Firat Univ Jour. of Exp. and Comp. Eng., 2(3), 124-133, 2023

Firat University Journal of Experimental and Computational Engineering
Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

Testing The Performance of Random Forest and Support Vector Machine Algorithms for Predicting Cyclist Casualty Severity

Bisikletli Yaralanma Derecesini Tahmin Etmek için Kullanılan Random Forest ve Support Vector Machine Algoritmalarının Performanslarının Test Edilmesi

Nurten AKGÜN[1]*

[1]Civil Engineering, Engineering and Natural Sciences, Bursa Technical University, Bursa, Turkey.
[1]nurten.akgun@btu.edu.tr

Received: 03.08.2023
Accepted: 26.09.2023

Revision: 10.09. 2023

doi: 10.5505/fujece.2023.57966
Research Article

Citation: Akgün N. “Testing the performance of random forest and support vector machine algorithms for predicting cyclist casualty severity”. *Firat University Journal of Experimental and Computational Engineering*, 2(3), 124-133, 2023.

Abstract
Traditional statistical regression models for predicting casualty severity have fundamental limitations. Machine learning algorithms for classifications have started to be applied in severity analysis in order to relax the assumptions and provide better accuracy in the models. However, the performances of highly advised classification algorithms for predicting cyclist casualty severity, which particularly occurred at roundabouts, have not been investigated comprehensively. Therefore, the study in this paper developed classification models for cyclist casualty severity prediction by applying the highest two advised algorithms in the literature namely Random Forest and Support Vector Machine. The dataset included 439 cyclist casualties which were recorded at give-way roundabouts in the North East of England. The predictive variables were sociodemographic information about cyclists, weather conditions, behavior-related contributory factors, speed limit, and roundabout geometrical parameters. 70% of the records were randomly selected for the training stage and 30% were used for the testing in both Random Forest and Support Vector Machine algorithms. After training the algorithm, the testing results showed that the Random Forest algorithm predicted the outcomes with 88.6% classification accuracy. On the other hand, Support Vector Machine algorithm predicted the testing values with 84.73% classification accuracy. The algorithms misestimated 18 and 20 of the casualties in Random Forest and Support Vector Machine, respectively. The outcomes suggested that both Random Forest and Support Vector Machine algorithms were applicable for cyclist casualty severity prediction models with high performance.

Keywords: Machine learning, Cyclist safety, Geometry, Roundabout, Classification accuracy

Özet
Kaza yaralanma derecesinin tahmininde kullanılan geleneksel istatistiksel regresyon modellerinin birtakım kısıtlamaları vardır. Varsayımları ortadan kaldırmak ve modellerde daha iyi doğruluk sağlamak amacıyla makine öğrenmesi tekniğinin sınıflandırma algoritmaları yaralanma derecesi analizinde uygulanmaya başlanmıştır. Ancak, uygulanması tavsiye edilen sınıflandırma algoritmalarının performansları, özellikle dönel kavşaklarda meydana gelen bisikletli kazalarının yaralanma derecesini tahmin etmek için, kapsamlı bir şekilde araştırılmamıştır. Bu sebeple, bu makaledeki çalışmada, literatürde en sık önerilen iki algoritma olan Random Forest ve Support Vector Machine'i bisikletli yaralanma derecesinin tahmininde kullanarak sınıflandırma modelleri geliştirmiştir. Veri seti, İngiltere'nin kuzey-doğu bölgesinde karma kullanımlı trafiğe sahip dönel kavşaklarda meydana gelen 439 bisikletli kazalarını içermektedir. Bağımsız değişkenler bisikletlilerin sosyodemografik bilgileri, hava koşulları, sürücü davranışı ile ilgili faktörler, hız limiti ve kavşak geometrik parametreleridir. Hem Random Forest hem de Support Vector Machine algoritmalarının eğitim aşamasında veri setinin %70'i, test aşamasında ise %30'u kullanılmıştır. Algoritmaların test aşamasından sonra ortaya çıkan sonuçlara göre, Random Forest yönteminin sınıflandırma doğruluğunun %88.6 olduğu belirlenmiştir. Support Vector Machine algoritmasının ise %84.73 sınıflandırma doğruluğu ile tahmin modeli oluşturduğu tespit edilmiştir. Yanlış tahmin edilen veri sayısı Random Forest yönteminde 18 iken Suppport Vector Machine yönteminde 20'dir. Sonuçlar, hem Random Forest hem de Support Vector Machine algoritmalarının, bisikletli kaza yaralanma derecesi tahmin modelleri oluşturmak için yüksek performansa sahip uygulanabilirliklerinin olduğunu göstermektedir.

