Factors propelling fatalities during road crashes: A detailed investigation and modelling of historical crash data with field studies

Malaya Mohanty a,*, Rachita Pandaa, Srinivasa Rao Gandupallib, Ritik Raj Aryaa, Sarthak Kumar Lenka a

Keywords: Road safety Accidents Severity index CART Logistic regression Clustering Crash fatality models

ABSTRACT

One of the major concerns in developing countries like India is to maintain traffic safety under mixed and heterogenous scenario. Although zero accidents is the need of the hour, the first step to attain it is ensuring zero deaths and no serious long-term disabling injuries in road crashes. To reduce the road crash fatalities, explicit and detailed studies have been conducted by utilising historical road crash data of two emerging smart cities of India - Bhubaneswar and Visakhapatnam. Traffic flow data and characteristics of road infrastructure has also been collected by performing field studies at accident prone locations. Various factors including vehicular characteristics, road user characteristics, and road infrastructure have been analyzed using various non-parametric tests to identify the contributing factors resulting in fatalities. It is observed that out of 14 variables used for study, 8 factors were significantly related to fatal crashes. These included categories of victim and accused, 85th percentile speed, presence of road markings, availability of sight distance, etc. The significant factors were subjected to binary logistic regression to determine the odd’s ratio of significant factors. The logistic regression predicted 79% of deaths correctly. Crash fatality prediction models are developed using both Classification and Regression Tree (CART) classification tree with 83% accuracy. Although CART classification led to higher accuracy, binary logistic regression is more robust as it considered more significant factors as compared to CART. Subsequently, a severity index has been proposed based on proportions of actual fatal crashes and usage of K-means clustering technique. The proposed indices shall be really helpful in traffic safety management, specifically in reduction of fatalities during road crashes.

1. Introduction

In a developing nation like India, Ministry of Road Transport and Highways (MORTH) has consistently reported around 150000 deaths per year due to road crashes since 2015 (MORTH, 2019). Similarly, World Health Organisation (WHO) has reported that more than 1.3 million deaths occur due to road crashes every year around the world. Further, 93% of fatalities due to road crashes occur in low-and middle-income countries even though they comprise only 60% of the total vehicles in the world (WHO, 2021). The high number of fatalities despite various steps taken to prevent them calls for more research. Secondly, with most of the studies depending on the road characteristics, its time the effect of driver behaviour and vehicle dynamics should also be studied (Mohanty et al., 2022). Secondly, in a country like India where we observe 5 rows of traffic movement on 3 lanes with various classes of vehicle movement (Mohapatra et al., 2016), the present study is even more important. Further, with various category of vehicles driving on the same road (mixed traffic condition), the chances and severity of road crashes greatly vary. More than road crashes, the fatalities in road crashes are very high in developing countries. In India, a total of around 4,50,000 accidents per year lead to 1,50,000 deaths (MORTH Accident data, 2019). The death to accident ratio is nearly 1:3. Whereas in a developed country like USA, the death to accident ratio is around 1:170 only (National Highway Traffic Safety Association, NHTSA). Although according to Vedagiri and Pragna (2013), have opined about under reporting of accidents in India, still the ratio of deaths to accidents look far-fetched. Therefore, it is required to study the road crashes, specifically the fatal crashes and assess the important factors that pose risk to occurrence of fatality in road crashes. The risk factors, if identified properly can provide more information about the

* Corresponding author.
E-mail address: malaya.mohantyfee@kiit.ac.in (M. Mohanty).
https://doi.org/10.1016/j.heliyon.2022.e11531
Received 7 March 2022; Received in revised form 24 June 2022; Accepted 3 November 2022
2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
nature of road accidents which may help to suggest better strategies and solutions in long run.

