https://archtrauma.kaums.ac.ir/
Archives of Trauma Research. 2023; 12(3): 137-149
https://doi.org/10.48307/ATR.2023.407607.1020

Research Article Open Access

A comprehensive evaluation of ensemble learning methods and decision trees for predicting trauma patient discharge status using real-world data

Zahra Kohzadi [1,2], **Ali Mohammad Nickfarjam** [1,2*], **Leila Shokrizadeh Arani** [1,2], **Zeinab Kohzadi** [3], **Mehrdad Mahdian** [4]

[1] Department of Health Information Management and Technology, Kashan University of Medical Sciences, Kashan, Iran
[2] Research Centre for Health Information Management, Kashan University of Medical Sciences, Kashan, Iran
[3] Medical Informatics Department, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4] Trauma Research Center, Kashan University of Medical Sciences, Kashan, Iran

[*] **Corresponding author:** Ali Mohammad Nickfarjam, Department of Health Information Management and Technology, Kashan University of Medical Sciences, Kashan, Iran. **Email:** nickfarjam-a@kaums.ac.ir

Received: 18 July 2023 **Revised:** 15 September 2023 **Accepted:** 11 October 2023 **e-Published:** 1 November 2023

Abstract

Background: Trauma registries collect and document data about the acute injury care in hospitals. The goal of trauma care systems is to reduce injury occurrence and enhance trauma patient survival rates.

Objectives: In this article, the Kashan trauma registry was used to predict trauma patient discharge status using machine learning.

Methods: This study employed 3930 Kashan Trauma Centre Registry entries after preprocessing. The study experimented with decision trees of varying complexity, using three separate metrics - information gain, Gini index, and gain ratio - to build and evaluate the trees. Finally, bagging, boosting and stacking ensemble learning techniques were implemented to evaluate their predictive performance. Ensemble learning models were developed based on decision trees of varying depths that utilized different learning measures/metrics. The predictive performance of the algorithms was evaluated using metrics such as accuracy, precision, recall, and the area under the receiver operating characteristic curve (AUC). This study aimed to compare ensemble-learning techniques like bagging, boosting and stacking to decision trees configured with various parameter settings, to assess their ability to predict trauma patients' discharge status outcomes.

Results: The stacking technique, which used decision tree algorithms (depth=5) that integrated parameters like information gain, gain ratio and Gini index at the base level along with KNN (k=12) using Euclidean distance, and then incorporated logistic regression as the meta-classifier, demonstrated superior predictive performance compared to using individual decision trees, bagging or boosting approaches alone.

Conclusion: However, while decision trees are straightforward algorithms and ensemble methods are more time-consuming and computationally complex, this study indicates that stacking learning is superior to single decision tree methods with a variety of parameters, bagging, and boosting.

Keywords: Data Mining, Ensemble Learning, Trauma, Decision Trees.

Introduction

As the term implies, trauma refers to physical or psychological injury resulting from an outside factor, such as an accident, violence, or natural disaster. There are physical and psychological injuries that can occur as a result of trauma, including fractured bones, burns, and wounds, as well as anxiety, depression, and post-traumatic stress disorder (PTSD).[1-3]

Globally, serious injury is the leading cause of death, and its effects on the economy, society, and individuals are significant.[4] WHO's Global Status of Road Safety Report 2018 cites traffic accidents as the leading cause of mortality for those aged 5 to 29. Most of the burden is placed on motorcyclists, cyclists, and pedestrians in underdeveloped countries. Due to established criteria, switching has a relatively high cost, according to the report. Taking serious action now is crucial to achieving any future global goals and saving lives.[5] Trauma prevention and treatment coordination present significant challenges to society because trauma affects prognosis.[4]

Using large national databases, we can identify risk factors and outcomes that are associated with better health. The most common data found in these databases is information about hospitalized patients.[6] Implementing trauma care systems, including trauma registries, has been largely responsible for the huge reduction in injury-related death and disability rates.[7]

Typically, trauma registries contain information about patient demographics, injury circumstances, pre-hospital treatment and transportation, emergency department visits, hospitalizations, descriptions of anatomic injuries, physiological measurements, complications, outcomes, and patient destinations. Furthermore, they increasingly contain information about pre-existing illnesses, which are recognized to be important independent of age and injury severity when predicting outcomes.[8-10] A researcher can use the data to discover trends and patterns in trauma, assess the effectiveness of various therapies, and generate new ideas regarding the causes and effects of trauma.[11-13] Due to the large volumes of data contained in trauma registries, the report notes that artificial intelligence (AI) techniques like machine learning are now being utilized to aid in diagnosis and knowledge discovery. Machine learning allows researchers to analyze trauma registry data at a scale not possible through manual review alone.[14,15]

Since healthcare relies on data and machine learning is capable of extracting information, it is crucial to the healthcare industry.

In the past 60 years, many industry pioneers have guided us in the right direction. Machine learning (ML) involves using various algorithmic approaches, with the specific type chosen dependent on the desired outcome. ML is commonly leveraged in software applications to improve the user experience through features like personalized recommendations. Healthcare data analytics is an area where ML is increasingly being applied, as companies work to gain insights that can help advance patient care and outcomes. A wide range of ML techniques are now being explored for medical and health-related uses. In order to give healthcare practitioners better-informed data, some startups are combining big data with machine learning.[16] In recent years, machine learning for healthcare has gained popularity with a growing number of studies in blood pressure,[17,18] psychology,[19] diabetes,[20] traumatic coagulopathy,[21] diabetic retinopathy,[22] and lymph node detection in head and neck.[23]

Consequently, ensemble learning is garnering growing research interest from investigators in recent years as a potent application of machine learning for healthcare and medical domains.[24-27]

In ensemble learning, multiple machine learning models are combined together rather than relying on one model alone to solve a particular problem. In this approach, several models work in parallel to solve a problem. The rationale behind ensemble learning is to generate a set of predictive models or hypotheses across different machine learning algorithms and then combine them in a way that aims to yield more accurate results than could be obtained from any single model or technique alone.

Integrating diverse algorithms enhances predictive performance compared to using a single algorithm alone by capitalizing on their individual strengths. Ensemble methods combine predictions from multiple models to create extremely powerful forecasting methods.

