SciencePG
Science Publishing Group

Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study

Tariq Al-Moqri[1, 2], Xiao Haijun[1, *], Jean Pierre Namahoro[1], Eshrak Naji Alfalahi[3], Ibrahim Alwesabi[4]

[1]School of Mathematics and Physics, China University of Geosciences, Wuhan, China

[2]Mathematics Department, Faculty of Applied Sciences, Thamar University, Dhamar, Yemen

[3]Ministry of Public Health and Population Yemen Field Epidemiology Training Program Almaqaleh St, Sana'a, Yemen

[4]School of Automation, China University of Geoscience, Wuhan, China

Email address:
tariq.almoqri@tu.edu.ye (T. Al-Moqri), xiaohj@cug.edu.cn (Xiao Haijun), jpnama1991@yahoo.com (J. P. Namahoro),
e.alfalahi@yahoo.com (E. N. Alfalahi), Ibrahimxiao@gmail.com (I. Alwesabi)
[*]Corresponding author

To cite this article:
Tariq Al-Moqri, Xiao Haijun, Jean Pierre Namahoro, Eshrak Naji Alfalahi, Ibrahim Alwesabi. Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study. *Applied and Computational Mathematics*. Vol. 9, No. 5, 2020, pp. 155-164. doi: 10.11648/j.acm.20200905.12

Abstract: This study focused on exploiting machine learning algorithms for classifying and predicting injury severity of vehicle crashes in Yemen. The primary objective is to assess the contribution of the leading causes of injury severity. The selected machine learning algorithms compared with traditional statistical methods. The filtrated second data collected within two months (August-October 2015) from the two main hospitals included 156 injured patients of vehicle crashes reported from 128 locations. The data classified into three categories of injury severity: Severe, Serious, and Minor. It balanced using a synthetic minority oversampling technique (SMOTE). Multinomial logit model (MNL) compared with five machine learning classifiers: Naïve Bayes (NB), J48 Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). The results showed that most of machine learning-based algorithms performed well in predicting and classifying the severity of the traffic injury. Out of five classifiers, RF is the best classifier with 94.84% of accuracy. The characteristics of road type, total injured person, crash type, road user, transport way to the emergency department (ED), and accident action were the most critical factors in the severity of the traffic injury. Enhancing strategies for using roadway facilities may improve the safety of road users and regulations.

Keywords: Crash Injury Severity, Machine Learning, Traditional Statistical Methods, SMOTE, WEKA

1. Introduction

Vehicles (cars, motorbikes, or bicycles) are globally safety tools for transportation. However, one of the negative results of its use is road traffic accidents (RTAs), which is one of the top ten leading causes of deaths. A global estimate of RTAs deaths is 1.35 million, and between 20 and 50 million suffer non-severe injuries. The most affected are the young pedestrians, cyclists, and motorcyclists [1, 2]. The behavior of riders, drivers, cyclists, and pedestrians are the crucial causes of accidents [3]. Driving/riding and using mobile phones or using drugs/alcohol simultaneously were reported

as major causes of RTAs for young drivers. On the other hand, cognitive, visual, and mobility injury are factors causing accidents for elderly drivers [4]. The rapid economic growth in low and middle-income countries is parallel to a huge number of injuries due to greater accessibility to transportation. Due to the limited technology of vehicle control [5], it leads to the highest rate of RTAs in Africa and the Southern part of Asia [6].

There were many strategies employed to decrease the rate of crash injuries and related consequences of RTAs [1]; some include road safety efforts and targets, problem solving-policies, and monitoring and evaluation [7]. The safety

management system practices linked with incident reduction have been proposed to be reliable and secure solutions that will lead to RTAs and injury prevention [8].

In Yemen, the violated traffic regulations, uncontrolled passenger movements, road conditions are the most probable causes of the accidents. In 2015, Sana'a traffic data showed that the trends of counted deaths per 10,000 vehicles were 27.5, and the injuries and fatalities caused by driving at high speed were 38.1% of total RTAs [9]. Further studies showed that RTAs cause 12 % of all mortality, and the trends of vehicle crash injuries still high [9-12]. After this year, due to the political affairs, the database of the RTAs was unavailable for use.

The causality models based on both statistical and reasonableness in viable decision making and regression models for predicting yearly deaths, fatalities, and injuries were applied [10, 12]. Unfortunately, these models could not show the RTAs trends in the presence of accident causes. Hospital case study shown that the majority of RTAs in Yemen affected the people less than 30 years old [13]. However, comprehensive studies combining road accidents and their relative leading causes are mostly needed in Yemen for boosting the safety and sustainable development of public health.

Recently, machine learning techniques have become modern methods used in transportation safety studies to identify the substantial influences associated with crash injury severity [14, 15]. It is able to show the proportionality of each factor influencing RTAs from each type of vehicle [16].

Other studies also used machine learning techniques to identify the most critical factors that might lead to crash accidents, investigate motorcycle crashes, and detect crash characteristics and influence of vehicle users in crash injuries [17-19]. In advance, these algorithms clarify the complex patterns associated with crash risk [20]. More detail in RTAs data, see [21] in counted data, regression model in prediction [22], artificial neural networks (ANN) [23], support vector machines (SVM) [24] and decision trees (DTs) [20].

Unlike statistical models, machine learning-based algorithms need no assumption for mathematical/statistical models that define the relationship between the dependent and independent variables. Machine learning algorithms can deal with nominal and discrete variables that have over two levels and handle well with multi-collinear explanatory variables [25].

In brief, previous studies partially contributed to the analysis and forecasting RTAs consequences by using the unspecified/counted data and traditional statistical methods.

