

RESEARCH

Open Access



The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study

Anna Larsson¹, Johanna Berg^{2,3}, Mikael Gellerfors^{4,5,6,7} and Martin Gerdin Wårnberg^{3,5*}

Abstract

Background: Accurate prehospital trauma triage is crucial for identifying critically injured patients and determining the level of care. In the prehospital setting, time and data are often scarce, limiting the complexity of triage models. The aim of this study was to assess whether, compared with logistic regression, the advanced machine learner XGBoost (*eXtreme Gradient Boosting*) is associated with reduced prehospital trauma mistriage.

Methods: We conducted a simulation study based on data from the US National Trauma Data Bank (NTDB) and the Swedish Trauma Registry (SweTrau). We used categorized systolic blood pressure, respiratory rate, Glasgow Coma Scale and age as our predictors. The outcome was the difference in under- and overtriage rates between the models for different training dataset sizes.

Results: We used data from 813,567 patients in the NTDB and 30,577 patients in SweTrau. In SweTrau, the smallest training set of 10 events per free parameter was sufficient for model development. XGBoost achieved undertriage rates in the range of 0.314–0.324 with corresponding overtriage rates of 0.319–0.322. Logistic regression achieved undertriage rates ranging from 0.312 to 0.321 with associated overtriage rates ranging from 0.321 to 0.323. In NTDB, XGBoost required the largest training set size of 1000 events per free parameter to achieve robust results, whereas logistic regression achieved stable performance from a training set size of 25 events per free parameter. For the training set size of 1000 events per free parameter, XGBoost obtained an undertriage rate of 0.406 with an overtriage of 0.463. For logistic regression, the corresponding undertriage was 0.395 with an overtriage of 0.468.

Conclusion: The under- and overtriage rates associated with the advanced machine learner XGBoost were similar to the rates associated with logistic regression regardless of sample size, but XGBoost required larger training sets to obtain robust results. We do not recommend using XGBoost over logistic regression in this context when predictors are few and categorical.

Keywords: Trauma, Prehospital triage, Undertriage, Overtriage, Clinical prediction model, Machine learning

Background

Accurate prehospital trauma triage is crucial for identifying critically injured patients and determining where to transport these patients [1]. In this context, undertriage (false negative) is when a patient requiring specialized trauma care is transferred to a lower-level trauma centre

*Correspondence: martin.gerdin@ki.se

³ Department of Global Public Health, Karolinska Institutet, 171 77 Solna, Sweden

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



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and is associated with higher mortality rates [1, 2]. Conversely, overtriage (false positive) occurs when a patient not in need of specialized trauma care is transferred to a higher-level trauma centre, resulting in extra costs and overutilization of resources [3].

The American College of Surgeons Committee on Trauma (ACS-COT) guidelines state that trauma systems must aim to attain a maximum of 5% undertriage and keep overtriage below 35% [4]. Extensive research has focused on the development of prediction models to assist emergency medical service (EMS) personnel in the early clinical decision making and triage process [5]. However, most existing trauma triage protocols perform poorly, leading to high rates of mistriage [6–8].

Prehospital trauma triage is particularly challenging due to a lack of time and patient information [9]. Previous research shows that prediction models incorporating several categories of predictors (physiological, demographic, anatomical, etc.) generally perform better than simpler models [5]. However, in the prehospital setting, patient information is often scarce, and health care providers need to prioritize stabilization and rapid transport over a thorough medical history [10, 11].

Extensive research has focused on how machine learning can improve predictions and diagnosis in medicine [12, 13]. Studies show that, given comprehensive input parameters, machine learning may outperform classical statistical methods in predicting the need for critical care and hospitalization in the emergency department and prehospital setting [14–17]. Other research indicates that logistic regression may perform as well as machine learners when used to develop clinical prediction models [18–20].

Little is known about how much data are needed and how complex the data must be to obtain accurate predictions with machine learners. Predictive accuracy is adversely affected when a model is transferred from one setting to another [21]; therefore, it is desirable to use models that can be developed from available local data. We also know that health care professionals are not willing to use complex systems disrupting their workflow and requiring time away from the patient [22].

We wanted to study whether an advanced machine learner could bring any added value to prehospital trauma triage, given limited input data complexity. Our aim was therefore to assess whether, compared with logistic regression, which is a classical modelling technique commonly used in this context, the advanced machine learner XGBoost (*eXtreme Gradient Boosting*) is associated with reduced prehospital trauma mistriage. To estimate how much data were needed for model development, we assessed the performance of the learners for different training data set sizes. We chose the machine

learner (XGBoost) because it has recently been dominating applied machine learning and Kaggle competitions [23, 24].