Ensemble methods combine predictions from multiple models to create extremely powerful forecasting methods. While exhibiting excellent generalizability across new data sets and scenarios, they omit overly specific or localized patterns that may not extend well to other contexts. The blend process leads to predictions that generalize well, even without detailed niche knowledge.[28,29]

The various target classes are not uniformly represented in imbalanced data, with some classes having many more examples than others. Machine learning models may favor the majority class over the minority class as a result of this. On datasets with class imbalance, where learning algorithms may otherwise struggle to accurately predict the underrepresented classes, ensemble methods are commonly used to assist with this problem.[30,31]

Decision trees are an effective machine learning technique for classification and predictive problems. They have several beneficial properties that make them well-suited for these tasks, including being straightforward to interpret, not requiring assumptions about the underlying data distribution, and being trained on large and diverse datasets containing different types of variables, such as continuous and categorical features.

In addition to their non-parametric nature, decision trees can handle a wide range of data types, making them an extremely versatile algorithm.[32-34]

Given the benefits of ensemble methods for improving performance on imbalanced data problems as well as the strengths of decision trees as a classification algorithm, this study employed an ensemble approach incorporating multiple decision tree models. The goal was to leverage both techniques to best address the challenge of the uneven class distribution present in the targeted predictive task. Considering that the use of machine learning has been useful in various studies[35-40] [Table 1], using this approach, we can gain a greater understanding of trauma and improve our treatment options-wide.

Table 1. Related studies

Study	Methods	Result
Abujaber et al. [35]	During January 2014 and February 2019, this study enrolled adult patients with TBIs admitted to a level 1 trauma facility. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) were evaluated using metrics such as accuracy, Area Under the Curve (AUC), sensitivity, precision, negative predictive value (NPV), specificity, and F-score.	Support Vector Machines (SVMs) demonstrated superior performance with accuracy and AUC of 95.6%.
Feng et al. [36]	A comparative study was conducted between twenty-two machine learning (ML) models and logistic regression in predicting survival among patients with severe traumatic brain injury (STBI).	A Cubic SVM, a Quadratic SVM, a Linear SVM, and a Linear Decision Tree all performed better than linear regression.
Benjamin et al. [37]	In their study, logistic regression (LR), lasso regression, and ridge regression were used, using fundamental predictors derived from IMPACT-II. There was also machine learning (ML) methods used simultaneously, including support vector machines, random forests, gradient boosting machines, and artificial neural networks.	The conventional regression method can perform comparably or better than machine learning algorithms in low-dimensional contexts of outcome prediction for moderate and severe traumatic brain injuries. In order to ensure ML algorithms' relevance across multiple populations, rigorous validation remains essential.
Lu et al. [38]	The purpose of this study was to integrate data mining techniques with serial Glasgow Coma Scale (GCS) scores and clinical and laboratory parameters in order to predict 6-month functional outcomes and mortality among patients with traumatic brain injury (TBI). Both mortality and functional outcomes were forecasted using artificial neural network (ANN), naïve Bayes (NB), decision tree, and logistic regression methodologies.	AUC values of 96.13%, 83.50%, and 89.73% were achieved by the artificial neural network (ANN) for functional outcome forecasting. This model generated the best mortality prediction results with an AUC of 91.14%, a sensitivity of 81.17%, and a specificity of 90.65%.
Ploeg et al. [39]	The study predicted 6-month mortality in traumatic brain injury (TBI) patients using various models, including logistic regression (LR), classification and regression trees (CART), random forests (RF), support vector machines (SVM), and neural networks (NN) to make prognostications.	Logistic regression models performed best among the complex predictive models, achieving the highest validated median area under the receiver operating characteristic curve value of 0.757.
Hertz et al. [40]	The study assessed various machine learning algorithms, including decision tree, logistic regression, naive Bayes, support vector machine (SVM), K-nearest neighbor (KNN), and ensemble classifiers, to classify whether bladder injury occurred based on factors during initial presentation of blunt pelvic trauma.	The Gaussian and Kernel Naive Bayes classifiers performed best, both achieving high accuracy of 97.8%, specificity of 99%, sensitivity of 83%, and AUC of 0.99. These optimal classification results were obtained using naive Bayes models.

Objectives

This study aimed to use ensemble and decision tree machine learning models to predict trauma patients' discharge status. There are several potential benefits to this approach, such as improved patient outcomes, more optimized resource utilization, cost savings, increased efficiency, and personalized care. The goal was to provide hospital care teams with predictive insights to inform patient care and resource planning. This study is divided into five sections:

Section 1 outlines the data used and discusses the machine learning architectures and algorithm selection.

Section 2 describes the evaluation method. Section 3 presents the results from running the algorithms on the data. Section 4 discusses the findings, limitations, and conclusions. Section 5 discusses potential future work. By covering the data, methods, results, discussion of findings and limitations, and avenues for future research, the six sections provide a comprehensive overview of the study.

Methods

We conducted a retrospective analysis of the Kashan Trauma Center registry data. The objective of this study was to compare decision trees and their combination with ensemble methods to predict the discharge status of trauma patients.

2.1 Dataset

In this study, data from March 2018 to February 2019 were employed, obtained from the Kashan Trauma Registry. We obtained 3930 records after pre-processing the data, including removing missing data and outlier records. As a result, missing values were imputed using the mean for numerical variables, and using the mode for categorical variables. The numerical variables in the data were normalized using the min-max normalization formula [Formula 1]. The categorical variables were encoded using one hot encoding to transform them into binary coded formats.[41, 42]

$$Y = \frac{X - X(MIN)}{X(MAX) - X(MIN)} \quad (1)$$

2.2 Ensemble Algorithm

Ensemble classifiers combine basis classifiers for the final classification. Any supervised classifier, such as decision trees, neural networks, or support vector machines, can be used as the base classifier. A brief overview of ensemble learning methods is provided here. One of the earliest ensemble algorithms developed was bagging (also known as bootstrap aggregating). Bagging helps to reduce the model's variance, which makes it effective for unstable models like decision trees. It is also easy to implement in parallel and distributed environments.[43-45] By using the boosting method, each base classifier is not built independently. Rather than doing it all at once, the basic classifiers are built one at a time, considering any mistakes made by the fundamental classifiers before them. This method is more effective than bagging, and it minimizes system bias. Additionally, it can be used with weaker models.[44-47] The stacking method, also referred to as stacked generalization, constructs ensembles differently than bagging or boosting. As described by Wolpert, stacking utilizes the outputs or predictions of the original base classifiers as inputs to train a new level of classifiers. Through this technique of leveraging one classifier's outputs to inform another, stacking is able to estimate and compensate for biases present in the individual base models.[44, 48]