Anahtar kelimeler: Makine öğrenmesi, Bisikletli güvenliği, Geometri, Dönel kavşak, Sınıflandırma doğruluğu

*Corresponding author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

1. Introduction

Give-way (non-signalized) roundabouts increase capacity [1], and delays are distributed more uniformly when traffic flows are balanced [2]. The "priority to right" give-way rule at roundabouts has been applied in the United Kingdom since 1996. Under this rule, drivers entering the roundabout must yield and give priority to circulating traffic [3]. Give-way roundabouts eliminate the number of conflict points which are potential collision locations [4]. In addition, they provide route deflection and force drivers to decrease their speed [2, 5, 6]. It was suggested that the recorded number of collisions and serious injuries for motor vehicle drivers reduced after converting signalized intersections to give-way roundabouts [7]. Therefore, there is a trend to convert signalized intersections to give-way roundabouts. However, give-way roundabouts are known to be less safe for cyclists compared to signalized intersections [8-10]. Converting signalized intersections to give-way roundabouts increases both collision number and severity [7, 11-13]. It has been suggested that approximately 25% of cyclists prefer to change their routes to avoid multilane give-way roundabouts [14] or accept the risk of casualty despite feeling unsafe [15].

Safety for cyclists at roundabouts depends on several factors, such as geometry, speed, pavement, markings, signage, driver/cyclist's age, gender, behavior, education, public awareness, and enforcement [8, 16, 17]. In particular, geometry and behavior-related contributory factors play a key role in improving safety for cyclists [18, 19]. Gaining a deeper understanding of parameters that affect cyclist crash severity at roundabouts is essential to improve safety. Prediction methods have been widely used to explore the influences on crash severities [20] and reduce the records [21]. Severity has an ordinal or binary data structure such as slight, serious and killed; therefore, logistic regression modeling approach has been applied in the majority of the former studies [1, 16, 18, 19, 22]. In addition, Poisson regression [16, 23], linear regression [9, 24] and gamma regression [23] models were applied to cyclist safety prediction research.

Traditional regression algorithms have inflexible assumptions such as linearity, independence of errors, normality, and multicollinearity [25]. The structure of the dataset collected from the real environment usually does not meet these assumptions. In addition, the number of variables in a regression model is important. Adding all variables into the regression model may cause overfitting and inefficiency. Adding a few variables may also cause underfitting and provide biased results [26]. These issues of traditional regression models reveal a significant limitation that offers a new prediction method, namely machine learning algorithms [27]. Therefore, machine learning algorithms have gained significant attention in developing accurate road crash severity prediction methods [28].

Developing crash severity prediction models using machine learning algorithms, which is a sub-division of artificial intelligence, has gained attention since 2001. The performances of different types of applied algorithms, such as Bayesian Network, K-Nearest Neighbors, Support Vector Machine, Random Forest, Multi-layer Perceptron, Artificial Neural Network, and Decision Tree, were reviewed [28]. It was suggested that Random Forest provided the highest accuracy in 70% of the applied prediction studies [28]. The second successful algorithm was Support Vector Machine and it was followed by Decision Tree and K-Nearest Neighbors. However, temporal instability was observed in the applied prediction algorithms in the literature. There is a significant need in this research field to explore the accuracy of the algorithms by using different types of casualty severity data [28]. This gap is even more significant in cyclist safety research because the majority of the studies focus on vehicle-vehicle crash records.

The number of applied cyclist casualty studies [29-35] was not sufficient in order to determine the best performance machine learning algorithms. An earlier Advanced Driver Assistance Systems based study was carried out considering vehicle cyclist realistic scenario [35]. Binary classifiers namely Support Vector Machine and Multiple Instance Learning algorithms were used to analyze 99 scenarios. The results suggested that these two algorithms were found to be applicable to detect and classify the scenarios with high accuracy. A comparison of these two algorithms suggested that Support Vector Machine gave the best performance by 87.9%. The outcomes revealed that different algorithms gave different performance levels to predict the safety for cyclists. Therefore, further studies focused on applying several classification algorithms in order to explore the best performance for prediction models.

A UK based study [33] aimed to analyze cyclist-vehicle crash severity by using Ordered Forest machine learning algorithm. Ordered Forest classifier, which was a subclass of traditional Random Forest, showed a fair performance by approximately 50%. The outcomes suggested that cyclist/driver' behavior, age and speed limit were statistically significant in the classification model. A similar Indian dataset including 160.597 cyclist casualty severity records was analyzed using Naive Bayes, Logistic Regression, Random Forest and Decision Tree classification techniques [32]. Random Forest provided the highest accuracy by 98%. The model showed that weather and road surface conditions and speed limit were found to be significant in predicting cyclist casualty severity. A comprehensive study [34] was carried out to develop a classification model for cyclist casualty severity. The study applied several algorithms which were Logistic Regression, Gradient Boosting, Naive Bayes, Random Forest, Gaussian Naive Bayes, Ridge, Support Vector Machine, Decision Tree, Extra Tree, Linear Support Vector Machine, Perceptron Algorithm, and K-nearest Neighbors. The compared performance outcomes suggested that Extra Tree algorithm gave the best performance among the applied techniques. The statistically significant variables were noted as sociodemographic, speed limit, traffic control, road user behavior, weather, road geometry and surface condition related predictive parameters.