Methods

Study design

We conducted a simulation study based on data from the US National Trauma Bank (NTDB) and the Swedish trauma registry (SweTrau). The NTDB cohort included a total of 813,567 patients enrolled in the NTDB in 2014. The SweTrau cohort included a total of 30,577 patients registered in the Swedish trauma registry between 2011 and 2016.

Variables

Outcome

The outcome of the study was the pairwise difference in over- and undertriage rates between the models. We used an ISS > 15 as the gold standard to define trauma severity as major trauma. Patients with an ISS ≤ 15 were considered to have minor trauma [4]. We defined the overtriage rate as the false-positive rate and the undertriage rate as the false-negative rate.

- Overtriage rate (false-positive rate) = Number of patients with an ISS ≤ 15 classified as major trauma / total number of minor trauma patients (ISS ≤ 15)
- Undertriage rate (false-negative rate) = Number of patients with an ISS > 15 classified as minor trauma / total number of major trauma patients (ISS > 15)

By definition: Specificity + False-positive rate = 1, Sensitivity + False-negative rate = 1.

Predictors

The predictors used to build our models were systolic blood pressure, respiratory rate, Glasgow Coma Scale (GCS) and age. Our rationale was that these parameters are known to be predictive of mortality after trauma and are easily collected by EMS personnel in the prehospital setting. We used the first recorded vital parameters at the scene of the injury. The vital parameters were categorized according to the Revised Trauma Score (RTS) [25]. Our rationale for categorizing these variables was that in severe trauma, it can be hard to obtain an exact count of the respiratory rate, GCS or an accurate measure of a very low blood pressure. We included age as a significant risk factor for mortality in trauma, with a significant increase from 57 years of age [26]. Additionally, due to the loss of regulatory and adaptive mechanisms, vital signs respond differently to stressors in elderly individuals [27].

Data

SweTrau is a nationally encompassing registry in which 92% of Swedish hospitals record trauma cases. The inclusion criteria were as follows: traumatic events with subsequent activation of the hospital trauma protocol, admitted patients with an NISS > 15 and patients transferred to the hospital within 7 days of traumatic events with an NISS > 15. SweTrau excludes patients if the only injury is chronic subdural haematoma or if the hospital trauma protocol is activated without traumatic events [28].

The US National Trauma Data Bank (NTDB) is the largest aggregation of U.S. trauma registry data ever assembled. Currently, the NTDB contains detailed data on over six million cases from over 900 registered U.S. trauma centres [29]. The NTDB includes any patient with at least one injury diagnosis code (ICD-9CM 800–959.9), excluding late effects of injury, superficial injuries, and foreign bodies [30]. In addition, patients must be admitted as trauma patients, transferred from another institution or have died as a result of their injury [30].

Eligibility criteria

The inclusion criteria of the present study were age above 15. Observations with unrealistic recordings of SBP > 300 and RR > 67 were excluded. Observations with systolic blood pressure of 0 were excluded due to uncertain underlying reasons ranging from unsuccessful recording to cardiac arrest. Observations with missing values were excluded.

Study size

As real-world data are often scarce, we wanted to estimate how much data are actually needed to develop a reliable model. We therefore performed the study for different training set sizes. The required study size for each phase of model development using logistic regression is fairly well established, i.e., for derivation at least 10 events per free parameter in the potentially most complex model and for validation at least 100 events, assuming that events are more common than non-events. No study size guidelines exist for the ML algorithm XGBoost. We therefore trained the models in training sets of sizes 10, 25, 100 and 1000 events per free parameter.

Model development

We used R for the statistical analyses and development of the models. The following steps were followed for each of the cohorts and training set size. First, a training set was created using a simple random draw from the complete cohort. The size of the training set was NX/P , where N is the number of free parameters, X is the number of events

per free parameter and P is the proportion of events in the cohort. Each of the models was trained in this training set. Second, the validation and test sets were created using simple random sampling from the complete cohort. The size of the validation and test sets was $200/P$, where P is the proportion of events in the cohort. We used the validation set to define the cut-off probability for major and minor trauma. The rationale for using the validation set and not the training set for this was to improve the generalization properties of the models.

