Construction of a predictive model based on MIV-SVR for prognosis and length of stay in patients with traumatic brain injury: Retrospective cohort study

Yifeng Pan[1],[*] (ID), Cheng Fang[2],[*], Xueling Zhu[1] and Jinghai Wan[2]

Abstract

Objective: To investigate the mean impact value (MIV) method for discerning the most efficacious input variables for the machine learning (ML) model. Subsequently, various ML algorithms are harnessed to formulate a more accurate predictive model that can forecast both the prognosis and the length of hospital stay for patients suffering from traumatic brain injury (TBI).

Design: Retrospective cohort study.

Participants: The study retrospectively accrued data from 1128 cases of patients who sought medical intervention at the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University, within the timeframe spanning from May 2017 to May 2022.

Methods: We performed a retrospective analysis of patient data obtained from the Neurosurgery Center of the Second Hospital of Anhui Medical University, covering the period from May 2017 to May 2022. Following meticulous data filtration and partitioning, 70% of the data were allocated for model training, while the remaining 30% served for model evaluation. During the construction phase of the ML models, a gamut of 11 independent variables—including, but not limited to, in-hospital complications and patient age—were utilized as input variables. Conversely, the length of stay (LOS) and the Glasgow Outcome Scale (GOS) scores were designated as output variables. The model architecture was initially refined through the MIV methodology to identify optimal input variables, whereupon five distinct predictive models were constructed, encompassing support vector regression (SVR), convolutional neural networks (CNN), backpropagation (BP) neural networks, artificial neural networks (ANN) and logistic regression (LR). Ultimately, SVR emerged as the most proficient predictive model and was further authenticated through an external dataset obtained from the First Hospital of Anhui Medical University.

Results: Upon incorporating the optimal input variables as ascertained through MIV, it was observed that the SVR model exhibited remarkable predictive prowess. Specifically, the Mean Absolute Percentage Error (MAPE) of the SVR model in predicting the GOS score in the test dataset is only 6.30%, and the MAPE in the external validation set is only 7.61%. In terms of predicting hospitalization time, the accuracy of the test and external validation sets were 9.28% and 7.91%, respectively. This error indicator is significantly lower than the error of other prediction models, thus proving the excellent efficacy and clinical reliability of the MIV-optimized SVR model.

Conclusion: This study unequivocally substantiates that the incorporation of MIV for selecting optimal input variables can substantially augment the predictive accuracy of machine learning models. Among the models examined, the MIV-SVR model emerged as the most accurate and clinically applicable, thereby rendering it highly conducive for future clinical decision-making processes.

[1]The School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei, China
[2]Department of Neurosurgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[*]Yifeng Pan and Cheng Fang are co-first authors

Corresponding author:
Jinghai Wan, Department of Neurosurgery, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China.
Email: wanjinghai@sina.com

Introduction

Traumatic brain injury (TBI) constitutes a critical source of morbidity and mortality, responsible for substantial adverse outcomes among various forms of injuries.[1,2] Globally, TBI afflicts over 50 million individuals annually, with epidemiological data suggesting that nearly half of the global populace is predisposed to experiencing one or more such injuries in their lifetime.[3] A seminal study led by Professor Jiang Jiyao's research group at Renji Hospital, affiliated with Shanghai Jiao Tong University School of Medicine, delineated the foundational characteristics of Chinese TBI patients.[4] Their research illuminated that the morbidity and mortality rates among TBI patients in intensive care units approximated 11.4%, while the majority of survivors demonstrated a reasonable timeframe for recovery. Despite these findings, resource constraints and relative paucity of medical expertise in low- and middle-income countries exacerbate the burden borne by TBI patients.[5] Consequently, the imperative for risk stratification of TBI patients—contingent on clinical variables and patients' physiological conditions—and the development of predictive tools for prognostic outcomes and length of stay cannot be overstated. Such measures are quintessential for achieving precision medicine, curtailing non-essential medical expenditure and mitigating societal healthcare burdens.

The contemporary ascendancy of machine learning has catalyzed the efficacious utilization of hitherto accumulated voluminous clinical datasets. Predictive models, engendered through machine learning algorithms, have gained pervasive applicability in the domain of neurocranial trauma, frequently yielding commendable outcomes.[6] However, the extant literature predominantly relies on past empirical analyses and heuristic experiences for the selection of input variables, often eschewing further rigorous scrutiny to ascertain the optimality of these predictive models.[7–13] Against this backdrop, the present investigation employs the mean impact value (MIV) methodology[14] as a novel screening mechanism for input variables. The yardstick for this screening is the predictive accuracy of the model under consideration. Subsequent to this initial filtration process, the curated input variables are harnessed for the construction of an assortment of machine learning models. The final stage involves a comparative evaluation aimed at identifying the most efficacious predictive model.

Methods

Study design and data source

This investigation is a retrospective cohort study and has received the requisite ethical clearance from the Institutional Review Board of the Second Affiliated Hospital of Anhui Medical University. Informed consent was diligently procured from the participants or their authorized representatives within a temporal window of 24 hours post-admission.

The study retrospectively accrued data from 1128 cases of patients who sought medical intervention at the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University, within the timeframe spanning from May 2017 to May 2022. The criteria delineating patient inclusion are systematically outlined in Table 1. Owing to the nature of our model's predictive focus—encompassing both patient prognosis and duration of hospital stay—discharge decisions necessitated interdisciplinary consultation involving the treatment team and rigorous evaluative measures by seasoned neurosurgical experts. Accordingly, exclusion criteria were judiciously

Table 1. Criteria for patient inclusion.

Inclusion criteria
1 Craniocerebral trauma as a result of external forces
2 Clinical diagnosis of craniocerebral trauma
3 Traumatic brain injury occurred between May 2017 and May 2022
4 Consulted at the Second Affiliated Hospital of Anhui Medical University
5 Complete clinical information such as cases, course records, imaging examinations and test reports are available

implemented to account for the following scenarios: (1) instances where the family of the patient opts to discontinue treatment, inclusive of economic constraints, resulting in self-discharge; (2) cases necessitating further intervention from specialized departments due to severe injuries in anatomical regions beyond the cranium; and (3) Patients with an antecedent history of neurocranial trauma. Post-application of these exclusion criteria, a refined dataset comprising 1001 patient records, was retained for subsequent analysis. In the process of emergency rescue, patients are often combined with severe trauma in other parts of the body in addition to TBI, and the specificity of the human brain also determines the sequence of emergency surgery. So TBI are dealt with first, and patients are often lost to follow-up when they need to be transferred to other hospitals or departments for surgery on other parts of the body after craniocerebral surgery, so this is one of the exclusion criteria. When patients have minor injuries elsewhere that do not require surgical treatment, they can be discharged from the neurosurgical ward with conservative treatment to heal (e.g. plaster immobilization after fracture), but these patients often require longer hospitalization for treatment and recovery, which will have an impact on the results of the length of hospital stay in this study, and so they are used as the input dataset for the prediction model, i.e. the input data (x8) in Table 2.

