Machine Learning in the Prediction of Trauma Outcomes: A Systematic Review

Timothy Zhang*; Anton Nikouline, MD; David Lightfoot, PhD; Brodie Nolan, MD, MSc

*Corresponding Author. E-mail: timothymed.zhang@mail.utoronto.ca.

Study objective: Machine learning models carry unique potential as decision-making aids and prediction tools for improving patient care. Traumatically injured patients provide a uniquely heterogeneous population with severe injuries that can be difficult to predict. Given the relative infancy of machine learning applications in medicine, this systematic review aimed to better understand the current state of machine learning development and implementation to help create a basis for future research.

Methods: We conducted a systematic review from inception to May 2021, using Embase, MEDLINE through Ovid, Web of Science, Google Scholar, and relevant gray literature, for uses of machine learning in predicting the outcomes of trauma patients. The screening and data extraction were performed by 2 independent reviewers.

Results: Of the 14,694 identified articles screened, 67 were included for data extraction. Artificial neural networks comprised the most commonly used model, and mortality was the most prevalent outcome of interest. In terms of machine learning model development, there was a lack of studies that employed external validation, feature selection methods, and performed formal calibration testing. Significant heterogeneity in reporting was also observed between the machine learning models employed, patient populations, performance metrics, and features employed.

Conclusion: This review highlights the heterogeneity in the development and reporting of machine learning models for the prediction of trauma outcomes. While these models present an area of opportunity as an ancillary to clinical decision-making, we recommend more standardization and rigorous guidelines for the development of future models. [Ann Emerg Med. 2022;■:1-16.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2022.05.011

INTRODUCTION

Background

The high rate of complications, health care costs, and overall burden to health care systems associated with trauma have prompted the development of various tools for predicting trauma patient outcomes, such as the Injury Severity Scale (ISS) and Trauma and Injury Severity Score (TRISS).[1,2] However, these tools were developed using outdated data sets and limited in the scope of data that could be processed at the time.[2] For this reason, some clinicians have raised concerns about the ability of the existing tools to provide accurate clinical predictions in contemporary trauma care.[3]

Machine learning provides an exciting opportunity in the development of sophisticated prediction models using large, complex data sets. Since its introduction to the medical field in the 1950s, machine learning's viability and potential uses have only grown with modern advancements in computing power.[4] The potential of machine learning in trauma-related prediction modeling is predicated on its ability to harness the vast amounts of patient-related data collected.

Compared with more traditional methods, machine learning is better able to handle large sets of data, particularly those that are unstructured or nonlinear or contain missing values.[5,6] Machine learning models have been able to predict which patients would develop severe sepsis in the future and who would deteriorate and require admission to the ICU.[7] Several studies have shown that predictive models using machine learning have outperformed traditional predictive tools and even field experts in some instances.[7,8] Due to the complexity and heterogeneity of trauma, predicting outcomes is difficult, and the potential complications are devastating for a largely young population; therefore, better prediction of trauma outcomes is paramount.

Importance

Despite their benefits, machine learning models still carry caveats. Because typical machine learning models are

Editor's Capsule Summary

What is already known on this topic

Machine learning has developed rapidly with many attempts to develop clinical decision tools and other forms of real-time decision support.

What question this study addressed

What does the published literature reveal about machine learning decision support predicting outcomes in trauma care?

What this study adds to our knowledge

Machine learning studies are too heterogenous to allow meta-analysis. Model development is varied and study quality either limited or indeterminate.

How this is relevant to clinical practice

While heralded and exciting, machine learning remains in its infancy in application to trauma care. Future uptake will likely require improved reporting standardization.

able to self-adjust as they "learn," this may lead to a "black-box" phenomenon, where the creators of the model may not be entirely certain of its operational methods. A machine learning model may also be susceptible to "overfitting," where a model's logic is too tightly conformed to its training data set and it performs poorly when introduced to new data.

A number of systematic reviews have already explored the use of machine learning in clinical prediction tools.[9-13] Liu and Salinas[6] conducted a 2017 systematic review investigating the potential of machine learning in predicting outcomes of traumatically injured patients and stated that further research was needed to establish common performance criteria. Since that time, although a large number of studies have been published, there has been little prospective implementation of machine learning models despite their potential utility in the clinical setting. As such, machine learning in health care remains in a relatively infantile stage. A lack of consistency across research reporting and standards may be hampering efforts to demonstrate the benefits and applications of machine learning.

Goals of This Investigation

We conducted this systematic review to elaborate on previous work, specifically focusing on performance criteria and feature selection. Our objective was to survey the studies that have used machine learning for predicting trauma outcomes and identify the methods used to provide insight for further standardization of machine learning research.

METHODS

We performed a systematic review according to the guidelines of the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) (Appendix E1, available at http://www.annemergmed.com). The study was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42021248580). The systematic review was conducted and reported according to the PRISMA statements.[14] Research ethics approval and informed consent were not required.

Eligibility Criteria and Outcomes

English articles using modern machine learning models and trauma outcomes were included, with no date restrictions, from any country of origin. Letters, editorials, case reports, case series, and review articles were excluded. Examples of machine learning models included artificial neural networks, support vector machines, and random forests. For our purposes, methods such as logistic regression and multivariate logistic regression were considered as statistical methods and not included. Clinically relevant outcomes included in-hospital mortality, the need for clinical interventions, and the development of complications. Studies that primarily focused on patients aged less than 16, orthopedic-only trauma, psychiatric trauma and outcomes, or out-of-hospital prediction (ie, need for hospital transport) were excluded. Articles centered on evaluating diagnostic modalities, such as applying machine learning–based computer vision to identify pathology on medical imaging, were also excluded. Finally, articles for which full-text versions could not be retrieved (eg, the author did not respond to a request for a full-text article, or a conference abstract existed without a full-text study) were also excluded.

Machine Learning Techniques

Artificial neural networks attempt to model human cognition through networks of intersignaling processing nodes, yielding summed probabilities to drive predictions. Support vector machines are known as a kernel method, which virtually projects data in multiple dimensions and generates "dividers," known as hyperplanes, to separate the data to create classifications.[15] A decision tree consists of a series, or "tree," of hierarchical binary nodes that are

organized based on their relative importance to the data and the given outcome. This makes decision trees sometimes more interpretable than other algorithms, such as artificial neural networks and support vector machines, but they can be very sensitive to small changes in training data and prone to overfitting.[16] A Bayesian network is a probabilistic graphic model that represents data through a set of variables and their conditional dependencies with a directed graph. Unlike in an artificial neural network, each input variable is assumed to be independent, and the accumulation of probabilities from all variables and their outcomes forms the final prediction. Random forests make predictions by creating hundreds or thousands of smaller decision trees (hence, "forest") and combining each of their outputs. This method helps combat the tendency of decision trees to overfit.[15]

Search Strategy

A search was conducted in consultation with an information specialist, among 4 databases (Embase, MEDLINE through Ovid, Web of Science, Google Scholar), from inception to May 14, 2021, alongside a gray-literature search reviewing trauma conference proceedings and clinical trial registries (eg, ClinicalTrials.gov) to capture all eligible articles that originated from both medical and technological backgrounds. The search terms used included subject headings and keywords associated with machine learning models and trauma. The bibliographies of identified articles were also reviewed for additional relevant articles. The full search strategy for MEDLINE is available in Appendix E2 (available at http://www.annemergmed.com).