The optimal cut-off was identified by performing a gridsearch on the output probabilities of the validation set, evaluating the over- and undertriage rates for every cut-off probability. The gridsearch was performed from 0 to 1 in steps of 0.001. This cut-off was set to aim at an undertriage rate below or equal to 5% with as low overtriage as possible to be in accordance with ACS-COT guidelines [4]. In the absence of undertriage below 5%, we set the cut-off to obtain as low undertriage as possible with an upper limit for overtriage below 50%. This cut-off was then applied to the output probabilities from the test set to compare the performance of the final models.

The exact performance of a model depends on the subsets of data used for development, validation and testing. We therefore repeated the above process 1000 times for every size of training set and cohort. The results are presented as the median and 2.5% and 97.5% percentiles across bootstrap samples.

Logistic regression

A logistic regression model was created using bootstrapping to limit overfitting and optimism according to current guidelines [31]. The following steps were followed. First, the model was fit using all data in the training set, creating an original model M_0 . Then, a bootstrap sample of size N was drawn with replacement from the training set, and the model was fit using this bootstrap sample, resulting in a bootstrap model $M_{bs,i}$. $M_{bs,i}$ was then used to predict the outcomes in the training set. The resulting linear predictor y was regressed on the outcomes Y using a logistic model. The coefficients of the linear predictor, denoted b_i , where i is the number of bootstrap samples, were stored. This procedure was repeated 1000 times, producing $v = b_1, b_2, b_3, \dots, b_{1000}$. \bar{v} was finally used to shrink the coefficients of the original model as $\bar{v} \times M_0$.

Extreme gradient boosting (XGBoost)

XGBoost is an implementation of gradient boosted decision trees [23]. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then combined to make the final prediction.

The XGBoost model was developed using the R package “xgboost”. The hyperparameters *nrounds*, *eta*, *max_depth* and *lambda* were tuned using the R package “MLR” [32]. For the other hyperparameters, we used default values. The hyperparameters were tuned using MLR functions following a well-defined procedure. We defined the searchspace for the parameters to be tuned and performed a random search using fivefold cross-validation. This approach implied that our training and validation data were combined and then split into 5 equally sized parts. The model was trained on four of the five parts and evaluated on the 5th. This process was repeated until each of the parts had been used as the validation set. The hyperparameters giving the best performance were selected for the training of the final XGBoost model. The final XGBoost model was trained using only the training data set.

Performance measures

Model performance was assessed in terms of over- and undertriage rates as defined in Sect. 2.2.1. Additionally, we calculated the sensitivity, specificity, area under the curve (AUC), calibration slope and calibration intercept for the models. The over- and undertriage rates are presented in detail in the results section. AUC and calibration properties are mentioned in the results section and presented in detail as supplementary material. As sensitivity and specificity can be directly calculated from the over- and undertriage rates, we have chosen to only present these measures as supplementary material.

Results

The study was conducted in the NTDB and SweTrau cohorts in parallel (Table 1). The NTDB cohort included a total of 813,567 patients enrolled in the National Trauma Data Bank (NTDB) during 2014. The SweTrau cohort included a total of 30,577 patients registered in the Swedish trauma registry between 2011 and 2016. After excluding observations with missing recordings and applying the inclusion criteria, we were left with 368,810 observations in the NTDB cohort and 16,547 observations in the SweTrau cohort. The proportion of major trauma events in the NTDB was larger than that in SweTrau, probably because it only included patients admitted to the hospital.

Performance of models

In SweTrau, the smallest training set of 10 events per free parameter was sufficient to achieve robust results (Table 2). XGBoost obtained undertriage rates in the range of 0.314–0.324 with corresponding overtriage rates of 0.322–0.319. In SweTrau logistic regression achieved undertriage rates ranging from 0.312 to 0.321, with

overtriage rates ranging from 0.323 to 0.321. The area under the curve (AUC), calibration slope and calibration intercept were calculated and are presented in Additional file 1: Tables S1 and S2. In SweTrau, XGBoost obtained a maximal AUC of 0.725 with a corresponding calibration slope of 1.056 and calibration intercept of 0.009. The best discrimination and calibration properties for logistic regression were an AUC of 0.725 with a calibration slope of 1.01 and calibration intercept of -0.017.

In NTDB, XGBoost required the largest training set size of 1000 events per free parameter to achieve robust results, whereas logistic regression achieved stable results from a training set size of 25 events per free parameter. For the training set size of 1000 events per free parameter, XGBoost achieved an undertriage rate of 0.406 with an associated overtriage rate of 0.463. The corresponding AUC was 0.611 with a calibration slope of 1.097 and calibration intercept of -0.021. For logistic regression, the corresponding undertriage was 0.395 with overtriage of 0.468 with an AUC of 0.614, calibration slope of 0.995 and calibration intercept of -0.015.

