

UCLA

UCLA Previously Published Works

Title

Machine learning without borders? An adaptable tool to optimize mortality prediction in diverse clinical settings.

Permalink

<https://escholarship.org/uc/item/57q05509>

Journal

The journal of trauma and acute care surgery, 85(5)

ISSN

2163-0755

Authors

Christie, S Ariane
Hubbard, Alan E
Callcut, Rachael A
[et al.](#)

Publication Date

2018-11-01

DOI

10.1097/ta.0000000000002044

Peer reviewed



Published in final edited form as:

J Trauma Acute Care Surg. 2018 November ; 85(5): 921–927. doi:10.1097/TA.0000000000002044.

Machine Learning Without Borders? An Adaptable Tool to Optimize Mortality Prediction in Diverse Clinical Settings

S. Ariane Christie, MD¹, Alan E Hubbard, PhD², Rachael A Callcut, MD MSPH¹, Morad Hameed, MD MPH³, Fanny Nadia Dissak-Delon, MD MPH⁴, David Mekolo, MD⁵, Arabo Saidou, MD⁵, Alain Chichom Mefire, MD⁶, Pierre Nsongoo, MD⁷, Rochelle A. Dicker, MD⁸, Mitchell Jay Cohen, MD⁹, and Catherine Juillard, MD MPH¹

¹Department of Surgery, University of California San Francisco; San Francisco, California.

²Department of Biostatistics, University of California Berkeley; Berkeley, California.

³Division of General Surgery, Vancouver General Hospital, University of British Columbia, Vancouver, BC, Canada.

⁴Littoral Regional Delegation of the Ministry of Public Health; Cameroon.

⁵Laquintinie Hospital of Douala; Douala, Cameroon.

⁶Regional Hospital of Limbe; Limbe, Cameroon.

⁷Catholic Hospital of Pouma; Pouma, Cameroon.

⁸Department of Surgery, University of California Los Angeles; Los Angeles, California.

⁹Denver Health Medical Center and the University of Colorado; Denver, Colorado.

Abstract

Background: Mortality prediction aids clinical decision-making and is necessary for quality improvement initiatives. Validated metrics rely on pre-specified variables and often require advanced diagnostics which are unfeasible in resource-constrained contexts. We hypothesize that machine learning will generate superior mortality prediction in both high-income and low and middle-income country cohorts.

Methods: SuperLearner (SL), an ensemble machine-learning algorithm, was applied to data from three prospective trauma cohorts: a highest-activation cohort in the United States (US), a high-volume center cohort in South Africa (SA), and a multicenter registry in Cameroon. Cross-validation was used to assess model discrimination of discharge mortality by site using receiver

Sabrinah.Christie@ucsf.edu, hubbard@berkeley.edu, Rachael.Callcut@ucsf.edu, morad.hameed@vch.ca, fannynadia@gmail.com, mekolodavid@yahoo.fr, arabosaidou@yahoo.fr, chichomefire@gmail.com, pierrensongoo@yahoo.fr, RDicker@mednet.ucla.edu, mitchell.cohen@dhha.org, Catherine.Juillard@ucsf.edu. Corresponding author: Catherine Juillard, MD MPH, 1001 Potrero Ave, 3A, San Francisco, CA 94110, Phone: 415-206-4622, Fax: (415) 206-5484, Catherine.Juillard@ucsf.edu.

Conflicts of interest: For all authors, no conflicts of interest declared.

AUTHOR CONTRIBUTIONS

SAC, AEH, RAC, RAD, MJC, and CJ contributed to study design, data collection, data analysis, data interpretation, writing and critical revision. FNDD, DM, AS, ACM, PN, and MH contributed to study design, data collection and critical revision.

operating characteristic curves. SL discrimination was compared with standard scoring methods. Clinical variables driving SL prediction at each site were evaluated.

Results: Data from 28,212 injured patients were used to generate prediction. Discharge mortality was 17%, 1.3%, and 1.7% among US, SA, and Cameroonian cohorts. SL delivered superior prediction of discharge mortality in the US (AUC 94–97%) and vastly superior prediction in Cameroon (AUC 90–94%) compared to conventional scoring algorithms. It provided similar prediction to standard scores in the SA cohort (AUC 90–95%). Context-specific variables (partial thromboplastin time in the US and hospital distance in Cameroon) were prime drivers of predicted mortality in their respective cohorts, while severe brain injury predicted mortality across sites.

Conclusions: Machine learning provides excellent discrimination of injury mortality in diverse settings. Unlike traditional scores, data-adaptive methods are well-suited to optimizing precise site-specific prediction regardless of diagnostic capabilities or dataset inclusion allowing for individualized decision-making and expanded access to quality improvement programming.