The contributing factors which effectively influence crash injury severities appeared to be the potential gaps, and applying machine learning-based algorithms became a leading tool to show the relationship between road accidents and associated causes. These algorithms were not used in the studies conducted in Yemen. Unlike traditional methods, these algorithms classified the observed data into several classes for having full access to every scenario that occurred

on the accident scene before and after the accidents, which is the deepest analysis needed in Yemen.

However, this study merely on assessing and analyzing the relative importance of characteristics and behavior of targeted variables in predicting injury severities of vehicle crashes, and then establish the effect rate of factors influencing RTAs. Due to the lack of studies focusing on contributing factors associated with crashes injuries in Yemen motivates this topic to relate with the comparison of selected machine learning algorithms (Support Vector Machines, Naïve Bayes, Random Forest, Multilayer Perceptron, and J48 decision tree) and tradition statistical methods (Multinomial logit model) for predicting crash injury severities.

2. Methods

2.1. Data

Despite several challenges in getting detailed data, we used secondary data collected in two ways. The records of the admitted crash injuries were obtained from emergency departments (EDs) in Sciences and Technology University Hospital (STUH) and Al-Gumhouri General Hospital (AGH) in the capital of Yemen (Sana'a city). The injured people were interviewed by a questionnaire that is designed by [26]. The data was collected between August 24, 2015, and October 8, 2015, and it was published in [13]. The injury severity scoring class is used in three injury levels: minor class noted for simple or no apparent injury, serious class noted for the injury that needs some treatment, and severe that requires intensive medical or surgical management.

2.2. Participants

The second data combining cases from recorded files and questionnaire were 156 (injuries and deaths with 128 males and 28 females). The data enclosed 128 roads and was collected from two main hospitals (STUH and AGH).

2.3. Data Analysis Tools

The data analyzed using the Waikato Environment for Knowledge Analysis (WEKA), which is a group of data mining algorithms implemented using java tools and created by the University of Waikato in New Zealand. This workbench was used in similar studies since it is an open-access software tool containing software that assists in functionalities of data mining such as data preprocessing, visualization, classification, feature selection, regression, clustering, and association rules [27]. WEKA workbench provides over 100 classification algorithms, 75 algorithms for preprocessing data, 25 algorithms for dimensionality reduction and evaluation metrics, and 20 algorithms for clustering methods. WEKA toolkit version 3.9 was freely downloaded and installed on a Lenovo IdeaPad S410p computer equipped with 4 GB RAM, 2.23 GHz, a 64-bit operating system, and an Intel (R) Core (TM) i5- 4200U CPU to perform the experiments.

2.4. Preprocessing of Dataset

In this study, first, we cleaned the data and transformed the continuous attributes into categorical attributes. The entire dataset contains 24 attributes that converted from numeric to nominal attributes. Those attributes included accident time, road types, road users, road conditions, vehicle types, and organs injured for victims. The response variable is multiclass, the minor injury takes the value (1), the serious injury takes the value (2), and the severe injury takes the value (3).

2.4.1. Balancing Data

Most road accident data are imbalanced data because of the minor injury class has more instances than serious or severe injury class. In our data, the minor class has 95 injured cases, whereas the serious and severe injury classes,

respectively, contain 35 and 26 cases. When the training dataset is excessively imbalanced, the predicted of the minority class will not detect the truth information and will affect the prediction accuracy [28]. In this study, a synthetic minority oversampling technique (SMOTE) [29] was used for balancing the data.

2.4.2. Attribute Selection

After the data became balanced, we applied CorrelationAttributeEvel tools in WEKA for selecting the most correlated attributes with class data. Among 24 categorical attributes, seven attributes were removed. The removed attributes are education status, occupation, accident time, vehicle type, cuts bites or open wound, sex, and road conditions. The concluding list of the attributes and their descriptions are presented in Table 1.

Table 1. *Descriptive of variables.*

	variables	Total (%)	Minor (%)	Serious (%)	Severe (%)
Hospital name	Science & Technology Hospital Al-	66.7	53	77.1	100
	Gomhouri General Hospital	33.3	46.3	22.9	0
Age	1: <25	59.6	61.1	48.6	69.2
	2: 25- 50	33.3	31.6	42.9	26.9
	3: Elder than 50	7.1	7.4	8.6	3.8
Sex	1: Male	82.1	84.2	80	76.9
	2: Female	17.9	15.8	20	23.1
Occupation	1: Public employ	17.3	18.9	20	7.7
	2: Private employ	7.1	9.5	5.7	0
	3: Business own	5.1	5.3	0	11.5
	4: Farmer	5.8	2.1	11.4	11.5
	5: Daily worker	12.8	13.7	14.3	7.7
	6: Student	44.2	44.2	37.1	63.8
	7: Other/unspecified	7.7	6.3	11.4	7.7
Educational Status	1: Illiterate	7.7	6.3	5.7	15.4
	2: Basic school in	31.4	27.4	37.1	38.5
	3: Secondary education	18.6	17.9	20	19.2
	4: University education	27.6	35.8	20	7.7
	5: Under university school-age	14.7	12.6	17.1	19.2
Accident time	$1: 00:00 -< 06:00$	1.3	2.1	0	0
	$2: 06:00 -< 12:00$	39.1	42.1	37.1	30.8
	$3: 12:00 -< 18:00$	48.1	45.3	48.6	57.7
	$4: 18:00 -< 00:00$	11.5	10.5	14.3	11.5
Accident total injured persons	1: One	55.1	48.4	62.9	69.2
	2: Two	13.5	17.9	2.9	11.5
	3: More than two	31.4	33.7	34.3	19.2
Transportation way to the hospital	1: Ambulance	12.8	8.4	11.4	30.8
	2: Tax	70.5	76.8	71.4	46.2
	3: Walking	1.9	3.2	0	0
	4: Private care \ Heavy vehicle	14.7	11.6	17.1	23.1
Road type (Infrastructure cause)	1: Urban highway	54.5	61.1	51.4	34.6
	2: Rural highway	28.2	22.1	28.6	50
	3: Urban subway	10.3	12.6	2.9	11.5
	4: Rural Subway	7.1	4.2	17.1	3.8
Road user	1: Vehicle's driver	7.7	8.4	2.9	11.5
	2: Car's driver	21.2	21.1	22.9	19.2
	3: Motor's driver	16.7	16.8	22.9	7.7
	4: Motor's passengers	13.5	20	2.9	3.8