2.3 Decision Trees

As a classification strategy, decision tree analysis is basically a divide-and-conquer approach. Decision trees can be used in large databases to find important features and patterns needed for distinguishing between things and predicting them. The features of decision trees as well as their natural interpretation have led to their widespread use for both exploratory data analysis and predictive modeling over the past two decades.[49] When constructing decision trees, several metrics can evaluate the purity or impurity of potential split points in the data, such as the Gini index, information gain, gain ratio, and misclassification rate. Information gain quantifies the reduction in uncertainty of the target variable resulting from a split. Gain ratio considers the inherent information contained in the splitting attributes. The Gini index measures the degree of impurity in the target variable after a split. Depending on the specific problem and characteristics of the available data, different metrics may be more suitable for selecting the best split rule. The properties of the situation and data should inform the choice of an appropriate impurity measure to guide tree construction.[50, 51]

2.4 Regression Logistic

For binary classification challenges, logistic regression is commonly used due to its statistical underpinnings.[52] Previous studies have incorporated logistic regression as the second layer in ensembles.[53-55] Similarly, logistic regression was also part of our proposed approach in this study.

2.5 KNN

KNN classification predicts queries based on the majority class of its k nearest neighbors in training data. Due to this capability, KNN was selected as one of the top 10 data mining algorithms [50, 56, 57]. Ensemble methods like random forests, gradient boosting and AdaBoost often use decision trees as base algorithms. These methods train multiple decision trees on subsets of training data or with randomization. Combining tree predictions improves accuracy by utilizing strengths of different algorithms while preventing overfitting.[58] Decision trees are commonly used as they find nonlinear relationships and are interpretable.[59]

2.6 Implemented Framework

This section explains how to select a model. The study's approach is categorized into four sections, as depicted in Figure 1.

- **In the initial section,** the study employs a total of 12 decision trees for classification. These decision trees are characterized by utilizing the Gini index and have depths of 10, 8, 6, and 5. Additionally, decision trees based on information gain and gain ratio are utilized, also with depths of 10, 8, 6, and 5.
- **The second section** involves a depth-based model. Within the framework of the bagging algorithm, decision trees are utilized for voting. For instance, decision trees employing the Gini index, information gain, and gain ratio indices are used for various depths, such as Bagging-10-depth, Bagging-8-depth, Bagging-6-depth, and Bagging-5-depth. This process is replicated for the boosting algorithms, resulting in the creation of four models named Boosting-depth-10, Boosting-depth-8, Boosting-depth-6, and Boosting-depth-5.
- **The third section** focuses on a learning metrics-based model. In this part, decision trees are utilized for voting based on each learning metric. For instance, decision trees with depths of 10, 8, 6, and 5 are employed for the information gain learning metric. This process is repeated for other learning metrics as well. Consequently, three models of the bagging algorithm are derived: Bagging-information gain, Bagging-gain ratio, and Bagging-Gini index. The same procedure is applied to the boosting algorithm, resulting in the creation of three models: Boosting-information gain, Boosting-gain ratio, and Boosting-Gini index.
- **The fourth section** involves the utilization of a stacking ensemble approach. The base algorithms employed in this approach were decision tree and k-nearest neighbors, while the Meta level or the top-level algorithm used was logistic regression.

Furthermore, in terms of measuring distance, the Euclidean distance exhibited superior performance compared to other metrics, particularly when k=12. As a result, all stacking models utilized the k-nearest neighbors (KNN) algorithm with 12 nearest neighbors and the Euclidean distance metric. The 10-fold cross-validation technique was employed throughout all stages of the study. As an example, the Stacking-Gini index model incorporates the following components:

For the Stacking-Gini index model:

Base Level:

- Decision trees with depths of 5, 6, 7, and 8, using the Gini index.
- K-nearest neighbors (KNN) with k=12, using the Euclidean distance metric.

Meta Level:

- Logistic regression.

For the Stacking-depth-10 model:

Base Level:

- Decision trees with a depth of 10, utilizing the Gini index, information gain, and gain ratio.
- K-nearest neighbors (KNN) with k=12, using the Euclidean distance metric.

Meta Level:

- Logistic regression.

Bagging is an ensemble technique that involves training an algorithm multiple times using different subsets of the training data. The final prediction is determined by averaging the predictions of all the sub-models. On the other hand, boosting is a sequential training approach where multiple models are trained iteratively, with each subsequent model learning from the mistakes made by its predecessors. In each iteration, data points are assigned weights, and the focus is shifted to reweighting misclassified points in the next round. The final classifier is a weighted sum of the ensemble predictions.

Stacking, another ensemble technique, combines diverse machine learning models (known as base learners) using an additional data mining method. First, the base learners are trained individually. Then, a combiner (referred to as the meta-classifier) is trained to generate the final prediction based on the predictions made by the base learners. The process of these algorithms is depicted in Figure 2.

3. Evaluation

A confusion matrix is a commonly used table for evaluating the performance of a classification model. It presents the counts of true positives, false positives, true negatives, and false negatives for each class in the dataset. These values in the confusion matrix are utilized to compute various performance metrics, including accuracy, precision, recall, and F1 score [60]. The ROC curve is a graphical representation of the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) as the decision threshold of a binary classification model is adjusted. The AUC value, ranging from 0 to 1, indicates the model's overall classification performance, with 1 representing perfect classification and 0.5 indicating random classification.

To evaluate the algorithms' performance, we calculated several metrics including accuracy, precision, recall, F-measure, and the area under the ROC curve (AUC). These metrics provide insights into the algorithms' ability to correctly classify different categories. Specifically, we define the following evaluation indices:

- TP: Represents a patient who is actually improved and has been correctly classified as such by the predictor.
- TN: Represents a patient who is actually non-improved and has been correctly classified as such by the predictor.
- FP: Represents an observation that is incorrectly classified as improved when it is actually non-improved.
- FN: Represents a patient who is actually non-improved but has been incorrectly classified as improved by the predictor.

These evaluation indices help assess the algorithms' performance in accurately identifying improved and non-improved cases.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (5)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (6)$$

$$\text{F-measure} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (7)$$

Considering the data imbalance, precision and recall criteria are also important, in addition to accuracy. In order to balance precision and recall, the F-measure criterion is used.

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki. Institutional Review Board approval (Ethics code: IR.KAUMS.NUHEPM.REC.1401.038) was obtained.

