www.surgicalneurologyint.com

Surgical Neurology International

Editor-in-Chief: Nancy E. Epstein, MD, Clinical Professor of Neurological Surgery, School of Medicine, State U. of NY at Stony Brook.

SNI: Trauma

Editor
Jutty Parthiban
Kovai Medical Center and Hospital, Tamil Nadu, India

ScientificScholar®
Knowledge is power
Publisher of Scientific Journals

SNI
Open Access

Review Article

Machine learning algorithms for predicting outcomes of traumatic brain injury: A systematic review and meta-analysis

Evan Courville[1], Syed Faraz Kazim[1], John Vellek[2], Omar Tarawneh[2], Julia Stack[2], Katie Roster[2], Joanna Roy[3], Meic Schmidt[1], Christian Bowers[1]

[1]Department of Neurosurgery, University of New Mexico, Albuquerque, New Mexico, [2]Department of Neurosurgery, School of Medicine, New York Medical College, Valhalla, New York, United States, [3]Department of Neurosurgery, Topiwala National Medical and B. Y. L. Nair Charitable Hospital, Mumbai, Maharashtra, India.

E-mail: *Evan Courville - evan.courville@gmail.com; Syed Faraz Kazim - skazim@salud.unm.edu; John Vellek - jvellek@student.nymc.edu; Omar Tarawneh - otarawne@student.touro.edu; Julia Stack - juliastackm@gmail.com; Katie Roster - katieroster@gmail.com; Joanna Roy - joanna.roy21@yahoo.in; Meic Schmidt - mhschmidt@salud.unm.edu; Christian Bowers - cabowers@salud.unm.edu

***Corresponding author:**
Evan Courville,
Department of Neurosurgery,
University of New Mexico,
Albuquerque, New Mexico,
United States.

evan.courville@gmail.com

Received : 08 April 2023
Accepted : 21 June 2023
Published : 28 July 2023

DOI
10.25259/SNI_312_2023

Quick Response Code:

ABSTRACT

Background: Traumatic brain injury (TBI) is a leading cause of death and disability worldwide. The use of machine learning (ML) has emerged as a key advancement in TBI management. This study aimed to identify ML models with demonstrated effectiveness in predicting TBI outcomes.

Methods: We conducted a systematic review in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis statement. In total, 15 articles were identified using the search strategy. Patient demographics, clinical status, ML outcome variables, and predictive characteristics were extracted. A small meta-analysis of mortality prediction was performed, and a meta-analysis of diagnostic accuracy was conducted for ML algorithms used across multiple studies.

Results: ML algorithms including support vector machine (SVM), artificial neural networks (ANN), random forest, and Naïve Bayes were compared to logistic regression (LR). Thirteen studies found significant improvement in prognostic capability using ML versus LR. The accuracy of the above algorithms was consistently over 80% when predicting mortality and unfavorable outcome measured by Glasgow Outcome Scale. Receiver operating characteristic curves analyzing the sensitivity of ANN, SVM, decision tree, and LR demonstrated consistent findings across studies. Lower admission Glasgow Coma Scale (GCS), older age, elevated serum acid, and abnormal glucose were associated with increased adverse outcomes and had the most significant impact on ML algorithms.

Conclusion: ML algorithms were stronger than traditional regression models in predicting adverse outcomes. Admission GCS, age, and serum metabolites all have strong predictive power when used with ML and should be considered important components of TBI risk stratification.

Keywords: Artificial intelligence, Head injury, Machine learning, Mortality, Outcomes, Traumatic brain injury

INTRODUCTION

Physicians are often presented with large quantities of complex data and limited processing time. This presents barriers to the real-time analysis and prediction of patient outcomes. In

computer science, complex algorithms designed to learn from data and create generalizations are known as machine learning (ML). The marked proliferation of electronic medical record systems during recent years has presented unique opportunities for ML to improve patient care. Several ML learning techniques have been used in clinical practice to predict deleterious events and alert appropriate care teams. This has led to an increase in the number of early interventions, reduced mortality, and decreased lengths of hospital stay.[3,9,18,20,39]

Traumatic brain injury (TBI) remains one most prevalent causes of death and disability throughout the world.[10,19,38] Robust prediction of outcomes in these patients is critical for clinical decision-making, family counseling, and for the need-based allocation of quality of care. In recent years, TBI research has employed several ML models for the prediction of patient events and outcomes; however, there exists much variability in their results.[40-42] Conflicting data continues to be reported in the literature; for example, while one study reported that the ML-based predictive models were more powerful than classic multivariate analysis in head trauma patients, another reported ML algorithms performed no better than conventional for prognostication in TBI.[15] To the best of our knowledge, there exists no systematic review comparing various ML models used for predictions in TBI. The present systematic review and meta-analysis were conducted to summarize and analyze the available clinical literature regarding ML-based prediction of TBI outcomes. We conducted a small meta-analysis of available studies to estimate the predictive performance of ML-based algorithms for TBI outcomes.

MATERIALS AND METHODS

The present systematic review and meta-analysis were performed per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[28,32] Figure 1 shows the PRISMA flow diagram for the study.

Literature search strategy

We conducted a literature search of studies reporting on ML-based prediction of TBI outcomes published until March 31, 2021. We searched the following three electronic bibliographic databases: PubMed, EMBASE, and Cochrane Library. We used the following MeSH (Medical Subject Heading) terms in combination with Boolean Operators OR and AND: "machine learning" OR "artificial intelligence" OR "neural network" OR "naive Bayes" OR "Bayesian learning" OR "random forest" OR "deep learning" OR "machine intelligence" OR "boosting" OR "nature language processing" OR "decision tree" AND "traumatic brain injury" OR "head injury." An additional search involving the following terms was also performed: "machine learning" OR "artificial intelligence" OR "neural network" OR "naive Bayes" OR "Bayesian learning" OR "random forest" OR "deep learning" OR "machine intelligence" OR "boosting" OR "nature language processing" OR "decision tree" AND "traumatic brain injury" OR "head injury" OR AND "outcome" OR "mortality" OR "morbidity."

Inclusion and exclusion criteria

We included peer-reviewed prospective and retrospective cohort studies published in the English language utilizing ML algorithms to predict outcomes of TBI in human patients. Single case reports, editorials, reviews, and conference/meeting abstracts were excluded from the study. Furthermore, TBI studies that used ML for a purpose other than predicting outcomes were also excluded. We also reviewed the reference lists of the selected articles for any additional articles related to the topic.

Data extraction

Three independent investigators (JV., OHT., and JS.) reviewed the full text of the included articles and extracted the data on a data collection form. Any disagreement between the three authors was resolved by discussion. The following data were extracted from each study: study design, TBI population characteristics, ML and comparative regression models used, ML input variables, outcome variables, study results, and predictive performance of various models used in the study.

Risk of bias assessment

We employed the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) in the Review Manager (RevMan) software version 5.4 to assess the quality of extracted studies [Figure 2]. The following four domains are included in the QUADAS-2: (1) patient selection; (2) index test; (3) reference standard; and (4) flow and timing.[46] We assessed each domain with regard to the risk of bias, and the first three also for concerns regarding applicability. We used the signaling questions to assess the risk of bias and applicability concerns. For each domain, we analyzed the risk of bias and concerns about applicability (the latter not applying to the domain of flow and timing) and rated each domain as low (+), high (–), or unclear (?) (could not be assessed due to missing information) risk. Studies rated as "low" on all domains regarding bias or applicability concerns were identified as having an overall low risk of bias or low concerns regarding applicability.[46] Contrarily, studies judged as having a high risk of bias in one or more domains were identified as having an overall high risk of bias or high concerns regarding applicability.[46]

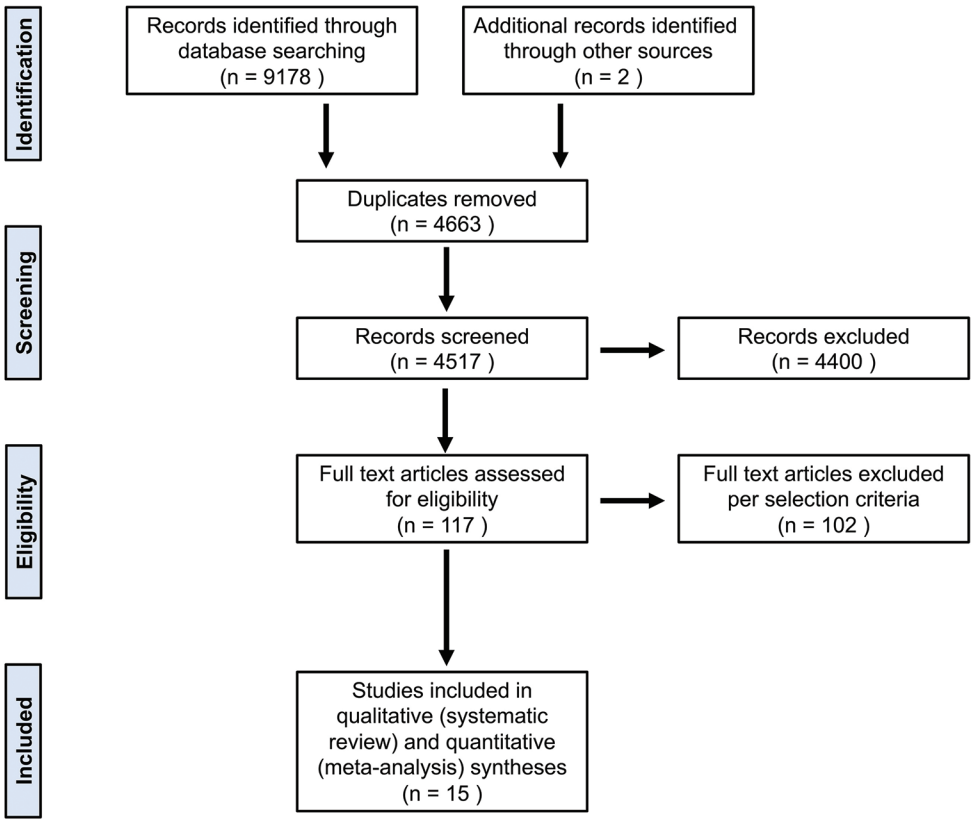

Figure 1: Flow diagram of literature selection process per PRISMA guidelines in the present systematic review and meta-analysis. n: Number of articles; PRISMA: Preferred reporting items for systematic reviews and meta-analyses.