Overall, as shown in Fig. 1 and Table 3, the predictive performance was better in SweTrau than in NTDB.

Comparison of model performance

The performance of XGBoost and logistic regression was compared in each of the 1000 runs, i.e., the models were compared when they had been developed on the same set of training data and evaluated on the same set of test data. Table 3 and Fig. 2 show the difference in under- and overtriage rates between the models for these 1000 repetitions (median and 2.5, 97.5 percentiles).

In SweTrau, there were only minimal differences in performance between the models. As seen from Table 3, the median of the difference in performance in SweTrau was 0 for all training set sizes. In NTDB, logistic regression achieved stable results from a training set size of 25 events per free parameter, whereas XGBoost required the largest training set size of 1000 events per free parameter to achieve stable results. Consequently, there were differences in performance for the smaller training sets with a clear advantage for logistic regression. For the largest training data set of 1000 events per free parameter, there were only minimal differences in performance between the models.

Discussion

The aim of this study was to determine whether, compared to logistic regression, the advanced machine learner XGBoost could bring any added value to pre-hospital trauma triage. We found that differences in mistriage rates were generally very small, using few

Table 1 Characteristics of study data and sizes of data sets

Characteristics	NTDB	SweTrau
Total number of observations	813,567	30,577
Number of missing observations	422,416	10,411
Number of included observations	368,810	16,547
Proportion major trauma	0.21	0.12
Proportion female	0.38	0.35
Age (median-IQR)	51 [30, 69]	41 [25, 59]
GCS Category	NTDB proportion of observations	SweTrau proportion of observations
13–15	0.9098	0.9173
9–12	0.0384	0.0401
6–8	0.0175	0.018
4–5	0.0061	0.009
3	0.0283	0.0156
RR Category	NTDB proportion of observations	SweTrau proportion of observations
30–67	0.0191	0.0543
10–29	0.9677	0.9407
6–9	0.0063	0.0036
0–5	0.0021	0.0011
SBP Category	NTDB proportion of observations	SweTrau proportion of observations
90–300	0.9707	0.9798
76–89	0.0192	0.0126
50–75	0.0091	0.0068
1–49	0.0011	0.0007
Size training sets (events per free parameter)	NTDB	SweTrau
10	714	1250
25	1786	3125
100	7143	12,500
1000	71,429	Missing
Size validation and test sets	NTDB	SweTrau
(200/proportion events)	952	1667

NTDB, National Trauma Data Bank; SweTrau, Swedish Trauma Registry; GCS, Glasgow Coma Scale; RR, Respiratory Rate; SBP, Systolic Blood Pressure

categorized predictors and regardless of sample size, but that XGBoost required more data to provide robust estimates.

The time-sensitive nature of trauma makes it difficult for EMS personnel to gather a full medical history and perform a meticulous exam. The prehospital decision to consider the patient seriously injured followed by field triage to a trauma unit is often based on limited information. It is therefore advantageous to develop reliable

models that depend only on easily accessible data such as vital parameters and age.

Both models achieved the best possible overtriage of 32% with the best possible undertriage of 31%. Thus, in SweTrau, the overtriage was in line with the ACS-COT recommendation of maximum 35% overtriage, but the undertriage of 31% was far from the ACS-COT recommendation of maximum 5%. The performance of the models was better in SweTrau than in NTDB. This is

Table 2 Under- and overtriage rates of logistic regression and XGBoost (median and 2.5 and 97.5 percentiles (calculation on 1000 runs))

Events per free parameter	Data set	Undertriage logistic regression	Overtriage logistic regression	Undertriage XGBoost	Overtriage XGBoost
10	SweTrau	0.321 [0.259, 0.389]	0.321 [0.299, 0.344]	0.324 [0.258, 0.683]	0.319 [0.057, 0.344]
10	NTDB	0.429 [0.338, 0.79]	0.453 [0.052, 0.501]	0.701 [0.35, 0.808]	0.08 [0.039, 0.494]
25	SweTrau	0.314 [0.257, 0.379]	0.322 [0.3, 0.346]	0.316 [0.258, 0.61]	0.321 [0.09, 0.345]
25	NTDB	0.405 [0.332, 0.771]	0.46 [0.06, 0.499]	0.436 [0.345, 0.792]	0.444 [0.045, 0.498]
100	SweTrau	0.312 [0.254, 0.373]	0.323 [0.301, 0.345]	0.314 [0.255, 0.4]	0.322 [0.291, 0.345]
100	NTDB	0.394 [0.324, 0.735]	0.466 [0.072, 0.503]	0.409 [0.327, 0.79]	0.459 [0.048, 0.497]
1000	NTDB	0.395 [0.327, 0.72]	0.468 [0.078, 0.507]	0.406 [0.328, 0.777]	0.463 [0.05, 0.504]