Results

A summary of the data employed for this research can be found in Table 2

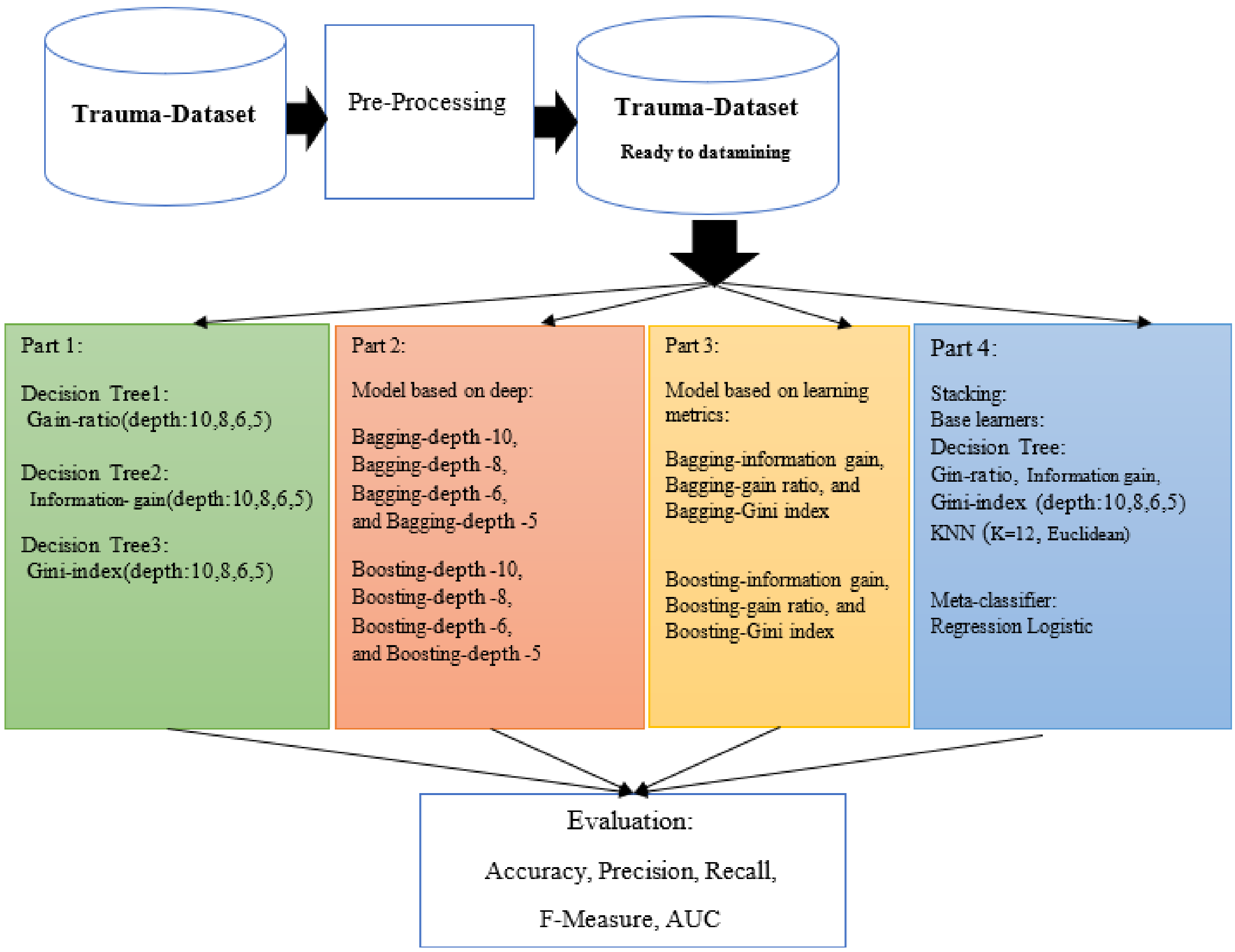

Figure 1. Architecture of a data mining technique and the process of algorithm selection

There is also accuracy, precision, recall, F-measures, and AUCs in Table 4. The highest and lowest values for accuracy were 83.89% and 82.62%, precision 90.11% and 84.69%, recall 92.96% and 83.98%, F-Measure 88.63% and 86.93%, and AUC 0.87 and 0.85, respectively. For the bagging algorithm, the highest and lowest values were 83.87% and 83.03 %, precision 89.78% and 84.84%, recall 92.28% and 84.39%, F-Measure 88.51% and 87.00%, and AUC 0.88 and 0.84, respectively. As a result of the boosting algorithm, the highest and lowest accuracy values were 83.87% and 83.08%, precision 89.88% and 84.78%, recall 92.77% and 84.34%, F-Measure 88.59% and 87.02%, and AUC 0.86 and 0.82, respectively.

The highest and lowest accuracy values for the stacking algorithm were 85.75% and 83.69%, precision 86.59% and 85.93%, recall 94.21% and 89.93%, F-Measure 89.89% and 88.11%, and AUC 0.89 and 0.87, respectively.

TP, TN, FP, and FN for various algorithms are shown in Table 3 based on the confusion matrix.