Quality assessment

The quality of the included studies was evaluated using the QUADAS-2 tool in RevMan version 5.4.1 software. Each study was assessed using 12 signaling questions (three from each domain) and three questions regarding study applicability (one each from the first three domains) [Figure 1]. The rating for each question was yes, no, or unclear. "No" indicates a small risk of bias, whereas "yes" indicates a high risk of bias for the specific question. "Unclear" indicates that the risk of bias could not be assessed due to missing information. We assessed agreement between both evaluators using three (yes, no, or unclear) and two (yes or combined unclear/no) response levels. The agreement was calculated for each question, for each domain, and for the overall assessment. Studies that were judged as "low" on all domains regarding bias or applicability were rated as having an overall low risk of bias or low concern regarding applicability. Studies that were judged as having a high risk of bias in one or more domains were rated as having an overall high risk of bias or high concern regarding applicability.

The domain "Patient Selection" addresses the following question: "Could the selection of patients or study participants have introduced bias?" The constitution of the study population is centrally important to a high-quality study. We distinguished three populations, study, source, and target. The study population is the population that was reported on in an article, sampled from a larger source population. Only two studies (Gravesteijn *et al*. 2020 and Raj *et al*. 2019) reported an unclear risk of bias while others answered a low risk of bias and only one study answered a high risk (Rizoli *et al*. 2016).[15,34,37]

The domain "Index Text" addresses the question: "Could the conduct or interpretation of the index test have introduced bias?" The index test results are one central component of a 2×2 table that is evaluated in diagnostic studies. The index test is the assay under investigation in the study, and a study may evaluate one or more index tests in the same population or among population subsets. Among the studies cohort in our systematic review and meta-analysis using ML and comparative regression models in the prediction of TBI, only one study qualified as high risk (Rizoli *et al*. 2016), four studies (Amorim *et al*. 2019, Rau *et al*. 2017, Kayhanian *et al*. 2019, and Raj *et al*. 2019) remained unclear risk while others reported a low risk of bias.[4,21,36,37]

The domain "Reference standard" addresses the question: "Could the reference standard, its conduct, or its interpretation

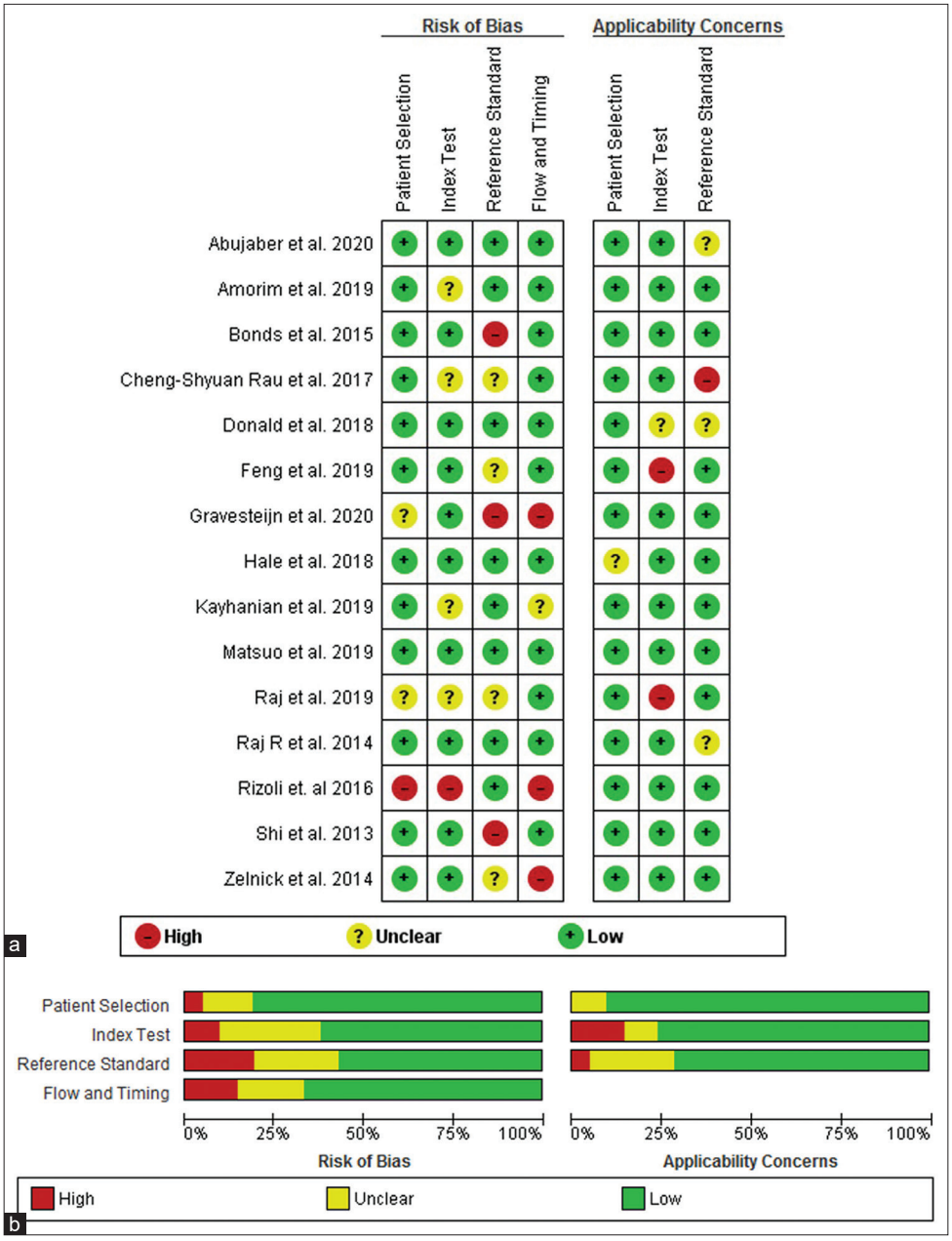

Figure 2: The risk of bias assessment.

have introduced bias?" Among all pooled studies, Gravesteijn *et al*. 2020, Shi *et al*. 2013, and Bonds *et al*. 2015 answered a high risk, Zelnick *et al*. 2014, Rau *et al*. 2018, Raj *et al*. 2019, and Feng *et al*. 2019 reported unclear risk while others pooled studies reported low risk.[7,14,15,34,36,43,49]

The domain "Flow and Timing" addresses the question: "Could the study flow and timing have introduced bias?" The methods and results sections should provide a clear description of clinical referral algorithms (i.e., patients who did/did not receive the index tests or reference standard, respectively) and of any patients excluded from the analyses. Three studies in our analysis (Zelnick *et al*. 2014, Gravesteijn *et al*. 2020, and Rizoli *et al*. 2016) reported a considerable risk of bias while one study (Kayhanian *et al*. 2019) answered unclear risk of bias, all others reported a low risk of bias. Figure 2 summarizes the overall risk of bias in our systematic review/meta-analysis studies.[15,21,37,49]

Statistical analysis

We recorded data from the included studies in a Microsoft Excel datasheet (Microsoft Corp., Redmond, Washington, USA). For the pooled mortality rate, we employed a

random-effects meta-analysis model in R statistical software version 4.02 (R Foundation for Statistical Computing, Vienna, Austria). We measured the heterogeneity between the included studies employing the Higgins I^2 statistic. We used a random-effects model due to the high statistical heterogeneity (defined as $I^2 > 25\%$) among studies included in the meta-analysis. Forest plots were generated using the function "metaforest" in R statistical software.[49] A meta-analysis of diagnostic accuracy with hierarchical modeling was carried out for each of the ML models across the selected studies where the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values were reported. The sensitivity and specificity with corresponding 95% confidence intervals (95% CIs) were calculated from the TP, FP, FN, and TN rates extracted through a 2 × 2 table from each included study. The "metandi" module in STATA Version 14.1 (StataCorp., 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP) was used for meta-analysis of diagnostic accuracy.[22,23]

RESULTS

The initial literature search identified 9180 articles. After removing duplicates (n = 4663) and screening the titles and abstracts (n = 4517), we excluded a total of 9063 studies. After the screening of full-text articles based on our selection criteria (n = 117), we included a total of 15 studies in the qualitative systematic review and quantitative meta-analysis. However, the actual number of included studies for the small meta-analysis varied depending on how many studies documented the data for a particular algorithm using a similar methodology. Figure 1 shows the flow diagram of the literature selection process per PRISMA guidelines. Figure 2 shows the risk of bias assessment in included studies. Tables 1 and 2 present a review of the major relevant findings of the included 15 studies.