	variables	Total (%)	Minor (%)	Serious (%)	Severe (%)
	5: Heavy truck's Driver/Passengers	0.6	0	0	3.8
	6: Bus's driver/passenger	4.5	4.2	5.7	3.8
	7: Cyclists	3.8	4.2	0	7.7
	8: Pedestrians	30.1	23.2	40	42.3
	9: Other/unspecified	1.9	2.1	2.9	0
Road conditions (E2)	1: Paved	91	92.6	85.7	92.3
	2: Unpaved	9	7.4	14.3	7.7
	1: Motorbike	37.2	42.1	31.4	26.9
Vehicle type	2: Car	45.5	43.2	48.6	50
	3: Bus	8.3	7.4	14.3	3.8
	4: Heavy vehicle	9	7.4	5.7	19.2
	1: Vehicle-vehicle	30.1	35.8	28.6	11.5
Crash type	2: Vehicle-person	36.5	28.4	45.7	53.8
	3: Vehicle- object	33.3	35.8	25.7	34.6
	1: Crossing the road	11.5	5.3	20	23.1
	2: Walking on the sideway	14.1	11.6	20	15.4
Accident action (Environmental condition-3)	3: Getting on/off a public transportation	3.8	4.2	5.7	0
	4: Getting on/off a private transportation	0.6	1.1	0	0
	5: Driving/passenger a 4 wheel vehicle	28.8	28.4	25.7	34.6
	6: Riding/passenger a 2 wheel vehicle	31.4	38.9	25.7	11.5
	7: Other	9.6	10.5	2.9	15.4
Head	1: Not injured	53.8	64.2	54.3	15.4
	2: Injured	46.2	35.8	45.7	84.6
Limbs	1: Not injured	35.3	30.5	25.7	65.4
	2: Injured	64.7	69.5	74.3	34.6
Trunk	1: Not injured	73.7	73.7	88.6	53.8
	2: Injured	26.3	26.3	11.4	46.2
Fracture	1: No	67.9	93.7	17.1	42.3
	2: Yes	32.1	6.3	82.9	57.7
Sprain strain or dislocation	1: No	90.4	89.5	94.3	88.5
	2: Yes	6.9	10.5	5.7	11.5
Cuts bites or open wound	1: No	78.8	81.1	80	69.2
	2: Yes	21.2	18.9	20	30.8
Bruise or super facial injury	1: No	26.3	12.6	34.3	65.4
	2: Yes	73.7	87.4	65.7	34.6
Concussion	1: No	93.6	98.9	97.1	69,2
	2: Yes	6.4	1.1	2.9	30.8
Organ system injury	1: No	84	100	85.7	23.1
	2: Yes	16	0	14.3	76.9
	1: Treated	59	91.6	14.3	0
	2: Admitted to ER	32.1	6.3	74.3	69.2
injured status in ED	3: Referred to another hospital	7.1	1.1	8.6	26.9
	4: Died	0.6	0	0	3.8
	5: Other	1.3	1.1	2.9	0

2.5. Machine Learning Classifiers

In this study, the target variable is crash injury severities containing one of three possible outcomes (severe, serious, and minor). The most suitable function of machine learning is classification. The supervised learning algorithms contain classifiers that are able to utilize and classify datasets and provide interesting results. Classification methods are predictive techniques used to forecast classes of a target variable from measurements of one or more attributes. The classification step is processed into three steps: Input has a defined set of known explanatory variables, classifier to forecast the explanatory variables whose value is unknown, and last, the output gives unknown explanatory variables determined by other known explanatory variables because of using a classification algorithm [30].

Several classifiers can handle various classification problems in WEKA according to their categories as divided into sub-packages (Bayesian classifiers, lazy classifiers, decision trees classifiers, functions algorithms, meta-learning algorithms, rule-based algorithms, and miscellaneous) [31]. WEKA was used in this study

because it is an open-access software freely available under the general public license. It is a Java programming language compatible with much modern computing platform and comprises a complete set of data preprocessing and modeling methods. In the following sub-sections, a short explanation of the some selected classification algorithms used in this work is presented.

2.5.1. J48 Decision Tree

J48 decision tree classifier is an open-source of the C4.5 algorithm implemented in WEKA that generates a decision tree through information entropy. This method uses a famous approach known as divide-and-conquer to solve the learning problem from a group of independent instances. It uses an endogenous attribute to choose a target value of a new instance. In this process, the interior node of the decision tree represents the unique attributes. The branches between nodes illustrate the possible values that the attributes can have in the observed instances, and the terminal nodes represent the final node of the class [32].