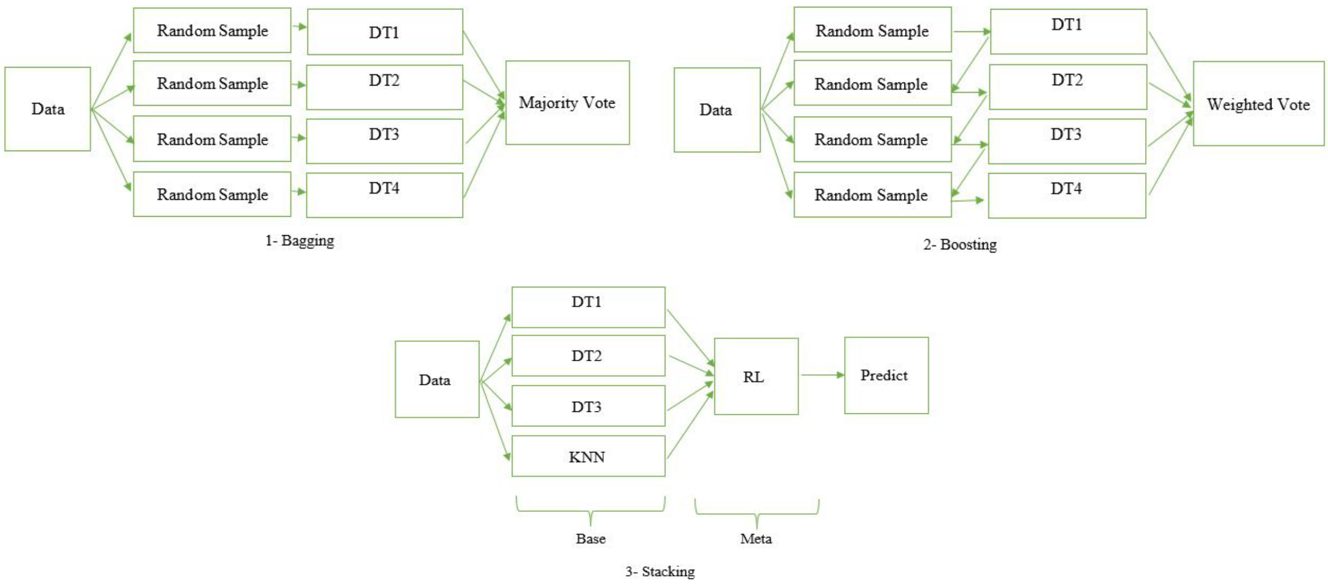

Figure 2. Process of ensemble algorithms. (DT1, DT2, DT3, DT4: Decision tree with different parameters, RL: regression logistic)

Table 2. Dataset used in data mining

Features	Feature type	Category	Percentage
Age	Numerical		
Place birth	Categorical	City	87.07
		Village	12.92
Type of insurance	Categorical	Treatment services	6.94
		Social security	26.53
		Military	1.17
		Bank	0.15
		Free	12.29
		Others	52.9
Sex	Categorical	Male	78.06
		Female	21.94
Occupation	Categorical	Child	4.86
		Staff	0.96
		Worker	14.32
		Farmer	3.81
		Unemployed	6.1
		Students	14.73
		Businessmen	8.54

		Housewives	15.67
		Others job	28.01
		Unknown	0.03
Education	Categorical	Child	4.86
		Illiterate	2.72
		School	13.89
		High school	72.41
		After diploma	6.1
Type of conveyance carrying to emergency	Categorical	Ambulance	71.83
		Taxi	0.17
		Personal vehicle	27.97
Total expenditures	Numerical	-	-
Number of days admitted	Numerical	-	-
ICD-external causes	Categorical	Pedestrian injured in transport accident	7.5
		Pedal cyclist injured in transport accident	1.8
		Motorcycle rider injured in transport accident	31.29
		Car occupant injured in transport accident	12.74
		Water transport accidents	22.59
		Slipping, tripping, stumbling and falls	18.72
		Exposure to electric current, radiation and extreme ambient air temperature and pressure	4.37
ICD-injuries	Categorical	Injuries to the head	12.49
		Injuries to the abdomen, lower back, lumbar spine, pelvis and external	1.52
		genitals	5.36
		Injuries to the shoulder and upper arm	11.19
		Injuries to the elbow and forearm	21.65
		Injuries to the wrist, hand and fingers	7.83
		Injuries to the hip and thigh	9.1
		Injuries to the knee and lower leg	6.87
		Injuries to the ankle and foot	
State of discharge	Categorical	No improvement	31.79
		Improvement	67.21

Table 3. True Positive, True Negative, False Positive, False Negative for different algorithms

	TP	TN	FP	FN
Decision Tree1-10*	2222	1039	244	424
Decision Tree2-10	2372	893	385	279
Decision Tree3-10	2351	916	365	298
Decision Tree1-8	2233	1031	253	414
Decision Tree2-8*	2383	888	382	277
Decision Tree3-8	2380	893	386	271
Decision Tree1-6	2235	1030	254	411
Decision Tree2-6	2448	844	431	206
Decision Tree3-6	2433	847	428	222
Decision Tree1-5	2300	947	300	383
Decision Tree2-5	2468	829	446	187
Decision Tree3-5*	2462	834	442	192
bagging-Gini	2433	859	421	217
Bagging-Information gain	2445	850	427	208

Bagging-Gain ratio	2232	1031	254	413
Bagging-depth-10	2350	946	334	300
Bagging-depth-8	2373	911	362	284
Bagging-depth-6	2393	884	381	272
Bagging-depth-5	2451	836	438	205
Boosting-Gini	2378	900	380	272
Boosting-Information gain	2384	901	370	275
Boosting-Gain ratio	2230	1035	251	414
Boosting- depth-10	2325	963	316	326
Boosting- depth-8	2357	919	350	304
Boosting- depth-6	2428	850	417	235
Boosting- depth-5	2462	834	442	192
Stacking-Gini	2376	913	375	266
Stacking-Information gain	2411	912	376	231
Stacking-Gain ratio	2468	884	404	174
Stacking- depth-10	2409	915	373	233
Stacking- depth-8	2453	901	387	189
Stacking- depth-6	2476	885	403	166
Stacking- depth-5	2489	881	407	153

*Decision Tree1-10: Gain-ratio(depth:10), Decision Tree2-8: Information-gain(depth:8), Decision Tree3-5: Gini index(depth:5)

Table-4. The performance of Decision Tress algorithms and Models based on the confusion matrix and the Area under the ROC Curve

	Accuracy	Precision	Recall	F Measure	AUC
Decision Tree1-10*	83.00%	90.11%	83.98%	86.93%	0.86
Decision Tree2-10	83.10%	86.04%	89.48%	87.72%	0.87
Decision Tree3-10	83.13%	86.56%	88.75%	87.64%	0.88
Decision Tree1-8	83.03%	89.82%	84.36%	87.01%	0.85
Decision Tree2-8*	83.23%	86.18%	89.59%	87.85%	0.88
Decision Tree3-8	83.28%	86.04%	89.78%	87.87%	0.88
Decision Tree1-6	83.08%	89.80%	84.47%	87.05%	0.85
Decision Tree2-6	83.79%	85.03%	92.24%	88.49%	0.88
Decision Tree3-6	83.46%	85.04%	91.64%	88.22%	0.87
Decision Tree1-5	82.62%	88.46%	85.72%	87.07%	0.85
Decision Tree2-5	83.89%	84.69%	92.96%	88.63%	0.86
Decision Tree3-5*	83.87%	84.78%	92.77%	88.59%	0.86
Bagging-Gini	83.77%	85.25%	91.81%	88.41%	0.88
Bagging-Information gain	83.84%	85.13%	92.16%	88.51%	0.88
Bagging-Gain ratio	83.03%	89.78%	84.39%	87.00%	0.84
Bagging-depth-10	83.87%	87.56%	88.68%	88.11%	0.88
Bagging-depth-8	83.56%	86.76%	89.31%	88.02%	0.88
Bagging-depth-6	83.38%	86.27%	89.79%	87.99%	0.86
Bagging-depth-5	83.64%	84.84%	92.28%	88.40%	0.86
Boosting-Gini	83.41%	86.22%	89.74%	87.94%	0.86
Boosting-Information gain	83.59%	86.56%	89.66%	88.08%	0.86
Boosting-Gain ratio	83.08%	89.88%	84.34%	87.02%	0.85
Boosting- depth-10	83.66%	88.03%	87.70%	87.87%	0.86