Many researchers (Mohanty and Gupta, 2016; Goel and Sachdeva, 2016; Peesapati et al., 2015; Laureshyn, 2006; Vedagiri and Pragna, 2013; Ashraf et al., 2019; Caliendo and Guida, 2012; Verster and Fourie, 2018; Aney and Ho, 2019; Ansari et al., 2000; Pirdavani et al., 2010; Mohanty et al., 2021-a, 2021-b) have tried to model the road crashes and have made attempts to study various solutions to increase road safety. Analysis of reported crash data is usually used for accident predictive and analysis modelling (Caliendo and Guida, 2012). Researchers use the historical crash data for a location and utilize various statistical analysis to predict the road crashes in future (Goel and Sachdeva, 2016; Kar et al., 2016; Edwards, 1996; Casado-Sanz et al., 2019; Ansari et al., 2000) However, over the years, numerous issues with the accident data have been reported by various researchers (Caliendo and Guida, 2012). The problem with road crash data is its availability as accidents are rare events. Secondly, all accidents are usually not reported unless it leads to serious property damage or loss of lives (Vedagiri and Pragna, 2013). However, it’s the fatal crashes that have been of keen interest to researchers in developing countries. Instead of reducing the overall number of accidents, which no doubt is also a prime parameter for increasing road safety, more focus is on reducing the number of fatal crashes on roads, or the accidents leading to death/s. Therefore, the causes behind fatal crashes should be examined and investigated properly. Sayed et al. (2007) had categorized the severity of crashes in terms of deaths, injury, and property damage only. They tried to analyze the effect of day and night on road crashes. Chen et al. (2007) analysed the impact of operating speeds on road crashes in China based on reported crashes. Further, few researchers (Mohammed, 2012; Divakaran and Sreelatha, 2013) have tried to utilize various factors for accidents by utilizing road crash data, but at specific locations like intersections. The usual techniques that are being used by most of the researchers to model road crashes or road crash deaths are Logit models, Probit Models, Bayesian networks, ANN, CART classification, etc (Mohapatra et al., 2021; Lee and Li, 2015;
Abdel-Aty, and Abdelwahab, 2004; De Oña et al., 2011). Unlike the Artificial Neural Network, regression trees in the CART model make it easier to interpret the results (Pande et al., 2010). The CART trees also remove the problems of multicollinearity (Chang and Wang, 2006). Chang and Wang (2006) examined the injury severity of crashes in Taiwan by using the CART model and found that vehicle category was strongly related to crash injury severity. Similarly, Montella et al. (2012) observed from the CART regression results that road type was significantly associated with severity of crashes in Italy. While proposing an accident severity index, mostly researchers (Soltani, and Askari, 2017; Iyanda, 2019; Tola et al., 2021) have used historical crash data only with few particular factors with clustering technique. However, such studies are less and not much study has been conducted regarding proposal of a severity index based on combination of historical crash data with field studies. Such an index would be really helpful to manage blackspots and accident-prone locations. Furthermore, road crash data haven’t been combined with the results from field studies much in past literature. This would be really helpful since the speeds of vehicles and road geometry are not provided in detail in police records. The combination of road crash data and field studies will help a more complete understanding of road accident-related deaths.

The present study deals with the investigation of all contributing factors for fatal road crashes based on the reported historical crash data. Along with historical crash data, field studies are conducted at accident locations to gather speed data. Thereafter, the significant factors are studied explicitly, and a severity prediction model has been developed by using CART regression tree. Finally, a severity index has been proposed by performing clustering analysis. The development of severity index can help understand the road crash fatalities better and help in framing better strategies for reducing the death rate during the road crashes.

2. Data collection and methodology

Data collection is an important aspect for the present study since better prediction models can be developed with a greater number of data points. In order to pursue the study, historical road crash data has been collected from 2 cities in India i.e., Bhubaneswar and Visakhapatnam from 2014 to 2017. These cities are emerging smart cities of the country nominated by Ministry of Urban Development, Government of India (2015). Both the cities are located near the east coast of India as shown in Figure 1 below. While Bhubaneswar is the capital city of Odisha, Visakhapatnam is one of the prominent coastal cities located in Andhra Pradesh. Both cities have similar demographics and geographical positioning. Both the cities have been categorized as tier – II city based on their population that lies between 0.5 to 5 million. In India maximum cities are in tier – II for their similar development in terms of population and many tier – III cities whose population is less than 0.5 million are subjected to major and swift developments and will soon join the tier – II within a short span (Aithal and Ramachandra, 2016). Since most of the Indian cities fall within this population demography of tier-II cities, therefore, in the present study, data from 2 such cities have been utilized to develop crash severity prediction models.

