Accident Analysis and Prevention 144 (2020) 105590

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

The novel approaches to classify cyclist accident injury-severity: Hybrid fuzzy decision mechanisms

Burak Yiğit Katanalp*, Ezgi Eren

Adana Alparslan Turkes Science and Technology University, Faculty of Engineering, Civil Engineering Department, Adana, Turkey

ARTICLE INFO

Keywords:
Fuzzy logic
Decision tree
Injury-severity
Machine learning
Cyclist safety

ABSTRACT

In this study, two novel fuzzy decision approaches, where the fuzzy logic (FL) model was revised with the C4.5 decision tree (DT) algorithm, were applied to the classification of cyclist injury-severity in bicycle-vehicle accidents. The study aims to evaluate two main research topics. The first one is investigation of the effect of road infrastructure, road geometry, street, accident, atmospheric and cyclist related parameters on the classification of cyclist injury-severity similarly to other studies in the literature. The second one is examination of the performance of the new fuzzy decision approaches described in detail in this study for the classification of cyclist injury-severity. For this purpose, the data set containing bicycle-vehicle accidents in 2013–2017 was analyzed with the classic C4.5 algorithm and two different hybrid fuzzy decision mechanisms, namely DT-based converted FL (DT-CFL) and novel DT-based revised FL (DT-RFL). The model performances were compared according to their accuracy, precision, recall, and F-measure values. The results indicated that the parameters that have the greatest effect on the injury-severity in bicycle-vehicle accidents are gender, vehicle damage-extent, road-type as well as the highly effective parameters such as pavement type, accident type, and vehicle-movement. The most successful classification performance among the three models was achieved by the DT-RFL model with 72.0 % F-measure and 69.96 % Accuracy. With 59.22 % accuracy and %57.5 F-measure values, the DT-CFL model, rules of which were created according to the splitting criteria of C4.5 algorithm, gave worse results in the classification of the injury-severity in bicycle-vehicle accidents than the classical C4.5 algorithm. In light of these results, the use of fuzzy decision mechanism models presented in this study on more comprehensive datasets is recommended for further studies.

1. Introduction

Keeping motor vehicles in the foreground during urban transportation planning causes non-motorized road users to face serious risk of injury or death at any time during their travels. Especially in developed countries, the high number of motor vehicles in traffic and high speeds pose serious risks for non-motorized road users. According to the European Road Safety Observatory (2017), 7466 fatal accidents involving pedestrians and cyclists occurred in the whole Europe (Highway Accident Statistics, 2017). In order to prevent or minimize these accidents, decision makers are considered as an important solution for the construction of facilities such as bicycle roads, pedestrian crossings, and sidewalks where pedestrians and cyclists feel safe (Eren and Uz, 2019; Nikiforiadis and Basbas, 2019; Soriguera and Jiménez-Meroño, 2020). However, despite the precautions taken, accidents sometimes cannot be prevented. For this reason, many studies in the literature have examined the relations among parameters including road geometry, socio-demographic characteristics, environmental factors, and users' behavioral characteristics as well as accident frequency and injury-severity in order to increase the safety of non-motorized road users (Dai, 2012; Ma et al., 2017; Pour-Rouholamin and Zhou, 2016). Defining these key factors plays an important role in planning to reduce the number of accidents and increase traffic safety.

In this paper, the DT-based two hybrid-fuzzy models used to classify injury-severity in accidents involving cyclists, who are non-motorized road users, are presented. One of the models called DT-CFL is structured by the rules of the classical C4.5 algorithm. Although this model exists in the literature (Banakar et al., 2017; Goel and Sehgal, 2015), its use for classifying injury-severity in an accident has not been encountered. The other model called DT-RFL is a completely novel approach based on the revising of the membership functions of each input by using the effect values determined via the C4.5 algorithm splitting criteria. Both models were structured with the C4.5 algorithm because it was seen that the C4.5 algorithm is successful in dealing with the injury-severity

* Corresponding author.
E-mail addresses: bkatanalp@atu.edu.tr (B.Y. Katanalp), eeren@atu.edu.tr (E. Eren).

https://doi.org/10.1016/j.aap.2020.105590
Received 16 March 2020; Received in revised form 9 May 2020; Accepted 10 May 2020
Available online 02 July 2020
0001-4575/

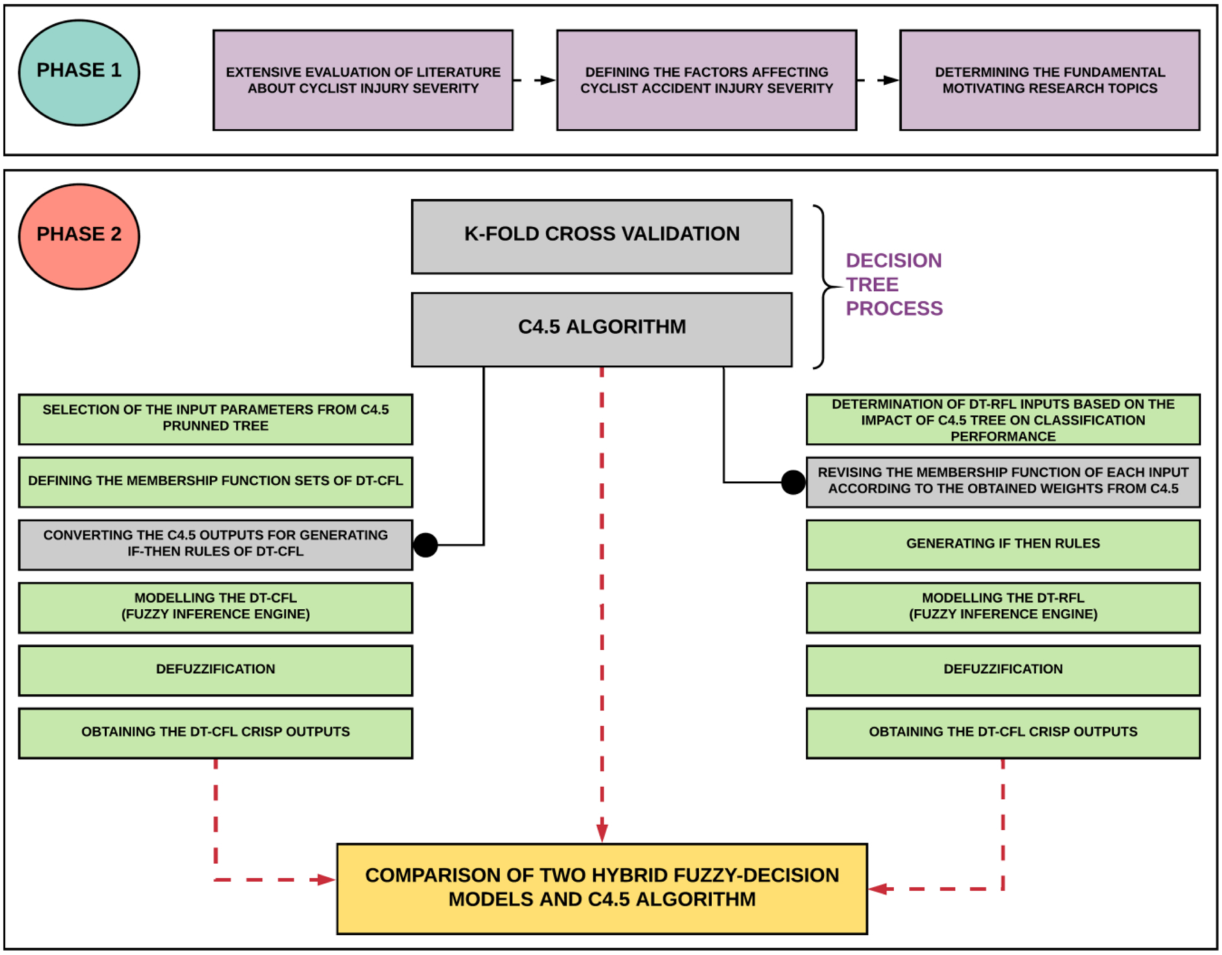

Fig. 1. Framework of the study.

classification (de Oña et al., 2013a, b) and it provides a visualization that reveals the importance of variables in the classification of injury-severity (Mafi et al., 2018). The organization of this study is comprised of two phases and is given in Fig. 1. In the first phase, a comprehensive evaluation of the existing literature was carried out and the research questions that motivated this study were identified by defining the factors affecting injury-severity in bicycle-vehicle accidents. At the beginning of the second phase, the C4.5 algorithm was used to classify accident injury-severity. Then two different fuzzy-hybrid models based on DT were created. Finally, among the infrastructure-related, road-related, environment-related, cyclist-related, accident-related, and safety-related parameters, those with high efficiency were evaluated considering the findings in the literature and the results of all three models were compared.

2. The state-of-the-art injury-severity classifications in accident research

To investigate the factors affecting injury-severity and road safety, ordered logit (OL) and probit (OP) models (Mujalli and de Oña, 2013; Pour-Rouholamin and Zhou, 2016), multinomial logit model (MNL) (Iranitalab and Khattak, 2017; Shankar and Mannering, 1996; Sun et al., 2019), mixed logit model (Haleem et al., 2015; Moore et al., 2011), logistic regression (LR) (Al-Ghamdi, 2002; Delen et al., 2017; Olszewski et al., 2015; Sarkar et al., 2011), latent class model (Behnood et al., 2014; Eluru et al., 2012; Sasidharan et al., 2015), markov switching model (Malyshkina and Mannering, 2009; Xiong et al., 2014), and mixed logit model with heterogeneity in means and variances (Behnood and Mannering, 2017; Seraneeprakarn et al., 2017; Xin et al., 2017) are commonly used in the literature. In addition, the use of machine learning-based models such as artificial neural networks (ANN) (Chimba and Sando, 2009; Delen et al., 2017; Katanalp et al., 2019; J. Lee et al., 2020; Zeng and Huang, 2014), classification and decision trees (Abellán et al., 2013; Karlaftis and Golias, 2002; Vilaça et al., 2019; Zhou et al., 2019) support vector machine (SVM) (Delen et al., 2017; Iranitalab and Khattak, 2017; Yu and Abdel-Aty, 2014) are frequently encountered in the literature.

In these studies, accident injury-severity was handled at different levels such as fatal, injury and no-injury. Then, road geometric features, socio-demographic features, traffic characteristics, environmental conditions, etc. parameters were defined as independent variables and their relationship with injury-severity was examined (Ye and Lord, 2014).

A study to investigate the factors affecting accident injury-severity (Eluru et al., 2008) stated that the most important factors affecting the injury-severity in accidents involving bicycles and pedestrians are age, speed limit, accident location and time of the day. In addition to these results, C. Lee and Abdel-Aty (2005) mentioned that alcohol use and atmospheric conditions have an important effect on the severity of injury.

Sze and Wong (2007) stated that accidents resulting in injury/fatality are related to high age, high-speed limits, head injuries and the place where the accident occurred. In another study Clifton et al. (2009) included the built-in environmental factors among the factors impacting the injury-severity and suggested that old age increases the probability of fatal accidents and road network connections are also effective on the possibility of injury accidents. Their study suggests that structural environmental factors should also be taken into account in traffic planning. Ukkusuri et al. (2012) supported the results of this

Table 1
Summary of the previous studies that investigate the factors affecting injury-severity.

Parameters	Studies & Year	Analyze Method
Age	Islam and Mannering (2006)	*Multinomial Logit Model*
	Yan, Radwan, and Birriel (2005)	*The Quasi-Induced Exposure Concept, Multiple Logistic Regression*
	Pour-Rouholamin and Zhou (2016)	*Ordered Logit (Proportional Odds) Model, Generalized Ordered Logit Model, Partial Proportional Odds Model*
Gender	Aziz et al. (2013)	*Random- Parameter Severity Model, Partial Derivative Of The Outcome Probabilities*
	Celik and Oktay (2014)	*Multinomial Logit Model*
Alcohol Consumption	Moudon et al. (2011)	*Binary Logistic Regression, Ordinal Logistic Regression*
	M. Kim et al. (2017)	*Hierarchical Ordered Model*
Day Status	Rifaat et al. (2011)	*Multinomial Logit Model*
	Zhang et al. (2013)	*Multivariate Stepwise Logistic Regression*
Weather	Celik and Oktay (2014)	*Multinomial Logit Model*
	J.-K. Kim et al. (2010)	*Mixed Logit Model*
	Haleem et al. (2015)	*Random Forest Technique, Mixed Logit Model*
Traffic Sign	Moudon et al. (2011).	*Binary Logistic Regression, Ordinal Logistic Regression*
	Prato et al. (2012)	*Questionnaire Survey, Field Studies*
Speed	Ulfarsson et al. (2010)	*Multinomial Logit Model*
	Haleem et al. (2015)	*Random Forest Technique, Mixed Logit Model*
	Rankavat and Tiwari (2016)	*Questionnaire Survey, Ordered Logit Model*
Road Geometry	Zheng et al. (2019)	*Gradient Boosting Decision Tree, Convolutional Neural Network*
	Celik and Oktay (2014)	*Multinomial Logit Model*
Lighting Conditions	Islam and Burton (2019)	*Mixed Logit Model*
	Fountas et al. (2020)	*Zero-Inflated Ordered Probit Model*
Vehicle Characteristics	Albalate and Fernandez-Villadangos (2010)	*Multinomial Logistic Regression*
	Moore et al. (2011)	*Multinomial Logit Model, Mixed Logit Model*
	Fountas and Anastasopoulos (2017)	*Random Parameters Hierarchical Ordered Probit Model*
Other Entities (e.g., animal)	*C.* Chen et al. (2016)	*Hierarchical Ordered Logit Model*
	Fountas and Anastasopoulos (2018)	*Zero-Inflated Hierarchical Ordered Probit*

study and emphasized that structural environmental factors such as number of lanes and location are effective on the injury-severity.