Boosting- depth-8	83.36%	87.07%	88.58%	87.82%	0.86
Boosting- depth-6	83.41%	85.34%	91.18%	88.16%	0.85
Boosting- depth-5	83.87%	84.78%	92.77%	88.59%	0.82
Stacking-Gini	83.69%	86.37%	89.93%	88.11%	0.87
Stacking-Information gain	84.55%	86.51%	91.26%	88.82%	0.87
Stacking-Gain ratio	85.29%	85.93%	93.41%	89.52%	0.88
Stacking- depth-10	84.58%	86.59%	91.18%	88.83%	0.87
Stacking- depth-8	85.34%	86.37%	92.85%	89.49%	0.88
Stacking- depth-6	85.52%	86.00%	93.72%	89.69%	0.89
Stacking- depth-5	85.75%	85.95%	94.21%	89.89%	0.89

Discussion

Until 2018, traffic accidents ranked eighth in terms of global mortality. These accidents result in millions of injuries and thousands of fatalities annually. However, it is important to acknowledge that all these deaths and injuries could have been prevented. Road safety, an issue that deserves more attention, holds immense potential for saving lives worldwide.[5] Consequently, this retrospective study explores the application of ensemble approaches in predicting the mortality of trauma patients.

The study findings indicate that data mining techniques exhibit satisfactory performance in predicting the mortality of trauma patients. One of the advantages of this study is the utilization of indigenous data. Furthermore, the scarcity of research on trauma data mining and ensemble learning adds to its significance. The results demonstrate that data mining approaches, particularly ensemble methods, exhibit strong predictive capabilities for patient mortality. Consequently, we can anticipate further research in this field. Previous studies have consistently shown that ensemble learning outperforms individual learning approaches.

In Raza's study,[26] it was demonstrated that ensemble learning outperforms individual classifiers in terms of accuracy. Similarly, in the present study, ensemble learning algorithms exhibited superior performance compared to individual algorithms. Moreover, another study[61] found that the performance of individual learning algorithms was weaker compared to ensemble learning algorithms. In a study[62] focused on categorizing brain tumors and auto-immune disease lesions using magnetic resonance imaging, an ensemble learning approach was proposed. The base learner consisted of a support vector machine classifier and a majority voting prediction model. The experimental results indicated an overall training accuracy of 97.95% and a testing accuracy of 97.744% for their proposed model. Another study[66] successfully utilized an ensemble learning approach to predict heart disease. Multiple classification techniques were employed in this investigation.

According to the study, the ensemble approach outperformed previous categorization methods with an accuracy of 86.32%. According to this study, algorithms that are working together to learn perform better than those that are working alone.

The results of this study and those of studies[35,36,38,39,40,63] showed that machine learning algorithms could be useful in trauma research.

According to the study,[37] machine learning algorithms are not more efficient than traditional regression algorithms, whereas the current study examined different parameters in the decision tree and then applied them to other collective algorithms, and the results showed that the algorithm could be used to predict trauma patients' discharge status.

In spite of the fact that the present study showed good results for machine learning methods, it also had some limitations. These limitations include: Data were for some years ago and therefore, accessing to PMH (Past medical history) is not possible for missing data. Therefore, we cannot determine the efficiency of imputation methods with respect to real values.

Moreover, conventional techniques like mean and mode were employed to handle missing data. Nevertheless, alternative approaches might yield diverse outcomes more effectively. The study primarily concentrated on utilizing decision trees and the ensemble method, which enables the integration of other data mining techniques. Subsequent research on trauma data could explore additional methods, such as oversampling and undersampling, to address imbalanced data. Moreover, alternative classification algorithms like SVM, neural networks, random forest, and Naïve Bayes were employed for diagnosing the discharge status of trauma patients.

Conclusions

The research findings revealed that the efficacy of decision trees is not solely determined by the depth of the tree, but also by the accurate selection of parameters such as information gain, gain ratio, and Gini index. Furthermore, the study demonstrated that employing ensemble learning techniques, such as stacking, can yield superior results compared to relying solely on the decision tree algorithm. The evaluation of a classifier's performance should not rely solely on precision and recall, as it depends on factors such as the study's objective, the cost of false positives and false negatives, and the balance or imbalance of classes. However, when precision, recall, and accuracy simultaneously exhibit high values, it indicates excellent performance of the classifier model. Ensemble learning, particularly through group voting, can outperform individual classification algorithms. The ensemble algorithm's acceptable performance is demonstrated when precision, recall, and accuracy all display high values. The selection of the best underlying algorithms directly impacts the performance of ensemble learning algorithms. By selecting appropriate metrics and depths, the ensemble algorithm effectively utilizes the distance and number of correct neighbors in KNN. Overall, the results indicate that machine learning techniques can effectively generate outputs for predicting the discharge status of trauma patients, and these outputs were deemed satisfactory in terms of their value.

Acknowledgment

This article is extracted from a research project supported by the Kashan University of Medical Sciences and we appreciate the collaboration of the trauma research center at Kashan University of Medical Sciences, Kashan, Iran.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

SVM: Support Vector Machine; ML: Machine learning; RF: Random Forest; KNN: K-Nearest Neighbor; NB: Naïve Bayes; TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; LR: Logistic regression; WHO: World Health Organization

Authors' contributions

AM. N, ZA. K, M. M: Conceived and designed the analysis; Collected the data; Revision. AM. N, ZA. K, ZE. K: Conceived and designed the analysis; Contributed data or analysis tools; Performed the analysis; Writing; Revision and Editing; Investigation; Methodology. AM. N, ZA. K, ZE. K, L. SH, M. M: Review of Related works in the field of trauma, Writing; Editing and Revision. All authors reviewed and approved the articleAll authors read and approved the final manuscript. All authors take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

This study was supported by a grant from the Research Council of Kashan University of Medical Sciences [grant number: 401062]. The authors did not receive any grants from nonprofit organizations or funding agencies either in public or commercial sectors.

Role of the funding source

None.