The road crash data were obtained from the local police stations of the studied locations. The road crash data has the following information about every accident.

1. Year, month, date, day, time and location of road crash.
2. Fatal/Non-fatal accident
3. Category of accused and victim vehicle
4. Whether the accident happened near intersections or other kind of traffic facilities

In order to pursue the objectives of the present study, the month of the accidents were used to identify the season in which a specific accident had occurred. Similarly, the location of the accident helped in collecting the speed data from the accident spot along with other geometrical observations like presence of service roads, and land use. The day and time of the accident helped in determining whether accidents happened in weekdays or weekends along with day or night and peak or off-peak time. The following factors have been used for testing their importance towards death in road crashes.

- Season in which accident had occurred.
- Type of day (weekday/weekend)
- Time of crash based on peak and off-peak hour
- Type of day (day/night and peak/off-peak periods)
- Category of accused vehicle (2-wheeler/3-wheelers/Cars and jeeps/Heavy vehicles)
- Category of victim (2-wheeler/3-wheelers/Cars and jeeps/Heavy vehicles/pedestrian/self-hit)
- Age of the people affected (died/injured)
- Type of road/intersection where accident happened
- 85th percentile speed (Safe speed limit)
- Availability of service roads
- Type of land use besides the location
- Availability of sight clearance
- Presence of proper traffic signs
- Presence of road markings

The flow chart provided in Figure 2 presents the methodology steps for the present study.

Modelling the road crash fatality is a complex procedure since the involvement of same category of vehicles may result in deaths and injuries at 2 different instances. Secondly, other factors like speed, type of traffic facility also affects the probability of fatality in a road crash. The variables considered in the present study are mostly ordinal/nominal in nature. For instance, vehicle type are categorical variables, and type of traffic facility/intersection is a nominal variable with yes and no as answers. Therefore, Phi & Cramer’s correlation test is first conducted among the nominal and ordinal factors to examine the presence of any statistical association/relationship between them.

Phi & Cramer’s value establishes the relationship between nominal variables by utilizing cross-tabulation. The value is predicted utilizing the random category assignment. It determines the improvement in predictability accuracy percentage of the dependent variable (here the result of road accident) when the value of assumed independent variables affecting the dependent variable are being used.

After performing the initial correlation test, two models i.e., binary logistic regression, and CART have been utilized in the present study to predict the probability of fatal crashes and fatality based on other road conditions.
and road user factors. Logistic/Logit models are probabilistic models that attempt to Many studies (Zhao et al., 2022; Zhang et al., 2021; Mohanty and Gupta, 2016; Tamakloe et al., 2022) have successfully utilized logit models to develop road safety related prediction models and assess the effect of each independent causative variable. The general equation of a binary logistic regression is provided in Eq. (1).

\[ P(\text{Event}) = \frac{e^{Y}}{1 + e^{Y}} \]  \hspace{1cm} (1)

Where.

\[ P(\text{Event}) = \text{Probability of occurrence of any event} \]

\[ Y = \text{General regression expression/equation} \]

CART is a regression tree technique which has also been utilized to predict road crash fatality with higher accuracy. Few researchers (Lee and Li, 2015; Abdel-Aty, and Abdelwahab, 2004) have also utilized decision trees like CART, and Random Forest Trees to model the safety of vehicles and pedestrians on road. Hu and Cai (2022) opined that logistic regression equation is primarily applied for establishing the relationship, and afterwards in case of successful model formation in logistic regression, the machine learning techniques like CART are applied for prediction modeling. Tamakloe et al. (2022) concluded that logistic regression identifies the individual effect of factors affecting severity of crashes. According to Pandel et al. (2010), decision trees like CART make it easier to interpret the results unlike ANN. The CART technique also remove the problems of multicollinearity. Therefore, in the present study, the logistic regression is first conducted and then the CART is employed for development of predictive models.

Finally, cluster analysis has been used for proposing a severity index. Cluster analysis is a common mathematical tool to classify the data into various ranges/groups. The objective of this analysis is to perform a partition where objects within a cluster form a group of similar structure and it is widely utilized in various fields of science and technology (Jain and Dubes, 1988; Sangngam et al., 2022).