Data Extraction and Analysis

The screening and subsequent data extraction processes were facilitated by Covidence (Veritas Health Information, 2016), a literature review streamlining software recommended by Cochrane. All identified articles underwent title and abstract screening by 2 independent reviewers (TZ, AN) with an interrater reliability of 98.5% and a Cohen kappa of 0.51. The full texts of relevant articles were then similarly screened. Screening conflicts were resolved by a third reviewer (BN). Bias and applicability assessments of the final included articles were performed by 2 independent reviewers (TZ, AN) with guidance from the Prediction Model Risk of Bias Assessment Tool (PROBAST), as described in previous systematic reviews of machine learning prediction models.[17]

A standardized form was used for data extraction and included the following fields: article title, first author, country of study, participant data collection type, model development or validation, patient population category, development cohort, derivation and internal validation, external validation, clinical outcome(s), total patients, percentage of males, average age, number of patients with outcome, number of model features considered, feature extraction method, final model features used, most important model features, machine learning algorithms used, presence of a comparator, and various model performance metrics (eg, area under the receiver operating characteristic curve [AUROC], accuracy, sensitivity, specificity).

Excel (Microsoft, 2021) was used to facilitate data organization and generate descriptive statistics of the extracted data. Individual machine learning models were classified by their broader types, and the specific clinical outcomes were classified into 6 categories. The features of each machine learning model were aggregated to identify common themes.

The heterogeneity of patient populations, clinical outcomes, and machine learning models made a meta-analysis infeasible.

RESULTS

Search Results

The literature search identified 14,694 studies. Following the removal of duplicates and addition of 15 articles through reference searches, 7,349 articles were screened by title and abstract. Full-text screening was performed for 95 articles, resulting in 67 studies that were eligible for data extraction and analysis. The full-text exclusion criteria included a lack of outcome predictions (n=12), not involving trauma patients (n=5), and a lack of clinically relevant outcomes (n=5). The details of the full screening process are described in Figure 1.

Characteristics of Study Subjects

Table 1 summarizes some key characteristics of the studies included in this review, with an expanded summary in Appendix E3 (available at http://www.annemergmed.com). The studies that were identified primarily originated from the United States (n=32, 48%) and countries in Europe and Asia. The publication dates ranged from 1993 to 2021, with nearly half (n=30, 45%) published after 2017. Ten (15%) studies used prospective data, 54 (82%) used retrospective data, and 2 (3%) employed both types. Most studies acquired their data from local electronic health records (n=41, 61%), followed by databases (n=19, 28%) and national electronic health records (n=6, 9%), with one (1%) study not specifying its data source. Excluding the 8 (12%) studies that were already focused on model validation, 8 (12%) studies included explicit external

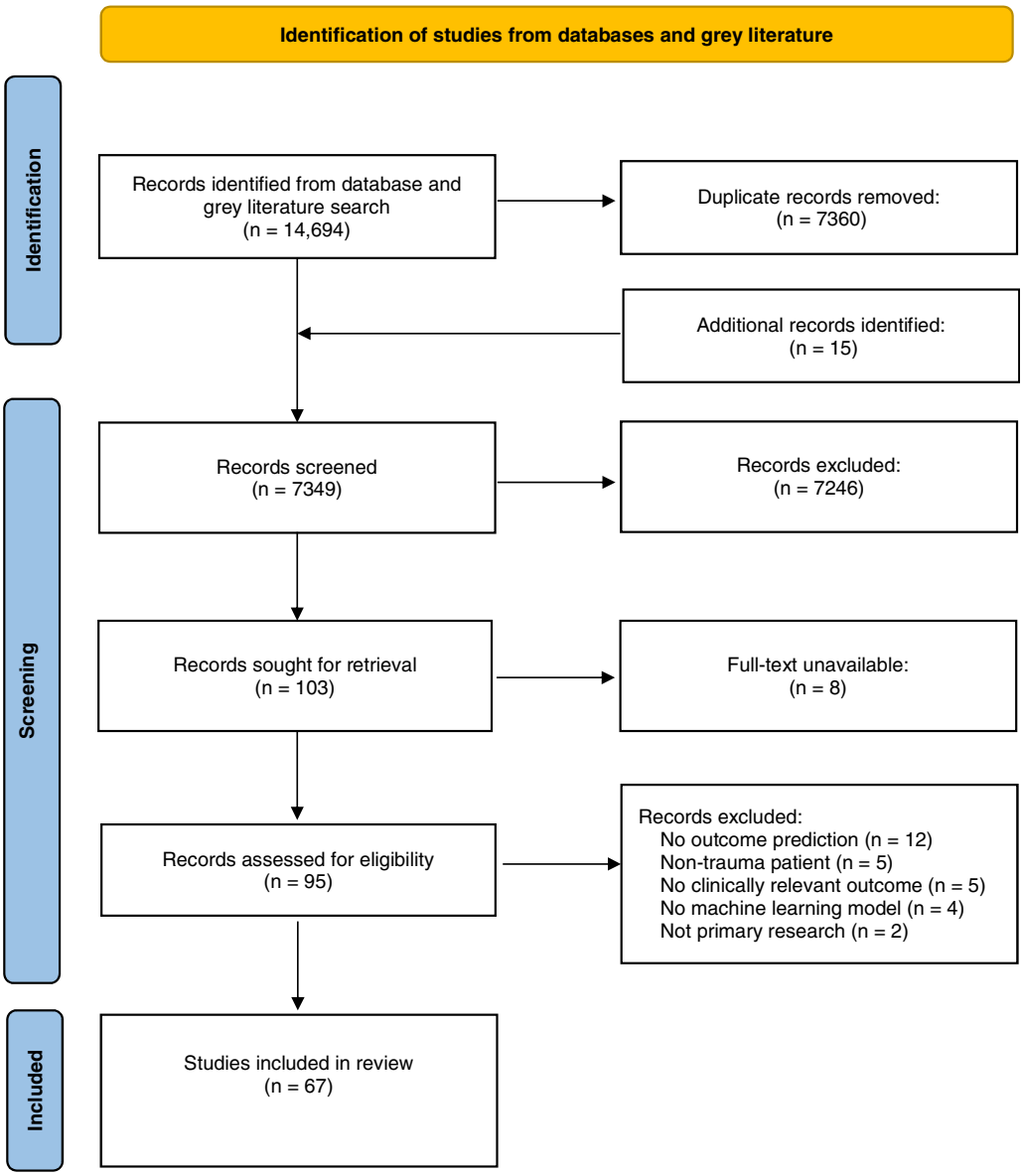

Figure 1. PRISMA flow diagram of the selection process for included studies.

validation in addition to internal validation in the development of their model. The articles were separated by focus on either multisystem trauma (n=42, 61%) or traumatic brain injury (n=25, 37%). Clinical outcomes involving mortality were the most common (n=38, 57%), followed by complications (n=18, 26%), functional outcomes (n=14, 20%), interventions (n=5, 7%), length of stay (n=5, 7%), and finances (n=1, 1%). A graphic summary of some key results is shown in Figure 2.

Main Results

The number of features considered for model development by various studies varied greatly, from 7 to 272. Out of the minority of articles that employed objective methods to extract the most important features (n=25, 36%), these methods included machine learning models, such as random forests, least absolute shrinkage and selection operator (LASSO), and elastic net regression, as well as statistical methods, including multivariate logistic regression, Fisher exact tests, and Pearson correlation coefficients. Appendix E4 (available at http://www.annemergmed.com) shows the features used by each of the machine learning models and, where available, up to 5 of their respective most important variables. The overall median number of features used was 9 (interquartile range 5 to 14), with similar medians in multisystem trauma and traumatic brain injury models, at 10 (interquartile range 5

Table 1. Baseline characteristics of included studies.