Prognostic factors for mortality and unfavorable outcomes

Although there was significant heterogeneity in the selected input variables used for the prediction of mortality and unfavorable outcomes, critical clinicopathological and imaging findings were identified from our review: abnormal serum glucose,[21,27,34] lactic acidosis,[21] older age and lower GCS at admission,[4,27,35] higher Marshall scores and decreased pupillary activity,[37] and high surgeon caseload and overall hospital workload.[43]

Diagnostic accuracy: Meta-analysis for ML algorithms

A small meta-analysis was conducted for studies using mortality as a primary outcome. Figure 3a illustrates the mortality data extracted from each study in which mortality was predicted using ML. A total of 32,721 patients were identified from nine studies, with an overall pooled mortality rate of 23%. Mortality rates within the individual studies ranged from 6%[16] to 54%,[15] with the majority falling in the range of 10–30%. A forest-funnel plot depicting mortality data is shown in Figure 3b, demonstrating a high degree of variability in the reported mortality; however, these values are in agreement with previously reported data.[24,48] Meta-analysis of diagnostic accuracy was conducted for recurring ML algorithms using receiver operating characteristic (ROC) curves. Figure 4 illustrates the findings for artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), and logistic regression (LR), respectively. Meta-analysis of diagnostic accuracy demonstrated that ANN results were consistent across studies and that its predictions were more accurate than traditional CT scanning models.[16]

In-hospital mortality

ANN and SVM have both been used to assess in-hospital mortality in a study containing 1620 patients.[1] The goal was to compare the performance of ML models to traditional in-hospital mortality measures that use multivariate regression. ML prediction variables included GCS, radiologic findings, arrival method, and time of day of presentation, among others. ANN and SVM predicted in-hospital mortality with an accuracy of>91%; however, SVM outperformed ANN with an accuracy of 95.6% and an area under curve (AUC) of 96%. SVM also outperformed traditional multivariate LR.[1] An additional study assessed the efficacy of various ML models versus LR in predicting mortality in TBI using a retrospective chart review.[14] The ML predictive variables included vital signs and GCS at admission and discharge. Linear, cubic, and quadratic SVM models all demonstrated an accuracy of 94%, whereas LR had an accuracy of 88%. Linear and quadratic SVM both showed an AUC of 0.93 with cubic SVM demonstrating an AUC of 0.94, in comparison to LRs value of 0.83.[14] All SVM ML models demonstrated a sensitivity of 0.98 or higher, the highest among the five studies in the meta-analysis for ML predicting mortality.[14] ANN was also used to predict in-hospital mortality using retrospective hospital data and patient comorbidities.[43] When comparing ANN to comparative regression model LR, ANN had a higher accuracy (95.23% vs. 82.44%), AUC (0.8961 vs. 0.7739), sensitivity (67.56% vs. 54.83%), specificity (95.23% vs. 92.67%), and positive (83.24% vs. 74.81%) and negative (89.35% vs. 87.64%) predictive values over LR. The study also included the Charlson comorbidity index, hospital volume, and surgeon volume as ML predictive variables.[43]

14-day mortality after TBI

An additional study used models to predict in-hospital and 14-day mortality. The models assessed mortality in 517 TBI patients in a low-middle-income country (LMIC).[4] Comparing

Table 1: Summary of included study findings.

Study	Outline	Design/Characteristics	TBI Population Characteristics	ML Models	Comparative Regression Models
Raj *et al*. 2019	Development and analysis of two simple dynamic algorithms for prediction of 30-day mortality of TBI patients using commonplace neuro-ICU measurements as predictive variables.	Retrospective multicenter study in Finland over range of 2003–2017	*n*=472, median age 48 years, 69% GCS 3–8 at admission, 79% had light reactive pupils bilaterally, 49% displayed mass lesion on CT 30-day mortality was 19% (*n*=92) Age 16+ reporting to ED within 24 h of trauma, must have had ICP monitoring for a minimum of 24 h Excluded patients dying within 36h of admission	ICP-CPP-MAP – 14 dynamic features + age included, refreshed every 8 h ICP-CPP-MAP-GCS – 13 dynamic features + age included, refreshed every 24 h	IMPACT-TBI – classified 30 patients as likely to die at 50% success Note – was not initially intended for this purpose, but was used as a linear predictor for comparison to a standard method in the field
Matsuo *et al*. 2019	Use of nine ML algorithms to determine effective prediction of poor outcome (based on Glasgow outcome score) and mortality.	Retrospective single center study in Japan over range of 2013–2016 Bootstrap analysis was used to amplify sample size in training sample	*n*=232 divided into groups 80:20 tuning: testing, mean age 59.4 years, 72.8% male, mean GCS 9.1, approximately half classified as severe TBI (GCS 3–8) Discharge were 7.8% good (GOS=5), 14.7% moderate disability (GOS=4), 77.6% poor outcome (GOS 1–3) Mean LOS 28.7 days, overall mortality 26.3% Age 10+ reporting to the ED, excluded if experienced cardiopulmonary arrest in the ED, pregnant, missing lab findings at admission	Ridge regression LASSO regression RF Gradient boosting Extra trees DT Gaussian NB Multinomial NB SVM	N/A – comparison was made to intrinsic training of ML algorithms
Rau *et al*. 2018	Design a ML model to predict mortality following moderate and severe TBI	Retrospective study based on data obtained from Trauma Registry System between 2009 and 2015	*n*=1734, 156 included in training set, 325 in test set. Hospitalized adult patients with head injuries characterized by Abbreviated Injury Score >3 points.	SVR NB ANN DT	LR
Amorim *et al*. 2019	Design and compare models of mortality in TBI patients in LIMC	Prospective and observational study of patients admitted to large trauma center who required ICU admission following TBI between 2012 and 2015	*n*=517 Patients ages 14+ in LMIC (Brazil) with intracranial abnormality on CT requiring ICU admission	RF neural network DT Stochastic gradient boosting Bayesian generalized linear model Partial least squares Multivariate adaptive regression splines NB Penalized discriminant analysis	Regularized least squares Linear regression

(Contd...)

Table 1: *(Continued).*

Study	Outline	Design/Characteristics	TBI Population Characteristics	ML Models	Comparative Regression Models
Feng *et al*. 2019	Compare the efficacy of various ML models versus LR in predicting mortality in TBI	Retrospective chart review of TBI patients in an urban hospital in China between 2009 and 2011	*n*=117 (85.5% male) Age 18–86 (mean 46), history of head injury and craniotomy Exclusion criteria: pregnancy, craniotomy at other site, hospitalization <24 h Mortality rate was 12% (*n*=14) Mean LOS was 28 days Median GCS at admission was 8 High rate of hospital-acquired pneumonia (33.3% *n*=39)	Cubic SVM Cubic KNN Complex tree Fine Gaussian SVM Weighted KNN Medium tree Medium Gaussian SVM Boosted trees Simple tree Coarse Gaussian SVM Bagged trees Linear discriminant Fine KNN Subspace discriminant Quadratic discriminant Medium KNN Subspace KNN Linear SVM Coarse KNN RUSBoosted trees Quadratic SVM Cosine KNN	LR
Abujaber *et al*. 2020	Compare performance of ML models ANN and SVM to traditional multivariate regression for prediction of in-hospital mortality following TBI.	Retrospective study of patients who sustained TBI and were admitted to level 1 trauma center (Qatar) between 2014 and 2019.	*n*=1620 (1417 survived, 203 deceased) Age 14+ (Mean: 34.4 years, SD 13.9) Most common mechanism of injury: fall from height (34%), MVA (30%) Most common finding: subdural hemorrhage (28.1%) and extradural hemorrhage (22.9%) Exclusion criteria: Pediatric patients (<14 years old) were excluded.	ANN SVM	Multivariate LR
Hale *et al*. 2018	Compare the use of ANN to traditional head CT analysis models to predict adverse outcomes in pediatric TBI patients.	Retrospective study of pediatric patients who sustained TBI and were admitted to an urban teaching hospital between 2006 and 2013	*n*=565 (533 favorable outcome and 32 unfavorable outcomes) Age <18 Admitted to hospital for TBI and underwent heat CT within 24 h of admission Exclusion criteria: fatality on arrival, no head CT within 24 h Follow-up at 6 months post-discharge Patients lost to follow-up with favorable GCS at admission were assigned a GOS of 5	ANN	Marshall CT Helsinki CT Rotterdam CT GCS

(Contd...)

Table 1: *(Continued).*

Study	Outline	Design/Characteristics	TBI Population Characteristics	ML Models	Comparative Regression Models
Kayhanian *et al*. 2019	Used serum metabolic markers to program a ML algorithm to predict unfavorable GOS in pediatric TBI patients	Retrospective study of pediatric patients admitted to a UK hospital for TBI from 2009 to 2013	*n*=94 Age <16 (mean 7.3) Admitted to hospital for severe TBI Inclusion criteria: confirmed TBI by CT or MRI, admission to the PICU after 24 h, invasive monitoring of ICP or arteriovenous pressure Follow-up at 6 months post-injury Assessed for GOS	SVM: Focused (only used pH, lactate, and glucose) Inclusive (used all blood variables)	LR: Focused (only used pH, lactate, and glucose) Inclusive (used all blood variables)
Donald *et al*. 2019[45]	Use of blood pressure values to develop a predictive model for hypotensive events in TBI patients in the neuro-ICU	Prospective phased trial of patients diagnosed with TBI across multiple centers in Europe between 2003 and 2011	Training set *n*=104 (2003–2005) Phase I *n*=30 (2009–2010) Phase II Stage I *n*=13 (2010–2011) Phase II Stage II *n*=36 (2010–2011) Final analysis group *n*=69 (75% male) Exclusion criteria: <24 h continuous monitoring, missing or incomplete data set Injury types: fall (30), car accident (25), pedestrian (4), unknown (4), sports related (3), assault (3)	BANN	N/A Training sample used
Bonds *et al*. 2015	Use of continuous VS for the prediction of secondary insult following severe TBI	Retrospective single center study of patients admitted to Level 1 trauma center with severe TBI GCS >9 between 2008 and 2010	*n*=132 adult patients Mean age: 40.2 (SD: 18.09) 96.97% blunt force trauma Mortality: 18 (13.63%)	NNR	Regression tree Simple shifting estimation
Shi *et al*. 2013	Use of retrospective hospital data and patient comorbidities to program a ML algorithm to predict in-hospital mortality of neurosurgical patients post-TBI	Retrospective database study of TBI patients undergoing neurosurgical treatment between 1998 and 2009 in Taiwan.	*n*=16956 adult patients Mean age 50.8 (SD 21.4) 73.5% male Mortality rate 26.8% Exclusion criteria: multiple TBI procedures, cerebrovascular disease, incomplete data, age under 18 years old	ANN	LR
Rizoli *et al*. 2016	Use of ML model to predict unfavorable outcomes at 6 months post-TBI	Retrospective analysis of data from multicenter, double blind, randomized, and placebo-controlled trial conducted between 2006 and 2009	*n*=1089 Ages 15+ Blunt trauma and Severe TBI Inclusion: GCS < 8 upon admission Exclusion criteria: mortality within 24 h of ED admission, evidence of hemorrhagic shock	DT	ROC curves

(Contd...)