The points in any cluster group are nearer to the center of the same group to which it has been assigned, as opposed to the center of other clusters. The clustering technique is said to be working when the points/data points have low intra-cluster distance and high inter-cluster distance (Jain and Dubes, 1988). The commonly used clustering techniques in the field of traffic and transportation engineering are k-means, and hierarchical clustering (Mohapatra et al., 2015). The present study uses K-means clustering for developing the crash severity indices which has been performed in the IBM SPSS software.

K-mean clustering is an unsupervised method for explaining the classification problems where the groups are classified based on their euclidean distances (Mohapatra et al., 2015). Precisely, the method attempts to sectorise the data in such a way that minimum variation within a cluster is observed. In K-means, the mechanism of clustering begins by randomly assigning data points to number of pre-assigned cluster groups. The data points are then iteratively reassigned to the cluster which has the least euclidian squared distance to its center. According to Mooi and Sarstedt (2014), when the data set exceeds 500, k-means clustering process provides better partitioning results. Many researchers (Mohapatra et al., 2015; Boora et al., 2017; Mohanty and Dey, 2019; Mohanty et al., 2021; Boora et al., 2017) have successfully utilised this technique for studying various aspects of traffic flow modelling like safety analysis, LOS determination, etc. The mechanism of clustering usually continues until a predetermined number of iterations are reached, also termed as convergence (Mooi and Sarstedt, 2014). Convergence is an important key to the K-means clustering technique. Convergence of cluster groups mean that there is no more change in the cluster memberships. This convergence is achieved by the help of several iterations which is usually performed by the help of SPSS software. Usually, Lloyd’s algorithm is used in K-means clustering to reach at convergence while finding out the center of clusters. A popular empirical technique for k-means clustering is Lloyd’s algorithm (Jain, 2010). The algorithm used in SPSS for convergence can also be written simply as follows.

**Input**

- K: number of desired clusters.
- D: \(\{d_1, d_2, ..., d_n\}\) a data set containing n objects.

**Output**

- A set of k clusters as specified in input.

**Method**

1. k data items are arbitrarily chosen from D dataset as initial cluster centres.
2. Each data item \(d_i\) is assigned to the cluster to which object is most similar based on the mean value of the object in cluster.
3. New mean value of the data items for each cluster is calculated and the mean value is updated.
4. The process is repeated until no change in mean value is observed in the next iteration.

Using this technique SPSS determines the lower and higher ranges of each clusters representing the severity index. Thus, in the present study, all the above-mentioned statistical techniques have been explored and used.

3. Results and analysis

More than 3000 accident data were collected for the study. Thereafter, they were checked for the completeness of documented information. 2425 number of accidents were found to be complete in all aspect and have been used for the present study. The accident locations were identified, and speed studies were conducted to determine the 85th percentile speed for the locations where accidents had occurred. Similarly, the locations were investigated physically as whether required sight distance is available to the road users, and whether any service roads are present besides the road where accidents had occurred. Based on the month of the accident occurrence, they were classified seasonally, i.e., November to February – Winter, March to June – Summer, and July to October – Rainy. Accidents occurring within morning 5 AM to evening 6 PM are denoted as daytime, and the rest as nighttime. Similarly, the accidents occurring in between 8 AM to 11 AM and 4 PM–7 PM are regarded as accidents during peak hours, whereas for accidents happening during other times are classified as off-peak hour accidents. This was classified after conducting traffic volume surveys which showed these times to have more vehicles on road as compared to others. Figure 3 demonstrated the general trend of traffic volume for an accident location for 7 days earmarking the peak and off-peak hours. Based on the graphs obtained from field studies, 8 am–11 am and 4 pm–7 pm have been classified as peak hours.

Thereafter, the composition and standard deviation of each variable with respect to road crash fatalities is ascertained to assess whether the variables can be actually contributing factors for road crash deaths. It was observed that out of 2425 reported crashes of road accidents, 550 accidents resulted in fatalities which amounts to 22.7%. In other words, the probability of a reported road crash resulting in death is around 0.23. The contributing factors and their importance in a road crash fatality needs to be assessed explicitly.