Study Reference	Data Collection Type	Study Type	Data Source	External Validation	Clinical Outcome Category	Machine Learning Algorithm(s)
Patients with Traumatic Brain Injuries						
Abbod 2011[18]	Retrospective	Development	Local EHR	No	Mortality	ANN
Abujaber 2020 (1)[8]	Retrospective	Development	Local EHR	No	Mortality	ANN, SVM
Abujaber 2020 (2)[19]	Retrospective	Development	Local EHR	No	Complications	RF, ANN, DT, SVM
Abujaber 2020 (3)[41]	Retrospective	Development	Local EHR	No	Mortality	ANN
Donald 2019[21]	Prospective	Validation	Database (BrainIT multicenter database)	N/A	Complications	BN
Eftekhar 2005[22]	Retrospective	Development	Multiple local EHRs	No	Mortality	ANN
Feng 2019[23]	Retrospective	Validation	Local EHR	N/A	Mortality	SVM, KNN, DT
Gravesteijn 2020[24]	Retrospective	Validation	Database (IMPACT-2, CENTER-TBI)	N/A	Mortality, functional outcome	LASSO, SVM, ANN, RF, gradient boosting
Hsu 2005[45]	Retrospective	Development	National EHR	No	Functional outcome	ANN
Kalpakis 2015[26]	Retrospective	Development	Local EHR	No	Mortality, functional outcome	Permutation entropy
Li 2000[27]	Retrospective	Development	National EHR	No	Intervention	ANN,
Lu 2015[28]	Retrospective	Development	Local EHR		Mortality, functional outcome	ANN, BN, DT
Matsuo 2020[29]	Retrospective	Development	Local EHR	No	Mortality, functional outcome	LASSO, RF, gradient boosting, DT, BN, SVM
Nikiforidis 1998[30]	Retrospective	Development	Local EHR	No	Functional outcome	BN
Pang 2007[31]	Prospective	Development	Local EHR	No	Functional outcome	DT, BN, ANN, discriminant analysis
Davis 2008[32]	Retrospective	Development	Local EHR	No	Mortality	ANN, SVM, DT
Lang 1997[33]	Prospective	Development	Local EHR	No	Mortality	ANN
Pourahmad 2016[34]	Retrospective	Development	Local EHR	No	Functional outcome	DT, ANN
Pourahmad 2019[35]	Retrospective	Validation	Local EHR	N/A	Functional outcome	Forward selection, MRMR, genetic algorithm
Rau 2018[36]	Retrospective	Development	Local EHR	No	Mortality	SVM, DT, BN, ANN
Rughani 2010[37]	Retrospective	Development	Database (National Trauma Data Bank)	No	Mortality	ANN, BN
Segal 2006[38]	Retrospective	Development	Database (Traumatic Brain Injury Model Systems database)	No	Functional outcome	ANN, DT
Shi 2013[39]	Retrospective	Development	National EHR	No	Mortality	ANN
van der Ploeg 2016[40]	Retrospective	Development	Database (IMPACT database)	Yes	Mortality	DT, RF, SVM, ANN
Vath 2000[41]	Retrospective	Development	Not stated	No	Functional outcome	ANN

Table 1. Continued.

Study Reference	Data Collection Type	Study Type	Data Source	External Validation	Clinical Outcome Category	Machine Learning Algorithm(s)
Patients with Multisystem Trauma						
Ahmed 2020[42]	Retrospective	Development	Database (MIMIC 3)	No	Mortality	BN
Becalick 2001[43]	Retrospective	Validation	Local EHR	N/A	Mortality	ANN
Bektas 2008[44]	Prospective	Development	Local EHR	No	Complications	ANN
Bradley 2020[45]	Retrospective	Development	Local EHR	No	Complications	RF
Bravo-Merodio 2019[46]	Prospective	Development	Dataset (Brain Biomarkers After Trauma Cohort Study)	No	Complications	Elastic net, LASSO
Chen 2013[47]	Retrospective	Development	National EHR	No	Complications	SVM
Christie 2018[48]	Prospective	Validation	Local EHR (United States, South Africa), national EHR (Cameroon)	N/A	Mortality	General additive models, RF, LASSO
Christie 2019[49]	Prospective	Validation	Dataset (Activation of Coagulation and Inflammation in Trauma)	N/A	Mortality, complications, intervention	RF, LASSO, generalized additive models
DiRusso 2000[50]	Retrospective	Development	Regional EHR	No	Mortality	ANN
Gelbard 2019[51]	Prospective	Development	Local EHR	No	Complications	RF
Gholipour 2015[52]	Retrospective	Development	Local EHR	No	Mortality, length of stay	ANN
Gorczyca 2019[53]	Retrospective	Development	Database (National Trauma Data Bank, Nationwide Readmission Database)	Yes	Mortality	BN, RF, ANN, gradient boosting
Hadzikadic 1996[54]	Retrospective	Development	Local EHR	No	Mortality	Concept formation
Hertz 2020[55]	Retrospective	Development	Local EHR	No	Complications	DT, BN, SVM, KNN
Hirshberg 2002[56]	Retrospective	Development	Local EHR	Yes	Intervention	ANN
Hubbard 2013[57]	Prospective	Development	Multiple local EHRs	No	Mortality	Generalized additive models, BN, LASSO, RF
Hunter 2000[58]	Retrospective	Development	National EHR	No	Mortality	ANN
Ji 2009[59]	Retrospective	Development	Regional EHR, Database (National Trauma Data Bank)	No	Mortality, functional outcome, length of stay	DT, adaptive boost, SVM
Kim 2018[60]	Retrospective	Development	Database (National Trauma Data Bank)	No	Mortality	RF, ANN
Fann 2007[61]	Retrospective	Development	Local EHR	No	Complications	ANN
Kuo 2018[62]	Retrospective	Development	Local EHR	No	Mortality	SVM, DT
Li 2020[63]	Retrospective	Development	Local EHR	No	Complications	RF
Liu 2014 (1)[64]	Retrospective	Development	Local EHR	No	Intervention	ANN
Liu 2014 (2)[65]	Retrospective & prospective	Development	Database (Trauma Vitals Database, Wireless Vital Signs Monitor trial)	No	Intervention	DT, SVM, ANN
Marble 1999[66]	Retrospective	Development	Database (TRACS database)	No	Complications	ANN
Maurer 2021[1]	Retrospective	Development	Database (ACS-TQIP database)	No	Mortality, complications	DT

McGonigal 1993[67]	Retrospective	Development	Local EHR	No	Mortality	ANN
Moinadini 2019[89]	Prospective	Development	Local EHR	No	Complications	ANN
Paydar 2021[90]	Retrospective	Development	Local EHR	No	Functional outcome, complications	SVM, KNN, ANN, adaptive boosting
Pearl 2006[70]	Retrospective	Development	Local EHR	No	Mortality	ANN
Pearl 2008[71]	Retrospective	Development	Database (National Trauma Data Bank)	No	Mortality	ANN
Pearl 2009[72]	Retrospective	Development	Database (National Trauma Data Bank)	No	Complications	ANN
Perkins 2020 (1)[73]	Retrospective	Development	Databases (US Department of Defense Trauma Registry, UK-JTTR)	Yes	Functional outcome	BN
Perkins 2020 (2)[74]	Retrospective & prospective	Development	Multiple Local EHRs	Yes	Complications	BN
Rutledge 1995[75]	Retrospective	Development	Regional EHR	No	Mortality	ANN
Rutledge 1998[76]	Retrospective	Development	Database (Health Care Policy Research's Health Care Utilization Project database)	No	Mortality	ANN
Schetinin 2018[98]	Retrospective	Development	Database (US National Trauma Data Bank)	No	Mortality	BN, DT
Servia 2020[78]	Retrospective	Development	Database (RETRAUCI database)	No	Mortality	ANN, DT, BN, adaptive boosting, RF, other
Staziaki 2021[2]	Retrospective	Development	Local EHR	No	Functional outcome, length of stay	ANN, SVM
Tsiklidis 2020[79]	Retrospective	Development	Database (National Trauma Data Bank)	Yes (Trauma Quality Programs participant use file—648,192 patient records)	Mortality	gradient boosting
Tsurumi 2020[80]	Retrospective	Development	Dataset (Inflammation and Host Response to Injury [Glue Grant] cohort)	Yes	Complications	LASSO, ANN
Wolfe 2006[81]	Retrospective	Both	Local EHR	Yes	Length of stay, mortality	DT, ANN

ACS-TQIP, American College of Surgeons-Trauma Quality Improvement Program; *ANN*, artificial neural network; *BN*, Bayesian network; *CENTER-TBI*, Collaborative European NeuroTrauma Effectiveness Research in traumatic brain injury; *DT*, decision tree; *EHR*, electronic health record; *IMPACT*, International Mission for Prognosis and Analysis of Clinical Trials in TBI; *IMPACT-2*, International Mission for Prognosis and Analysis of Clinical Trials in TBI-2; *KNN*, K-nearest neighbor; *MIMIC 3*; Medical Information Mart for Intensive Care; *MRMR*, maximum relevance — minimum redundancy; *RETRAUCI*, Spanish Intensive Care Unit Trauma Registry; *RF*, random forest; *SVM*, support vector machine; *TRACS*, National Trauma Registry System (US); *UK-JTTR*, UK joint theatre trauma register.