NTDB, National Trauma Data Bank; SweTrau, Swedish Trauma Registry

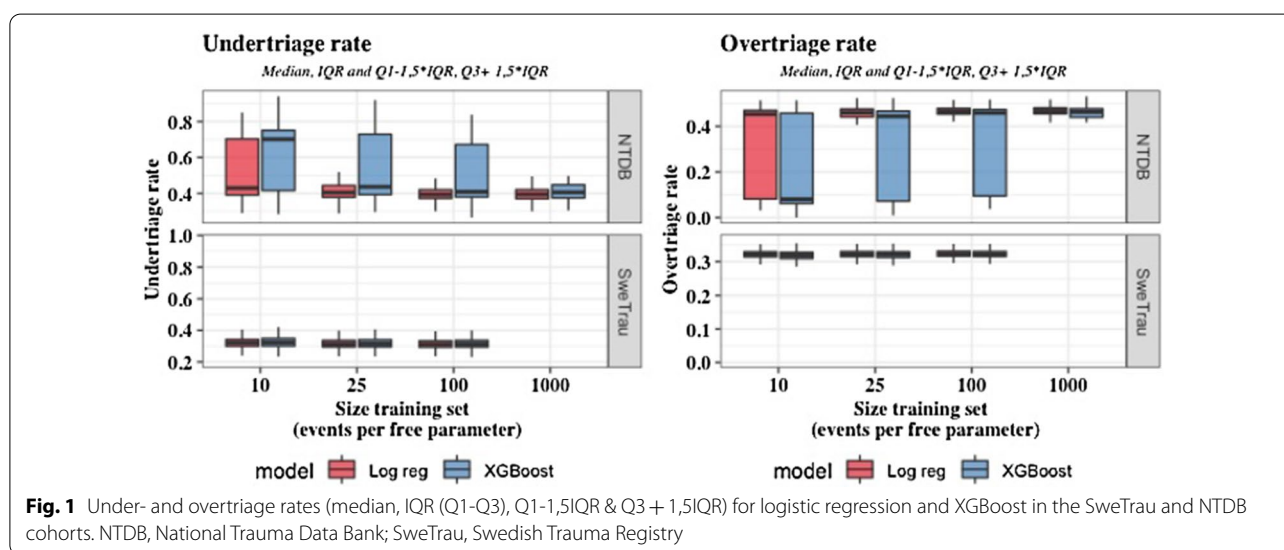


Fig. 1 Under- and overtriage rates (median, IQR (Q1-Q3), Q1-1.5*IQR & Q3+ 1.5*IQR) for logistic regression and XGBoost in the SweTrau and NTDB cohorts. NTDB, National Trauma Data Bank; SweTrau, Swedish Trauma Registry

Table 3 Median and 2.5 and 97.5 percentiles of difference in under- and overtriage rates between learners (calculation on 1000 runs)

Events per free parameter	Data set	Difference in undertriage LogReg-XGBoost	Difference in overtriage LogReg-XGBoost
10	SweTrau	0 [-0.354, 0.025]	0 [-0.005, 0.265]
10	NTDB	-0.005 [-0.398, 0.035]	0.004 [-0.017, 0.433]
25	SweTrau	0 [-0.301, 0.015]	0 [-0.003, 0.238]
25	NTDB	-0.005 [-0.39, 0.023]	0.003 [-0.015, 0.427]
100	SweTrau	0 [-0.025, 0.005]	0 [-0.001, 0.008]
100	NTDB	0 [-0.396, 0.005]	0 [-0.003, 0.427]
1000	NTDB	0 [-0.386, 0]	0 [-0.001, 0.422]

