Performance of prediction models for opportunities for improvement in trauma care across patient cohorts

Anna Thoss

# Introduction

Trauma, defined as the clinical entity composed of the combination of physical injury and the body’s associated responses (1) is a significant public health concern worldwide. It accounts for a substantial portion of morbidity and mortality rates, being the cause of 4.3 million fatalities each year (2). It is the leading cause of death for people aged 18-39 (3). Trauma, and care of patients with traumatic injuries, have long been a significant target point for healthcare improvement programs (4). Multidisciplinary morbidity and mortality conferences form a cornerstone of initiatives dedicated to enhancing the quality of trauma care and consequently, improving patient outcomes (5).

### Opportunities for improvement

Audit filters, a system used to flag abnormal clinical parameters as possible opportunities for improvement (OFI), are applied to determine possible events in patient care that are associated with unfavourable outcomes. The occurrence of such an event triggers a review process, which in turn leads to a meeting of the multidisciplinary review board during mortality and morbidity conferences. An OFI is a consensus decision made during a conference and includes proposals for corrective actions, and when appropriate implementation of corrective measures to address systematic errors and feedback to individual practitioners (6). At mortality and morbidity conferences, representatives from different professions and specialities adjacent to trauma care come together to discuss care provided to a specific patient and compare it to care given under optimal conditions. The findings of such a conference are whether OFI can be found in any given case. Examples of OFI may include lack of resources and management errors (7). The use of audit filters has its drawbacks, being associated with high rates of false positives (8–10).

### Artificial intelligence

In recent years, artificial intelligence (AI) has made great advances, making it an exciting opportunity for the development of more sophisticated prediction models. This opens up a wide field of possible future healthcare applications. Research on AI implementation for clinical use has become commonplace and has already been tried in some clinical contexts, such as aiding in the diagnosis of pathologies through analysis of radiological images (11) as well as epidemiological prediction of disease patterns, an example being during the COVID-19 pandemic (12).

With AI, computing is used to mimic human intelligence, characterized by its ability to learn and reason. Through predictive modelling, AI can be used as a tool to problem solve. Computation has the advantage of analysing great quantities of data in a short time, giving it an edge (11). The primary way to achieve AI is through machine learning (ML), in which statistical modelling is applied to a computer system where it learns from previous data, and thus circumvents human programming. The approach with ML has increasingly been utilized to analyse data and can be used to gain further information from data, such as making predictions of outcomes(13). The ability to improve modelling algorithms with the use of ML thus has the ability to improve said algorithms autonomously, which in turn opens the door for a wide array of applications in healthcare. One such application is to develop more accurate prediction models that can further quality improvement programs.

### AI in trauma care

Following the general trend towards implementation of machine learning models within healthcare, there has been research into possible application within the wider field of trauma care. Previous methods to predict outcomes within trauma, such as Injury severity score (ISS) and Trauma and Injury severity score (TRISS) were developed using much smaller quantities of data than processing allows for these days(14), this has lead some to call their accuracy of prediction into question.(15). There is an implicit heterogeneity to trauma patients, which lends itself well to machine learning modelling, having not only the ability to make predictions based on a greater dataset, but also circumvents problems as it can handle unstructured and non-linear data and account for missing values.(14)

Previous studies have found that delay of treatment along with errors in judgement are the leading causes of mortality in level 1 trauma centres.(16,17) Machine learning has been found to outperform clinical tools for prediction of mortality.(18) Research on morbidity in trauma patients has been less conclusive, and fewer studies have been done on the topic, with most focusing on mortality.(14)

To enhance the precision of the selection process for mortality and morbidity conferences, there have been efforts to introduce trauma mortality prediction models. Nevertheless, the performance of these models has been unsatisfactory, being poor predictors for OFI (9,19). This is likely because these models were originally developed with a focus on predicting mortality rather than morbidity or instances of care failures within the realm of trauma care. Machine learning-based prediction models have been found to outperform conventional methods to predict opportunities for improvement, and the learner XGBoost showed particular promise (20).