Not only the importance of the type of algorithm but also the sampling technique was observed. The study [31] used Decision Tree and logistic regression algorithm to explore the impacts on vulnerable road users' crash severity. The study applied different resampling techniques namely under, over and synthetic sampling. The results suggested that over sampling increased the performance of used classification algorithms. The outcomes of the prediction showed that road infrastructure and sociodemographic variables had an influence on casualty severity. A study [30] aimed to explore the performance of hybrid models. These hybrid models are a mix of Fuzzy Logit and Decision Tree algorithms, namely Decision Tree based converted Fuzzy Logit (DT-CFL) and Decision Tree based revised Fuzzy Logit (DR-RFL). The dataset considered road infrastructure, geometry, weather conditions and behavior related contributory factors as predictive variables. The outcomes suggested that gender, vehicle damage-extent, road and pavement type, and vehicle-movement were found statistically significant in both models. Regarding the algorithm comparison analysis, the performance of DT-RFL was higher than DT-CFL with 69.96 % and 59.22% accuracy levels, respectively. The other current study [29] applied Neural Network algorithm and suggested that a joint correlation analysis using machine learning techniques might be a novel approach for cyclist casualty severity studies.

The current knowledge in the literature suggests that different algorithms and sampling techniques provided different level of model performance and a solid approach for analysis of casualty severity has not been determined yet. The second important gap in the literature was the target groups in the data. The majority of the studies analyzed either vehicle-vehicle or motorcyclist-vehicle crash severities. However, vulnerable road user crash severities, particularly cyclists, have not been analyzed comprehensively. A third significant gap in the literature was none of the studies focused on intersections, particularly give-way roundabouts as known dangerous for cyclists. Considering these gaps, the research in this paper aimed to apply the two highest recommended machine learning algorithms, Random Forest and Support Vector Machine, to explore the impact on cyclist casualty severity that occurred at give-way roundabouts with mixed traffic.

Machine learning algorithm has a learning process from a set of data and develops a classification or prediction model [36]. The advantage of machine learning is making no assumption about the structure or relationship between the predictive variables. Therefore, it is more preferable than traditional statistical techniques [34]. It has three types of learning supervised, unsupervised and semi-supervised [27]. Both of the algorithms used in this research, Random Forest and Support Vector Machine, are supervised classification techniques. Random Forest is developed from decision tree technique that creates a forest (Figure 1). The input values create the trees in the forest and the output shows the accuracy of the model. The higher number of variables in the model increases the accuracy of the algorithm [32]. Random Forest is a subdivision of Decision Tree method but the knowledge in the literature suggests that Random Forest gives better performance than Decision Tree technique [37]. Support Vector Machine is a type of supervised learning binary model [35]. The algorithm segregates and classifies the input and builds a hyperplane to separate the classes [38]. Support vectors are the points that are close to the dividing line and the margin is the distance between support vectors and the dividing line [39]. These two algorithms were widely applied in modeling crashes by severity because severity has bivariate (non-injured, injured) or multivariate (slight, serious, killed) levels [27]. Random Forest outperformed Decision Tree in the research conducted by [40, 41]. Despite the applied alternative methods such as K-Nearest Neighbor and Support Vector Machine in these studies, RF consistently demonstrated superior performance.

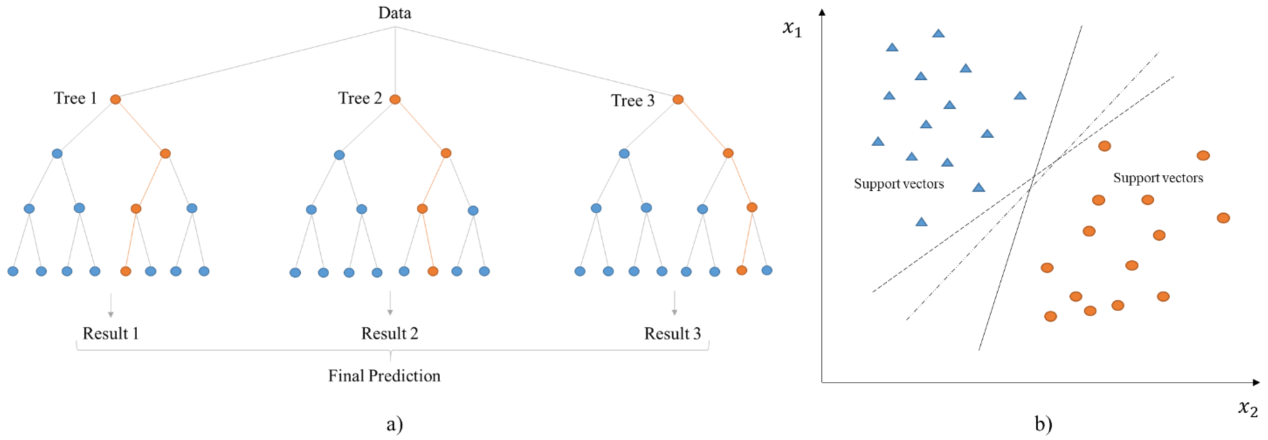

Figure 1. a) Random Forest algorithm, trees in a forest; b) Support Vector Machine algorithm, possible separating hyperplanes

The objectives of the research are to develop a cyclist casualty prediction model and compare the performance of applied algorithms in order to contribute to the gap in the literature. Section 2 explains the methodology of the research and the structure of the data. The results are presented in Section 3. Finally, discussion, recommendations and limitations are given in Section 4.