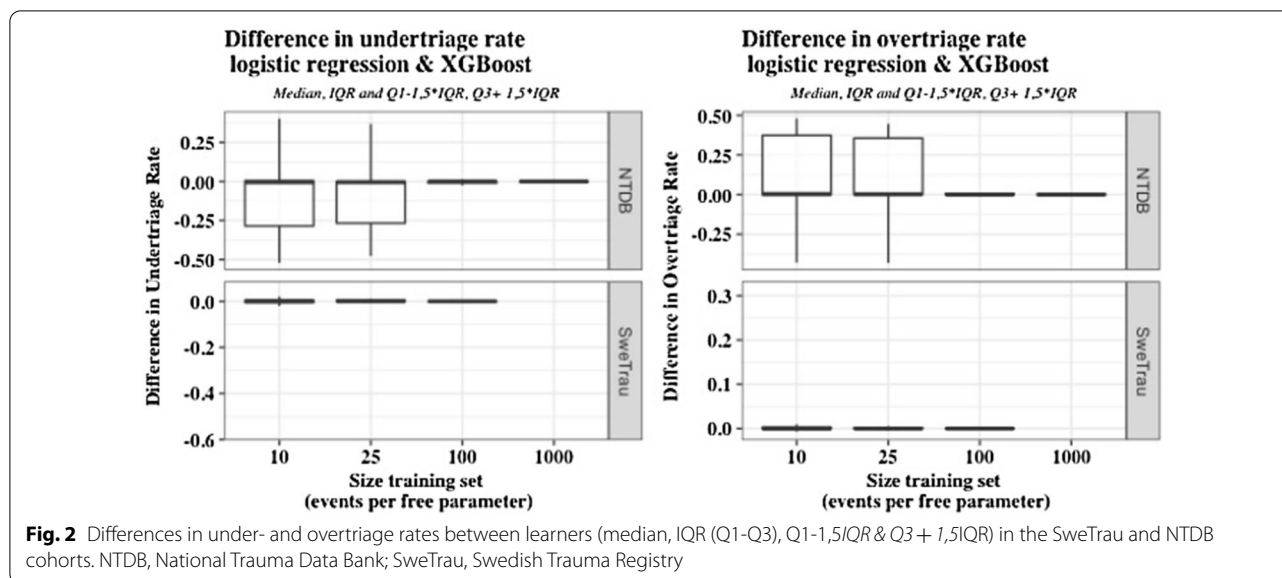
NTDB, National Trauma Data Bank; SweTrau, Swedish Trauma Registry

likely a reflection of a different distribution of the input vital parameters in relation to the proportion of major trauma output in the two cohorts. Despite the higher

proportion of events in the NTDB cohort, the categorical distribution of vital parameters was similar in the two cohorts. This naturally implies different conditions for the modelling task. This also enlightens the complexity of modelling, showing how the performance of the same learner may vary from one setting to another.

In SweTrau, both models produced robust results for all training set sizes, and there were only minimal differences in performance. In NTDB, however, logistic regression produced reliable results from the training set of 25 events per free parameter, whereas XGBoost required the largest data set of 1000 events per parameter to stabilize. Thus, the results indicate that XGBoost may require larger and more comprehensive training data than logistic regression to produce robust results. In addition, as opposed to XGBoost, logistic regression has the advantage of being more transparent.

Previous studies have shown variable results when comparing logistic regression to machine learning for clinical prediction models. Overall, it seems that for



smaller studies with a limited number of predictors, logistic regression performs as well as more advanced machine learners [19, 20], whereas for larger studies with many predictors, more advanced machine learners may have an advantage [14–17]. A recent review including 71 studies with a median of 19 predictors and 8 events per predictor showed no benefits of machine learning over logistic regression [18], but the included studies did not investigate which factors influenced performance. Further studies are needed to establish determinants of the performance of different algorithms.

Our results show that static vital parameters and age are not sufficient as input parameters to achieve mis-triage rates in line with the ACS-COT recommendations. To improve predictive accuracy, vital parameters could be encoded differently, for instance, using the New Trauma Score (NTS) [33] and shock index [34]. Additional predictors, such as trauma type, mechanism of injury or anatomical location of injury, could also be introduced. However, increased model complexity and creation of comprehensive data registries need to be balanced against the clinical reality of EMS personnel in the prehospital setting.

Identifying the optimal prediction model for trauma team activation may enable the development of prehospital and early in-hospital decision-making tools to assist EMS personnel in field triage. This tool may also reduce the prehospital provider interindividual variation in the assessment of injury severity, creating a more solid trauma team activation field triage system.

For example, a recent study showed that only 50% of seriously injured patients in Norway are treated by anaesthesiology-manned prehospital critical care teams [35].

Prehospital undertriage may lead to negative patient outcomes. Analysing selected dispatch data with previously described learners may assist dispatch centres in correctly dispatching a high tier unit to a trauma scene with severely injured patients.

Limitations

Neither SweTrau nor NTDB are perfect populations for developing a trauma triage tool. The ideal population would have been all trauma cases reported to the EMS. Additionally, we excluded data with missing observations. Thus, there were possible sources of selection bias; however, as our objective was to compare two learners, this bias is unlikely to have any major impact on our findings. If the goal was to develop a new model, we would recommend a different approach for handling missing observations, such as multiple imputation.