### Trauma care in different patient cohorts

Trauma is a wide field with a heterogeneous patient population as well as encompassing an array of different injuries and responding mechanisms of the body to preserve homoeostasis.(21) Trauma cases can be divided into cohorts in different ways, depending on mechanism of injury, damage to the nervous system or response of the body. Not only does the presentation of these patients wary, but management differers both due to severity of injury, as well as type of injury. (22) Thus, clinical judgement is an important component in treatment. As previous studies have not only found that judgement errors are common in OFI, but that the time frame between arrival at the emergency room and diagnostic tests such as CT or corrective procedures like surgery is longer in more critically ill patients (23), the importances to not discriminate against a cohort in a machine learning model is vital in finding OFI for quality improvement across the board. If models are to be implemented in the clinical context in the future, it is important to establish that they perform well across different trauma cohorts.

# Aim

To assess the difference in performance of a machine learning based prediction model for opportunities for improvement in trauma care across different clinical trauma cohorts.

# Methods

### Source of data

We conducted a registry-based study, encompassing all trauma patients enlisted in both the Karolinska University Hospital trauma registry and the trauma care quality database spanning from 2012 to 2022. The trauma registry encompasses approximately 12,000 patients treated from 2012 to 2022. The trauma care quality database in turn is a subset of the trauma registry, including about 6000 patients selected for review in the years 2013-2022.

### Participants

The Karolinska University Hospital in Solna is equivalent to a level 1 trauma centre and manages approximately 1,500 acute trauma patients each year (9,24). The trauma registry comprises all patients admitted to the hospital with trauma team activation, irrespective of their Injury Severity Score (ISS), as well as patients with an ISS exceeding 9. Included in the registry is data on vital signs, injuries, interventions, times, and patient demographics.

All patients reviewed for possible OFI between the years 2013 and 2022 are included in the study, with the exception of patients under the age of 15 due to differing clinical pathways in this subgroup.

### Outcome

The outcome in our study is OFI, as previously defined by Albaaj et. al: “The outcome was the presence of OFI, as decided by the morbidity and mortality conference. An OFI is any failure of care including, but not limited to, any potentially preventable or preventable death, delay in treatment, clinical judgment error, missed diagnosis and technical error as decided by the mortality and morbidity conference. The study outcome is binary with the levels”Yes - At least one OFI identified” and “No - No OFI identified”.”(23)

An OFI is a consensus decision made during a conference and includes proposals for corrective actions. OFI are identified through meetings of the multidisciplinary review board during mortality and morbidity conferences. An OFI is a consensus decision made during a conference and includes proposals for corrective actions. These in turn are recorded in the trauma care quality database. The trauma care quality database also includes data relevant to mortality and morbidity conferences, such as audit filters. The mortality and morbidity conferences at Karolinska University Hospital involve professionals from various specialities within trauma care, including surgery, neurosurgery, orthopaedics, anaesthesia, intensive care, nursing, and radiology. Cases considered for a conference go through multiple rounds of review. Mortality leads to direct inclusion.

Between 2013 and 2017, non-fatal poor outcomes were identified through reviews conducted by a specialized trauma nurse to identify potential OFI. Starting in 2017, the process was formalized, involving an initial review by a specialized trauma nurse along with the application of audit filters. All cases that were flagged either during the initial nurse review or by the audit filters underwent a secondary review conducted by two specialized nurses. If the second review identified potential areas for improvement (OFI), the case was marked for discussion in a multidisciplinary conference.

Figure illustrating process of mortality and morbidity conference at Karolinska University Hospital.

