IJTE, Vol. 11, No. 2, Autumn 2023, pp. 1401-1418
Research Paper

A Probabilistic Density Approach for Evaluating Factors Contributing to Injury Severity (Case Study: Borujerd-Khorramabad Rural Highway)

Shahab Hassanpour[1,*], Shahriar Afandizadeh[2]

Received: 2022/07/17 *Accepted: 2023/01/29*

Abstract

Road accidents reduce traffic safety due to injuries and fatalities. Investigating and prioritizing factors contributing to road accidents have been based on deficient traditional ways as they do not consider the probability density of factors contributing to road accidents. Accordingly, an examination of accident road factors based on the probability density seems necessary. Thus, this paper first aimed at using principle components analysis (PCA) as a statistical prioritization tool for identifying the main and sub-main factors that contribute to injury severity on Borujerd-Khorramabad as a four-lane rural highway during the years 2015 to 2017. Secondly, the multivariate Gaussian probability model was used as a probabilistic density approach to estimate the probability density based on the relationship between factors that contributing to injury severity and the Pearson correlation. The results obtained through PCA indicated that factors contributing to injury severity were ranked in terms of Eigen values and rotated component matrix. Findings from the PCA model showed that OS, PSL, AADT, SL, R, and S, as 6 important factors affecting the accident occurrence relevant to injury severity. The results of the probability density also showed that the relation between operating speed with posted speed limits, and the relation among segment length, operating speed and radius are considerable due to increasing the probability density of accident occurrence. Moreover, AADT with operating speed and operating speed with slope and radius have significant effects on the probability density of occurring accidents. The results of the present study show that applying a multivariate Gaussian probability model helps to estimate the probability density of the accident occurrence of factors contributing to road accidents based on their values.

Keywords: Road traffic accidents, statistical method, PCA method, multivariate Gaussian model

* Corresponding author email address: sh.hassanpour@abru.ac.ir
[1]Assistant Professor, Faculty of Civil Engineering, Ayatollah Boroujerdi University, Boroujerd, Iran
[2] Full Professor, Faculty of Civil Engineering, Iran University Science and Technology Tehran, Iran

1. Introduction

Road traffic accidents (RTAs) are increasingly known as one of the most concerning issues for public health due to imposing fatalities, injuries and severe socioeconomic costs worldwide (Gopaul et al., 2016; Tang et al., 2019). Haagsma et al. (2016) indicate that, in developing countries, rapid economic growth and a significant increase in motor vehicles have caused many fatalities, serious injuries, and financial costs for citizens. Recently, Iran, a developing country in the Eastern Mediterranean region, has had high mortality rates, compared to the other countries in the region, due to road traffic accidents (Bahadorimonfared et al., 2013; Sadeghi-Bazargani et al., 2016; Hosseinian et al., 2021). Road traffic accidents are ranked the second leading cause of death in Iran (Forouzanfar et al., 2014). According to a report released by the World health organization (WHO) (WHO, 2015), 24,896 death incidents, approximately 32.1 per 100,000 population of Iran, were involved in road traffic accidents in 2015. Based on comparisons across the road-traffic accidents death rates in Iran and the global average rate reported that statistically speaking, one out of 200 men dies in road traffic accidents (Forouzanfar et al., 2014). Road accidents are the most ten cause of death in the world. The global statistics about the consequences of road accidents on human life reported by WHO show that 1.35 million people died, and 50 million people injured (WHO, 2019). Furthermore, road safety issue has been involved with human, vehicle, road, and environmental characteristics (Shirmohammadi et al., 2018). Moreover, the analysis of road accidents has been a great interest of safety researchers and road safety managers. On rural roads, the number of fatality in Iran was 17,000 significantly in 2017 in which 11,000 of the accidents belonged to rural roads (LMO, 2017). Specifically, Lorestan province has been dramatically observed an increase in the number of road accidents on rural roads, and it is ranked as one of the ten provinces of the country in terms of road traffic accidents (Hejazi and Alipour, 2016; Mohammadfam et al., 2016). Accident analysis, due to having complex factors, has been overestimated, and all safety strategies and prevention plans have been carried out deficiently. So, in order to execute safety strategies and reduce accident accurately, there is a need to investigate factors contributing to the probability of accident occurrence regarding the influencing factors (Mahdieh Rad et al., 2016; Liu et al., 2018; Afandizadeh and Hassanpour, 2020).

A combination of several complex and interacting factors, including road users, the vehicle, roadway, the environment, and driver's reaction have been identified as the main reasons for road accidents (Eisenberg, 2004; Mahdieh Rad et al., 2016; Das et al., 2021; Khedher et al., 2022). Thus, in road safety, evaluating the occurrence probability of factors with each other should be an important priority for safety engineers involved in road accidents to examine the relation between factors contributing to traffic accidents and propose safety countermeasures to reduce accidents and improve safety on roads. To evaluate the factors contributing to road accidents and, thereby improve safety, the present study firstly aims at investigating the number of factors that are commonly reported as the most important factors influencing road accidents. These are described as operating speed (OS), posted speed limits (PSL), segment length (SL), and annual average daily traffic (AADT), slope (S), and the radius (R) of curvature. It is thought that understanding the effect of these factors and their relationship with the probability of accident occurrence is necessary in taking steps to reduce accidents. Secondly, the present study examines the effect of these factors and their relationship with injury severity using the PCA method and multivariate Gaussian probability model. Further, based on the evaluation of the proposed methods, the most influencing factors and their relationship among factors are examined.

The rest of the paper is organized as follows: The second section focuses on the recent works related to factors influencing accidents and provides a discussion of the factors that are commonly reported as important in road accidents. The third section proposes a new framework for identifying factors influencing injury severity and their impacts on the probability of accidents causing injury severity. Then, the estimation results along with a discussion on the findings are presented based on the proposed methods. At the end, the concluding remarks are explained regarding the obtained results.

2. Literature Review

Many researchers attempted to investigate factors leading to road accidents. Road-related accidents' factors are categorized as speed, and geometric characteristics. Finding these factors helped researchers to identify countermeasure solutions related to reducing accidents and increasing traffic safety on rural roads (Kononov et al., 2008; Mergia et al., 2013; Arévalo-Támara et al., 2020; Fu and Sayed, 2021). Though, Boodlal et al. (2015) found that geometric characteristics have not directly caused accident severity. They found these characteristics affect the likelihood of accident occurrence. In another study, Haghighi et al. (2018) examined the impact of roadway geometric characteristics on accident severity on a rural two-lane. They indicated a strong association between some geometric characteristics and accident severity outcomes.

Among traffic conditions, average speed, speed limit, and speed variation are the most influencing factors leading to the concurrence of accident frequency (Saha et al., 2016; Wu et al., 2021). Traffic volume, due to its relation to density, and congestion, has also been expressed as a common cause of accident rates on highways (Golob and Recker, 2003; Wang et al., 2009). Other studies (Wang et al., 2014; Retallack and Ostendorf, 2019) reported that by increasing traffic volume, the possibility for accident frequency is also increased significantly. Retallack and Ostendorf (2019) reviewed some papers that explored the role of traffic congestion on accidents. Their study disclosed an inverse relationship between accidents and congestion.

The relationship between operating speed and accident was also evaluated. Bird and Hashim (2006) reported fewer accidents when operating speeds were increased. Yu et al. (2018) investigated the impact of data aggregation approaches on the relationships between operating speed and traffic safety. They found that there is a positive relationship between speed and accident frequency. Ma et al. (2010) also indicated that roads with heavier traffic volume, more road lanes and higher speed limits have a high tendency to increase accident frequency.

Koorey (2009) also attempted to show segment length, including fixed and variable segments, as important factors in accident rates. Shirmohammadi et al. (2018), for instance, found out that segment lengths have significant effects on traffic accidents. Recently, probabilistic models have been proposed to predict accident numbers based on the influencing factors on urban segments (Ferreira et al., 2015; Shirmohammadi et al., 2019; Sangare et al., 2021).

In order to prioritize road traffic factors associated with traffic accidents, Elyasi et al. (2017) focused on prioritizing the influential factors in sub-urban accidents using a network analysis process. Sadeghi et al. (2013) also used the data envelopment analysis (DEA) technique for prioritizing accident-prone sections (APSs) with regard to traffic, geometric and environmental characteristics of road. Later, studies were developed based on the identification and ranking of hotspot segments for road safety improvement using statistical and participatory methods (Coll et al., 2013; Shen et al., 2019; Xie et al., 2020; Zhang et al., 2021). For modifying statistical prioritization methods, dynamic segmentation methods were used to identify accident prone segments. The results indicated that dynamic segmentation methods have better performance in comparison with statistical prioritization methods (Boroujerdian et al., 2014; Fountas et al., 2018; Assi, 2020). Martins and Garcez (2021) applied a multidimensional and multi-period analysis of road safety to prevent and mitigate the risks of traffic accidents. They showed that different dimensions that can influence traffic accidents are related to the characteristics of the road and its traffic. Ullah et al. (2021) evaluated factors influencing injury severities of motor vehicle crashes on the National highways of Pakistan. They indicated that the most important factors influencing road accidents are driver characteristics, crash characteristics, highway characteristics, temporal characteristics, and environmental characteristics. Moreover, the results were helpful for traffic managers and road engineers in prioritizing road sections for improvements and implementing suitable road safety interventions. Drosu et al. (2021) also proposed a fatal injury risk model as a probabilistic approach for rainy conditions on highways based on driver characteristics, crash characteristics, highway characteristics, and temporal characteristics. They showed that highways are dangerous during rains since fatal accidents are 3.340 times higher than on streets. de Moura et al. (2022) also found that identification and prioritization techniques are necessary on urban roads to determine the most hazardous locations and segments for improving road safety and further modification regarding countermeasures. These identification and prioritization techniques are categorized as the statistical and dynamic prioritization methods for road safety engineers to evaluate accidents.

Studies reviewed in the present paper reveal a significant gap in the previous research with regard to identifying and prioritizing factors causing accidents, Moreover, understanding the relationship between traffic and roadway characteristics that contribute to road accident is missing in the literature. Research studies have not, yet, comprehensively examined a reasonable level of certainty occurrence of factors causing accident. Accordingly, the current investigation is an attempt to investigate the relation between factors influencing road accidents to prioritize factors regarding PCA as a statistical method. However, mere statistical prioritization is not enough due to lack of occurrence likelihood of factors with each other. This need motivated researchers to look up for other exact solutions to discover ways of the probabilistic models for the probability occurrence of factors affecting accident. In addition, the present study applies new models for increasing the level of certainty based on the probability. The novelty of this paper is to firstly prioritize the most effective factors in injury severity using PCA as a statistical prioritization method on Borujerd-Khorramabad as a rural highway during years 2015 to 2017. Secondly, multivariate Gaussian probability model, to estimate the likelihood of occurrence for factors contribute to accident severity. Finally, the results are compared based on the literature review.

3. Research Method

Evaluating factors causing road accidents in current study, encompassed the methodological approach and procedures briefly displayed as a flowchart in Figure 1. According to Figure 1, it can be shown that the present study identifies and prioritizes factors influencing injury severity based on the following steps as given in Figure 1:

- Data collection and examination of the accident factors and injury severity for data accidents of Borujerd-Khorramabad as a rural highway during years 2015 to 2017.

- Initial statistical analysis using Pearson correlation to investigate the relation among accident factors, and using PCA method model to prioritize accident factors causing injury severity.
- Applying multivariate Gaussian probability to estimate the probability density of the factors relevant to accidents of injury severity based on the correlation values of Pearson analysis.

3.1. Data Collection

The accident data were collected from the database of the traffic police on Borujerd - Khorramabad which is a four-lane rural highway during years 2015 to 2017 as shown in Figure 2(a). The location map of the study area is shown in Figure 2(b). The length of the roads is 108 km. In total, there were 987 accidents belonged to the injury, respectively. Due to the importance of accidents affecting injury in the present study, the accidents factors contributing to injury severity based on the datasets are posted speed limits (PSL), Operating speed (OP), segment length (SL), and annual average daily traffic (AADT), slope (S), and radius (R). Further, the data for injury severity and other factors were examined and coded for using in the initial analysis, Pearson analysis and PCA model. The segment length was also evaluated based on division of the total road into 106 segments with minimum and maximum lengths 0.1 and 5.2 km considering homogenous characteristics, respectively.

The initial statistical analysis for each variable is summarized in Table 1. Table 2 also indicates a correlation analysis between factors influencing injury severity. From Table 2, it can be shown OS, PSL, AADT, SL have significant effects on injury severity. However, S, and R have an indirect effect on injury severity. Thus, based on the summary statistics in Tables 2 and 3, one can infer that the accident occurrences are statistically dependent on factors influencing road accidents, in other words, this correlation influences the probability of accident occurrence among factors. Thus, OS, PSL, AADT, and SL have direct correlations with increasing accident occurrence and S and R have indirect correlations with accident occurrence. The direct correlation shows that an increase in OS, PSL, AADT, and SL leads to an increase in the probability of accident occurrence on injury severity. However, the indirect correlation shows a reverse relation between factors and the probability of accident occurrence.

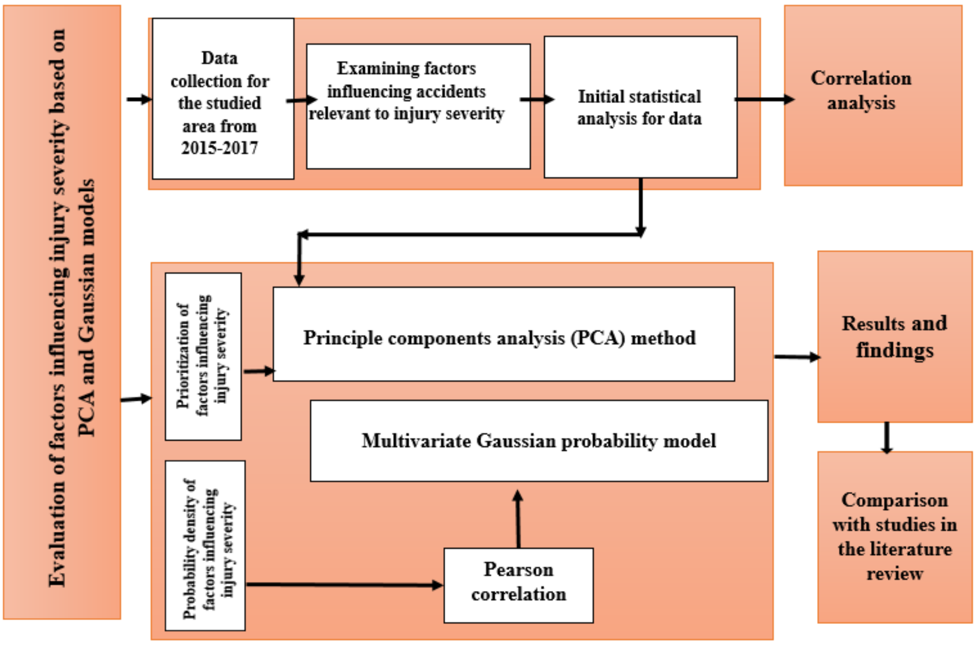
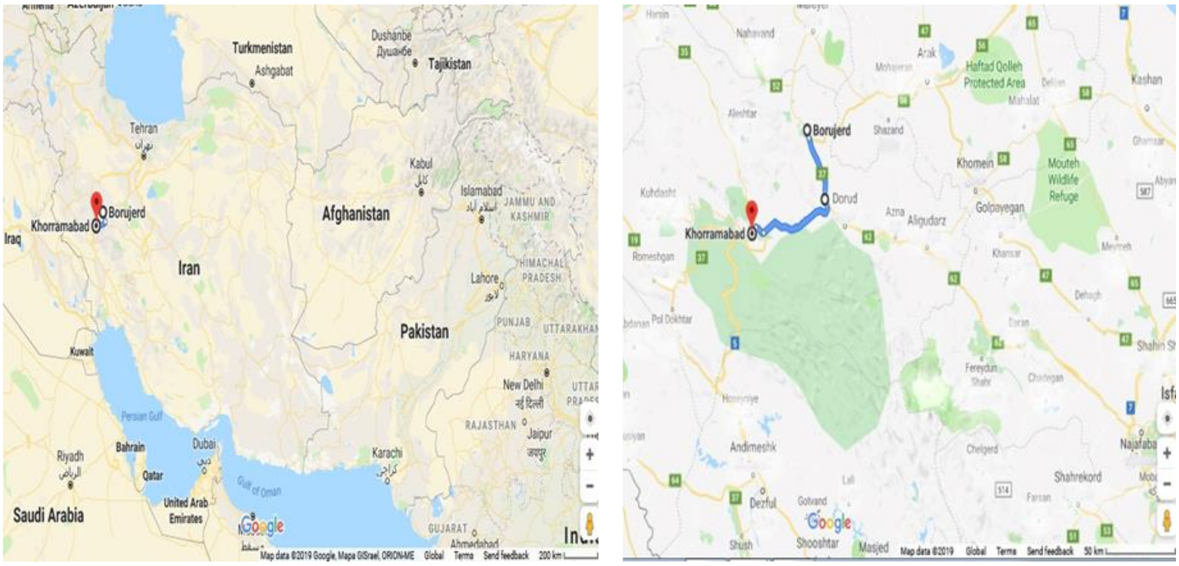


Figure 1. The Flowchart of the proposed method



(a) Location map of the study area (b) Plan of Borujerd -Khorramabad road

Figure 2. Schematic map of the case study (Source: Google map)

Table 1. Statistical analysis for factors influencing injury severity

Variables	Minimum	Maximum	Mean	S.D
PSL (km/hr)	50.100	106.000	83.871	12.391
SL (km)	0.100	5.200	2.683	0.879
AADT (flow*1000 veh/day)	10.835	22.267	12.576	4.519
OS (km/hr)	48.000	113.00	86.689	11.694
S (%)	-5.00	5.00	-0.83	1.06
R (m)	0.00	680.00	334.87	177.87

Note. Posted speed limits (PSL); Segment length (SL); Annual Average Daily Traffic (AADT); Operating speed (OS); Slope (S); Radius (R); S.D (Standard Deviation).

Table 2. Pearson correlation analysis between each factor and injury severity

Factors	Correlation value
PSL	+0.893
SL	+0.820
AADT	+0.861
OS	+0.910
S	-0.783
R	-0.720

Note. Posted speed limits (PSL); Segment length (SL); Annual Average Daily Traffic (AADT); Operating speed (OS); Slope (S); Radius(R)

Table 3. Pearson correlation analysis between factors contributing to injury severity

Factors	PSL	SL	AADT	OS	S	R
PSL	1	0.20	- 0.10	0.85^{*}	-0.31	0.15
SL	---	1	0.23	0.70^{*}	0.16	0.56^{*}
AADT	---	---	1	-0.75^{*}	0.11	0.44
OS	---	---	---	1	-0.64^{*}	0.78^{*}
S	---	---	---	---	1	0.18
R	---	---	---	---	---	1

Note. Posted speed limits (PSL); Segment length (SL); Annual Average Daily Traffic (AADT); Operating speed (OS); Slope (S); Radius(R); $^{*}P<0.01$

3.2. PCA Method

Principal component analysis has been widely used by researchers to investigate influencing factors on outcome experiments. This method has a mathematical basic, which is known as a statistical data reduction method (Yi et al., 2015). PCA has many applications in traffic safety and analyzing the most important factors influencing road accidents (Timmerman, 2010; Youming et al., 2018; Kassu and Hasan, 2019). Figure 3 represents a summary of the PCA procedure undertaken to identify the factors influencing accidents of injury severity. According to Figure 3, PCA method relies on the following basic four steps:

1. The mean of data is subtracted and they are assumed as (x_1, x_2,..,x_m). In this method, it is necessary to evaluate the suitability of data regarding Kaiser-Meyer-Olkin (*KMO>0.5*), Bartlett Test ($p<0.05$) and Bartlett Test ($p<0.05$).

2. The covariance matrix of this mean dataset is subtracted.

3. The eigen values and eigen vectors of the covariance matrix found in step 2 are calculated. In this step, it is useful to make a rotation of the vectors (Ledesma et al., 2007).

4. A feature vector is formed by selecting the eigenvectors with the largest eigenvalues. The eigen values greater than 1 (eigen>1) are selected during the construction of the model (Ledesma et al., 2007). By considering PCA method, a relationship between independent and dependent variables is determined. In this method, multi-linear regression analysis is obtained as the result of multiplying original independent variables x_{ik} by eigenvectors as shown in Eq. 1.

$$\begin{cases} y_t = \beta_0 + \left(\beta_1 PC_1 + \beta_2 PC_2 + ... + \beta_k PC_k\right), \\ y_t = \beta_0 + \sum_{k=1} \beta_k PC_k + u_t. \end{cases} \tag{1}$$

Where, β_0: is intercept, β_1, β_2,..,β_k: are the regression coefficients. PC_1, PC_2,...PC_n: are basic components, and u_i: is the error during regression.

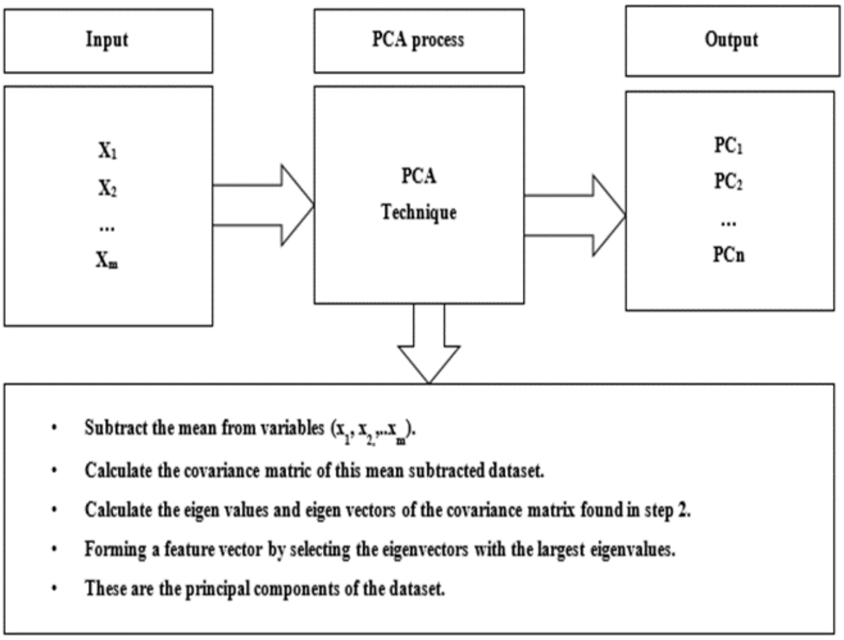


Figure 3. PCA procedure for identification of factors influencing injury severity

3.3. Multivariate Gaussian Model

In mathematics, probability theory has been mostly used as a framework of modelling and simulating the likelihood of occurrence of different possible outcomes in an experiment. To extend the feasibility of this theory, some researchers have developed multivariate Gaussian probability density model to identify the likelihood of phenomenon and events (Ribeiro, 2004; Chen et al., 2008; Stroock, 2010; Sangare et al., 2021). In order to understand the form of this model, it is needed to define it in the following steps (Ahrendt, 2005; Nawri, 2019): The first step is defining a vector of variables as independent parameters $X=\left[X_1,X_2,...,X_N\right]$ for the dependent variable. If we assume that the independent variable contributed in a linear equation, the following equation can be written as follows:

$$Y=a_1X_1+a_2X_2+...+a_NX_N. \qquad (2)$$

Where, $a_i \in R$ is a (univariate) Gaussian distribution.

The second step is to obtain the mean values between independent and dependent variables. In this study, mean values are assumed as $\mu_1,\mu_2,...,\mu_N$.

The third step is to calculate the standard deviation of the independent variables in the equation. In our study, standard deviations of variables are defined as $\sigma_1,\sigma_2,...,\sigma_N$.

Finally, the multivariate Gaussian probability density is defined as $x_1,...,x_k$, with in the Eq. (3) (Ahrendt, 2005):

$$f_x\left(x_1,...,x_k\right)=\frac{1}{\sqrt{(2\pi)^k \det\sum.}}\exp\left(-\frac{1}{2}(x-m)^T\sum{}^{-1}(x-m)\right). \qquad (3)$$

where x is a real k-dimensional column vector and $\det\sum.$ is the determinant of $\sum.$, also known as the generalized variance. The equation above reduces to that of the univariate normal distribution if $\sum.$ is a 1×1 matrix (i.e. a single real number).

Eq. (3) can be simplified as Eq. (4) (Prince, 2012):

$$f(z)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y}+\frac{(y-m_y)^2}{\sigma_y^2}\right)\right). \qquad (4)$$

Where, ρ is the correlation values, σ_x and σ_y are the standard deviations in the equation for variables x, and y, respectively. Furthermore, m_x and m_y are the mean values in the equation for variables x, and y, respectively.

4. Results and Discussions

The obtained results regarding the PCA and the probabilistic models for finding the most influential factors on the accident occurrence for injury severity are explained as given:

4.1. PCA Method

Principal component method is the most commonly applied for determining a first set of loadings and seeks values of the loadings and estimates the observed variances. The correlation matrix obtained through the PCA method clarified the existence of two factors, represented by the number of eigen values, to be greater than one. The extraction factors that accounted for 63.411% of the total variance of the variables are shown in Table 4. It is inferred that the 6 principal components can be used to fit the regression model regarding the variables into the model. In summary, two main and sub-factors with their correlation values, summarized in Table 5, were obtained as given:

- **The First Main Factor**

The most important factor (total variance= 42.514) with the greatest impact on accident occurrence included four variables: OS, PSL, and SL with components (0.899, 0.849, 0.772), respectively (See Table 5).

- **The Second Main Factor**

As illustrated in Table 5, the second main factor is in terms of importance in the interpretation of the relationship between the variables, with 20.897 of the total variance, contains two variables, namely, AADT, R, and S with components (0.835, -0.740, -0.710), respectively. Thus, based on PCA method outcomes, the most influential factors contributing to accident occurrence of injury severity regarding the maximum correlation values were identified and prioritized as (1) OS, (2) PSL, (3) AADT, (4) SL, (5) R, and (6) S, respectively.

Table 4. Initial Eigen values and rotation sums of squared loadings

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative (%)	Total	% of Variance	Cumulative %
1	2.150	43.004	43.004	2.126	42.514	42.514
2	1.020	22.100	65.104	1.045	20.897	63.411
3	0.926	13.120	78.224			
4	0.631	11.600	89.824			
5	0.273	6.450	96.274			
6	0.187	3.726	100			

Table 5. Rotated component matrix

Rotated Component Matrix		
Accident factors	**Component**	
	1	**2**
PSL	0.849*	0.017
SL	0.772*	0.017
AADT	0.063	0.835*
OS	0.899*	0.156
S	0.131	-0.710*
R	0.065	-0.740*

Note. Posted speed limits (PSL); Segment length (SL); Annual Average Daily Traffic (AADT); Operating speed (OS); Slope (S); Radius (R); $^{*}P<0.01$

4.2. Multivariate Gaussian Model

In the present study, multivariate Gaussian probability was used to estimate the probability density of two factors based on the Pearson correlation analysis in Table 3. The results are obtained and displayed in Figures 4 to 8.

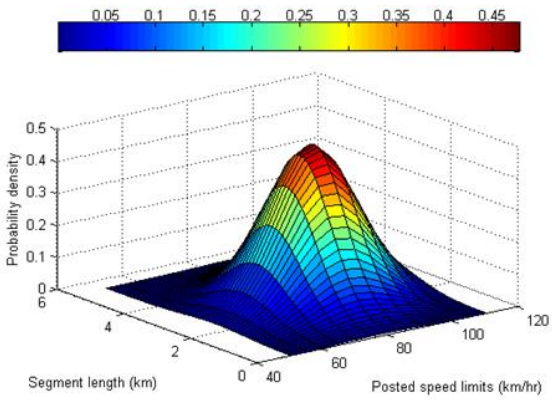
Regarding the Gaussian model outcomes in Figures 4(a) to 4(e), the posted speed limits have the maximum probability density with the operating speed. Thus, these factors have a strong relationship with each other to occur accidents causing injury severity. As the posted speed limits and the operating speed increase, the probability density for accident occurrence regarding these factors increases. Further, a decrease in slope and an increase in the posted speed limits lead to an increase in the probability density for accident occurrence. Thus, regarding Figure 4(d), the highest probability density regions can be seen approximately at speeds 80 to 100 km/hr and slopes less than zero. Other factors have less probability density compared with the operating speed. This shows that the probability density of accident occurrence for these parameters with the posted speed limit is below 0.5. Thus, operating speed and slope strongly affect posted speed limits to occur traffic accidents causing injury severity.

Figure 5 indicates the probability density of the relation between segment length and others. Figure 5 shows that among the factors, segment length has a strong relation with operating speed and radius. Thus, regarding this figure, the highest probability density regions can be seen approximately in segment lengths between 2 to 4 km and speeds 80 to 100 km/hr. Further, the probability density of accident occurrence increases as the segment length increases and slope decreases.

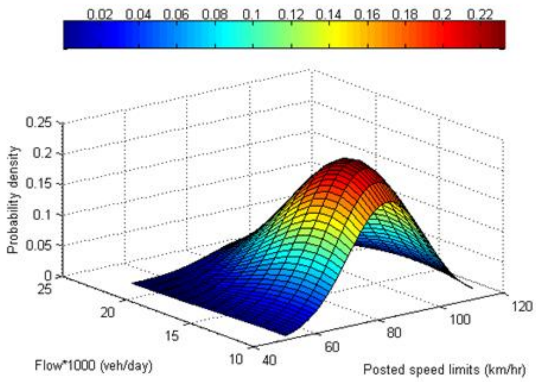
Figure 6 also shows the relation of probability density between AADT and other factors. According to this figure, the probability density of accident occurrence between AADT and operating speed has the maximum value compared with others. Thus, Figure 6 shows that the probability density of AADT and operating speed in the occurrence of accident increases. the maximum probability density is observed between 10000 and 12000 veh/day and speeds between 80 to 100 km/hr, as shown in Figure 6(b). However, AADT with slope and radius leading lower probability density values of accident occurrence. Figure 7 illustrates the relation between operating speed and slope and operating speed and radius. Based on Figure 7, it can be inferred that decreasing slope and increasing radius cause vehicles to increase their speed and thereby the probability density to increase accident occurrence. Thus, the maximum probability density for operating speed is observed between 80 and 100 km/hr, slopes less than zero, and radius between 400 to 700 m as shown in Figures 7(a), and 7(b). However, the probability density of accident occurrence between slope and radius in Figure 8 is low, and almost no significant relation is observed for the occurrence of accident by these factors with each other. Thus, based on Figures 4 to 8, it is shown that relation between posted speed limits with operating speed have considerable effects on the probability density to occur accidents causing injury. As Hashim (2006) and Yu et al., (2018) reported that operating speed is one of the most influencing factors in the occurrence of road accidents.

Further, based on the obtained results, the relation among segment length, operating speed and radius needs to be examined. As some studies showed that the frequency of accidents on higher radii and long segments increases because of the increasing operating speed of drivers (Shirmohammadi et al. 2018; Ferreira et al., 2015; Shirmohammadi et al., 2019; Sangare et al., 2021). Based on the obtained results, AADT with operating speed, and operating speed with slope and radius have considerable effects on the probability density of accidents occurrence as studies (Golob and Recker, 2003; Wang et al., 2009; Sadeghi et al. (2013) indicated that traffic factors such as AADT, speed, and geometric characteristics have significant effects on the probability occurrence of road accidents.

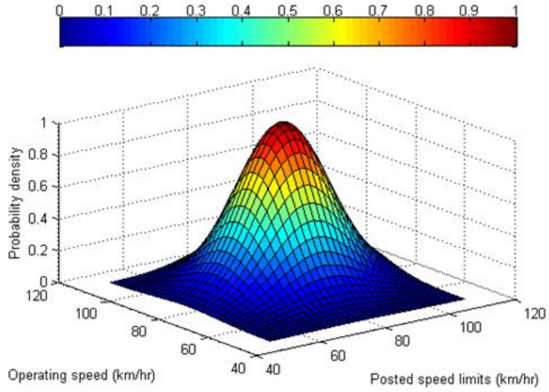
Therefore, based on the results obtained in the present study, drivers should follow speed limits and have an operating speed lower than 80 km/hr in free flow conditions while entering a radius of 400 to 700 m. Further, drivers should decrease their speed while entering slopes less than zero and long segments since the probability of accident occurrence increases due to increasing the maximum probability density. Thus, the relation between posted speed limits with operating speed, and the relation among segment length, operating speed and slope are essential on the occurrence of accidents.



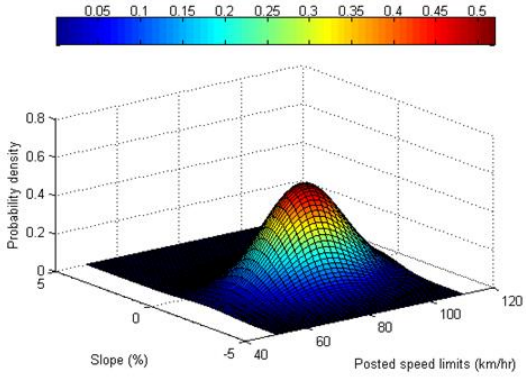
b



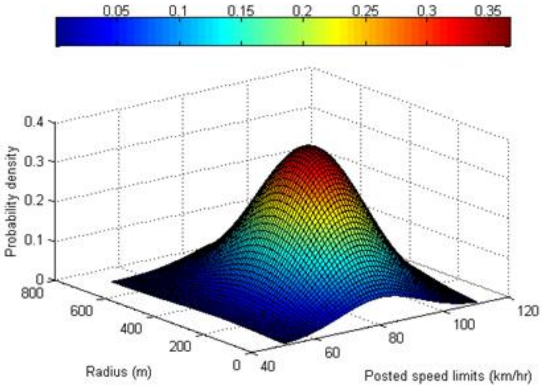
a



d

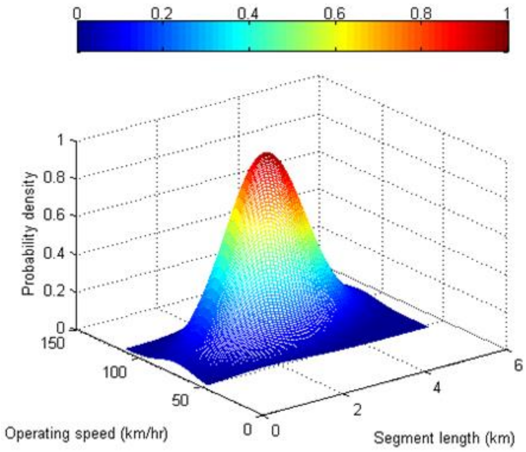


c

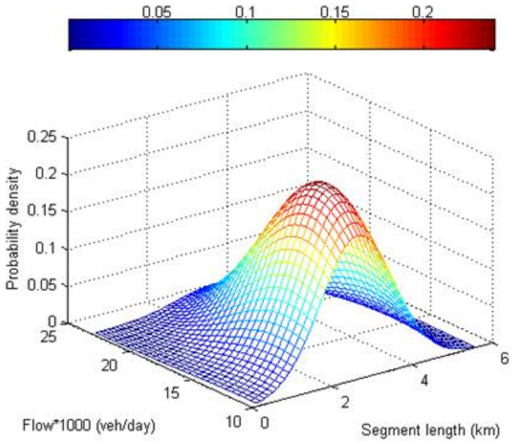


e

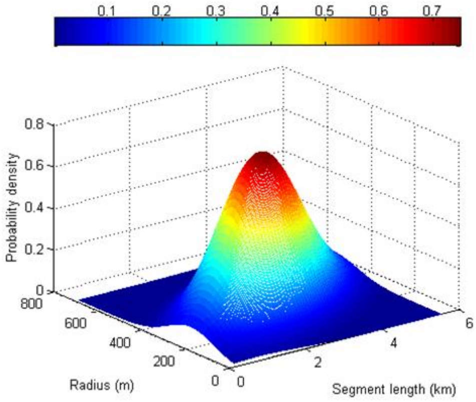
Figure 4. Relationship between posted speed limits and other factors on probability density



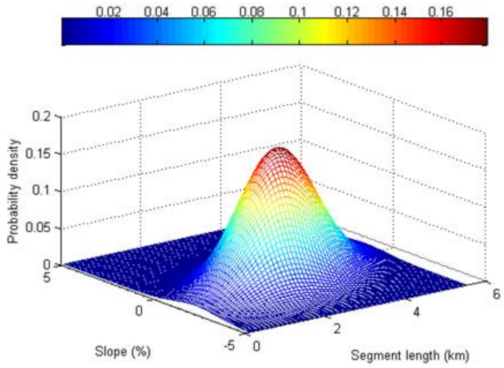
b



a

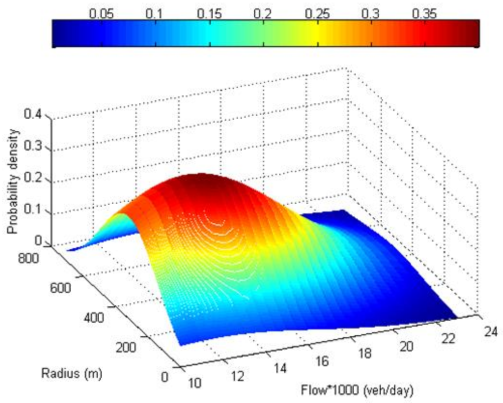


d

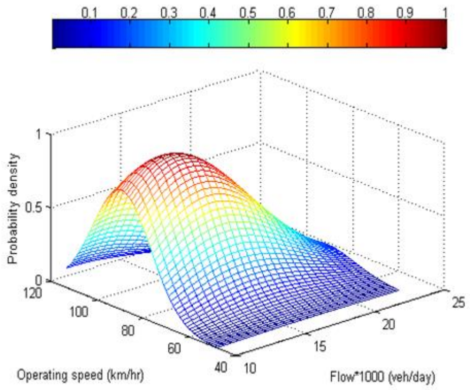


c

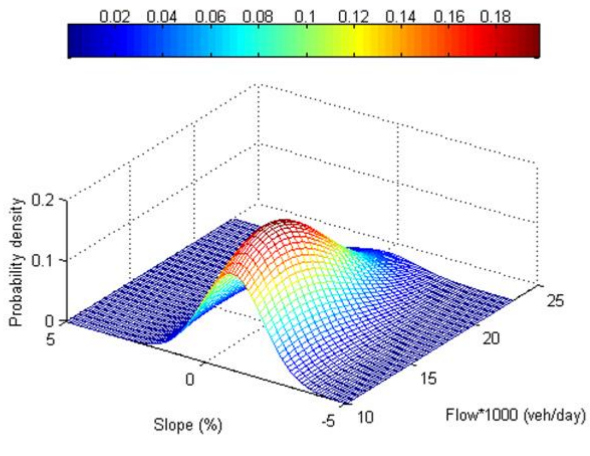
Figure 5. Relationship between segment length and other factors on probability density



b

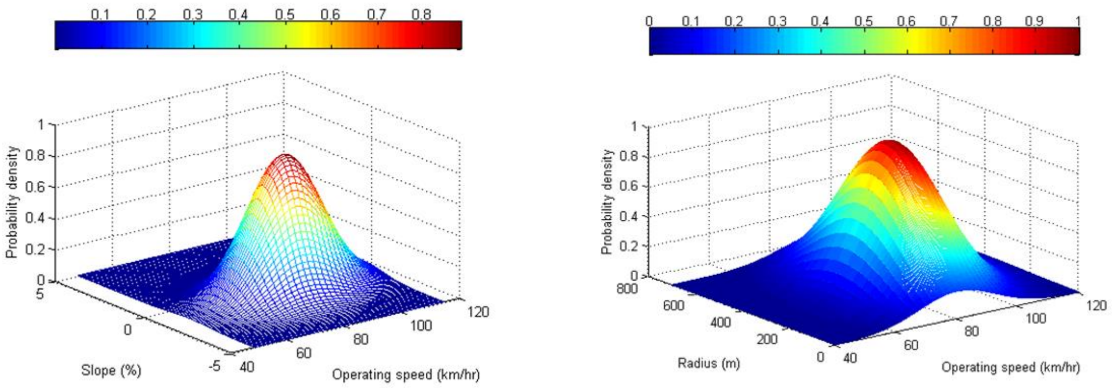


a



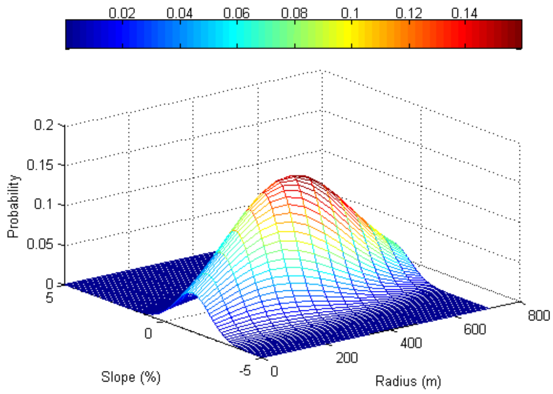
c

Figure 6. Relationship between AADT and other factors on probability density



b a

Figure 7. Relationship between operating speed and other factors on probability density



(a)

Figure 8. Relationship between slope and radius on probability density

5. Conclusion

Prioritization of factors contributing to accidents is essential to identify the most important factors in road safety. However, it is inappropriate due to a lack of reasonability and probability. Thus, looking up other methods in a way that probability helps safety researchers to reduce accidents seems to be warranted. This study first applies PCA as a statistical prioritization tool. Second, based on the Pearson correlation, the probabilistic method with a combination of probability theory is selected to estimate the probability density of the relation between factors causing injury. The accident datasets for injury severity on Borujerd -Khorramabad highway from 2015 to 2017 are evaluated for the proposed models. The obtained results are summarized as follows:

1- With the help of PCA as a statistical prioritization tool, we identified and ranked OS, PSL, AADT, SL, R, and S, respectively, as 6 critical factors affecting injury severity regarding the maximum correlation values.

2- The result of the Multivariate Gaussian model indicated that the relationship between posted speed limits with operating speed, and the relation among segment length, operating speed and radius are considerable to occur accident. Moreover, AADT with operating speed and operating speed with slope and radius significantly affect the probability of occurrence of accidents relevant to injury severity.

3- According to the obtained results in the present study, drivers should follow speed limits and have operating speeds lower than 80 km/hr in free flow conditions while entering a radius of 400 to 700 m. Further, drivers should decrease their speed while entering slopes less than zero and long segments since the probability of road accident occurrence increases due to increasing the maximum probability density of the relationship between posted speed limits with operating speed; the relation among segment length, operating speed and slope.

4- The results of the present research can help safety researchers and organizations concerning the factors contributing to road accident severity to examine the probability of factors influencing accidents to introduce countermeasures based on the maximum probability values.

In the present study, the accident datasets on various weather conditions, the days of the week, road surface conditions, and other possible variables could be applied to examine the most influential factors of accidents on different kinds of roads using the PCA method and multivariate Gaussian probability model. Future studies may use complete dataset regarding accident severity to investigate the relation between factors influencing accident occurrence.

6. References

– Afandizadeh, S., & Hassanpour, S. (2020). Evaluating the Effect of Roadway and Development Factors on the Rural Road Safety Risk Index. Advances in Civil Engineering, 2020.

– Ahrendt, P. (2005). The multivariate gaussian probability distribution. Technical University of Denmark, Tech. Rep.

– Arévalo-Támara, A., Orozco-Fontalvo, M., & Cantillo, V. (2020). Factors influencing crash frequency on colombian rural roads. Promet-Traffic &Transportation, 32(4), 449-460.

– Assi, K. (2020). Traffic crash severity prediction—A synergy by hybrid principal component analysis and machine learning models. International journal of environmental research and public health, 17(20), 7598.

– Bahadorimonfared, A., Soori, H., Mehrabi, Y., Delpisheh, A., Esmaili, A., Salehi, M., & Bakhtiyari, M. (2013). Trends of fatal road traffic injuries in Iran (2004–2011). PloS one, 8(5): e65198.

– Bird, R., & Hashim, I. (2006). Exploring Relationship between Safety and Consistency of Geometry and Speed on British Roads (No. 06-1509).

– Boodlal, L., Donnell, E.T, Porter, R.J., Garimella, D, Le. T., Croshaw, K., Himes, S., Kulis, P.N., & Wood, J. (2015). Factors Influencing Operating Speeds and Safety on Rural and Suburban Roads. Turner-Fairbank Highway Research Center.

– Boroujerdian, A.M., Saffarzadeh, M., Yousefi, H., & Ghassemian, H. (2014). A model to identify high crash road segments with the dynamic segmentation method. Accident Analysis & Prevention, 73, 274-87.

– Chen, P., Chen, Z., Bak-Jensen, B. (2008). Probabilistic load flow: A review. In2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (pp. 1586-1591). IEEE.

– Coll, B., Moutari, S., & Marshall, A. H. (2013). Hotspots identification and ranking for road safety improvement: An alternative approach. Accident Analysis & Prevention, 59, 604-617.

– Das, S., Geedipally, S. R., & Fitzpatrick, K. (2021). Inclusion of speed and weather measures in safety performance functions for rural roadways. IATSS research, 45(1), 60-69.

– De Moura, F. S., Garcia, L. F., Torres, T. B., de Oliveira, L. P., & Nodari, C. T. (2022). Identification and Prioritization Techniques Applied to Urban Accident-Prone Locations: A Scoping Review.

– Drosu, A., Cofaru, C., & Popescu, M. V. (2021). Fatal Injury Risk Model (FIRM) of the Road Accidents that Occurred in Rainy Conditions—A Probabilistic Approach. International journal of automotive technology, 22(5), 1415-1426.

– Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident analysis & prevention, 36(4), 637-647.

– Elyasi, M.R., Saffarzadeh, M., Boroujerdian, A.M., Semnarshad, M., Mazaheri, M. (2017). Prioritization of suburban accident factors based on analytical network process. International Journal of Transportation Engineering, 5(2), 197-209.

– Ferreira, S., & Couto, A. (2015). A probabilistic approach towards a crash risk assessment of urban segments. Transportation Research Part C: Emerging Technologies, 50, 97-105.

– Forouzanfar, M.H., Sepanlou, S.G., Shahraz, S., BESc, P.N., Pourmalek, F., Lozano, R., Asadi-Lari, M., Sayari, A.A., Dphi, C.J., Naghavi, M. Evaluating causes of death and morbidity in Iran, global burden of diseases, injuries, and risk factors study 2010. Archives of Iranian medicine, 17(5), 304.

– Fountas, G., Anastasopoulos, P. C., & Mannering, F. L. (2018). Analysis of vehicle accident-injury severities: A comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. Analytic methods in accident research, 18, 15-32.

– Fu, C., & Sayed, T. (2021). Multivariate Bayesian hierarchical Gaussian copula modeling of the non-stationary traffic conflict extremes for crash estimation. Analytic methods in accident research, 29, 100154.

– Golob, T.F., & Recker, W.W. (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions, Journal of transportation engineering.

– Gopaul, C. D., Singh-Gopaul, A., Sutherland, J. M., Rostant, L., Ebi, K. L., & Chadee, D. D. (2016). The epidemiology of fatal road traffic collisions in Trinidad and Tobago, West Indies (2000–2011). Global health action, 9(1), 32518.

– Haagsma, J. A., Graetz, N., Bolliger, I., Naghavi, M., Higashi, H., Mullany, E. C. & Phillips, M. R. (2016). The global burden of injury: incidence, mortality, disability-adjusted life years and time trends from the Global Burden of Disease study 2013. Injury prevention, 22(1), 3-18.

– Haghighi, N., Liu, X. C., Zhang, G., & Porter, R. J. (2018). Impact of roadway geometric features on crash severity on rural two-lane highways. Accident Analysis & Prevention, 111, 34-42.

– Hejazi, S., & Alipour, M. (2016). Prioritization of Reform of Road Accident Prone Sections Based On Road Safety Audit (Case Study Route between the City Of Khorramabad and Aleshtar in Lorestan Province). Journal of Transportation Research, 13(3), 65-85.

– Hosseinian, S. M., Najafi Moghaddam Gilani, V., Tahmasbi Amoli, H., Nikookar, M., & Orouei, A. (2021). Presentation of analytical methods for better decision making about the most important factor influencing rural accidents. Mathematical Problems in Engineering, 2021.

– Kassu, A., & Hasan, M. (2019). Identifying the principal factors influencing traffic safety on interstate highways. SN Applied Sciences, 1(12), 1-10.

– Khedher, M. B. B., & Yun, D. (2022). Generalized Linear Models to Identify the Impact of Road Geometric Design Features on Crash Frequency in Rural Roads. KSCE Journal of Civil Engineering, 26(3), 1388-1395.

– Kononov, J., Bailey, B., & Allery, B.K. (2008). Relationships between safety and both congestion and number of lanes on urban freeways. Transportation research record, 2083(1), 26-39.

– Koorey G. (2009). Road data aggregation and sectioning considerations for crash analysis. Transportation research record, 2103(1), 61-8.

– Ledesma, R.D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. Practical assessment, research & evaluation, 12(2).

– Liu, G., Chen, S., Zeng, Z., Cui, H., Fang. Y., Gu, D., Yin, Z., Wang, Z. (2018). Risk factors for extremely serious road accidents: Results from national Road Accident Statistical Annual Report of China. PLoS one, 13(8).

– Ma, M., Yan, X., Abdel-Aty, M., Huang, H., & Wang, X. (2010). Safety analysis of urban arterials under mixed-traffic patterns in Beijing. Transportation Research Record, 2193(1), 105-115.

– Mahdieh Rad, A.L., Ansari-Moghaddam, A., Mohammadi, M, Rashedi, F, Ghasemi, A. (2016). The pattern of road traffic crashes in South East Iran. Global journal of health science, 8(9), 149.

– Martins, M. A., & Garcez, T. V. (2021). A multidimensional and multi-period analysis of safety on roads. Accident Analysis & Prevention, 162, 106401.

– Mergia, W.Y., Eustace, D., Chimba, D., Qumsiyeh, M. (2013). Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. Accident Analysis & Prevention, 55, 202-210.

– Mohammadfam, I., Naserkhani, R. K., & Soltanian, A. (2016). The analysis of deaths caused by driving accidents in Ilam province, western Iran and the related factors by using the method of time series. International journal of occupational hygiene, 8(4), 200-207.

– Nawri, N. (2019). Berechnung von Kovarianzellipsen" (PDF).

– Office of Public Relations and International Affairs Legal Medicine Organization of Iran (LMO). (2017). Statistics of deaths and injuries caused by traffic accidents in 2017.

– Retallack, A.E., & Ostendorf, B. (2019). Current understanding of the effects of congestion on traffic accidents. International journal of environmental research and public health, 16(18), 3400.

– Ribeiro, M.I. (2004). Gaussian probability density functions: Properties and error characterization. Institute for Systems and Robotics, Lisboa, Portugal.

– Sadeghi, A., Ayati, E., & Neghab, M.P. (2013). Identification and prioritization of hazardous road locations by segmentation and data envelopment analysis approach. Promet-Traffic & Transportation, 25(2):127-36.

– Sadeghi-Bazargani, H., Ayubi, E., Azami-Aghdash, S., Abedi, L., Zemestani, A., Amanati, L., Moosazadeh, M., Syedi, N., Safiri, S. (2016). Epidemiological patterns of road traffic crashes during the last two decades in Iran: a review of the literature from 1996 to 2014. Archives of trauma research, 5(3).

– Saha, S., Schramm, P., Nolan, A., & Hess, J. (2016). Adverse weather conditions and fatal motor vehicle crashes in the United States, 1994-2012. Environmental health, 15(1), 1-9.

– Sangare, M., Gupta, S., Bouzefrane, S., Banerjee, S., & Muhlethaler, P. (2021). Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. Expert Systems with Applications, 167, 113855.

– Shen, L., Lu, J., Long, M., & Chen, T. (2019). Identification of accident blackspots on rural roads using grid clustering and principal component clustering. Mathematical Problems in Engineering, 2019.

– Shirmohammadi H, Najib AS, Hadadi F. (2018). Identification of Road Critical Segments Using Wavelet Theory and Multi-Criteria Decision-Making Method. EUROPEAN TRANSPORT-TRASPORTI EUROPEI, 1(68).

– Shirmohammadi, H., Hadadi, F., & Saeedian, M. (2019). Clustering analysis of drivers based on behavioral characteristics regarding road safety. International Journal of Civil Engineering, 17(8), 1327-1340.

– Prince, S. J. (2012). Computer vision: models, learning, and inference. Cambridge University Press.

– Stroock DW. (2010). Probability theory: an analytic view. Cambridge university press.

– Tang, H., Gayah, V. V., & Donnell, E. T. (2019). Evaluating the predictive power of an SPF for two-lane rural roads with random parameters on out-of-sample observations. Accident Analysis & Prevention, 132, 105275.

– Timmerman ME. (2010). Principal component analysis. 2nd Edition. Berlin: Springer.

– Ullah, H., Farooq, A., & Shah, A. A. (2021). An empirical assessment of factors influencing injury severities of motor vehicle crashes on national highways of Pakistan. Journal of advanced transportation, 2021.

– Wang, C., Quddus, M.A., Ison, S.G. (2009). Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. Accident Analysis & Prevention, 41(4), 798-808.

– Wang, X., Song, Y., Yu, R., Schultz, G.G. (2014). Safety modeling of suburban arterials in shanghai, china. Accident Analysis & Prevention, 70, 215-224.

– World Health Organization (WHO). (2015). Global status report on road safety 2015.

– World Health Organization. (2019). Global status report on alcohol and health 2018. World Health Organization.

– Wu, P., Song, L., & Meng, X. (2021). Influence of built environment and roadway characteristics on the frequency of vehicle crashes caused by driver inattention: a comparison between rural roads and urban roads. Journal of safety research, 79, 199-210.

– Xie, G., Shangguan, A., Fei, R., Hei, X., Ji, W., & Qian, F. (2020). Unmanned System Safety Decision-Making Support: Analysis and Assessment of Road Traffic Accidents. IEEE/ASME Transactions on Mechatronics, 26(2), 633-644.

– Yi, T., Zhu, L., Peng, W.L., He, X.C., Chen, H.L., Li, J., Yu, T., Liang, Z.T., Zhao, Z.Z., & Chen, H.B. (2015). Comparison of ten major constituents in seven types of processed tea using HPLC-DAD-MS followed by principal component and hierarchical cluster analysis. LWT-Food Science and Technology, 62(1), 194-201.

– Youming, T., Deliang, Z., Xinyu, Z., & Na, L. (2018). Principal Component Analysis of Fatal Traffic Accidents Based on Vehicle Condition Factors. In2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA) 2018 Sep 22 (pp. 315-317). IEEE.

– Yu, R., Quddus, M., Wang, X., Yang, K. (2018). Impact of data aggregation approaches on the relationships between operating speed and traffic safety. Accident Analysis & Prevention, 120, 304-310.

– Zhang, H., Hou, N., Zhang, J., Li, X., & Huang, Y. (2021). Evaluating the safety impact of connected and autonomous vehicles with lane management on freeway crash hotspots using the surrogate safety assessment model. Journal of advanced transportation, 2021.