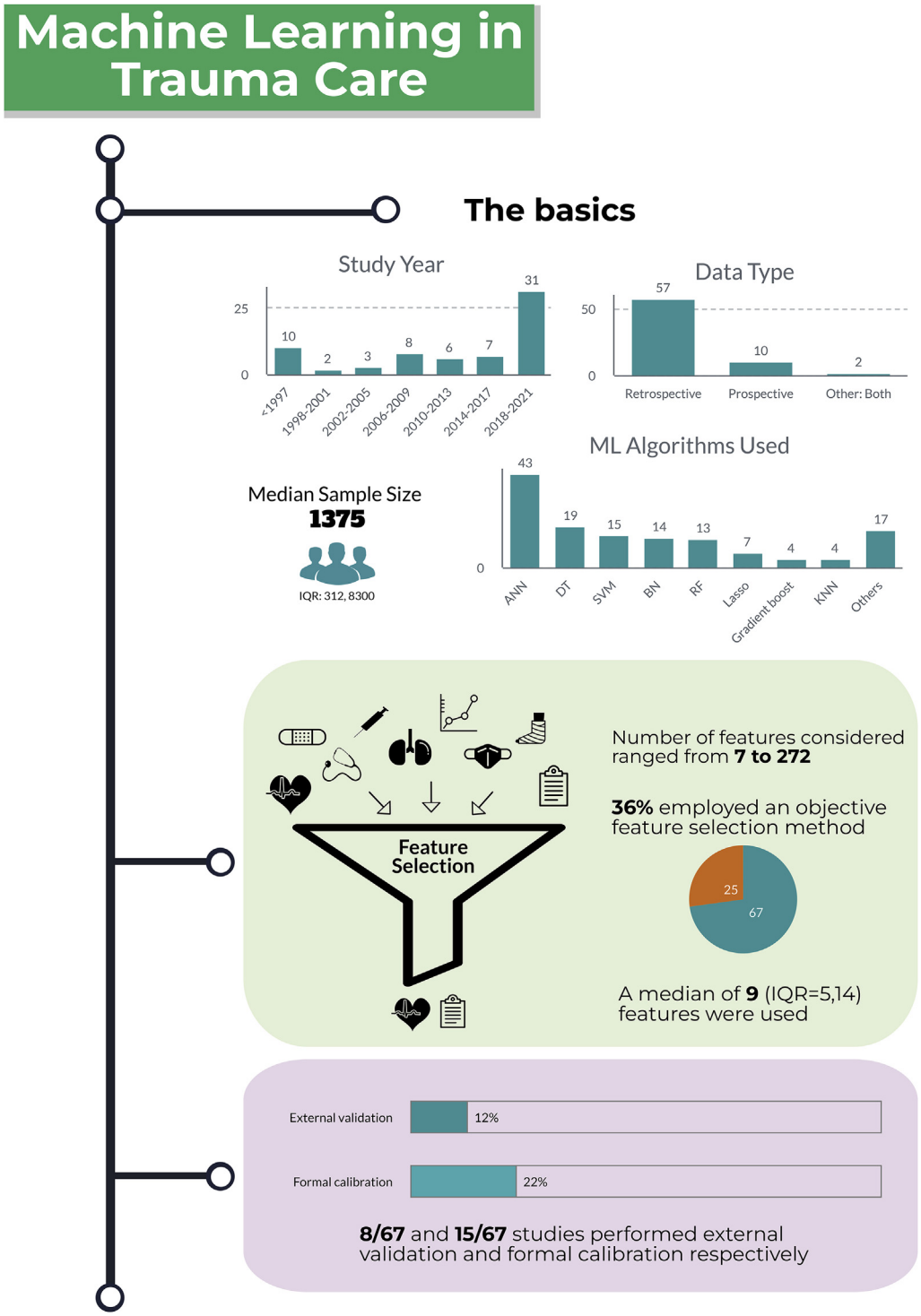

Figure 2. Infographic of select key study results.

to 14) and 8.5 (interquartile range 6 to 11.25), respectively. The Glasgow Coma Scale (GCS) score, age, and various injury severity scores were the most commonly included. Building upon this, Table 2 consists of a database of all features used by at least 2 of the included studies for predicting mortality in patients with traumatic brain injuries or multi-system trauma.

A summary of all studies' machine learning models and their respective performance metrics are presented in Appendix E5 (available at http://www.annemergmed.com).

Table 2. Features used by at least 2 studies to predict mortality in patients with either traumatic brain injury or multisystem trauma.

TBI	Multisystem Trauma
Age	Age
Blood transfusion	Age above 55
Body temperature	AIS abdomen
CT findings	AIS chest
Diabetes mellitus	AIS extremity
GCS	AIS head
GCS motor	Body temperature
Glucose	Coagulopathy
Head AIS	Comorbidities
Heart rate	Creatinine
Hematoma presence	Hemoglobin
Intubation	GCS
Injury Severity Score	GCS eye
Length of stay	GCS motor
Mode of transport	GCS verbal
Mechanism of injury	Glucose
Presence of other complications	Hematocrit
Pupillary response	Heart rate
Sex	ICD-9 code
SpO_2	INR
Systolic blood pressure	Injury Severity Score
TBI diagnosis	Low respiratory rate
Venous thromboembolism	Low systolic blood pressure
	Mechanism of injury
	Potassium
	Race
	Respiratory rate
	Sex
	Sodium
	SpO_2
	Systolic blood pressure

AIS, Abbreviated injury score; *CT*, computed tomography; *ICD*, International Classification of Diseases; *INR*, International Normalized Ratio; *SpO_2*, oxygen saturation; *TBI*, traumatic brain injury.

The number of machine learning models employed in each study ranged from 1 to 23. In terms of broader categories, artificial neural networks were the most commonly used (n=42, 63%), followed by decision trees (n=18, 27%), support vector machines (n=15, 22%), Bayesian networks (n=14, 21%), random forests (n=13, 19%), LASSO (n=7, 10%), gradient boosting (n=4, 6%), and K-nearest neighbors (n=3, 5%). Others included generalized additive models, adaptive boosting, elastic net, and concept formation. As previously stated, logistic regression and multivariate regression methods were explicitly excluded from counts of machine learning models, as these were labeled as statistical methods. Out of the 25 studies that directly compared machine learning models to logistic regression, 48% (n=12) reported no notable differences in performance, while machine learning outperformed logistic regression in 10 (40%) studies and underperformed in 3 (12%) studies. Three (4.5%) studies compared their model performance to clinicians. Calibration statistics were formally reported in 15 (21%) articles.