### Predictors

All variables from the trauma registry are considered potential predictors in our analysis. The models have been developed and trained on the data set previous to this study (20). Eight models were included in the original study, all being developed using the Tidymodels framework (25). These include logistic regression (LR), random forest (RF), decision tree (DT), support vector machine with a radial basis kernel (SVM), XGBoost, LightGBM, CatBoost, and k-nearest neighbor (k-NN). In this study, we focused on the model developed using XGBoost, as this model performed the best in the original paper.

Our variables include information spanning the pre-hospital, initial care, and subsequent in-hospital phases, which include initial vital signs, timing, and types of procedures and interventions, duration and level of care, injury details, mechanisms of injury, injury types, and standard demographic information. This array of predictors comprises both continuous and categorical variables, ultimately resulting in the final models utilizing a total of 17 predictors.

### Sample size

All available data was used, with a sample size of 8309 patients. This is due to previously developed prediction models being based on all available data, as well as including all 17 predictors regardless of the learner used for the development of the model.

### Cohorts

The different models’ performance were assessed in the following cohorts(22):

1. Isolated severe TBI: Injury isolated to the area of the brain with AIS > 2 and:
   1. pre-hospital GCS of < 9
   * or
   1. Pre- or in-hospital intubation.
2. Blunt multi-system trauma without TBI: Blunt trauma with AIS > 2 and injuries in at least two of the following AIS body regions: head, face, neck, thorax, abdomen, spine, or upper and lower extremities and no injury to the brain with AIS >2 and:
   1. pre-hospital GCS of < 9
   * or
   1. Pre- or in-hospital intubation.
3. Blunt multi-system trauma with TBI: Blunt trauma with AIS > 2 and injuries in at least two of the following AIS body regions: head, face, neck, thorax, abdomen, spine, or upper and lower extremities and with injury to the brain with AIS >2 and:
   1. pre-hospital GCS of < 9
   * or
   1. Pre- or in-hospital intubation
4. Severe penetrating: At least one AIS > 3 injury in any of the following AIS body regions: neck, thorax, abdomen.
5. Other cohort: Patient does not fall into any of the above.

### Statistical analysis

All statistical analyses were conducted using R. To extract probabilities for each patient case in the dataset, a 10-fold cross-validation strategy was employed, during which predictions for the holdout fold were recorded for each data point. We assessed performance within predefined cohorts and compared it across the different cohorts in terms of discrimination. This comparison was based on the area under the receiver operating characteristic curve (AUC). To estimate the 95% confidence interval for AUC performance, we employed bootstrapping on the extracted probabilities, repeating the process 1000 times. Additionally, an Analysis of Variance (ANOVA) test was conducted to compare differences between cohorts. A significance level of 5% was employed.

### Ethical considerations

The nature of this study is non-invasive. However, handling of registry data constitutes a risk for breach of patients integrity. The registry used in this study does contain personal identification. In order to minimize the risk of breach, patients have been semi-anonymized after collection of data and been assigned a new identification number upon entry into dataset. Due to the nature of the patient population, patients have not consented previous to entry into the registry, thus, there is an opt out principle applied here. This is common with registry data in Sweden, and opens up opportunities for studies in a wide range of fields. In turn, more scientific knowledge has great potential for benefit, both on the societal scale, as well as the individual. Thus, the risk for the individual is comparably small. Further, no individual patients data is presented in the study, mitigating risk of violating personal integrity. The study has been granted ethical permit, with the record number 2021-02541 and 2021-03531.

# Results

### Participants

The Karolinska trauma registry includes 14022 patients, of which 8309 are reviewed for OFI, either through selection by individual review or application of audit filters, as well as an unknown process (n=2). Through this, 7571 patients were selected for the morbidity conference, and 727 patients for the mortality conference.