Encoding of data as categorical variables implies loss of information and could introduce bias. For instance, it is unlikely that the prognosis of an individual would change drastically on the day of the 57th birthday. However, categorical encoding is the norm in published models for prehospital trauma triage, as it can be difficult to count an exact respiratory rate or GCS in a stressful situation.

There are other machine learners for binary classification that we could have assessed, but we used XGBoost because of its current popularity and applicability. Our results cannot be generalized to other machine learners and are limited to XGBoost, as it is implemented in the R package with the same name. It is possible that more extensive hyperparameter tuning, including data re-balancing techniques, would have improved the performance of both learners, but the use of the R package

MLR allowed us to replicate the analysis process in all bootstrap samples. This would not have been possible if we had tuned hyperparameters manually.

Conclusion

The results showed that the advanced machine learner XGBoost did not bring any added value over logistic regression for prehospital trauma triage. In contrast, we observed that this advanced machine learner required larger data sets to produce reliable results. Thus, when predictors are few and categorical, logistic regression may be preferable to this more advanced machine learner.

Abbreviations

ACS COT: The American College of Surgeons Committee on Trauma; AUC: Area Under the Curve; EMS: Emergency Medical Service (Ambulance services, paramedic services); GCS: Glasgow Coma Scale; NTDB: National Trauma Data Bank; RTS: Revised Trauma Score; SweTrau: Swedish Trauma Registry; XGBoost: Extreme Gradient Boosting.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-021-01558-y>.

Additional file 1. Additional model performance results.

Authors' contributions

All authors conceived the study, and AL and JB prepared the data. AL and MGW conducted the analyses. AL wrote the first and subsequent drafts. JB, MG, and MGW read and revised the drafts. All authors read and approved the final manuscript.

Funding

Open access funding provided by Karolinska Institute. This research was funded by the Swedish National Board of Health and Welfare, grant numbers 23745/2016 and 22464/2017.

Availability of data and materials

The data that support the findings of this study are available from the US American College of Surgeons and the Swedish Trauma Registry, but restrictions apply to the availability of these data, which were used under licence for the current study and so are not publicly available. To access the data from this study, please request the appropriate datasets from the US American College of Surgeons (<https://www.facs.org/quality-programs/trauma/tqp/center-programs/ntdb>) and the Swedish Trauma Registry (<http://rcsyd.se/swetrau/om-swetrau/about-swetrau-in-english/swetrau-the-swedish-trauma-registry>) or contact Martin Gerdin Wärnberg (martin.gerdin@ki.se) for more information.

Declarations

Ethics approval and consent to participate

This study was approved by the Regional Ethical Review Board in Stockholm. Ethical review numbers 2015/426-31 and 2016/461-32. Individual informed consent was waived.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Emergency Department, Södersjukhuset, Sjukhusbacken 10, 11883 Stockholm, Sweden. ²Department of Emergency Medicine, Skåne University Hospital Malmö, Inga Marie Nilssons gata 47, 21421 Malmö, Sweden. ³Department of Global Public Health, Karolinska Institutet, 171 77 Solna, Sweden. ⁴Department of Physiology and Pharmacology, Karolinska Institutet, 171 77 Solna, Sweden. ⁵Function Perioperative Medicine and Intensive Care, Karolinska University Hospital, Solna, Stockholm, Sweden. ⁶Swedish Air Ambulance (SLA), Mora, Sweden. ⁷Rapid Response Cars, Stockholm, Sweden.