2.5.2. Random Forest (RF)

The RF approach uses the decision tree algorithm for parameterization and mixing a sampling procedure, subspace technique, and an ensemble strategy to optimize the model building. RF principle is to aggregate many binary decision trees as following: the use of bootstrap samples (obtained by randomly selecting n observations with replacement from the learning set L) instead of the whole sample L and the construction of a randomized tree predictor instead of CART on each bootstrap sample [33]. In classification problems, it selects a majority vote among all individual tree predictions and predicts new instances to the majority vote class.

2.5.3. Naïve Bayes (NB)

NB is a robust learning algorithm for classification in WEKA, based on Bayes' rule with a strong assumption that the attributes are conditionally independent in a given class. The advantages of naïve Bayes classification are often to deliver a competitive classification accuracy, computational efficiency, and many other desirable features, including multi-classification [34].

2.5.4. Support Vector Machine (SVM)

SVM is a powerful binary classification method that primarily performs classification step by constructing hyper-planes in a high or multidimensional space that separates the instances that are belonging to different class labels [35]. To achieve the classification target for multiclass by applying SVM, three popular methods (one against all, one against one, and directed acyclic graph) used, and due to classification accuracy, directed acyclic graph-SVM marked to be more robust [36]. In this study, SVM implemented using the SMO algorithm in WEKA software that uses one against one method.

2.5.5. Multilayer Perceptron (MLP)

MLP is a feed-forward neural network classifier with more than one hidden layer. It uses linear or nonlinear activation functions for computing the weighted sum of input and biases, that is used to decide if a neuron can be fired or not [37]. This algorithm involves an iterative procedure for minimization of the error of the function for achieving good prediction with adjustments of weights [38].

2.5.6. Multinomial Logit Model (MNL)

The MNL model is an upgraded version of binary logit regression, and both are traditional statistical methods applied to predict the probability of class association on a predicted variable depends on numerous predictor variables. The desirable variable in inquiry is nominal and for which there are over two categories, while the predictor variables can be dichotomous or continuous. MNL is famous in multi-classification since its ability to tolerate two or more categories of the outcome variables and applies maximum likelihood estimation to assess the probability of categorical membership. MNL relies on independence normality, and multicollinearity assumptions [39].

2.6. Parameter Setting

The data were balanced in WEKA using SMOTE oversampling technique. After the data became balanced, randomly split into train data (70%) and test data (30%). The parameters are shown in Table 2. After selecting the best parameters, we applied the selected machine learning algorithms for the classification step.

Table 2. Parameter settings.

Algorithm parameter settings	
NB	Weka parameter default values (W. P. D. V) [40]
MLP	L= 0.2; M = 0.1; W.P.D.V
SMO	Built calibration models = True;
	Calibrator LibSVM; W.P.D.V
J48	BinarySplit = True; W.P.D.V
RF	K = 5; I = 200; W.P.D.V
MNL	ridge = 1.0E-12; W.P.D.V

2.7. Evaluation

For evaluating the classifier performance, the following evaluation metrics are used: confusion matrix, accuracy, precision, recall, F-measure, and Kappa statistics. The confusion matrix shows the well predicted and misclassified instances. The accuracy measures the ratio of correctly predicted instances over the total number of the whole dataset. Recall calculates the rate of the instances that are predicted correctly positive. Precision measures the instances that are predicted correctly positive from the total predicted instances in the positive class, F-measure measures the harmonic mean between the recall value and precision value [41]. Kappa statistic shows the goodness of the observed agreement (P_o) in the classifier over the predicted agreement (P_e) that is predicted by chance [42].

Table 3. *Binary confusion matrix.*

		Predicted class	
		Yes	No
Actual class	Yes	True positive (TP)	False Negative (FN)
	No	False-positive (FP)	True Negative (TN)

The columns in the Table 3 denotes the predicted class instances, rows indicate the actual class instances, and diagonal elements represent the accurate prediction. Thus, the performance of a classifier can be visualized in the confusion matrix [32]. This confusion matrix can be generalized to be used for multiclass problems.

According to [27], True positives (TP) and true negatives (TN) are correctly classified. A false positive (FP) is noted for the cases that are wrongly classified as "Yes" and a false negative (FN) is noted for the cases that are improperly classified as "No". TPR measures the rate of the cases that are correctly identified, whereas FPR measures the rate of the cases that are incorrectly classified. The evaluation metrics can be given as follows:

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

$$FPR = \frac{FP}{FP+TN} \tag{2}$$

$$Precision(P) = \frac{TP}{TP+FP} \tag{3}$$

$$Recall(R) = \frac{TP}{TP+TN} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$F-measure = \frac{2*P*R}{P+R} \tag{6}$$

$$Kappa\ St = \frac{P_o - P_e}{1 - P_e} \tag{7}$$

3. Experiment and Analysis

In this study, machine learning classifiers and MNL were applied to model crash injury severities. The injury severity attribute is the class attribute and takes three values as target values (Minor, Serious, and Severe). Distributing values in the dataset are presented in Table 1. The numbers (1, 2 …10) present values assigned to each variable; rows represent attributes, and columns represent the classes of these attributes. Some attributes were environmental leading causes (E1-E3): accident action, road conditions and accident time, and infrastructure leading causes as characteristics of the road type. In total, 156 crash records were reported between August 24, 2015, and October 8, 2015.