Additionally, for purposes of external model validation, clinical data from 111 patients treated at the First Affiliated Hospital of Anhui Medical University were collected.

Predictor variables

The conceptual framework guiding the initial selection of both input and output variables for the model was informed by our antecedent research endeavors.[15] The dataset for model construction incorporated a multifaceted array of parameters, encapsulating patients' general condition as well as their clinical and imaging data post-admission for craniocerebral trauma. Specifically, we operationalized 11 variable groups, including gender, age, pre-existing medical history, mechanism underlying the traumatic brain injury, presence or absence of post-injury loss of consciousness, admission Glasgow Coma Scale (GCS) score, cranial computed tomography (CT) findings at admission, surgical interventions pertaining to the brain, concomitant injuries in anatomical sites other than the cranium, the administration of intensive care treatment and its associated duration. These variables constituted the input data for the machine learning model. The categorical taxonomy and definitional attributes of the variables employed in model construction are elaborated upon in Table 2.

The predictive variables in this study are all discrete variables, and it is necessary to perform one-hot encoding on the discrete data. One-hot encoding is the representation of categorical variables as binary vectors. Firstly, the classification values are mapped to integer values, and then each integer value is represented as a binary vector, with zero values except for the index of integers. One-hot encoding has the advantage of calculating the distance between features more reasonably, which is conducive to the establishment and development of the prediction model in this study, and is suitable for processing discrete data. Therefore, this study uses one-hot encoding to process predicted variables, using the processed feature vectors as inputs to the model.

Outcome indicators

The outcome indicators predicted by the model include length of stay and Glasgow Outcome Scale (GOS) score. Length of stay usually refers to the duration of a patient's stay from admission to discharge, usually measured in days. This study directly extracted the length of stay of patients with traumatic brain injury from the hospital's electronic medical record system. GOS is a scoring system used to evaluate the recovery and functional outcomes after brain injury. It is mainly used to evaluate the overall functional recovery of patients with severe brain injury. The GOS score in this study was obtained at the time of discharge through detailed assessments of patients with traumatic brain injury conducted by physicians, nurses or other medical professionals. It is worth noting that the GOS score serves as a robust and widely accepted prognostic instrument for assessing global disability and recovery trajectories following traumatic brain injury.[16] The score employs a categorical scale, where a rating of 1 denotes mortality, while scores of 4 and 5 are indicative of favorable recovery outcomes. Conversely, scores of 2 and 3 portend a more deleterious prognosis.

Research design

Section 2.1 provides a detailed overview of our data collection methods. This section will describe the broader research design, highlighting the framework, objectives and guiding principles of our study. The overall design can be visualized in Figure 1.

To start, the raw dataset is filtered using the mean impact value (MIV) method. This method ensures that we have an optimal dataset for machine learning applications. In essence, MIV evaluates the significance of each data feature by calculating its mean impact value over multiple iterations.[17,18] With MIV's insights, we can better understand the importance of each feature, simplifying feature selection and model tuning. The filtered data is then used to train various machine learning models, focusing on predicting the GOS scores and the duration of a patient's hospital stay. We chose to test the following models: support vector regression machine (SVR), convolutional neural network (CNN), backpropagation (BP) neural network,

Table 2. Variables used to construct the model.

Variables	Total (n = 701)
Age (years)	
≤17, n (%)	42 (6.0)
18–44, n (%)	202 (28.8)
45–64, n (%)	289 (41.2)
65–74, n (%)	114 (16.2)
≥75, n (%)	54 (7.7)
Gender	
Male, n (%)	493 (70.3)
Female, n (%)	208 (29.7)
Previous medical history	
Hypertension, n (%)	127 (18.1)
Diabetes, n (%)	38 (5.4)
Coronary artery disease, n (%)	13 (1.9)
Chronic renal failure, n (%)	5 (0.7)
Cerebral infarction, n (%)	16 (2.3)
Respiratory disorders, n (%)	9 (1.3)
Mechanism of traumatic brain injury	
Fall on the same plane, n (%)	197 (28.1)
Fall from high place, n (%)	138 (19.7)
Road accident, n (%)	340 (48.5)
Object striking the head, n (%)	26 (3.7)
Presence or absence of loss of consciousness after injury	
Yes, n (%)	323 (46.1)
No, n (%)	378 (53.9)
Glasgow Coma Scale score	
13–15, n (%)	416 (59.3)
9–12, n (%)	107 (15.3)
3–8, n (%)	178 (25.4)
Admission cranial CT examination results	
Epidural hematoma, n (%)	234 (33.4)
Subdural hematoma, n (%)	418 (59.6)

(continued)

Table 2. Continued.

| Variables	Total (n = 701)
Subarachnoid hemorrhage, n (%)	403 (57.4)
Skull fracture, n (%)	472 (67.3)
Diffuse axonal injury, n (%)	13 (1.9)
Brain herniation, n (%)	18 (2.6)
Whether brain surgery was performed	
No, n (%)	171 (24.4)
Yes, n (%)	530 (75.6)
Injury at other sites besides the cranium	
Fractures in other areas, n (%)	224 (32.0)
Visceral contusions, n (%)	17 (2.4)
Traumatic wet lung, n (%)	100 (14.3)
Pneumothorax, n (%)	14 (2.0)
Duration of intensive care treatment (days)	
≤5, n (%)	97 (16.1)
6–15, n (%)	91 (13.0)
≥16, n (%)	28 (4.0)
Whether intensive care treatment was performed	
No, n (%)	485 (69.2)
Yes, n (%)	216 (30.8)
GOS score	
1, n (%)	0 (0)
2, n (%)	82 (11.7)
3, n (%)	90 (12.8)
4, n (%)	351 (50.1)
5, n (%)	178 (25.4)
Length of stay in hospital (days)	
≤10, n (%)	215 (30.1)
11–20, n (%)	301 (42.9)
21–30, n (%)	130 (18.5)
31–40, n (%)	50 (7.0)
≥41, n (%)	5 (0.7)
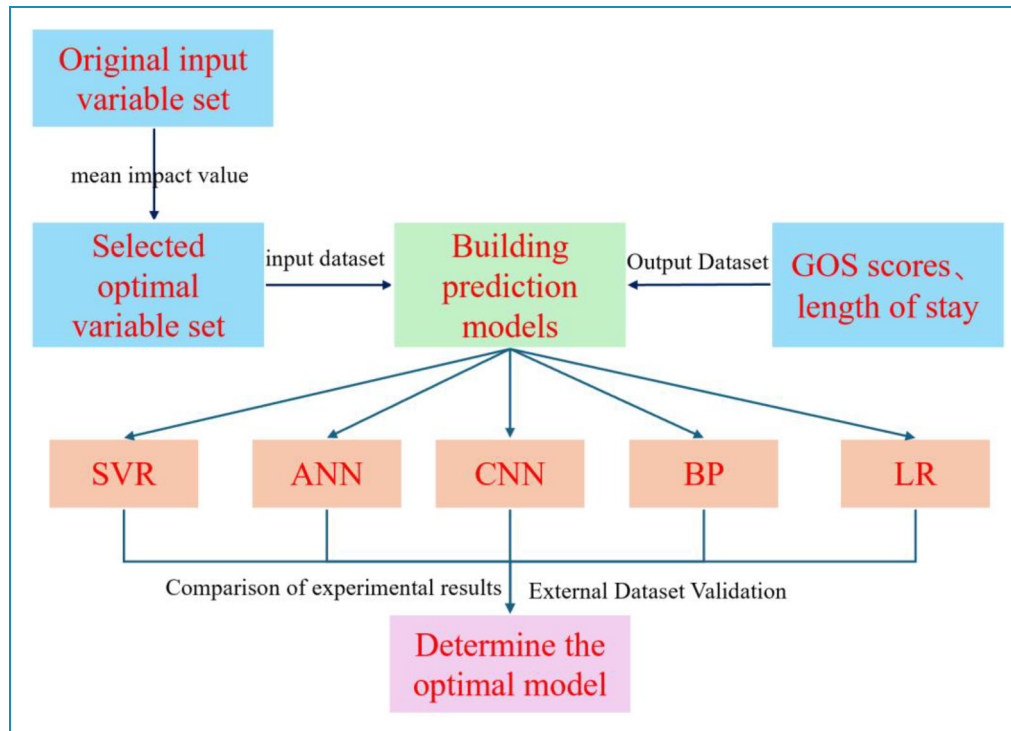
Figure 1. Research design.