Mohamed et al. (2013) stated that heavy vehicles, lighting conditions and arterials increase the frequency of fatal accidents. Obeng and Rokonuzzaman (2013) mentioned that similar to previous studies, vehicle type, gender, land use and road geometric features increase the severity of injury. Olszewski et al. (2015) highlighted the parameters affecting the possibility of fatal accident the most in the form of divided roads, roads with a speed limit higher than 70 km/h, roads with insufficient lighting conditions, and areas other than the residential ones. Ravishankar and Nair (2018) argued that, in addition to age and gender, driver behavior affects accident injury-severity. Sun et al. (2019) determined that alcohol use, high age, gender and bad weather conditions increase the probability of fatal accidents. P. Chen and Shen (2019) emphasized that age was in positive correlation with possible injury and that the use of helmets positively correlated with no-injury, and speed parameter showed positive correlation with possible injury and fatal.

Sivasankaran and Balasubramanian (2020) analyzed twenty factors in total including temporal, environmental, vehicle and cyclist characteristics that affect bicyclists injury-severity using logistic regression. In their work, they aimed to eliminate the heterogeneity in the data by dividing the cyclist vehicle accidents that occurred in nine years in the state of Tamilnadu, India into five classes using the latent class clustering (LCC) method. Defining bicyclists as vulnerable road users, J. Liu et al. (2020) proposed the geographically weighted ordinal logistic regression model, which takes into account spatial heterogeneity to reveal non-stationary correlations of bicycle-vehicle injury-severity. As a result, they found that cyclist age and behavior, high travel speeds and alcohol use increase the severity of the injury. Robartes and Chen (2017) analyzed bicyclist, automobile driver, vehicle, environmental, and roadway characteristics that affect the severity of the cyclist injury in single-vehicle single-bicycle accidents using an ordered probit model. Model results show that use of alcohol by the driver increases the likelihood of serious or fatal injury to the cyclist, as well as bicycle and vehicle speeds and vertical roadway grades and horizontal curves increase the bicyclist injury-severity.

In all these studies, model performance was expressed on the basis of the accuracy percentage of the classification. In addition, there are studies in the literature comparing performances of models that are used to determine severity of injury.

The performance of multilayer perceptron (MLP) and fuzzy adaptive resonance theory models in determining the injury-severity was compared by Abdelwahab and Abdel-Aty (2001) and it was emphasized that the MLP model gave better results. In the study where cluster analysis was performed, Mohamed et al. (2013) stated that the LCC is more successful compared to k-means clustering (KMC). In another study where MNL, nearest neighbour classification (NNC), SVM and random forests (RF) were compared, Iranitalab and Khattak (2017) argued that the most successful classification performance belonged to the NNC model. When comparing three machine learning-based models, RF, C4.5 and NNC, and multinomial logit statistical model, Wahab and Jiang (2019) concluded that the algorithm with the highest accuracy is RF.

Besides, there are also studies in the literature comparing hybrid models and classical models where several models are used to increase model performance. In order to classify traffic accident severity, Kunt et al. (2011) have used a genetic algorithm-pattern search (GA-PS) combined model, GA, and ANN. The study showed that the ANN model gave the most successful results in comparison with the others. Siamidoudaran and İşçioğlu (2019) emphasized that the MLP-SVM model is significantly more successful than the other models in the classification of injury-severity by the SVM, MLP and MLP-SVM hybrid model. Another comparison of a novel model called TASP-CNN with NNC, naive bayes (NB), DT, LR, gradient boosting (GB) and SVM was made by Zheng et al. (2019) and it showed that TASP-CNN performed a classification with fewer errors than other models. In the study using the combination of SVM and rough set theory (RST), Jianfeng et al. (2019) stated that the SVM-RST model was more successful than classical SVM in classifying injury-severity.

As a result of large-scale literature review, it was seen that the age (Islam and Mannering, 2006; Pour-Rouholamin and Zhou, 2016; Yan et al., 2005a, b), gender (Al-Shammari et al., 2009; Aziz et al., 2013; Celik and Oktay, 2014), alcohol consumption(M. Kim et al., 2017; Moudon et al., 2011), day status (Rifaat et al., 2011; Zhang et al., 2013), weather (Celik and Oktay, 2014; Haleem et al., 2015; J.-K. Kim et al., 2010), traffic sign (Moudon et al., 2011; Prato et al., 2012), speed (Haleem and Gan, 2015; Kadali and Vedagiri, 2013; Rankavat and Tiwari, 2016; Ulfarsson et al., 2010) and road geometry (Celik and Oktay, 2014; Karlaftis and Golias, 2002) parameters are commonly emphasized in the literature. The parameters in the results of previous studies that have had a large impact on injury-severity are given in Table 1.

It is observed that the trend of machine learning-based algorithms has increased due to the fact that these algorithms provide better results in classifying accident injury-severity. However, studies that use the fuzzy decision mechanisms used in this study to classify injuries in bicycle-vehicle accidents have not been encountered in the literature.

As a result of an extensive literature review, two fundamental research topics that motivated this study are given below.

˙First, investigation of the effect of road infrastructure, road geometry, street, accident, atmospheric and cyclist-related parameters on the classification of cyclist injury-severity, similar to other studies in the literature.

˙Second, examination of the performance of the new fuzzy decision approach described in detail in this study on the classification of cyclist injury-severity.

3. Methodology

In this section, the classic C4.5 algorithm and DT-based two hybrid-fuzzy models that are used to classify cyclist injury-severity in bicycle-vehicle accidents are presented.

3.1. Decision tree

DT algorithm, which is one of the frequently-used machine-learning based algorithms, gives successful results in classification and prediction problems (J Ross Quinlan, 1987). In Fig. 2, DT flow-chart is given. The fact that this algorithm is easy to understand, the models that are installed can be visualized and can be used in cases where the dependent variable is multi-level enables this algorithm to be preferred by decision makers (Rokach and Maimon, 2008). In the literature, Quinlan's ID3 algorithm which uses information gain (IG) for splitting criteria, forms the basis of these algorithms although various algorithms are used to determine the splitting criteria in DT (J. Ross Quinlan, 1986). In this study, the C4.5 algorithm, an improved version of the ID3

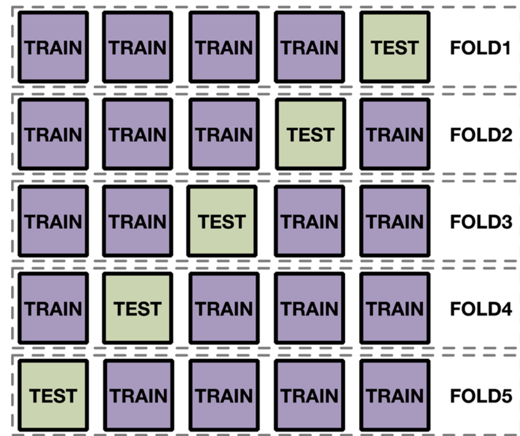

Fig. 3. 5-fold cross validation illustration.

algorithm introduced by Quinlan, was used. Differently from ID3, C4.5 algorithm determines the splitting criteria according to information gain ratio (IGR). The expressions of IG and IGR for attribute variable A and given class Y are given below (Chang and Chien, 2013; de Oña et al., 2013a, b) (Fig. 3).

$$IG(Y, A) = H(Y) - H(YA) \tag{1}$$

$$IGR(Y, A) = \frac{IG(Y, A)}{H(A)} \tag{2}$$

$$H(Y) = \sum_j p(y_j)\log_p^{(y_j)} \tag{3}$$

where $H(Y)$ is the entropy of Y, $p(y_j) = p(Y = y_j)$ is the probability of each value in the dataset. Detailed information can be found (J. Ross Quinlan, 1986; Watanabe and Rendell, 1991).

3.1.1. Importance of variables

The following equation was used (Abellán et al., 2013) for the purpose of determining the effect of independent variables on the classification of cyclist injury-severity.

$$IoV(A) = \sum_{i=1}^{h} \frac{na_i}{n} I((YA) = a_i) - (Y)) \tag{4}$$

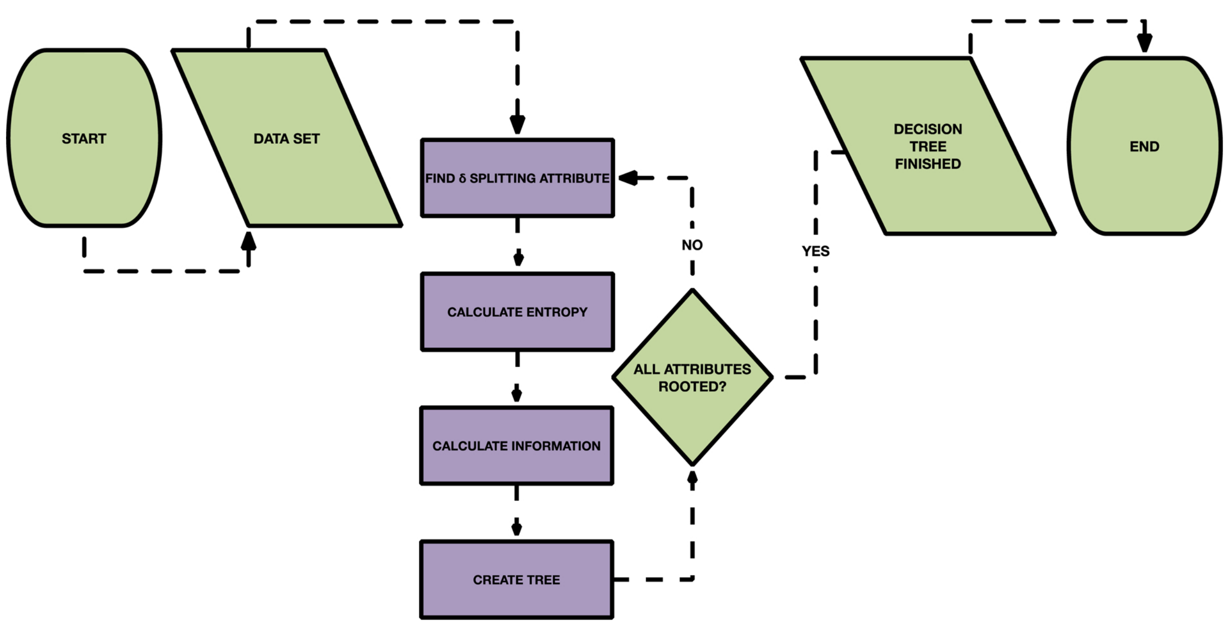

Fig. 2. DT flow-chart.

where Y is the injury-severity class, na_i is the number of situations that $A = a_i$, and n is the total situation number. IoV refers to the effect of an input parameter A on the gain of the output parameter of Y. The change of splitting criteria used in the model changes IoV. Since C4.5 algorithm uses as a splitting criterion, I, represents the IGR. According to the IoV values normalized to the 0–100 range, the parameters in the 40–60 % range were evaluated as highly effective, and the parameters over 60 % were considered to be the most effective.

3.1.2. K-fold cross validation

In machine learning techniques, it is important to ensure that the model applied to the dataset is consistent and to prevent the model from overfitting the dataset. To avoid overfitting, there are several methods in the literature that are applied to divide the dataset into parts and use some of them in the training and the rest in the test phase (Domingos, 2012; Lison, 2015; I. H. Witten et al., 2005). There are studies in the literature where K is determined as 5 and 10 (Delen et al., 2017; Rezapour et al., 2019). In this study, the K value was applied by choosing from 2 to 10 and the best model performance was obtained when K = 5.

3.2. Establishment of DT-based two fuzzy mechanisms

3.2.1. Fuzzy logic

It was the first time that Zadeh (1988) mentioned the FL approach and stated that it can be utilized FL in defining and solving complex systems and real-life problems which were ambiguous. On the other hand, FL, which can also be used in the solution of these problems, is used for the solution of problems where there is a non-linear relationship between inputs and outputs (Al-Mousa and Faza, 2019; Larimian et al., 2013).

A typical FL system (Fig. 4) consists of three phases which are fuzzification, fuzzy inference engine, and defuzzification (Mehran, 2008). In the first phase, actual data is expressed in qualitative values or linguistic terms (Khosravanian et al., 2016). Moreover, its membership function type and the number of fuzzy sets are determined and these fuzzy sets take any number between 0 and 1 (Ross, 2005). The membership function types are triangular, trapezoidal and Gaussian, etc. In this study, the most preferred triangular and trapezoidal membership functions in the literature were used (Rahim, 2017). In the second stage, the fuzzy rule base, which is expressed in IF-THEN form, is created according to all possible fuzzy relationships between inputs and outputs and the fuzzy inference system (FIS) applies judgment to compute the fuzzy outputs. There are two types of FIS that Mamdani-type and Sugeno-type(Ross, 2005). Mamdani-type FIS, which was proposed by Mamdani and Assilian (1993), is utilized for the establishment of DT-based two fuzzy mechanisms. Finally, the defuzzification phase converts fuzzy outputs into crisp values.

The reason the FL approach was used in this study is that the bicycle-vehicle accident severity class corresponding to the combination of conditions in the form of attribute-value pairs in the dataset is ambiguous, vague and unclear. In other words, although all conditions related to the attributes are the same, different bicycle-vehicle accident injury-severity can occur, which causes complexity and confusion in predicting accident injury-severity. In this study, it is thought that this complexity can be characterized by a fuzzy methodology. Hosseinpour et al. (2013) supported the argument that the relationship between traffic accidents and explanatory variables is more complex, its nature uncertain. The results of the study showed that neuro-fuzzy systems are more effective in dealing with uncertain human behavior due to the flexible and adaptable abilities when compared to Poisson, negative binomial, and non-linear exponential regression models. Similarly, Meng et al. (2009) developed a FL-based model, creating 41 fuzzy rules for urban traffic accident prediction. They reported a positive correlation between the observed and predicted data proving the feasibility of the proposed FL method.