Level of Evidence: Level III

Study Type: Prognostic and Therapeutic

Keywords

Precision medicine; Low and middle-income countries; Prediction; Machine learning

BACKGROUND

Mortality prediction after injury enables patient risk stratification, aids in clinical decision-making, and is a necessary benchmark for the development of quality improvement initiatives (1, 2). Over the past 60 years, the search for superior, widely-applicable trauma prediction has led to the development of dozens of severity algorithms. These models have incorporated a variety of different clinical variables including indicators of anatomic severity (3–5), physiologic predictors (6–8), and combinations of both anatomic and physiologic parameters (9, 10).

Despite these many approaches, no single score has been successful in generating ideal prediction over time and across diverse clinical contexts (11). While most scores are initially published with high reported predictive capacity, historically, performance has degraded over time, particularly when applied to different trauma populations. This, in turn, fuels the development of still more severity scoring systems and modifications.

One substantial problem has been the application of severity scores to trauma populations in low- and middle-income countries (LMICs). Scoring metrics developed in the US and other high-income countries routinely rely upon advanced diagnostic imaging or operative exploration for accurate anatomic assessment, making utilization unfeasible in resource-constrained contexts. Even physiologic parameters, such as vital signs, are often sparse in settings with extreme staffing and equipment limitations. Scoring systems designed to generate prediction in LMICs, such as the Kampala Trauma Score, have performed inconsistently in different datasets (12–14).

In other fields, machine learning algorithms have been utilized to generate excellent prediction of complex outcomes by using a data-driven approach to model selection with many potential predictors (15, 16). Due to the fact that these methods do not rely on a static variable set, they have the potential to actually adapt prediction to different clinical contexts (17, 18). Application of one such machine learning algorithm called SuperLearner (SL) has previously been demonstrated to improve prediction of complex outcomes in human immunodeficiency virus (15, 16) and in a single-center cohort SL has demonstrated superior discrimination of death after injury compared to conventional statistical approaches (19, 20).

However, the promise of machine learning is wide applicability of the precision medicine approach. To date, the predictive capacity of machine learning in diverse trauma datasets remains unknown. In this study, we hypothesize that machine learning, specifically so-called ensemble learning using all available admission variables will generate superior prediction of death in both high-income country and LMIC trauma cohorts. Identification of a translatable method of generating reliable prediction in diverse contexts would enable setting-specific tailored severity estimation for improved resource triage and clinical decisionmaking.

METHODS

SuperLearner (SL), an ensemble machine-learning algorithm, was applied to all admission data from three large prospective trauma cohorts: a highest-level activation urban trauma center in the United States (US), a high-volume trauma center in South Africa (SA), and three referral hospitals in Cameroon.

US COHORT

The US data utilized was extracted from prospective data collected on a severely-injured patient cohort presenting to San Francisco General Hospital between February 2005 and April 2015. Specifically, the Activation of Coagulation and Inflammation in Trauma (ACIT), which has been described in detail in other studies, was a single-center prospective cohort study which followed severely injured trauma patients from emergency department admission through hospitalization (21, 22). All adult trauma patients presenting to San Francisco General Hospital between February 2005 and April 2015 who met criteria for highest triage activation level were included into the US cohort. Exclusion criteria included patient age less than 15 years, pregnancy, incarceration, thermal mechanism, and transfer from outside hospital. Baseline data collected included patient demographics, past medical history, substance use, and injury characteristics and physiologic variables including vital signs, laboratory monitoring including an extensive panel of coagulation and inflammation markers, ventilator parameters, input/output data, and fluid, colloid, blood product, and medication administration. Outcomes data including hospital mortality were collected.

Although ACIT collected longitudinal data from admission throughout the first 28 days of hospitalization, for the purposes of this study, only admission variables were used for SL prediction. The study was carried out with the approval of the University of California San Francisco Institutional Review Board.

SOUTH AFRICAN COHORT

The South African data utilized was extracted from prospective data collected on injured patients presenting to Groote Schuur Hospital between April 2014 and November 2016. South Africa is an upper-middle-income country with high rates of injury deaths due to homicide and vehicular trauma (23). Groote Schuur Hospital is a government-funded, high-volume, tertiary teaching hospital in Cape Town, South Africa with an estimated census of 9000 trauma patients per year (24). Between April 2014 and November 2016 prospective data were collected on injured patients presenting to Groote Schuur Hospital using a clinician-entered point-of-care mobile electronic Trauma Health Record (eTHR) system. This system and cohort has been described in detail in prior publications (24). Briefly, all admitted patients older than 12 years of age with primary blunt or penetrating trauma mechanisms were included in the cohort and followed from admission through hospitalization. Patients were excluded if they were determined to be dead on arrival, presented with thermal injuries, were discharged home or transferred to a step-down facility from the emergency department, or if the patient's discharge and/or hospital disposition was unknown. Concurrent with patient care clinicians entered data on demographic and injury characteristics, clinical parameters on admission, and patient outcomes including disposition and death.

For the purposes of this study only admission variables were used for SL prediction. Ethical approval for this study was obtained from the Human Research Ethics Committee of the University of Cape Town and Institutional Review Board of the Department of Surgery, Groote Schuur.