Among the 8309 patients, the majority (n=5386) (69%) were male, and the mean age was 45 (SD 21). Overall, the mortality in the group was 727 (9%). In the group identified as OFI (n = 512), the mean age was somewhat higher than in the group without OFI (n=7797) (mean 49 vs 45 years). For OFI patients, they were most commonly treated at an intensive care unit (33%), whereas patients without OFI were most commonly treated at a general ward (38%). ISS was also higher for OFI patients (mean: 18, SD:11 vs mean: 12, SD:13). “Respiratory rate” was the most common missing variable at 20% (n=1644), followed by “ED GCS” at 10% (n=866).

Out of 727 deaths reviewed, 41 were found to be preventable or possibly preventable. The remaining 686 patients were thus rendered non-preventable without OFI. Further, out of the 7571 alive patients were included during a morbidity conference, 471 (6%) were identified as cases with OFI.

The biggest cohort in the study was “other cohort” with 6707 (81%) of patients, followed by “blunt multisystem without TBI” at 9% (n=708). These were also the cohorts with the most OFI, making up 64% (n=329) and 23% (n=116) respectively. See table 1 for patient characteristics.

### Main result

We compared the performance of the ML model in the different cohorts with AUC as our performance metric (See table 2). The ML model performed best in the cohort of patients with isolated severe TBI, in which it had an AUC of 0.811 (0.807-0.814), followed by “Other cohort” (AUC 0.761 0.76-0.762). The ML model performed worst in the cohort of patients with blunt multisystem trauma with TBI (AUC 0.696, 0.692-0.701). Figure 2 shows details of the receiver operating curves. The P-value between the different cohorts was 1 x 10 ^-6, meaning that there were significant differences in AUC between cohorts.

# Discussion

We set out to assess the performance of a model based on the machine learner XGBoost’s viability as a tool to aid clinicians in their effort to improve trauma care quality, and specifically how they would perform across different trauma cohorts. We found that there is some variation in the model’s performance between the different cohorts, with “Blunt multisystem without TBI” performing the lowest, despite being the largest clearly defined trauma cohort in the study. The largest group “Other cohorts”, which is made up of cases falling outside of the boundaries for the other formalized cohorts performs the best over all, despite possibly encompassing the largest heterogeneity.

A previous study on the same dataset has found that audit filters, with an AUC of 0.624, are outperformed by ML models when the dataset is not categorised by cohorts.(20) It is unclear how audit filters perform in this context across cohorts as this was outside of the scope the study. Interestingly, the model performed better on the “Isolated severe TBI” cohort in terms of AUC then the best model in the aforementioned study (AUC 0.811 vs 0.789). A 2020 study found similarly good performance of machine learning models in TBI patients looking to predict poor outcomes. (26) This may suggest that some models inclination to perform well in prediction of traumatic brain injury, whereas outcomes linked to other trauma cohorts may be harder to predict.

Overall the model perform well in the prediction of OFI across the cohorts. The application of machine learning models to outcomes within trauma is commonly more widely studied with more straight forward prediction of mortality, rather than the homogeneous reasons that make up the outcome of OFI. (14,27) Therefore, it is hard to make direct comparisons to other studies findings in this case, as they are generally either focused on a outcome far less complex, or not specifically tied to performance in outcomes in a particular cohort.

### Limitations

The study had several limitations. Firstly, the nature of OFI is defined as a binary- variable. However, events leading up to a case being positive for OFI contains a multitude of variables, and can be determined by outcomes ranging from bad documentation to preventable death. Behind these are a diverse set of possible clinical events, which in turn can be determined by different predictors. It has been suggested that data with higher resolution, such as vital signs series rather then single values could be a solution to this. However this comes with a drawback, as the current collection of data is based on the Upstein template (Uniform Reporting of Data following Major Trauma), and as such opens the door to a wider set of application in other contexts. Moving away from this has important implications for external validity of the model. (20)

Secondly, our learner is dependent on the current selection process for OFI, as this makes out the training data and thus makes out the groundwork for what is considered OFI in our dataset. As a consequence, the models are prone to some of the same limitations as the current system of audit filters. The review process is dependent both the individual review of specialized nurses, along with the current set of audit filters, which may not be the optimal parameters for determining OFI. This introduces an inherent bias. The audit filter system favour some types of error, while possibly missing others. At Karolinska University Hospital, the current system is more apt at identifying severe clinical outcomes, such as preventable death, or patient cases with more interventions or longer hospital stays, and flagging them as possible OFI. Higher ISS scores and “Days in hospital” have been found to be predictors of high importance when identifying OFI (23). This in turn introduces possible selection bias, giving a skewed view of errors within the trauma care, and as such not a holistic view of the trauma care quality.