3.1. Effect of various factors on road crash fatalities using descriptive statistics and binary logistic regression

To establish any relationship between age groups and road crash fatalities, a histogram has been prepared (as shown in Figure 4). It can be observed that all age groups (from 15 to 60 years) contribute nearly...
Figure 8. Effect of time of accident occurrence on road crash fatalities.

Figure 9. Effect of category of accused on road crash fatalities.

Figure 10. Effect of category of victim on road crash fatalities.

Figure 11. Effect of occurrence of accident at/near intersection on road crash fatalities.

Figure 12. Effect of presence of service road on road crash fatalities.

Figure 13. Effect of land use on road crash fatalities.
equally to fatalities due to road crashes. The average age of deceased victims was determined to be 41 years with an SD of 17 years which confirms the almost equal involvement of all age groups towards deaths on roads. Therefore, a single age group can’t be considered as a major factor for road crash fatalities.

Next, the factors (as mentioned in methodology) are analysed one by one by conducting cross tabulations in SPSS software and checking their significance value of Phi-Cramer’s statistics which is usually presented for nominal data. Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18 showcase the variation of various factors with respect to road accident injuries and deaths. Table 1 represents the significance value of Phi-Cramer’s statistics. Factors having significance values of <0.10 (10% significance level) are considered as significant variables contributing to road crash fatalities and are considered for further modelling. The brackets besides a factor contain the components into which the factor is divided for analyses.

Table 1 clearly reveals that out of the total 14 factors considered at start for the study, 8 factors have significant effect on the outcome of road crash in terms of injury/fatality. The same can also be detected from Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18. It can be observed that for the factors like day of accident occurrence, season
of occurrence, traffic type (peak or off-peak), day or night, etc. which don’t have significant effect on road crash outcomes have similar percentage of fatal road crashes with respect to total number of road crashes. For example, in case of seasons it is observed that out of total 2425 accidents, 727 accidents have occurred in summer season, while 829 and 869 number of accidents have occurred in rainy and winter respectively which are statistically similar to each other. Further, the percentage of fatal crashes in every season lies in the range of 21–23% which is not significantly different. Also, the overall fatality rate for all reported road crashes is also 22.7%. No season exhibited significantly more or less percentage of fatal crashes which is in principle with the findings of Phi and Cramer’s statistics. Similarly, if we consider the type of traffic in terms of peak and off-peak hours, it can be observed that since peak hours are only for 6 h in a day and rest are off-peak hours, therefore a greater number of accidents (1644) have occurred during off-peak hours as compared to peak hours (781). However, the percentage of fatal crashes are almost the same during both peak and off-peak hours at around 22–23% due to which the factor is not deemed significant with respect to road crash fatalities. However, if we consider a factor which has come significant in Phi & Cramer’s statistics like category of accused vehicle, it can be observed that for various accused vehicle category, the death percentage largely vary. Two wheelers (2W) are accused vehicles in 610 road accident incidents, in which 160 incidents have resulted in fatalities which is 26.23%. Similarly, the percentage of fatal crashes for three wheelers (3W), cars and jeeps, and heavy vehicles (HV) are 19.4%, 22.5%, and 21.3% respectively. Since the percentage of death vary from 19% to 26% for various accused category of vehicles, therefore the factor is deemed to be significant with respect to crash type (injury/fatality).

Table 2 below presents the percentage of fatal road crashes for factors that have been found to be significant in affecting the outcome of a road crash (injury/fatal).

Some salient points that are observed from Tables 1 and 2 are discussed below.