2. Methodology

The supervised classification technique was selected to apply because the response variable of the dataset was categorical values as casualty severity (slight and serious). The records did not have killed casualties; therefore, the response values were binary classes. Random Forest and Support Vector Machine algorithms were applied using the same dataset and the obtained classification accuracies from both methods were compared. Classification accuracy is calculated using a confusion matrix. The outcome variable has binary values such as slight and serious; therefore, the confusion matrix has two categories with a 2x2 matrix. There are four predicted options on the matrix which are True Positives: slight is predicted and the real output is slight, True Negatives: serious is predicted and the real output is serious, False Positives: slight is predicted and the real output is serious, False Negatives: serious is predicted and the real output is slight. False Positive is type 1 error and False Negative is type 2 error. Classification accuracy is the ratio of the correct number of predictions to the total number of predictions across all three classes. The accuracy ranges between 0-1 and it should be as high as possible. The overall accuracy can be calculated as given in Equation 1 [34].

$$Classification\ Accuracy\ =\ (True\ Positive\ +\ True\ Negative)\ /\ Total\ Sample \quad (1)$$

The Northumbria region located in the North East of England was chosen as a case study (See Figure 2). In the case study area, cyclists have to share the road with private motor vehicles due to the lack of cycling infrastructure facilities. The cyclist casualty at give-way roundabout data from the UK STATS19 police records was provided through the Traffic and Accident Data Unit which is held by Gateshead Council. In the study area, comprehensive collision records have been collected by the local authorities since 2010; therefore, the data for the period of 2011-2016 were analyzed in this work. The data includes geographical coordinates of the roundabouts where the collision occurred. The physical data for the roundabouts where the collisions occurred was obtained as maps from Digimap in the form of AutoCAD files, and the geometric variables and coordinates of collisions were imported to these maps. Roundabout geometric design variables were determined by considering the regulations given in the UK road design manual (Figure 2) [42]. The dataset included 370 slight and 69 serious cyclist casualty records. There was no missing value in the dataset. The predictor variables were cyclist's gender and age, light, weather and road surface conditions, speed limit, and behavior related to contributory factors such as failed to look properly, careless, passing closed to cyclist, failed to judge other person's path or speed, and poor turn or maneuver. Data also included roundabout geometric design variables namely approach number of lanes, approach width, entry path radius, number of arms, inner circle radius, entry width, entry number of lanes, and circulating number of lanes. The descriptive statistics, Random Forest and Support Vector Machine algorithms were applied in the R statistical analysis program.

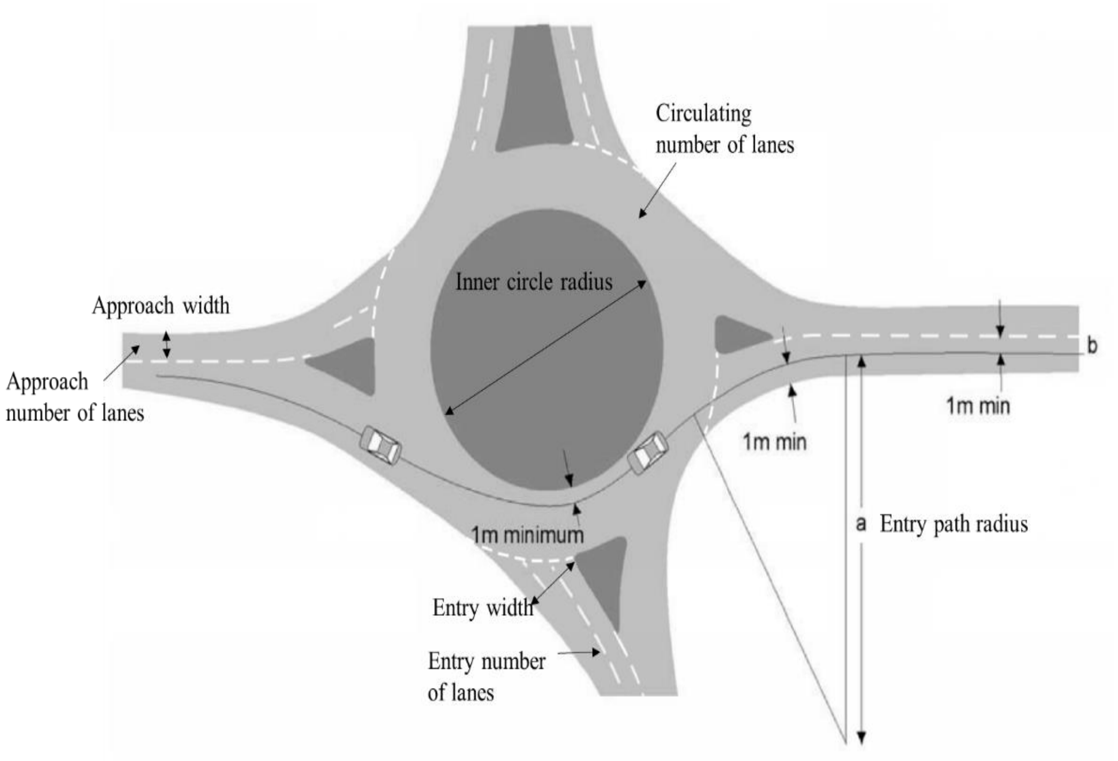

Figure 2. Roundabout geometric design parameters [42].