CAMEROONIAN COHORT

The Cameroon data utilized was extracted from the Cameroon Trauma Registry which collected prospective data on injured patients between July 2015 and January 2017. Cameroon is a primarily francophone lower-middle income country in central Africa with among the lowest life expectancies on earth. In 2015 a Cameroon Trauma Registry was established to collect prospective data on all injured patients presenting to three referral hospitals: Laquintinie Hospital of Douala, a high-volume urban hospital in Cameroon's largest city; Limbe Regional Hospital, a regional medical center with mixed urban/rural catchment in the Anglophone Southwest Region of Cameroon; and the Catholic Hospital of Pouma, a hospital on the highly-trafficked road between Douala and Yaoundé, Cameroon's capital. Between July 2015 and January 2017 all injured patients presenting to these three hospitals for trauma were followed from emergency admission to hospital discharge. Specifically, trained Cameroonian trauma registrars recorded patient data at the time of patient admission using paper trauma registry forms (collecting data regarding demographics, mechanism of injury, past medical and surgical history, injuries and injury severity, vital signs, radiology studies, resuscitation measures, treatments, and disposition) and subsequently on paper follow-up forms (collecting data regarding operative interventions, laboratory and radiology studies obtained, medications and treatments administered, complications, and hospital outcomes). The paper forms were initially stored in secured lock boxes on-site at each hospital and then were subsequently entered by trauma registrars into a REDCap database hosted by the University of California San Francisco

server. All patients presenting with injury mechanisms were included in the cohort. No exclusions were made on the basis of injury severity, trauma mechanism, hospital disposition, or age.

For the purposes of this analysis, only hospital indicators available at presentation were utilized for SL prediction. Ethical clearance was obtained from the University of California, San Francisco Committee on Human Research and the National Ethics Committee of the Republic of Cameroon.

SUPERLEARNER PREDICTION

SL is a previously-validated ensemble machine learning algorithm which has been described in detail in prior publications (18). SL can be downloaded as a package within the R coding language (25). Both the R software and the SL package are open-source and can be accessed without charge to the user (<https://cran.r-project.org/> and <https://cran.r-project.org/web/packages/SuperLearner/index.html>, respectively). This means that this technology is widely accessible to clinicians and researchers in both high-income and LMIC contexts.

Rather than pre-specifying a single statistical approach, SL simultaneously investigates multiple algorithms ranging from simple logistic regression to highly complex machine learning (e.g., neural nets) in order to optimally predict outcomes of interest from complex datasets. SL uses cross-validation to tailor a weighted (convex) combination of learners to optimize prediction on new data from the same data-generating distribution. Embedded cross-validation eliminates the risk of over-fitting (18).

In this study SL was applied to all admission variables of the US, SA, and Cameroonian cohorts to generate setting- specific prediction of hospital mortality. We used a set of algorithms including: logistic and linear regression, generalized additive models with various levels of smoothing (26), random forest (27), lasso (28) and systems-based on sieves of parametric models (e.g. polyclass). To report the ability of the resulting SL fit to future data, we estimated the cross-validated area under the curve (AUC) of receiver-operator characteristic curves (ROC) as well as using cv-AUC as the objective function, so the procedure optimizes prediction towards more clinically-relevant measures of performance (in this case, a function of specificity and sensitivity). SL prediction and cross-validated risk was used to evaluate performance of each model (29).

COMPARISON OF SUPERLEARNER AND STANDARD SEVERITY SCORING METRICS

Feasibility, calculated by present missingness in each dataset, was calculated for standard severity scoring metrics, including Trauma and Injury Severity Score (TRISS) (9), Kampala Trauma Score (KTS) (8), Revised Trauma Score (RTS) (30), Glasgow Coma Scale, Age, Pressure Score (GAP) (6), Mechanism, Age, and Pressure Score (MGAP) (31), and the Glasgow Coma Score (GCS) (32).

Discrimination of hospital mortality using SL was compared with these standard scoring metrics (TRISS, KTS, RTS, GAP, MGAP, and GCS). Given differences in the baseline variables, not all scores could be calculated for each dataset. Specifically, TRISS was only calculated in the US cohort whereas KTS was only calculated in the Cameroonian cohort.

VARIABLE IMPORTANCE MEASURES

Many choices are available for assessing the relative importance of predictor variables in the overall ability of the final model to predict. One such technique is available as a byproduct of another ensemble-type machine learning algorithm called random forest. In this case an average of regression trees is used as the final predictor. Specifically, random forest variable importance is measured as the change in the out-of-sample (samples not used in training) fit of the model when each variable is randomly permuted (and thus made independent of the outcome). We report the five most important predictor variables within each study as chosen by this metric (27, 33). Top influential predictors for hospital mortality after injury were compared across clinical settings.