- Crash due to 2W are associated with higher fatal crashes (26%) in comparison to other vehicle categories. However, crashes due to HVs and Car and Jeeps are also leading to fatalities for 22–23% of incidents which is quite similar with 2W. Since, the number of crashes due to HVs are too high, the percentage alone cannot determine the actual fatality rate in case of accidents due to an individual vehicle.
- In case of victim category, vehicles who self-hit and pedestrians have the highest and almost equal share of road crash fatalities with more than 35% chance of crashes resulting in deaths.
- It was observed that crashes at intersections and not at intersections both have similar contribution to road crash fatalities within 20–23%. However, the factor has come significant since the number of crashes at intersections (652) are much less than the crashes at midblock sections and other traffic facilities (1773).
- On roads with no land use besides them there tends to be higher risk of fatality (27.4%) during road crashes as compared to the roads with mixed land use (21.1%). The reason can be attributed to the fact that due to absence of any structures on both sides of roads, vehicles tend to be travel at higher speed as compared to the locations where some land use in terms of shops/offices, etc. are present.
- A unique phenomenon was observed that the presence of earthen shoulder increases the chances of fatal road crashes (25.6%) when compared with accidents on roads with no shoulders. One of the prime reasons behind this is the carefree nature of road users after knowing that the road has a shoulder. Due to presence of shoulder, drivers tend to travel with higher speed due to a psychological thinking that the road is wider. Secondly, vehicles also tend to crash with the mountable kerbs during night, morning or foggy weathers if they suddenly appear besides carriageway.
- Availability of sight distance plays a big role in fatal road crashes since the fatality decreases to 18% where required sight distance is provided, whereas unavailability of sight distances lead to 24% chance of fatality.
Table 3. Contingency table for Hosmer and Lemeshow test.

| Type of injury – Injuy | Type of injury – Death | Total |
|------------------------|------------------------|-------|
| Observed   | Expected | Observed   | Expected | Observed   | Expected |
| 1          | 235      | 237.229   | 8        | 5.771     | 243    |
| 2          | 231      | 229.852   | 9        | 10.148    | 240    |
| 3          | 211      | 205.364   | 24       | 29.636    | 235    |
| 4          | 209      | 203.243   | 34       | 39.757    | 243    |
| 5          | 205      | 194.118   | 41       | 51.882    | 246    |
| 6          | 173      | 184.157   | 73       | 61.843    | 246    |
| 7          | 170      | 176.468   | 78       | 71.532    | 248    |
| 8          | 153      | 164.816   | 90       | 78.184    | 243    |
| 9          | 147      | 153.477   | 97       | 90.523    | 244    |
| 10         | 141      | 126.276   | 96       | 110.724   | 237    |

Table 4. Classification Table for Binary Logistic Regression model.

| Observed   | Predicted |
|------------|-----------|
| Type of injury | Percentage Correct |
| Injury      | Death     |     |
| 1092       | 783       | 58.2 |
| 116        | 434       | 78.9 |
| Overall    |           | 62.9 |

Table 5. Odd’s Ratio for Binary Logistic Regression model for road crash severity.

| Sub component variables | B | Exp(B) or Odd’s ratio |
|-------------------------|---|-----------------------|
| Accused Vehicle – Heavy Vehicle (Base variable) | - | - |
| Accused Vehicle – 2-Wheeler | -.642 | .526 |
| Accused Vehicle – 3-Wheeler | -.809 | .445 |
| Accused Vehicle – Cars and Jeeps | -.484 | .617 |
| Victim Vehicle – Heavy Vehicle (Base variable) | - | - |
| Victim Vehicle – Self hit | 2.285 | 9.825 |
| Victim Vehicle – Pedestrian | 2.202 | 9.041 |
| Victim Vehicle – 2-Wheeler | 1.310 | 3.706 |
| Victim Vehicle – 3-Wheeler | 1.311 | 3.708 |
| Victim Vehicle – Cars and Jeeps | -.994 | .370 |
| Intersection Crash – Yes (Base variable) | - | - |
| Intersection Crash – No | -.008 | 1.008 |
| Land use – Mixed land use (Base variable) | - | - |
| Land use – No land use | .019 | 1.019 |
| Earthen shoulder – Yes (Base variable) | - | - |
| Earthen shoulder – No | -.131 | .877 |
| Sight clearance of the drivers – Available (Base variable) | - | - |
| Sight clearance of the drivers – Not available | .150 | 1.162 |
| Adequate Road Markings – Present (Base variable) | - | - |
| Adequate Road Markings – Not present | .138 | 1.148 |
| Speed – 60-80 kmph (Base variable) | - | - |
| Speed – 20-40 kmph | -.607 | .545 |
| Speed – 40-60 kmph | -.245 | .782 |
| Constant | -2.218 | .109 |

- The absence of road markings at required locations lead the drivers to make uninformed decisions which results in fatalities in case of 27.9% of total crashes.
- Rise in 85th percentile speed on roads also rises the chances of fatality. The percentage of road crash resulting in fatality increases from 16.7% to almost 30% when the average 85th percentile speed of the vehicles increases from 20 to 80 kmph.