3. Results

Descriptive statistic was applied and the results showed that the number of male cyclists was higher than females (Table 1). As stated in the state of the art review, cyclists avoided using roundabouts with mixed traffic [14]. This situation might be even more significant considering the gender. In addition, the unpopularity of cycling among females might reduce the number of female cyclists and therefore the number of female records might be significantly less than males. The mean age was 39 for slight and 41 for serious casualties. This might be the higher cycling rates for middle age male road users. The majority of the casualties occurred in daylight, fine weather and dry road surface conditions. It should be noted that cyclists might not prefer cycling under heavy environmental conditions. Approximately 80% of the casualties had failed to look properly issue at roundabouts. This is followed by not judging other road user's path by %23. The majority of the records were observed at roundabouts with a 30mph speed limit and four arms due to the high rate of urban cycling demand.

Table 1. Descriptive statistics of variables

Variable	Slight Casualty	Serious Casualty
Gender	Female (45); Male (325)	Female (10); Male (59)
Age	Min. (5); Max. (77); Mean (39); S.D. (14)	Min. (15); Max. (80); Mean (41); S.D. (14)
Light condition	Darkness (83); Daylight (287)	Darkness (15); Daylight (54)
Weather	Fine (313); Rain (45); Other (12)	Fine (60); Rain (9); Other (0)
Road surface	Dry (261); Wet (100); Ice (9)	Dry (51); Wet (18); Ice (0)
Failed to look properly	Yes (291); No (79)	Yes (58); No (11)
Careless	Yes (85); No (285)	Yes (14); No (55)
Passing too closed to cyclist	Yes (49); No (321)	Yes (9); No (60)
Failed to judge other person's path or speed	Yes (91); No (279)	Yes (10); No (59)
Poor turn or maneuver	Yes (45); No (325)	Yes (6); No (63)
Speed limit (mph)	20 (3); 30 (280); 40 (33); 50 (12); 60 (33); 70 (9)	20 (2); 30 (43); 40 (8); 50 (1); 60 (9); 70 (6)
Approach number of lanes	1 (274); 2 (90); 3 (6)	1 (32); 2 (36); 3 (1)
Approach width (meter)	Min. (3); Max. (11.37); Mean (5.15); S.D. (1.79)	Min. (3); Max. (8.78); Mean (5.81); S.D. (1.66)
Entry path radius (meter)	Min. (19.23); Max. (99.83); Mean (64.36); S.D. (20.58)	Min. (23.77); Max. (99.98); Mean (80.74); S.D. (20.35)
Arms (3; 4; 5; 6)	3 (60); 4 (245); 5 (53); 6 (12)	3 (13); 4 (41); 5 (12); 6 (4)
Inner circle radius (meter)	Min. (1.00); Max. (124.79); Mean (18.36); S.D. (23.06)	Min. (1.00); Max. (124.79); Mean (20.01); S.D. (23.71)
Entry width (meter)	Min. (3.00); Max. (20.27); Mean (7.68); S.D. (2.60)	Min. (4.00); Max. (12.66); Mean (7.87); S.D. (2.26)
Entry number of lanes	1 (168); 2 (180); 3 (22)	1 (21); 2 (41); 3 (6); 4 (1)
Circulating number of lanes (1; 2; 3)	1 (237); 2 (127); 3 (6)	1 (33); 2 (35); 3 (1)

It is suggested in the literature that, approximately 60-80% of the data is used for training and the rest of the data is used for testing. More in detail, the majority of the crash severity analysis used 70% for training and 30% for testing [27]. With respect to the applied casualty severity studies in literature, in the Random Forest algorithm, 70% of the data were randomly selected for the training stage and 30% was used for the testing. All the predictive variables were used in the algorithm. After training the algorithm, the testing procedure started and the results showed that the Random forest algorithm predicted the outcomes with 88.6% of classification accuracy. In the Support Vector Machine algorithm, 70 % of the data were used for the training and 30 were used for the tasting stages. The results showed that the model predicted the testing values with 84.73% classification accuracy. It only misestimated 18 and 20 of the casualties in Random Forest and Support Vector Machine algorithms, respectively (Table 2).

Table 2. Random Forest and Support Vector Machine outcomes

Algorithm	Testing Outcomes (Confusion Matrix)			Classification Accuracy
Random Forest		Slight	Serious	88.60%
	Slight	136	3*	
	Serious	15**	4	
Support Vector Machine		Slight	Serious	84.73%
	Slight	111	20*	
	Serious	0**	0	

*Type 1 Error
**Type II Error

An example tree of the Random Forest is given in Figure 3. It can be seen that the root of the tree started with the approach number of lanes. If the lane number was 1, the branch grew in the left direction; and if no, in the right direction. The

Random Forest algorithm created several trees which generated a forest. The average results of these forests provided a final prediction. Regarding the Support Vector Machine algorithm, the hyperplane was not drowned because the model was multi-dimensional. It was aimed to improve the performance of the Support Vector Machine and the C parameters were customized in classification. Different C parameters (i.e. 0, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, and 2.5) were coded in the algorithm. The model was trained with the new C parameters and the testing was carried out. The higher C parameter reduced the accuracy of the model (Figure 4). This result was expected because a larger C parameter considers the outliers while creating a hyperplane. Hyperplane should separate the values correctly and the margin should be minimum. However, both cases may not occur at the same time. When the C parameter is low, the hyperplane is located with a large minimum margin but some outliers can be observed. Therefore, a higher C parameter may change the location of the hyperplane and the accuracy of the model can be affected. In this analysis, the increase in the C parameter reduced the classification accuracy.