artificial neural network (ANN) and logistic regression (LR). The choice of these models is grounded in their proven effectiveness for diverse data types. Here's a brief overview of each: CNNs are adept at processing data like images and speech. They extract data features using convolutional and pooling layers. Tools like TensorFlow and PyTorch offer modules for CNNs, with models like ResNet, VGG, and AlexNet being popular choices for image tasks. ANNs are versatile neural networks suitable for tasks ranging from classification to regression. They don't adhere to a fixed structure, allowing flexibility in application. Frameworks such as TensorFlow, PyTorch, and Keras support ANN development. BP networks are feedforward networks that include hidden layers to handle complex data relations. They leverage backpropagation algorithms for training. LR is a straightforward linear model used for classification and regression. It efficiently combines linear equations and logistic functions without demanding intricate computations or hyperparameter tuning. SVR is a novel small-sample learning method underpinned by a solid theoretical foundation. At its core, it doesn't rely on probabilistic measures or the law of large numbers, setting it apart from conventional statistical approaches. Essentially, SVR sidesteps the traditional process of moving from induction to deduction, enabling efficient 'transductive inference' from training samples to prediction samples. This approach considerably streamlines typical regression prediction challenges.

Our objective in using these models is to comprehensively evaluate their performance in our specific context. This comparative approach strengthens the validity of our findings. All computations for this research were conducted using Python 3.9.

Lastly, the top-performing model underwent further testing using external datasets sourced from various healthcare facilities. This external validation tests the model's broader applicability, its resilience to overfitting and its performance across different scenarios. By this comprehensive approach, we aim to enrich the existing literature, offering deeper insights into risk stratification for traumatic brain injuries and enhancing the accuracy of predictions.

Filtering of predictor variables

The veracity of any predictive model is fundamentally anchored to the salient attributes represented by its input variables. In the extant literature, an uncritical direct inclusion of original input variables into predictive models has been the norm, with scant consideration for the potential diminution of predictive accuracy as a result thereof. This study, therefore, aims to institute a robust filtration protocol for input variables with a view to enhancing both computational efficiency and empirical accuracy.

The mean impact value (MIV) method was elected for this purpose, predicated on its ubiquity in the field and its computationally unencumbered principles. Given the

absence of a predetermined framework for variable selection specific to this research context, the MIV methodology stands as a logical choice. In the context of a given sample S, this method modulates a single input variable by a factor of ± 10%, while maintaining the remainder of the variables at their original values. This manipulation yields two derivative samples, S1 and S2, which are then subjected to predictive modeling to obtain the corresponding regression output sets R1 and R2. The magnitude of the deviation between R1 and R2, averaged over numerous experimental iterations, serves as the MIV metric that quantifies the influence of individual input variables on the predictive model (Figure 2).

Conforming to the procedural framework delineated in Figure 2, the MIV algorithm is synergistically integrated with the machine learning model to compute MIV magnitudes for each of the scrutinized input variables, as demonstrated in Figure 3.

The absolute magnitudes of the mean impact value (MIV) serve as the determinative metrics for assessing the relative significance of various input variables in influencing the predictive accuracy of the model. In descending order of their importance, these variables are classified as follows: time to intensive care treatment ($x1$), admission Glasgow Coma Scale (GCS) score ($x2$), the mechanism underlying traumatic brain injury ($x3$), the administration

of intensive care treatment ($x4$), initial cranial computed tomography (CT) evaluation ($x5$), the execution of brain surgery ($x6$), the incidence of post-traumatic loss of consciousness ($x7$), injuries to anatomical locales extraneous to the cranial cavity ($x8$), patient age ($x9$), antecedent medical history ($x10$) and gender ($x11$).

To discern the optimal configuration of input variables for this model, a process of methodical elimination was implemented, based on the ascending values of MIV. Concurrently, the fluctuation in predictive error rates across various combinations of input variables was rigorously studied. The most favorable configuration was identified as the one corresponding to the minimum predictive error. As an evaluative criterion, the root-mean-square error (RMSE) was enlisted, as articulated in Equation (1), where y_a signifies the actual value and y_p designates the predicted value:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_a - y_p)^2} \qquad (1)$$

The RMSE under different input variable regimes is shown in Table 3.

According to the data elucidated in Table 2, a consistent decline in RMSE was observed across the four variable systems, denoted as $F1$ to $F4$. This decline was correlated with the sequential exclusion of variables such as gender, antecedent medical history, age and non-cranial injuries, thereby revealing an ongoing enhancement in model precision. However, additional excisions performed within variable systems $F5$ to $F11$—which included non-cranial injuries, post-traumatic consciousness status, surgical interventions, cranial CT findings, intensive care treatment, the mechanism of traumatic brain injury and GCS score—induced an increment in the RMSE values, thereby indicating a deterioration in predictive accuracy. The empirical evidence unequivocally corroborates that the model achieves its minimal RMSE when configured with the input variable system $F4$. Consequently, this study ascertains the optimal set of input variables to include injuries extraneous to the cranial region, post-traumatic consciousness status, whether surgical intervention was conducted, initial cranial CT evaluations, the administration of intensive care, the mechanism of traumatic brain injury, GCS scores upon admission and the duration of intensive care treatment.