RESULTS:

DATA AND COHORT CHARACTERISTICS

Data from 28,212 injured patients were evaluated and used to generate prediction. (Table 1) Available admission variables differed by site, with the quantity of variables used for prediction ranging from 133 (Cameroon) to 212 (US).

Cohort composition also varied by dataset (Table 1); however, all three cohorts were predominantly young (median age 29–36) and male (72.4–82.0%). Rates of blunt trauma were lower for the SA and US cohorts (45.6% and 57%) than for the Cameroonian cohort (82.8%). Discharge mortality ranged from 1.28% in the SA cohort to 17.4% in the US Cohort.

FEASIBILITY OF CONVENTIONAL SEVERITY SCORES

Compared to the SL algorithm, which was able to utilize data from all patients at each site, all conventional scores were limited by missing composite patient data, which ranged from 0.94% for patient age in the US cohort to 100% for injury severity score in the Cameroonian cohort and KTS in the US cohort. Overall data missingness for conventional scores was higher for resource-constrained settings compared to the US site. (Table 2)

SL DISCRIMINATION OF MORTALITY

SL demonstrated excellent prediction of discharge mortality in the US cohort (AUC 94–97%) with improved discrimination compared to TRISS and far superior discrimination compared to physiologic scores, including the GAP, MGAP, GCS, and RTS. (Figure 1) KTS could not be calculated in this cohort due to lack of composite score recording of the Glasgow Coma Scale.

Among the SA cohort, SL generated prediction with similar discrimination to conventional scoring using GAP and RTS (AUC 90–95%). (Figure 2) KTS and TRISS could not be calculated for this cohort with the available variables.

SL demonstrated far superior prediction of mortality in the Cameroonian cohort (AUC 90–94%) compared to KTS, GAP, MGAP, GCS, and RTS. (Figure 3) Anatomic scoring was not

available for this cohort given limitations of diagnostic imaging, record-keeping, and operative capacity in this setting.

VARIABLE IMPORTANCE MEASURES

For each cohort, influential drivers of prediction were identified. (Figure 4) Unique, context-specific variables found to drive prediction of mortality in each cohort included admission partial thromboplastin time in the US cohort and distance to the hospital and patient education level in the Cameroonian cohort. Regardless of site differences, indicators of severe head injury, including admission GCS (US, SA) and image-guided or clinician-estimated head injury severity (US, Cameroon), portend high risk of mortality after injury. Advanced patient age was also found to drive prediction of trauma mortality in diverse clinical contexts (US and Cameroon).

DISCUSSION

In this study, we demonstrate that machine learning generates excellent prediction of trauma mortality across extremely diverse clinical settings. Unlike conventional statistical approaches, which rely on relatively complete collection of pre-specified variables, machine learning was able to discriminate mortality on the basis of different baseline admission variables. Additionally, the feasibility of conventional algorithms in all settings was limited by variable missingness. We further demonstrate that variables driving prediction can lend important insights into context-specific and potentially modifiable targets for clinical intervention. These capabilities of machine learning address many of the challenges that have plagued prior attempts at severity scoring and suggest that machine learning could be an adaptable benchmark for establishing optimal prediction to aid clinical decision making and evaluate trauma systems in diverse contexts.

The datasets utilized in this study reflect trauma care in areas with substantial differences in resource availability, patient population, injury patterns, prehospital systems, and trauma practice standard of care. Prior attempts to apply algorithms developed in high-income contexts to LMIC contexts have been limited by feasibility constraints and these underlying differences in patient trajectory in different health systems (34). Lack of reliable scoring has hindered efforts to generate quality improvement programming in LMICs. In high-income countries, clinicians have yet to harness the full power of the vast amounts of patient data available on modern trauma patients.

Machine learning offers potential solutions to many of these obstacles by approaching the problem in a novel fashion. Traditional modeling has attempted to generate a single model capable of assessing all trauma patients. It can be argued that in addition to patient physiology, injured patient outcomes are a reflection of the overall fitness of a trauma system. Moderate injury may confer minimal mortality risk in the US, but it confers a much higher risk in contexts where resources such as blood transfusions, prehospital care, critical care, and other therapeutic interventions are not available. Machine learning does not attempt to find a single model for all patients but rather to identify the risk to a given patient in their given context. Tailored prediction is achievable because machine learning, unlike conventional scoring, is not constrained by the need for algorithm simplicity and is able to

adjust for the complex relationships between many more potential predictors. These features largely explain why SL prediction remains robust despite differences in care.

Although excellent prediction was generated at all sites, the discrimination margin between SL and conventional scoring narrows where mortality prediction is already very good, as seen in the SA cohort. This is because SL will predict future outcomes using the same data-generating distribution as well (asymptotically) *or better* than the best single algorithm in the library of learners (18). If one uses a broad suite of learners, from very smooth, traditional modeling approaches (e.g., logistic regression) to very flexible, nonparametric routines (like random forest), then the cross-validated estimate of fit can provide a standard by which to compare other, simpler competing approaches. In contrast, where traditional prediction is less robust, either because of underfitting when a large number of physiologic parameters are available (US cohort), or where feasibility of advanced imaging and laboratory measurements has limited application (Cameroon), machine learning offers substantial improvements in discrimination with important implications on health systems development.