Table 1: *(Continued).*

Study	Outline	Design/Characteristics	TBI Population Characteristics	ML Models	Comparative Regression Models
Gravesteijn *et al*. 2020	Use of standard predictors for outcome of TBI to program ML on a large scale in comparison to traditional regression	Retrospective database study of TBI patients in the Netherlands between 1984 to 2004 and 2014 to 2018.	IMPACT-II *n*=11,002 (median age 31) Mortality rate 32% Unfavorable outcome rate 48% CENTER-TBI *n*=1375 (median age 48) Mortality rate 29% Unfavorable outcome rate 54%	SVM RF GBM ANN	LR Lasso Regression Ridge Regression
Zelnick *et al*. 2014	Secondary analysis of a multi-center, randomized, placebo controlled TBI clinical trial to evaluate patterns of missing outcome data, changes in functional status between hospital discharge and 6 months. Three prognostic models to predict long-term functional outcome from covariates available at hospital discharge (functional measures, demographics, and injury characteristics).	A secondary analysis of data from a multi-center, double-blind, randomized, and placebo-controlled trial conducted by ROC and administered under exception from informed consent. Thirteen regional clinical centers, 75 EMS agencies, and 53 hospitals in the US and Canada between May 2006 and May 2009	2 cohorts, 1 cohort with TBI, 1 w/hypovolemic shock. *n*=1282 enrolled patients w/blunt trauma, a prehospital GCS score of ≤8 and without hypovolemic shock. Patients ≥15 years old with blunt trauma and an out-of-hospital GCS score ≤8 were randomized to receive hypertonic saline/dextran, hypertonic saline, or normal saline in the out-of-hospital setting.	Model 1: included discharge GOSE only covariate Model 2 included discharge GOSE and length of hospital stay, both thought to be clinically important predictors of long-term functional outcome. Model3:included multiple covariates selected through exhaustive search using AIC values using nine predictors (Age, sex, discharge disposition, discharge disability rating scale (range, 0–30), discharge GOSE (range, 1–8), length of hospital stay	3 LR Models/le Cessie-van Houwelingen goodness-of-fit test Injury severity score (range, 0–75) Maximum head abbreviated injury severity score (range, 0–6) # of days alive out of ICU through day 28.
Raj *et al*. 2014	Comparison of a simple two-variable predictive model to more in-depth programs to predict mortality in adult moderate-severe TBI patients	Retrospective database study of TBI patients admitted to the ICU in Finland from 2003 to 2013	*n*=1625 (median age 55) Overall 6-month mortality 33% Exclusion criteria: age <16 years old, non-neurosurgical hospital, admission GCS >13, missing data, missing outcome 64% of 6-month mortality occurred during	APACHE II SAPS II SOFA SOFA adjusted)	LR

(Contd...)

Table 1: *(Continued).*

Study	Outline	Design/Characteristics	TBI Population Characteristics	ML Models	Comparative Regression Models
			hospital stay Split into two cohorts for development and validation Development *n*=844 (median age 56, mortality 33%) Validation *n*=781 (median age 54, mortality 34%)		

ML: Machine learning, VS: Vital signs, LMICs: Low-middle-income countries, TBI: Traumatic brain injury, ICU: Intensive care unit, GCS: Glasgow coma scale, GOS: Glasgow outcome scale, EMS: Emergency medical service, BANN: Bayesian artificial neural network, CT: Computed tomography, ED: Emergency department, LOS: Length of stay, MVA: Motor vehicle accident, SD: Standard deviation, APACHE II: Acute physiology and chronic health evaluation II, SAPS II: Simplified acute physiology score II, SOFA: Sequential organ failure assessment, ICP: Intracranial pressure, MAP: Mean arterial pressure, CPP: Cerebral perfusion pressure, LASSO: Least absolute shrinkage and selection operator, SVM: Support vector machine, SVR: Support vector machine, NB: Naïve Bayes, ANN: Artificial neural networks, DT: Decision tree, NNR: Nearest neighbor regression, RF: Random forest, GBM: Gradient boosting machine, AIC: Akaike information criterion, GOSE: Glasgow outcome scale extended, LR: Logistic regression, ROC: Receiver operating characteristic, PICU: Pediatric intensive care unit, KNN: K nearest neighbor, N/A/NA: Not applicable, *n*: Number of patients

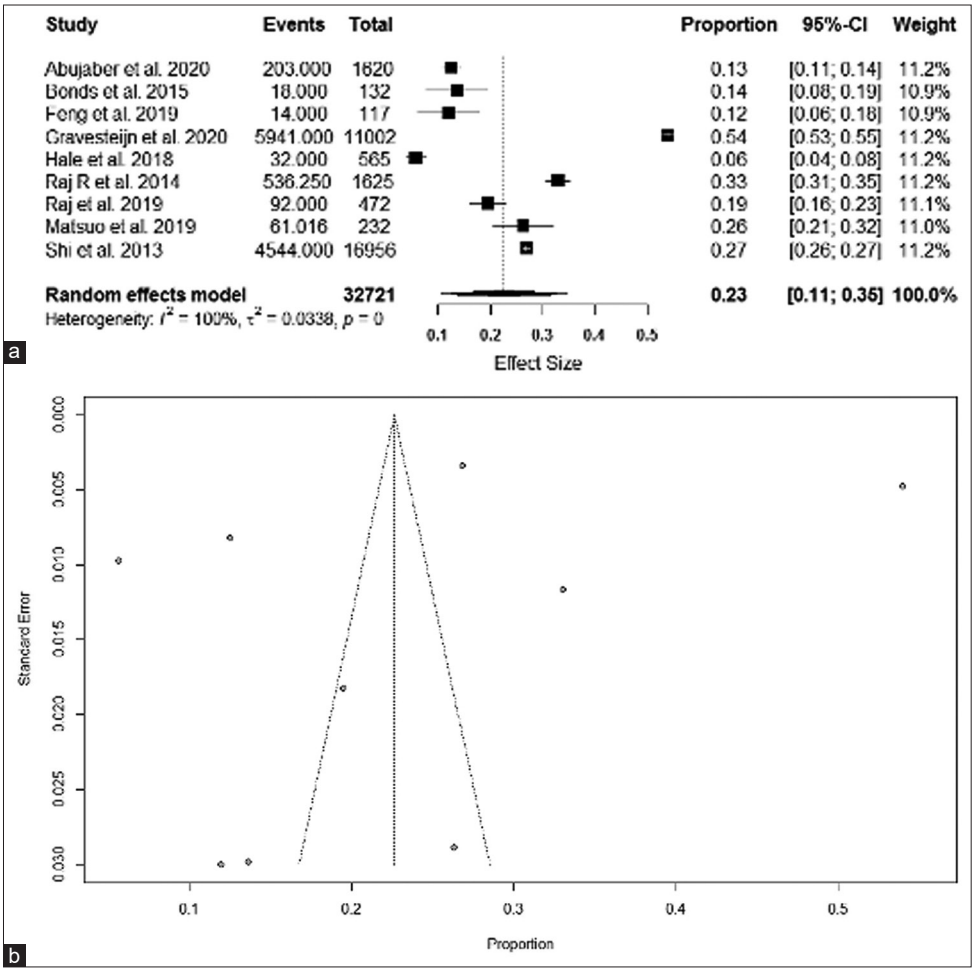

Figure 3: The forest (a) and funnels (b) plots of mortality data extracted from studies in which mortality was predicted using machine learning.

Table 2: Summary of included study findings.

Study	ML predictive variables	Outcome variables	ML model results	Study findings
Raj *et al.* 2019	ICP, MAP, CPP – measured as 5-min medians over 5 days GCS – measured every 24 h using ocular and motor components, verbal omitted due to intubation of patients	30-day all-cause mortality	ICP-CPP-MAP – AUC increased from 67% to 81% day 1 to day 5, 18 false positives potentially caused by decompressive craniectomy, 2 false positives, and 9 false negatives on mortality prediction ICP-CPP-MAP-GCS – AUC increased from 72% to 84% day 1 to day 5, 1 false positive and 4 false negatives on mortality prediction	Decreased mortality seen in younger patients presenting without mass lesion, with active pupillary reflex, with higher GCS at admission Decreased mortality seen in lower blood glucose, higher hemoglobin Increased mortality seen with higher ICP, lower CPP, no correlation to MAP Both dynamic algorithms were more effective at predicting 30 day mortality than the static variable model Limited mortality data in this study due to exclusion criteria of death within 36 h (withdrawal of treatment within last 12 h skewed ICP values)
Matsuo *et al.* 2019	Age, GCS, SBP, abnormal pupillary response, major extracranial injury, CT findings, blood lab values (glucose, C-reactive protein, fibrinogen degradation products) Most important variables identified (in order): Poor outcome: Age, GCS, FDP Mortality: FDP, GCS, Abnormal pupillary response	Unfavorable outcome as defined by GOS 1–3, death=1, persistent vegetative state=2, severe disability=3	Top training results – Poor outcome: Sensitivity 97.2% – RF Specificity 82.8% – Gaussian NB Accuracy 87.5% – Gradient boosting AUC 0.894 – SVM Top sample results – Poor outcome: RF – AUC and sensitivity Top training results – Mortality: Sensitivity 85.1% – ridge regression Specificity 99.3% – RF Accuracy 89.8% – SVM AUC 0.960 – RF Top sample results – Mortality ridge regression – sensitivity and AUC	ML adaptive algorithms were capable of predicting poor outcome and mortality following TBI at varying rates. RF was most effective at predicting poor outcome (100% sensitivity, 72.3% specificity, 91.7% accuracy, 0.895 AUC), ridge regression was most effective at predicting mortality (88.4% sensitivity, 88.2% specificity, 88.6% accuracy, 0.875 AUC). Dependence on CT findings was a major limitation due to time constraints. Cannot be implemented in LMIC due to dependence on CT technology. Age, GCS, FDP, and glucose were the most important parameters between poor outcome and mortality
Rau *et al.* 2018	Age Sex Helmet wearing status Coronary artery disease Congestive heart failure Cerebrovascular accident Diabetes mellitus End-stage renal disease Hypertension GCS Temperature SBP HR Respiratory rate Serum panel: WBC, RBC, hemoglobin, hematocrit, platelets, BUN, creatinine, ALT, AST, sodium, potassium, glucose Epidural hematoma Subdural hematoma Subarachnoid hematoma Intracerebral hematoma Injury severity score	Mortality	Training set: LR, SVR, NB, and ANN models had specificity >90%, accuracy >90%. ANN highest mortality prediction sensitivity (80.59%), highest AUC (0.968). Test set: ANN highest mortality prediction sensitivity (84.3%), followed by SVM, LR, NB, and DT	ANN model yielded best prediction of mortality for patients with moderate and severe TBI
Amorim *et al.* 2019	Gender Age Level of pupil reactivity at admission Prehospital GCS (at scene of injury) GCS at admission Motor component score of GCS Hypoxia Hypotension Midline shift bigger than 5 mm Brain herniation detected on CT (defined as effacement of the third ventricle or the basal cisterns)	Primary: 14-day mortality Secondary: hospital mortality, ICU length of stay, hospital length of stay	14-day mortality prediction NB had the best predictive performance (AUC=0.906), Bayesian generalized linear model (AUC=0.881) RF (AUC=0.880) Penalized discriminant analysis (AUC=0.880) In-hospital mortality prediction RF was the best performing model (AUC=0.838) Generalized partial least squares (AUC=0.831) Stochastic gradient boosting (AUC=0.823 Penalized discriminant analysis (AUC=0.803)	Prehospital GCS, GCS at admission, age, Glasgow Motor Score most impactful predictors of 14-day mortality. GCS at admission, age, prehospital GCS, and