Table 3 summarizes the bias and applicability results, guided by the PROBAST tool.[17] Nineteen articles (28%) had high risk of bias, 35 (52%) had low risk of bias, and the remaining 13 (19%) had "unsure" risk of bias. It was observed that most articles that were not assessed to have low risk of bias faltered in the "analysis" portion of bias, with high-risk and unsure ratings being given to 18 (26%) articles each. These evaluations largely originated from the improper handling of continuous and categorical variables, not utilizing methods to select model features, lacking adequate performance measures, and/or lacking model calibration. Studies were evaluated to be much more favorable in terms of applicability, with 54 (81%) articles rated as highly applicable, 13 (19%) as unsure, and none as low applicability.

LIMITATIONS

This review carried some limitations. First, although the heterogeneity in the studies was an important finding, this also rendered it difficult to compare models and outcomes more directly and correspondingly; a meta-analysis could not be performed. This review was also not designed to examine the suitability between the chosen machine learning model(s), patient population, and clinical outcome(s) for any given study. The sheer amount of different machine learning models that are available creates difficulties in comparison, as there may be no one optimal model for a given clinical scenario. Aside from the in-depth understanding that would be required of each article's methodology, the myriad of subtypes which also exist for each model type adds to this complexity.[15] We attempted a broad scope for our search, but some articles were likely missed, owing to language barriers, and not being accessible. For some studies, it was also difficult to assess whether machine learning methods were used for outcome prediction or for other instances. More standardized frameworks for conducting and reporting machine learning research will mitigate this. Publication bias is potentially present, as the literature may lean toward models that generate positive results. However, selecting for the best-

Table 3. Bias and applicability assessments of included studies by review authors based on PROBAST tool.

	Risk of Bias					Applicability			
Study Reference	**Participants**	**Predictors**	**Outcome**	**Analysis**	**Overall**	**Participants**	**Predictors**	**Outcome**	**Overall**
Abbod 2011[18]	+	+	+	?	+	+	+	+	+
Abujaber 2020 (1)[8]	?	+	+	-	-	+	+	+	+
Abujaber 2020 (2)[19]	-	?	-	-	-	+	-	+	?
Abujaber 2020 (3)[20]	-	-	+	-	-	+	+	+	+
Ahmed 2020[42]	+	+	+	+	+	+	+	+	+
Becalick 2001[43]	?	+	+	?	+	+	+	+	+
Bektas 2008[44]	+	+	+	+	+	+	+	+	+
Bradley 2020[45]	+	+	+	?	?	+	+	+	+
Bravo-Merodio 2019[46]	+	+	+	+	+	+	+	+	+
Chen 2013[47]	-	+	-	-	-	?	+	+	?
Christie 2018[48]	+	?	+	+	+	+	+	+	+
Christie 2019[49]	+	+	+	?	+	+	+	+	+
DiRusso 2000[50]	+	+	+	+	+	+	+	+	+
Donald 2019[21]	+	+	+	?	+	+	+	+	+
Eftekhar 2005[22]	?	?	+	-	-	?	+	+	?
Feng 2019[23]	+	+	+	+	+	+	+	+	+
Gelbard 2019[51]	+	+	+	-	?	+	+	?	?
Gholipour 2015[52]	+	+	+	-	-	+	+	+	+
Gorczyca 2019[53]	+	+	+	+	+	+	+	+	+
Gravesteijn 2020[24]	?	+	+	+	+	+	+	+	+
Hadzikadic 1996[54]	?	+	+	-	-	+	+	+	+
Hertz 2020[55]	?	+	+	-	?	?	+	+	?
Hirshberg 2002[56]	-	?	+	?	-	+	+	+	+
Hsu 2005[25]	+	+	+	-	-	+	+	+	+
Hubbard 2013[57]	+	+	+	+	+	+	+	+	+
Hunter 2000[58]	+	+	+	?	?	+	+	+	+
Ji 2009[59]	+	+	?	+	+	+	+	+	+
Kalpakis 2015[26]	+	+	+	+	+	+	+	+	+
Kim 2018[60]	+	+	+	+	+	+	+	+	+
Fann 2007[61]	?	+	?	-	-	?	-	+	?
Kuo 2018[62]	+	+	+	?	?	+	+	+	+
Li 2000[27]	+	+	+	?	+	+	+	+	+
Li 2020[63]	?	+	?	+	?	?	+	?	?
Liu 2014 (1)[64]	+	?	+	?	?	+	?	+	?
Liu 2014 (2)[65]	+	+	+	+	+	+	+	+	+
Lu 2015[28]	+	+	+	+	+	+	+	+	+
Marble 1999[66]	+	+	?	-	-	+	+	+	?
Matsuo 2020[29]	+	+	+	?	?	+	+	+	+
Maurer 2021[1]	+	?	+	+	?	+	+	+	+
McGonigal 1993[67]	+	+	+	?	?	+	+	+	+
Moinadini 2019[68]	+	+	?	?	-	?	+	+	+
Nikiforidis 1998[30]	+	?	+	-	-	+	+	+	+
Pang 2007[31]	+	+	+	+	+	+	+	+	+
Paydar 2021[69]	+	+	?	?	?	+	+	+	+
Pearl 2006[70]	+	+	+	-	-	+	+	+	+

Table 3. Continued.

	Risk of Bias					Applicability			
Study Reference	**Participants**	**Predictors**	**Outcome**	**Analysis**	**Overall**	**Participants**	**Predictors**	**Outcome**	**Overall**
Pearl 2008[71]	+	+	+	-	-	+	+	+	?
Pearl 2009[72]	+	?	?	-	-	+	?	?	?
Davis 2008[53]	+	+	+	?	?	+	+	+	+
Perkins 2020 (1)[73]	+	+	+	+	+	+	+	+	+
Perkins 2020 (2)[74]	+	+	+	+	+	+	+	+	+
Lang 1997[33]	+	+	+	-	-	+	+	+	+
Pourahmad 2016[34]	+	+	+	+	+	+	+	+	+
Pourahmad 2019[35]	+	+	+	?	+	+	+	+	+
Rau 2018[36]	+	+	+	+	+	+	+	+	+
Rughani 2010[37]	+	+	+	?	?	+	+	+	+
Rutledge 1995[75]	+	+	+	+	+	+	?	+	+
Rutledge 1998[76]	-	+	+	?	-	?	+	+	+
Schetinin 2018[77]	+	+	+	+	+	+	+	+	+
Segal 2006[38]	+	?	+	+	+	+	+	+	+
Servia 2020[78]	+	+	+	+	+	+	+	+	+
Shi 2013[39]	+	+	+	+	+	+	+	+	+
Staziaki 2021[2]	?	+	+	+	+	+	+	+	+
Tsiklidis 2020[79]	+	+	+	+	+	+	+	+	+
Tsurumi 2020[80]	+	+	+	+	+	+	?	+	?
van der Ploeg 2016[40]	+	+	+	+	+	+	+	+	+
Vath 2000[41]	-	+	?	-	-	?	+	?	?
Wolfe 2006[81]	+	+	+	+	+	+	+	+	+

+, Favorable result (low bias/high applicability); -, unfavorable result (high bias/low applicability);?, unsure.

performing model is a natural step in machine learning development, as poorly performing models are irrelevant, and the previous findings of the effect of publication bias are mixed.[82-84]

DISCUSSION

Machine learning has extensive potential to help clinicians navigate the complexities of trauma medicine, improve patient outcomes, and reduce health care costs.[85] As the role of machine learning will only continue to grow within medicine, this systematic review was conducted to help support the future development of predictive machine learning models by gleaning insight from the current state of affairs. A portion of this consisted of a partial update to the 2017 review by Liu and Salinas.[6] Nearly half of the studies included were published after their review, which speaks volumes about the rapid growth of machine learning in trauma care. While we, similarly, found that machine learning models were often successful in accurately predicting patient outcomes, this review highlights the extent of heterogeneity in the development and evaluation of machine learning models, even within the single field of trauma medicine. Differences were widespread in areas including model development (ie, feature selection, data sampling, features used), algorithms used, model validation, performance metrics, and research reporting.