Importantly, the embedded use of cross-validation within the SL algorithm is the essential component in the generation of reliable context-tailored prediction. It is possible that simplistic models may report discrimination higher than that generated by SL, but as these models do not inherently embed cross-validation, they are naturally prone to overfit (predicting the building data well), and require calibration on a prospective hold-out sample to better gauge performance.

Comparative variable importance reveals conserved and context-variable predictors driving prediction of mortality in different settings. While it is not possible to definitively determine why this is the case, we would hypothesize that this is not due simply to differences in what variables are being measured but also in part due to underlying differences in the patient population, health system, and in care and treatment practices. Distance to the hospital is a top variable importance measure in Cameroon where there is no formal pre-hospital emergency response system, catchment areas are large and treatment delays are often substantial, whereas this measure is not predictive of outcome in the US cohort as transit times within the catchment region of the center analyzed were negligible. Conversely, markers of trauma-induced coagulopathy rank highly in the US cohort. We would expect that, if measured, these markers would also predict mortality in LMIC cohorts. This highlights how, in addition to revealing potentially modifiable context-specific targets for intervention, variable importance generates an evidence-based means of triaging variable inclusion which could potentially add predictive power in resource-constrained settings.

Although this approach appears promising, there are several notable limitations that should be considered. First, machine learning is most appropriately utilized with large and complex datasets. Without adequate size and variability, predictive capacity will be limited. With simple predictor sets, SL may be effective, but may not be necessary and more traditional statistical approaches may yield similar results. Although machine learning can maximize data utilization and address variable missingness, it will not be able to compensate for poor data quality. It is likely that some prediction-driving variables in any context remain unmeasured, and these will limit the prediction capacity of any prospective algorithm,

including SL. Particularly among LMIC settings, acquisition of high quality data inputs will likely remain the most significant barrier to use. Continued progression toward ubiquitous trauma registry data collection on all injured patients will be essential to facilitate improvements in trauma care. Finally, despite being open source, the authors anticipate that trauma clinicians may not routinely have time to download the software and learn a coding language in order to use SL to aid clinical decision making. With this in mind, in order to facilitate point-of-care use in the clinical setting it will be important to develop open-source cloud-based applications of the SL software which allow clinicians to simply upload their trauma dataset in order to obtain context-tailored prediction without needing advanced coding skills.

Currently, the surgical field remains skeptical of black-box approaches like machine learning, as most clinicians and scientists have been trained under a strict paradigm of hypothesis testing based on clinical relevance; data mining, therefore, can seem unfamiliar. Despite their novelty, clinically agnostic approaches offer several other potential advantages. First, SL and similar programs allow for context-tailored prediction and target identification with minimal risk of overfitting. Additionally, these programs can easily be translated into user-friendly open-source applications with the potential for rapid, point-of-care use in diverse contexts. The algorithms are adaptable over time; in contexts with longitudinal and streaming data capacity, they could establish a foundation for real-time individualized decision support. Finally, it is important to note that this method can be readily adapted for clinically-relevant outcomes other than trauma mortality, providing great potential for augmenting clinical decision to optimize performance.

CONCLUSIONS

Machine learning provides superior prediction of mortality after injury in diverse clinical contexts offering discrimination without the need for sophisticated diagnostic data or a common variable set. It is readily scalable and can be used to identify site-specific factors that drive prediction, showing potential as a benchmark for outcomes scoring and risk stratification to improve injury care.

Acknowledgments

Funding: Supported by PCORI R-IMC-1306-02735 (MJC), NIH #K01ES026834 (RAC).

REFERENCES

1. Lefering R Trauma scoring systems. *Curr Opin Crit Care*. 2012;18(6):637–40. [PubMed: 22918259]
2. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Keast SL, Bain LW, Jr, Flanagan ME, Frey CF The Major Trauma Outcome Study: establishing national norms for trauma care. *J Trauma*. 1990;30(11):1356–65. [PubMed: 2231804]
3. Baker SP, O'Neill B, Haddon W, Jr., Long WB The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14(3):187–96. [PubMed: 4814394]
4. Civil ID, Schwab CW. The Abbreviated Injury Scale, 1985 revision: a condensed chart for clinical use. *J Trauma*. 1988;28(1):87–90. [PubMed: 3339667]