Table 4. *Expermintal confusion matrix.*

Classifier		NB			SVM			MLP			J48			RF			MNL		
Class	Predicted class	M	S	D	M	S	D	M	S	D	M	S	D	M	S	D	M	S	D
Actual class	M	21	1	0	21	1	0	21	1	0	21	1	0	22	0	0	20	2	0
	S	3	25	4	2	30	0	4	27	1	4	27	1	2	28	2	4	27	1
	D	1	4	26	1	4	26	1	2	28	0	1	30	1	1	29	1	3	27

Table 5. *Classifiers evaluation metrics.*

Classifier	TPR	FPR	Precision	Recall	F-Measure	Kappa Statistic	Accuracy (%)
NB	0.847	0.079	0.847	0.847	0.846	0.769	84.7
SVM	0.906	0.048	0.914	0.906	0.906	0.857	90.59
MLP	0.894	0.049	0.900	0.894	0.895	0.840	89.41
J84	0.918	0.037	0.921	0.918	0.917	0.875	91.76
RF	0.948	0.025	0.949	0.948	0.948	0.922	94.84
MNL	0.871	0.063	0.876	0.871	0.872	0.823	87.05

Table 6. *Classifiers evaluation metrics by class.*

Classifier	Class	TP Rate	FP Rate	Precision	Recall	F-Measure
NB	Minor	0.840	0.063	0.955	0.894	0.894
	Serious	0.833	0.094	0.781	0.806	0.806
	Severe	0.867	0.074	0.839	0.852	0.852
SVM	Minor	0.875	0.048	0.955	0.913	0.913
	Serious	0.857	0.094	0.938	0.896	0.896
	Severe	1	0.000	0.839	0.912	0.912
MLP	Minor	0.808	0.079	0.955	0.875	0.875
	Serious	0.900	0.057	0.844	0.871	0.871
	Severe	0.966	0.019	0.903	0.933	0.933
J48	Minor	0.840	0.063	0.955	0.894	0.894
	Serious	0.931	0.038	0.844	0.885	0.885
	Severe	0.968	0.019	0.968	0.968	0.968
RF	Minor	0.926	0.028	0.962	0.943	0.943
	Serious	0.941	0.032	0.941	0.928	0.928

Classifier	Class	TP Rate	FP Rate	Precision	Recall	F-Measure
	Severe	0.972	0.016	0.972	0.972	0.972
	Minor	0.800	0.079	0.909	0.851	0.851
MNL	Serious	0.844	0.094	0.844	0.844	0.844
	Severe	0.876	0.063	0.871	0.872	0.872

104 cases were from the Emergency department of Science and Technology University Hospital (49% minor, 26% serious, and 25% severe sub-classes respectively), and 52 collected from Al-Gomhouri General Hospital (84.6% minor, 15.4% serious and 0% severe sub-classes respectively). Approximately 60.9%, 22.4%, and 16.7% were classified into minor, serious, and severe classes respectively. Min-age=1, average=23, and Max-age=65, students=61 (nursery and school pupils), University students=10, workers=47 (public and daily workers), private employees = 34, and unemployed=10, drivers=40, passengers=59, and pedestrians=50. After the preprocessing step, the dataset was loaded as an Attribute Relation File Format (ARFF) file into WEKA tools. Seventeen predictor factors (attributes) were applied with the class variable to create models for forecasting the level of crash injury severities. Table 4 presents the confusion matrix for each classifier where M, S, and D are respectively noted for minor, serious, and severe injury. Performance metrics of all the five model types, specifically, it shows the confusion matrix, TPR, FPR, precision, recall, F-measure, Kappa statistics, and classification accuracy obtained using a randomly splitting method for each of the five classifiers. For each class, the confusion matrix tells how instances from that class recognized the classifications used in this study. All correctly classified are in the diagonal of the contingency table. Therefore, it is possible to examine the matrix for errors visually. Table 4 indicates the confusion matrix of predicted class with all correctly classified are in the diagonal. Table 5 showed the performance metrics (TPR, FPR, Precision, recall, and accuracy) for each classifier. Table 5 indicates the performance metrics by class for each class. The NB classifier was the worst classifier that achieved an accuracy of 84.7%, with a precision of 0.84, 0.833, and 0.867 for minor, serious, and severe, respectively. For the SVM, the accuracy achieved was 90.59%, with a precision of 0.875, 0.857, and 1 for minor, serious, and severe, respectively. For MLP, the accuracy was 89.41%, with a precision of 0.808, 0.900, and 0.966 for minor, serious, and severe, respectively. For J48 classifier achieved an accuracy of 91.76%, with the precision of 0.840, 0.931, and 0.968 for minor, serious and severe, respectively. The best classifier that achieved high performance is RF with an accuracy of 94.84%, with a precision of 0.926, 0.941, and 0.972 for minor, serious, and severe, respectively. MNL classifier achieved an accuracy of 87.05%, with a precision of 0.909, 0.844, and 0.871 for minor, serious, and severe, respectively. This table shows the comparison amongst classes for each classifier through TP and FP rates, precision, recall, and F-measure.