Received: 16 January 2021 Accepted: 9 June 2021

Published online: 21 June 2021

References

1. Staudenmayer K, Weiser TG, Maggio PM, Spain DA, Hsia RY. Trauma center care is associated with reduced readmissions after injury. *J Trauma Acute Care Surg.* 2016;80(3):412–8.
2. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med.* 2006;354(4):366–78.
3. Zocchi MS, Hsia RY, Carr BG, Sarani B, Pines JM. Comparison of mortality and costs at trauma and nontrauma centers for minor and moderately severe injuries in California. *Ann Emerg Med.* 2016;67(1):56–67.
4. American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patient: 2014. 6th ed. Chicago: American College of Surgeons; 2014.
5. de Munter L, Polinder S, Lansink KW, Cnossen MC, Steyerberg EW, de Jongh MA. Mortality prediction models in the general trauma population: a systematic review. *Injury.* 2017;48(2):221–9.
6. van Rein EA, Houwert RM, Gunning AC, Lichtveld R, Leenen LP, van Heijl M. Accuracy of prehospital triage protocols in selecting major trauma patients. *J Trauma Acute Care Surg.* 2017;83(2):328–39.
7. Newgard CD, Fu R, Zive D, Rea T, Malveau S, Daya M, et al. Prospective validation of the national field triage guidelines for identifying seriously injured persons. *J Am Coll Surg.* 2016;222(2):146–58.
8. Newgard CD, Zive D, Holmes JF, Bulger EM, Staudenmayer K, Liao M, et al. A multisite assessment of the American College of Surgeons Committee on trauma field triage decision scheme for identifying seriously injured children and adults. *J Am Coll Surg.* 2011;213(6):709–21.
9. Bashiri A, Savareh BA, Ghazisaeedi M. Promotion of prehospital emergency care through clinical decision support systems: Opportunities and challenges. *Clin Exp Emerg Med.* 2019;6(4):288–96.
10. Harmsen AM, Giannakopoulos GF, Moerbeek PR, Jansma EP, Bonjer HJ, Bloemers FW. The influence of prehospital time on trauma patients outcome: a systematic review. *Injury.* 2015;46(4):602–60.
11. Brown JB, Rosengart MR, Forsythe RM, Reynolds BR, Gestring ML, Hallinan WM, et al. Not all prehospital time is equal: Influence of scene time on mortality. *J Trauma Acute Care Surg.* 2016;81(1):93–100.
12. Deo RC. Machine learning in medicine. *Circulation.* 2015;132(20):1920–30.
13. Battineni G, Sagarro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *J Pers Med.* 2020;10(2):21.
14. Raita Y, Goto T, Faridi MK, Brown DF, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care.* 2019;23(1):64.
15. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE.* 2018;13:1–13.
16. Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. *Shock.* 2017;48(5):504–10.
17. Kang DY, Cho KJ, Kwon O, Kwon JM, Jeon KH, Park H, Lee Y, Park J, Oh BH. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. *Scand J Trauma Resuscit Emerg Med.* 2020;28(1):17.
18. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine

- learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.
19. Nusinovic S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, Wong TY, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol*. 2020;122:56–69.
 20. Lynam A, Dennis JM, Owen KR, Oram R, Jones A, Shields B, Ferrat LA. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnos Prognos Res*. 2020;4:6.
 21. Gerdin M, Roy N, Felländer-Tsai L, Tomson G, Von Schreeb J, Petzold M, et al. Traumatic transfers: calibration is adversely affected when prediction models are transferred between trauma care contexts in India and the United States. *J Clin Epidemiol*. 2016;74:177–86.
 22. Castillo RS, Kelemen A. Considerations for a successful clinical decision support system. *CIN Comput Inform Nurs*. 2013;31(7):319–28.
 23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.
 24. Morde, Vishal VAS. XGBoost Algorithm: Long May She Reign! 2019.
 25. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the trauma score. *J Trauma*. 1989;29(5):623–9.
 26. Goodmanson NW, Rosengart MR, Barnato AE, Sperry JL, Peitzman AB, Marshall GT. Defining geriatric trauma: when does age make a difference? *Surgery*. 2012;152:668–75.
 27. Chester JG, Rudolph JL. Vital signs in older patients: age-related changes. *J Am Med Dir Assoc*. 2011;12(5):337–43.
 28. Manual Svenska Traumaregistret. 2015.
 29. American College of Surgeons. National Trauma Data Bank Research Data Set User Manual and Variable Description List. 2017; July.
 30. American college of Surgeons Committee Of Trauma. NTDB Data dictionary 2020. 2020; August 2019.
 31. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. 2013.
 32. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. Mlr: machine learning in R. *J Mach Learn Res*. 2016;17(1):5938–42.
 33. Jeong JH, Park YJ, Kim DH, Kim TY, Kang C, Lee SH, et al. The new trauma score (NTS): a modification of the revised trauma score for better trauma mortality prediction. *BMC Surg*. 2017;17(1):77.
 34. Montoya KF, Charry JD, Calle-Toro JS, Núñez LR, Poveda G. Shock index as a mortality predictor in patients with acute polytrauma. *J Acute Dis*. 2015;4(3):202–4.
 35. Wisborg T, Ellensen EN, Svege I, Dehli T. Are severely injured trauma victims in Norway offered advanced pre-hospital care? National, retrospective, observational cohort. *Acta Anaesthesiol Scand*. 2017;61(7):841–7.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