5. Copes WS, Lawnick M, Champion HR, Sacco WJ. A comparison of Abbreviated Injury Scale 1980 and 1985 versions. *J Trauma*. 1988;28(1):78–86. [PubMed: 3339666]
6. Kondo Y, Abe T, Kohshi K, Tokuda Y, Cook EF, Kukita I. Revised trauma scoring system to predict in-hospital mortality in the emergency department: Glasgow Coma Scale, Age, and Systolic Blood Pressure score. *Crit Care*. 2011;15(4):R191. [PubMed: 21831280]
7. 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. [PubMed: 2657085]
8. Kobusingye OC, Lett RR. Hospital-based trauma registries in Uganda. *J Trauma*. 2000;48(3):498–502. [PubMed: 10744292]
9. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. *J Trauma*. 1987;27(4):370–8. [PubMed: 3106646]
10. Champion HR, Copes WS, Sacco WJ, Frey CF, Holcroft JW, Hoyt DB, Weigelt JA. Improved predictions from a severity characterization of trauma (ASCOT) over Trauma and Injury Severity Score (TRISS): results of an independent evaluation. *J Trauma*. 1996;40(1):42–8; discussion 8–9. [PubMed: 8576997]
11. Chawda MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: which scoring system? *Injury*. 2004;35(4):347–58. [PubMed: 15037369]
12. Bruijns SR, Wallis LA. The Kampala Trauma Score has poor diagnostic accuracy for most emergency presentations. *Injury*. 2017;48(10):2366–7. [PubMed: 28855083]
13. Akay S, Ozturk AM, Akay H. Comparison of modified Kampala trauma score with trauma mortality prediction model and trauma-injury severity score: A National Trauma Data Bank Study. *Am J Emerg Med*. 2017;35(8):1056–9. [PubMed: 28222915]
14. Haac B, Varela C, Geyer A, Cairns B, Charles A. The utility of the Kampala trauma score as a triage tool in a sub-Saharan African trauma cohort. *World J Surg*. 2015;39(2):356–62. [PubMed: 25315093]
15. Houssaini A, Assoumou L, Marcelin AG, Molina JM, Calvez V, Flandre P. Investigation of Super Learner Methodology on HIV-1 Small Sample: Application on Jaguar Trial Data. *AIDS Res Treat*. 2012;2012:478467. [PubMed: 22550568]
16. Petersen ML, LeDell E, Schwab J, Sarovar V, Gross R, Reynolds N, Haberer JE, Goggin K, Golin C, Arnsten J, et al. Super Learner Analysis of Electronic Adherence Data Improves Viral Prediction and May Provide Strategies for Selective HIV RNA Monitoring. *J Acquir Immune Defic Syndr*. 2015;69(1):109–18. [PubMed: 25942462]
17. CM B Pattern Recognition And Machine Learning. New York: Springer; 2006. 740 p.
18. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
19. Hubbard A, Munoz ID, Decker A, Holcomb JB, Schreiber MA, Bulger EM, Brasel KJ, Fox EE, del Junco DJ, Wade CE, et al. Time-dependent prediction and evaluation of variable importance using superlearning in high-dimensional clinical data. *J Trauma Acute Care Surg*. 2013;75(1 Suppl 1):S53–60. [PubMed: 23778512]
20. Diaz I, Hubbard A, Decker A, Cohen M. Variable importance and prediction methods for longitudinal problems with missing variables. *PLoS One*. 2015;10(3):e0120031. [PubMed: 25815719]
21. Cohen MJ, Bir N, Rahn P, Dotson R, Brohi K, Chesebro BB, Mackersie R, Carles M, Wiener-Kronish J, Pittet JF. Protein C depletion early after trauma increases the risk of ventilator-associated pneumonia. *J Trauma*. 2009;67(6):1176–81. [PubMed: 20009664]
22. Cohen MJ, Brohi K, Calfee CS, Rahn P, Chesebro BB, Christiaans SC, Carles M, Howard M, Pittet JF. Early release of high mobility group box nuclear protein 1 after severe trauma in humans: role of injury severity and tissue hypoperfusion. *Crit Care*. 2009;13(6):R174. [PubMed: 19887013]
23. Norman R, Matzopoulos R, Groenewald P, Bradshaw D. The high burden of injuries in South Africa. *Bull World Health Organ*. 2007;85(9):695–702. [PubMed: 18026626]
24. Nicol A, Knowlton LM, Schuurman N, Matzopoulos R, Zargarani E, Cinnamon J, Fawcett V, Taulu T, Hameed SM. Trauma Surveillance in Cape Town, South Africa: An Analysis of 9236 Consecutive Trauma Center Admissions. *JAMA Surg*. 2014;149(6):549–56. [PubMed: 24789507]