(Contd...)

Table 2: *(Continued).*

Study	ML predictive variables	Outcome variables	ML model results	Study findings
	Subarachnoid hemorrhage Epidural hemorrhage Subdural hemorrhage Intracerebral hemorrhage Trauma severity Prothrombin time Partial thromboplastin time			thromboplastin time partial test most impactful predictors of in-hospital mortality. Prehospital GCS most important parameter across all models.
Feng *et al.* 2019	Age GCS (hospitalization) Injury severity score Temperature Systolic pressure Diastolic pressure Open brain injury or not Concussion presence Brain contusion presence Brain-stem injury presence Contrecoup presence Epidural hematoma presence Subdural hematoma presence Hematoma volume Intracerebral hematoma presence Brain hernia presence Oxygen saturation Infection complication presence Presence of other complications Number of surgeries Length of stay Length of ICU stay Multiple trauma presence Tracheotomy presence Period of mechanical ventilation Aspiration presence GCS (discharge) Previous TBI occurrence Hypothermia presence Acidosis presence Presence of hospital-acquired pneumonia WBC count Dose of glucose Glucocorticoid use Nasogastric tube use Coagulation change Parenteral nutrition use Lipid emulsion use Enteral nutrition time Sequelae or not	Mortality	**Algorithm / AUC / Accuracy / F Score / P / R** Quadratic discriminant 0.89 86.3% 0.92 96% 88% Linear discriminant 0.92 88.00% 0.93 96% 90% Fine KNN 0.72 88.90% 0.94 93% 94% Subspace KNN 0.84 88.90% 0.94 90% 98% Coarse KNN 0.47 88.0% 0.94 88% 100% Coarse Gaussian SVM 0.93 88.00% 0.94 88% 100% Medium Gaussian SVM 0.93 89.70% 0.94 90% 100% Fine Gaussian SVM 0.57 88.00% 0.94 88% 100% Cubic KNN 0.83 89.70% 0.94 90% 100% Boosted trees 0.30 88.0% 0.94 88% 100% Subspace discriminant 0.94 90.60% 0.95 95% 94% Simple tree 0.76 90.60% 0.95 94% 95% Medium tree 0.76 90.60% 0.95 94% 95% Complex tree 0.76 90.60% 0.95 94% 95% Cosine KNN 0.94 90.60% 0.95 93% 96% Medium KNN 0.88 90.60% 0.95 91% 99% RUSBoosted trees 0.91 92.30% 0.96 97% 94% Bagged trees 0.95 93.20% 0.96 94% 98% Weighted KNN 0.88 93.20% 0.96 94% 99% Logistic regression 0.83 88.00% 0.93 95% 91% Cubic SVM 0.94 94.00% 0.97 95% 98% Quadratic SVM 0.93 94.00% 0.97 95% 98% Linear SVM 0.93 94.00% 0.97 94% 99%	ML was more effective at predicting outcome of TBI than logistic regression. Cubic SVM, Quadratic SVM, and Linear SVM performed the best of the ML models in this study.
Abujaber *et al.* 2020	Race (Asian or other) Mechanism of injury Arrival method (ambulance or other) CT findings (SDH, EDH, SAH, hemorrhagic contusion, diffuse axonal injury) GCS (range: 15–13, 12–9, 8–3) Hospital shift (day or night) Presence of: midline shift, cerebral edema, facial bone fractures, lung contusion, hemothorax, pneumothorax, abdominal organ injury, known comorbidities, intubation, venous thromboembolism prophylaxis, blood transfusion	In-hospital mortality	Both ANN and SVM >91% accuracy. SVM outperformed ANN with accuracy 95.6% and AUC 96%. SVM outperformed traditional multivariate LR. SVM performance consistent with previous study (Senders *et al.*) indicative of good external validity. **Algorithm / Acc / AUC / PPV / NPV / Sens / Spec / F Score** SVM 95.6% 96% 88% 97% 73% 99% 0.8 ANN 91.6% 93.5% 66% 96% 62% 96% 0.64	SVM indicated endotracheal intubation most important predictor of in-hospital mortality. Positive relationship between HR and in hospital mortality. Mean HR for admission (93 bpm), survived (90.8 bpm), deceased (108.5 bpm)
Hale *et al.* 2018	GCS Serum glucose Serum hemoglobin Pupillary response Admission head CT results: SDH, ICH, IVH, cistern integrity, midline shift	6-month GOS Favorable >3 Unfavorable ≤3 6-month Mortality	Accuracy >91% AUC >93% 6-month GOS (AUC) Marshall CT – 0.663 Helsinki CT – 0.717 Rotterdam CT – 0.748 GCS – 0.855 ANN – 0.9462 6-month mortality (AUC) Marshall CT – 0.781 Helsinki CT – 0.814 Rotterdam CT – 0.838 GCS – 0.920 ANN – 0.9462	After training, ANN was more accurate than existing head CT models at predicting outcomes of pediatric TBI patients

(Contd...)

Table 2: *(Continued).*

Study	ML predictive variables	Outcome variables	ML model results	Study findings
Kayhanian *et al.* 2019	Serum glucose Serum hemoglobin Serum albumin Serum C-reactive protein Serum sodium Serum urea Serum magnesium Serum lactate Venous pH White cell count Neutrophil count Hematocrit Prothrombin time Activated partial thromboplastin time	GOS at 6 months Favorable >3 Unfavorable ≤3	LR (AUC, Specificity, and Sensitivity) – Prediction of favorable outcomes Focused: 0.83, 0.99, 0.71 Inclusive: 0.90, 0.99, 0.75 SVM (AUC, Specificity, and Sensitivity) – Prediction of favorable outcomes Focused: N/A, 0.99, 0.80 Inclusive: N/A, 1, 0.63 Note – AUC was not calculated for SVM	The ML model SVM using the three most meaningful serum measurements (H+, lactate, glucose) had greater specificity in predicting a favorable outcome in pediatric TBI patients than LR models High serum levels of H+ (acidosis), lactate (acidosis), and glucose (hyperglycemia) at admission were correlated with unfavorable outcomes
Donald *et al.* 2019	Age Sex MAP SBP HR	Hypotensive event Defined as SBP ≤90 mmHg or MAP ≤70 mmHg sustained for at least 5 min Event determined over when blood pressure returned to level above threshold for at least 5 min	AUC Test set: 0.74 With false-positive correction: 0.68 Without false-positive correction: 0.63 Target sensitivity of >30% was achieved Target specificity of >90% was achieved	The use of BANN was able to predict the presence of hypotensive events in at least 1 of every 3 occurrences before onset.
Bonds *et al.* 2015	HR SBP MAP SI ICP	ICP	Bland-Altman plots=NNR provides good agreement in predicting actual ICP with a bias of 0.02 (±2 SD=4 mm Hg) for the subsequent 5 min and −0.02 (±2 SD=10 mm Hg) for the subsequent 2 hours. Continuous VS were collected on 132 adult patients over a minimum of 3 h/patient (5,466 h total; 65,600 data points).	NNR predicts future ICP levels following severe TBI
Shi *et al.* 2013	Age Sex Carlson comorbidity Index Length of stay Hospital volume Surgeon volume	In-hospital mortality	ANN vs. LR Accuracy: 95.23% vs. 82.44% Hosmer-Leschow C statistic: 43.9 vs. 53.18 AUC: 89.61% vs. 77.39% Sensitivity: 67.56% vs. 54.83% Specificity: 95.23% vs. 92.67% PPV: 83.24% vs. 74.81% NPV: 89.35% vs. 87.64%	ANN outperformed LR in all metrics of comparison when looking at the ability to predict in-hospital mortality of TBI patients The most impactful variables input to the ML models were surgeon volume, hospital volume, and CCI in that order
Rizoli *et al.* 2016	Age Sex SBP Pupil reactivity at admission AIS severity Initial CT scan Marshall score (scale of 1–6)	eGOS 6 months post-injury Poor outcome: eGOS ≤4; (severe disability or death) Acceptable outcome: eGOS >4; (moderate or no disability)	Proposed Model vs. Core Model vs. Extended Model Sensitivity: 72.3% vs. 83.8% vs. 92.7% Specificity: 62.5% vs. 47.7% vs. 44.3% PPV: 74.0% vs. 70.3% vs. 71.1% NPV: 60.4% vs. 66.7% vs. 80.4%	Higher head AIS, Marshall score, and absence of pupillary reactivity among those with poor outcome eGOS≤4 Proposed model higher specificity, that is, more apt at classifying poor outcome
Gravesteijn *et al.* 2020	Age GCS Pupillary response CT classification Traumatic SAH Epidural hematoma Hypoxia Hypotension Serum glucose Serum sodium Serum hemoglobin	Mortality Unfavorable outcome (GOS <4)	Mortality: internal-external cross-validation LR: 0.81 Lasso regression: 0.81 Ridge regression: 0.81 SVM: 0.81 RF: 0.79 GBM: 0.81 ANN: 0.81 Unfavorable Outcome: internal-external cross-validation LR: 0.81 Lasso regression: 0.81 Ridge regression: 0.81 SVM: 0.80 RF: 0.79 GBM: 0.80 ANN: 0.80	ML models were no more or less effective than traditional regression in predicting mortality and unfavorable outcome in TBI patients Population of study seems to play a role in the effectiveness of ML for outcome prediction
Zelnick *et al.* 2014	Age, sex, ISS, maximum head AIS score, discharge GOSE score, discharge DRS score, length of hospital stay, number	Primary outcome measure was 6-mo	Poor outcomes (GOSE <4) 14 the primary outcome of 6-mo GOSE score was missing in 15% of enrolled patients. The characteristics of TBI patients with and without missing functional outcome	Performance was excellent (C-statistics ranged between 0.88 and 0.91) for all 3 prognostic models