Establishment and evaluation of the prediction model

There is no missing data in the dataset of this study. After applying the exclusion criteria, a sample set containing 1001 patient records was retained for model training. A cohort of 1001 patients was partitioned in a stochastic manner into two distinct subsets: 70% were allocated to
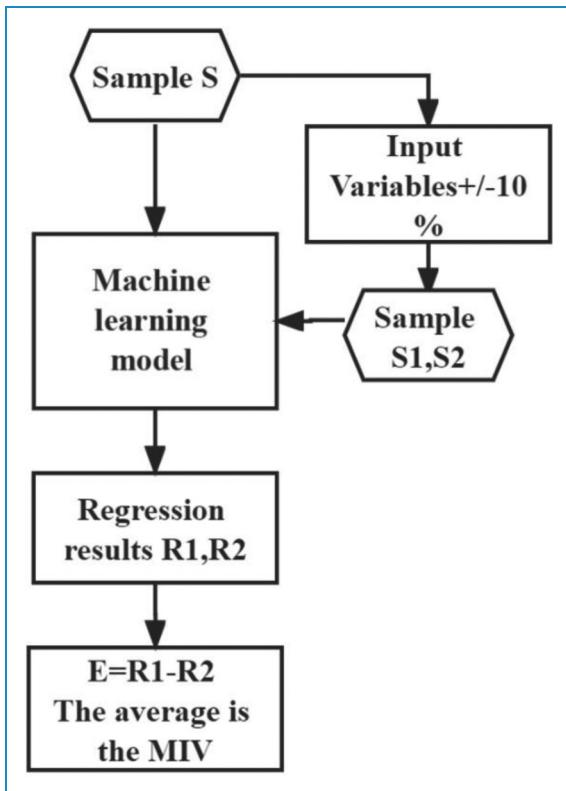

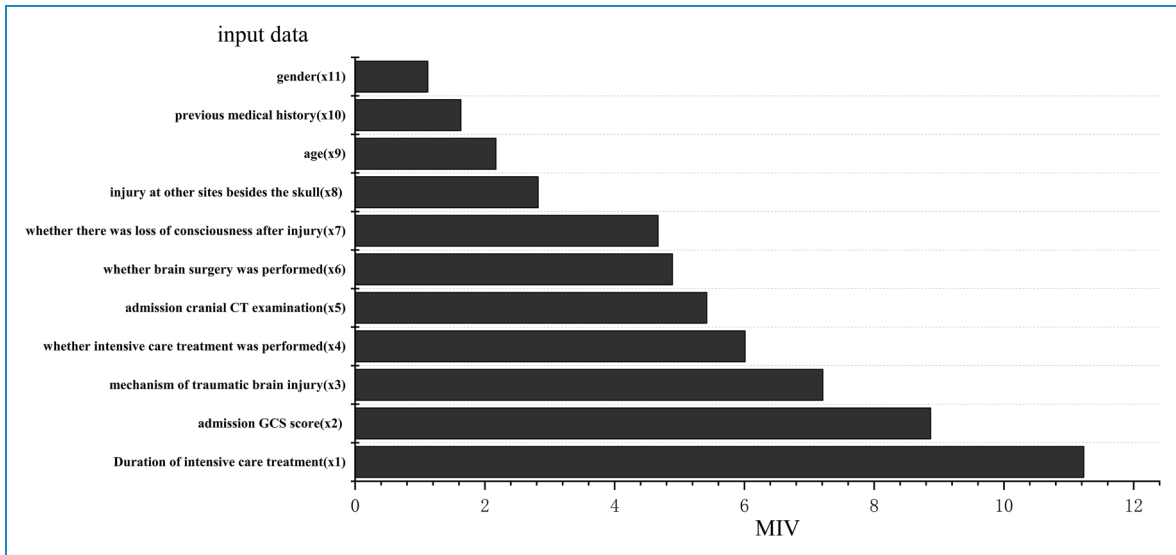**Figure 2.** Flow chart of MIV method.

Figure 3. Size of MIV values for different input variables.

Table 3. RMSE values under different input variable systems.

System	Input variable system	RMSE
F1	x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11	0.484
F2	x1, x2, x3, x4, x5, x6, x7, x8, x9, x10	0.413
F3	x1, x2, x3, x4, x5, x6, x7, x8, x9	0.375
F4	x1, x2, x3, x4, x5, x6, x7, x8	0.331
F5	x1, x2, x3, x4, x5, x6, x7	0.489
F6	x1, x2, x3, x4, x5, x6	0.517
F7	x1, x2, x3, x4, x5	0.574
F8	x1, x2, x3, x4	0.588
F9	x1, x2, x3	0.612
F10	x1, x2	0.627
F11	x1	0.656

Table 4. The architecture of convolutional neural networks.

Network layer	Model parameter setting
Input layer	700×11 spectral data matrix
Convolution layer 1	64 1×1 convolution kernels; kernel_size = 5
Convolution layer 2	128 1×1 convolution kernels; kernel_size = 5
Pool layer 1	MaxPool; kernel_size = 1; stride = 2
Convolution layer 3	128 1×1 convolution kernels; kernel_size = 5
Pool layer 2	MaxPool; kernel_size = 1; stride = 2
Convolution layer 4	256 1×1 convolution kernels; kernel_size = 5
Pool layer 3	MaxPool; kernel_size = 1; stride = 2
Convolution layer 5	512 1×1 convolution kernels; kernel_size = 5
Pool layer 4 (adaptive pooling layer)	Output one-dimensional vector
Full connection layer	Result

the training set, and the remaining 30% comprised the test set. Obtain the prediction effect of the model through cross-validation method. A quintet of machine learning paradigms—namely, SVR, CNN, BP neural network, ANN and LR—were instantiated. These models were calibrated utilizing the empirically determined optimal vectors of input and output variables, specifically the GOS scores

and length of hospital stay. Subsequently, the efficacy of each model in prognosticating these outcome variables was rigorously assessed to facilitate the selection of the most efficacious machine learning model for the ultimate predictive schema.

The evaluative metrics enlisted for this analytical endeavor included the Mean Absolute Percentage Error (MAPE), which was employed to quantify the prediction discrepancies concerning GOS scores and length of stay, as articulated in Equation (2). Additionally, the coefficient of determination (R^2) served as an ancillary measure to gauge the congruence between the true and predicted values emanating from the model:

$$MAPE = \frac{1}{n}\sum_{i=1}^{N}\left|\frac{y_a - y_p}{y_p}\right| \times 100\% \qquad (2)$$

With respect to hyperparameter optimization for each model, the settings were delineated as follows. For SVR, a Radial Basis Function (RBF) kernel was utilized, with a penalty factor (p) of 100 and a kernel parameter (γ) set to 0.01. In the case of CNN, the corresponding network architectures and hyperparameter configurations are elaborated in Tables 4 and 5. For ANN, a triad of hidden layers was incorporated, with the learning epochs stipulated at 1,000, a learning rate of 0.01 and a batch size parameterized at 64. In the BP model, the architecture encompassed quintuple hidden layers with 2000 learning epochs and a learning rate of 0.005. Lastly, the logistic regression model incorporated an $L2$ regularization term, with the convergence threshold definitively fixed at 0.001.

Validation

In order to better evaluate the generalization ability of the model, the risk of overfitting has been reduced, ensuring the practicality of the prediction model and improving the credibility of the model. This study used the clinical data of 111 patients from the First Affiliated Hospital of Anhui Medical University throughout 2022 as the external

Table 5. Setting hyperparameters for convolutional neural networks.

Hyperparameter	Setting
Activation function	ReLu
Optimizer	Adam
Learning rate	0.001
Batch size	64
Dropout	0.5

Table 6. Variables used to validate the model.

Variables	Total (n = 111)
Age (years)	
≤17, n (%)	4 (3.6)
18–44, n (%)	34 (30.7)
45–64, n (%)	43 (38.7)
65–74, n (%)	20 (18.0)
≧75, n (%)	10 (9.0)
Gender	
Male, n (%)	72 (64.9)
Female, n (%)	39 (35.1)
Previous medical history	
Hypertension, n (%)	15 (13.5)
Diabetes, n (%)	7 (6.3)
Coronary artery disease, n (%)	2 (1.8)
Chronic renal failure, n (%)	3 (2.7)
Cerebral infarction, n (%)	2 (1.8)
Respiratory disorders, n (%)	2 (1.8)
Mechanism of traumatic brain injury	
Fall on the same plane, n (%)	33 (29.7)
Fall from high place, n (%)	9 (8.1)
Road accident, n (%)	61 (55.0)
Object striking the head, n (%)	8 (7.2)
Presence or absence of loss of consciousness after injury	
Yes, n (%)	47 (42.3)
No, n (%)	64 (57.7)
Glasgow Coma Scale score	
13–15, n (%)	67 (59.3)
9–12, n (%)	17 (15.3)

(continued)

Table 6. Continued.

Variables	Total (n = 111)
3–8, n (%)	27 (25.4)
Admission cranial CT examination results	
Epidural hematoma, n (%)	41 (36.9)
Subdural hematoma, n (%)	53 (47.7)
Subarachnoid hemorrhage, n (%)	57 (51.4)
Skull fracture, n (%)	77 (69.4)
Diffuse axonal injury, n (%)	3 (2.7)
Brain herniation, n (%)	2 (1.8)
Whether brain surgery was performed	
No, n (%)	80 (72.1)
Yes, n (%)	31 (27.9)
Injury at other sites besides the cranium	
Fractures in other areas, n (%)	15 (13.5)
Visceral contusions, n (%)	5 (4.5)
Traumatic wet lung, n (%)	3 (2.7)
Pneumothorax, n (%)	2 (1.8)
Duration of intensive care treatment (days)	
≤5, n (%)	17 (15.3)
6–15, n (%)	12 (10.8)
≥16, n (%)	2 (1.8)
Whether intensive care treatment was performed	
No, n (%)	80 (72.1)
Yes, n (%)	31 (27.9)
GOS score	
1, n (%)	0 (0)
2, n (%)	8 (7.2)
3, n (%)	21 (18.9)

(continued)

Table 6. Continued.

Variables	Total (n = 111)
4, n (%)	60 (54.1)
5, n (%)	22 (19.8)
Length of stay in hospital (days)	
≤10, n (%)	35 (31.5)
11–20, n (%)	52 (46.8)
21–30, n (%)	20 (18.0)
31–40, n (%)	3 (2.8)
≥41, n (%)	1 (0.9)

Table 7. MAPE under four prediction models.

Model	MAPE (%)	
	Training error	Test error
SVR	5.89	6.30
BP	21.17	23.51
CNN	10.78	12.59
ANN	17.41	22.05
LR	24.23	29.93

dataset for this study. The inclusion criteria, outcome indicators and predictive indicators are identical to those in the modeling dataset. The dataset used for model validation has no missing data. Table 6 provides a detailed explanation of the classification and definition attributes of variables used in model validation. Comparing the data in Tables 1 and 6, there are differences in the distribution of important variables between the model validation dataset and the model development dataset.

Results

Prediction of GOS scores

In an endeavor to ascertain the most efficacious predictive model for craniocerebral trauma outcomes, a comparative evaluation was executed across five distinct machine

learning frameworks: SVR, CNN, BP neural network, ANN and LR. Empirical observations drawn from the MAPE of these models, as delineated in Table 7, reveal that the SVR model manifests the most diminutive predictive error. Specifically, its MAPE is 17.21%, 6.29%, 23.63% and 15.75% inferior relative to the BP, CNN, LR and ANN paradigms, respectively.

Further quantitative substantiation is provided by the coefficients of determination (R^2) between the predicted and the actual GOS scores across both the training and test datasets. As elucidated in Table 8, the SVR model attains a preeminent R^2 value, measuring 93.2% in the training set and 91.7% in the test set. This lends compelling evidence to assert that the SVR model exhibits superior predictive fidelity compared to its counterparts.

Predicting length of stay

A parallel analytical exercise was conducted to discern the optimal predictive model for the length of hospital stay. Once again, a competitive analysis was undertaken across the aforementioned quintet of machine learning algorithms. The MAPE values, enumerated in Table 9, evince that the SVR model outperforms its peers by registering the least predictive error. Specifically, the MAPE values for the SVR model are 7.29%, 11.62%, 18.25% and 16.34% lower than those exhibited by the BP, CNN, logistic regression and ANN models, respectively.

Augmenting these findings, the R^2 values, encapsulated in Table 10, reinforce the preeminence of the SVR model in terms of predictive accuracy. The SVR model boasts R^2 values of 91.8% and 90.6% for the training and test sets, respectively, thereby outclassing the alternative models in capturing the nuances of hospital length of stay.

Validation set results

To further bolster the empirical validation of the SVR model's efficacy in predicting the length of hospital stay among patients with craniocerebral trauma, we utilized an external validation set of 111 patients sourced from the First Affiliated Hospital of Anhui Medical University.

The octet of input variables, meticulously sieved through the mean impact value (MIV) methodology, was incorporated into the five comparative models to generate predictive GOS scores for these cases. The resultant outputs are delineated in Figure 4. The predictive acumen of the SVR model exhibited closer congruence to the actual GOS scores when juxtaposed against logistic regression, which demonstrated the most pronounced disparity from the veridical values. Upon computational scrutiny, the error rates for these five models (SVR, CNN, ANN, BP and LR) in the external dataset were ascertained to be 7.61%, 15.08%, 18.86%, 24.27% and 28.56%, respectively.

A refined set of eight input variables, meticulously selected via the mean impact value (MIV) methodology, were incorporated into each of the five comparative models to generate prognostications for the duration of hospital stay for this patient cohort. The outcomes of this investigation are graphically represented in Figure 5. The SVR model demonstrated a higher degree of accuracy, closely approximating the true lengths of stay, while the logistic regression model manifested the most considerable deviation from the actual values. Subsequent to computational analysis, the error rates associated with these five models (SVR, CNN, ANN, BP and LR) in the external validation dataset were quantified as 7.91%, 15.07%, 22.55%, 10.94% and 26.84%, respectively. These metrics compellingly indicate that the SVR model possesses the lowest error rate among the models examined.

In the concluding phase of our empirical assessment, we sought to rigorously ascertain the extent to which the MIV methodology enhances the predictive capabilities of the established models. Accordingly, we introduced all 11 variable types from the external validation dataset, sans any MIV-mediated filtering, directly into the quintet of models under study (SVR, CNN, ANN, BP and LR). The consequent outcomes are tabulated in Table 11. The data

Table 8. R^2 under four prediction models.

Model	R^2 (modeling set)	R^2 (testing set)
SVR	93.2%	91.7%
BP	84.5%	82.7%
CNN	89.6%	85.4%
ANN	79.3%	76.5%
LR	72.1%	70.4%

Table 9. MAPE under four prediction models.

Model	MAPE (%)	
	Training error	Test error
SVR	7.83	9.28
BP	13.46	16.57
CNN	16.38	20.90
ANN	21.71	25.62
LR	21.36	27.53

Table 10. R^2 under four prediction models.

Model	R^2 (modeling set)	R^2 (testing set)
SVR	91.8%	90.6%
BP	87.2%	84.3%
CNN	83.4%	81.2%
ANN	75.2%	73.7%
LR	70.3%	68.5%

reveals that the error rates for each of these five models exhibited varying degrees of amelioration in their predictive performance for both GOS scores and lengths of stay when devoid of MIV screening.

Discussion

Principal findings

Subsequent to the meticulous employment of the mean impact value (MIV) methodology for discerning the most salient input variables from an extensive corpus of clinical data, five machine learning models (SVR, CNN, ANN, BP, LR) were instantiated for evaluation. Comparative analysis of these algorithms revealed variances in their predictive capacities. The SVR model showed significant superiority, with a MAPE of 5.89% in the training dataset, 6.30% in the testing set and 7.61% in external validation, especially in predicting the GOS score. In the predictive hospitalization time model, the SVR model also demonstrated commendable fidelity, recording 7.83% of MAPE in the training set, 9.28% in the testing set and 7.91% in the external validation set. These empirical metrics unequivocally establish the SVR model as the most efficacious predictive algorithm for this study, attesting both to its robust reliability and its clinical applicability.

Moreover, to empirically ascertain the contributive impact of the MIV method on the predictive accuracy of the machine learning algorithms, we engaged in a secondary experiment wherein unscreened variables from external datasets were directly integrated into the five models. The outcomes compellingly revealed that the algorithms which incorporated MIV-screened variables exhibited varying but consistently positive enhancements in predictive accuracy for both GOS scores and lengths of hospital stay. These observations serve to underscore the augmentative role of MIV in optimizing input variables for predictive modeling.

To the best of our scholarly awareness, this investigation represents a pioneering effort in the amalgamation of MIV and machine learning techniques to enhance predictive accuracy, specifically within the domain of TBI.

Comparison with prior work

In antecedent investigations, the modus operandi for the identification of pertinent clinical data for algorithmic modeling predominantly emanated from a synthesis of extant scholarly publications and empirical clinical wisdom.[6] In a marked departure from this conventional methodology, the current study leveraged the mean impact value (MIV) approach for a priori filtration of an initial pool of 11 data categories. This selection protocol was designed to augment the model's predictive fidelity. The implementation of MIV offered a transparent and granular elucidation of the relative salience and inadequacy of diverse input variables. Employing RMSE as the evaluative metric, we opted for the assemblage of input data with the most diminutive RMSE value as the quintessential input variable set for our machine learning model. Upon rigorous analysis, variables such as the presence of injuries outside of cranial regions, the occurrence or absence of post-injury loss of consciousness, the undertaking of surgical brain intervention, age, administration of intensive care treatment, the underlying mechanism of the TBI, initial GCS scores upon admission and the duration of intensive care were adjudicated as the most propitious input variables for the model. Crucially, specific factors like the length of intensive care, the admission GCS scores, the mechanistic etiology of the TBI and cranial CT scans upon admission wielded a disproportionate influence on the model's predictive accuracy. This alignment with extant clinical contexts and neurosurgical experiences further underscores the pragmatic utility and clinical viability of the adopted methodological framework.

Presently, a diverse array of machine learning (ML) algorithms have been deployed in the realm of medical prognosis, among which ANN have gained prominence as the de facto standard for disease prediction.[19–25] The ascendancy of ANNs is largely attributed to their superior capability to model intricate nonlinear relationships between input and output variables in high-dimensional data spaces, coupled with their facility for optimal model selection based on accuracy metrics.[26] However, despite these advantages, ANNs are not without limitations. For instance, they necessitate the configuration of an extensive array of hyperparameters prior to model training, the selection of which is predominantly reliant on the subjective expertise of the researcher.[27,28] Furthermore, ANNs are often criticized for their protracted learning cycles and their susceptibility to local minima, which could potentially undermine the model's predictive accuracy.[29] In juxtaposition, SVR models proffer distinct advantages, most notably their capacity to converge towards a global optimum, thereby obviating the challenges associated with local minima inherent in ANN models.[30] SVR achieves this by projecting the data into a high-dimensional feature space through nonlinear transformations and
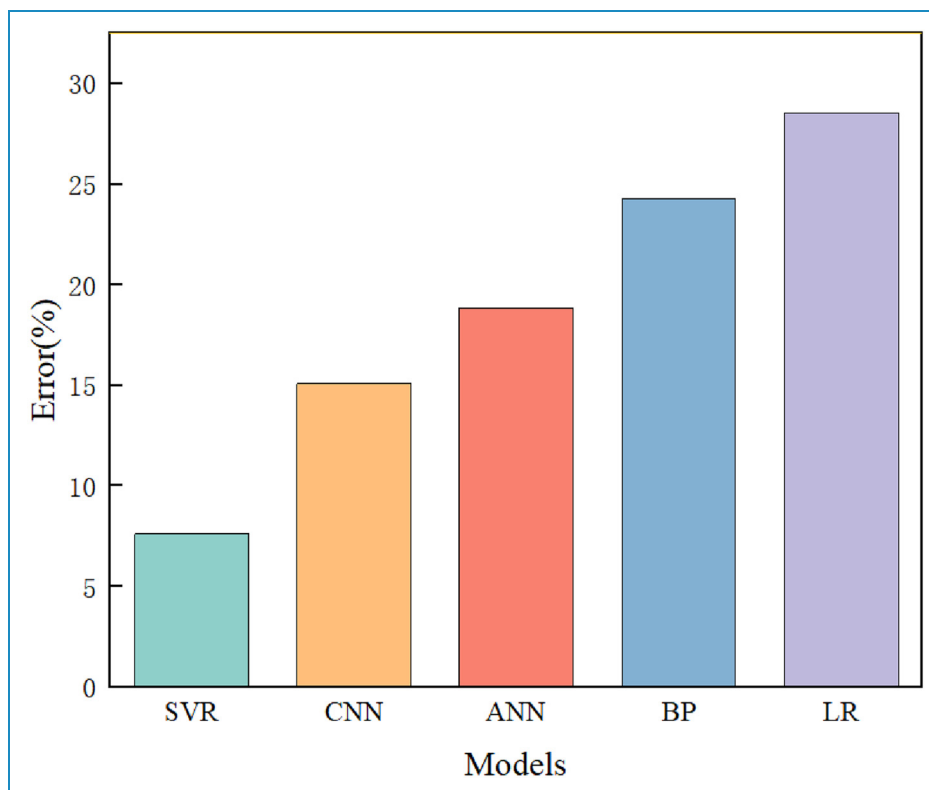
Figure 4. Error of five models in predicting GOS scores in external datasets.
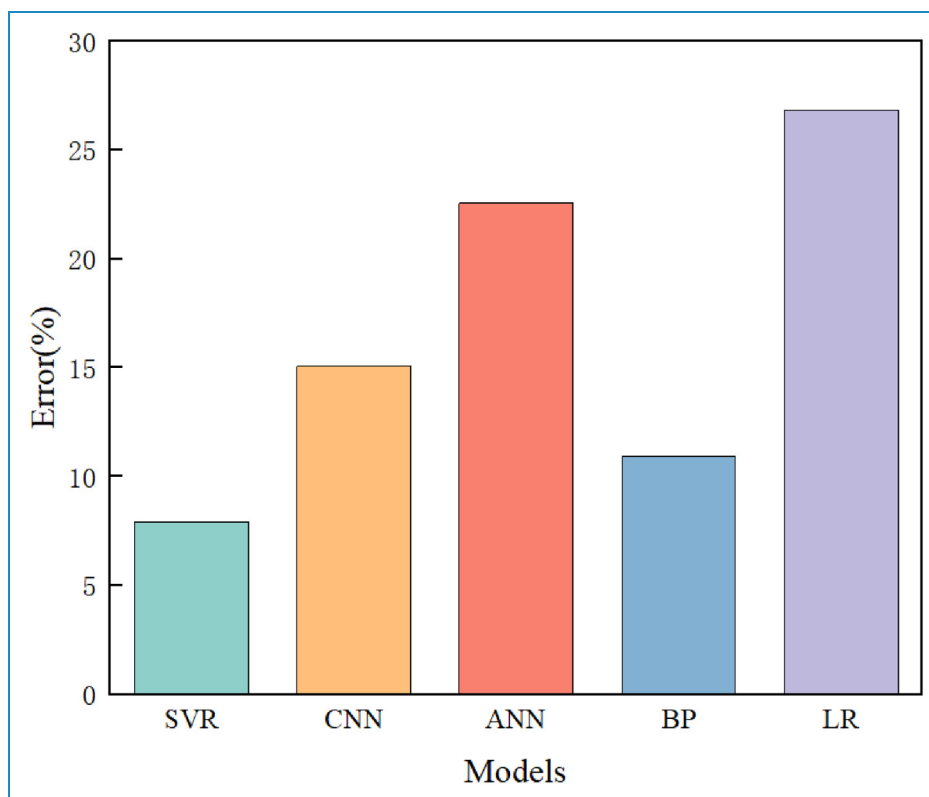

Figure 5. Error of five models in predicting length of stay in external datasets.

Table 11. Error of the five models not screened by MIV in predicting GOS scores and length of stay.

Model	MAPE for predicting GOS scores (%)	MAPE for predicting length of stay (%)
SVR	11.27	12.74
CNN	18.64	20.17
ANN	25.18	24.26
BP	28.79	15.41
LR	33.05	34.63

subsequently constructing a linear decision boundary in this augmented space, thereby realizing a nonlinear decision function in the original data space.[31] One of the most salient attributes of this algorithmic approach is its inherent complexity being independent of the dimensionality of the sample space. This ensures both high predictive accuracy and robust generalization capabilities, virtues that are often conspicuously absent in ANN models.[32] The importance of a machine learning model's generalizability cannot be overstated, particularly in its translational applicability to clinical settings. Empirical evidence demonstrates that the predictive outcomes of the SVR model, as validated on an external dataset from the First Affiliated Hospital of Anhui Medical University, were superior to those of other comparative models. Notably, these results were congruent with the initial dataset sourced from the Second Affiliated Hospital of Anhui Medical University, further corroborating the robust generalizability and predictive efficacy of the SVR model across disparate clinical environments.

Future deployments of this model have the potential to contribute significantly to targeted interventions for patients with TBI who exhibit poor prognostic indicators. Utilizing hypothetical datasets, healthcare providers could anticipate future trajectories of TBI patients in terms of GOS scores and duration of hospital stay. Additionally, the SVR model offers clinicians a data-driven decision support tool for making evidence-based prognostic assessments. For example, if a patient's high GOS score and short projected length of stay align with established discharge criteria, clinicians could elect to discharge the patient, thus mitigating the economic burden and the deleterious impacts associated with prolonged hospitalization.

Limitations

While our investigation pioneers a novel paradigm in the utilization of ML methodologies within the context of TBI, it is not devoid of certain limitations that warrant explicit mention. Firstly, the adoption of ML techniques remains a relatively unfamiliar territory for the majority of practicing clinicians, potentially rendering these approaches susceptible to dismissal as esoteric or empirically unvalidated modalities, particularly in high-stakes domains such as diagnostic and surgical procedures. However, it is critical to note that the present study circumscribes the application of ML solely to the prediction of patients' GOS scores and length of hospital stay. This judicious scope mitigates attendant risks while preserving predictive accuracy, serving principally to furnish clinicians with actionable insights and to enable healthcare institutions to allocate resources with enhanced precision.

Secondly, in the variable selection phase employing the MIV methodology, we observed that age exerted a significant influence on predictive outcomes. Pediatric TBI constitutes a focal point of our ongoing inquiry. Given the presence of specialized pediatric facilities within our geographical jurisdiction, the dataset concerning severe pediatric TBI may lack adequate representation. Future efforts may necessitate collaborative endeavors with pediatric medical centers to amass a more comprehensive dataset, thereby refining the model's applicability to this demographic.

Lastly, our model construction was predominantly based on epidemiological factors, encompassing demographic attributes, clinical manifestations and radiological indices, but did not integrate pertinent laboratory metrics, such as electroencephalogram (EEG) readings. This omission may introduce a degree of bias, particularly when viewed through the lens of retrospective study designs. Additional variables from external studies, such as the psychiatric status of patients, were also not accounted for. These enumerated caveats will form the nucleus of our forthcoming research initiatives.

Conclusions

In summary, the present study represents an inaugural effort to amalgamate MIV methodologies with ML algorithms to enhance predictive accuracy within the domain of TBI research. Our findings corroborate the hypothesis that the MIV-SVR model demonstrates superior accuracy in comparison to traditional statistical approaches. Consequently, the MIV-SVR algorithm holds substantial promise in ascertaining patient prognosis and length of stay, with potential for impactful integration into clinical workflows. Our conclusion is based on adherence to the TRIPOD standard, as detailed in the Appendix Table.

Abbreviations

BPNN	Backpropagation neural networks
CNN	Convolutional neural network
SVR	Support vector regression
CT	Computed tomography

MAPE   Mean absolute percentage error
ML   Machine learning
MSE   Mean square error
TBI   Traumatic brain injury
ANN   Artificial neural network
LR   Logistic regression
MIV   Mean impact value

ORCID iD: Yifeng Pan   https://orcid.org/0000-0002-5280-4938

References

1. Kim E, Lauterbach EC, Reeve A, et al. Neuropsychiatric complications of traumatic brain injury: a critical review of the literature (a report by the ANPA committee on research) [J]. *J Neuropsychiatry Clin Neurosci* 2007; 19: 106–127.
2. Hyder AA, Wunderlich CA, Puvanachandra P, et al. The impact of traumatic brain injuries: a global perspective [J]. *NeuroRehabilitation* 2007; 22: 341–353.
3. Pickelsimer EE, Selassie AW, Gu JK, et al. A population-based outcomes study of persons hospitalized with traumatic brain injury: operations of the South Carolina Traumatic Brain Injury Follow-up Registry [J]. *J Head Trauma Rehabil* 2006; 21: 491–504.
4. Gao G, Wu X, Feng J, et al. Clinical characteristics and outcomes in patients with traumatic brain injury in China: a prospective, multicentre, longitudinal, observational study [J]. *Lancet Neurol* 2020; 19: 670–677.
5. Dewan MC, Rattani A and Gupta S. Estimating the global incidence of traumatic brain injury [J]. *J Neurosurg* 2018; 130(4): 1080–1097.
6. Hale AT, Stonko DP, Lim J, et al. Using an artificial neural network to predict traumatic brain injury [J]. *J Neurosurg: Pediatr* 2018; 23: 219–226.
7. Feng J, Wang Y, Peng J, et al. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries [J]. *J Crit Care* 2019; 54: 110–116.
8. Zhou Z, Huang C, Fu P, et al. Prediction of in-hospital hypokalemia using machine learning and first hospitalization day records in patients with traumatic brain injury [J]. *CNS Neurosci Ther* 2023; 29(1): 181–191.
9. Mohamed M, Alamri A, Mohamed M, et al. Prognosticating outcome using magnetic resonance imaging in patients with moderate to severe traumatic brain injury: a machine learning approach [J]. *Brain Inj* 2022;36(3):353–358.
10. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review [J]. *World Neurosurg* 2018; 109: 476–486.e1.
11. Wang X, Zhong J, Lei T, et al. An artificial neural network prediction model for posttraumatic epilepsy: retrospective cohort study [J]. *J Med Internet Res* 2021; 23: e25090.
12. Li L, Zhong C, Wang G, et al. Artificial intelligence-enabled automated medical prediction and diagnosis in trauma patients [J]. *Recent Adv AI-enabled Autom Med Diagnosis* 2022; 1: 135–145.
13. 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]. *J Clin Epidemiol* 2019; 110: 12–22.
14. Dombi GW, Nandi P, Saxe JM, et al. Prediction of rib fracture injury outcome by an artificial neural network [J]. *J Trauma Acute Care Surg* 1995; 39: 915–921.
15. Fang C, Pan Y, Zhao L, et al. A machine learning-based approach to predict prognosis and length of length of stay in adults and children with traumatic brain injury: retrospective cohort study. *J Med Internet Res* 2022; 24: e41819.
16. Jennett B and Bond MR. Outcome after severe brain damage: a practical scale [J]. *Lancet* 1975; 1: 480–483.
17. Yong-Yan LU and Wei-Guo W. Variable selection of financial distress prediction—The SVM method based on mean impact value [J]. *Syst Eng* 2011; 29(8): 73–78.
18. Chen Y and Li M. Evaluation of influencing factors on tea production based on random forest regression and mean impact value [J]. *Agric Econ* 2019; 65(7): 340–347.
19. Ellethy H, Chandra SS and Nasrallah FA. The detection of mild traumatic brain injury in paediatrics using artificial neural networks [J]. *Comput Biol Med* 2021; 135: 104614.
20. Huang S, Xu Y, Yue L, et al. Evaluating the risk of hypertension using an artificial neural network method in rural residents over the age of 35 years in a Chinese area [J]. *Hypertens Res* 2010; 33: 722–726.
21. Çelik G, Baykan ÖK, Kara Y, et al. Predicting 10-day mortality in patients with strokes using neural networks and multivariate statistical methods [J]. *J Stroke Cerebrovasc Dis* 2014; 23: 1506–1512.
22. Numan N, Abuelenin S and Rashad M. Prediction of lung cancer using artificial neural network [J]. *Int J Intell Comput Inf Sci* 2016; 16: 1–19.
23. Lundin M, Lundin J, Burke HB, et al. Artificial neural networks applied to survival prediction in breast cancer [J]. *Oncology* 1999; 57: 281–286.

24. Petalidis LP, Oulas A, Backlund M, et al. Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data [J]. *Mol Cancer Ther* 2008; 7: 1013–1024.

25. Bhandari A, Koppen J and Agzarian M. Convolutional neural networks for brain tumour segmentation [J]. *Insights Imaging* 2020; 11: 1–9.

26. Renganathan V. Overview of artificial neural network models in the biomedical domain [J]. *Bratisl Lek Listy* 2019; 120: 536–540.

27. Tran-Ngoc H, Khatir S, De Roeck G, et al. An efficient artificial neural network for damage detection in bridges and beam-like structures by improving training parameters using cuckoo search algorithm [J]. *Eng Struct* 2019; 199: 109637.

28. Bergstra J, Yamins D and Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures[C]. In: International conference on machine learning, 2013, pp.115–123. PMLR.

29. Schuster M and Paliwal KK. Bidirectional recurrent neural networks [J]. *IEEE Trans Signal Process* 1997; 45: 2673–2681.

30. Hong WC. Hybrid evolutionary algorithms in a SVR-based electric load forecasting model [J]. *Int J Electr Power Energy Syst* 2009; 31: 409–417.

31. Baudat G and Anouar F. Generalized discriminant analysis using a kernel approach [J]. *Neural Comput* 2000; 12: 2385–2404.

32. Singh A, Thakur N and Sharma A. A review of supervised machine learning algorithms[C]. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp.1310–1315. IEEE.

Appendix

TRIPOD declaration list.

Part	Entry	Is this entry included
Title	1	Yes
Abstract	2	Yes
Background and objectives	3a	Yes
	3b	Yes
Data sources	4a	Yes
	4b	Yes
Research object	5a	Yes
	5b	Yes
	5c	No
Outcome indicators	6a	Yes

(continued)

Continued.

Part	Entry	Is this entry included
	6b	No
Predictor	7a	Yes
	7b	No
Sample size	8	Yes
Missing data	9	Yes
Statistical method	10a	Yes
	10b	Yes
	10c	Yes
	10d	Yes
	10e	No
Risk stratification	11	No
Dataset comparison	12	Yes
Research object	13a	Yes
	13b	Yes
	13c	Yes
Modeling	14a	No
	14b	No
Model specification	15a	Yes
	15b	No
Model performance	16	Yes
Model updating	17	No
Boundedness	18	Yes
Elaboration	19a	Yes
	19b	Yes
Indication	20	Yes
Supplementary information	21	No
Finances	22	No