Further, machine learning models are suboptimal when applied to cases with rare events, as they are limited to what they have previously been trained on. This means that not only are they limited when finding OFI in cases with novel predictors or events that lead with error, but the models are also prone to creating a self-reinforcing cycle, making them specialised on finding a set of OFI with accuracy. The current system can circumvent such by the individual review followed by a mortality and morbidity conference. If the model approach to predicting OFI is to be implemented in clinical practice, some other form, either individual review or other form of identification would be needed to supplement the machine learning model to assure prediction previously unknown basis for OFI.

Finally, given that the aim of identifying OFI at mortality and morbidity conferences is the correction of errors within the system of trauma care, with time and implementation of corrective actions, the the scenarios and following this the parameters that predict OFI may change. Inherent to machine learning models is the potential for heavy bias.(14) To mitigate this, a model should be subject to scrutiny of data sets and features selected, and external validation.

### Interpretation

At this point it worth to note that perfect performance of the model was neither expected or the end goal. As previously stated, OFI is a heterogeneous outcome, and comparing a machine learning model to the current system of both quantitative screening and subsequent human reviews is in some ways unfair. The goal in developing a model to perform this analysis is its potential in aiding human efforts, to streamline the selection process for mortality and morbidity conferences. If implemented in a clinical context, it would be less resource intensive and could potentially be adjusted to new clinical patterns in the future that could be associated with OFI.

In this sense, as it stands state of the art machine learning models have immense potential to aid in these types of selection processes. Time and costs are finite resources, and as such, there should always be the question of if they are adequately spent and if there is room for optimisation. This does not only apply to the selection process itself, but also to mortality and morbidity conferences. Therefore, efforts to optimise the selection of possible OFI to be discussed is imperative. With the current selection process resources are spent on discussing cases without OFI as the outcome. To limit the time spent on such cases would be a win.

A recent guideline stresses the need for more consideration of aspects such as usability, trust in algorithm output and human factor human factors to enable implementation of decision systems based on AI into clinical practice. (28) Key barriers to entry for these kinds of model into the medical community remain.

In this study, we set out to investigate the viability of a model based on the machine learner XGBoost across cohorts. The findings suggest that the model does not perform as well across all patient types, and could need further improvement before implementation in selection of cases to be discussed at a conference. This is important from a fairness perspective, as the type of injury should not dictate the opportunity to find and correct errors in clinical practise.

### Further research

Overall, further research is needed to increase performance of the models. Under optimal circumstances, any method to screen for possible OFI should minimise the set of false positives and also capture any true OFI, and as such minimise the clinical cost and time needed for trauma case quality improvement.

It is important to highlight that these findings are based on single a single centre study, and as such only illustrates the situation in Stockholm. To establish a model that has validity in a wider context, further research would have to establish how the performance may differ in a different clinical context, and take a wider set of data and circumstances into account.

Further, men make up a disproportionally large group compared to women when it comes to trauma. Each year, almost twice as many men die due to traumatic injury when compared to women, with the distribution of different types of injury varying by demographic and region (29). Females have been shown to have a higher survival following trauma in higher income countries, whereas those in lower income countries show no difference compared to men. Differences in cell-mediated immune response following haemorrhage, differences in post-traumatic cytokine levels, and potential protective effects of oestrogen have been proposed to be the cause of the disparity between these groups (30). Research on audit filters has found no correlation between audit filters and OFI. (23) It is important, that due to the differences between men and women, that the model for prediction has similar performance in these subgroups, and does not discriminate. For future studies, it will be important to establish that any prediction model implemented in clinical practice performs equally well for both genders.