Liu and Salinas[6] also emphasized the importance of utilizing information that could be realistically and quickly collected at the bedside for prediction and called for a common database of potential features that could be used to predict trauma outcomes. To this end, our feature database (Table 3) demonstrated, similarly to TRISS, that features that were commonly predictive of mortality included GCS score, age, and ISS score. Other features that were included in model development were body temperature, oxygen saturation, and age. The only variable within TRISS that was not commonly represented in machine learning models was the respiratory rate for patients with traumatic brain injuries. One of the benefits

of machine learning is the inclusion of a larger and more complex set of features in outcome prediction that would simply be too unwieldly to mirror with a traditional scoring tool. Second, this is paired with variable feature weighing, something that is not always possible with standard prediction scores.

This review highlights a lack of external validation. External validation is a substantially consequential step in developing effective and applicable machine learning models but was only performed in a minority of studies. This process not only helps improve accuracy but also represents a crucial step to help ensure that overfitting is avoided. This is compounded with the finding that most studies did not employ objective methods in feature selection for their machine learning models. Properly weeding out redundant features can reduce the number of parameters required, reduce training time, enhance generalizability, and help avoid overfitting.[86] Two other systematic reviews have also found that calibration tests are rarely performed, signaling this as an important weakness in the field of predictive machine learning.[87,88] This becomes especially important when considering the ethical implications of machine learning and the negative effects that can arise. The algorithms that form the basis of machine learning models are far from infallible and may overfit data that carry institutional biases, however subtle they may be. Zech et al[89] found that their model performed poorly on external data sets, as it had incorporated confounding factors associated with institutional biases as part of its predictions alongside more objective evaluated pathology. When evaluating a machine learning algorithm developed to identify patients in need of extra care in the United States, Obermeyer et al[90] found that it frequently classified Black patients at the same risk level as White patients, even when they were considerably sicker. Examples such as these underscore the importance of the proper development of machine learning models and the scrutiny of input data. The technological superiority of machine learning compared to more traditional paper-and-pencil screening tools does not exempt it from crucial steps, such as external validation to mitigate downstream bias.

The relative complexity of machine learning as a field also means that it is inherently difficult to explain in layman's terms and, at times, even more difficult to show due to its "black-box" nature. A number of papers have explored the unique nature of machine learning methodology.[91-93] It is, therefore, unsurprising that there is hesitation in the uptake of machine learning.[6] Although most developed models show great accuracy and outperform their conventional counterparts, there are still studies that identify more negative results.[94,24,95] It is difficult to pinpoint, within such a study, where exactly the problem may lie—whether there was a suboptimal decision in the development of the model or if machine learning is simply maladapted to a specific clinical situation. Even within machine learning, scrutiny should be given to the exact machine learning algorithm being used. As found in this review, while more novel methods may address downfalls in their predecessors, more "traditional" methods, such as logistic regression, can outperform within some contexts. Questions still exist around adapting practice to machine learning predictions and which situations machine learning models might be best suited. For instance, a machine learning model that predicted sepsis was observed to create some ambiguity for clinicians who were unsure how to change the management of patients who were already receiving optimal therapy and were clinically stable but predicted to later develop sepsis.[96] However, improving clinical efficacy by incorporating standard practices into the development and reporting of predictive machine learning models should be a first step in beginning to address some of these barriers.[7]

Additionally, many deviations from the prescribed best practices of machine learning model development were observed, which—although they may reflect the infancy of the field in medicine—are harmful to its overall growth. Underlining the importance of high-quality data in the development of machine learning, some patient cohorts were relatively small, while in other articles, issues were observed regarding the handling of missing data and inappropriate conversions of continuous features into categorical values. Other shortcomings, such as a paucity of studies employing prospective data collection and/or performing formal calibration tests, emphasize the need for guidelines to follow.

As such, there is a strong need for additional resources for future researchers to follow, including a database of machine learning algorithms and, perhaps, the standardization of outcomes and metrics for machine learning uses. For instance, while the AUROC currently appears to be widely accepted as a performance criterion, not all studies in this review reported AUROCs; some opted for metrics such as sensitivity and specificity instead. Developers should also strive to include external validation metrics as a standard, normalized practice so that issues of generalizability are identified early. Within individual medical fields, it may be important to set specific priority topics to better focus research efforts.[7] Setting proper research priorities and standardizing the reporting of research through the use of machine learning reporting guidelines would allow for increased ease in directly comparing machine learning models, both individually and

through meta-analyses. This would potentially produce stronger evidence for real-time machine learning applications, fueling the rationale for clinical implementation, as the necessary technology and need for machine learning in trauma care are already present. The unique plasticity of machine learning to continue adapting after implementation also carries a distinct potential advantage over traditional decisionmaking tools, but not without regulatory and ethical concerns to explore. Looking to the future, breaking the glass on the real-time clinical applications of machine learning would open avenues to allow for dynamic machine learning–supported decision tools that could be vital in trauma evaluation and treatment.

Ultimately, targeting the issues associated with heterogeneity highlights the need for the establishment of clearer guidelines to help promote the best practices of machine learning model development. Although it is a clear upside that many models were assessed to be effective in predicting their respective outcomes, there are clearly many growing pains in the field. Some tools, such as PROBAST, CONSORT, and SPIRIT-AI, exist, but these are less oriented to aid in model development or they may not entirely apply to machine learning model development.[17,97,98] TRIPOD-AI is a tool that is currently being developed to address these issues in the development of machine learning prediction models.[99,100] We recommend following these guidelines to help ensure quality artificial intelligence research, as they provide in-depth checklists for every step from machine learning development to reporting, similarly to how the PRISMA outlines steps and considerations for a review.[14] Our systematic review demonstrated a significant lack of database scrutiny (as seen in the PROBAST bias guidelines), feature selection, and external validation. Improvements in these areas would help reduce bias in model output and improve reproducibility. The "black-box" aspect of artificial intelligence research heavily influences our ability to bring it into the mainstream, as most clinicians are hesitant without knowing the inner workings. These strategies would help mitigate this hesitancy and improve the adoption of these powerful models.

In conclusion, machine learning has great potential in trauma care and can be applied to a variety of patient outcomes. To help improve further research, we have developed a feature database related to mortality prediction for both traumatic brain injury and multisystem trauma. We recommend the use of existing guidelines, such as CONSORT-AI, SPIRIT-AI, and PROBAST, and look forward to TRIPOD-AI to standardize study methods and reporting. Machine learning has the potential for heavy bias, and the scrutiny of data sets, feature selection, and external validation are imperative for further implementation. Our study demonstrates the current paucity of these, and future work is needed to ensure the proper development, accuracy, and applicability of machine learning models.

Supervising editor: Stephen Schenkel, MD, MPP. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Temerty Faculty of Medicine (Zhang) and Division of Emergency Medicine, Department of Medicine, University of Toronto (Nikouline, Nolan); and the Health Science Library, Unity Health Toronto (Lightfoot), Li Ka Shing Knowledge Institute (Nolan), and Department of Emergency Medicine, St. Michael's Hospital, Toronto, Ontario, Canada (Nolan).

Author contributions: AN, TZ, and BN conceived and designed the study. DL drafted the initial electronic search strategy. All authors participated in article screening. TZ and AN performed the data collection and extraction. TZ, AN, and BN carried out the data analysis. TZ drafted the initial manuscript, all authors contributed substantially to its writing and revision. TZ takes responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist. The authors report this article did not receive any outside funding or support.

Publication dates: Received for publication November 19, 2021. Revision received March 20, 2022. Accepted for publication May 4, 2022.

Trial registration number: CRD42021248580.