Several features of visualizations of the threshold curves are presented in Tables 5 and 6, and Figures 1 and 2. These tables

and figures indicate that for each classifier, the accuracy, recall, F-measure, and Kappa statistics were significantly higher than 0.7, and some fall in the acceptable range of substantial (0.75-0.85), and another fall in perfect agreement ranges (0.857-1.0). However, the values of these statistical tests indicate that most of the corresponding classifiers have a greater ability and useful to classify crash injury severities correctly than traditional statistical methods represented by MNL in this study. Additionally, four machine learning techniques are better in predicting all injury severity classes. Results indicate that RF was the most accurate classifier with the highest TPR, precision, and recall, and lowest FPR.
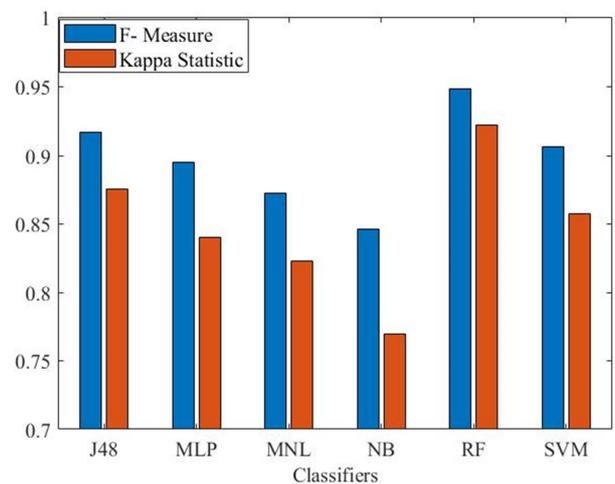

Figure 1. *Classifiers comparison based on F-measure and Kappa statistics.*
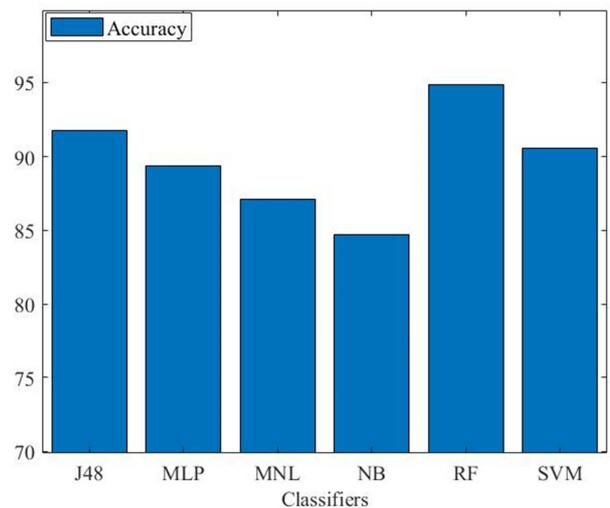

Figure 2. *Classifiers comparison based on the accuracy.*

The second was j48 and then followed by SVM and MLP with closely in the accuracy but differ with FPR, with slightly better performance metrics than the traditional

method MNL and NB. RF performed well in other studies that obtained by [27, 14], but it slightly differs from those studies by achieving higher accuracy. The goodness of the accuracy that we got might be achieved well due to the resampling SMOTE method.

For checking the performance of RF against the other classifiers based on weighted average F- Measure, paired t-test has proved reliable for comparing machine learning algorithms in related studies [43, 44], and it used for the same purpose in this study. After checking the normality of the data, a paired t-test is used for the mean comparison of the weighted average F- Measure for RF and the other proposed classifiers. The results are shown in Table 7, and it indicates that RF performed better than the other classifiers, especially NB and MNL.

Table 7. Mean comparison results.

Pair	Mean difference	t-Statistic	Sig
RF vs NB	0.0467	5.764	<0.0001
RF vs SVM	0.0199	2.645	0.027
RF vs MLP	0.0223	3.631	0.005
RF vs J48	0.0259	2.967	0.016
RF vs MNL	0.0664	3.836	0.004

Root mean square error (RMSE) also can evaluate the model and show the error of the classification process and the misclassification cases [45]. The classification process was repeated ten times, as shown in Figure 3. It indicates that RF has less RMSE than the other classifiers and is more stable than the other classifiers. Table 8 shows the importance of the attributes that contribute to determining the relation between the class (injury severity scoring) and the independent attributes. It was obtained in WEKA– RF tree-compute Attribute Importance (set True). The remaining seven attributes were deleted in the feature selection step.
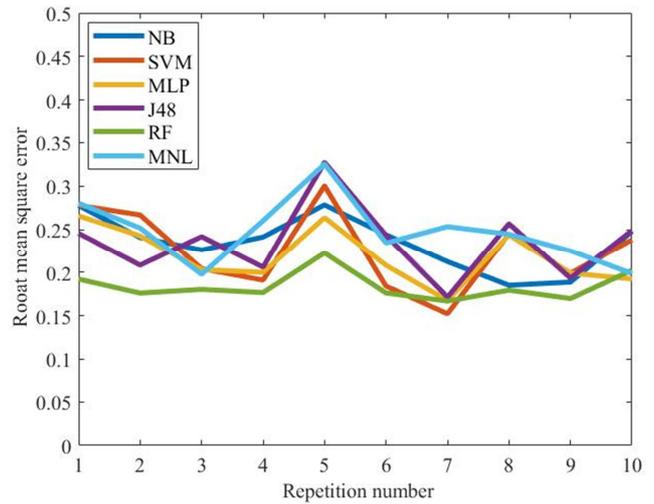

Figure 3. Root mean square error of the classifiers.

Table 8. The importance of the attributes.

Injury leading causes	Variable	Road type	Total injured persons	Crash type	Road user	Transport way to ED	Accident action
	Rank	0.56	0.49	0.48	0.47	0.47	0.47
Injury-related factors	Variable	Fracture	injured status in ED	Concussion	Organ system injury	Bruise or superficial injury	Age
	Rank	0.6	0.57	0.55	0.54	0.51	0.5
	Variable	Trunk	Hospital name	Limps	Head	Sprain strain or dislocation	
	Rank	0.49	0.47	0.46	0.41	0.31	

According to the contributing factors, this study shows that road type, crash type, road user, accident action, characteristics of road type, and collision partners had a similar impact on road accidents in different vehicles in other countries [27, 46] The predicted leadings causes (environmental and infrastructure causes): the behavior of road users (cyclist and pedestrians) identified as common causes, the activity during an accident, characteristics of road-type, ways of transport crash injuries to the hospital, collision type, vehicle type (motors are exposing their users to high risks of crash injuries and fatalities) and road shoulder condition; detected to have a significant impact in vehicle crash injury severity. Some of these predicted environmental and infrastructure leading causes, and others are similar to those predicted in the cyclist crash severity in Spain [46]. The factors that are related to the injured situation have an impact on injury severity scoring, and it can be recognized from Tables 1 and 8.

This study has several limitations. There was a lack of access to the road traffic police database due to the political situation in Yemen. The previous studies only used the counted mortalities data published by the Central Statistical Organization in Yemen (Yearly statistical books: 1991–2013), but RTAs causes were not identified in these books, and its data were not suitable for this study. Regarding the data collection procedures and under-conditions of participants, some environmental causes (weather conditions), visibilities of the streets, violating road safety regulations (speed, over-weights), and other clarification of streets (junctions, crosses) which probably lead to the vehicle crashes were not reported.

Regardless of these challenges, some strategies are recommended to diminish the severity of injuries in vehicle crashes in Yemen. These safety strategies include the use of roadway facilities such as road signage and speed bump at junctions, implementation of laws on red-light violations and speed limit, and road safety and behavior management of road users in either urban or rural villages. These are essential strategies for improving and achieving sustainable development goals in declining mortality. Developing visibility on the roadway, especially with street lighting and visible road allocation, also can be applied. Attention must be taken to the behavior of road users as established in some studies [47-49].

4. Conclusions

This study compared machine learning algorithms and traditional statistical methods for classifying and predicting injury severity of vehicle crashes. Based on performance and errors as revealed by evaluators, machine learning algorithms were substantially classified and predicted vehicle crash injury severity better than traditional methods. Most crash injury severity data are imbalanced, and the severe injury level has the least instances comparing with the remaining levels. The oversampling method SMOTE used for balancing crash injury severity data, and it helped in improving the classification accuracy. RF was the best classifier that got high accuracy and low misclassification cases. The determined effect rate of relative factors (environmental and infrastructural leading causes) that influence injury severity in the vehicle crash shows that awareness and policymakers should improve its conditions.

Funding

Conflicts of Interest

The authors declare that they have no competing interest.

Acknowledgements

References

[1] W. H. Organization, "Global status report on road safety 2018: Summary," World Health Organization 2018.

[2] N. Dhakal, "Using Naturalistic Data for Bicycle Safety Analysis: An Application of Cycle Philly Data to Assess Wrong-Way Riding," 2017.

[3] S. P. Wall et al., "The effect of sharrows, painted bicycle lanes and physically protected paths on the severity of bicycle injuries caused by motor vehicles," Safety, vol. 2, no. 4, p. 26, 2016.

[4] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," Accident Analysis & Prevention, vol. 115, pp. 11-24, 2018.

[5] D. Khorasani-Zavareh and H. Sadeghi-Bazargani, "Epidemiological pattern of motorcycle injuries with focus on riding purpose: Experience from a middle-income country," Journal of Research in Clinical Medicine, vol. 3, no. 3, pp. 149-156, 2015.

[6] V. M. Nantulya and M. R. Reich, "The neglected epidemic: road traffic injuries in developing countries," Bmj, vol. 324, no. 7346, pp. 1139-1141, 2002.

[7] M. Peden et al., "World report on road traffic injury prevention. World Health Organization," in Hyder AA, Jarawan E, Mathers C, 2004: Citeseer.

[8] J. K. Wachter and P. L. Yorio, "A system of safety management practices and worker engagement for reducing and preventing accidents: An empirical and theoretical investigation," Accident Analysis & Prevention, vol. 68, pp. 117-130, 2014.

[9] A. A. Al-Thaifani, N. A. Al-Rabeei, and A. M. Dallak, "Study of the injured persons and the injury pattern in road traffic accident in Sana'a city, Yemen," Advances in Public Health, vol. 2016, 2016.

[10] J. R. Ameen and J. A. Naji, "Causal models for road accident fatalities in Yemen," Accident Analysis & Prevention, vol. 33, no. 4, pp. 547-561, 2001.

[11] F. M. Karim, "Road traffic accidents in Yemen," International journal of injury control and safety promotion, vol. 15, no. 3, pp. 165-166, 2008.

[12] F. M. Karim, A. A. Saleh, A. Taijoobux, and M. Ševrović, "Developing a Model for Forecasting Road Traffic Accident (RTA) Fatalities in Yemen," Slovak Journal of Civil Engineering, vol. 25, no. 4, pp. 12-18, 2017.

[13] E. Alfalahi, A. Assabri, and Y. Khader, "Pattern of road traffic injuries in Yemen: a hospital-based study," Pan African medical journal, vol. 29, no. 1, pp. 1-9, 2018.

[14] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," Accident Analysis & Prevention, vol. 108, pp. 27-36, 2017.

[15] R. E. Al Mamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of machine learning algorithms for predicting traffic accident severity," in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, pp. 272-276: IEEE.

[16] M. B. Anvari, A. T. Kashani, and R. Rabieyan, "Identifying the most important factors in the at-fault probability of motorcyclists by data mining, based on classification tree models," International Journal of Civil Engineering, vol. 15, no. 4, pp. 653-662, 2017.

[17] H. Jiang, Y. Zou, S. Zhang, J. Tang, and Y. Wang, "Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model," Mathematical Problems in Engineering, vol. 2016, 2016.

[18] S. Kumar and D. Toshniwal, "Severity analysis of powered two wheeler traffic accidents in Uttarakhand, India," European transport research review, vol. 9, no. 2, p. 24, 2017.

[19] A. Montella, M. Aria, A. D'Ambrosio, and F. Mauriello, "Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery," Accident Analysis & Prevention, vol. 49, pp. 58-72, 2012.

[20] L.-Y. Chang and J.-T. Chien, "Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model," Safety science, vol. 51, no. 1, pp. 17-22, 2013.

[21] Z. Halim, R. Kalsoom, S. Bashir, and G. Abbas, "Artificial intelligence techniques for driving safety and vehicle crash prediction," Artificial Intelligence Review, vol. 46, no. 3, pp. 351-387, 2016.

[22] D. W. Kononen, C. A. Flannagan, and S. C. Wang, "Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes," Accident Analysis & Prevention, vol. 43, no. 1, pp. 112-122, 2011.

[23] F. R. Moghaddam, S. Afandizadeh, and M. Ziyadi, "Prediction of accident severity using artificial neural networks," International Journal of Civil Engineering, vol. 9, no. 1, p. 41, 2011.

[24] I. Aghayan, M. H. Hosseinlou, and M. M. Kunt, "Application of support vector machine for crash injury severity prediction: A model comparison approach," Journal of Civil Engineering and Urbanism, vol. 5, no. 5, pp. 193-199, 2015.

[25] S. H.-A. Hashmienejad and S. M. H. Hasheminejad, "Traffic accident severity prediction using a novel multi-objective genetic algorithm," International journal of crashworthiness, vol. 22, no. 4, pp. 425-440, 2017.

[26] W. H. Organization, Injury surveillance guidelines (no. WHO/NMH/VIP/01.02). World Health Organization, 2001.

[27] L. Wahab and H. Jiang, "A comparative study on machine learning based algorithms for prediction of motorcycle crash severity," PLoS one, vol. 14, no. 4, p. e0214966, 2019.

[28] T. Cai, H. He, and W. Zhang, "Breast cancer diagnosis using imbalanced learning and ensemble method," Applied and Computational Mathematics, vol. 7, no. 3, pp. 146-154, 2018.

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.

[30] C. Sugetha, L. Karunya, E. Prabhavathi, and P. K. Sujatha, "Performance evaluation of classifiers for analysis of road accidents," in 2017 Ninth International Conference on Advanced Computing (ICoAC), 2017, pp. 365-368: IEEE.

[31] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[32] J. R. Quinlan, C4. 5: programs for machine learning. Elsevier, 2014.

[33] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix, "Random forests for big data," Big Data Research, vol. 9, pp. 28-46, 2017.

[34] J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, "Naive bayes classification of uncertain data," in 2009 Ninth IEEE International Conference on Data Mining, 2009, pp. 944-949: IEEE.

[35] M. Stitson, J. Weston, A. Gammerman, V. Vovk, and V. Vapnik, "Theory of support vector machines," University of London, vol. 117, no. 827, pp. 188-191, 1996.

[36] B. Kijsirikul and N. Ussivakul, "Multiclass support vector machines using adaptive directed acyclic graph," in Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290), 2002, vol. 1, pp. 980-985: IEEE.

[37] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," arXiv preprint arXiv: 1811.03378, 2018.

[38] A. M. Kalteh and R. Berndtsson, "Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP)," Hydrological sciences journal, vol. 52, no. 2, pp. 305-317, 2007.

[39] J. S. Long and J. Freese, Regression models for categorical dependent variables using Stata. Stata press, 2006.

[40] L. Kotthoff, C. Thornton, and F. Hutter, "User guide for auto-WEKA version 2.6," Dept. Comput. Sci., Univ. British Columbia, BETA lab, Vancouver, BC, Canada, Tech. Rep, vol. 2, 2017.

[41] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," International Journal of Data Mining & Knowledge Management Process, vol. 5, no. 2, p. 1, 2015.

[42] B. D. Eugenio and M. Glass, "The kappa statistic: A second look," Computational linguistics, vol. 30, no. 1, pp. 95-101, 2004.

[43] P. Hajek and R. Henriques, "Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods," Knowledge-Based Systems, vol. 128, pp. 139-152, 2017.

[44] J. Sun, H. Fujita, P. Chen, and H. Li, "Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble," Knowledge-Based Systems, vol. 120, pp. 4-14, 2017.

[45] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," Geoscientific model development, vol. 7, no. 3, pp. 1247-1250, 2014.

[46] R. Aldred, S. García-Herrero, E. Anaya, S. Herrera, and M. Á. Mariscal, "Cyclist injury severity in Spain: a Bayesian analysis of police road injury data focusing on involved vehicles and route environment," International journal of environmental research and public health, vol. 17, no. 1, p. 96, 2020.

[47] Y. Zheng, Y. Ma, N. Li, and J. Cheng, "Personality and behavioral predictors of cyclist involvement in crash-related conditions," International journal of environmental research and public health, vol. 16, no. 24, p. 4881, 2019.

[48] V. P. Nyakyi, "Modelling Assessment on Causes of Road Accidents Along Kilimanjaro-Arusha Highway in Tanzania," Applied and Computational Mathematics, vol. 7, no. 2, pp. 71-74, 2018.

[49] C. E. Anderson, A. Zimmerman, S. Lewis, J. Marmion, and J. Gustat, "Patterns of cyclist and pedestrian street crossing behavior and safety on an urban greenway," International journal of environmental research and public health, vol. 16, no. 2, p. 201, 2019.