Heliyon 10 (2024) e32469

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identifying interactions among factors related to death occurred at the scene of traffic accidents: Application of "logic regression" method

Milad Jamali-dolatabad [a], Homayoun Sadeghi-bazargani [b], Saman Salemi [c], Parvin Sarbakhsh [d,*]

[a] Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran
[b] Road Traffic Injury Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[c] Department of Medicine, Islamic Azad University Tehran Medical Sciences, Tehran, Iran
[d] Road Traffic Injury Research Center, Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Keywords:
Traffic accidents
Logic regression
Boolean combinations
Annealing algorithm
Mortality
Interaction effects
Pedestrians
Passengers
Drivers

ABSTRACT

Aim: Traffic accidents are caused by several interacting risk factors. This study aimed to investigate the interactions among risk factors associated with death at the accident scene (DATAS) as an indicator of the crash severity, for pedestrians, passengers, and drivers by adopting "Logic Regression" as a novel approach in the traffic field.

Method: A case-control study was designed based on the police data from the Road Traffic Injury Registry in northwest of Iran during 2014–2016. For each of the pedestrians, passengers, and drivers' datasets, logic regression with "logit" link function was fitted and interactions were identified using Annealing algorithm. Model selection was performed using the cross-validation and the null model randomization procedure.

Results: regarding pedestrians, "The occurrence of the accident outside a city in a situation where there was insufficient light" (OR = 6.87, *P*-value<0.001) and "the age over 65 years" (OR = 2.97, *P*-value<0.001) increased the chance of DATAS. "Accidents happening in residential inner-city areas with a light vehicle, and presence of the pedestrians in the safe zone or on the non-separate two-way road" combination lowered the chance of DATAS (OR = 0.14, *P*-value<0.001). For passengers, "Accidents happening in outside the city or overturn of the vehicle" combination (OR = 8.55, *P*-value<0.001), and "accidents happening on defective roads" (OR = 2.18, *P*-value<0.001) increased the odds of DATAS; When "driver was not injured or the vehicle was two-wheeled", chance of DATAS decreased for passengers (OR = 0.25, p-value<0.001). The

Abbreviations: (PA>65), Pedestrian Aged>65; (NDL), No-daylight; (NCW), Non clear weather; (UA), Uneven area; (NRA), Non-residential area; (SUA), Suburban accident; (NSTR), Non-separate two-way road; (PUS), Pedestrian in an unsafe position; (MP), Male pedestrians; (HV), Heavy vehicle; (COT), Car overturning; (UCR), Unclean road surface; (HOC), Head-on collision; (DR), Defective Road; (MSR), Main street or road; (TWV), two-wheeler vehicle; (BCD), Blame the car driver; (MPA), Male passenger; (LED), Low educated driver (primary and illiterate); (IND), Injured driver (dead or injured); (AWCV), accident with counterpart vehicle; (AORW), Accident on the runway; (RWS), Road without shoulders; (ABS), Vehicle without airbags and ABS brakes; (MD), Male driver; (MCA), Multi-car accident.

* Corresponding author. Road Traffic Injury Research Center, Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Golgasht St, Attar Neyshabori St, Tabriz, Iran.

E-mail addresses: miladjamali1994@gmail.com (M. Jamali-dolatabad), homayoun.sadeghi@gmail.com (H. Sadeghi-bazargani), sam.salemi12@gmail.com (S. Salemi), p.sarbakhsh@gmail.com (P. Sarbakhsh).

https://doi.org/10.1016/j.heliyon.2024.e32469
Received 19 September 2023; Received in revised form 2 June 2024; Accepted 4 June 2024
Available online 5 June 2024

odds of DATAS were higher for "drivers who had a head-on accident, or drove a two-wheeler vehicle, or overturned the vehicle" (OR = 4.03, P-value<0.001). "Accident on the roads other than runway or the absence of a multi-car accident or an accident in a non-residential area" (OR = 6.04, P-value<0.001), as well "the accident which occurred outside the city or on defective roads, and the drivers were male" had a higher risk of DATAS for drivers (OR = 5.40, P-value<0.001).

Conclusion: By focusing on identifying interaction effects among risk factors associated with DATAS through logic regression, this study contributes to the understanding of the complex nature of traffic accidents and the potential for reducing their occurrence rate or severity. According to the results, the simultaneous presence of some risk factors such as the quality of roads, skill of drivers, physical ability of pedestrians, and compliance with traffic rules play an important role in the severity of the accident. The revealed interactions have practical significance and can play a significant role in the problem-solving process and facilitate breaking the chain of combinations among the risk factors. Therefore, practical suggestions of this study are to control at least one of the risk factors present in each of the identified combinations in order to break the combination to reduce the severity of accidents. This may have, in turn, help the policy-makers, road users, and healthcare professionals to promote road safety through prioritizing interventions focusing on effect size of simultaneous coexistence of crash severity determinants and not just the main effects of single risk factors or their simple two-way interactions.

Definitions

Head-on collision: a crash of two vehicles that are moving directly toward each other.

Car overturning: An overturn is a type of vehicle crash in which a vehicle tips over onto its side or roof.

A two-wheeler: a vehicle that runs on two wheels.

Separate two-way road: a dividing structure or strip located in the middle of a road to separate two-way lanes, preventing vehicles traveling on lanes going in the opposite direction from colliding.

Heavy Vehicle: a motor vehicle with or without load which alone or together with any trailer, semi-trailer or other vehicle being towed, weighs 4.5 tons or more.

Defective road: refers to any roadway that has been impacted by a hazardous condition or defect that makes it dangerous for drivers and their passengers.

Main road: A major road for any form of motor transport.

Blame the car driver: Blame or fault in legal terms means failing to take reasonable care.

Counterpart vehicle: a counterpart vehicle is a vehicle that has similar stats to another vehicle.

Road shoulder: The shoulder is a strip of pavement outside an outer lane; it is provided for emergency use by traffic and to protect the pavement edges from traffic damage.

Multiple car accident: a road traffic collision involving many vehicles.

1. Introduction

Traffic accidents are considered a major public health problem worldwide about 1.35 million people die annually due to traffic accidents [1]. These accidents lead to more tragic consequences in underdeveloped and developing countries compared to developed countries [2].

One of the main reasons behind the high death rate of traffic accidents, especially in undeveloped countries, is the failure to accurately identify the factors contributing to the occurrence or severity of the accidents.

Numerous factors (e.g., factors associated with human, environment, and vehicle) can play a role in the occurrence or severity of a traffic accident [3–5]. Often, not only do the main effects of these factors contribute to the occurrence of an accident, but they also interact and even combine with each other, thereby increasing the chances of an accident occurrence or severity. Furthermore, sometimes a single factor may not be enough to cause an outcome (such as death at the traffic accident scene), but its simultaneous occurrence with other factors is the reason for the occurrence of the outcome. In other words, the risk factor is not a single variable but a special combination of several variables. Take, for example, lack of lighting and bad weather condition as two important factors affecting the occurrence of an accident or its severity. Each one of these factors alone has a certain main effect on the occurrence or severity of the accident, while a combination of these two risk factors and the simultaneous presence of them, called interaction, leads to their synergy and produces a more catastrophic effect [6,7].

Given the significance of traffic accidents and the presence of several risk factors, it is imperative to prioritize the examination and management of combinations of risk factors. Identification of these interactions plays a significant role in problem-solving process and facilitates breaking the chain of combinations among the variables by interfering and controlling the factors possible to intervene or control which can lead to a reduction in the occurrence rate or severity of the traffic accidents [8].

Classic regression models are basically used to examine the relationships among variables and predictions. These methods are usually used to identify the main effects of variables and ultimately low-order interactions (i.e., two-way and three-way at most). But often, especially when all predictors are binary, the interaction between many predictors is what causes the differences in response. It is not possible to manually identify and consider the higher-order interactions in these models due to their complexity. To address this problem and identify these complex interactions, "Logic Regression" method can be used [9].

Logic regression, not to be confused with logistic regression, is one of the generalized and relatively new methods in which the predictor variables are made by the Boolean combination of the initial binary variables. In this method, an attempt is made to create new binary variables that are combinations of primary variables and have higher predictive ability for the response variable among other Boolean combinations. In this method, moreover, search algorithms (e.g., Annealing or Greedy algorithms) are used to achieve a combination of variables that minimizing the model score function [9,10].

While lower order interactions can be evaluated by traditional regressions, these interactions must be known in advance, and used as input variables in the model. Conversely, in logic regression, desired interactions need not be known in advance, quite the contrary, the detection of important higher order interactions is the main goal of logic regression [9,11].

Furthermore, one of the main potencies of Logic Regression is the ease of interpretation of the fitted models. This makes it separated from other methods such as neural networks, where the focus is on prediction instead of interpretation [11]. Although some of the machine learning methods in the traffic context such as neural networks (NN) [12], random forest [13], SVM [14], and decision tree [15], are able to calculate the importance of individual variables, they do not directly quantify the importance of interactions of variables [14,15]. Logic regression can quantify the importance of the interactions according to their predictive abilities [16].

Logic regression offers a great potential for identifying the interaction effects and has been used in various disciplines of medical sciences [17–19], particularly, in genetics studies [20–22]. In the field of traffic accidents, however, it is an unknown method and has received insufficient attention despite its huge potential; it has just been employed by a study in the field of traffic in order for investigating the role of interactions in fatal accidents [23].

Given the above discussion, it is highly important to identify the interactions among the factors affecting the occurrence of a traffic accident or the severity of an injury. Previous studies thoroughly investigated the main effect of risk factors associated with traffic accidents [24–44], but despite the importance of the interactions among traffic accident risk factors, in the literature review, rare articles were found that specifically seek to investigate these interaction and deal with efficient methods to identify them [6,7,13].

Our hypothesis in this study was that risk factors of traffic accidents can interact with each other to increase the chance of death at the accident scene (DATAS) outcome that is an indicator of the severity of accidents.

So, this study, aimed to identify the effects of complex interactions among the factors contributing to road traffic accidents on the DATAS outcome for three groups of road users (i.e., drivers, passengers, and pedestrians) based on the police data for three north-western provinces of Iran (i.e., East Azerbaijan, West Azerbaijan, and Ardabil) from March 21, 2014 to March 19, 2016 through logic regression as a novel approach in the traffic field.

Therefore, the specific contribution of this study and the new insights it brings to the field of traffic accident research is that, in addition to discussing the interactions among the risk factors of traffic accidents as a noticeable issue, the logic regression method, which is a powerful statistical learning technique for finding complex interactions among variables, is used to find such interactions and quantify their importance.

2. Materials and method

2.1. Data

This study used the police station data of the Integrated Road Traffic Injury Registry (IRTIR) as defined earlier. The IRTIR study was approved by the Ministry of Health and Medical Education (MOHME), Iranian Trustee for Traffic Knowledge Development and Road Traffic Injury Research Center (approval no.: 700/1482). The ethical approval for this registry was obtained from the Ethical Committee in Tabriz University of Medical Sciences under number IR.TBZMED.REC.1396.465 [45].

Data for crash scene police station of IRTIR is collected using "KAM114" form filled out by the police at the accident scene. This eight-page form contains several variables that are used by the Iranian police to report all traffic accidents in the country. The data are deposited in two sections: general crash characteristics and road user-specific information sections. General crash characteristics includes data about time, place, road conditions, weather conditions, lighting conditions, vehicle characteristics, personal characteristics, and causes of the accident. The road user-specific section, on the other hand, includes information about the driver, passengers, and pedestrians. These data are registered in the police registry system and can be accessed according to the desired time period and geographic region.

In the current study, the police dataset was examined separately for three groups of accidents involving the pedestrians, drivers, and passengers in East Azerbaijan, West Azerbaijan, and Ardabil Provinces from March 21, 2014 to March 19, 2016 and the analyses were performed separately for each dataset.

Desired risk factors were identified according to the literature review, experts' opinions, and availability of variables in our datasets. Then, a set of binary variables (i.e., having risk factor/not having risk factor) were created as follows: Pedestrian aged>65 (PA>65), no-daylight (NDL), non-clear weather (NCW), uneven area (UA), non-residential area (NRA), suburban accident (SUA), non-separate two-way road (NSTR), pedestrian in an unsafe position (PUS), male pedestrians (MP), heavy vehicle (HV), car overturning (COT), unclean road surface (UCR), head-on collision (HOC), defective road (DR), main street or road (MSR), two-wheeler vehicle (TWV), blame the car driver (BCD), male passenger (MPA), low educated (primary or illiterate) driver (LED), injured driver (dead or injured) (IND), accident with counterpart vehicle (AWCV), accident on the runway (AORW), road without shoulders (RWS), vehicle without airbags and ABS brakes (ABS), male driver (MD), multi-car accident (MCA). Details of dividing variables into binary are presented in Tables 1–3.

Data pre-processing was performed by eliminating the duplicate data (using the unique accident code) and variables with a high missing rate (above 20 %). Multiple imputation by chained equations (MICE) method [46] was used to impute the missing data for the rest of the variables. After data preparation, a case-control design was applied to prevent misinterpretations due to sole statistical significance of non-meaningful clinical and practical interactions because of the large sample size of police data. Therefore, out of all traffic accidents recorded in the police system during the intended period and region (separately for driver, pedestrian, and passenger datasets), all the people who, according to the police report, was killed at the crash scene were considered as the cases, and from the non-fatal accidents, almost three times the number of cases, were randomly selected as the controls. The flowchart of the research methodology is presented in Fig. 1.

2.2. Logic regression

Logic regression is a tree-based method and powerful statistical learning technique that tries to construct new, better predictors for the response by considering Boolean combinations of primary binary predictors. It makes a set of decision trees with binary predictors linked by logical operators.

Suppose $X_1, X_2, \ldots, X_k$ are binary predictor variables and y is the response variable. The purpose of logic regression would be to fit a regression model of the form:

$$g(E(y)) = \beta_0 + \sum_{j=1}^{t} \beta_j L_j \tag{1}$$

Where L_j is a Boolean combination (tree) of the binary predictor variables X_i (leaves), and $g(E(y))$ is a link function. These combinations are called a "Boolean logic expression". Boolean logic expression is denoted by using the logic operators $\vee$ ("OR"), $\wedge$ ("AND"), and the superscript c refers to the complement ("not"). An example for a Boolean expression is:

$$L = \left(\left[X_1 \vee X_2^c\right]\right) \wedge X_3 \tag{2}$$

Table 1
Descriptive statistics of risk factors by cases and controls and the results of the chi-square test for pedestrian data set.

Variable	Coding	Category	Case(158)	Control(474)	P-value
Age of pedestrian	Primary	quantitative	47.23(SD = 27.62)	37.73(SD = 23.49)	<0.001
	Binary	Pedestrian Aged>65 (PA>65)	48(30.38 %)	73(15.40)	<0.001
Light condition	Primary	day	99(62.65 %)	334(70.46 %)	0.138
		night	49(31.01 %)	118(24.89 %)	
		sunrise or sunset	10(6.32 %)	20(4.22 %)	
	Binary	No daylight (NDL)	59(37.34 %)	138(29.11 %)	0.054
weather condition	Primary	Non-clear	16(10.13 %)	26(5.49 %)	0.046
		clear	142(89.87 %)	448(94.51 %)	
	Binary	Non-clear weather (NCW)	16(10.13 %)	26(5.49 %)	0.046
Unevenness condition	Primary	smooth	145(93.67 %)	454(95.87 %)	0.025
		the hill	5(3.16 %)	5(1.05 %)	
		the mountain	8(5.06 %)	15(3.16 %)	
	Binary	Uneven area (UA)	13(8.23 %)	20(4.22 %)	0.054
Area type	Primary	residential	98(62.02 %)	423(89.24 %)	<0.001
		non-residential	54(34.17 %)	41(8.64 %)	
		other	6(3.79 %)	9(1.89 %)	
	Binary	Non-residential area(NRA)	60(37.97 %)	50(10.55 %)	<0.001
The place of the accident	Primary	Suburban	67(42.41 %)	45(9.49 %)	<0.001
		urban	89(56.32 %)	428(90.29 %)	
	Binary	Suburban accident(SUA)	67(42.41 %)	45(9.49 %)	0.001
road type	Primary	One- way	18(11.39 %)	51(10.75 %)	0.516
		Separated two-way	83(52.53 %)	227(47.89 %)	
		Non-separated two-way	57(36.08 %)	195(41.13 %)	
	Binary	Non-separate two-way road (NSTR)	57(36.08 %)	195(41.13 %)	0.261
Pedestrian position	Primary	Crossing the road from the permitted route	64(40.50 %)	301(63.50 %)	<0.001
		Crossing the road in an illegal way	21(13.29 %)	32(6.75 %)	
		Standing on the side of the road	11(6.96 %)	9(1.89 %)	
		other items	62(39.24 %)	132(27.84 %)	
	Binary variable	Pedestrian in an unsafe position(PUS)	94(59.49 %)	173(36.49 %)	<0.001
gender	Primary	Male	119(75.32 %)	323(68.14 %)	0.090
		Female	39(24.68 %)	151(31.85 %)	
	Binary	Male pedestrians(MP)	119(75.32 %)	323(68.14 %)	0.090
Vehicle type	Primary	Light Vehicle	122(77.21 %)	448(94.51 %)	<0.001
		Heavy vehicle	36(22.78 %)	26(5.49 %)	
	Binary	Heavy vehicle(HV)	36(22.78 %)	26(5.49 %)	<0.001

Table 2
Descriptive statistics of risk factors by cases and controls and results of chi-square test for passenger data set.

Variable	Coding	Category	Case(158)	Control(474)	P-value
age	Primary	quantitative	32.33(SD = 19.40)	30.32(SD = 15.99)	0.049
	Binary	Passenger Aged>65(PA>65)	26(6.48 %)	40(3.33)	0.007
light	Primary	day	215(53.61 %)	638(53.03 %)	0.169
		night	87(21.69 %)	335(27.84 %)	
		sunrise or sunset	57(14.21 %)	92(7.8 %)	
	Binary	No daylight(NDL)	144(35.91 %)	472(35.49 %)	0.236
weather condition	Primary	Clear	286(71.32 %)	1007(83.70 %)	0.001
		Non clear	115(28.68 %)	196(16.29 %)	
	Binary	Non clear weather(NCW)	115(28.68 %)	196(16.29 %)	<0.001
Uneven condition	Primary	smooth	325(81.04 %)	1080(89.87 %)	<0.001
		the hill	26(6.48 %)	61(5.07 %)	
		the mountain	50(12.46 %)	62(5.13 %)	
	Binary	Uneven area (UA)	76 (18.95 %)	123(10.22 %)	<0.001
Area type	Primary	other	88(21.94 %)	158(13.13 %)	<0.001
		residential	46(11.47 %)	600(49.87 %)	
		non-residential	267(66.58 %)	445(36.99 %)	
	Binary	Non-residential area (NRA)	355(88.53 %)	603(50.12 %)	<0.001
mechanism	Primary	A collision between a vehicle and a motorcycle	11(2.74 %)	180(14.96 %)	<0.001
		Collision of a vehicle with a vehicle	186(46.38 %)	530(44.05 %)	
		Collision of a vehicle with several vehicles	26(6.48 %)	93(7.73 %)	
		Collision with a fixed object	19(4.73 %)	97(8.06 %)	
		Car overturning	130(32.42 %)	194(16.13 %)	
		Leaving from the road	26(6.48 %)	80(6.65 %)	
		other	3(0.74 %)	29(2.41 %)	
	Binary	Car overturning(COT)	130(32.42 %)	194(16.13 %)	<0.001
Road surface	Primary	dry	276 (68.82 %)	957(79.55 %)	0.015
		wet	24 (5.98 %)	63(5.23 %)	
		Frost and snow	12(2.99 %)	13(1.08 %)	
		other	89(22.19 %)	170 (14.13 %)	
	Binary	Unclean road surface(UCR)	125(31.17 %)	246(20.45 %)	<0.001
The manner of the accident	Primary	Head-on collision	111(37.50 %)	185(15.38 %)	<0.001
		Front to back	20(8.89 %)	205(17.04 %)	
		Front to the right side	43(21.61 %)	156(12.96 %)	
		Front to the left side	14(11.20 %)	111(9.22 %)	
		Right side to right side	3(60 %)	2(0.16 %)	
		Left side to left side	7(14.29 %)	42(5.33 %)	
		other	203(28.88 %)	502(41.72 %)	
	Binary	Head-on collision(HOC)	111(27.68 %)	185(15.38 %)	<0.001
The place of the accident	Primary	out of town	45(11.22 %)	666(55.36 %)	<0.001
		inner city	348(86.78 %)	533(44.31 %)	
		rural road	8(1.99 %)	4(0.33 %)	
	Binary	Suburban accident(SUA)	348(86.78 %)	533(44.31 %)	<0.001
road defect	Primary	Non defected	223(55.61 %)	915(76.06 %)	<0.001
		defected	178(44.39 %)	288(23.94 %)	
	Binary	Defective Road (DR)	178(44.39 %)	288(23.94 %)	<0.001
Road type	Primary	freeway	4(0.99 %)	10(0.83 %)	<0.001
		highway	19(4.73 %)	30(2.49 %)	
		main street	20(4.98 %)	544(45.22 %)	
		main road	210(52.36 %)	283(23.52 %)	
		sub street	5(1.24 %)	51(4.23 %)	
		secondary road	112(27.93 %)	220(18.28 %)	
		rural road	30(7.48 %)	57(4.73 %)	
		other	1(0.24 %)	8(0.66 %)	
	Binary	Main street or road(MSR)	253(63.09 %)	867(72.07 %)	0.001
Vehicle type	Primary	light	349(87.03 %)	956(79.46 %)	<0.001
		Heavy	39(9.72 %)	50(4.15 %)	
		motorcycle	13(3.24 %)	197(16.38 %)	
	Binary	two-wheeled vehicle(TWV)	13(3.24 %)	197(16.38 %)	<0.001
The fault of the driver	Primary	driver at fault	318(79.30 %)	737(61.26 %)	<0.001
		driver not at fault	83(20.70 %)	466(38.74 %)	
	Binary	Blame the car driver(BCD)	318(79.30 %)	737(61.26 %)	<0.001
Passenger sex	Primary	Male	394(98.25 %)	1159(96.34 %)	0.065
		Female	7(1.75 %)	44(3.66 %)	
	Binary	Male passenger(MPA)	394(98.25 %)	1159(96.34 %)	0.065
Driver education	Primary	illiterate	13(3.24 %)	53(4.40 %)	0.205
		primary	27(6.73 %)	102(8.47 %)	
		Under diploma	87 (21.69 %)	175 (14.54 %)	
		diploma	263(65.58 %)	826(68.66 %)	

(*continued on next page*)

Table 2 (*continued*)

Variable	Coding	Category	Case(158)	Control(474)	P-value
		Academic education	11(2.74 %)	47(3.91 %)	
	Binary	Low educated driver (primary and illiterate) (LED)	40(9.98 %)	155(12.88 %)	0.124
Injury status of the driver	Primary	not injured	155(38.65 %)	783(83.48 %)	<0.001
		injured	125(31.17 %)	406(76.46 %)	
		dead	121(30.17 %)	14(10.37 %)	
	Binary	Injured driver (dead or injured)(IND)	246(61.34 %)	420(34.91 %)	<0.001
vehicle Involved in the accident	Primary	No other vehicle	184(45.88 %)	405(33.66 %)	<0.001
		Multiple vehicles	15(3.74 %)	69(5.76 %)	
		Light vehicle	113(28.17 %)	630(52.36 %)	
		Heavy vehicle	89(22.19 %)	99(8.22 %)	
	Binary	Accident with counterpart vehicle(AWCV)	217(54.11 %)	798(66.33 %)	<0.001

where 1 equals “L is True”, 0 equals “L is False.”

The above-given format (1) can include many forms of generalized linear models such as linear regression (identity link function: $g(E(y)) = E(y)$), binary logistic regression (logit link function: $g(E(y)) = \log\left[\frac{E(y)}{1-E(y)}\right]$), and Cox proportional hazard function as long as the score function is defined [9,10]. In general, a score function is defined for each model, which indicates the quality of the assumed model; for example, for linear regression, the score function could be the residual sum of squares, and for logistic regression, the score could be the deviance. In logic regression, the goal is to find the Boolean expression that minimizes the assigned score function. β_j is estimated simultaneously by Annealing algorithm when calculating the expression L_j.

2.2.1. Search algorithms

In practice, many logic combinations can be constructed using a given set of variables, and there is no direct method for listing all logic combinations that can facilitate evaluating all combinations and selecting the best model. Therefore, some search algorithms are needed to find the best scoring models. By defining the move set by a set of standard operations like splitting and pruning the tree, simulated annealing or greedy algorithms are executed to discover the best logic combination. While the greedy algorithm is quick, it does not always find the best scoring model. The simulated annealing algorithm usually does, but it is time-consuming. In practice, the simulated annealing algorithm is usually used to ensure the best-fitted models are found. The simulated annealing algorithm is a stochastic search algorithm in the state space S (each state represents a configuration of the problem under investigation), which seeks the best combination based on the specified score function [9].

2.2.2. Null model randomization test

Nevertheless, as usual, the best scoring model generally overfits the data, and techniques are required to separate signal and noise in the candidate models discovered by the search algorithms. A null model randomization test is an overall test of the signal in the data.

The relationship between the response and the independent variables can be checked using this test by comparing the scores obtained from the response’s random fit and the algorithm’s best logic model. We first find the best scoring model, given the data. The null hypothesis suggests that there is no relationship between X and Y. If no association is found between X and Y, the best model fit on the data with the response randomly permuted should yield about the same score as the best model fit on the original data. By repeating this test as often as desired, the proportion of scores better than the score of the best model on the original data can be considered an exact *P*-value, indicating evidence against the null hypothesis.

2.2.3. Cross-validation test

The optimal number of logic combinations and model variables can be determined using the cross-validation test to avoid over-fitting the model [9,47]. When searching through the annealing algorithm, it is possible to encounter a model with many combinations (i.e., logic trees) and variables (i.e., leaves), which increase the complexity of the model. The cross-validation method and sets of test and training are used to find the optimal number of trees/leaves and deal with the model’s complexity, resulting in the optimal size of the logic model with the best predictability [9,47].

The cases in the dataset are divided into m equally sized sets to examine the performance of the best model of size k compared to models of different sizes. For each m set, the cases from set i are removed from the data. The best scoring model of size k is found, using only the data from the remaining m - 1 sets, and its score is calculated by the cases in set i under this model. This yields score ε_{ki}. The cross-validated score for model size k is $\varepsilon_k = \left(\frac{1}{m}\right)\sum_i \varepsilon_{ki}$. So, the cross-validated scores for models of various sizes can be compared [9,47].

2.3. Statistical analysis

The frequency and percentage were reported as descriptive statistics. To analyze the data, the chi-square test was used for between-group comparisons. The logic regression model with the “logit” link function, deviance as the score function, and the annealing algorithm as the search algorithm was used to identify the interactions.

Table 3
Descriptive statistics of risk factors by cases and controls and results of chi-square test for driver data set.

Variable	Coding	Category	Case(158)	Control (474)	P-value
age	Primary		36.27(SD = 13.43)	035.40(SD = 11.89)	0.180
		Driver Aged>65(PA>65)	16(3.17 %)	18(1.17 %)	0.004
light	Primary	day	304(60.19 %)	1175(76.64 %)	<0.001
		night	162(32.07 %)	319(20.80 %)	
		sunrise	21(4.01 %)	3(0.19 %)	
		sunset	18(3.56 %)	36(2.34 %)	
	binary	No daylight(NDL)	201(39.80 %)	358(23.35 %)	<0.001
weather condition	Primary	Clear	428(84.75 %)	1430(93.28 %)	<0.001
		Non-clear	77(15.25 %)	103(6.72 %)	
	binary	Non-clear weather (NCW)	77(15.25 %)	103(6.72 %)	<0.001
Uneven condition	Primary	smooth	427(84.55 %)	1457(95.04 %)	<0.001
		The hill	22(4.35 %)	26(1.69 %)	
		The mountains	56(11.08 %)	50(3.26 %)	
	binary	Uneven area (UA)	78(15.45 %)	76(4.96 %)	<0.001
Area type	Primary	other	62(12.27 %)	34(2.21 %)	<0.001
		residential	88(17.42 %)	1147(74.82 %)	
		Agricultural area	105(20.79 %)	75(4.89 %)	
		non-residential	250(49.50 %)	277(18.06 %)	
	binary	Non-residential area(NRA)	417(82.57 %)	386(25.18 %)	<0.001
Accident mechanism	Primary	other	14(2.77 %)	47(3.06 %)	<0.001
		A collision between a vehicle and a motorcycle	83(16.43 %)	247(16.11 %)	
		Collision of a vehicle with a vehicle	204(40.39 %)	850(55.44 %)	
		Collision of a vehicle with several vehicles	41(8.11 %)	242(15.78 %)	
		Collision of a vehicle with a fixed object	28(5.54 %)	68(4.43 %)	
		Overturning and falling	109(21.58 %)	53(3.46 %)	
		Leaving the road	26(5.14 %)	26(1.69 %)	
	binary	Car overturning(COT)	109(21.58 %)	53(3.46 %)	<0.001
The location of the accident	Primary	Rider band	421(83.37 %)	1476(96.28 %)	<0.001
		the shoulder	13(2.57 %)	10(0.65 %)	
		roadside	53(10.49 %)	42(2.73 %)	
		outside the road	18(3.56 %)	5(0.32 %)	
	binary	Accident on the runway(AORW)	421(83.37 %)	1476(96.28 %)	<0.001
Road surface	primary	dry	456(90.29 %)	1435(93.60 %)	0.073
		wet	40(7.95 %)	75(4.89 %)	
		Frosty and snowy	9(1.78 %)	23(1.50 %)	
	binary	Unclean road surface(UCR)	49(9.70 %)	98(6.39 %)	0.013
The manner of the accident	primary	Head-on collision	178(35.25 %)	189(12.33 %)	<0.001
		Front to back	30(5.94 %)	492(32.09 %)	
		Front to the right side	45(8.91 %)	200(13.04 %)	
		Front to the left side	24(4.75 %)	198(12.91 %)	
		Back to the left side	4(0.79 %)	28(1.82 %)	
		Right side to right side	2(0.39 %)	23(1.50 %)	
		Left side to left side	5(0.99 %)	16(1.04 %)	
		Left side to right side	9(1.78 %)	118(7.69 %)	
		other	199(39.40 %)	248(16.17 %)	
		Front with a fixed object	9(1.78 %)	21(1.36 %)	
	binary	Head-on collision(HOC)	178(35.25 %)	189(12.33 %)	<0.001
The place of the accident	primary	in town	67(13.26 %)	1144(74.62 %)	<0.001
		suburban	429(84.95 %)	385(25.11 %)	
		rural road	9(1.78 %)	4(0.26 %)	
	binary	Suburban accident(SUA)	429(84.95 %)	385(25.11 %)	<0.001
road defect	primary	Non-defective Road	404(80.00 %)	1442(94.06 %)	<0.001
		defective Road	101(20.00 %)	91(5.94 %)	
	binary	defective Road (DR)	101(20.00 %)	91(5.94 %)	<0.001
type of road	primary	freeway	5(0.99 %)	15(0.97 %)	<0.001
		highway	15(2.97 %)	30(1.95 %)	
		main Street	48(9.50 %)	1001(65.29 %)	
		main road	245(48.51 %)	234(15.26 %)	
		sub street	6(1.18 %)	99(6.45 %)	
		other	1(0.19 %)	12(0.78 %)	
		side road	135(26.73 %)	102(6.65 %)	
		rural road	50(9.90 %)	40(2.60 %)	
	binary	Main street or road(MSR)	313(61.98 %)	1280(83.50 %)	<0.001
Road type	primary	one sided	42(8.31 %)	168(10.95 %)	<0.001
		Separated two-way	125(24.75 %)	774(50.48 %)	
		Non-separated two-way	338(66.93 %)	591(38.55 %)	

(*continued on next page*)

Table 3 (*continued*)

Variable	Coding	Category	Case(158)	Control (474)	P-value
	binary	Non-separate two-way road(NSTR)	338(66.93 %)	591(38.55 %)	<0.001
Vehicle type	primary	light	329(65.14 %)	1154(75.27 %)	<0.001
		Heavy	35(6.93 %)	117(7.63 %)	
		Motorcycle	123(24.36 %)	262(17.09 %)	
		Agricultural and industrial	18(3.56 %)	0	
	binary	two-wheeled vehicle (TWV)	123(24.36 %)	262(17.09 %)	<0.001
The fault of the driver	primary	Driver at fault	355(70.30 %)	808(52.71 %)	<0.001
		Driver not at fault	150(29.70 %)	725(47.29 %)	
	binary	Blame the car driver(BCD)	355(70.30 %)	808(52.71 %)	<0.001
Driver sex	primary	Male	499(98.81 %)	1432(93.41 %)	<0.001
		Female	6(1.18 %)	101(6.58 %)	
	binary	Male driver(MD)	499(98.81 %)	1432(93.41 %)	<0.001
shoulder	primary	Without shoulder	143(28.32 %)	1209(78.86 %)	<0.001
		Earthen shoulder	268(53.06 %)	209(13.63 %)	
		asphalt shoulder	94(18.61 %)	115(7.50 %)	
	binary	Road without shoulders(RWS)	143(28.32 %)	1209(78.86 %)	<0.001
Safety option	primary	No special safety equipment	488(96.63 %)	1430(93.28 %)	0.011
		ABS brake	10(1.98 %)	73(4.76 %)	
		Air Bag	7(1.38 %)	30(1.95 %)	
	binary	Vehicle without airbags and ABS brakes (ABS)	488(96.63 %)	1430(93.28 %)	0.007
The number of vehicles involved in the accident	primary	1	179(35.44 %)	186(12.13 %)	<0.001
		2	300(59.40 %)	1175(76.64 %)	
		More than two vehicles	26(5.14 %)	172(11.21 %)	
	binary	Multi-car accident(MCA)	326(64.55 %)	1347(87.87 %)	<0.001

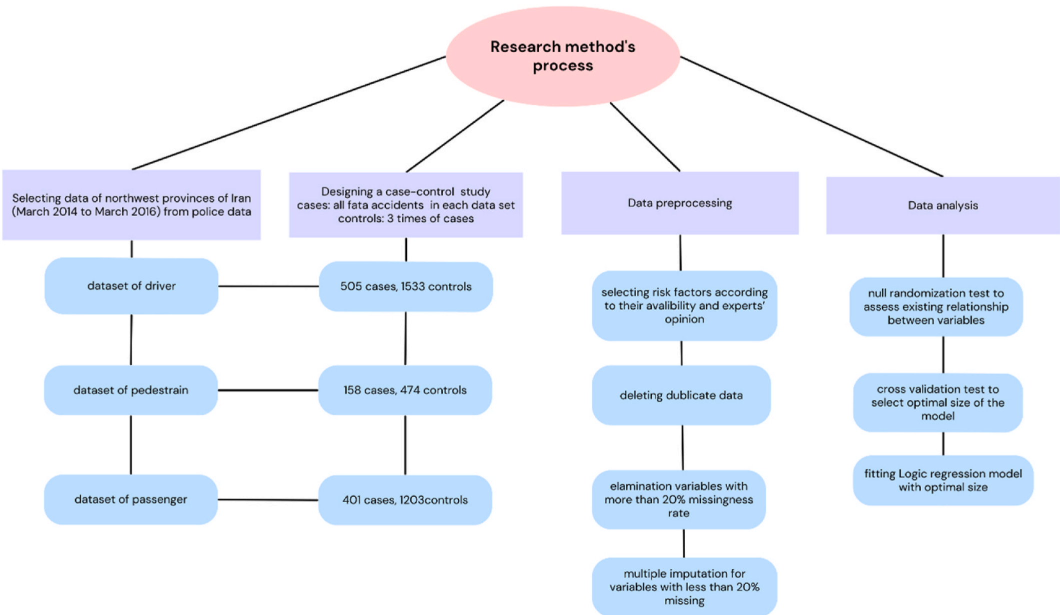

Fig. 1. Flowchart of the research method's process.

The null model randomization test with 500 replications was used to investigate and confirm the statistical relationship between the mentioned risk factors and DATAS. In addition, the optimal number of logic combinations and model variables were determined by adopting the 10-fold cross-validation method. Since we had no apriori idea of how many trees we maximally needed to fit, as a conservative choice, the ranges for the number of trees and leaves to search for optimal combinations by cross-validation test were considered 1–5 and 1–15, respectively.

All analyses were performed using R version 4.1.1 statistical software package as well as the "LogicReg" R package to fit Logic Regression models [48] and the "MICE" package for Multiple Imputation [49] in R.

3. Results

In this study, three case-control datasets were constructed. The first dataset was for pedestrians with 632 individuals (158 cases, 474 controls), the second one was for vehicle passengers with 1604 accidents (401 cases, 1203 controls), and the last one was for drivers with 2038 accidents (505 cases, 1533 controls). Overall, about 25 % of the investigated accidents were fatal (cases) and the rest were non-fatal (controls). Tables 1–3 present the descriptive statistics on the risk factors behind the fatal and non-fatal accidents as well as the results from chi-square test.

3.1. *Confirmation of a signal's existence through null model randomization test results*

Ten binary risk factors associated with fatalities in pedestrians, 17 binary risk factors associated with fatalities in passengers, and 19 binary risk factors associated with fatalities in drivers were entered into the logic regression model. First, a randomization test was performed for "Null model" to determine the relationship between response variable (fatal/non-fatal accident) and predictor variables. To perform these tests, three trees with the possibility of ten leaves were considered as the best model (i.e., a model with a more appropriate computational cost was selected because the existence of a relationship in this model guarantees the existence of a relationship in models with higher complexity). The Null model was fitted with a constant coefficient β_0 without any predictor variables. Table 4 shows the results from the null model randomization tests. The number of replications in this test was 500 times, and the deviance statistic of the best model in the dataset was obtained using the Annealing algorithm. All scores from the random models were worse than the score of the best model fitted to the original data; the results of one-sample *t*-test also suggested a significant difference between the mean score of random models with the score of the zero model. The Histogram of scores for the pedestrian dataset is presented in Fig. 2. These results, in other words, confirmed that the predictor variables contained some information to predict the status of DATAS in the three studied datasets.

3.2. *Results of cross-validation test to determine the optimal number of trees and leaves*

The 10-fold cross-validation test was used to determine the appropriate size (i.e., number of the Boolean combinations and variables) of the model. The best model has the lowest score in the cross-validation test, leading to the model with the best predictive ability.

Fig. 3 displays the effect of the model size on the deviance statistics of logistic regression corresponding to logic model fitted with 1–5 trees and 1–15 leaves in each one of the studied datasets; (Fig. 3 (A): Scores related to models with different number of trees and leaves in pedestrian data; Fig. 3 (B): Scores related to models with different number of trees and leaves in passenger data, Fig. 3 (C): Scores for models with different numbers of trees and leaves in the drivers' data).

According to Fig. 2, cross-validation scores had the lowest deviance in the pedestrian dataset of 5 trees with 7 leaves, in the passenger dataset of 4 trees with 11 leaves, and in the drivers' database of 5 leaves with 10 leaves. On the other hand, the differences among the deviance of these models with the 3-tree and 8-leaf model in the pedestrian dataset, with the 3-tree and 5-leaf model in the passenger dataset, and finally with the 3-tree and 9-leaf model in the driver dataset were small. Those models with fewer trees were considered as the final models due to the easier and practical interpretation.

3.3. *Results of fitting the "logic regression" models with optimal sizes*

The final models were fitted using the optimal sizes obtained from the results of cross-validation tests by using the Annealing algorithm introduced in the previous section. Tables 5–7 indicate the Boolean combinations obtained by the Annealing algorithm as well as the odds ratio of each combination based on the studied datasets.

The results of the fitted models for three datasets and listed in Tables 5–7 are reported below.

3.3.1. *Pedestrians*

"Suburban accident (SUA) or No-daylight (NDL)" combination ($L_1 = SUA \wedge NDL$) increased the pedestrians' chances of dying at the scene by almost seven times (OR = 6.87, P-value<0.001). Pedestrians aged >65 (PA>65) were about three times more likely to die at the scene (OR = 2.97, P-value<0.001). Also, in accidents which were "on a residential area and the urban accident, and the Pedestrian was in safe position or accident occurred on the two-way road without a separator (NSTR), and the colliding vehicle is not a heavy vehicle (HV)" ($L_3 = ((NRA)^c \wedge (SUA)^c) \wedge \left(\left((PUS)^C \vee NSR\right) \wedge (HV)^c\right)$), the chance of dying of pedestrian at accident scene was about

Table 4
The results of the null model randomization test to investigate the existence of a relationship between the risk factors and the death at the accident scene in the investigated data sets.

Dataset	Mean of deviance for 500 repetitions	Null model deviances	Best model deviances	P-value[a]
Pedestrian	687.56(SD = 5.75)	710.79	556.24	<0.001
Passenger	1774.82(SD = 5.93)	1803.97	1364.22	<0.001
Driver	2252.43(SD = 6.28)	2282.14	1446.65	<0.001

[a] P-value of one sample *t*-test for comparison of the average deviances obtained from 500 replicates with the zero model.

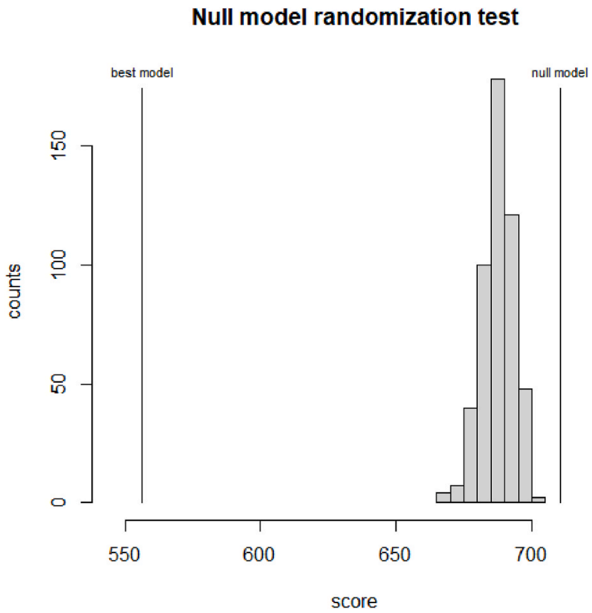

Fig. 2. Null model randomization test results for pedestrian deaths at the scene. Histogram of scores are obtained from the logic model with 3 trees and the possibility of the presence of 10 leaves in the model.

86 % lower (OR = 0.14, P-value<0.001) (Table 5).

3.3.2. Passengers

"Suburban accident (SUA) or car overturning (COT)" combination ($L_1 = SUA \vee COT$) increased the chances of vehicle passengers dying at the scene by approximately 8.5 times (OR = 8.55, *P*-value<0.001). "Occurrence of an accident on a defective road (DR)" increased the chance of DATAS for vehicle passengers by about 2.2 times (OR = 2.18, *P*-value<0.001). As for those accidents where "the vehicle was a two-wheeled vehicle (TWV) or the driver of vehicle was not injured (IND)" ($L_3 = (IND)^C \vee TWV$)), the chances of the passengers dying at the scene were about 75 % lower (OR = 0.25, *P*-value<0.001) (Table 6).

3.3.3. Drivers

"Head-on collision (HOC) or car overturning (COT) or two-wheeled vehicle (TWV)" ($L_1 = HOC \vee (TWV \vee COT)$) combination increased the chance of DATAS for vehicle drivers by about four times (OR = 4.03, *P*-value<0.001). "Accident on the roads other than runway (AORW) or not collision with more than one vehicle (MCA) or accident in a non-residential area (NRA)" ($L_2 = (AORW)^C \vee \left((MCA)^C \vee NRA\right)$) increased the chance of DATAS for vehicle drivers by about 6.5 times (OR = 6.40, *P*-value<0.001). Where the driver was male (MD) and accident occurred in suburban areas (SUA) or on defective roads (DR) ($L_3 = MD \wedge (SUA \vee DR)$) the odds ratio of DATAS for drivers was about 5.5 (OR = 5.40, *P*-value<0.001) (Table 7).

4. Discussion

This study mainly aimed to identify the interactions of the factors affecting DATAS as an indicator of the severity of the accident by adopting the logic regression method.

To this end, a case-control study was designed based on the data recorded by the police in three northwestern provinces of Iran during 2014–2016. The reason for adopting the logic regression method in this study was the great ability of this method and its search algorithms to identify complex interaction effects among the data. Our study results based on logic regression are discussed in the following section.

4.1. Pedestrians' dataset

According to the results from the logic regression of the identified interaction effects on DATAS, the occurrence of "the accident outside the city in a situation where there was insufficient light" significantly increased the chance of pedestrian death at the scene, which may have been attributable to the great severity of collisions in low light conditions in, especially, suburban accidents due to the high speed, failure to predict the pedestrians' behaviors, and insufficient visibility of the drivers. Several studies have documented the increase in the severity of pedestrian injuries in non-urban environments [37–39] and the effect of insufficient light on the severity of pedestrian accidents [50,51].

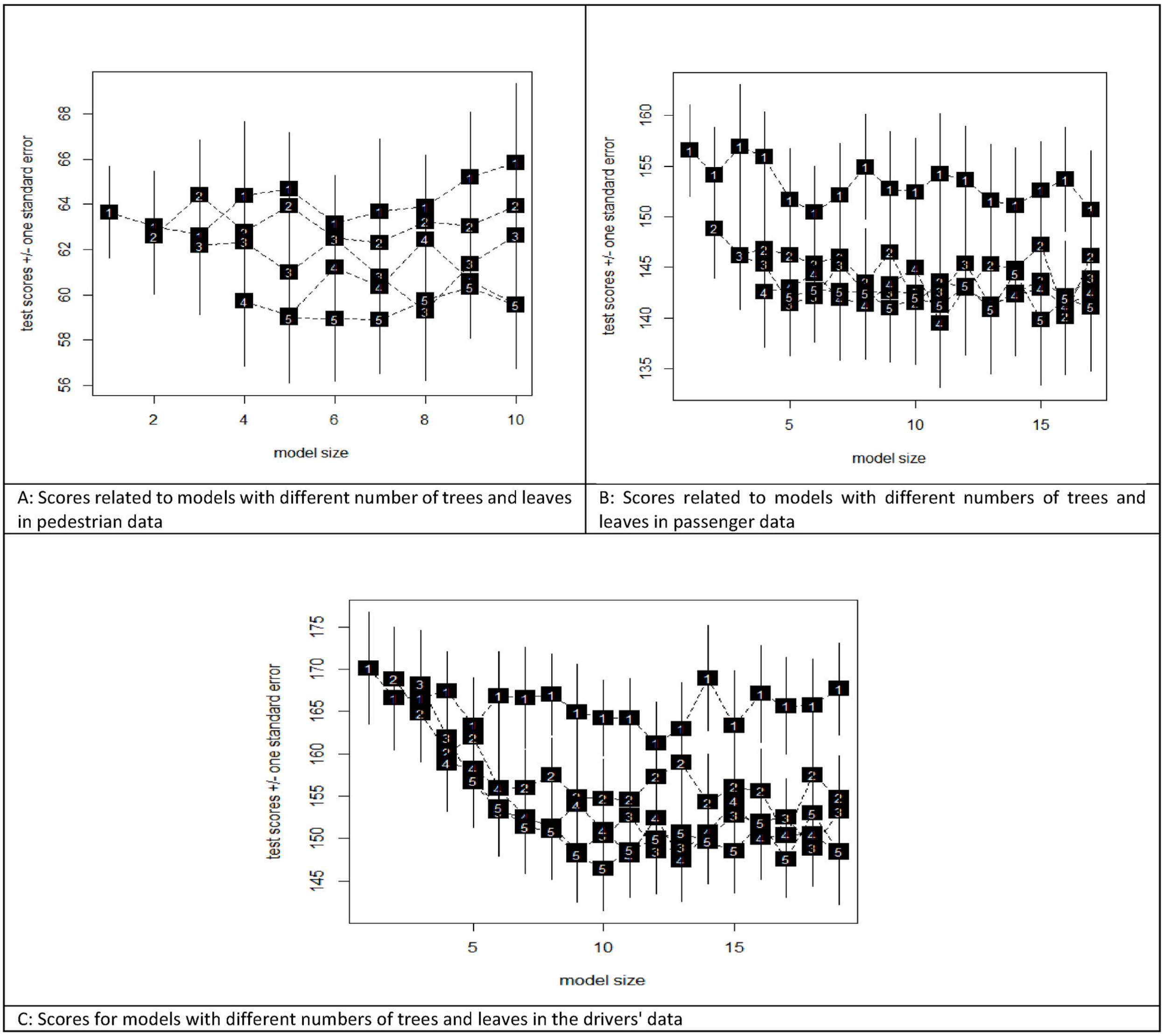

Fig. 3. The results of the 10-fold cross-validation test with different numbers of trees and leaves in the examined datasets.

Table 5
Boolean combinations found by the Annealing algorithm and their odds ratio for predicting the occurrence of death at the accident scene for pedestrians.

Dataset	combination	Boolean combination	OR	CI 95 %	P-value
Pedestrian	1	$L_1 = SUA \wedge NDL$	6.87	3.02–15.65	<0.001
	2	$L_2 = PA > 65$	2.97	1.65–4.39	<0.001
	3	$L_3 = ((NRA)^c \wedge (SUA)^c) \wedge \left(\left((PUS)^C \vee NSTR\right) \wedge (HV)^c\right)$	0.14	0.09–0.23	<0.001

Suburban accident (SUA), No-daylight (NDL), Pedestrian Aged>65 (PA>65), Non-residential area (NRA), Pedestrian in an unsafe position (PUS), Non-separate two-way road (NSTR), Heavy vehicle (HV).

Furthermore, “accidents happening in residential areas and presence of the pedestrians in the safe zone or roads without dividers and collision with a light vehicle” were less likely to cause pedestrian death at the accident scene.

According to the results from the previous studies, these factors reduced the chance of DATAS for pedestrian accidents and the simultaneous occurrence of these factors significantly lowers the given chance.

Pedestrian passing through an unsafe route was one of the factors that increased the chance of DATAS. Our study results were consistent with the findings from previous studies reporting that the severity of the accident was increased when pedestrians were in

Table 6
Boolean combinations found by the Annealing algorithm and their odds ratio for predicting the occurrence of death at the accident scene for passengers.

Dataset	combination	Boolean combination	OR	CI 95 %	P-value
Passenger	1	$L_1 = SUA \vee COT$	8.55	5.93–12.34	<0.001
	2	$L_2 = DR$	2.18	1.66–2.86	<0.001
	3	$L_3 = (IND)^C \vee TWV$	0.25	0.19–0.32	<0.001

Car overturning (COT), Suburban accident (SUA), Defective Road (DR), Injured driver (dead or injured) (IND) two-wheeler vehicle (TWV).

Table 7
Boolean combinations found by the Annealing algorithm and their odds ratio for predicting the occurrence of death at the accident scene for drivers.

Dataset	combination	Boolean combination	OR	CI 95 %	P-value
Driver	1	$L_1 = HOC \vee (TWV \vee COT)$	4.03	3.11–5.23	<0.001
	2	$L_2 = (AORW)^C \vee \left((MCA)^C \vee NRA\right)$	6.40	4.40–9.31	<0.001
	3	$L_3 = MD \wedge (SUA \vee DR)$	5.40	3.83–7.62	<0.001

Head-on collision (HOC), two-wheeler vehicle (TWV), Car overturning (COT), Accident on the runway (AORW), Multi-car accident (MCA), Non-residential area (NRA), Male driver (MD), Suburban accident (SUA), Defective Road (DR).

unsafe situations [24–27].

Crashing into a heavy vehicle, compared to other vehicles, significantly increased the chance of death, which may have been due to a greater severity of the accident. Several studies have also reported the effect of this factor on the severity of traffic accidents [24–27].

4.2. *Passengers' dataset*

According to results of the logic regression model, "occurring the accident outside the city together with overturning the vehicle" significantly increased the chance of DATAS in passengers, which was in agreement with the findings from other studies indicating that being outside the city and overturning the vehicle [30,31] increased the severity of an accident.

Where "The driver was not injured or driving two-wheeled vehicles", as other identified interaction effects, was found to lower the chance of the passenger dying at the scene due to the low severity of these accidents. Other studies also demonstrated that injury of driver [52,53] and driving two-wheeled vehicles [28,29] affected the severity of injury in passengers.

4.3. *Drivers' dataset*

According to the results of logic regression, the chance of dying at the accident scene was higher when "drivers had a head-on accident or the used vehicle was a two-wheeler or the mechanism of the accident was overturning".

Previous studies also revealed that these factors alone increased the chance of dying at the scene.

Overturning the car was one of the factors that increased the chance of DATAS, which was due to its severe impact on the driver and because of other factors (e.g., the vehicle roof collapse on the driver). Statistics show that more than 31 % of the deaths on highways in the United States are caused by the accidents associated with vehicle rollovers, although they account for a much smaller proportion of all accidents [54]. Other studies also reported the increased risk of the death in car overturning accidents, which was in line with our study result [30,31].

Head-on collision of a vehicle increased the chance of DATAS for drivers, which was attributed to the increased force sustained by both drivers. Our study results in this regard were consistent with the findings from other studies suggesting an increase in the severity of the incident in head-on collisions [32–34].

The drivers of two-wheeled vehicles had a higher chance of dying at the scene of the accidents, which was in line with the results from previous studies [28,29].

Our study results in this section suggested that a combination of these factors may also have increased the chance of dying at the accident scene.

Furthermore, "the occurrence of an accident in a place other than the road lane or the absence of a multi-car accident or an accident in a non-residential area" increased the chance of death.

The previous studies showed the impact of accidents in non-residential areas [35,36]. The chance of DATAS was higher when the "accident occurred outside the city or on the defective roads with male drivers". Also, according to the previous studies, accidents happening outside the city [37–39] or in the defective roads [40,41] increased the severity of them. Previous studies, moreover, reported that the male drivers had higher crash severity than female drivers [42–44].

In this regard, our study results suggested that the educational interventions on male drivers together with the improvement to safety features of the non-urban roads and defective roads may have reduced the severity of accidents in this group of people.

A handful of studies assessed the interaction effects related to traffic accidents and, to our best knowledge, only one study by Rohani-Rasaf et al. [55] investigated the interactions among various factors associated with fatal road traffic accidents by employing

Logic regression method used in our study. The given study found significant interactions among the road and driver factors, and concluded that the roads with poor design may have caused a driver to make mistakes and increased the fatal accident rate, which was relatively consistent with our study results.

Our study shares a similarity with the study conducted by Rohani-Rasaf et al. in that both studies employ logic regression to examine the impact of interactions associated with fatalities in road accidents. Also, in this study, police data were used. The difference between our study and the mentioned study is that they have also examined the death 30 days after the accident using the forensic database, in which many variables have been omitted due to the limited common variables in the police and forensic databases. However, our study examines the effects of interaction on death at the accident scene using police data, which collects more specific and complete information. Also, they assessed the data for one year (2014) for the whole country. Our study included accidents occurred in the northwestern provinces of the country for 3 years (2016–2014). Also, our study was designed as a case-control, which is different from the design of the aforementioned study. Furthermore, in the aforementioned study, the analyses were performed for all accidents without considering the type of road user. In contrast, in our study, due to the difference in the mechanism of the accident according to the road users, the analyses were performed separately for the driver, pedestrian, and passenger.

Another study by Chen et al. evaluated the main and interaction effects of commercial vehicle mix and roadway attributes on crash rates using a Bayesian random-parameter Tobit model. Results of this study demonstrated that the interaction between commercial vehicle and roadway attributes was strong enough to mediate the effect of commercial vehicles on crash rates, and the magnitude and direction of such mediation varies across the vehicle classes, crash severity levels, and roadway attribute type [6]. Their findings were in line with our study results, indicating the interaction effects between vehicle and road characteristics on DATAS.

Ahmed et al. [7] successfully predicted the road accidents and assessed their contributing factors by using machine learning models in order to consider the relationship and interaction among the features. To this end, they used the Shapley Additive exPlanation (SHAP) dependence plot to investigate the relationship and interaction of the factors with the outcome variable, and concluded that the factors like road category, the number of vehicles involved in an accident, and driver's age resulted in severe injuries.

Due to the dissimilarities in the methods, a direct comparison between our results and those of other studies is not feasible. However, their findings align with ours in terms of the elements that contribute to the severity of accidents, including the age of road users, road category, and type of vehicles.

In sum, as it mentioned, a traffic accident might be caused by several interacting factors. Specifically, the coexistence of some risk factors and the certain combination of them have a greater impact than their separate presence. For instance, according to the results of our study, if a pedestrian has an accident in a rural area during nighttime, the likelihood of him to have severe accident and die at the scene of accident is about seven times greater than the states that this combination is absent. This high-odds ratio has practical significance and can be used in accident prevention policymaking. By identifying effective combinations of factors, a number of factors involved in these combinations can be selected for control or preventive intervention to break combinations.

Our study revealed some significant interactions among risk factors associated with DATAS, which may have helped traffic policymakers, road users, and healthcare professionals reduce road injuries and enhance safety.

Moreover, logic regression, has shown great potential in finding such interaction effects. Employing logic regression method may have provided the researchers with results from a different perspective and shown the complex interactions among the risk factors.

This study aimed to encourage the researchers to pay more attention to interaction effects and use efficient methods such as logic regression to uncover the new aspects of traffic accidents. Future research can be designed more specifically and with detailed variables for a more accurate assessment of interactions, which will reduce road injuries.

5. Conclusion

In sum, the occurrence of an accident outside the city in a situation where there was insufficient light significantly increased the severity of the injuries for pedestrians. Accidents happening in residential inner-city areas and collision with a light vehicle together with presence of the pedestrians in the safe zone or on the roads without dividers, lowered the chance of dying at the scene for pedestrians. The age over 65 years increased the chance of dying on the scene; in other words, it increased the severity of the injury for pedestrians.

For passengers, "accidents happening in outside the city or overturn of the vehicle" combination and accidents on the roads with defects increased the chance of DATAS for passengers. When driver was not injured or the vehicle was two-wheeled, chance of DATAS decreased for passengers.

The chances of dying at the scene were higher for those drivers who had a head-on accident, drove a two-wheeler vehicle, or overturned their vehicles. The occurrence of an accident on the roads other than runway or the absence of a multi-car accident or an accident in a non-residential area increased the chance of death. The chance of DATAS was higher when the accident occurred outside the city or on defective roads and the drivers were male.

Based on our findings, it can be concluded that some of the obtained results are related to management factors and some are related to road users. Therefore, practical suggestions and recommendations resulting from the results of this study can be presented to reduce the severity of crashes in both the management and road users' domains. In the management sector, the quality, safety, and lighting of roads and in the domain of road users, sufficient skill and ability for road traffic, speed limit, and compliance with traffic rules play an important role in the severity of the accident, which leads to death at the scene. A combination of these factors and the simultaneous presence of several risk factors increase the severity of the injury. The practical recommendation of this study is to control at least one or more of the risk factors present in each of the identified multiple-factor combinations in order to break the combination and reduce the severity of the damage. This may help the policy-makers, road users, and healthcare professionals to promote road safety through prioritizing interventions focusing on the effect size of the simultaneous coexistence of crash severity determinants and not just the main effects of single risk factors or their simple two-way interactions. Future research is recommended for including the vehicle factors including the vehicle make, model, installed vehicle safety features, and vehicle body style to introduce new combinations with potentially higher effect sizes while adjusting for the number of passengers in each vehicle.

5.1. Strengths and limitations of the study

The novelty of this study lies in the fact that it used the Logic regression method for identifying the interactions among the risk factors related to fatal accidents. Our study, however, had some limitations. First, the variables examined in this study only included the information about the accident scene, and only the deaths at the accident scene were considered as the severe accidents. In addition, the study fails to account for other elements that could potentially impact the results, such as visibility information and speed management. Despite their significance, these variables receive minimal attention in the existing registration system.

Second, there were some issues with the data that was used, which was general registration data that might not have been accurate or full in all cases of traffic accidents.

Moreover, data collection accuracy and consistency may vary depending on the police officers involved. These could introduce potential biases or limitations in the generalizability of the findings. Additionally, we used multiple imputation techniques to impute missing data since we believed that the missing data were missing at random (MAR). If the missing data mechanism is not random, this could bring biases or inaccuracies into the study.

The variables must be binary in order for logic regression to find higher-order interactions among them. Therefore, we had to divide some variables—like age— into binary.

Funding

This study was funded by the Road Traffic Injury Research Center, Tabriz University of Medical Sciences under grant number 59390.

Ethics approval and consent to participate

Ethics approval and consent to participate in the national integrated Road Traffic Injury Registry system

The National Integrated Road Traffic Injury Registry (IRTIR) study was approved by the Ministry of Health and Medical Education (MOHME), Iranian Trustee for Traffic Knowledge Development and Road Traffic Injury Research Center (approval no.: 700/1482). The funding from MHOME came under contract number 700/D/581, signed by the Secretary for Research and Technology. The ethical approval for this study was obtained from the Ethics Committee of Tabriz University of Medical Sciences under number IR.TBZMED.REC.1396.465. IRTIR was also supported by WHO under contract number 2017/742294–0 and the related concept note. Verbal informed consent was obtained from all participants (or their proxies/legal guardians) of the study.

Ethics approval and consent to participate in the present study

The current study that was performed based on the police's registered data (IRTIR study), was approved by the Ethics Committee of Tabriz University of Medical Sciences under the code of ethics IR.TBZMED.REC.1396.1170. Participants (or their proxies/legal guardians) had provided informed consent to participate in the IRTIR study. The data used in the current study were anonymized before its application. All methods were performed following the relevant guidelines and regulations.

Consent for publication

Not applicable.

Data availability statement

The authors do not have permission to share data. So, data associated with our study has not been deposited into a publicly available repository.

CRediT authorship contribution statement

Milad Jamali-dolatabad: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Homayoun Sadeghi-bazargani:** Writing – original draft, Supervision, Investigation. **Saman Salemi:** Writing – review & editing, Writing – original draft, Visualization. **Parvin Sarbakhsh:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization.

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.

Acknowledgment

The authors would like to thank research consultation of "Clinical Research Development Unit of Alzahra Educational and Treatment Center" Tabriz University of Medical Sciences for kind supports. We also acknowledge Traffic Police of the Islamic Republic of Iran for sharing the required database.

References

[1] Organization W.H., Road Safety, 2020.
[2] M. Sengoelge, L. Laflamme, Z. El-Khatib, Ecological study of road traffic injuries in the eastern Mediterranean region: country economic level, road user category and gender perspectives, BMC Publ. Health 18 (1) (2018) 1–9.
[3] M. Belloumi, F. Ouni, Factors affecting the severity of motor vehicle traffic crashes in Tunisia, SAE International journal of transportation safety 7 (2) (2019) 99–128.
[4] A. Razzaghi, et al., Risk factors of deaths related to road traffic crashes in World Health Organization regions: a systematic review, Archives of Trauma Research 8 (2) (2019) 57–86.
[5] G. Robb, S. Sultana, S. Ameratunga, R. Jackson, A systematic review of epidemiological studies investigating risk factors for work-related road traffic crashes and injuries, Inj. Prev. 14 (1) (2008) 51–58.
[6] T. Chen, et al., Analysing the main and interaction effects of commercial vehicle mix and roadway attributes on crash rates using a Bayesian random-parameter Tobit model, Accid. Anal. Prev. 154 (2021) 106089.
[7] S. Ahmed, et al., A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance, Transp. Res. Interdiscip. Perspect. 19 (2023) 100814.
[8] P.R. Lucek, J. Ott, Neural network analysis of complex traits, Genet. Epidemiol. 14 (6) (1997) 1101–1106.
[9] I. Ruczinski, C. Kooperberg, M. LeBlanc, Logic regression, J. Comput. Graph Stat. 12 (3) (2003) 475–511.
[10] H. Schwender, I. Ruczinski, Logic regression and its extensions, Adv. Genet. 72 (2010) 25–45.
[11] I. Ruczinski, C. Kooperberg, M.L. LeBlanc, Exploring interactions in high-dimensional genomic data: an overview of Logic Regression, with applications, J. Multivariate Anal. 90 (1) (2004) 178–195.
[12] M.E. Shaik, M.M. Islam, Q.S. Hossain, A review on neural network techniques for the prediction of road traffic accident severity, Asian Transport Studies 7 (2021) 100040.
[13] J. Wang, et al., Analyzing the risk factors of traffic accident severity using a combination of random forest and association rules, Appl. Sci. 13 (14) (2023) 8559.
[14] C. Chen, et al., Investigating driver injury severity patterns in rollover crashes using support vector machine models, Accid. Anal. Prev. 90 (2016) 128–139.
[15] J. Abellán, G. López, J. de Oña, Analysis of traffic accident severity using decision rules via decision trees, Expert Syst. Appl. 40 (15) (2013) 6047–6054.
[16] C.M. Rocco, E. Hernandez-Perdomo, J. Mun, Application of logic regression to assess the importance of interactions between components in a network, Reliab. Eng. Syst. Saf. 205 (2021) 107235.
[17] A. Bellavia, et al., The use of logic regression in epidemiologic studies to investigate multiple binary exposures: an example of occupation history and amyotrophic lateral sclerosis, Epidemiol. Methods 9 (1) (2020).
[18] A.-M. Noone, et al., Machine learning methods to identify missed cases of bladder cancer in population-based registries, JCO Clinical Cancer Informatics 5 (2021) 641–653.
[19] D. Yoneoka, et al., Identification of optimum combinations of media channels for approaching COVID-19 vaccine unsure and unwilling groups in Japan, The Lancet Regional Health-Western Pacific 18 (2022) 100330.
[20] K.K. Nicodemus, J.D. Malley, Predictor correlation impacts machine learning algorithms: implications for genomic studies, Bioinformatics 25 (15) (2009) 1884–1890.
[21] K.K. Nicodemus, W. Wang, Y.Y. Shugart, Stability of variable importance scores and rankings using statistical learning tools on single-nucleotide polymorphisms and risk factors involved in gene× gene and gene× environment interactions, in: BMC Proceedings, 2007. BioMed Central.
[22] C. Im, et al., Genome-wide haplotype association analysis of primary biliary cholangitis risk in Japanese, Sci. Rep. 8 (1) (2018) 1–11.
[23] M. Rohani-Rasaf, et al., The role of interaction-based effects on fatal accidents using logic regression, Archives of Trauma Research 7 (4) (2018) 140–145.
[24] M. Jamali-Dolatabad, H. Sadeghi-Bazargani, P. Sarbakhsh, Predictors of fatal outcomes in pedestrian accidents in Tabriz Metropolis of Iran: application of PLS-DA method, Traffic Inj. Prev. 20 (8) (2019) 873–879.
[25] J. Liu, et al., Pedestrian injury severity in motor vehicle crashes: an integrated spatio-temporal modeling approach, Accid. Anal. Prev. 132 (2019) 105272.
[26] C. Xin, et al., 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, Analytic Methods in Accident Research 16 (2017) 117–132.
[27] X. Xu, X. Luo, C. Ma, D. Xiao, Spatial-temporal analysis of pedestrian injury severity with geographically and temporally weighted regression model in Hong Kong, Transport. Res. F Traffic Psychol. Behav. 69 (2020) 286–300.
[28] S.-B. Homayoun, J.-D. Milad, G. Mina, S. Parvin, Predictors of pre-hospital vs. hospital mortality due to road traffic injuries in an Iranian population: results from Tabriz integrated road traffic injury registry, BMC Emerg. Med. 22 (1) (2022) 1–8.
[29] Y. Katayama, et al., Factors associated with prehospital death among traffic accident patients in Osaka City, Japan: a population-based study, Traffic Inj. Prev. 19 (1) (2018) 49–53.
[30] R.J. Alharbi, V. Lewis, C. Miller, A state-of-the-art review of factors that predict mortality among traumatic injury patients following a road traffic crash. Australasian Emergency Care, 2021.
[31] C. Dengler, et al., Vehicle rollover detection using recurrent neural networks, in: 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2019. IEEE.
[32] P. Liu, W. Fan, Modeling head-on crash severity on NCDOT freeways: a mixed logit model approach, Can. J. Civ. Eng. 46 (6) (2019) 322–328.
[33] P. Liu, W. Fan, Modeling head-on crash severity with drivers under the influence of alcohol or drugs (DUI) and non-DUI, Traffic Inj. Prev. 21 (1) (2020) 7–12.
[34] P. Liu, W. Fan, Analysis of head-on crash injury severity using a partial proportional odds model, J. Transport. Saf. Secur. 13 (7) (2021) 714–734.
[35] F. Jahanjoo, H. Sadeghi-Bazargani, M. Asghari-Jafarabadi, Prediction of Road Traffic Fatalities in the Six Most Populous Provinces of Iran, 2022, pp. 2015–2016.
[36] P. Salamati, et al., High crash areas resulting in injuries and deaths in Tehran traffic areas from november 2011 through february 2012: a geographic information system analysis, Med. J. Islam. Repub. Iran 29 (2015) 214.
[37] M. Jamali-Dolatabad, H. Sadeghi-bazargani, S. Mousavi, Applying count time series to assess 13-year pedestrian mortality trend caused by traffic accidents in East-Azerbaijan province, Iran, Int. J. Inj. Control Saf. Promot. 29 (2) (2022) 239–246.
[38] V. Rahmanian, et al., Risk factors of mortality following road accident in southern Iran, Trauma Mon. 26 (4) (2021) 199–205.
[39] M. Yousefifard, et al., Risk factors for road traffic injury-related mortality in Iran; a systematic review and meta-analysis, Arch Acad Emerg Med 9 (1) (2021) e61.

[40] Y. Darma, M.R. Karim, S. Abdullah, An analysis of Malaysia road traffic death distribution by road environment, Sādhanā 42 (9) (2017) 1605–1615.
[41] K.B. Lankarani, et al., The impact of environmental factors on traffic accidents in Iran, J Inj Violence Res 6 (2) (2014) 64–71.
[42] G. González-Sánchez, et al., Road traffic injuries, mobility and gender. Patterns of risk in Southern Europe, J. Transport Health 8 (2018) 35–43.
[43] S. Mafi, Y. AbdelRazig, R. Doczy, Machine learning methods to analyze injury severity of drivers from different age and gender groups, Transport. Res. Rec. 2672 (38) (2018) 171–183.
[44] K. Obeng, Gender differences in injury severity risks in crashes at signalized intersections, Accid. Anal. Prev. 43 (4) (2011) 1521–1531.
[45] H. Sadeghi-Bazargani, et al., Developing a national integrated road traffic injury registry system: a conceptual model for a multidisciplinary setting, J. Multidiscip. Healthc. 13 (2020) 983–996.
[46] S. Van Buuren, J.P. Brand, C.G. Groothuis-Oudshoorn, D.B. Rubin, Fully conditional specification in multivariate imputation, J. Stat. Comput. Simulat. 76 (12) (2006) 1049–1064.
[47] I. Ruczinski, Logic Regression and Statistical Issues Related to the Protein Folding Problem, 2001.
[48] C. Kooperberg, I. Ruczinski, M.C. Kooperberg, Package 'LogicReg', 2021.
[49] S. van Buuren, K. Groothuis-Oudshoorn, Mice: multivariate imputation by chained equations in R, J. Stat. Software 45 (3) (2011) 1–67.
[50] D. Li, et al., Analyzing pedestrian crash injury severity under different weather conditions, Traffic Inj. Prev. 18 (4) (2017) 427–430.
[51] M. Nasri, K. Aghabayk, A. Esmaili, N. Shiwakoti, Using ordered and unordered logistic regressions to investigate risk factors associated with pedestrian crash injury severity in Victoria, Australia, J. Saf. Res. 81 (2022) 78–90.
[52] R. Kakkar, et al., Road traffic accident: retrospective study, Indian J. Sci. Res. 5 (1) (2014) 59.
[53] N. Sae-Tae, A. Lim, N. Dureh, Determinants of severe injury and mortality from road traffic accidents among motorcycle and car users in Southern Thailand, Int. J. Inj. Control Saf. Promot. 27 (3) (2020) 286–292.
[54] C. Wang, et al., A vehicle rollover evaluation system based on enabling state and parameter estimation, IEEE Trans. Ind. Inf. 17 (6) (2020) 4003–4013.
[55] Rohani-Rasaf Marzieh, Mehrabi Yadollah, Hashemi-Nazari Seyed Saeed, Azizmohammad Looha Mehdi, S. Hamid, The role of Interaction?Based effects on fatal accidents using logic regression, Archives of Trauma Research. 140–145 (2018) سال هفتم.