Before proceeding for development of a crash severity prediction model using CART, it is important to identify the factors, specifically the subcomponents of factors that lead to higher severity rates. Although Table 2 depicts a clear picture of the effect of each variable under the significant factors, traditionally logistic regressions/logit models are being used to determine the odd’s ratio of different significant factors (Yan et al., 2005; Mohanty and Gupta, 2016; Mokhtarimousavi, 2019; Rahman et al., 2021). In the present study, the obtained significant factors have been subjected to a binary logistic regression to know the odd’s ratio of each subcomponent variable. Binary logistic regression is used since the outcome is either 1 or 0 (Fatal or non-fatal). In order to confirm the goodness of fit of the logistic regression, Hosmer-Lemeshow test statistics is observed. Table 3 provides the contingency table for Hosmer-Lemeshow test. It can be observed that the obtained and expected values are very close to each other. Further, since the average percentage of fatal crashes is 23%, therefore it would be inappropriate to keep 0.5 as the cut off for representing deaths. Any value above 0.23 would mean higher risk of fatality as compared to total average value of fatal crashes. Therefore, after changing the cut value to 0.23, the classification table for the binary logistic regression model as shown in Table 4 confirms that, for 63% of the data, the model predicts the type of injury correctly. Further, the model predicts fatal crashes correctly for 79% of data. This is a better fit since it is better to predict an injury as fatality rather than predicting a fatality as injury. Table 5 shows the coefficients and Odd’s ratio for each significant sub component variable.

The salient observations from Table 5 are detailed below.

- In case of accused vehicle category, as compared to crashes due to HVs, it is observed that the crashes due to 2W shall have 0.5 or 50% less chance of fatality. Similarly, 3W lead to fatal crashes which is 0.4 times the rate of fatal crashes by HVs. Cars and jeeps also lead to a smaller number of fatal crashes which is 0.6 (or 40% less) times of fatal crashes due to HVs.
- However, in case of victim, with respect to HVs, pedestrian and self-hit victims have 9 to 9.8 times more chance of death due to road crashes. Similarly, 2W and 3W show similar odd’s ratio as crash with 2W or 3W leads to 3.7 times more chances of fatality as compared to HVs. Cars and Jeeps seem to be the safest on road as they exhibit 0.3 times (70% less) fatality as compared to crash with HVs.
- Crashes at/not at intersection doesn’t affect the severity of crash much as exhibited from Phi and Cramer’s statistics previously. Severity increases by only 1.008 times when the crash happens on normal roads when compared to intersections. Those are quite similar and might not prove useful while developing an accident severity prediction model.
- Similarly, no land use increases the chance of fatality by only 1.019 times as compared to mixed land use. This increase is not very high and severity prediction model might not be affected by it.
- However, absence of earthen shoulders besides the accident location decreases the rate of fatality by 0.877 times as compared to the presence of shoulder. This unique finding was also observed in general descriptive statistics. This is mostly because of the carefree nature of drivers on seeing a wider road without able to distinguish between road and shoulder technically.
- The absence of adequate road markings and unavailability of sight distance leads to higher fatalities which is 1.15 times as compared to roads with proper sight distance and road markings.
- 85th percentile speed on road reaches also affects the crash severity. Road stretches with speeds around 40–60 kmph and 20–40 kmph have fatalities rate 0.78 times and 0.54 times respectively when compared with road stretches with speeds 60–80 kmph. Thus, higher, the speed much more chances of fatal crashes.
3.2. Crash severity prediction model using CART

Binary logistic regression revealed the important factors contributing significantly to the road crash fatalities. After careful analysis of the odd’s ratio results, the following factors have been considered for development of crash severity prediction model.