# Conclusion

State of the art machine learning models show promise for future implementation into a clinical context to predict OFI. Our findings show that although they overall perform well across cohorts, some disparity exists in the models accuracy across different cohorts.

# References

1. Gerdin M. The risk of dying: Predicting trauma mortality in urban indian hospitals. Stockholm: Karolinska Institutet; 2015.

2. GBD 2019 Diseases and Injuries Collaborators. Injuries—level 1 cause. The Lancet. 2020 Oct;396.

3. Peng HT, Siddiqui MM, Rhind SG, Zhang J, Luz LT da, Beckett A. Artificial intelligence and machine learning for hemorrhagic trauma care. Mil Med Res. 2023 Feb;10(1):6.

4. Organization WH. Guidelines for trauma quality improvement programmes. Available from <https://apps.who.int/iris/handle/10665/44061> (2023/09/21);

5. Santana MJ, Stelfox HT. Development and evaluation of evidence-informed quality indicators for adult injury care. Annals of Surgery [Internet]. 2014 Jan;259(1):186–92. Available from: <https://doi.org/10.1097/sla.0b013e31828df98e>

6. McDermott FT. TRAUMA AUDIT AND QUALITY IMPROVEMENT. ANZ Journal of Surgery [Internet]. 1994 Mar;64(3):147–54. Available from: <https://doi.org/10.1111%2Fj.1445-2197.1994.tb02168.x>

7. Lazzara EH, Salisbury M, Hughes AM, Rogers JE, King HB, Salas E. The morbidity and mortality conference: Opportunities for enhancing patient safety. Journal of Patient Safety [Internet]. 2020 Aug;18(1):e275–81. Available from: <https://doi.org/10.1097/pts.0000000000000765>

8. Sanddal TL, Esposito TJ, Whitney JR, Hartford D, Taillac PP, Mann NC, et al. Analysis of preventable trauma deaths and opportunities for trauma care improvement in utah. Journal of Trauma: Injury, Infection, Critical Care [Internet]. 2011 Apr;70(4):970–7. Available from: <https://doi.org/10.1097/ta.0b013e3181fec9ba>

9. Ghorbani P, Strömmer L. Analysis of preventable deaths and errors in trauma care in a scandinavian trauma level-i centre. Acta Anaesthesiologica Scandinavica [Internet]. 2018 May;62(8):1146–53. Available from: <https://doi.org/10.1111/aas.13151>

10. Evans C, Howes D, Pickett W, Dagnone L. Audit filters for improving processes of care and clinical outcomes in trauma systems. Cochrane Database of Systematic Reviews [Internet]. 2009 Oct; Available from: <https://doi.org/10.1002/14651858.cd007590.pub2>

11. Cellina M, Cè M, Irmici G, Ascenti V, Caloro E, Bianchi L, et al. Artificial intelligence in emergency radiology: Where are we going? Diagnostics (Basel). 2022 Dec;12(12):3223.

12. Chen J, See KC. Artificial intelligence for COVID-19: Rapid review. J Med Internet Res. 2020 Oct;22(10):e21476.

13. Jung A. Machine learning, the basics [Internet]. Springer; 2022. Available from: <https://books.google.se/books/about/Machine_Learning.html?hl=it&id=1IBaEAAAQBAJ&redir_esc=y>

14. Zhang T, Nikouline A, Lightfoot D, Nolan B. Machine learning in the prediction of trauma outcomes: A systematic review. Annals of Emergency Medicine [Internet]. 2022 Nov;80(5):440–55. Available from: <http://dx.doi.org/10.1016/j.annemergmed.2022.05.011>

15. Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over decision tree models: An application for estimating uncertainty in trauma severity scoring. International Journal of Medical Informatics [Internet]. 2018 Apr;112. Available from: <http://dx.doi.org/10.1016/j.ijmedinf.2018.01.009>

16. Teixeira PGR, Inaba K, Hadjizacharia P, Brown C, Salim A, Rhee P, et al. Preventable or potentially preventable mortality at a mature trauma center. Journal of Trauma: Injury, Infection &amp; Critical Care [Internet]. 2007 Dec;63(6):1338–47. Available from: <http://dx.doi.org/10.1097/TA.0b013e31815078ae>

17. Teixeira PGR. Preventable morbidity at a mature trauma center. Archives of Surgery [Internet]. 2009 Jun;144(6):536. Available from: <http://dx.doi.org/10.1001/archsurg.2009.82>

18. Kareemi H, Vaillancourt C, Rosenberg H, Fournier K, Yadav K. Machine learning versus usual care for diagnostic and prognostic prediction in the emergency department: A systematic review. Mitchell AM, editor. Academic Emergency Medicine [Internet]. 2021 Jan;28(2):184–96. Available from: <http://dx.doi.org/10.1111/acem.14190>

19. Heim C, Cole E, West A, Tai N, Brohi K. Survival prediction algorithms miss significant opportunities for improvement if used for case selection in trauma quality improvement programs. Injury [Internet]. 2016 Sep;47(9):1960–5. Available from: <https://doi.org/10.1016%2Fj.injury.2016.05.042>

20. Attergrim J, Szolnoky K, Strömmer L, Brattström O, Whilke G, Jacobsson M, et al. Predicting opportunities for improvement in trauma using machine learning: A registry based study. 2023 Jan; Available from: <https://doi.org/10.1101/2023.01.19.23284654>

21. Dumovich J SP. Physiology, trauma. 2022.

22. Hornor MA, Hoeft C, Nathens AB. Quality benchmarking in trauma: From the NTDB to TQIP. Current Trauma Reports [Internet]. 2018 Apr;4(2):160–9. Available from: <http://dx.doi.org/10.1007/s40719-018-0127-1>

23. Albaaj H, Attergrim J, Strömmer L, Brattström O, Jacobsson M, Wihlke G, et al. Patient and process factors associated with opportunities for improvement in trauma care: A registry-based study. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine [Internet]. 2023 Nov;31(1). Available from: <http://dx.doi.org/10.1186/s13049-023-01157-y>

24. SweTrau. Årsrapporter. Available at <https://rcsyd.se/swetrau/om-swetrau/arsrapporter> (2023/09/21);

25. Kuhn M WH. Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles. Available at <https://www.tidymodels.org/> (2023/09/21); 2020.

26. Matsuo K, Aihara H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. Journal of Neurotrauma [Internet]. 2020 Jan;37(1):202–10. Available from: <http://dx.doi.org/10.1089/neu.2018.6276>

27. Gauss T, Perkins Z, Tjardes T. Current knowledge and availability of machine learning across the spectrum of trauma science. Current Opinion in Critical Care [Internet]. 2023 Oct;29(6):713–21. Available from: <http://dx.doi.org/10.1097/MCC.0000000000001104>

28. 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. Nature Medicine [Internet]. 2022 May;28(5):924–33. Available from: <http://dx.doi.org/10.1038/s41591-022-01772-9>

29. Organization WH. Injuries and violence: The facts 2014. Available at [http://
https://iris.who.int/handle/10665/149798](http:// https://iris.who.int/handle/10665/149798) (2005/06/12);

30. Pendleton AA, Sarang B, Mohan M, Raykar N, Wärnberg MG, Khajanchi M, et al. A cohort study of differences in trauma outcomes between females and males at four indian urban trauma centers. Injury [Internet]. 2022 Sep;53(9):3052–8. Available from: <https://doi.org/10.1016/j.injury.2022.07.022>