REFERENCES

1. Maurer LR, Bertsimas D, Bouardi HT, et al. Trauma outcome predictor: an artificial intelligence interactive smartphone tool to predict outcomes in trauma patients. *J Trauma Acute Care Surg*. 2021;91:93-99.
2. Staziaki PV, Wu D, Rayan JC, et al. Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma. *Eur Radiol*. 2021;31:5434-5441.
3. Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over decision tree models: an application for estimating uncertainty in trauma severity scoring. *Int J Med Inform*. 2018;112:6-14.

4. Hassanipour S, Ghaem H, Arab-Zozani M, et al. Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: a systematic review and meta-analysis. *Injury*. 2019;50:244-250.
5. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019;19:64.
6. Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. *Shock*. 2017;48:504-510.
7. Kareemi H, Vaillancourt C, Rosenberg H, et al. Machine learning versus usual care for diagnostic and prognostic prediction in the emergency department: a systematic review. *Acad Emerg Med*. 2021;28:184-196.
8. Abujaber A, Fadlalla A, Gammoh D, et al. Prediction of in-hospital mortality in patients with post traumatic brain injury using National Trauma Registry and machine learning approach. *Scand J Trauma Resusc Emerg Med*. 2020;28:44.
9. Aleman R, Patel S, Sleiman J, et al. Cardiogenic shock and machine learning: a systematic review on prediction through clinical decision support softwares. *J Card Surg*. 2021;36:4153-4159.
10. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg*. 2018;109:476-486.e1.
11. Lisboa PJG. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw*. 2002;15:11-39.
12. Schöllhorn WI. Applications of artificial neural nets in clinical biomechanics. *Clin Biomech*. 2004;19:876-898.
13. Ohno-Machado L, Rowland T. Neural network applications in physical medicine and rehabilitation. *Am J Phys Med Rehabil*. 1999;78:392-398.
14. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.
15. Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning. *Clin Pharmacol Ther*. 2020;107:871-885.
16. Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol*. 2008;26:1011-1013.
17. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170:51-58.
18. Abbod MF, Cheng KY, Cui XR, Huang SJ, Han YY, Shieh JS. Ensembled neural networks for brain death prediction for patients with severe head injury. *Biomed Signal Process Control*. 2011;6:414-421.
19. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Using trauma registry data to predict prolonged mechanical ventilation in patients with traumatic brain injury: Machine learning approach. *PLoS One*. 2020;15:e0235231.
20. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Prediction of in-hospital mortality in patients on mechanical ventilation post traumatic brain injury: machine learning approach. *BMC Med Inform Decis Mak*. 2020;20:1-11.
21. Donald R, Howells T, Piper I, et al. Forewarning of hypotensive events using a Bayesian artificial neural network in neurocritical care. *J Clin Monit Comput*. 2019;33:39-51.
22. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak*. 2005;5:3.
23. Feng JZ, Wang Y, Peng J, Sun MW, Zeng J, Jiang H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. *J Crit Care*. 2019;54:110-116.
24. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J Clin Epidemiol*. 2020;122:95-107.
25. Hsu MH, Li YC, Chiu WT, Yen JC. Outcome prediction after moderate and severe head injury using an artificial neural network. *Stud Health Technol Inform*. 2005;116:241-245.
26. Kalpakis K, Yang S, Hu PF, et al. Permutation entropy analysis of vital signs data for outcome prediction of patients with severe traumatic brain injury. *Comput Biol Med*. 2015;56:167-174.
27. Li YC, Liu L, Chiu WT, Jian WS. Neural network modeling for surgical decisions on traumatic brain injury patients. *Int J Med Inform*. 2000;57:1-9.
28. Lu HY, Li TC, Tu YK, Tsai JC, Lai HS, Kuo LT. Predicting long-term outcome after traumatic brain injury using repeated measurements of Glasgow Coma Scale and data mining methods. *J Med Syst*. 2015;39:14.
29. 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. *J Neurotrauma*. 2020;37:202-210.
30. Nikiforidis GC, Sakellaropoulos GC. Expert system support using Bayesian belief networks in the prognosis of head-injured patients of the ICU. *Med Inform (Lond)*. 1998;23:1-18.
31. Pang BC, Kuralmani V, Joshi R, et al. Hybrid outcome prediction model for severe traumatic brain injury. *J Neurotrauma*. 2007;24:136-146.
32. Davis DP, Peay J, Good B, et al. Air medical response to traumatic brain injury: a computer learning algorithm analysis. *J Trauma*. 2008;64:889-897.
33. Lang EW, Pitts LH, Damron SL, Rutledge R. Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neurol Res*. 1997;19(3):274-280.
34. Pourahmad S, Hafizi-Rastani I, Khalili H, Paydar S. Identifying important attributes for prognostic prediction in traumatic brain injury patients. A hybrid method of decision tree and neural network. *Methods Inf Med*. 2016;55:440-449.
35. Pourahmad S, Rasouli-Emadi S, Moayyedi F, Khalili H. Comparison of four variable selection methods to determine the important variables in predicting the prognosis of traumatic brain injury patients by support vector machine. *J Res Med Sci*. 2019;24:97.
36. Rau CS, Kuo PJ, Chien PC, Huang CY, Hsieh HY, Hsieh CH. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PLoS One*. 2018;13:e0207192.
37. Rughani AI, Dumont TSM, Lu Z, et al. Use of an artificial neural network to predict head injury outcome. *J Neurosurg*. 2010;113:585-590.
38. Segal ME, Goodman PH, Goldstein R, et al. The accuracy of artificial neural networks in predicting long-term outcome after traumatic brain injury. *J Head Trauma Rehabil*. 2006;21:298-314.
39. Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J Neurosurg*. 2013;118:746-752.
40. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *J Clin Epidemiol*. 2016;78:83-89.
41. Väth A, Meixensberger J, Dings J, Meinhardt M, Roosen K. Prognostic significance of advanced neuromonitoring after traumatic brain injury using neural networks. *Zentralbl Neurochir*. 2000;61:2-6.
42. Ahmed FS, Ali L, Joseph BA, Ikram A, Ul Mustafa R, Bukhari SAC. A statistically rigorous deep neural network approach to predict mortality in trauma patients admitted to the intensive care unit. *J Trauma Acute Care Surg*. 2020;89:736-742.
43. Becalick DC, Coats TJ. Comparison of artificial intelligence techniques with UKTRISS for estimating probability of survival after trauma. UK trauma and injury severity score. *J Trauma*. 2001;51:123-133.