(Contd...)

Table 2: *(Continued).*

Study	ML predictive variables	Outcome variables	ML model results	Study findings
	of days alive out of ICU through day 28 (ICU-free days), and discharge disposition (death, home, inpatient rehabilitation, skilled nursing facility, left against medical advice, or other). 6 mo GOSE outcome: discharge GOSE score could be calculated from responses from the patient alone, caregiver alone, patient and caregiver together or from hospital records. All highly predictive (C-statistic >0.88) of 6-month GOSE.	GOSE (gold standard for functional neurological outcome in study. severe disability or death: GOSE <4 (≤4: bad functional outcome) and moderate or no disability: GOSE ≥5 (good functional outcome) and DRS picks up more subtle changes over GOSE for functional outcomes. Discharged DRS is included in model 1 and model 2	(GOSE) at 6 months differed. Missing 6-mo GOSE generally being less severely injury. Comparing functional status at hospital discharge versus 6 months post-injury suggests that changes in mental function continue after the initial hospitalization. Findings also suggest that patient characteristics, injury severity, and function at hospital discharge can be used to accurately predict 6-month functional status for patients with missing primary outcome data	and calibration adequate for two models (P-values, 0.22 and 0.85). Results suggest multiple imputation of the standard 6-month GOSE may be reasonable in TBI research when the primary outcome cannot be obtained through other means. Pts with TBI enrolled in a trauma trial w/and w/o measured GOSE obtained at 6 months post-injury appear to be inherently different. Explored the reasons for missing outcomes and found that a majority were secondary to refusal or inability to obtain consent and a smaller portion being truly lost to follow-up. Thus, the mechanism of missingness for functional outcome was not "missing completely at random," where missing values have no association with observed or unobserved values. Although simplistic methods for handling missing data appear flawed, our results demonstrate that patient information available at or before hospital discharge adequately predicts 6-month GOSE. using a discharge GOSE in place of a 6-month GOSE appears not to be adequate by itself
Raj *et al.* 2014	Reference model: Age, lowest GCS within 24 h of admission APACHE II: Immunocompromization status, age, temperature, MAP, blood pH, HR, respiratory rate, serum values (sodium, potassium, and creatinine), hematocrit, WBC, acute renal failure status, FiO_2 ≤50%, and GCS SAPS II: Age, HR, SBP, temperature, GCS, mechanical ventilation pressure (if applicable), BUN, urine output, serum values (sodium, potassium, bicarbonate, and bilirubin), WBC, chronic disease, and type of admission SOFA: PaO_2, FiO_2, mechanical ventilation status, platelets, MAP, serum values (creatinine, bilirubin) SOFA Adjusted: SOFA + age + GCS	6-month mortality Secondary outcome: in-hospital mortality		SAPS II and APACHE II had significantly higher predictive power of in-hospital mortality as compared to LR by measure of AUC (P=0.011, P=0.013 respectively) There was no significant difference in predictive power of 6-month mortality by measure of AUC With both variables, the SOFA Adjusted scale outperformed the SOFA scale Overall the predictive power of the in-depth models was only slightly improved compared to traditional models sing only GCS and age

Algorithm	Discrimination		Calibration		Precision
	AUC	95% CI	H-L *P*-value	GiViTI *P*-value	Brier score
Development					
APACHE II	0.81	0.78, 0.84	0.153		0.160
SAPS II	0.81	0.77, 0.84	0.343		0.160
SOFA	0.68	0.64, 0.72	0.282		0.201
Adjusted SOFA	0.78	0.75, 0.81	0.444		0.175
Reference	0.75	0.72, 0.78	0.144		0.185
Validation					
APACHE II	0.79	0.75, 0.82	0.062	0.653	0.167
SAPS II	0.80	0.77, 0.83	0.775	0.782	0.166
SOFA	0.68	0.64, 0.72	0.691	0.710	0.201

ICP: Intracranial pressure, MAP: Mean arterial pressure, CPP: Cerebral perfusion pressure, GCS: Glasgow coma scale, SBP: Systolic blood pressure, CT: Computed tomography, WBC: White blood cell, RBC: Red blood cell, BUN: Blood urea nitrogen, ALT: Alanine transaminase, AST: Aspartate transaminase, SDH: Subdural hemorrhage, EDH: Extradural hemorrhage, SAH: Subarachnoid hemorrhage, IVH: Intraventricular hemorrhage, ICH: Intracerebral hemorrhage, MAP: Mean arterial pressure, HR: Heart rate, SI: Shock index, ISS: Injury severity score, AIS: Abbreviated injury scale, DRS: Disability rating scale, ICU: Intensive care unit, GOSE: Glasgow outcome scale extended, AUC: Area under curve, LMICs: Low-middle-income countries, APACHE II: Acute physiology and chronic health evaluation II, TBI: Traumatic brain injury, ML: Machine learning, SAPS II: Simplified acute physiology score II, LR: Logistic regression, SVM: Support vector machine, GOS: Glasgow outcome scale, PPV: Positive predictive value, NPV: Negative predictive value, NB: Naïve Bayes, RF: Random forest, SD: Standard deviation, KNN: K nearest neighbor, ANN: Artificial neural networks, eGOS: Glasgow outcome scale extended, VS: Vital signs, SOFA: Sequential organ failure assessment, , FDP: Fibrin Degradation Products, DT: Decision Tree, LR: Logistic Regression, NNR: Neural Network Regression, H+: Hydrogen ion, GBM: Gradient boosting machine, CCI: Charlson Comorbidity Index, AIS: Abbreviated Injury Scale, FiO2: Fraction of Inspired Oxygen, PaO2: Partial Pressure of Oxygen, CI: Confidence Interval, H-L: Hosmer-Lemeshow Test: TBI: Traumatic Brain Injury, SVR: Support vector machine, GiViTI: The Italian Group for the Evaluation of Intervention in Intensive Care Medicine calibration belt.

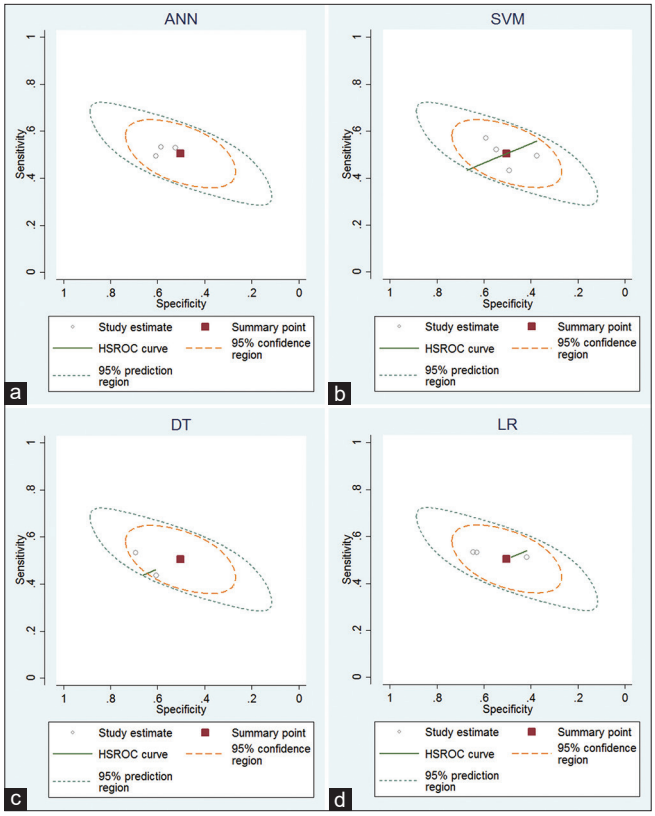

Figure 4: (a-d) Meta-analysis of diagnostic accuracy with hierarchical modeling for machine learning models across the selected studies. ANN: Artificial neural network, SVM: Support vector machine, DT: Decision tree, LR: Logistic regression.

regularized least squares and linear regression to 9 ML models, the models consistently demonstrated an AUC above 0.80. Naïve Bayes (NB) had the highest predictive performance model for 14-day mortality prediction with an AUC of 0.906. In-hospital mortality prediction was best predicted by random forest (RF) with an AUC value of 0.838.[4] In a separate study, NB was again used for mortality prediction with an accuracy of >90% in the training set.[36] ANN had the highest AUC (0.968) with a prediction sensitivity of 80.59%. In the test set, ANN remained the highest predictor of mortality followed by the additional three models used in the study: SVR, NB, and DT.[36]

