injury.2016.04.024>.
 31. Liu NT, Holcomb JB, Wade CE, Batchinsky AI, Cancio LC, Darrah MI, et al. Development and validation of a machine learning algorithm and hybrid system to predict the need for life-saving interventions in trauma patients. *Med Biol Eng Comput*. 2014;52:193–203. <https://doi.org/10.1007/s11517-013-1130-x>.
 32. Perkins ZB, Yet B, Marsden M, Glasgow S, Marsh W, Davenport R, et al. Early Identification of Trauma-induced Coagulopathy: Development and Validation of a Multivariable Risk Prediction Model. *Ann Surg*. 2021;274:e1119–28. <https://doi.org/10.1097/SLA.0000000000003771>.
 33. James A, Abback P-S, Pasquier P, Ausset S, Duranteau J, Hoffmann C, et al. The conundrum of the definition of haemorrhagic shock: a pragmatic exploration based on a scoping review, experts' survey and a cohort analysis. *Eur J Trauma Emerg Surg*. 2022;48:4639–49. <https://doi.org/10.1007/s00068-022-01998-9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.