25. RCoreTeam. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013
26. Hastie T, Tibshirani R. Generalized Additive Models: An Introduction With R. New York, NY: Chapman and Hall; 1990.
27. Breiman L Random Forests- Random Features. Berkeley, Ca : Department of Statistics, University of California; 1999.
28. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1–22. [PubMed: 20808728]
29. Dudoit SavdL, Mark J Asymptotics of Cross-Validated Risk Estimation in Estimator Selection and Performance Assessment. *Stat Meth.* 2005;2(2):131–54.
30. Gilpin DA, Nelson PG. Revised trauma score: a triage tool in the accident and emergency department. *Injury.* 1991;22(1):35–7. [PubMed: 2030028]
31. Sartorius D, Le Manach Y, David JS, Rancurel E, Smail N, Thicoipe M, Wiel E, Ricard-Hibon A, Berthier F, Gueugniaud PY, et al. Mechanism, glasgow coma scale, age, and arterial pressure (MGAP): a new simple prehospital triage score to predict mortality in trauma patients. *Crit Care Med.* 2010;38(3):831–7. [PubMed: 20068467]
32. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet.* 1974;2(7872):81–4. [PubMed: 4136544]
33. Variable HI importance in Binary regression trees and forests. *Electron J Stat.* 2007:519–37.
34. O'Reilly GM, Joshipura M, Cameron PA, Gruen R. Trauma registries in developing countries: a review of the published experience. *Injury.* 2013;44(6):713–21. [PubMed: 23473265]

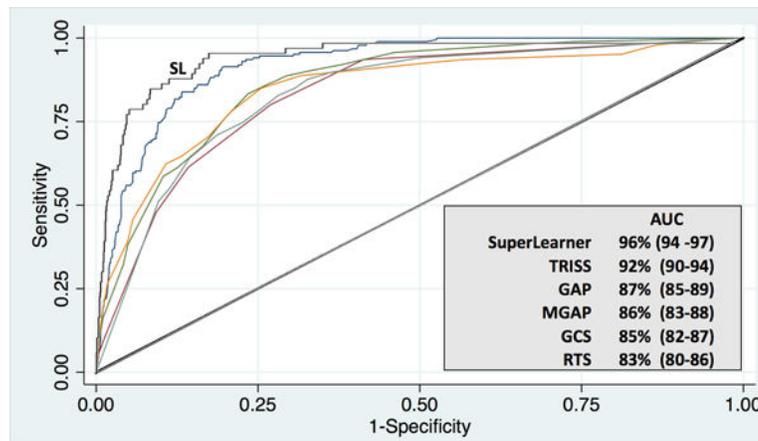


Figure 1: Comparative prediction of mortality in a US trauma cohort. (N=1,494). Receiver operating characteristic curves demonstrating discrimination of mortality using a machine learning algorithm (SuperLearner) compared with conventional severity scoring in a single center, highest activation US trauma cohort. Data in parentheses represent 95% confidence intervals. AUC, Area Under the Curve; US, United States; SL, SuperLearner; TRISS, Trauma and Injury Severity Score; RTS, Revised Trauma Score; GAP Glasgow Coma Scale, Age, Pressure Score; MGAP, Mechanism, Glasgow Coma Scale, Age, and Pressure Score; and GCS, Glasgow Coma Scale.

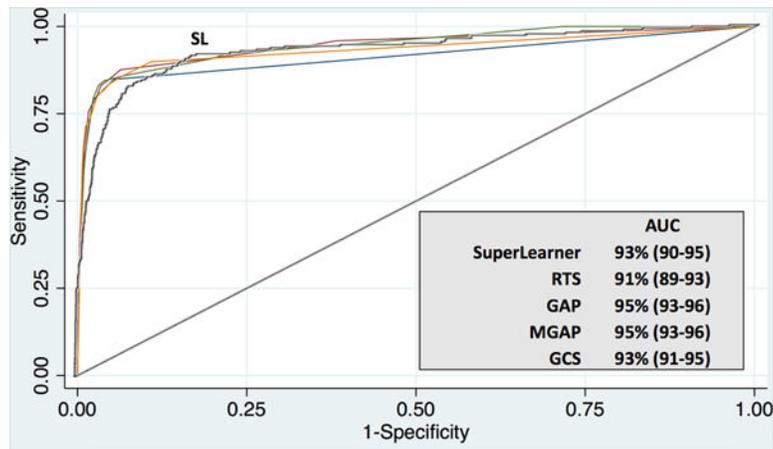


Figure 2: Comparative prediction of mortality in a South African trauma cohort. (N=18,821). Receiver operating characteristic curves demonstrating discrimination of mortality using a machine learning algorithm (SuperLearner) compared with conventional severity scoring in a high-volume urban South African trauma cohort. Data in parentheses represent 95% confidence intervals. AUC, Area Under the Curve; SA, South Africa; SL, SuperLearner; RTS, Revised Trauma Score; GAP Glasgow Coma Scale, Age, Pressure Score; MGAP, Mechanism, Glasgow Coma Scale, Age, and Pressure Score; and GCS, Glasgow Coma Scale.