30-day mortality after TBI

Two custom models were developed to predict the 30-day mortality of TBI patients using commonplace neurointensive care unit measurements as predictive variables.[34] ML model variables include a combination of intracranial pressure (ICP), mean arterial pressure (MAP), and cerebral perfusion pressure (CPP) all measured in 5 min medians over 5 days. The second model included ICP, CPP, and MAP and GCS. Model 1 (ICP-CPP-MAP) had an AUC that increased from 67% to 81% from day 1 to day 5. False positives and false negatives also influenced mortality prediction. There were 18 false positives potentially caused by decompressive craniectomy, two additional false positives, and nine false negatives on mortality prediction. The second model (ICP-CPP-MAP-GCS) had an AUC that increased from 72% to 84% from day 1 to day 5, with 1 false positive and 4 false negatives on mortality prediction.[34] Another study performed a secondary analysis of a multi-center, randomized, placebo-controlled clinical trial of TBI patients to evaluate patterns of missing outcome data, and changes in functional status between hospital discharge and 6 months follow-up.[49] Three novel prognostic models were developed to predict long-term functional outcome from covariates available at hospital discharge. The ML predictive variables included the Glasgow outcome scale extended (eGOS) and the disability rating scale (DRS). An adverse outcome was defined as eGOS less than or equal to four. In both models, discharge DRS was used. ML model results included missing data for poor outcomes for 15% of enrolled patients. The model performance was excellent (C-statistic between 0.88 and 0.91) for all three prognostic models and calibration was adequate for two models (P = 0.22 and 0.85). A two-variable predictive model was compared to more in-depth programs to predict mortality in adult moderate-severe TBI patients in a retrospective database study.[35] Notably, 64% of 6-month mortality occurred during hospital stay.[35] Data were split for development and validation. ML models used to predict mortality were Acute Physiology and Chronic Health Evaluation II (APACHE II), Simplified Acute Physiology Score II (SAPS II), Sequential Organ Failure Assessment (SOFA), and SOFA Adjusted. The reference model, LR, scored an AUC of 0.75 in development. During validation, SAPS II and APACHE II scored higher than LR, whereas SOFA scored AUC of 0.68. The AUC of the SAPS II was 0.80 (95% CI 0.77–0.83).[35]

Unfavorable outcomes at 6 months

DT was used as the ML model to predict unfavorable outcomes at 6 months post-TBI.[37] The comparative model was ROC curves. ML predictive variables included various vital signs, pupil reactivity, AIS severity, initial CT scan, and Marshall score (scale of 1–6). When considering eGOS 6 months post-injury, an acceptable outcome was labeled as an eGOS score of greater than four indicating moderate or no disability. A poor outcome indicated a severe disability or death, a score of four or less. The proposed model had a specificity of 62.5%, which was higher than the core model (47.7%) and extended model (44.3%).[37] The proposed model had the highest positive predictive value of 74.0% and the extended model had a negative predictive value of 80.4%. The sensitivity was also higher in the extended model (92.7%) when compared to the proposed model (72.3%) and core model (83.8%).[37] ML and LR were used to assess mortality as an unfavorable outcome using the GOS score of <4.[15] ML

predictive values include vital signs, GCS, pupillary response, and CT Classification. Prediction of mortality was measured using internal-external cross-validation, with all regressions and ML models scoring 0.81 except RF, which scored 0.79. The cross-validation of unfavorable outcomes included all regression models scoring 0.81 and all ML models scoring 0.80 except RF, which scored 0.79.[15] Various ML models were used to predict unfavorable outcomes and mortality in an additional study.[27] Unfavorable outcomes were defined as a GOS score of 1–3 (death = 1, persistent vegetative state = 2, and severe disability = 3). The RF ML model was the most effective at predicting poor outcomes (100% sensitivity, 72.3% specificity, 91.7% accuracy, and 0.895 AUC). Ridge regression (RR) was most effective at predicting mortality (88.4% sensitivity, 88.2% specificity, 88.6% accuracy, and 0.875 AUC).[27]

Prediction of outcomes in the pediatric population

ANN was compared to traditional head computed tomography (CT) analysis (i.e., Marshall CT, Helsinki CT, and Rotterdam CT) and GCS to predict adverse outcomes and mortality in pediatric TBI patients.[16] ML predictive variables included GCS, serum glucose, serum hemoglobin, pupillary response, and admission head CT results: subdural hematoma, intracranial hemorrhage, intraventricular hemorrhage, cistern integrity, and midline shift. The AUC using ANN was 0.9462 when predicting mortality and adverse outcomes, defined as a 6-month GOS ≤3.[35] The CT results ranged from an AUC of 0.781–0.838. The GCS had an AUC of 0.920.[16] An additional study used serum metabolic markers to program an ML algorithm to predict unfavorable GOS in pediatric TBI patients with both SVM and LR.[21] Both models were programmed as both a focused (only used pH, lactate, and glucose) and an inclusive algorithm using serum metabolic markers for prediction. AUC was not calculated for SVM. When predicting favorable outcomes, SVM scored a specificity of 0.99 and sensitivity of 0.80 using the focused model. The inclusive model for SVM had higher specificity with a value of 1 and the sensitivity was lower with a value of 0.63. LR predicted favorable outcomes using the focused model with an AUC of 0.83, specificity of 0.99, and sensitivity of 0.75.[21] The LR inclusive model scored higher across AUC, specificity, and sensitivity.

Prediction of secondary insults: ICP, hypotensive events, and shock index (SI)

Two studies used vital signs as outcome variables for ML. Bayesian Artificial Neural Network (BANN) was used to assess blood pressure values to develop a predictive model for hypotensive events in TBI patients in the neuro-intensive care unit.[45] Hypotensive events were described as an SBP ≤90 mmHg and MAP ≤70 mmHg sustained for at least 5 min. A hypotensive event ceases when blood pressure returns to a level above threshold/baseline for at least 5 min.[45] Vital signs were also used for the prediction of secondary insult following severe TBI.[7] Using the Nearest Neighbor Regression (NNR), ML model predictive variables such as SI and ICP were assessed.[7] Both studies assessed HR, SBP, and MAP. AUC values for BANN were as follows: test set 0.74, false-positive correction 0.68, without false-positive correction 0.63. Finally, the target sensitivity of >30% and specificity of >90% were achieved.[45] Using NNR, the other study found good agreement in predicting actual ICP with a bias of 0.02 (±2 standard deviation [SD] = 4 mm Hg) for the subsequent 5 min and −0.02 (±2 SD = 10 mm Hg) for the subsequent 2 h. The patient's vital signs were continuously collected on 132 adult patients over a minimum of 3 h/patient (*5,466 h total; 65,600 data points*).[7] However, ANN was the most effective model for the prediction of hypotensive events in critical care patients.[45]

DISCUSSION

In the present systematic review and meta-analysis, we evaluate the predictive power of various ML algorithms for unfavorable outcomes and mortality in patients with TBI. Several studies have demonstrated the utility of ML in medicine; however, most TBI studies were focused on diagnosis and classification.[17,30,44,48] The 15 studies included in this review sought to expand the use of ML in TBI patients with a focus on mortality and unfavorable outcome prediction. ML algorithms encountering in this review including SVM, ANN, RF, NNR, and NB.

TBI remains one of the leading causes of death and disability throughout the world.[10,19,38] It is estimated that as many as 50 million people experience TBI each year.[10,13,19,25] TBI is a trimodal class of injury, affecting young children (falls), adults (motor vehicle accidents), and the elderly (falls) at high rates compared to other injuries.[13,24,31,47] TBI can result in multiple deficits, ranging from motor to sensory, and often affects cognition and memory.[33,47] Causes of brain damage can also include hemorrhagic infarct, cerebral edema, and crush injuries to the brain and brainstem.[31,47] Clinical practice has also shown that immediate treatment is dependent on clear and accurate neuroimaging, and ML algorithms have been designed to diagnose and classify TBI using radiological findings.[8,44] Depending on the severity of the injury, deficits from TBI can be permanent, or require intensive care and extensive rehabilitation.[24,29,31,47] However, neurointensive care and rehabilitation are expensive, time-consuming, and require significant effort from both the patient and their caregivers. In addition, TBI mortality is high, reaching up to 30–40% in severe TBI, and lifelong deficits are reported in approximately 60% of patients who recover.[5,6,11,13,24-26,29,47]

ML algorithms including RF, RR, and NB were all identified as effective prediction models for the unfavorable outcome or in-hospital mortality in TBI patients.[4,12] SVM was identified in multiple studies as more effective than LR in predicting mortality and unfavorable outcome using GOS.[2,14,21] Interestingly, using the NNR model, ICP fluctuations were more effectively predicted compared to traditional LR models.[7] ANN outperformed LR and other ML models in the prediction of mortality in moderate and severe TBI patients.[36,43] A large-scale database study identified no difference in the predictive power of both SVM and ANN; however, the authors did hypothesize that the predictive power identified in other studies may be population-dependent.[15]

The use of scoring algorithms SAPS II, APACHE II, and SOFA was found to have increased predictive power of in-hospital mortality over LR, but no significant difference with overall 6-month mortality.[35] These results indicate that the benefit of ML in the prediction of outcomes may be limited to short-term complications such as in-hospital mortality and major complications. However, this information is still valuable when making clinical decisions surrounding how to treat these patients. In addition, these models outperformed the predictive power of the IMPACT II TBI database. However, these models found little influence from the input of MAP values. Furthermore, the inclusion of GCS improved the accuracy.[34] The C-statistic of models for prediction of eGOS at 6 months after discharge improved with the addition of new input variables. Discharge eGOS was used as a baseline, and with the addition of hospital length of stay as well as age, predictive power improved.[29] These findings lend further support to the importance of age and admission GCS in TBI prognosis.