44. Bektaş F, Eken C, Soyuncu S, Kilicaslan I, Cete Y. Artificial neural network in predicting craniocervical junction injury: an alternative approach to trauma patients. *Eur J Emerg Med*. 2008;15:318-323.
45. Bradley M, Dente C, Khatri V, et al. Advanced modeling to predict pneumonia in combat trauma patients. *World J Surg*. 2020;44:2255-2262.
46. Bravo-Merodio L, Acharjee A, Hazeldine J, et al. Machine learning for the detection of early immunological markers as predictors of multi-organ dysfunction. *Sci Data*. 2019;6:328.
47. Chen W, Cockrell CH, Ward K, Najarian K. Predictability of intracranial pressure level in traumatic brain injury: features extraction, statistical analysis and machine learning-based evaluation. *Int J Data Min Bioinform*. 2013;8:480-494.
48. Christie SA, Hubbard AE, Callcut RA, et al. Machine learning without borders? An adaptable tool to optimize mortality prediction in diverse clinical settings. *J Trauma Acute Care Surg*. 2018;85:921-927.
49. Christie SA, Conroy AS, Callcut RA, Hubbard AE, Cohen MJ. Dynamic multi-outcome prediction after injury: applying adaptive machine learning for precision medicine in trauma. *PLoS One*. 2019;14:e0213836.
50. DiRusso SM, Sullivan T, Holly C, Cuff SN, Savino J. An artificial neural network as a model for prediction of survival in trauma patients: validation for a regional trauma area. *J Trauma*. 2000;49:212-223.
51. Gelbard RB, Hensman H, Schobel S, et al. Random forest modeling can predict infectious complications following trauma laparotomy. *J Trauma Acute Care Surg*. 2019;87:1125-1132.
52. Gholipour C, Rahim F, Fakhree A, Ziapour B. Using an Artificial Neural Networks (ANNs) model for prediction of intensive care unit (ICU) outcome and length of stay at hospital in traumatic patients. *J Clin Diagn Res*. 2015;9:19-23.
53. Gorczyca MT, Toscano NC, Cheng JD. The trauma severity model: An ensemble machine learning approach to risk prediction. *Comput Biol Med*. 2019;108:9-19.
54. Hadzikadic M, Hakenewerth A, Bohren B, Norton J, Mehta B, Andrews C. Concept formation vs. logistic regression: predicting death in trauma patients. *Artif Intell Med*. 1996;8:493-504.
55. Hertz AM, Hertz NM, Johnsen NV. Identifying bladder rupture following traumatic pelvic fracture: A machine learning approach. *Injury*. 2020;51:334-339.
56. Hirshberg A, Wall MJ, Mattox KL. Bullet trajectory predicts the need for damage control: an artificial neural network model. *J Trauma*. 2002;52:852-858.
57. Hubbard A, Munoz ID, Decker A, et al. Time-dependent prediction and evaluation of variable importance using superlearning in high-dimensional clinical data. *J Trauma Acute Care Surg*. 2013;75:S53-S60.
58. Hunter A, Kennedy L, Henry J, Ferguson I. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Comput Methods Programs Biomed*. 2000;62:11-19.
59. Ji SY, Smith R, Huynh T, Najarian K. A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries. *BMC Med Inform Decis Mak*. 2009;9:2.
60. Kim D, You S, So S, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. *PLoS One*. 2018;13:e0206006.
61. Fann SA, Kosciusko RD, Yost MJ, et al. The use of prognostic indicators in the development of a statistical model predictive for adrenal insufficiency in trauma patients. *Am Surg*. 2007;73:210-214.
62. Kuo PJ, Wu SC, Chien PC, et al. Derivation and validation of different machine-learning models in mortality prediction of trauma in motorcycle riders: a cross-sectional retrospective study in southern Taiwan. *BMJ Open*. 2018;8:e018252.
63. Li K, Wu H, Pan F, et al. A machine learning-based model to predict acute traumatic coagulopathy in trauma patients upon emergency hospitalization. *Clin Appl Thromb*. 2020:26.
64. Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock*. 2014;42:108-114.
65. Liu NT, Holcomb JB, Wade CE, et al. Development and validation of a machine learning algorithm and hybrid system to predict the need for life-saving interventions in trauma patients. *Med Biol Eng Comput*. 2014;52:193-203.
66. Marble RP, Healy JC. A neural network approach to the diagnosis of morbidity outcomes in trauma care. *Artif Intell Med*. 1999;15:299-307.
67. Mc Gonigal MD, Cole J, Schwab CW, Kauder DR, Rotondo MF, Angood PB. A new approach to probability of survival scoring for trauma quality assurance. *J Trauma*. 1993;34:863-870.
68. Moinadini S, Tajoddini S, Hedayat AA. Prediction of the central venous pressure in trauma patients on the basis of non-invasive parameters using artificial neural network. *Hong Kong Journal of Emergency Medicine*. 2019;28(3):152-164.
69. Paydar S, Parva E, Ghahramani Z, et al. Do clinical and paraclinical findings have the power to predict critical conditions of injured patients after traumatic injury resuscitation? Using data mining artificial intelligence. *Chin J Traumatol*. 2021;24:48-52.
70. Pearl A, Caspi R, Bar-Or D. Artificial neural network versus subjective scoring in predicting mortality in trauma patients. *Stud Health Technol Inform*. 2006;124:1019-1024.
71. Pearl A, Bar-Or R, Bar-Or D. An artificial neural network derived trauma outcome prediction score as an aid to triage for non-clinicians. *Stud Health Technol Inform*. 2008;136:253-258.
72. Pearl A, Bar-Or D. Using Artificial Neural Networks to predict potential complications during Trauma patients' hospitalization period. *Stud Health Technol Inform*. 2009;150:610-614.
73. Perkins ZB, Yet B, Marsden M, et al. Early identification of trauma-induced coagulopathy: development and validation of a multivariable risk prediction model. *Ann Surg*. 2021;274:E1119-E1128.
74. Perkins ZB, Yet B, Sharrock A, et al. Predicting the outcome of limb revascularization in patients with lower-extremity arterial trauma: development and external validation of a supervised machine-learning algorithm to support surgical decisions. *Ann Surg*. 2020;272:564-572.
75. Rutledge R. Injury severity and probability of survival assessment in trauma patients using a predictive hierarchical network model derived from ICD-9 codes. *J Trauma*. 1995;38:590-601.
76. Rutledge R, Osler T. The ICD-9-based illness severity score: a new model that outperforms both DRG and APR-DRG as predictors of survival and resource utilization. *J Trauma*. 1998;45:791-799.
77. Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over Decision Tree models for trauma severity scoring. *Artif Intell Med*. 2018;84:139-145.
78. Serviá L, Montserrat N, Badia M, et al. Machine learning techniques for mortality prediction in critical traumatic patients: anatomic and physiologic variables from the RETRAUCI study. *BMC Med Res Methodol*. 2020;20:1-12.
79. Tsiklidis EJ, Sims C, Sinno T, Diamond SL. Using the National Trauma Data Bank (NTDB) and machine learning to predict trauma patient mortality at admission. *PLoS One*. 2020;15:e0242166.
80. Tsurumi A, Flaherty PJ, Que YA, et al. Multi-biomarker prediction models for multiple infection episodes following blunt trauma. *iScience*. 2020;23:101659.
81. Wolfe R, McKenzie DP, Black J, Simpson P, Gabbe BJ, Cameron PA. Models developed by three techniques did not achieve acceptable prediction of binary trauma outcomes. *J Clin Epidemiol*. 2006;59:26-35.
82. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. 2021;375:n2281.

83. Xiong Z, Liu T, Tse G, et al. A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. *Front Physiol*. 2018;9:835.
84. De Silva K, Lee WK, Forbes A, et al. Use and performance of machine learning models for type 2 diabetes prediction in community settings: a systematic review and meta-analysis. *Int J Med Inform*. 2020;143:104268.
85. Rosenberg AL. Recent innovations in intensive care unit risk-prediction models. *Curr Opin Crit Care*. 2002;8:321-330.
86. Chen R-C, Dewi C, Huang S-W, et al. Selecting critical features for data classification based on machine learning methods. *J Big Data*. 2020;7:52.
87. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.
88. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9:e1001221.
89. Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 2018;15:e1002683.
90. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447-453.
91. Price WN. Big data and black-box medical algorithms. *Sci Transl Med*. 2018;10; eaao5333.
92. Poon AIF, Sung JJY. Opening the black box of AI-medicine. *J Gastroenterol Hepatol*. 2021;36:581-584.
93. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep*. 2019;49:15-21.
94. Zhang S, Lau BPH, Ng YH, et al. Machine learning algorithms do not outperform preoperative thresholds in predicting clinically meaningful improvements after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. Published online July 10, 2021. https://doi.org/10.1007/S00167-021-06642-4
95. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. 2021;181:1065-1070.
96. Giannini HM, Ginestra JC, Chivers C, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med*. 2019;47:1485-1492.
97. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26:1364-1374.
98. Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26:1351-1363.
99. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1-W73.
100. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162:55-63.