3.2.2. DT-based converted FL prediction model

The DT-based converted FL (DT-CFL) model was used by Omid (2011) to classify open and closed-shell pistachio nuts. The basis of this study is the creation of an expert system that combines traditional DT learning techniques (C4.5 algorithm) and the superior capabilities of the FL. Based on the model results, Omid (2011) emphasized that the DT-based FL shows high performance. In addition to this study, Sakthivel et al. (2010) used DT and rough sets to create rules from the statistical features of the mono-block centrifugal pump. In their study, the rules obtained from DT were converted into if-then rules in the FL. The study results revealed that the overall classification accuracy achieved by the DT-based FL is better than the rough set-based fuzzy hybrid system. The basic working principle of the DT-CFL model is based on converting the obtained DT rules into if-then rules in FL. The working principle of the DT-CFL model can be expressed in four steps as follows:

Step 1. Data pre-processing and building DT
Step 2. Converting the obtained DT rules into if-then rules in FL
Step 3. Creating the DT-based converted FL
Step 4. Prediction of bicycle-vehicle accident injury-severity

In DT which consist of nodes, branches, and leaves, every path from leaves to the main node is interpreted as a rule. All the rules obtained from the DT are converted to if-then rules in FL. DT-CFL hybrid model, together with the establishment of DT-based two fuzzy mechanisms applies judgment to compute the fuzzy outputs. Besides, the fuzzification, FIS, and defuzzification phases gave in Section 3.2.1 are followed in DT-CFL hybrid model and the model performance is evaluated by comparing with the actual data as well as the other predict models.

3.2.3. DT-based revised FL prediction model

In the DT-based revised FL (DT-RFL) model, the membership functions of each input are revised by using the effect values determined by the DT algorithm. Instead of being used to define each effect value

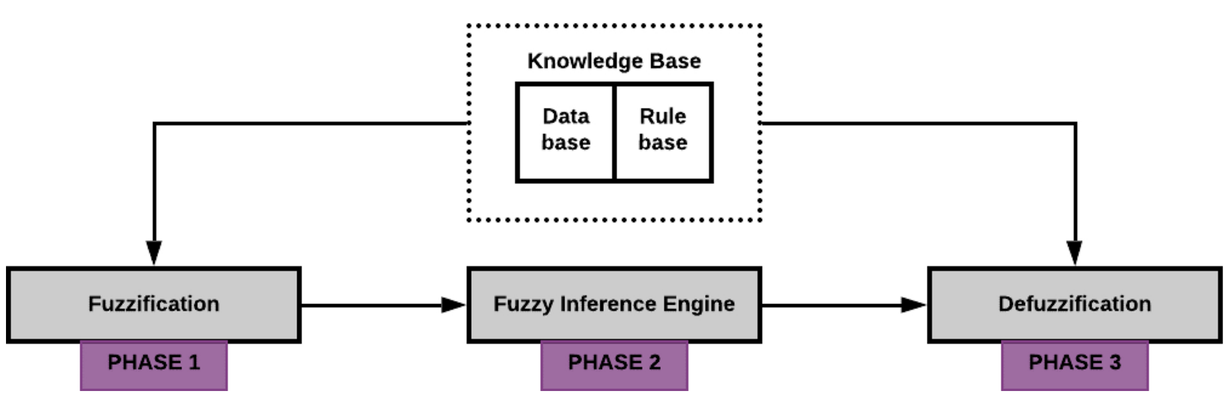

Fig. 4. General structure of a typical FL system.

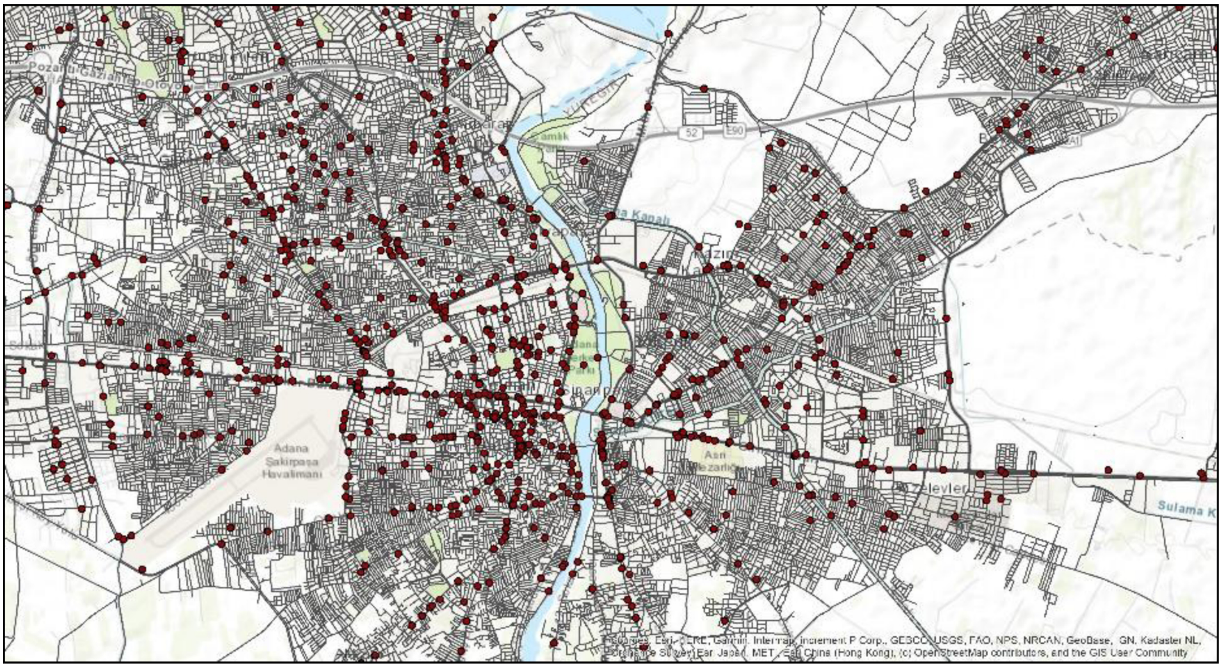

Fig. 5. Crashes in the five-year period of 2013-2017.

obtained from the DT algorithm as the rule weights in the fuzzy mechanism, these values were utilized to revise the membership functions of the inputs for determining the fuzzy sets of outputs and to obtain the results of IF-THEN conditions. There are many classification or prediction studies in the literature that define rule weights and optimize rule weights using learning algorithms such as artificial neural networks, ANFIS, FL-DT, etc.

The fuzzy rule weights, which could be called "measure of importance or influence", can be applied either to the entire or consequent part of the related fuzzy rule. The FL designer of MATLAB software enables the rule weight to be applied to complete the related rule. However, according to Nauck and Kruse (1998), defining rule weights in the FL model gives rise to problems such as “non-normal fuzzy sets and different representations for the same linguistic term”. To deal with these problems, they emphasized that rather than defining the rule weight, an equivalent effect could be achieved by making modifications membership functions in the antecedents. Similarly, C. Wang (2015) has created fuzzy sets and fuzzy rules in designing a timing system for an automotive engine by revising the membership functions in the antecedents rather than using the rule weights.

In the DT-RFL model, the antecedent score of each rule is obtained by multiplying the fuzzy set score by the weight of the corresponding input variable and adding these contributions from each input parameter that creates the rule. For example; When 'a' and 'b' are assigned as weights to inputs A and B, if “If the fuzzy set score of A is Low, represented by "1″ and the fuzzy set score of B is Medium, represented by "2″, the antecedent score of the rule is calculated as follows:

$$\text{Antecedent Score:}(1 \times a) + (2 \times b) \tag{5}$$

Finally, the antecedent score of each rule is expanded over the range of bicycle-accident severity, which is called an extended score. The boundaries of the fuzzy sets of the output and the consequents of the fuzzy rules are determined according to the extended score. The working principle of the DT-CFL model can be expressed in five steps as follows:

Step 1. Data pre-processing and building DT

Step 2. Obtaining the impact values for each parameter from DT algorithm

Step 3. Revising the membership function of each input according obtained the effect values

Step 4. Creating the DT-based revised FL model

Step 5. Prediction of bicycle-vehicle accident injury-severity

There are studies in the literature where DT and FL models are used in traffic safety and transportation engineering (Gaber et al., 2017; H. Wang et al., 2011). However, the DT-RFL model presented in this study is a completely new approach that assigns the coefficients obtained according to IGR, which determines the splitting criteria in C4.5 algorithm, to the input variables by weight and determines the result of the IF-THEN rules according to the expanded score obtained from the added weights of the inputs without expert opinion. The DT-CFL model, on the other hand, was created by defining all the conditions occurring from root to leaf according to the splitting criteria of the C4.5 algorithm as a rule to FL. Although this model has been previously used by Sakthivel et al. (2010) and Omid (2011), the use of the model in traffic safety and accident analysis is a new approach.

4. Data description

The dataset was obtained from the General Police Directorate, Turkish Statistical Institute and hospital records and includes 970 bicycle-vehicle crashes in Adana city, East Mediterranean region, in a five-year period between the years 2013–2017. Since 11 pieces of accident data did not contain some selected parameters, these accidents were removed from the dataset and the remaining 959 data were examined in the models. Crashes in the dataset were mapped with ArcGIS (ArcView, 1996) and are given in Fig. 5. The parameters related with road infrastructure and geometry, environmental, safety, accident, atmospheric and cyclist have been defined as independent variables which are unique for every single accident. The dependent variable has been defined as cyclist injury-severity which was divided into three levels such as fatal (FTL), injury (INJ), no-injury (NOI). The distribution of FTL, INJ, and NOI levels in the whole dataset is 4.48 %, 70.17 %, and 25.33 %, respectively.

Detailed information such as description, statistic and type about dependent and independent variables used in this study are given in Table 2.

Table 2
Descriptive information of dependent and independent variables.

Variable			Type	Description	Summary statistics
Independent variables	Road infrastructure and road geometry related	Pavement type	Dummy	Asphalt = 1; Else = 0	Asphalt = 72.67(%); Else = 27.32(%)
		Junction type	Categorical	1 = Three leg; 2 = Four leg; 3 = Other; 4 = No-junction	No-junction = 39.52(%); Three leg = 27.94(%); Four leg = 14.38(%); Other = 18.14(%)
		Curve	Dummy	Presence of curve = 1; Else = 0	Curve = 65.38(%); Else = 34.72(%)
		Slope	Dummy	Presence of slope = 1; Else = 0	Slope = 6.67(%); Else = 93.32(%)
		Shoulder	Dummy	Presence of shoulder = 1; Else = 0	Shoulder = 11.36(%); Else = 88.63(%)
		Road type	Categorical	1 = Divided; 2 = One-way; 3 = Two-way	Divided = 66.42(%); One-way = 6.67(%); Two-way = 26.90(%)
	Environment and Safety related parameters	Traffic sign	Dummy	Presence of traffic sign = 1; Else = 0	Traffic sign = 48.69(%); Else = 51.30(%)
		Traffic light	Dummy	Presence of traffic light = 1; Else = 0	Traffic light = 40.04(%); Else = 59.95(%)
		Street light	Categorical	Presence of street light = 1; Else = 0	Street light = 51.92(%); Else = 48.07(%)
		Sidewalk	Dummy	Presence of sidewalk = 1; Else = 0	Sidewalk = 81.12(%); Else = 18.87(%)
		Pedestrian crossing	Dummy	Presence of pedestrian crossing 1; Else = 0	Ped. Crossing = 30.96(%); Else = 69.03(%)
	Crash related parameters	Crash type	Categorical	1 = Rear-end collision; 2 = Side collision; 3 = Head-on collision; 4 = Other	Rear-end collision = 11.47(%); Side collision = 65.69(%); Head-on collision = 9.17 (%); Other = 13.66(%)
		Impact side (bicycle)	Categorical	1 = Front; 2 = Rear; 3 = Left; 4 = Right	Front = 63.92(%); Rear = 6.98(%); Left = 15.84(%); Right = 13.24(%)
		Damage extent	Categorical	1 = Non; 2 = Light; 3 = Functional; 4 = Heavy	Non = 13.76(%); Light = 42.25(%); Functional = 30.03(%); Heavy = 10.94(%)
		Vehicle movement	Categorical	1 = Move straight; 2 = Lane change; 3 = Turn right; 4 = Turn left; 5 = Other	Move straight = 79.24(%); Lane change = 3.96(%); Turn Right = 2.91(%); Turn left = 9.38(%); Other = 4.48(%)
	Cyclist related parameters	Cyclist gender	Categorical	1 = Male; 2 = Female	Male = 93.74(%); Female = 6.25(%)
		Cyclist education	Categorical	1 = Grade; 2 = High; 3 = Higher-education	Grade = 69.13(%); High = 25.54(%); Higher-education = 5.31(%)
	Additional parameters	Weather	Dummy	If weather is clear = 1; Else = 0	Clear = 93.01(%); Else = 6.98(%)
		Day of week	Categorical	1 = Monday; 2 = Tuesday; … 7 = Sunday	Monday = 13.86(%); Tuesday = 14.49(%); Wednesday = 16.16(%); Thursday = 16.26(%); Friday = 14.07(%); Saturday = 13.45(%); Sunday = 11.67(%)
Dependent variable		Cyclist Injury-severity	Categorical	1 = No-injury; 2 = Injury; 3 = Fatal	No-injury = 25.33(%) Injury = 70.17(%); Fatal = 4.48(%)

5. Results and comparison of two hybrid fuzzy-decision models and C4.5 algorithm

In this section, the prediction results for each model and comparison of two hybrid fuzzy-decision models and C4.5 algorithm are given. Additional details about creation of C4.5 algorithm-based models, including two different fuzzy approaches, are also presented in this section.

5.1. C4.5 algorithm results

Weka software was used to build the C4.5 algorithm (I. Witten and Frank, 2005). While applying the C4.5 algorithm to the dataset, 5-fold cross-validation procedure was used to prevent overfitting. In each fold, the dataset was divided into five parts and one of that parts was selected for usage in the test and the rest of them were used in training. As seen in Fig. 6, DT created with C4.5 algorithm consists of 70 nodes and 48 terminal nodes. While the DT was being built, the C4.5 algorithm created a new branch for each variable. At this stage, the algorithm used IGR as the splitting criteria, which are mentioned in section 3.1. The

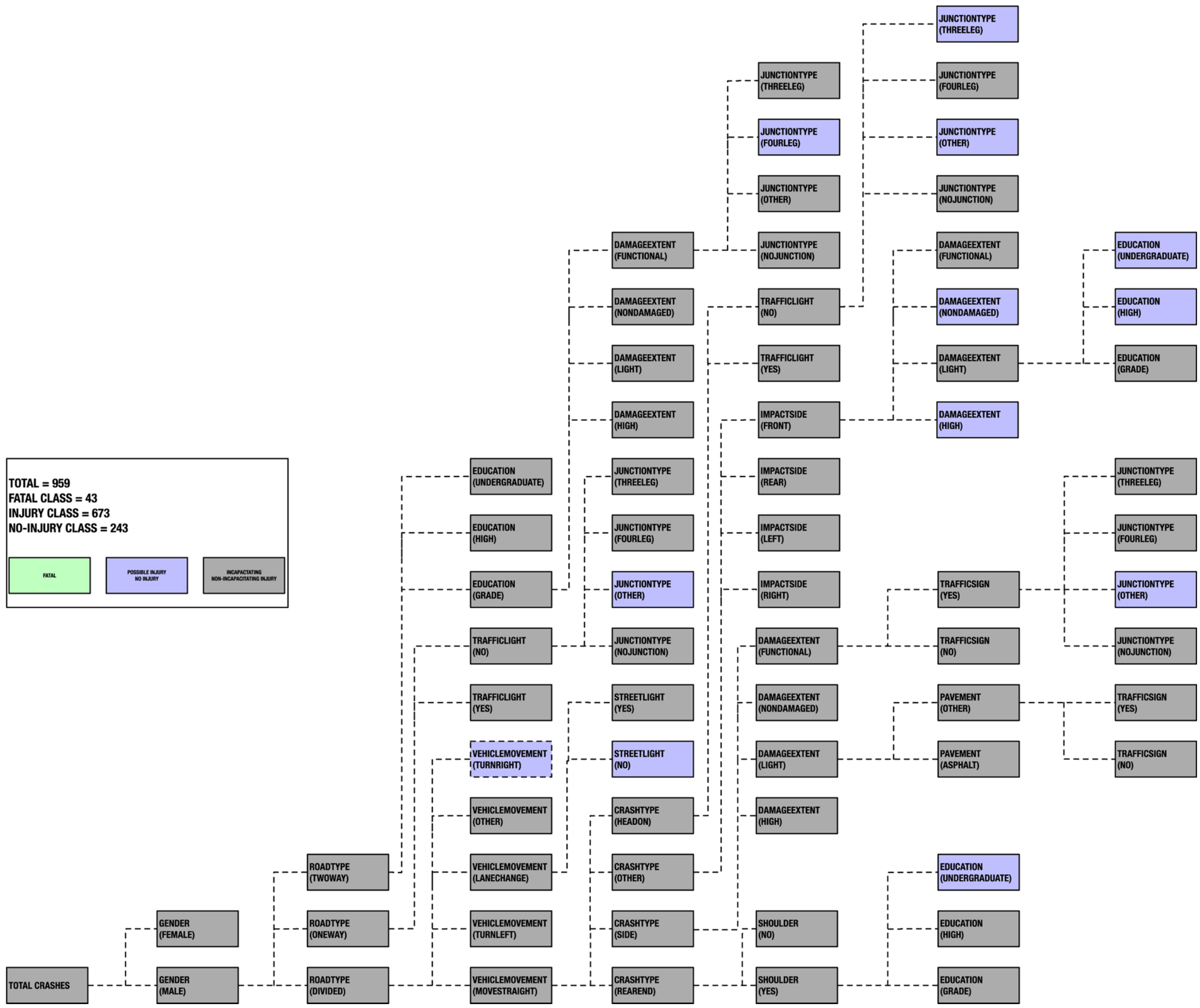

Fig. 6. DT Model created with C4.5 algorithm.

classification accuracy of the DT model was 67.46 %. When the model results are examined, DT model correctly classifies 647 cases for a total of 959 cases. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were obtained at 0.39 and 0.28, respectively, which are similar to some of the previous model performances (Fountas et al., 2018a, b; J. Lee et al., 2020; Mukoko and Pulugurtha, 2019; Ye and Lord, 2011) and better than some ((Kunt et al., 2011; Olutayo and Eludire, 2014). Even though 48 rules were created while going from the starting node to each terminal node, only 14 of these rules (Table 3.) met the threshold value of Support, Probability and Population (Section 5.1.1.) values.

5.1.1. Rules extraction

To better understand the potential useful information of DT structure, the rule extraction method, which has been applied many times in previous studies (Agrawal et al., 1993; Pande and Abdel-Aty, 2009), was used. DT rules are created in a logical conditional structure. For a set of different levels of X variables, the Y dependent variable turns into a single class. For example:

IF ((**GENDER** (MALE), **CRASHTYPE** (HEAD-ON), **CYCLISTEDUCATION** (HIGH)), → THEN (**INJURYSEVERITY** (NOI)).

All these sets of IF rules are called the antecedent and THEN is called the consequent. In order to evaluate the factors affecting the severity of injury in bicycle-vehicle accidents, the 3 criteria described below were used to extract the most important IF → THEN logical conditions.

Support (S): The support of the entire rule.

Probability (P): The percentage of situations where an IF-THEN rule is accurate.

Population (Po): The support of the antecedent of the rule. Percentage of occurrence of IF conditions in an IF-THEN logical-conditional state. (Pande and Abdel-Aty, 2009).

In previous studies, the selection of threshold values differs. While examining the rules, it is acceptable the rules have high S and P values. Pande and Abdel-Aty (2009) used 0.009 for S and 0.1 for P values, Montella et al. (2012) used lower values (S = 0.001, P = 0.01) for their study. In another study, de Oña, López, and Abellán (2013) and Abellán et al. (2013) defined S = 0.006 and P = 0.06 values for threshold value. In this study, to extract significant rules, threshold values for S and P are determined as 0.0075, 0.75 for INJ class and 0.006, 0.60 for NOI class respectively. When the rules obtained by considering the threshold values are examined, it is seen that the rule, which has the highest S (8.9 %) and Po (11.5 %) values (rule 7 in Table 3.), gives the result of INJ. This result shows that the combination of male cyclists, two-lane roads, vehicles going straight (presumably in the main arterials), side collision and asphalt pavement has a significant effect on accidents with injury. Almost all the important rules obtained consist of male cyclists and are in line with the road type parameter. It can be seen that 10 of the 14 rules materialized under divided road conditions, 3 of which resulted in NOI and 7 of which resulted in INJ. The rules which have the highest S and Po values for both NOI and INJ classes include side-collision and vehicle-going-straight conditions. This is important because this finding confirms the results of previous studies (Behnood and

Table 3
Information about the significant rules according to C4.5 algorithm.

R. N.	Rules (IF…)	THEN	S (Support)	P (Probability)	Po (Population)
2	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLEMOVEMENT** (MOVESTRAIGHT) AND **CRASH TYPE** (REARENDCOLLISION) AND **SHOULDER** (YES)	**INJ**	135	9285	145
3	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (MOVESTRAIGHT) AND **CRASHTYPE** (REARENDCOLLISION) AND **SHOULDER** (NO) AND **CYCLIST EDUCATION** (GRADE)	**INJ**	364	7776	468
6	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLEMOVEMENT** (MOVESTRAIGHT) AND **CRASHTYPE** (SIDECOLLISION) AND **DAMAGE EXTENT** (NONDAMAGED)	**INJ**	333	8003	416
7	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLEMOVEMENT** (MOVESTRAIGHT) AND **CRASHTYPE** (SIDECOLLISION) AND **DAMAGE EXTENT** (LIGHT) AND **PAVEMENT** (ASPHALT)	**INJ**	896	7747	1157
9	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (MOVESTRAIGHT) AND **CRASH TYPE** (SIDECOLLISION) AND **DAMAGE EXTENT** (LIGHT) AND **PAVEMENT** (OTHER) AND **TRAFFIC SIGN** (NO)	**NOI**	104	6250	166
18	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (MOVESTRAIGHT) AND **CRASH TYPE** (OTHER) AND **IMPACT SIDE** (FRONT) AND **DAMAGE EXTENT** (LIGHT) AND **CYCLIST EDUCATION** (HIGH)	**NOI**	072	7779	092
25	**GENDER** (MALE) AND **ROAD TYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (MOVESTRAIGHT) AND **CRASH TYPE** (HEADONCOLLISION) AND **TRAFFIC LIGHT** (YES)	**INJ**	114	9166	125
30	**GENDER** (MALE) AND **ROAD TYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (LANECHANGE) AND **STREET LIGHT** (YES)	**INJ**	114	8461	135
32	**GENDER** (MALE) AND **ROAD TYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (TURNRIGHT)	**NOI**	071	6366	111
33	**GENDER** (MALE) AND **ROADTYPE** (DIVIDED) AND **VEHICLE MOVEMENT** (TURNLEFT)	**INJ**	427	8039	531
41	**GENDER** (MALE) AND **ROADTYPE** (TWOWAY) AND **CYCLIST EDUCATION** (GRADE) AND **DAMAGE EXTENT** (LIGHT)	**INJ**	761	8295	917
42	**GENDER** (MALE) AND **ROADTYPE** (TWOWAY) AND **CYCLIST EDUCATION** (GRADE) AND **DAMAGE EXTENT** (FUNCTIONAL) AND **JUNCTION TYPE** (THREELEG)	**INJ**	166	7619	218
45	**GENDER** (MALE) AND **ROADTYPE** (TWOWAY) AND **CYCLIST EDUCATION** (GRADE) AND **DAMAGE EXTENT** (FUNCTIONAL) AND **JUNCTION TYPE** (NOJUNCTION)	**INJ**	166	8421	198
47	**GENDER** (MALE) AND **ROAD TYPE** (TWOWAY) AND **CYCLIST EDUCATION** (HIGH)	**INJ**	417	8163	510

Mannering, 2017; Kaplan et al., 2014) and shows that side-collision crashes on arterials, where vehicles move straight, is an important case for safety researchers to consider. 5 rules result in INJ and contain the cyclist-education parameter, 4 of which occurred under the condition of low education. This finding confirms the positive correlation between low education level and accident severity as concluded in previous works (Clifton et al., 2009; Fountas et al., 2018a, b; Waseem et al., 2019).

Fig. 7 illustrates the normalized importance of variables according to C4.5.IGR split criteria and indicates that variables such as gender (100 %), damage-extent (74.32 %) and road-type (74.05 %) with a normalized value above 60 % have the greatest impact on cyclist injury-severity in bicycle-vehicle accidents. However, such a high value of importance for a factor with highly heterogeneous effects like gender can be questionable. Moreover, this ambiguity is further extended by the fact that almost 94 % of the cyclists are male. This may induce further bias to the classification, as the ratio of the control group (woman cyclists) is very low. When the injury-severity predicted in the presence of unobservable heterogeneity is as it is in the DT model, the distribution of parameter values between observations can become an important specification problem. This problem can be expressed as biased parameter estimates and marginal effects. However, this result may serve as a proxy of other effects, which cannot be captured through the specific dataset (mainly effects of socio-demographic and behavioral traits). It should be noted that regulating the heterogeneity will have a positive effect on the estimation results. Also, focusing on how heterogeneity affects the variance of parameter distribution and dealing with unobserved heterogeneity in traffic engineering are potentially important topics. (Guo et al., 2018; Heydari, 2018; Mannering et al., 2016). Pavement-type, accident-type and vehicle-movement parameters are also evaluated to be important since their normalized values were more than 40 %. These results substantiate previous studies. Many researchers (Abellán et al., 2013; Behnood and Mannering, 2017; De Oña, López, Mujalli, et al., 2013; Kockelman and Kweon, 2002) have noted that the *crash-type* parameter is the most important factor in identification of injury-severity. *Gender* is also considered as a highly affective factor in crash injury-severity (Behnood and Mannering, 2017; Islam and Hossain, 2015; Obeng, 2011; Rash-ha Wahi et al., 2018), i.e. Men more likely tend to get involved in an accident because they behave in a more risky manner than women in traffic. The results of some previous works supported the findings of this study and pointed out that the *vehicle-movement* parameter is another significant factor that related to injury-severity in bicycle-vehicle accidents (Pai and Saleh, 2008; Tang et al., 2020; Wen et al., 2020; Zahabi et al., 2011). Another two important and interrelated factors are the *road-type* (Kaplan et al., 2014; Wen et al., 2020; Zahabi et al., 2011) and *pavement-type* (Li et al., 2018; Vilaça et al., 2019). The studies stressed that the roads with better pavement conditions (Asphalt roads, concrete roads) and roads with over two lanes are more dangerous for cyclists. This may be related to higher quality roads encouraging car drivers to go faster. Contrary to many studies in the literature (Abellán et al., 2013; Hagel et al., 2015; Kweon and Lee, 2010) the effect of *streetlight, slope, traffic sign* parameters on cyclist injury-severity in bicycle-vehicle accidents was found non-significant in this study. In addition, variables such as *speed-limits*, *age*, *traffic exposure parameters*, etc., which have been found to affect the accident cyclist injury-severity as a result of previous studies (Adanu et al., 2020; Clarry et al., 2019; P. Liu and Fan, 2020) were not examined in this study since they were not included in the dataset.

5.2. DT-CFL model results

DT-CFL model has provided a fuzzy reasoning possibility in the prediction of bicycle-vehicle accident cyclist injury-severity (Fig. 6). This ability of the model is supported by the learning techniques of DT. In other words, the rules obtained from the C4.5 algorithm were converted to fuzzy rules in the DT-CFL model.

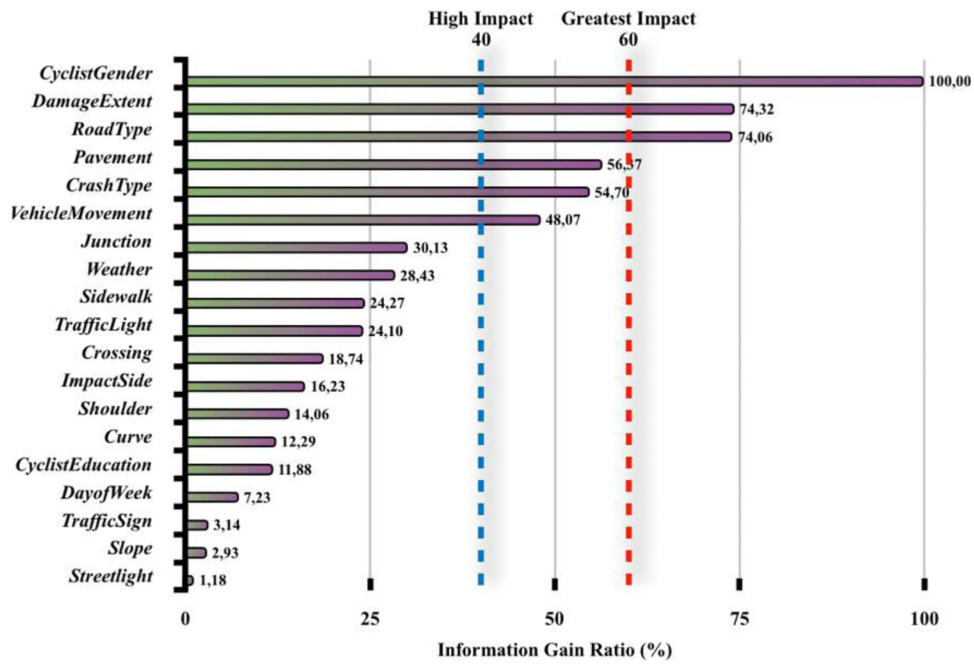

Fig. 7. Importance of variables according to IGR.

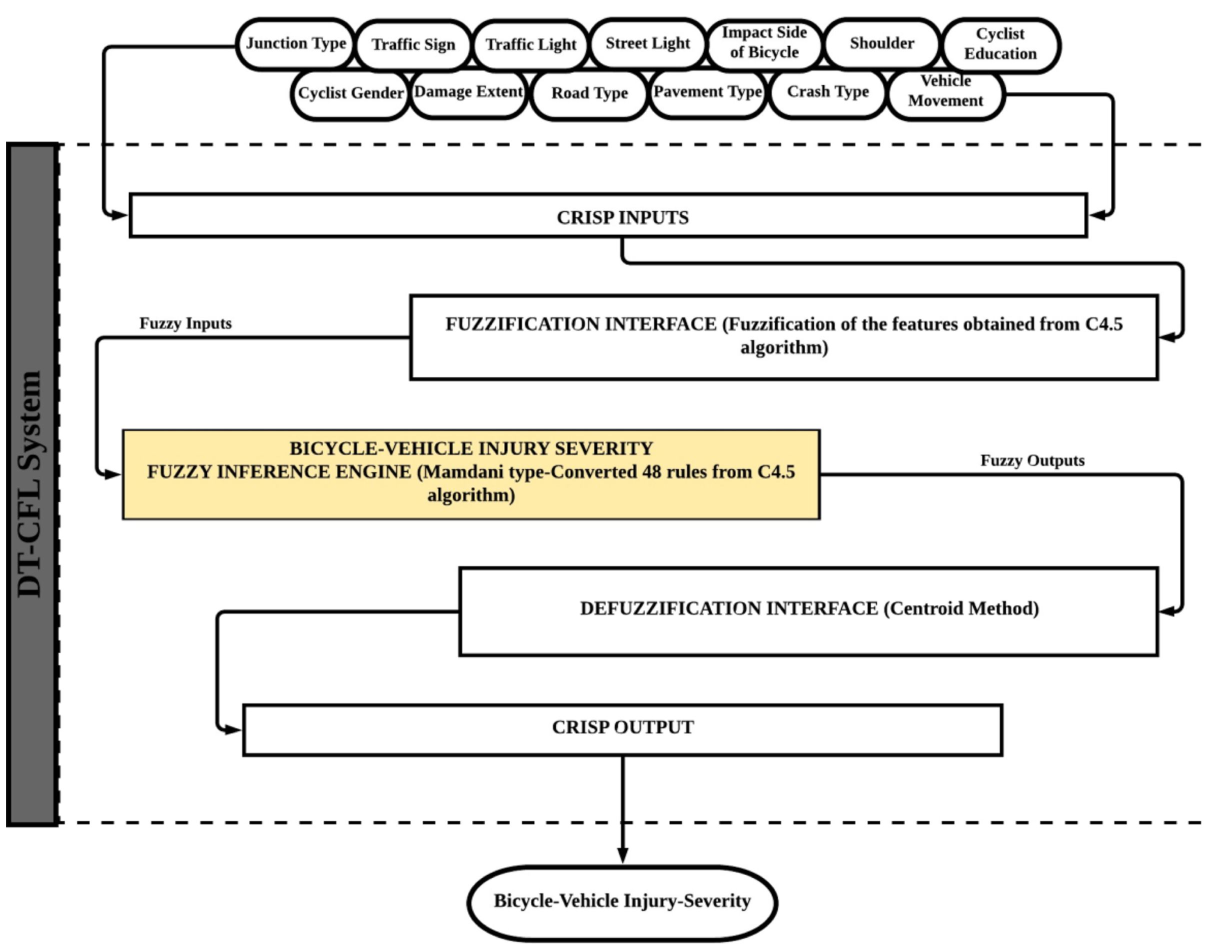

Fig. 8. The general structure of DT-CFL model.

As seen in Fig. 8, the model included thirteen inputs and one output determined according to the C4.5 algorithm's splitting criteria. The inputs are parameters that affect the injury-severity defined by means of DT experience. These inputs could be defined as cyclist gender and education, damage extent, road and pavement type, vehicle movement, shoulder, impact side of the bicycle, street and traffic light, traffic sign

Table 4
The membership function properties of the DT-CFL model.

Inputs	Membership functions types	Number of membership functions	Ranges
Cyclist Gender	Triangular	2	All membership function range is [0,1]
Damage Extent	Trapezoidal	4	
Road Type	Triangular	3	
Pavement Type	Triangular	2	
Crash Type	Trapezoidal	4	
Vehicle Movement	Trapezoidal	5	
Junction Type	Trapezoidal	4	
Traffic Sign	Triangular	2	
Traffic Light	Triangular	2	
Street Light	Triangular	2	
Impact Side of Bicycle	Trapezoidal	4	
Shoulder	Triangular	2	
Cyclist Education	Trapezoidal	3	

and junction type. The details of membership functions such as range, numbers, types are given in Table 4 for the variables converted from the DT in the prediction of bicycle-vehicle accident cyclist injury-severity, which is expressed by the inputs. In this paper, based on the C4.5 algorithm's splitting criteria, the form of membership functions and their boundaries were determined.

Additionally, degrees of the membership functions of each input, which are given in Table 4, were expressed in linguistic terms by a set of values. In other words, to establish membership functions for each input variable, a value between zero and one was given depending on whether the data is categorical and/or ordered. For example, the damage-extent variable in the accident data was defined sets of degree: non-damaged, light-damaged, functional-damage, and heavy-damage. However, in the not-ordered features such as whether the coating type is asphalt or whether the cyclist gender is female or male, etc. "1″ has been assigned to the cluster with the highest number of accident data (asphalt, male, etc.) according to the number of accidents in the data set (Fig. 9).

In the general structure of a typical FL system (Fig. 4), after the fuzzification phase, a total of 48 rules with the IF-THEN structure were developed according to rule-based Mamdani method, considering the importance of DT rule base (knowledge-based). These rules covered all the conditions between root and leaf obtained from DT without expert opinion. Some of the IF-THEN rules of DT-CFL are presented in Table 5.

According to the working principle of the DT-CFL model, the last step is to predict the injury-severity level using to obtain crisp outputs. The defuzzification systems aim to produce crisp (non-fuzzy) outputs using the fuzzification outputs of the model. There are many defuzzification methods. In this study, centroid method, which is the most preferred technique in the literature, was used.

During the defuzzification step, due to availability of a large number of the actual bicycle-vehicle accident data, the classification results were obtained from the MATLAB FL workspace instead of the MATLAB defuzzification monitor (Sivanandam et al., 2007). The result of the DT-CFL model is a numerical value (non-fuzzy output) that represents cyclist injury-severity level ranging from one to three. The non-fuzzy output is the cyclist injury-severity which can be categorized into three levels: no-injury (1–1.5); injury (1.5–2.5); fatal (2.5–3).

5.3. DT-RFL model results

The DT-RFL model (See in Fig. 10) is a novel hybrid fuzzy mechanism that can establish fuzzy reasoning with the support of the C4.5

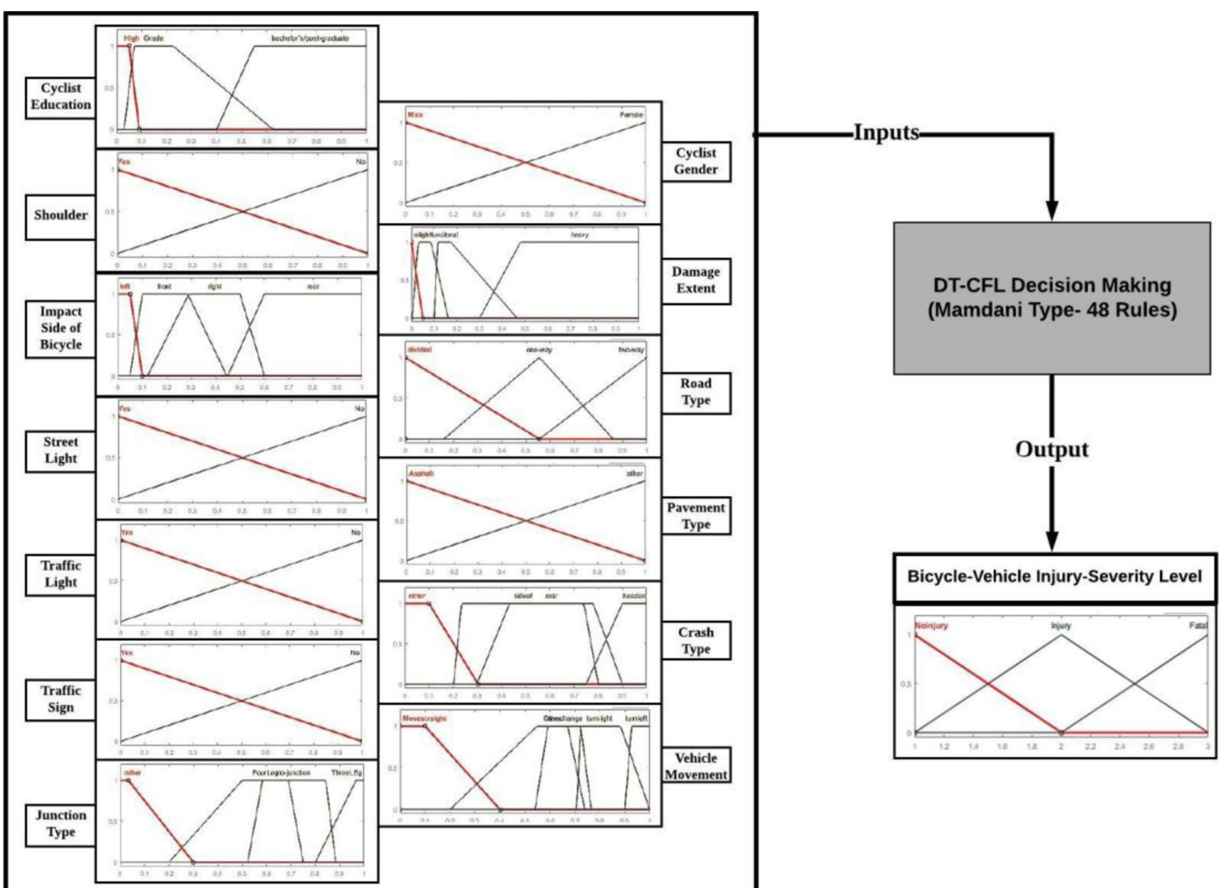

Fig. 9. The detailed structure of DT-CFL model.

Table 5
Some of the IF-THEN rules of DT-CFL.

Rule Number	Rule
19	IF ((**GENDER** (MALE), **ROADTYPE** (DIVIDED), **VEHICLE MOVEMENT** (STRAIGHTMOVEMENT), **CRASHTYPE** (OTHER), **IMPACT SIDE OF BICYCLE** (FRONT), **DAMAGE EXTENT** (LEIGHT), **CYCLIST EDUCATION** (HIGHEREDUCATION)→THEN (**NO-INJURYSEVERITY** (NOI)).
10	IF ((**GENDER** (MALE), **ROAD TYPE** (DIVIDED), **VEHICLE MOVEMENT** (STRAIGHTMOVEMENT), **CRASH TYPE** (SIDE-OF), **DAMAGE EXTENT** (FUNCTIONAL), **TRAFFIC SIGN** (YES), **JUNCTION TYPE** (THREE-LEG) → THEN (**INJURY SEVERITY** (INJ)).
46	IF ((**GENDER** (MALE), **ROADTYPE** (TWO-WAY), **CYCLIST EDUCATION** (GRADE), **DAMAGE EXTENT** (HEAVY) → THEN (**INJURY SEVERITY** (INJ)).

DT-CFL model are presented in Fig. 9 (Section 3.2.3. Step 4).

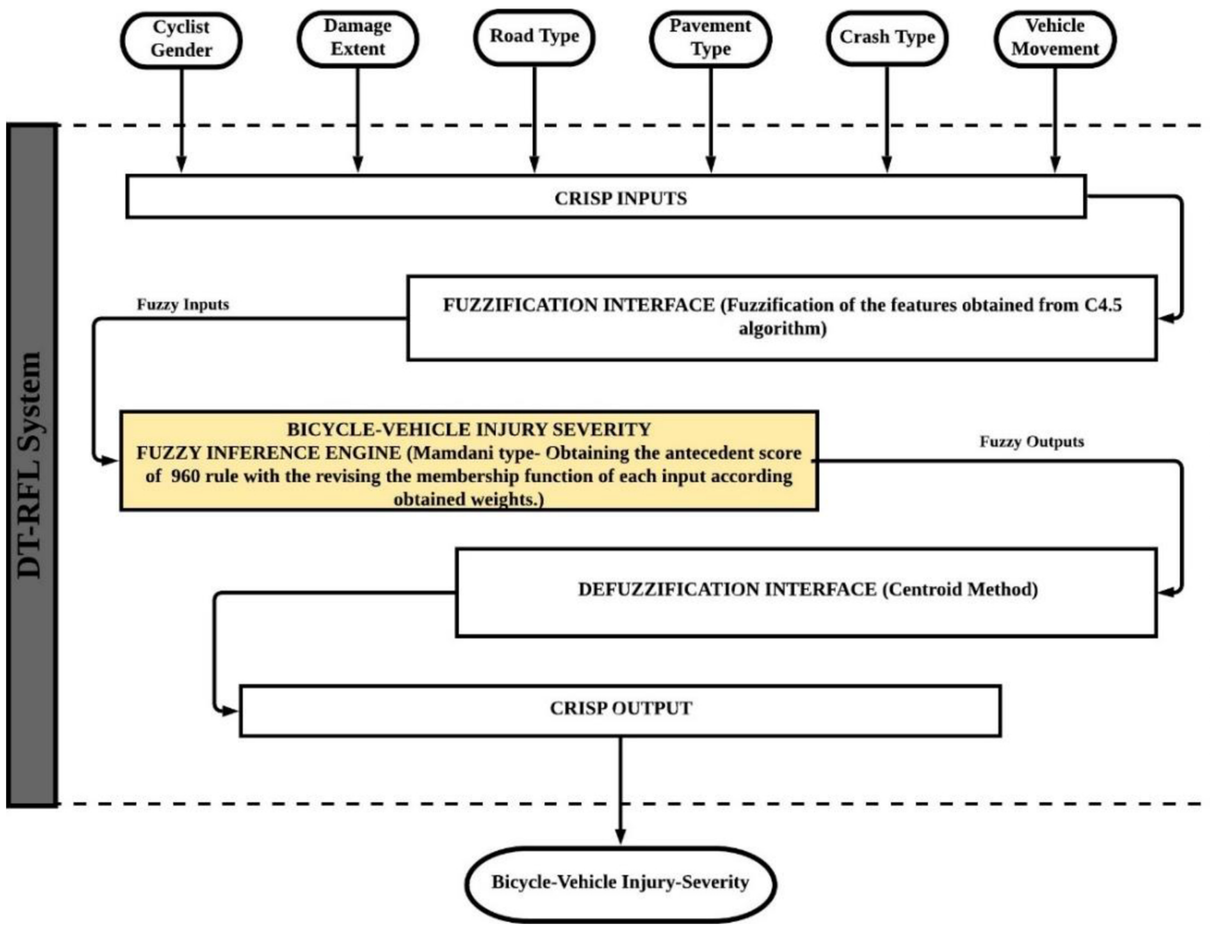

Fig. 10. The general structure of DT-RFL model.

algorithm (Section 3.2.3. Step 1). In this hybrid fuzzy mechanism, unlike the DR-CFL model, six parameters with the highest impact (net value above 35 %) on cyclist injury-severity as a result of the IGR are defined as input variables. In this model, also the antecedent of membership functions of the input parameters was revised according to the weights obtained from IGR in Table 1 (Section 3.2.3. Step 2). Besides, all combinations determined according to the number of membership functions of the input parameters such as cyclist gender, damage extent, road type, pavement type, crash type, and vehicle movement were defined as "IF-THEN" structure in the DT-RFL model.

In the model, a total of $2 \times 4 \times 3 \times 2 \times 4 \times 5 = 960$ IF-THEN rules were created. The antecedent score of each rule was calculated using Eq. 1, and then an extended score of 1–3 of each rule was determined. While defining the "THEN" part of the rule, the extended score of the fuzzy rule was utilized. The result of the extended score: 1–1.5 are defined in "no-injury", $1-5$-2.5 in "injury" and 2.5–3 "fatal" in the "THEN" part of the fuzzy rule (Section 3.2.3. Step 3).

The process of obtaining the antecedent and extended score is further clarified in Table 6. The association of the fuzzy revised rules is the process of a hybrid decision-making mechanism with bases on strong associations with the fuzzy rule base and DT. Nevertheless, the Mamdani rule-based fuzzy inference engine in the DT-RFL decides or draws conclusions without the expert opinion of the number of observations that the rule combination should contain and consequents of the rules.

The membership function properties of the DT-RFL model are given in Table 7. DT-RFL model where the membership functions and IF-THEN rules are detailed in the first three steps are presented in Fig. 11 (Section 3.2.3. Step 4).

Finally, the bicycle-vehicle injury-severity level was utilized as an output fuzzy set (among one to three levels). It created a rule base by revising the antecedent of membership functions of the input parameters according to the weights obtained from IGR. The rule base that includes 960 fuzzy rules was created to represent the association between inputs and output. Using these 960 fuzzy rules with the Mamdani centroid defuzzification method, non-fuzzy outputs to predict the injury-severity level were obtained. During the defuzzification step, the classification results were obtained from the MATLAB FL workspace.

Table 6
The process of obtaining the antecedent score of 960 rule by revising the membership function of each input according obtained weights.

Variables	Rule 1	Rule 2	Rule 3
Damage Extent (0,016,058)	NON	FUNCTIONAL	HEAVY
Fuzzy Set Score	0	06,667	1
Cyclist Gender (0,021,606)	FEMALE	MALE	MALE
Fuzzy Set Score	0	1	1
Pavement Type (001,218)	ASPHALT	ASPHALT	ASPHALT
Fuzzy Set Score	1	1	1
Road Type (0,016,001)	ONE-WAY	TWO-WAY	TWO-WAY
Fuzzy Set Score	0	0,5	0,5
Crash Type (0,011,819)	HEAD-ON	REAR-END	SIDE
Fuzzy Set Score	0	03,333	1
Vehicle Movement (0,010,387)	LANECHANGE	TURN-RIGHT	TURN-LEFT
Fuzzy Set Score	025	0	075
Antecedent Score	001,478	005,643	007,745
Extended Score	133,564	228,179	275,929
"THEN" Part of the Rule	133 < 1,5 so **NOI**	228 < 2,5 so **INJ**	275 > 2,5 so **FTL**
Crash Type	Trapezoidal	4	
Vehicle Movemen	Trapezoidal	5	

Table 7
The membership function properties of the DT-RFL model.

Inputs	Membership functions types	Number of Membership functions	Ranges
Cyclist Gender	Triangular	2	All membership function range is [0,1]
Damage Extent	Trapezoidal	4	
Road Type	Triangular	3	
Pavement Type	Triangular	2	
Crash Type	Trapezoidal	4	
Vehicle Movement	Trapezoidal	5	

5.4. Comparison of two hybrid fuzzy-decision models and C4.5 algorithm

To classify injury-severity in bicycle-vehicle accidents, DT-CFL and DT-RFL models were used in addition to the classical C4.5 algorithm in this study. While nineteen independent variables in Table 2 are used in the C4.5 algorithm, in the DT-RFL model, independent variables whose normalized values are below 40 % according to IGR, which determines the splitting criterion of C4.5 algorithm, were excluded. Thus, according to IGR, a total of six parameters were used as inputs to the DT-RFL model: three parameters with high effect (40 % < x ≤ 60 %) and three parameters with the greatest effect (x > 60 %) (See in Table 7.).

While creating the DT-CFL model, the rules of the C4.5 algorithm were used. Therefore, thirteen parameters in the rules of the C4.5 algorithm are used as input (See in Table 4). In this section, with a view to comparing two fuzzy-decision models and C4.5 algorithm Accuracy, Precision, Recall, and F-measure (Eq. 6) values, which are described below, were used by means of the confusion matrix (Table 8).

Accuracy: Percentage of all classes correctly classified by the model.

Precision: The ratio of correctly classified samples of a class to all classes specified to belong to that class.

Recall: The ratio of correctly classified samples of a class to all samples of that class.

$$\text{F}- \text{ measure: } 2 \times \frac{precision \times recall}{precision + recall} \tag{6}$$

When the Confusion matrix is examined, it is seen that C4.5 algorithm correctly classifies 647 out of 959 samples. Therefore, the accuracy value of the model was obtained as 67.46 %. This value was 69.96 % for DT-RFL and 59.22 % for DT-CFL. In previous studies, the average accuracy value in the study of Abdelwahab and Abdel-Aty (2001) was 62.5 %, the average accuracy value in the study of de Oña et al. (2011) was 61.5 % and Chen et al. (2015) was 66.3 %.When models in the literature are evaluated according to their accuracy, it is seen that the models presented in this study are quite acceptable. On the other hand, the DT-CFL model showed a relatively lower accuracy compared to the other two models. However, when evaluating model performances, it is necessary to examine the precision and recall values carefully as well as the accuracy values. In C4.5 model, the precision values of the NOI, INJ and FTL classes were 37.14 %, 71.83 %, 11.11 %, respectively, and the weighted average precision value of the model was 60.2 %. In the DT-RFL model, precision values are 48.48 %, 76.68 %, 31.57 % and the weighted average precision value is 67.51 %, while in the DT-CFL model the values are 26.29 %, 69.95 %, 11.76 % and weighted average value is 56.27 %. It is clear that the DT-RFL model performs more successfully than the other two models presented in the classification of cyclist injury-severity in bicycle vehicle-accidents.

The results of the DT-RFL fuzzy-decision mechanism were found to be consistent, especially since they have higher precision values in the classification of NOI and FTL classes. Regarding the Recall values of the models, the C4.5 algorithm gave better results with a weighted average Recall value of 67.5 % compared to DT-CFL (59.27 %) and DT-RFL (62.96 %). In accordance with these results, it is possible to say that the classification performance of one model for NOI, INJ, and FTL classes is better than another. For example, the DT-RFL model showed the best performance in classification of the NOI class with a recall value of 48.48 %, while it showed the worst classification performance with a recall value of 74.57 % in the classification of the INJ class. On the other hand, the C4.5 algorithm, which showed excellent classification performance for the INJ class, performed very poorly with a recall value of 2.32 % for the FTL class and correctly classified only 1 out of the 43 samples. For this reason, to evaluate the performance of the models clearly, the weighted average F-measure value (Table 9) of each model was calculated and compared. Consequently, it was found that the DT-RFL model demonstrated considerable classification performance with 72.0 % F-measure compared to other models.

6. Conclusions and discussion

This study aims to introduce two DT-based hybrid-fuzzy models (DT-CFL and DT-RFL) that are used while performing the classification of injury-severity in accidents involving cyclists, who are non-motorized road users. For this purpose, five- year accident data set of 2013–2017 was utilized to assess the results of the proposed classification models. Firstly, a total of nineteen parameters affecting the cyclist injury-severity were determined in the light of the literature, and the injury-severity level was classified using the C4.5 algorithm. Secondly, to build the DT-CFL model, all splitting conditions in the DT (from root to leaf) were obtained as a result of the C4.5 algorithm. Then, all splitting conditions were converted into the IF-THEN structure in the FL. Hence, the first fuzzy prediction model based on the DT was created. Thirdly, according to IGR, which also determines the splitting criteria in the C4.5 algorithm, the greatest effective parameters were determined as cyclist gender, damage extent, road type as well as highly effective parameters pavement type, crash type, and vehicle movement. Thanks to the IGR, the effect of each six parameters on bicycle-vehicle injury-severity were calculated. These parameters were defined as input variables in the DT-RFL model and the membership function of each input was revised according to the obtained weights. The final (extended) value (between 1 and 3) of multiplying the fuzzy cluster scores of the respective weights and parameters allowed the consequence of the 960 IF-THEN rules in the DT-RFL model to be

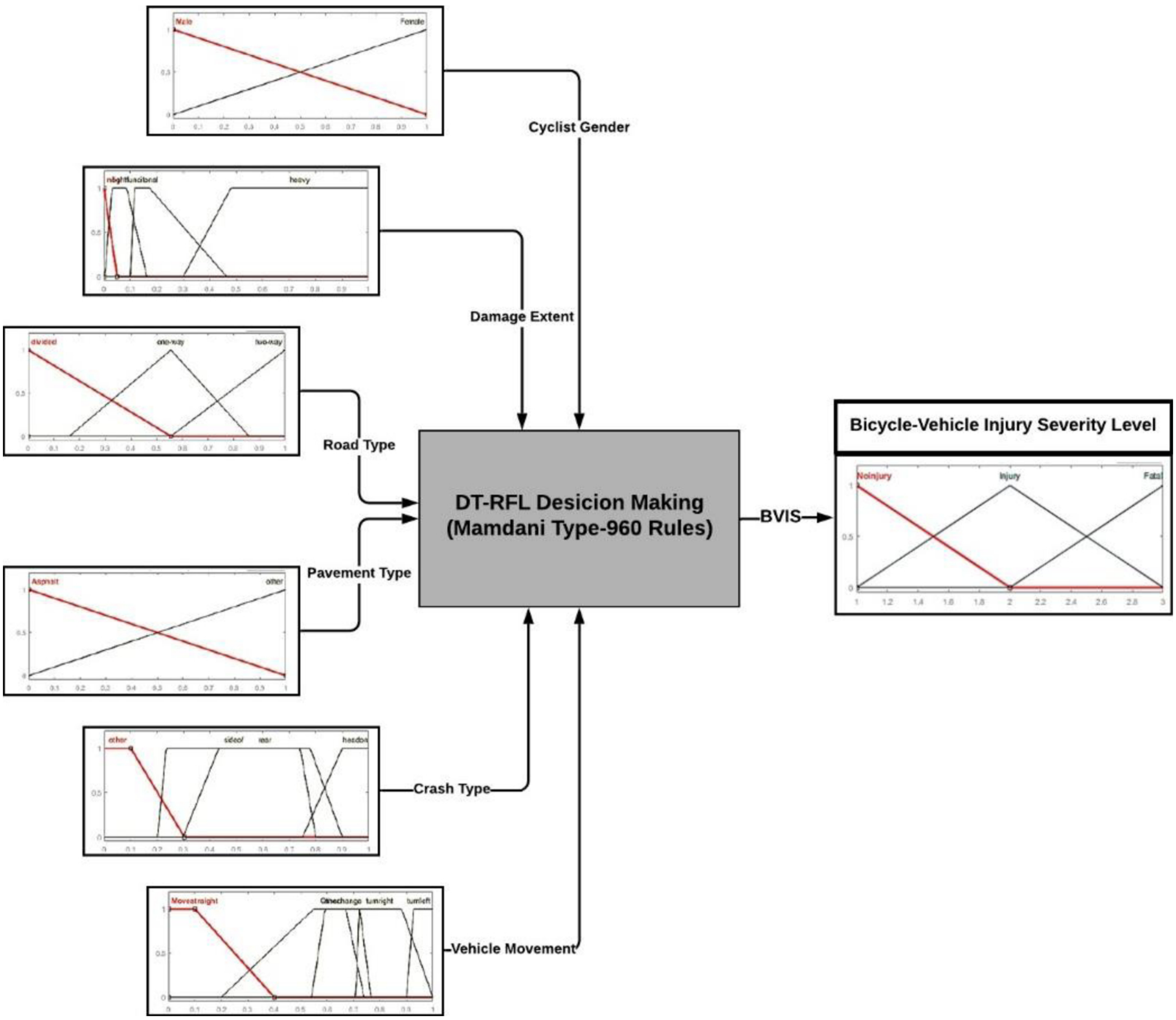

Fig. 11. The detailed structure of DT-RFL model.

Table 8
Confusion matrix of two fuzzy-decision models and C4.5 algorithm.

C4.5	NOI	INJ	FTL	**DT-RFL**	NOI	INJ	FTL	**DT-CFL**	NOI	INJ	FTL
NOI	**39**	202	2	NOI	**96**	143	4	NOI	**56**	185	2
INJ	60	**607**	6	INJ	95	**569**	9	INJ	150	**510**	13
FTL	6	36	**1**	FTL	7	30	**6**	FTL	7	34	**2**

Table 9
Performance evaluation criteria of models.

Models	Accuracy (%)	Weighted Average Precision (%)	Weighted Average Recall (%)	Weighted Average F-measure (%)
C4.5	67.46	60.2	**67.5**	62.0%
DT-RFL	**69.96**	**67.51**	62.96	72.0 %
DT-CFL	59.22	56.27	59.22	57.5%

defined as NOI, INJ or FTL without expert opinion. Finally, the bicycle-vehicle injury-severity classification performance of the proposed models was analyzed with respect to their accuracy, precision, recall, and F-measure values. Among the presented models, the C4.5 algorithm and DT-RFL model were found to be successful in classifying the cyclist injury-severity of bicycle-vehicle accidents with 67.46 % and 69.96 % accuracy values. DT-CFL model was determined to give relatively worse results than other models.

The main contributions of this study to the literature are given as follows:

- Handling the classification of accident cyclist injury-severity with a fuzzy approach can significantly improve the model performance.
- A Hybrid model which is created by combining several different models can give better results rather than classical machine learning algorithms.
- In the FL approach, it can be used in DT-FL models in addition to ANN-FL models, which are widely encountered in the literature to reduce or eliminate the influence of the expert opinion in the creation of rules.

In conclusion, this study presents important findings regarding the use of hybrid fuzzy-decision mechanisms to classify cyclist injury-severity in accidents. When compared to the existing literature, the analysis of cyclist injury-severity on vehicle related accidents is studied for the first time in the East Mediterranean region.

Besides the aforementioned contributions, some limitations that can change the performance of the models are observed during the process-stage of the models. The dataset consists of five-year bicycle-vehicle accidents. However, the number of fatal classes is very low compared to the number of injury and no-injury classes. This situation may negatively affect the classification performance of the models. Furthermore, in Turkey, accidents resulting in material damage are not put in the police records due to the agreement between the parties involved in the accident. Naturally, this affects the total number of accidents. Therefore, the dataset used in this study is unbalanced. Moreover, exposure and spatial parameters, which are important in previous studies, were not included in the models since they were not included in the dataset in this study. Model performances are affected by the number of parameters used in the model, performing the analysis with more parameters, and increasing the number of samples in the data set (for example, using 10-year accident data instead of 5). However, as it is known, traffic accidents are primarily of human origin and are significantly affected by human behavior. In this study, it is thought that model performance will increase with the addition of driver behavior and cyclist behavior parameters. On the other hand, since the data set does not contain information about driver and cyclist behavior, it could not be examined in models.

For further studies, it is strongly recommended that researchers can utilize the fuzzy-decision mechanisms on more balanced and comprehensive datasets that contain more independent variables.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Burak Yiğit Katanalp: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Ezgi Eren:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization.

References

Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transp. Res. Rec. 1746 (1), 6–13.
Abellán, J., López, G., De OñA, J., 2013. Analysis of traffic accident severity using decision rules via decision trees. Expert Syst. Appl. 40 (15), 6047–6054.
Adanu, E.K., Jones, S., Odero, K., 2020. Identification of factors associated with road crashes among functionally classified transport modes in Namibia. Scientific African, e00312.
Agrawal, R., Imieliński, T., Swami, A., 1993. Mining Association Rules Between Sets of Items in Large Databases. Paper Presented at the Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data.
Albalate, D., Fernandez-Villadangos, L., 2010. Motorcycle injury severity in Barcelona: the role of vehicle type and congestion. Traffic Inj. Prev. 11 (6), 623–631.
Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 34 (6), 729–741.
Al-Mousa, A., Faza, A., 2019. A fuzzy-based customer response prediction model for a day-ahead dynamic pricing system. Sustain. Cities Soc. 44, 265–274.
Al-Shammari, N., Bendak, S., Al-Gadhi, S., 2009. In-depth analysis of pedestrian crashes in Riyadh. Traffic Inj. Prev. 10 (6), 552–559.
ArcView, G., 1996. The Geographic Information System for Everyone. Environmental Systems Research Institute, Inc. Version,3.
Aziz, H.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accid. Anal. Prev. 50, 1298–1309.
Banakar, A., Zareiforoush, H., Baigvand, M., Montazeri, M., Khodaei, J., Behroozi-Khazaei, N., 2017. Combined application of decision tree and fuzzy logic techniques for intelligent grading of dried figs. J. Food Process Eng. 40 (3), e12456.
Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: a random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 35–47.
Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Anal. Methods Accid. Res. 3, 56–91.
Celik, A.K., Oktay, E., 2014. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. Accid. Anal. Prev. 72, 66–77.
Chang, L.-Y., Chien, J.-T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf. Sci. 51 (1), 17–22.
Chen, P., Shen, Q., 2019. Identifying high-risk built environments for severe bicycling injuries. J. Safety Res. 68, 1–7.
Chen, C., Zhang, G., Huang, H., Wang, J., Tarefder, R.A., 2016. Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. Accid. Anal. Prev. 96, 79–87.
Chimba, D., Sando, T., 2009. The prediction of highway traffic accident injury severity with neuromorphic techniques. Adv. Transp. Stud.: Int. J. 19, 17–26.
Clarry, A., Imani, A.F., Miller, E.J., 2019. Where we ride faster? Examining cycling speed using smartphone GPS data. Sustain. Cities Soc. 49, 101594.
Clifton, K.J., Burnier, C.V., Akar, G., 2009. Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment? Transp. Res. D Transp. Environ. 14 (6), 425–436.
Dai, D., 2012. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. J. Transp. Geogr. 24, 206–214.
de Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prev. 43 (1), 402–411.
de Oña, J., López, G., Abellán, J., 2013a. Extracting decision rules from police accident reports through decision trees. Accid. Anal. Prev. 50, 1151–1160.
De Oña, J., López, G., Mujalli, R., Calvo, F.J., 2013b. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accid. Anal. Prev. 51, 1–10.
Delen, D., Tomak, L., Topuz, K., Eryarsoy, E., 2017. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. J. Transp. Health 4, 118–131.
Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.
Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accid. Anal. Prev. 40 (3), 1033–1054.
Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accid. Anal. Prev. 47, 119–127.
Eren, E., Uz, V.E., 2019. A review on bike-sharing: the factors affecting bike-sharing demand. Sustain. Cities Soc., 101882.
Fountas, G., Anastasopoulos, P.C., 2017. A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. Anal. Methods Accid. Res. 15, 1–16.
Fountas, G., Anastasopoulos, P.C., 2018. Analysis of accident injury-severity outcomes: the zero-inflated hierarchical ordered probit model with correlated disturbances. Anal. Methods Accid. Res. 20, 30–45.
Fountas, G., Anastasopoulos, P.C., Abdel-Aty, M., 2018a. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Anal. Methods Accid. Res. 18, 57–68.
Fountas, G., Anastasopoulos, P.C., Mannering, F.L., 2018b. Analysis of vehicle accident-injury severities: a comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. Anal. Methods Accid. Res. 18, 15–32.
Fountas, G., Fonzone, A., Gharavi, N., Rye, T., 2020. The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents. Anal. Methods Accid. Res., 100124.
Gaber, M., Wahaballa, A.M., Othman, A.M., Diab, A., 2017. Traffic accidents prediction model using fuzzy logic: aswan desert road case study. J. Eng. Sci. Assiut Univ 45, 2844.
Goel, N., Sehgal, P., 2015. Fuzzy classification of pre-harvest tomatoes for ripeness estimation–An approach based on automatic rule learning using decision tree. Appl. Soft Comput. 36, 45–56.
Guo, Y., Li, Z., Wu, Y., Xu, C., 2018. Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. Accid. Anal. Prev. 115, 118–127.
Hagel, B.E., Romanow, N.T., Enns, N., Williamson, J., Rowe, B.H., 2015. Severe bicycling injury risk factors in children and adolescents: a case–control study. Accid. Anal. Prev. 78, 165–172.
Haleem, K., Gan, A., 2015. Contributing factors of crash injury severity at public highway-railroad grade crossings in the US. J. Safety Res. 53, 23–29.
Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Accid. Anal. Prev. 81, 14–23.
Heydari, S., 2018. A flexible discrete density random parameters model for count data: embracing unobserved heterogeneity in highway safety analysis. Anal. Methods Accid. Res. 20, 68–80.
Highway Accident Statistics, 2017. Retrieved From.
Hosseinpour, M., Yahaya, A.S., Ghadiri, S.M., Prasetijo, J., 2013. Application of Adaptive Neuro-fuzzy Inference System for road accident prediction. Ksce J. Civ. Eng. 17 (7), 1761–1772.
Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36.
Islam, S., Burton, B., 2019. A comparative injury severity analysis of rural intersection crashes under different lighting conditions in Alabama. J. Transp. Saf. Secur. 1–22.
Islam, S., Hossain, A.B., 2015. Comparative analysis of injury severity resulting from pedestrian–Motor vehicle and bicycle-motor vehicle crashes on roadways in Alabama. Transp. Res. Rec. 2514 (1), 79–87.
Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. J. Safety Res. 37 (3), 267–276.
Jianfeng, X., Hongyu, G., Jian, T., Liu, L., Haizhu, L., 2019. A classification and recognition model for the severity of road traffic accident. Adv. Mech. Eng. 11 (5), 1687814019851893.
Kadali, B.R., Vedagiri, P., 2013. Modelling pedestrian road crossing behaviour under mixed traffic condition. Eur. Transp. Res. Rev. 55 (3), 1–17.

Kaplan, S., Vavatsoulas, K., Prato, C.G., 2014. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. J. Safety Res. 50, 75–82.
Karlaftis, M.G., Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. Accid. Anal. Prev. 34 (3), 357–365.
Katanalp, B.Y., Yıldırım, Z.B., Karacasu, M., İbrikçi, T., 2019. Atik Kömür Katkili Asfalt Betonu performans KARAKTERİSTİKLERİNİN Yapay SİNİR AĞLARI ve MERKEZİ KOMPOZİT Tasarim YÖNTEMLERİ Kullanilarak Karşilaştirilmasi. Mã¼hendislik Bilim. Ve Tasarä±m Derg. 7 (3), 680–688. https://doi.org/10.21923/jesd.512018.
Khosravanian, R., Sabah, M., Wood, D.A., Shahryari, A., 2016. Weight on drill bit prediction models: sugeno-type and Mamdani-type fuzzy inference systems compared. J. Nat. Gas Sci. Eng. 36, 280–297.
Kim, J.-K., Ulfarsson, G.F., Shankar, V.N., Mannering, F.L., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accid. Anal. Prev. 42 (6), 1751–1758.
Kim, M., Kho, S.-Y., Kim, D.-K., 2017. Hierarchical ordered model for injury severity of pedestrian crashes in South Korea. J. Safety Res. 61, 33–40.
Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: an application of ordered probit models. Accid. Anal. Prev. 34 (3), 313–321.
Kunt, M.M., Aghayan, I., Noii, N., 2011. Prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. Transport 26 (4), 353–366.
Kweon, Y.-J., Lee, J., 2010. Potential risk of using General Estimates System: bicycle safety. Accid. Anal. Prev. 42 (6), 1712–1717.
Larimian, T., Zarabadi, Z.S.S., Sadeghi, A., 2013. Developing a fuzzy AHP model to evaluate environmental sustainability from the perspective of Secured by Design scheme—a case study. Sustain. Cities Soc. 7, 25–36.
Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. Accid. Anal. Prev. 37 (4), 775–786.
Lee, J., Yoon, T., Kwon, S., Lee, J., 2020. Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: seoul City study. Appl. Sci. 10 (1), 129.
Li, Z., Chen, C., Wu, Q., Zhang, G., Liu, C., Prevedouros, P.D., Ma, D.T., 2018. Exploring driver injury severity patterns and causes in low visibility related single-vehicle crashes using a finite mixture random parameters model. Anal. Methods Accid. Res. 20, 1–14.
Lison, P., 2015. An Introduction to Machine Learning. Language Technology Group, Edinburgh, UK.
Liu, P., Fan, W., 2020. Exploring injury severity in head-on crashes using latent class clustering analysis and mixed logit model: a case study of North Carolina. Accid. Anal. Prev. 135, 105388.
Liu, J., Khattak, A.J., Li, X., Nie, Q., Ling, Z., 2020. Bicyclist injury severity in traffic crashes: a spatial approach for geo-referenced crash data to uncover non-stationary correlates. J. Safety Res.
Ma, X., Chen, S., Chen, F., 2017. Multivariate space-time modeling of crash frequencies by injury severity levels. Anal. Methods Accid. Res. 15, 29–40.
Mafi, S., Abdelrazig, Y., Doczy, R., 2018. Machine learning methods to analyze injury severity of drivers from different age and gender groups. Transp. Res. Rec. 2672 (38), 171–183.
Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: an application to accident-injury severities. Accid. Anal. Prev. 41 (4), 829–838.
Mamdani, E.H., Assilian, S., 1993. An Experiment in Linguistic Synthesis With a Fuzzy Logic Controller. In Readings in Fuzzy Sets for Intelligent Systems. Elsevier, pp. 283–289.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.
Mehran, K., 2008. Takagi-sugeno Fuzzy Modeling for Process Control. Industrial Automation, Robotics and Artificial Intelligence (EEE8005). pp. 262.
Meng, X.-h., Zheng, L., Qin, G.-m., 2009. Traffic accidents prediction and prominent influencing factors analysis based on fuzzy logic [J]. J. Transp. Syst. Eng. Inf. Technol. 2, 015.
Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: a comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. Saf. Sci. 54, 27–37.
Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. 49, 58–72.
Moore, D.N., Schneider, I.V.W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accid. Anal. Prev. 43 (3), 621–630.
Moudon, A.V., Lin, L., Jiao, J., Hurvitz, P., Reeves, P., 2011. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. Accid. Anal. Prev. 43 (1), 11–24.
Mujalli, R.O., de Oña, J., 2013. Injury Severity Models for Motor Vehicle Accidents: a Review. Paper Presented at the Proceedings of the Institution of Civil Engineers-transport.
Mukoko, K.K., Pulugurtha, S.S., 2019. Examining the influence of network, land use, and demographic characteristics to estimate the number of bicycle-vehicle crashes on urban roads. Iatss Res.
Nauck, D., Kruse, R., 1998. How the Learning of Rule Weights Affects the Interpretability of Fuzzy Systems. Paper Presented at the 1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36228).
Nikiforiadis, A., Basbas, S., 2019. Can pedestrians and cyclists share the same space? The case of a city with low cycling levels and experience. Sustain. Cities Soc. 46, 101453.
Obeng, K., 2011. Gender differences in injury severity risks in crashes at signalized intersections. Accid. Anal. Prev. 43 (4), 1521–1531.
Obeng, K., Rokonuzzaman, M., 2013. Pedestrian Injury Severity in Automobile Crashes.
Observatory, E.R.S., 2017. Annual Accident Report. in: European Commission Brussels.
Olszewski, P., Szagała, P., Wolański, M., Zielińska, A., 2015. Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. Accid. Anal. Prev. 84, 83–91.
Olutayo, V., Eludire, A., 2014. Traffic accident analysis using decision trees and neural networks. Int. J. Inf. Technol. Comput. Sci. 2, 22–28.
Omid, M., 2011. Design of an expert system for sorting pistachio nuts through decision tree and fuzzy logic classifier. Expert Syst. Appl. 38 (4), 4339–4347.
Pai, C.-W., Saleh, W., 2008. Modelling motorcyclist injury severity by various crash types at T-junctions in the UK. Saf. Sci. 46 (8), 1234–1247.
Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf. Sci. 47 (1), 145–154.
Pour-Rouholamin, M., Zhou, H., 2016. Investigating the risk factors associated with pedestrian injury severity in Illinois. J. Safety Res. 57, 9–17.
Prato, C.G., Gitelman, V., Bekhor, S., 2012. Mapping patterns of pedestrian fatal accidents in Israel. Accid. Anal. Prev. 44 (1), 56–62.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.
Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man. Stud. 27 (3), 221–234.
Rahim, R., 2017. Comparative analysis of membership function on Mamdani fuzzy inference system for decision making. Paper presented at the Journal of Physics Conference Series.
Rankavat, S., Tiwari, G., 2016. Pedestrians risk perception of traffic crash and built environment features–Delhi, India. Saf. Sci. 87, 1–7.
Rash-ha Wahi, R., Haworth, N., Debnath, A.K., King, M., 2018. Influence of type of traffic control on injury severity in bicycle–motor vehicle crashes at intersections. Transp. Res. Rec. 2672 (38), 199–209.
Ravishankar, K.V.R., Nair, P.M., 2018. Pedestrian risk analysis at uncontrolled midblock and unsignalised intersections. J. Traffic Transp. Eng. 5 (2), 137–147.
Rezapour, M., Molan, A.M., Ksaibati, K., 2019. Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. Int. J. Transp. Sci. Technol.
Rifaat, S.M., Tay, R., De Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accid. Anal. Prev. 43 (1), 276–283.
Robartes, E., Chen, T.D., 2017. The effect of crash characteristics on cyclist injuries: an analysis of Virginia automobile-bicycle crash data. Accid. Anal. Prev. 104, 165–173.
Rokach, L., Maimon, O.Z., 2008. Data Mining With Decision Trees: Theory and Applications Vol. 69 World scientific.
Ross, T.J., 2005. Fuzzy Logic With Engineering Applications. John Wiley & Sons.
Sakthivel, N., Sugumaran, V., Nair, B.B., 2010. Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. Mech. Syst. Signal Process. 24 (6), 1887–1906.
Sarkar, S., Tay, R., Hunt, J.D., 2011. Logistic regression model of risk of fatality in vehicle–pedestrian crashes on national highways in Bangladesh. Transp. Res. Rec. 2264 (1), 128–137.
Sasidharan, L., Wu, K.-F., Menendez, M., 2015. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. Accid. Anal. Prev. 85, 219–228.
Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: a random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 15, 41–55.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J. Safety Res. 27 (3), 183–194.
Siamidoudaran, M., İşçioğlu, E., 2019. Injury severity prediction of traffic collision by applying a series of neural networks: the City of London Case Study. PrometTrafficTransportation 31 (6), 643–654.
Sivanandam, S., Sumathi, S., Deepa, S., 2007. Introduction to Fuzzy Logic Using MATLAB Vol. 1 Springer.
Sivasankaran, S.K., Balasubramanian, V., 2020. Exploring the severity of bicycle–Vehicle crashes using latent class clustering approach in India. J. Safety Res. 72, 127–138.
Soriguera, F., Jiménez-Meroño, E., 2020. A continuous approximation model for the optimal design of public bike-sharing systems. Sustain. Cities Soc. 52, 101826.
Sun, M., Sun, X., Shan, D., 2019. Pedestrian crash analysis with latent class clustering method. Accid. Anal. Prev. 124, 50–57.
Sze, N.-N., Wong, S., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accid. Anal. Prev. 39 (6), 1267–1278.
Tang, T., Wang, H., Ma, J., Zhou, X., 2020. Analysis of crossing behavior and violations of electric bikers at signalized intersections. J. Adv. Transp. 2020.
Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Saf. Sci. 50 (4), 1141–1151.
Ulfarsson, G.F., Kim, S., Booth, K.M., 2010. Analyzing fault in pedestrian–motor vehicle crashes in North Carolina. Accid. Anal. Prev. 42 (6), 1805–1813.
Vilaça, M., Macedo, E., Coelho, M.C., 2019. A rare event modelling approach to assess injury severity risk of vulnerable road users. Safety 5 (2), 29.
Wahab, L., Jiang, H., 2019. A multinomial logit analysis of factors associated with severity of motorcycle crashes in Ghana. Traffic Inj. Prev. 20 (5), 521–527.
Wang, C., 2015. A Study of Membership Functions on Mamdani-type Fuzzy Inference System for Industrial Decision-making.
Wang, H., Zheng, L., Meng, X., 2011. Traffic Accidents Prediction Model Based on Fuzzy Logic. In Advances in Information Technology and Education. Springer, pp. 101–108.
Waseem, M., Ahmed, A., Saeed, T.U., 2019. Factors affecting motorcyclists' injury severities: an empirical assessment using random parameters logit model with heterogeneity in means and variances. Accid. Anal. Prev. 123, 12–19.
Watanabe, L., Rendell, L.A., 1991. Learning Structural Decision Trees From Examples. Paper Presented at the IJCAI.
Wen, H., Tang, Z., Zeng, Y., Zhang, K., 2020. A comprehensive analysis for the

heterogeneous effects on driver injury severity in single-vehicle passenger Car and SUV rollover crashes. J. Adv. Transp. 2020.
Witten, I., Frank, E., 2005. The Weka machine learning workbench. Data Mining Practical Machine Learning Tools and Techniques 363–427.
Witten, I.H., Frank, E., Hall, M.A., 2005. Practical Machine Learning Tools and Techniques. Morgan Kaufmann, pp. 578.
Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P.-S., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: a random parameters generalized ordered probability model with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 117–132.
Xiong, Y., Tobias, J.L., Mannering, F.L., 2014. The analysis of vehicle crash injury-severity data: a Markov switching approach with road-segment heterogeneity. Transp. Res. Part B Methodol. 67, 109–128.
Yan, X., Radwan, E., Abdel-Aty, M., 2005a. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accid. Anal. Prev. 37 (6), 983–995.
Yan, X., Radwan, E., Birriel, E., 2005b. Analysis of red light running crashes based on quasi-induced exposure and multiple logistic regression method. Transp. Res. Rec. 1908 (1), 70–79.
Ye, F., Lord, D., 2011. Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. Transp. Res. Rec. 2241 (1), 51–58.
Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85.
Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Saf. Sci. 63, 50–56.
Zadeh, L.A., 1988. Fuzzy logic. Computer 21 (4), 83–93.
Zahabi, S.A.H., Strauss, J., Manaugh, K., Miranda-Moreno, L.F., 2011. Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. Transp. Res. Rec. 2247 (1), 81–90.
Zeng, Q., Huang, H., 2014. A stable and optimized neural network model for crash injury severity prediction. Accid. Anal. Prev. 73, 351–358.
Zhang, G., Yau, K.K., Chen, G., 2013. Risk factors associated with traffic violations and accident severity in China. Accid. Anal. Prev. 59, 18–25.
Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Wang, Z., 2019. Traffic accident's severity prediction: a deep-learning approach-based CNN network. IEEE Access 7, 39897–39910.
Zhou, B., Li, Z., Zhang, S., Zhang, X., Liu, X., Ma, Q., 2019. Analysis of factors affecting hit-and-Run and Non-Hit-and-Run in Vehicle-Bicycle crashes: a non-parametric approach incorporating data imbalance treatment. Sustainability 11 (5), 1327.