The heterogeneity of input variables between ML models and studies limits the potential for cross-comparison. Those studies that had compatible methodologies were included in a small meta-analysis in an attempt to draw a quantitative conclusion regarding which model best predicts mortality. However, with the heterogeneity of input variables, inconsistency of outcome measurement, and variable criteria for TBI classification, this cannot be generalized to all TBI mortality predictions. A further prospective study with an increased sample size is necessary to definitively state, in which ML model is objectively most effective at mortality prediction. Furthermore, future studies should seek to standardize the necessary input variables for the operation of ML models. There is great inconsistency among the presented studies in the selection of input variables, with some studies only utilizing a few simple serum studies. While convenient for the provider, this limited input data may fail to capture a complete picture of the patient's current condition. On the other hand, multiple studies employ a myriad of input variables including information that may not be easily accessible in an emergent situation, such as detailed radiological findings and hospital staffing statistics.[21,24,26,27] While many of these variables are employed for training the model, in practice, this level of detail is not feasible in emergent cases where these models could be most beneficial, such as emergency room triage. Based on the common variables between the analyzed studies and their individual analysis of which variables were most impactful, we would recommend studying the efficacy of models when programmed with patient age, admission GCS, serum lactate, and serum glucose.[18-22] While multiple studies within the review employed blood pressure measurements for programming, these were not found to be significant prognostic factors when programming the ML models to predict adverse events.[19]

CONCLUSION

TBI continues to be one of the leading causes of death and disability worldwide. This study reiterates the clinical utility of ML as an adjunct in patients with TBI. The use of ML to predict outcomes following TBI is entering clinical practice at an increasing rate and the present study reinforces the utility of these models. Using these models, simple admission data can be used to accurately predict the prognosis for individual patients. This can ultimately enhance the clinical decision-making process in terms of whether surgical intervention, medical management, or palliative care is most appropriate. There was a lack of consistency among the investigated studies with the selection of input variables used for predictive models; as a result, some models simply had more data to utilize for prediction, making inter-study comparisons more difficult. Further, research should utilize the core prediction variables identified in this review and apply these markers across a wide range of models and in multiple clinical settings. Given that the described models have demonstrated a robust ability to predict outcomes, there exists a significant degree of untapped potential in implementing ML to aid in neurosurgical decision-making. It is conceivable that these tools can be further expanded to guide and optimize patient treatment and perhaps alert neuro-care providers of patients at high risk of early neurological deterioration. Despite the increased use and predictive power of ML, it remains to be seen whether clinicians will routinely incorporate these models to guide clinical care following TBI.

Declaration of patient consent

Patient's consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. 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;20:336.
2. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. 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.
3. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, *et al*. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med 2022;28:1455-60.
4. Amorim RL, Oliveira LM, Malbouisson LM, Nagumo MM, Simoes M, Miranda L, *et al*. Prediction of early TBI mortality using a machine learning approach in a LMIC population. Front Neurol 2020;10:1366.
5. Andelic N, Howe EI, Hellstrøm T, Sanchez MF, Lu J, Løvstad M, *et al*. Disability and quality of life 20 years after traumatic brain injury. Brain Behav 2018;8:e01018.
6. Areas FZ, Schwarzbold ML, Diaz AP, Rodrigues IK, Sousa DS, Ferreira CL, *et al*. Predictors of hospital mortality and the related burden of disease in severe traumatic brain injury: A prospective multicentric study in Brazil. Front Neurol 2019;10:432.
7. Bonds BW, Yang S, Hu PF, Kalpakis K, Stansbury LG, Scalea TM, *et al*. Predicting secondary insults after severe traumatic brain injury. J Trauma Acute Care Surg 2015;79:85-90.
8. Carlsen RW, Daphalapurkar NP. The importance of structural anisotropy in computational models of traumatic brain injury. Front Neurol 2015;6:28.
9. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44:368-74.
10. Dewan MC, Rattani A, Gupta S, Baticulon RE, Hung YC, Punchak M, *et al*. Estimating the global incidence of traumatic brain injury. J Neurosurg 2019;130:1080-97.
11. Dikmen SS, Machamer JE, Powell JM, Temkin NR. Outcome 3 to 5 years after moderate to severe traumatic brain injury 11 No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated. Arch Phys Med Rehabil 2003;84:1449-57.
12. Fartoumi S, Emeriaud G, Roumeliotis N, Brossier D, Sawan M. Computerized decision support system for traumatic brain injury management. J Pediatr Intensive Care 2015;5:101-7.
13. Faul M, Coronado V. Epidemiology of traumatic brain injury. In: Handbook of Clinical Neurology. Vol. 127. USA: Elsevier; 2015. p. 3-13.
14. Feng J, Wang Y, Peng J, Sun M, Zeng J, Jiang H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. J Crit Care 2019;54:110-6.
15. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, *et al*. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95-107.
16. Hale AT, Stonko DP, Brown A, Lim J, Voce DJ, Gannon SR, *et al*. Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury. Neurosurg Focus 2018;45:E2.
17. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. Stroke 2019;50:1263-5.
18. Herasevich V, Tsapenko M, Kojicic M, Ahmed A, Kashyap R, Venkata C, *et al*. Limiting ventilator-induced lung injury through individual electronic medical record surveillance. Crit Care Med 2011;39:34-9.
19. Hyder AA, Wunderlich CA, Puvanachandra P, Gururaj G, Kobusingye OC. The impact of traumatic brain injuries: A global perspective. NeuroRehabilitation 2007;22:341-53.
20. Hyland SL, Faltys M, Hüser M, Lyu X, Gumbsch T, Esteban C, *et al*. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med 2020;26:364-73.
21. Kayhanian S, Young AM, Mangla C, Jalloh I, Fernandes HM, Garnett MR, *et al*. Modelling outcomes after paediatric brain injury with admission laboratory values: A machine-learning approach. Pediatr Res 2019;86:641-5.
22. Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers-part I. General guidance and tips. Korean J Radiol 2015;16:1175-87.
23. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers-part II. Statistical methods of meta-analysis. Korean J Radiol 2015;16:1188-96.
24. Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. Lancet Neurol 2008;7:728-41.
25. Majdan M, Plancikova D, Brazinova A, Rusnak M, Nieboer D, Feigin V, *et al*. Epidemiology of traumatic brain injuries in Europe: A cross-sectional analysis. Lancet Public Health 2016;1:e76-83.
26. Martins ET, Linhares MN, Sousa DS, Schroeder HK, Meinerz J, Rigo LA, *et al*. Mortality in severe traumatic brain injury: A multivariated analysis of 748 Brazilian patients from Florianópolis city. J Trauma 2009;67:85-90.
27. 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;37:202-10.
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. Open Med 2009;3:e123-30.
29. Muballe KD, Sewani-Rusike CR, Longo-Mbenza B, Iputo J. Predictors of recovery in moderate to severe traumatic brain injury. J Neurosurg 2019;131:1648-57.

30. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am Heart J 2020;229:1-17.
31. Pearn ML, Niesman IR, Egawa J, Sawada A, Almenar-Queralt A, Shah SB, *et al*. Pathophysiology associated with traumatic brain injury: Current treatments and potential novel therapeutics. Cell Mol Neurobiol 2017;37:571-85.
32. PRISMA-P Group, Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.
33. Rabinowitz AR, Levin HS. Cognitive sequelae of traumatic brain injury. Psychiatr Clin North Am 2014;37:1-11.
34. Raj R, Luostarinen T, Pursiainen E, Posti JP, Takala RS, Bendel S, *et al*. Machine learning-based dynamic mortality prediction after traumatic brain injury. Sci Rep 2019;9:17672.
35. Raj R, Skrifvars M, Bendel S, Selander T, Kivisaari R, Siironen J, *et al*. Predicting six-month mortality of patients with traumatic brain injury: Usefulness of common intensive care severity scores. Crit Care 2014;18:R60.
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. Rizoli S, Petersen A, Bulger E, Coimbra R, Kerby JD, Minei J, *et al*. Early prediction of outcome after severe traumatic brain injury: A simple and practical model. BMC Emerg Med 2016;16:32.
38. Rusnak M. Giving voice to a silent epidemic. Nat Rev Neurol 2013;9:186-7.
39. Sawyer AM, Deal EN, Labelle AJ, Witt C, Thiel SW, Heard K, *et al*. Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. Crit Care Med 2011;39:469-73.
40. Senders JT, Arnaout O, Karhade AV, Dasenbrock HH, Gormley WB, Broekman ML, *et al*. Natural and artificial intelligence in neurosurgery: A systematic review. Neurosurgery 2018;83:181-92.
41. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman ML, *et al*. Machine learning and neurosurgical outcome prediction: A systematic review. World Neurosurg 2018;109:476-86.e1.
42. Senders JT, Zaki MM, Karhade AV, Chang B, Gormley WB, Broekman ML, *et al*. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir 2018;160:29-38.
43. 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;118:746-52.
44. Stone JR, Wilde EA, Taylor BA, Tate DF, Levin H, Bigler ED, *et al*. Supervised learning technique for the automated identification of white matter hyperintensities in traumatic brain injury. Brain Inj 2016;30:1458-68.
45. The Brain IT Group, Donald R, Howells T, Piper I, Enblad P, Nilsson P, *et al*. Forewarning of hypotensive events using a Bayesian artificial neural network in neurocritical care. J Clin Monit Comput 2019;33:39-51.
46. Whiting PF. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529.
47. Williams OH, Tallantyre EC, Robertson NP. Traumatic brain injury: Pathophysiology, clinical outcome and treatment. J Neurol 2015;262:1394-6.
48. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: Lessons from machine learning. Perspect Psychol Sci 2017;12:1100-22.
49. Zelnick LR, Morrison LJ, Devlin SM, Bulger EM, Brasel KJ, Sheehan K, *et al*. Addressing the challenges of obtaining functional outcomes in traumatic brain injury research: Missing data patterns, timing of follow-up, and three prognostic models. J Neurotrauma 2014;31:1029-38.

How to cite this article: Courville E, Kazim SF, Vellek J, Tarawneh O, Stack J, Roster K, *et al*. Machine learning algorithms for predicting outcomes of traumatic brain injury: A systematic review and meta-analysis. Surg Neurol Int 2023;14:262.