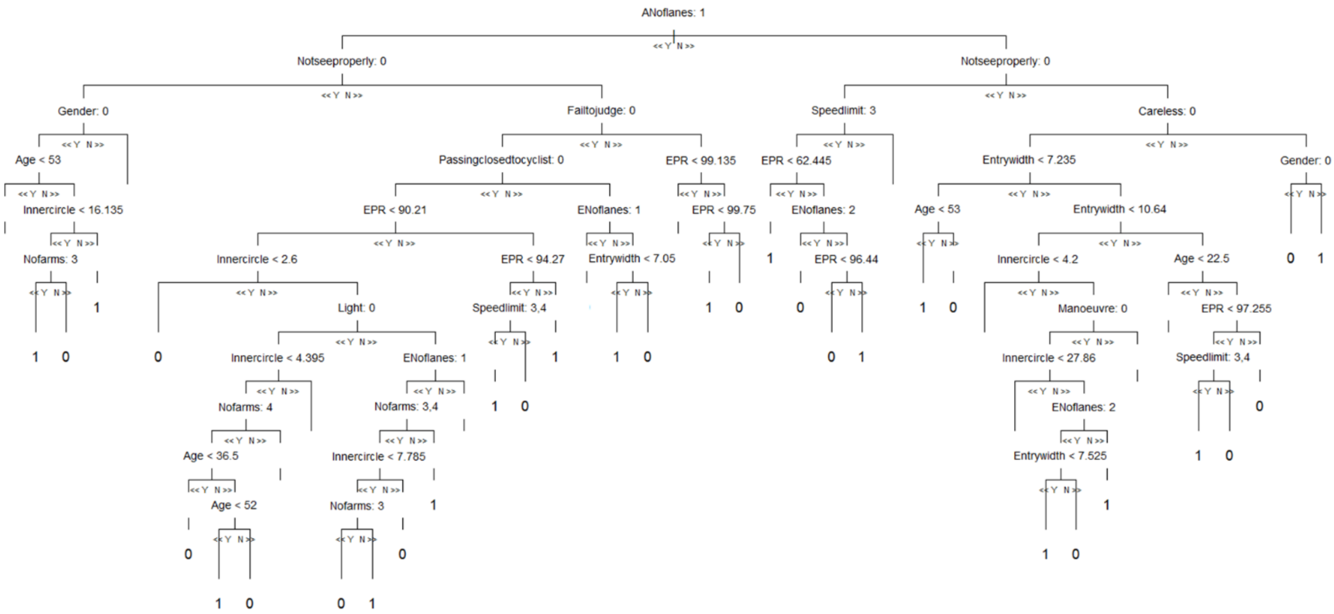

Figure 3. Accuracy outcomes for different C parameters

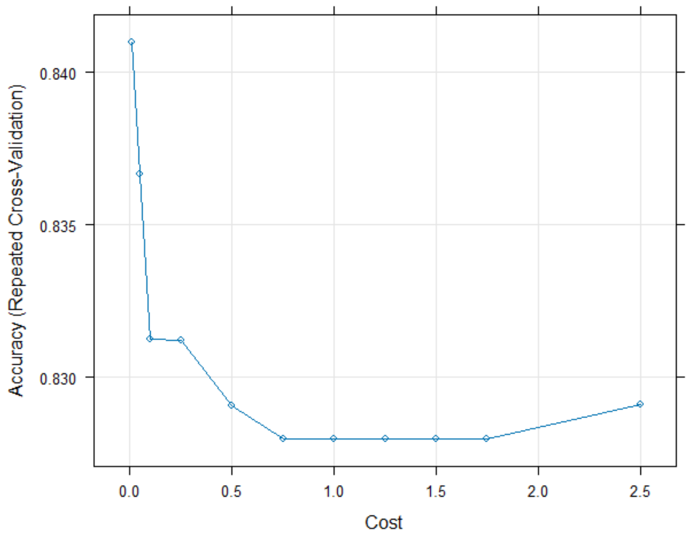

Figure 4. Accuracy outcomes for different C parameters

4. Discussion

The study in this paper explored the performance of Random Forest and Support Vector Machine algorithms to predict cyclist casualty severities. The results suggested that Random Forest performed slightly better than Support Vector Machine. This outcome was expected since the majority of the studies in the literature suggested that Random Forest was the top recommended algorithm with the highest accuracy among the other machine learning algorithms [28]. However, further studies should consider other types of algorithms. For instance, research suggested [34] that Extra Tree algorithm performed better than Random Forest. It is clearly seen that Random Forest has been the most advised algorithm but the results might change for every dataset.

The outcomes of this paper showed that the classification accuracy was 88.6% for the Random Forest algorithm. However, the former study which applied Random Forest reached higher accuracy with 98% [32]. This study [32] and the research in this paper used a similar dataset structure; however, the number of sample size was 160.597 in the former study [32]. This suggests that increasing the sample size can increase the classification accuracy of the algorithm. A similar conclusion was observed for Support Vector Machine algorithm. The results in this analysis suggested that classification accuracy was 84.73%; while a previous study reached 87.9% [35]. However, the study in this paper suggests that high accuracy can be obtained with 439 casualty records. It should be noted here that the number of variables used in the algorithm is also important because more variables in the model increase the accuracy of the machine learning algorithms [32].

Considering the given knowledge in the literature [27], the majority of the former studies used 70% of the data for training and 30% for testing in the analysis. This study in this paper also used a 70/30 ratio in both algorithms. However, the 70/30 ratio is a rule of thumb and it can be extended from 60/40 to 80/20. Future studies can consider different types of ratios in the analysis. In addition, a comparative study can be conducted considering different ratios for different classification algorithms.

The limitation of this study was the limited number of casualty records. Future studies may consider applying Random Forest and Support Vector Machines classifiers using around 440 observations; however, increasing the sample size is suggested. Furthermore, machine learning algorithms are still in development process. Some derivative algorithms, such as Ordered Forest [33], can be considered in further research.

5. Acknowledgement

The casualty severity data was obtained from Gateshead Council, England.

6. Author Contribution Statement

In the study, Author 1 contributed to literature review, methodology, analysis and writing up the article.

7. Ethics Committee Approval and Conflict of Interest

Ethics committee approval is not needed for preparing the article.
There is no conflict of interest for this article.

8. References

[1] Silvano AP, Ma X, Koutsopoulos HN. "When do drivers yield to cyclists at unsignalized roundabouts". *Transportation Research Record: Journal of the Transportation Research Board*, 2520, 2015.
[2] Silvano AP, Linder, A. "Traffic safety for cyclists in roundabouts: Geometry, traffic, and priority rules". *Swedish National Road and Transport Research Institute*, 2017.
[3] Bruce W, Rodegerdts L, Scarborough W, Kittelson W, Troutbeck R, Brilon W, Bondzio L, Courage K, Kyte M, Mason J, Flannery A, Myers E, Bunker J, Jacquemart G. "Roundabouts: an informational guide". US Department of Transport: Federal Highway Administration, AASHTO, 2000.
[4] Poudel N, Singleton PA. "Bicycle safety at roundabouts: a systematic literature review". *Transport Reviews*, 41 (5), 617-642, 2021.
[5] Retting RA, Persaud BN, Garder PE, Lord D. "Crash and injury reduction following 17 installation of roundabouts in the United States". *American Journal of Public Health*, 91 (4), 628-31, 2001.
[6] Gross F, Lyon C, Persaud B, Srinivasan R. "Safety effectiveness of converting signalized intersections to roundabouts". *Accident Analysis & Prevention*, 50, 234–241, 2013.
[7] De Brabander B, Vereeck L. "Safety effects of roundabouts in Flanders: signal type, speed limits and vulnerable road users". *Accident Analysis & Prevention*, 39 (3), 591-599, 2007.
[8] Furtado G. "Accommodating vulnerable road users in roundabout design". Annual Conference of the Transportation, Canada, Quebec City, 2004.
[9] Daniels S, Brijs T, Nuyts E, Wets G. "Injury crashes with bicyclists at roundabouts: influence of some location characteristics and the design of cycle facilities". *Journal of Safety Research*, 40 (2), 141-148, 2009.
[10] Jensen SU. "Safe roundabouts for cyclists". *Accident Analysis & Prevention*, 105, 30-37, 2017.
[11] Robinson BW, Rodegerdts L, Scarborough W, Kittelson W, Troutbeck R, Brilon W, Bondzio L, Courage K, Kyte M, Mason J, Flannery A, Myers E, Bunker J, Jacquemart G. "Roundabouts: an informational guide". FHWA-RD-00-067, Project 2425, Informational Guide Book, 2000.
[12] Persaud BN, Retting RA, Garder PE, Lord D. "Observational before-after study of the safety effect of U.S. roundabout conversions using the empirical Bayes method". *Annual Meeting of the Transportation Research Board*, TRB ID: 01-0562, 2001.
[13] Elvik R. "Effects on road safety of converting intersections to roundabouts: review of evidence from non-US studies". *Transportation Research Record: Journal of the Transportation Research Board*, 1847 (1), 2003.
[14] Arnold LS, Flannery A, Ledbetter L, Bills T, Jones MG, Ragland DR, Spautz L. "Identifying factors that determine bicyclist and pedestrian: involved collision rates and bicyclist and pedestrian demand at multi-lane roundabouts". UC Berkeley Safe Transportation Research & Education Center, I.o.T.S. University of California, Berkeley, 2010.
[15] Davies DG, Taylor MC, Ryley TJ, Halliday ME. "Cyclists at roundabouts — the effects of 'Continental' design on predicted safety and capacity". *Transport Research Laboratory*, 1997.
[16] Hels T, Orozova-Bekkevold I. "The effect of roundabout design features on cyclist accident rate". *Accident Analysis & Prevention*, 39 (2), 300-307, 2007.
[17] Montella A. "Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types". *Accident Analysis & Prevention*, 43 (4), 1451-1463, 2011.
[18] Akgün N, Dissanayake D, Thorpe N, Bell MC. "Cyclist casualty severity at roundabouts – To what extent do the geometric characteristics of roundabouts play a part?". *Journal of Safety Research*, 67, 83–91, 2018.
[19] Akgün N, Daniels S, Bell MC, Nuyttens N, Thorpe N, Dissanayake D. "Exploring regional differences in cyclist safety at roundabouts: A comparative study between the UK (based on Northumbria data) and Belgium". *Accident Analysis & Prevention*, 150, 105902, 2021.
[20] Wang C, Quddus MA, Ison SG. "Predicting crash frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model". *Accident Analysis & Prevention*, 43 (6) (2011), 1979-1990, 2011.
[21] Savolainen P, Mannering F, Lord D, Quddus MA. "The statistical analysis of crash-injury severities: a review and assessment of methodological alternatives". *Accident Analysis & Prevention*, 43 (5) (2011), 1666-1676, 2011.
[22] Daniels S, Brijs T, Nuyts E, Wets G. "Externality of risk and crash severity at roundabouts". *Accident Analysis & Prevention*, 42 (6), 1966-1973, 2010.
[23] Daniels S, Brijs T, Nuyts E, Wets G. "Extended prediction models for crashes at roundabouts". *Safety Science*, 49 (2), 198-207, 2011.

[24] Møller M, Hels T. "Cyclists' perception of risk in roundabouts". *Accident Analysis & Prevention*, 40 (3), 1055-1062, 2008.
[25] Field A. Discovering Statistics Using SPSS. 3rd Edition, SAGE Publications Ltd, 2009.
[26] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. "A simulation study of the number of events per variable in logistic regression analysis". *Journal of Clinical Epidemiology*, 49 (12), 1373-1379, 1996.
[27] Silva PB, Andrade M, Ferreira S. "Machine learning applied to road safety modeling: A systematic literature review". *Journal of Traffic and Transportation Engineering* (English Edition), 7 (6), 775-790, 2020.
[28] Santos K, Dias JP, Amado C. "A literature review of machine learning algorithms for crash injury severity prediction". *Journal of Safety Research*, 80, 254-269, 2022.
[29] Janstrup KH, Kostic B, Møller M, Rodrigues F, Borysov S, Pereira FC. "Predicting injury-severity for cyclist crashes using natural language processing and neural network modelling". *Safety Science*, 164, 106153, 2023.
[30] Katanalp BY, Eren E. "The novel approaches to classify cyclist accident injury-severity: Hybrid fuzzy decision mechanisms". *Accident Analysis & Prevention*, 144, 105590, 2020.
[31] Vilaça M, Macedo E, Coelho MC. "A Rare event modelling approach to assess injury severity risk of vulnerable road users". Safety, 5(2), 29, 2019.
[32] Yamparala R, Challa R, Valeti P, Chaitanya PS. "Prediction of cyclist road accidents in india using machine learning and visualization techniques". Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 476-481, 2022.
[33] Zhang Y, Li H, Ren G. "Analyzing the injury severity in single-bicycle crashes: An application of the ordered forest with some practical guidance". *Accident Analysis & Prevention*, 189, 107126, 2023.
[34] Birfir S, Elalouf A, Rosenbloom T. "Building machine-learning models for reducing the severity of bicyclist road traffic injuries". *Transportation Engineering*, 12, 100179, 2023.
[35] Cara I, De Gelder E. "Classification for safety-critical car-cyclist scenarios using machine learning". IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 1995-2000, 2015.
[36] Fan J, Ma C, Zhong Y. "A selective overview of deep learning". Statistical Science, 36(2), 264-290, 2021.
[37] Li L, Shrestha S, Hu G. "Analysis of road traffic fatal accidents using data mining techniques". Ieee 15th International Conference on Software Engineering Research, Management and Applications (Sera), 363–70, 2021.
[38] Boser BE, Guyon IM, Vapnik VN. "A training algorithm for optimal margin classifiers". Proceedings of the 5th Annual Workshop on Computational Learning Theory, 144-152, 1992.
[39] Suthaharan S. "Support vector machine - In Machine learning models and algorithms for big data classification". Integrated Series in Information Systems, 207–235, Springer, 2016.
[40] Wahab L. Jiang H. "A comparative study on machine learning based algorithms for prediction of motorcycle crash severity". PLoS one, 14-4, 2019.
[41] Zhang J, Li Z, Pu Z, Xu C. "Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods". IEEE Access, 6, 60079-60087, 2018.
[42] DfT. "TD 16-07 Geometric design of roundabouts". Department for Transport, the UK, 2007.