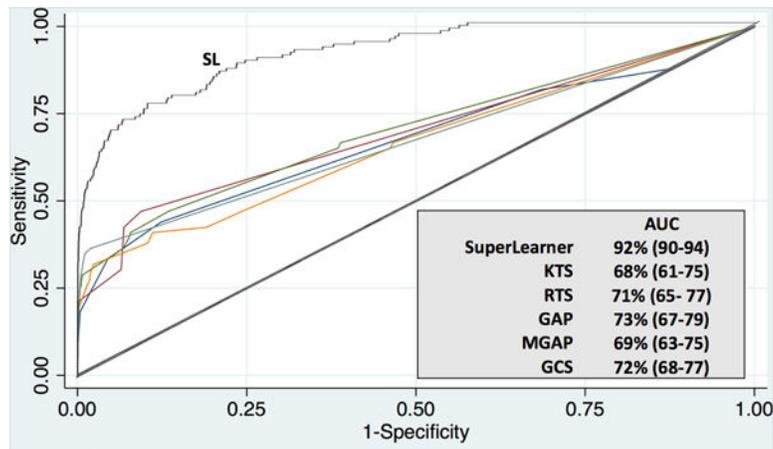


Figure 3: Comparative prediction of mortality in a multisite Cameroonian trauma cohort. (N=7,897).

Receiver operating characteristic curves demonstrating discrimination of mortality using a machine learning algorithm (SuperLearner) compared with conventional severity scoring in a multi-center Cameroonian trauma cohort. Data in parentheses represent 95% confidence intervals. AUC, Area Under the Curve; SL, SuperLearner; KTS, Kampala Trauma Score; RTS, Revised Trauma Score; GAP, Glasgow Coma Scale, Age, Pressure Score; MGAP, Mechanism, Glasgow Coma Scale, Age, and Pressure Score; and GCS, Glasgow Coma Scale.

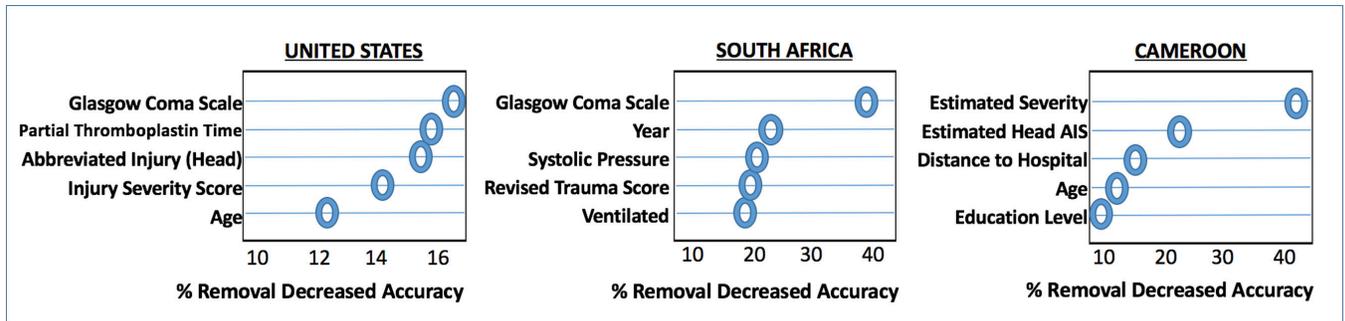


Figure 4: Context-specific Clinical Admission Variables Driving SuperLearner Prediction of Mortality.

Top variable importance measures by context using random forest estimation. Estimated severity, ordinal ranking of perceived injury severity by provider at patient presentation (1–6); AIS, Abbreviated Injury Score.

Table 1:

Dataset and Trauma Cohort Characteristics

	US Cohort <i>N=1,494</i>	SA Cohort <i>N=18,821</i>	Cameroon Cohort <i>N=7,897</i>
Data Characteristics			
Data Collection (years)	2005–2015	2014–2016	2015–2017
Medical Centers (n)	1	1	3
Admission Predictor Variables (n)	212	144	131
Observations (n)	853,074	17,416,200	2,574,422
Cohort Characteristics			
Age (years, IQR)	36 (25–52)	30 (24–32)	29 (22–40)
Pediatric Patients (%)	0	5.1	14.2
Male Patients (%)	82.0	75.6	72.4
Blunt Mechanism (%)	57.0	45.6	82.8
Discharge Mortality (%)	17.4	1.28	1.74

Age is presented as median (IQR), all other variables reported as specified. US, United States Cohort; SA, South African Cohort.

Table 2:

Data Missingness by Trauma Severity Score and Cohort (%)

	US Cohort <i>N=1,494</i>	SA Cohort <i>N=18,821</i>	Cameroon Cohort <i>N=7,897</i>
SuperLearner	0%	0%	0%
Kampala Trauma Score	100%	0%	46.6%
Revised Trauma Score	6.2%	6.8%	46.5%
Trauma and Injury Severity Score	9.4%	0%	100%
GCS, Age, Pressure Score	3.4%	8.3%	20.6%
Mechanism, GCS, Age, Pressure Score	3.7%	9.4%	24.2%
Glasgow Coma Score	1.4%	4.7%	3.9%

US, United States Cohort; SA, South African Cohort; GCS, Glasgow Coma Score.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript