Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Survival Prediction in Traumatic Brain Injury Patients Using Machine Learning Algorithms

Hosseinali Khalili
Shiraz University of Medical Sciences
Maziyar Rismani
Fasa University of Medical Sciences
Mohammad Ali Nematollahi
Fasa University
Mohammad Sadegh Masoudi
Shiraz University of Medical Sciences
Arefeh Asadollahi
Fasa university of medical sciences
Reza Taheri (reza.neuro@gmail.com)
Shiraz University of Medical Sciences
Hossein Pourmontaseri
Fasa University of Medical Sciences
Adib Valibeygi
Fasa University of Medical Sciences
Mohamad Roshanzamir
Fasa University
Roohallah Alizadehsani
Deakin University
Amin Niakan
Shiraz University of Medical Sciences
Aref Andishgar
Fasa University of Medical Sciences
Sheikh Mohammed Shariful Islam
Deakin University
U. Rajendra Acharya
Ngee Ann Polytechnic

Article

Keywords:

Posted Date: August 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1916615/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Predicting treatment outcomes in traumatic brain injury (TBI) patients is challenging worldwide. The present study aimed to achieve the most accurate machine learning algorithms to predict the outcomes of TBI treatment by evaluating demographic features, laboratory data, imaging indices, and clinical features. We used data from 3347 patients admitted to a tertiary trauma centre in Iran from 2016 to 2021. After the exclusion of incomplete data, 1653 patients remained. We used machine learning algorithms such as Random Forest (RF) and Decision Tree (DT) with ten-fold cross-validation to develop the best prediction model. Our findings reveal that among different variables included in this study, the motor component of the Glasgow Coma Scale, condition of pupils, and condition of cisterns were the most reliable features for predicting in-hospital mortality, while the patients' age takes the place of cisterns condition when considering the long-term survival of TBI patients. Also, we found that the RF algorithm is the best model to predict the short-term mortality of TBI patients. However, the generalized linear model (GLM) algorithm had the best performance (with an accuracy rate of 82.03 ± 2.34) in predicting the long-term survival of patients. Our results showed that using appropriate markers, and machine learning algorithms can provide a reliable prediction of TBI patients' survival in the short- and long-term with reliable and easily accessible features of patients.

Introduction

Traumatic brain injury (TBI) is among the most common causes of in-hospital death and neurological disabilities [1]. Recent observations showed that the mortality and morbidity of TBI are growing, and the incidence of age has decreased worldwide [2,3]. Over the last two decades, several studies have been dedicated to investigating the risk factors related to TBI morbidity and mortality. For instance, it has been found that age, gender, and the severity of TBI play essential roles in 10-year mortality [4]. Further investigations also introduced multiple risk factors for TBI mortality, such as intracranial pressure (ICP), using alcohol, the intensity of care, oxidative stress imbalance, and grouping complications. Although different risk factors have been distinguished in recent years, we still have a long way to go to achieve accurate assessment scales to manage patients with TBI [5–7].

Glasgow Coma Scale (GCS) is a commonly used tool to assess the neurological condition of patients with different brain injuries, especially TBI. Although GCS provides a reliable measurement for clinicians to manage the TBI, more efficient predictors are required to predict the outcomes of these cases [8]. Glasgow outcome scale (GOS) was another tool recruited to monitor the long-term recovery of patients, which has been extended from 5 classes to 8 classes (extended GOS or GOSE) to provide a more detailed follow-up [9]. It appropriately depicts the clinical outcomes in the hospital setting and even several months after patient discharge. Recent studies demonstrated that other factors, such as age, the motor component of GCS, pupillary reactivity, and type of injury, significantly influence the prediction of clinical outcomes [10,11]. In recent years, novel analytical tools and computational methods have been developed and applied to the field to precisely predict the patients' survival rate and outcome regarding their conditions. Developing complex nonlinear algorithms in the machine learning era to analyze clinical and para-clinical databases was one of the first promising attempts to predict patients' outcomes [12]. The low sensitivity and specificity of conventional analysis changed the trend of investigations [13]. Different artificial intelligence (AI) based techniques recently provided more accurate results by applying state-of-the-art algorithms to TBI datasets [14]. Machine learning (ML) scientists developed highly accurate novel approaches to control and monitor the patient conditions of multiple diseases. The extensive capability of ML algorithms allows it to be efficiently applied to a huge dataset with highly complex feature spaces and accurately model the future behaviour of data points [15]. ML significantly improved the prediction mechanism compared to conventional methods [16]. Prediction of clinical outcomes of patients with TBI contains several complexities and challenges. Annually, multiple new risk factors confirm and show significant effects on the GOS score of TBI [7,17]. To date, some studies have evaluated supervised learning algorithms, such as artificial neural networks (ANN), to predict the clinical outcomes of patients with TBI. However, the wide ability of a deep neural network (DNN) to weigh the parameters without supervision allows us to achieve more accurate models.

Literature Review

In 2009, Guler et al. [18] investigated the application of ANN to develop a diagnostic system and determine the severity of TBI. This small study analyzed simple clinical features among 32 cases, including vital signs, GCS, and electroencephalography (EEG), using a 3-layered ANN to find the similarities. This study showed that neurological and systematic features of TBI cases are similar by more than 90%.

Rughani et al. [19] used 11 clinical inputs to predict hospital survival in individuals with head injury by an ANN and compared it with clinician diagnosis and regression models. The data analysis of 7769 patients showed that ANN models are more accurate, sensitive, and discriminating than clinicians and regression models. The specificity, however, was the same across all models. Although this study showed that ANN would represent a more efficient model for predicting the outcomes of patients with head injuries, there is still a significant gap between the present models and the actual clinical scenarios.

In a study by Shi et al. [5], ANN was used to develop more accurate predictor models for in-hospital mortality after TBI surgery. The clinical inputs of 16,956 patients were analyzed to compare the performance of ANN and logistic regression (LR) models. Like previous observations, this study showed that ANN model is significantly more accurate, sensitive, and specific. Moreover, the ANN model demonstrated a higher area under the curve (AUC), positive predictive value (PPV), and negative predictive value (NPV). The findings showed that hospital volume, Charlson comorbidity index, length of stay, sex, and age would represent the best prediction of in-hospital mortality after TBI surgery.

Chong et al. [20] compared the efficiency of ML and LR in predicting TBI. This retrospective case-control study included 39 TBI cases and 156 age-matched controls hospitalized from 2006 to 2014. Then, the performance of ML and LR in the prediction of TBI were compared using receiver operating characteristics (ROC). The findings indicated that analysis of four novel features (involvement in road traffic accidents, loss of consciousness, vomiting, and signs of a base of skull fracture) by ML improved diagnostic parameters (sensitivity (94.9% vs 82.1%), specificity (97.4% vs 92.3%), PPV (90.2% vs 72.7%), NPV (98.7% vs 95.4%), and area under the curve (0.98 vs. 0.93)) in comparison with LR.

In 2015, Lu et al. [21] investigated the application of ANN in predicting long-term outcomes in TBI cases. This study included different clinical variables, such as GCS (at admission, 7th day, and 14th day), gender, blood sugar, white blood cells, history of diabetes and hypertension, pupil size, diagnosis to predict the 6-month GOS using ANN, Naïve Bayes (NB), DT, and LR. The findings of 128 adult participants showed that ANN has the best performance among different models (AUC of 96.13%, sensitivity of 83.5%, and specificity of 89.73%).

Another study by Beliveau et al. [22] tried to optimize the prediction models of one-year functioning of patients with TBI. Using clinical data from 3142 cases, this prospective study increased the diagnostic parameters of AI through novel techniques, including a subset of train and tests. The results indicated that ANN and other models, like LR, generally have high accuracy with the same AUC. Therefore, applying these models would develop reliable tools to manage further healthcare services for vulnerable TBI cases.

The study by Pourahmad et al. [23] was another attempt to optimize the predictive models of prediction in TBI patients. The clinical features of 410 cases (including age, gender, CT scan findings, pulse rate, respiratory rate, pupil size, reactivity, and the cause of injury) were admitted to Shahid Rajaee Hospital with GCS $\leq$ 10 were analysed by a 4-layered ANN combined with DT. This hybrid model could improve accuracy (86.3 vs. 82.2), sensitivity (55.1 vs. 47.6), specificity (93.6 vs. 91.1), and AUC (0.705 vs. 0.695) of the prediction of 6-month GOS in patients with TBI.

In 2019, Hale et al. [24] applied computed tomography (CT) scans in broadly diagnosing TBI. In this study, six clinical features and 17 different variables of CT scan of 480 patients (< 18 years old) were included in an analysis by a two-layer feed-forward ANN with 11 sigmoid hidden and softmax output neurons. The results of this study showed that the application of CT scan in the diagnosis of clinically relevant TBI would significantly increase all diagnostic parameters and achieve a highly optimized predictive model in the future.

A recent study by Abujaber et al. [25] investigated the application of ML models to predict in-hospital mortality for patients with TBI. The clinical and demographic features of 1620 patients, alongside their CT scan findings, were included in this study to develop efficient models using ANN and support vector machines (SVM). The results showed that SVM is more sensitive (73 vs. 62), accurate (95.6 vs. 91.6), and specific (99 vs. 96) than ANN and has a higher AUC (96 vs. 93.5) and F-score (0.8 vs. 0.64) in predicting the in-hospital mortality.

Recently, Thara et al. [26] conducted a novel study comparing ML and nomogram performance in predicting intracranial injury in children with TBI. Initially, the clinical parameters of 964 young patients with mild TBI, such as age, sex, road traffic injury, loss of consciousness, amnesia, hemiparesis, scalp injury, bleeding per nose or ear, hypotension, bradycardia, seizure, GCS at ED, pupillary light reflex were included in different algorithms (SVM, LR, NB, k-nearest neighbors, DT, random forest classifier (RFC), gradient boosting classifier, ANN, and nomogram). The findings showed that RFC represents the best performance in predicting pediatric TBI using different clinical features, especially CT scans.

In 2021, Hodel et al. [27] explored databases such as EBSCOhost CINAHL Complete, PubMed, and IEEE Xplore, to find all publications that developed prediction models for spinal cord injury (SCI). The searches showed that twelve different predictive models were developed in seven unique studies to predict the following clinical outcomes in patients with SCI. This review clearly showed that providing a comprehensive overview of patients with neurological traumas using different ML models would improve our clinical decision-making in the future to make the least mistakes.

Mawdsley et al. [28] conducted a study to systematically review the efficiency of ML models in predicting different psychosocial aspects of TBI cases. This comprehensive study found nine studies that included eleven types of ML to predict various outcomes. The findings showed that although these models could successfully develop predictive models, there is a lack of evidence to choose ML algorithms as a reliable tool in clinical decision makings.

In 2017, a critical review by Alanazi et al. [29] evaluated the quality of ML models in predicting patients' outcomes with different disorders. This study showed that AI could provide several promising models to predict these outcomes using patients' multiple clinical, demographical, and imaging data. But, still, we face some limitations in applying these models in clinical situations. Some studies indicated that these novel models would demonstrate significant errors and low efficiency even using the same database. Therefore, further studies are required to increase the reliability of provided models in the future.

In 2022, Choi et al. [30] developed new models to predict the diagnosis and prognosis of TBI patients at the prehospital stage. This multi-center retrospective study included 1169 TBI cases that were admitted from 2014 to 2018 in different hospitals in Korea. Various features, such as intracranial hemorrhage, admission with/without emergency department, and other demographic characteristics, were applied in five ML models, including LR, extreme gradient boosting, SVM, RF, and elastic net (EN). The findings of this study confirmed that EN would significantly develop our overview of the prediction of TBI outcomes at the prehospital stage by increasing AUC, specificity, and sensitivity.

In this year, Daley et al. [31] tried to provide effective ML-based models to predict severe TBI in admitted patients. This study tried to purposefully applicate neurological and biological data, such as partial thromboplastin time (PTT), motor component of GCS, serum glucose, the fixed pupil(s), platelet count, and creatinine to evaluate the predictive performance of different ML algorithms in the prediction of TBI in 196 admitted children. The findings of this study showed that the optimized models achieve the highest available accuracy (82%) and AUC (0.90).

There are inconsistencies in choosing the best clinical or para-clinical features and the most accurate machine learning model to predict the TBI patients' outcomes. Hence, the present study is designed to address these problems by recruiting a large population and a wide range of variables using different machine learning and regression algorithms are studied.

Dataset description

We used data from 3347 patients in the present study collected from admitted patients at Shahid Rajaee Hospital (Tertiary Trauma Centre), Shiraz, from 2016 to 2021. After the exclusion of patients with incomplete data, 1653 patients remained. The mean ± SD age of the final studied population was 39.55 ± 19.41, which consisted of 1371 men (82.9%). The demographic features, health status, and condition of TBI-induced hemorrhage in studied patients are indicated in Table 1.

To use the dataset in this research regarding diagnostic and therapeutic purposes, institutional approval was granted. Approval was granted on the grounds of existing datasets. Patients were informed about the data collection process and their consent were obtained. All methods were compliant with relevant guidelines and regulations. To use data, ethical approval was obtained from Shahid Rajaee Hospital (Tertiary Trauma Centre), Shiraz, Iran.

Table 1
The details of demographic features, health status, and traumatic brain injury-induced hemorrhage condition.

Variable	Frequency	percent
Demographic features		
Male	1371	82.9
Smoking	132	8.0
Opium	111	6.7
Health status		
Hypertension	124	7.5
Diabetes mellitus	83	5.0
Cardiovascular disease	52	3.2
Condition of traumatic brain injury		
Subarachnoid Hemorrhage	571	34.5
Intraventricular Hemorrhage	173	10.5
Epidural Hematoma	469	28.4
Subdural Hematoma	509	30.8
Intracerebral Brain Hemorrhage	755	45.7
Outcomes		
Mortality during hospitalization	319	19.3
Mortality up to 6 months	396	24.0

The demographic features included age, gender, smoking (smoker, non-smoker), opium (addicted, non-addicted), health status, hypertension, diabetes mellitus, and cardiovascular disease by asking the patients while take history. Also, GCS and pupil condition (anisocoric/brisk/fixed/sluggish/unable to check/bilateral non-reactive) were measured during a physical exam. The laboratory data of patients, including international normalized ratio (INR), blood sugar (BS), and fibrinogen level, were recorded from reported measurements in electronic documents. The Marshall score, subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), epidural hematoma (EDH), subdural hematoma (SDH), intracerebral hemorrhage (ICH), base of skull fracture, depressed skull fracture, and cisterna were evaluated in CT-scan imaging. The GOS (1 = dead/ 2 = vegetative state/ 3 = severe disability/ 4 = moderate recovery/ 5 = good recovery) and GOSE (1 = dead/ 2 = Vegetative State/ 3. Lower Severe Disability/ 4. Upper Severe Disability/ 5 = Lower Moderate Disability/ 6 = Upper Moderate Disability/ 7 = Lower Good Recovery/ 8 = Upper Good Recovery) were measured at the discharge day (GOSE0) and after 6 months (fGOSE) by trained specialists. The validity and equality of the specialist measurements were confirmed in a session to evaluate 10 cases.

Methodology

We tested few state-of-the-art ML algorithms on the dataset according to the flowchart shown in Fig. 1. The target features of our dataset (i.e. the GOS-extended of recovered TBI patients on the GOSE0 and fGOSE) have eight values from 1 to 8 that shows the level of consciousness. There is no consciousness, and the patient dies when the target feature is 1. On the other hand, when the target feature is 8, the patient can take care of his/her personal affairs. Unfortunately, when the target feature has 8 values (8 classes are defined), the performance of classification algorithms was so weak. So, we converted it to 5-class-dataset according to the physician's suggestion. Accordingly, classes 3 and 4, 5 and 6, and 7 and 8 were combined. In this case, the performance of classification algorithms was improved significantly.

Results

The results of applying some of the most important classification algorithms to our patients just when they leave the hospital are shown in Table 2. The classification algorithms used in this work are NB [32], RF [33], KNN(k = 5) [34], KNN(k = 6), DT [35], Rule Induction (RI) [36], Deep Learning (DL) [37] and Gradient Boosting Trees (GBT) [38] implemented in RapidMiner v9.10 [39]. Rapidminer is a comprehensive data science platform with visual workflow design and full automation. It is one of the most popular data science tools. This platform was run on a system which Intel(R) Core(TM) i5-4570, a 3.20GHz processor with 4 GB of RAM.

According to the obtained result, GBT, DL, and RF have the best accuracy rate of 47.67 ± 2.65, 46.22% ± 1.60%, and 45.37% ± 1.53%, respectively, while KNN (K = 5) has the worst with an accuracy rate of 33.82% ± 2.07%. Accuracy is defined as shown in Eq. (1).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

1

where TP shows true positive. It shows the number of positive instances the model correctly predicts as the positive class. TN shows true negative. It shows the number of negative instances that the model correctly predicts as the negative class. FP shows false positives. It shows the number of negative instances that the model incorrectly predicts as the positive class. Finally, FN shows a false negative. It shows the number of positive instances the model incorrectly predicts as the negative class.

As we have more than two classes in this test, only recall for each class was calculated. Therefore, the recall of each class is calculated as shown in Eq. (2).

$$recall = \frac{TP}{TP + FN}$$

2

Recall of class *i* is the fraction of instances where we correctly classified out of all es where their true label is *i*. Both accuracy and recall of investigated algorithms are shown in Table 2.

Table 2
The performance of different classification algorithms on 5-class-dataset according to GOS0

Algorithm	Accuracy (%)	Accuracy Rank	Recall1 (%)	Recall2 (%)	Recall3 (%)	Recall4 (%)	Recall5 (%)
NB	44.76 ± 3.94	4	42.95	33.51	11.51	9.60	81.66
RF	45.37 ± 1.53	3	52.66	10.47	1.19	1.66	94.06
KNN(k = 5)	33.82 ± 2.07	8	34.48	9.42	8.33	17.55	60.61
KNN(k = 6)	36.36 ± 2.55	7	40.44	10.99	7.54	17.88	64.18
DT	43.86 ± 1.96	6	53.29	5.24	1.59	1.32	91.17
RI	44.71 ± 3.02	5	44.20	17.80	11.90	8.28	86.42
DL	46.22 ± 1.60	2	55.49	8.90	11.90	11.92	85.57
GBT	**47.67 ± 2.65**	**1**	59.25	27.23	8.73	10.60	83.70
Mean of Accuracy:	42.85						

Table 3 shows the top 10 features with a higher role in classification and their weights. The weights are calculated by information gain [40]. The GCS motor component on admission (GCSM0), pupil, and Cisterns are the most significant features in classification, respectively.

Table 3
Feature weights calculated by information gain applied on 5-classes-dataset of gos0

Row	feature	weight
1	GCSM0	0.1956
2	pupil	0.1265
3	Cisterns	0.0809
4	DC	0.0725
5	Marshall	0.0718
6	age	0.0651
7	INR	0.0425
8	IVH	0.0385
9	1st BS	0.0378
10	Shift	0.0277

GCSM0: motor component of GCS on admission, DC: decompressive craniotomy, INR: international normalized ratio, IVH: intraventricular hemorrhage, BS: blood sugar.

After six months of leaving the hospital, when the target feature is fGOSE, the patients' conditions were investigated again. As it was shown in Table 4, GBT, RF, and DL have the best accuracy rate of 64.97% ± 1.62%, 64.97% ± 2.72%, and 64.37% ± 1.56%, respectively, while KNN (K = 5) has the worst with an accuracy rate of 55.89% ± 3.72%. As we have more than two classes in this test, only recall for each class was calculated. Therefore, the recall of each class is shown in Table 4.

Table 4
The performance of different classification algorithms on 5-class-dataset according to fgose

algorithm	Accuracy (%)	Accuracy Rank	Recall1 (%)	Recall2 (%)	Recall3 (%)	Recall4 (%)	Recall5 (%)
NB	59.46 ± 3.22	6	50.51	11.11	0.99	2.02	83.95
RF	64.97 ± 2.72	2	46.21	0.00	0.00	0.00	96.64
KNN(k = 5)	55.89 ± 3.72	8	35.86	0.00	0.99	10.61	82.43
KNN(k = 6)	56.20 ± 1.74	7	35.35	0.00	0.00	9.09	83.62
DT	63.16 ± 1.58	4	55.05	0.00	0.00	1.52	89.26
RI	62.67 ± 2.33	5	47.47	0.00	0.00	5.05	90.89
DL	64.37 ± 1.56	3	55.30	0.00	1.98	3.54	90.67
GBT	**64.97 ± 1.62**	**1**	59.85	0.00	0.99	1.01	90.46
Mean of Accuracy:	61.46						

In addition, comparing the mean of accuracy in Table 2 and Table 4 shows that predicting the future condition of the patients according to the selected features is more reliable after 6 months.

Table 5 shows the top 10 features with a higher role in classification and their weights. The weights are calculated by information gain. GCSM0, pupil, and age are the most significant features in classification, respectively. Compared to Table 2, the importance of age has increased, and now its role is more important than Cisterns.

Table 5
Feature weights calculated by information gain applied on 5-classes-dataset of fGOS

Row	feature	weight
1	GCSM0	0.1144
2	pupil	0.1037
3	age	0.0702
4	Marshall	0.0598
5	Cisterns	0.0580
6	DC	0.0383
7	INR	0.0335
8	1st BS	0.0261
9	IVH	0.0248
10	Shift	0.0237

GCSM0: motor component of GCS on admission, DC: decompressive craniotomy, INR: international normalized ratio, BS: blood sugar, IVH: intraventricular hemorrhage.

We also checked the system's performance when the patients were classified into only two groups, dead and alive. In this case, in addition to the classification mentioned above, two more algorithms LR and GLM were also investigated, which can be applied to only two-class classification problems. In this case, the performance of classification algorithms was again improved compared with the 5-class-dataset. The results of applying classification algorithms to our patients just when they leave the hospital when the patients are classified into two classes, dead and alive, are shown in Table 6. Accordingly, the accuracy rates of all algorithms are more than 80% which shows significant improvement compared with classification algorithms applied on the 5-class-dataset. In addition, there is no significant difference between the accuracy rates of most of these algorithms. All algorithms have a performance rate between 80% and 85%. The precision, recall, and AUC are also shown in this table. Precision is defined as Eq. (3).

$$precision = \frac{TP}{TP + FP}$$

3

AUC indicates the area under the receiver operating characteristic (ROC) curve. ROC is an evaluation metric for binary classification problems. A probability curve plots the TP rate against the FP rate at various threshold values.

According to the results shown in Table 6, RF, GLM, and RI have the best accuracy rate, respectively. The confusion matrix of best performing RF classifier is shown in Table 7.

Table 6
The performance of different classification algorithms on 2-class-dataset according to gose0

algorithm	**Accuracy (%)**	**Accuracy Rank**	**Precision (%)**	**Recall(%)**	**AUC**
NB	81.67 ± 1.29	7	52.61% ± 3.42	51.71% ± 7.88	0.820 ± 0.033
RF	**84.45 ± 1.29**	**1**	**76.72% ± 12.75**	27.89% ± 7.10	0.827 ± 0.046
KNN(k = 5)	80.64 ± 2.40	10	50.11% ± 13.99	24.14% ± 7.53	0.659 ± 0.048
KNN(k = 6)	81.07 ± 2.43	9	51.48 ± 13.05	24.13 ± 9.87	0.679 ± 0.056
DT	82.46 ± 1.15	6	59.98 ± 6.95	30.14 ± 8.62	0.703 ± 0.038
RI	83.24 ± 2.94	4	61.84 ± 12.20	39.82 ± 8.62	0.797 ± 0.067
DL	81.13 ± 2.77	8	52.81 ± 8.79	**55.18 ± 12.29**	**0.845 ± 0.029**
GBT	82.82 ± 1.72	5	55.62 ± 3.88	51.72 ± 13.27	0.827 ± 0.046
LR	84.03 ± 1.76	2	64.80 ± 8.94	40.08 ± 8.01	0.842 ± 0.043
GLM	83.91 ± 2.08	3	63.39 ± 9.43	41.08 ± 6.06	0.841 ± 0.039
Mean of Accuracy:	82.52				

Table 7
Confusion matrix obtained using RF classifier.

		Predicted as	
		alive	dead
Actually	alive	1307	27
	dead	230	89

Table 8 shows the top 10 features with a higher role in classification and their weights. The weights are calculated by information gain. Like Table 3 and Table 5, the pupil has a significant role in classification. The order of other features does not have a substantial difference between Table 3 and Table 5.

Table 8
Feature weights calculated by information gain applied on 2-classes-dataset of GOS0.

Row	**feature**	**weight**
1	pupil	0.0614
2	Cisterns	0.0560
3	age	0.0539
4	GCSM0	0.0472
5	Marshall	0.0414
6	INR	0.0345
7	DC	0.0209
8	1st BS	0.0192
9	Shift	0.0186
10	SAH	0.0172

GCSM0: motor component of GCS on admission, INR: international normalized ratio, DC: decompressive craniotomy, BS: blood sugar, SAH: subarachnoid hemorrhage.

The results of applying the classification algorithms on the 2-class-dataset after six months of leaving the hospital are shown in Table 9. Table 11 shows the importance f the feature role in classification. Comparing the Mean of accuracy in Table 6 and Table 9 shows that the accuracy rate does not change significantly after 6 months of the patient’s discharge. Finally, the confusion matrix of best performing GLM algorithm is shown in Table 10.

Table 9
The performance of different classification algorithms on 2-class-dataset according to fGOS

Algorithm	**Accuracy (%)**	**Accuracy Rank**	**Precision (%)**	**Recall(%)**	**AUC**
NB	78.65 ± 3.93	7	55.95 ± 8.59	53.82 ± 8.36	0.812 ± 0.039
RF	80.88 ± 1.86	3	**77.32 ± 10.90**	29.04 ± 4.50	0.807 ± 0.035
KNN(k = 5)	76.95 ± 1.66	9	54.49 ± 7.76	28.54 ± 2.95	0.661 ± 0.060
KNN(k = 6)	76.95 ± 1.96	10	53.56 ± 7.68	29.27 ± 7.31	0.675 ± 0.052
DT	78.22 ± 1.23	8	63.26 ± 9.17	25.33 ± 8.72	0.683 ± 0.039
RI	80.70 ± 2.51	4	66.69 ± 9.83	41.18 ± 7.83	0.758 ± 0.054
DL	78.95 ± 2.74	6	55.82 ± 5.73	**59.37 ± 10.79**	0.821 ± 0.038
GBT	79.25 ± 3.02	5	57.51 ± 6.86	56.56 ± 7.68	0.823 ± 0.015
LR	81.61 ± 2.58	2	67.61 ± 8.10	45.99 ± 4.21	**0.834 ± 0.031**
GLM	**82.03 ± 2.34**	**1**	68.00 ± 6.36	47.22 ± 8.41	0.834 ± 0.038
Mean of Accuracy:	79.42				

Table 10
Confusion matrix obatined using GLM algorithm according to fGOS.

		Predicted as	
		alive	dead
Actually	alive	1169	88
	dead	209	187

Table 11
Feature weights calculated by information gain applied on 2-classes-dataset of fGOS

Row	feature	weight
1	age	0.0631
2	pupil	0.0611
3	GCSM0	0.0547
4	Cisterns	0.0491
5	Marshall	0.0402
6	INR	0.0302
7	DC	0.0232
8	SAH	0.0196
9	1st BS	0.0185
10	IVH	0.0173

GCSM0: motor component of GCS on admission, INR: international normalized ratio, DC: decompressive craniotomy, SAH: subarachnoid hemorrhage, BS: blood sugar, IVH: intraventricular hemorrhage.

Overall, according to the results shown in Table 2 and Table 4, GBT has the best performance. RFs and DL are in the next ranks. Meanwhile, the ranks of accuracy in Table 6 and Table 9 show that GLM, LR, and RF have better performance than other compared algorithms in the classification of these data. Finally, it should be noted that DL has the best Recall among all of the investigated algorithms in both Table 6 and Table 9. The rank-based analysis of investigated algorithms is shown in Table 12.

Table 12
Rank-based analysis of investigated algorithms

	5 Classes			2 Classes			
Algorithm	Rank for gos0	Rank for fgos	Mean of Ranks	Rank for gos0	Rank for fgos	Mean of Ranks	Overall Mean of Ranks
NB	4	6	5	7	7	7	6
RF	3	2	2.5	1	3	2	2.25
KNN(k = 5)	8	8	8	10	9	9.5	8.75
KNN(k = 6)	7	7	7	9	10	9.5	8.25
DT	6	4	5	6	8	7	6
RI	5	5	5	4	4	4	4.5
DL	2	3	2.5	8	6	7	4.75
GBT	1	1	1	5	5	5	3
LR	—	—	—	2	2	2	2
GLM	—	—	—	3	1	2	2

Discussion

The present longitudinal study primarily aimed to predict the GOS of recovered TBI patients at discharge and six months after discharge. Our findings showed that different machine learning algorithms applied in this study provide acceptable accurate prediction models using collected health status, demographic features, clinical physical exams, and laboratory data.

The first steps of prediction begin with classifying TBI cases' severity by baseline features. There have been controversies about whether ML can be a more accurate classifier than the neurologist's categorizations. It has been previously claimed that ML algorithms were not more efficient than neurologists [18]. One year later, another study investigated the performance of ANN in providing predictive models using 11 clinical baseline features. This study carried out by Rughani et al. in 2010 showed that ANN could achieve better outcomes than regression models and clinicians' categorizations in predicting the survival of TBI patients, with accuracy equal to 73% [19]. Since then, ML algorithms have been advancing and achieving higher accuracy in predicting the outcomes.

This investigation's first aim was to find the most reliable prognostic markers following TBI. Additionally, widening the range of clinical features is another key point in achieving more efficient models. In different observations during recent years, several features have been introduced as the most reliable variables. In a study by Shi et al., some clinical and demographic features such as length of hospitalization, sex, age, and Charlson comorbidity index developed acceptable predictive DL models using ANN for in-hospital mortality in patients with TBI [5]. Other evaluations introduced some other features, including vomiting, signs of a skull base fracture, loss of consciousness (LOC), and a history of traffic accidents [20]. However, our assessments on wide background, clinical, and paraclinical features with various models indicated that the condition of pupils, the condition of cisterns (whether they are present, absent, or compressed), and the patient's age are the best predictors of in-hospital mortality, while the condition of the pupils, GCSM, and age are the most important clinical features in predicting the long-term mortality [41]. Some factors may stand for different findings among the studies, such as entering different variables into the analysis. For instance, our study used the motor component of GCS rather than the total GCS, which is broadly used in various trials [21]. Supporting our findings, previous studies confirmed that using the motor component of GCS would provide more accurate models in comparison with the total GCS [31].

The second aim of the present study was to provide efficient ML and statistical models to predict the short- and long-term outcomes of TBI patients. The outcomes of TBI would be appropriately predicted using the clinical features of the first day of admission [9], as discussed earlier. The first evaluations emphasized that all prediction models, based on ML or LR, would achieve a high success rate [22]. According to our findings, it would be concluded that the RF, LR and GLM model are the most accurate models to predict the in-hospital mortality of patients (based on the 2-class GOS).

On the other hand, GLM (with an accuracy of 82%) was found to be the most accurate predictor of 6-months mortality. Instead, when using 5-class GOS, GBT was the most accurate predictor of both in-hospital and 6-months follow-up morbidity and mortality. However, as described in the results, the accuracy of the 5-class GOS is lower as compared to the 2-class GOS. In line with our results, in 2019, Matsuo et al. found that RF is the best model for predicting in-hospital outcomes following TBI [42]. In 2015, Lu et al. conducted a study to compare the efficacy of different ML models and LR in predicting 6-month GOS. Using clinical features, ANN showed the best performance, with AUC equal to 0.96 [21].

Applying CT scans in prediction models based on two-layer feed-forward ANN revealed promising outcomes to forecast the TBI prognosis [24]. For instance, in a study by Abujaber et al. in 2020, which included CT scans in their variables, it was shown that SVM was the best tool to predict in-hospital mortality of TBI patients [25]. Also, a model by Steyerberg et al. introduced the Marshal score (a CT scan index) as a major feature of predicting TBI outcomes, alongside glucose, hemoglobin, hypotension, and hypoxia [10]. In our study, Marshal's score ranked 5 among 10 variables regarding the weight of information gain.

Choi et al. provided a novel ML model, named elastic net (EN), to predict the TBI outcomes. This study showed that EN would be a promising model in the future [30]. Although the efficiency of ML models has been developed case by case till today, it is not yet precisely clear how accurate the predictions made by AI can be and how much we can rely on them. There are, therefore, doubts that we can leave some clinical decision makings to AI. A systematic review by Mawdsley et al. indicated a long distance to achieve reliable prediction models [28].

On the other hand, new studies are also constantly developing specified models for outcome prediction in TBI, such as a recent study by Warman et al., which can outperform previously used models [43]. Furthermore, Pourahmad et al. developed a 4-layered ANN with more accuracy, sensitivity, and specificity [23]. Interestingly, by training ML algorithms, Lang et al. also provided clinical decision support for TBI patients that could decrease the 7-day mortality of these patients. This finding highlights the value of ML algorithms for application in critical clinical decision makings [44].

Controversially, in another study in 2020 on a large database of patients referred to the hospital due to moderate to severe TBI, it was observed that ML algorithms could not outperform the LR in predicting the outcome of these patients, which questions the superiority of ML over LR. In this study, the authors suggested that instead of focusing on ML algorithms, we should focus on including more valuable prognostic markers [45]. Another recent investigation also came up with the same conclusion that LR and ML have similar performance. However, this study used a more limited number of predictor variables and did not include serologic markers [46]. Finally, another study in 2022 shows ML provides similar results to correlation and multiple linear regression analysis. However, this study recruited only 168 patients with severe TBI [47].

The strengths of our proposed model are as follows:

1. We have obtained high performance using simple ML algorithms.

2. Employed maximum number of patients and used more features.

The limitations of our automated system are as follows:

1. Several missing data had to be omitted in this work.

2. Using our model for assessments and quick examinations in the emergency department on a critical TBI patient may introduce misevaluations and cause biases, which is inevitable due to the nature of the investigation. Hence, the model must be tested with a huge database before clinical usage.

Conclusion

The results of our study revealed that the prognosis and survival outcome of TBI patients in short- and long-term periods could be reliably predicted using machine learning algorithms, such as RF and GLM. According to our findings, the condition of pupils, GCSM, condition of cisterns, and the patients’ age are the best predictors of their survival.

As the future work, to perform further validations of all these models on other datasets, we plan to collect bigger and multicentre datasets. This makes prediction models more robust. In addition, investigating the health conditions of patients in smaller time interval makes our study results more precise. It may lead to make a faster prediction that can more carefully guide medical attention. In this research, we only focused on the mortality rate. However, efforts can also be made toward predicting the functionality of the patients after a predefined amount of time. It leads to better medical cares as it provides more information for physicians.

Declarations

Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Competing interests

The authors declare no competing interests

Acknowledgements

There is not acknowledgement to declare.

Author contributions

Contributed to prepare the first draft: H.K., M.R., M.A.N., M.S.M., A.As., R.T., and H.P. Contributed to editing the final draft: A.V., M.R., R.A., A.N., A.An., S.M.S.I., and U.R.A. Contributed to all analysis of the data and produced the results accordingly: H.K., M.R., M.A.N., M.S.M., and R.A.

Searched for papers and then extracted data: A.As., R.T., H.P., A.V., M.R., A.N., and A.An. Provided overall guidance and managed the project: A.An., S.M.S.I., U.R.A., and R.T.

References

1. Khokhar B, Simoni-Wastila L, Slejko JF, Perfetto E, Zhan M, Smith GS. In-hospital mortality following traumatic brain injury among older medicare beneficiaries, comparing statin users with nonusers. Journal of pharmacy technology 2017; **33**(6): 225–36.
2. Palchak MJ, Holmes JF, Vance CW, et al. Does an isolated history of loss of consciousness or amnesia predict brain injuries in children after blunt head trauma? Pediatrics 2004; **113**(6): e507-e13.
3. Lu J, Marmarou A, Choi S, et al. Mortality from traumatic brain injury. Intracranial Pressure and Brain Monitoring XII: Springer; 2005: 281-5.
4. Flaada JT, Leibson CL, Mandrekar JN, et al. Relative risk of mortality after traumatic brain injury: a population-based study of the role of age and injury severity. Journal of neurotrauma 2007; **24**(3): 435–45.

5. 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**(4): 746–52.
6. Thompson HJ, Rivara FP, Jurkovich GJ, Wang J, Nathens AB, MacKenzie EJ. Evaluation of the effect of intensity of care on mortality after traumatic brain injury. Critical care medicine 2008; **36**(1): 282.
7. Ho C-H, Liang F-W, Wang J-J, Chio C-C, Kuo J-R. Impact of grouping complications on mortality in traumatic brain injury: A nationwide population-based study. PloS one 2018; **13**(1): e0190683.
8. Becker A, Peleg K, Olsha O, Givon A, Kessel B, Group IT. Analysis of incidence of traumatic brain injury in blunt trauma patients with Glasgow Coma Scale of 12 or less. Chinese journal of traumatology 2018; **21**(03): 152–5.
9. Rubin ML, Yamal J-M, Chan W, Robertson CS. Prognosis of six-month Glasgow Outcome Scale in severe traumatic brain injury using hospital admission characteristics, injury severity characteristics, and physiological monitoring during the first day post-injury. Journal of neurotrauma 2019; **36**(16): 2417–22.
10. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS medicine 2008; **5**(8): e165.
11. Hukkelhoven CW, Steyerberg EW, Habbema JDF, et al. Predicting outcome after traumatic brain injury: development and validation of a prognostic score based on admission characteristics. Journal of neurotrauma 2005; **22**(10): 1025–39.
12. Majdan M, Brazinova A, Rusnak M, Leitgeb J. Outcome prediction after traumatic brain injury: comparison of the performance of routinely used severity scores and multivariable prognostic models. Journal of neurosciences in rural practice 2017; **8**(01): 020–9.
13. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology 1996; **49**(11): 1225–31.
14. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Using trauma registry data to predict prolonged mechanical ventilation in patients with traumatic brain injury: Machine learning approach. PLoS one 2020; **15**(7): e0235231.
15. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future healthcare journal 2019; **6**(2): 94.
16. Feng J-z, Wang Y, Peng J, Sun M-w, Zeng J, Jiang H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. Journal of critical care 2019; **54**: 110–6.
17. Muballe KD, Sewani-Rusike CR, Longo-Mbenza B, Iputo J. Predictors of recovery in moderate to severe traumatic brain injury. Journal of Neurosurgery 2018; **131**(5): 1648–57.
18. Güler İ, Gökçil Z, Gülbandilar E. Evaluating of traumatic brain injuries using artificial neural networks. Expert Systems with Applications 2009; **36**(7): 10424–7.

19. Rughani AI, Dumont TM, Lu Z, et al. Use of an artificial neural network to predict head injury outcome. J Neurosurg 2010; **113**(3): 585–90.
20. Chong S-L, Liu N, Barbier S, Ong MEH. Predictive modeling in pediatric traumatic brain injury using machine learning. BMC Med Res Methodol 2015; **15**: 22-.
21. Lu H-Y, Li T-C, Tu Y-K, Tsai J-C, Lai H-S, Kuo L-T. Predicting long-term outcome after traumatic brain injury using repeated measurements of Glasgow Coma Scale and data mining methods. Journal of medical systems 2015; **39**(2): 1–10.
22. Belliveau T, Jette AM, Seetharama S, et al. Developing Artificial Neural Network Models to Predict Functioning One Year After Traumatic Spinal Cord Injury. Arch Phys Med Rehabil 2016; **97**(10): 1663-8.e3.
23. Pourahmad S, Hafizi-Rastani I, Khalili H, Paydar S. Identifying important attributes for prognostic prediction in traumatic brain injury patients. Methods of Information in Medicine 2016; **55**(05): 440–9.
24. Hale AT, Stonko DP, Lim J, Guillamondegui OD, Shannon CN, Patel MB. Using an artificial neural network to predict traumatic brain injury. J Neurosurg Pediatr 2018; **23**(2): 219–26.
25. 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. Scandinavian journal of trauma, resuscitation and emergency medicine 2020; **28**(1): 1–10.
26. Tunthanathip T, Duangsuwan J, Wattanakitrungroj N, Tongman S, Phuenpathom N. Comparison of intracranial injury predictability between machine learning algorithms and the nomogram in pediatric traumatic brain injury. Neurosurgical focus 2021; **51**(5): E7.
27. Hodel J, Stucki G, Prodinger B. The potential of prediction models of functioning remains to be fully exploited: A scoping review in the field of spinal cord injury rehabilitation. Journal of Clinical Epidemiology 2021; **139**: 177–90.
28. Mawdsley E, Reynolds B, Cullen B. A systematic review of the effectiveness of machine learning for predicting psychosocial outcomes in acquired brain injury: Which algorithms are used and why? Journal of Neuropsychology 2021; **15**(3): 319–39.
29. Alanazi HO, Abdullah AH, Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. Journal of medical systems 2017; **41**(4): 1–10.
30. Choi Y, Park JH, Hong KJ, Ro YS, Song KJ, Do Shin S. Development and validation of a prehospital-stage prediction tool for traumatic brain injury: a multicentre retrospective cohort study in Korea. BMJ open 2022; **12**(1): e055918.
31. Daley M, Cameron S, Ganesan SL, et al. Pediatric severe traumatic brain injury mortality prediction determined with machine learning-based modeling. Injury 2022; **53**(3): 992–8.
32. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Computer

methods and programs in biomedicine 2017; **141**: 19–26.

33. Khozeimeh F, Sharifrazi D, Izadi NH, et al. Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. Scientific Reports 2021; **11**(1): 1–18.
34. Sharifrazi D, Alizadehsani R, Roshanzamir M, et al. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomedical Signal Processing and Control 2021; **68**: 102622.
35. Joloudari JH, Mojrian S, Nodehi I, et al. Application of Artificial Intelligence Techniques for Automated Detection of Myocardial Infarction: a Review. 2022.
36. Khozeimeh F, Alizadehsani R, Roshanzamir M, Khosravi A, Layegh P, Nahavandi S. An expert system for selecting wart treatment method. Computers in biology and medicine 2017; **81**: 167–75.
37. Alizadehsani R, Sharifrazi D, Izadi NH, et al. Uncertainty-Aware Semi-Supervised Method Using Large Unlabeled and Limited Labeled COVID-19 Data. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 2021; **17**(3s): 1–24.
38. Alizadehsani R, Abdar M, Roshanzamir M, et al. Machine learning-based coronary artery disease diagnosis: A comprehensive review. Computers in biology and medicine 2019; **111**: 103346.
39. RapidMiner. https://rapidminer.com/.
40. Alizadehsani R, Habibi J, Hosseini MJ, et al. A data mining approach for diagnosis of coronary artery disease. Computer methods and programs in biomedicine 2013; **111**(1): 52–61.
41. Rocha TAH, Elahi C, da Silva NC, et al. A traumatic brain injury prognostic model to support in-hospital triage in a low-income country: a machine learning–based approach. Journal of neurosurgery 2019; **132**(6): 1961–9.
42. Matsuo K, Aihara H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. Journal of neurotrauma 2020; **37**(1): 202–10.
43. Warman PI, Seas A, Satyadev N, et al. Machine learning for predicting in-hospital mortality after traumatic brain injury in both high-income and low-and middle-income countries. Neurosurgery 2022; **90**(5): 605–12.
44. Lang E, Neuschwander A, Favé G, et al. Clinical decision support for severe trauma patients: machine learning based definition of a bundle of care for hemorrhagic shock and traumatic brain injury. Journal of trauma and acute care surgery 2022; **92**(1): 135–43.
45. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. Journal of clinical epidemiology 2020; **122**: 95–107.
46. Bruschetta R, Tartarisco G, Lucca LF, et al. Predicting Outcome of Traumatic Brain Injury: Is Machine Learning the Best Way? *Biomedicines* 2022; **10**(3): 686.

47. Kazim SF, Sursal T, Schmidt MH, Bowers CA, Cole CD. Predictors of in-hospital mortality and length of stay after severe traumatic brain injury: conventional regression versus machine learning-based analyses. 2021.

Figures

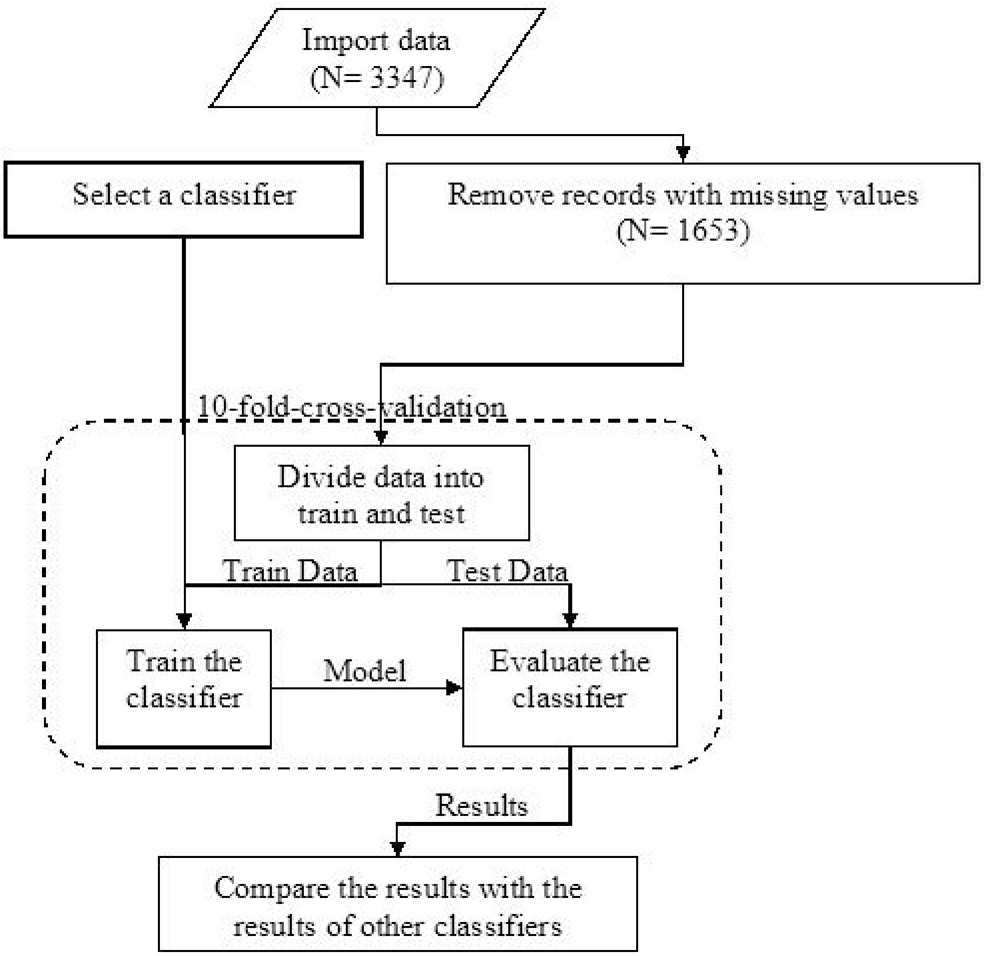

Figure 1

The flowchart of analyzing the data with different classifier algorithms