Availability of data and materials

The data used in this study are available from the corresponding author on request.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. Institutional Review Board approval (Ethics code: IR.KAUMS.NUHEPM.REC.1401.038) was obtained.

Consent for publication

By submitting this document, the authors declare their consent for the final accepted version of the manuscript to be considered for publication.

References

1. APA. Trauma 2021. https://www.apa.org/topics/trauma.
2. CDC. Traumatic Brain Injury & Concussion 2021. https://www.cdc.gov/traumaticbraininjury/index.html.
3. WHO. Injuries and violence 2021 [Available from: https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence.
4. Nathens AB, Brunet FP, Maier RV. Development of trauma systems and effect on outcomes after injury. The Lancet 2004; 363 (9423):1794-801. doi:10.1016/S0140-6736(04)16307-1 PMid:15172780
5. WHO. Global status report on road safety 2018. Geneva 2018.
6. Pfeifer R, Tarkin IS, Rocos B, Pape H-C. Patterns of mortality and causes of death in polytrauma patients-has anything changed? Injury 2009;40(9):907-11. doi:10.1016/j.injury.2009.05.006 PMid:19540488
7. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. lancet 1997;349(9064):1498-504. doi:10.1016/S0140-6736(96)07492-2 PMid:9167458
8. McGwin Jr G, MacLennan PA, Fife JB, Davis GG, Rue III LW. Preexisting conditions and mortality in older trauma patients. J Trauma Acute Care Surg 2004;56(6):1291-6. doi:10.1097/01.TA.0000089354.02065.D0 PMid:15211139

9. Milzman DP, Boulanger BR, Rodriguez A, Soderstrom CA, Mitchell KA, Magnant CM. Pre-existing disease in trauma patients: a predictor of fate independent of age and injury severity score. J Trauma 1992;32(2):236-43. doi:10.1097/00005373-199202000-00021
10. Morris JA, MacKenzie EJ, Edelstein SL. The effect of preexisting conditions on mortality in trauma patients. JAMA 1990;263 (14): 1942-6. doi:10.1001/jama.1990.03440140068033 PMid:2313871
11. Lefering R, Ruchholtz S. Trauma registries in Europe. Eur J Trauma Emerg Surg 2012;38:1-2. doi:10.1007/s00068-011-0169-3 PMid:26815665
12. Rutledge R. The goals, development, and use of trauma registries and trauma data sources in decision making in injury. Surg Clin N Am 1995;75(2):305-26. doi:10.1016/S0039-6109(16)46590-4 PMid:7900000
13. Moore L, Clark DE. The value of trauma registries. Injury 2008;39 (6):686-95. doi:10.1016/j.injury.2008.02.023 PMid:18511052
14. Stonko DP, Guillamondegui OD, Fischer PE, Dennis BM. Artificial intelligence in trauma systems. Surgery 2021;169(6): 1295-9. doi:10.1016/j.surg.2020.07.038 PMid:32921479
15. Farrow L, Zhong M, Ashcroft GP, Anderson L, Meek RMD. Interpretation and reporting of predictive or diagnostic machine-learning research in Trauma & Orthopaedics. Bone Joint J 2021;103-b(12):1754-8. doi:10.1302/0301-620X.103B12.BJJ-2021-0851.R1 PMid:34847720
16. El Naqa I, Murphy MJ. What is machine learning?: Springer; 2015. doi:10.1007/978-3-319-18305-3_1
17. Santhanam P, Ahima RS. Machine learning and blood pressure. J Clin Hypertens 2019;21(11):1735-7. doi:10.1111/jch.13700 PMid:31536164 PMCid:PMC8030505
18. Simjanoska M, Gjoreski M, Gams M, Madevska Bogdanova A. Non-invasive blood pressure estimation from ECG using machine learning techniques. Sensors 2018;18(4):1160. doi:10.3390/s18041160 PMid:29641430 PMCid:PMC5949031
19. Lindsay GW. Attention in psychology, neuroscience, and machine learning. Front Comput Neurosci 2020;14:29. doi:10.3389/fncom.2020.00029 PMid:32372937 PMCid:PMC7177153
20. Makroum MA, Adda M, Bouzouane A, Ibrahim H. Machine learning and smart devices for diabetes management: Systematic review. Sensors 2022;22(5):1843. doi:10.3390/s22051843 PMid:35270989 PMCid:PMC8915068
21. Li K, Wu H, Pan F, Chen L, Feng C, Liu Y, et al. A machine learning-based model to predict acute traumatic coagulopathy in trauma patients upon emergency hospitalization. Clin Appl Thromb Hemost 2020;26:1076029619897827. doi:10.1177/1076029619897827 PMid:31908189 PMCid:PMC7098202
22. Akbari M, Atoof F, Nazari-Alam A, Fatemi Nasab ZS, Miranzadeh MB, Mirzaei N. Assessment of bacterial bioaerosols and particulate matter characteristics in indoor air of dentistry clinics. Int Arch Health Sci 2023; 10(3): 130-136. doi: 10.48307/iahsj.2023.179403
23. Tang H, Li G, Liu C, Huang D, Zhang X, Qiu Y, et al. Diagnosis of lymph node metastasis in head and neck squamous cell carcinoma using deep learning. Laryngoscope Investig Otolaryngol 2022; 7 (1):161-9. doi:10.1002/lio2.742 PMid:35155794 PMCid:PMC8823170
24. Sarmadi H, Entezami A, Saeedi Razavi B, Yuen KV. Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. Struct Control Health Monit 2021;28 (2): e2663. doi:10.1002/stc.2663
25. Cao L, Li Y, Zhang J, Jiang Y, Han Y, Wei J. Electrical load prediction of healthcare buildings through single and ensemble learning. Energy Rep 2020;6:2751-67. doi:10.1016/j.egyr.2020.10.005 PMCid:PMC7560125
26. Raza K. Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. U-Healthcare Monitoring Systems: Elsevier; 2019. p. 179-96. doi:10.1016/B978-0-12-815370-3.00008-6
27. Kalmady SV, Greiner R, Agrawal R, Shivakumar V, Narayanaswamy JC, Brown MR, et al. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. NPJ Schizophrenia 2019;5(1):1-11. doi:10.1038/s41537-018-0070-8 PMid:30659193 PMCid:PMC6386753
28. Zhang C, Ma Y. Ensemble machine learning: methods and applications: Springer; 2012. doi:10.1007/978-1-4419-9326-7
29. Zhou Z-H. Ensemble methods: foundations and algorithms: CRC press; 2012. doi:10.1201/b12207
30. Sahid MA, Hasan MZ, Akter N, Tareq MMR. Effect of Imbalance Data Handling Techniques to Improve the Accuracy of Heart Disease Prediction using Machine Learning and Deep Learning. IEEE Region 10 Symposium 2022:1-6. doi:10.1109/TENSYMP54529.2022.9864473
31. Ghorpade SJ, Chaudhari RS, Patil SS. Enhancement of Imbalance Data Classification with Boosting Methods: An Experiment. ECS Transactions 2022. doi:10.1149/10701.15923ecst
32. Yao M, Zhu Y, Li J, Wei H, He P. Research on Predicting Line Loss Rate in Low Voltage Distribution Network Based on Gradient Boosting Decision Tree. Energies 2019. doi:10.3390/en12132522
33. Lee M-W, Chen SY, editors. D Decision Tree Applications for Data Modelling2016. doi:10.4018/978-1-59904-849-9.ch067
34. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015;27(2):130-5. doi:10.11919/j.issn.1002-0829.215044.
35. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Prediction of in-hospital mortality in patients with post traumatic brain injury using National Trauma Registry and Machine Learning Approach. Scand J Trauma Resusc Emerg Med 2020;28:1-10. doi:10.1186/s13049-020-00738-5 PMid:32460867 PMCid:PMC7251921
36. 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. J Crit Care 2019; 54:110-6. doi:10.1016/j.jcrc.2019.08.010 PMid:31408805
37. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, Van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95-107. doi:10.1016/j.jclinepi.2020.03.005 PMid:32201256
38. 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. J Med Syst 2015;39:1-10. doi:10.1007/s10916-014-0187-x PMid:25637541

39. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. J Clin Epidemiol 2016;78:83-9. doi:10.1016/j.jclinepi.2016.03.002 PMid:26987507
40. Hertz AM, Hertz NM, Johnsen NV. Identifying bladder rupture following traumatic pelvic fracture: a machine learning approach. Injury 2020;51(2):334-9. doi:10.1016/j.injury.2019.12.009 PMid:31866131
41. Al-Shehari T, Alsowail RA. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. Entropy 2021;23(10):1258. doi:10.3390/e23101258 PMid:34681982 PMCid:PMC8535057
42. Rodríguez P, Bautista MA, Gonzalez J, Escalera S. Beyond one-hot encoding: Lower dimensional target embedding. Image Vis Comput 2018;75:21-31. doi:10.1016/j.imavis.2018.04.004
43. Breiman L. Bagging predictors. Machine Learning 1996;24(2):123-40. doi:10.1007/BF00058655
44. Ensemble methods 2023 [Available from: https://scikit-learn.org/stable/modules/ensemble.html.
45. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Mach Lear Res 2014; 15 (1): 3133-81.
46. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55(1):119-39. doi:10.1006/jcss.1997.1504
47. Caruana R, Niculescu-Mizil A, editors. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning; 2006. doi:10.1145/1143844.1143865
48. David HW. Stacked generalization. Neural Netw 1992;5(2): 241-59. doi:10.1016/S0893-6080(05)80023-1
49. Russell S, Norvig P. Artificial intelligence: A modern approach prentice-hall. Englewood cliffs 1995.
50. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14(1):1-37. doi:10.1007/s10115-007-0114-2
51. Carbonell JG, Michalski RS, Mitchell TM. An overview of machine learning. Mach Lear 1983:3-23. doi:10.1016/B978-0-08-051054-5.50005-4
52. Sperandei S. Understanding logistic regression analysis. Biochem Med 2014;24(1):12-8. doi:10.11613/BM.2014.003 PMid:24627710 PMCid:PMC3936971
53. Yang Y, Wei L, Hu Y, Wu Y, Hu L, Nie S. Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. J Neurosci Methods 2021; 350:109019. doi:10.1016/j.jneumeth.2020.109019 PMid:33321153
54. Chaurasia V, Pal S. Stacking-Based Ensemble Framework and Feature Selection Technique for the Detection of Breast Cancer. SN Comput Sci 2021;2(2):67. doi:10.1007/s42979-021-00465-3
55. Dhanya R, Paul IR, Akula SS, Sivakumar M, Nair JJ. F-test feature selection in Stacking ensemble model for breast cancer prediction. Procedia Comput Sci 2020;171:1561-70. doi:10.1016/j.procs.2020.04.167
56. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992;46(3):175-85. doi:10.1080/00031305.1992.10475879
57. Alpaydin E. Design and analysis of machine learning experiments. 2010. https://ieeexplore.ieee.org/abstract/document/6284951/metrics#metrics
58. Carreira-Perpiñán MÁ, Zharmagambetov A. Ensembles of Bagged TAO Trees Consistently Improve over Random Forests, AdaBoost and Gradient Boosting. Proceedings of the ACM-IMS on Foundations of Data Science Conference 2020. doi:10.1145/3412815.3416882
59. García EM, Alberti MG, Arcos Álvarez AA. Measurement-While-Drilling Based Estimation of Dynamic Penetrometer Values Using Decision Trees and Random Forests. Appl Sci 2022. doi:10.3390/app12094565
60. Powers M. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. J Mach Lear Technol. doi:10.48550/arXiv.2010.16061.
61. Singh A, Mehta JC, Anand D, Nath P, Pandey B, Khamparia A. An intelligent hybrid approach for hepatitis disease diagnosis: Combining enhanced k-means clustering and improved ensemble learning. Expert Syst 2021;38(1):e12526. doi:10.1111/exsy.12526
62. Shafi A, Rahman MB, Anwar T, Halder RS, Kays HE. Classification of brain tumors and auto-immune disease using ensemble learning. Inform Med Unlocked 2021;24:100608. doi:10.1016/j.imu.2021.100608
63. Gelbard RB, Hensman H, Schobel S, Khatri V, Tracy BM, Dente CJ, et al. Random forest modeling can predict infectious complications following trauma laparotomy. J Trauma Acute Care Surg 2019;87(5):1125-32. doi:10.1097/TA.0000000000002486 PMid:31425495

How to Cite this Article:
Kohzadi Z, Nickfarjam AM, Shokrizadeh Arani L, Kohzadi Z, Mahdian M. A comprehensive evaluation of ensemble learning methods and decision trees for predicting trauma patient discharge status using real-world data. Arch Trauma Res. 2023;12(3):137-149 doi:10.48307/ATR.2023.407607.1020