- Category of accused vehicles
- Category of victim
- 85th percentile speed range
- Presence/absence of earthen shoulders
- Presence/absence of adequate road markings
- Availability/non-availability of sight distance.

Crash at intersection, and land use are the factors which have been omitted from the prediction model since their presence or absence doesn’t change the outcome of a crash much as observed from Table 5 from their values of odd’s ratio. In the present study, regression tree

**Figure 19.** Relative importance of variables for CART prediction model.

**Figure 20.** Optimal tree diagram.
(CART) is used to develop the crash severity model. As mentioned earlier in the Introduction section, regression trees in the CART model make it easier to interpret the results as compared to the Artificial Neural Network (Pande et al., 2010). Moreover, the CART trees also remove the problems of multicollinearity (Chang and Wang, 2006). Minitab software has been used to develop the CART prediction model. Deaths being predicted as injuries is more dangerous than injuries being predicted as deaths. To minimize this error, the response variable was selected as ‘death’ so that at least number of deaths are predicted as injuries. The CART model sidelined most of the variables that were considered as significant for estimating fatal crashes. It utilized only the victim and accused categories for the prediction of fatal crashes and considered them as major contributing variables. This is in similar lines with the odd’s ratio from logistic regression where odd’s ratio of accused category and victim category are either too high than 1 or very less than 1. Figure 19 shows the relative importance of variables in the CART prediction model and Figure 20 shows the CART tree. It can be clearly observed from Figure 19 that category of victim and accused are the significant variables that have relatively higher importance in prediction of road crash fatalities. Other variables don’t seem to play significant role in the model development as being observed from Figures 19 and 20. Further, Figure 20 clearly reveals a dominant effect of victim category in deciding whether a crash will be fatal or non-fatal. While, pedestrians, self-hit, 2-wheeler and 3 wheelers have a higher probability of fatality when hit (28%), cars, jeeps, and heavy vehicles seem to be the safest categories (3.8%) when hit by others as can be observed from node 2 and terminal node 4. Terminal node 1 and node 3 again classify the node 2 category victims into 2 parts: pedestrian and self-hit with higher fatality rate (35%), and 2 wheelers and 3 wheelers with comparatively lower fatality rate (19.6%). Finally, node 3 is divided into 2 parts, terminal node 2 and terminal node 3 based on category of accused with respect to victim vehicles 2 wheelers and 3 wheelers. It is observed that when accused vehicle is HV, the fatality rate is 30.6% in case of 2 wheelers and 3 wheelers. For other categories of vehicles hitting 2 wheelers and 3 wheelers, the fatality drastically reduces to 12.8%. Thus, according to CART prediction model, these 2 variables are sufficient to predict whether a crash will result in fatality. Moreover, the model predicts 83.7% of fatal crashes correctly which is higher as compared to binary logistic regression model which predicted 79% of fatal incidents correctly. Thus, CART tree predicts the deaths with more precision by omitting the unnecessary variables.

### 3.3. Development of severity prediction index through two-stage clustering technique

Clustering technique is widely used by various researchers to classify data points (Jain and Dubes, 1988; Mohapatra et al., 2015; Boora et al., 2017; Mohanty and Dey, 2019; Mohanty et al., 2021-a). An efficient clustering method will produce clusters with property that their intra-cluster distance is small, and their inter-cluster distance is large (Jain and Dubes, 1988). The commonly used clustering algorithms are k-means, k-medoid, and hierarchical agglomerative. In this study, K-means clustering technique has been used for determining the ranges of crash severity indices since it is one of the better clustering techniques for large data sets of more than 500 (Boora et al., 2017). K-mean clustering is one of the unsupervised hard partitioning methods for explaining the classification problems (Mohapatra et al., 2015). Moreover, K-mean clustering technique uses the variation within every cluster as a quantity to form homogenous clusters.

Reported road crash data have a 22.7% chance of fatal crashes as mentioned earlier. Therefore, if obtained probability values of fatality (from CART model or logistic regression model) is less than 0.227, it can be adjudged as safe with respect to deaths or it can be termed as accidents with only injuries, since it is less than the overall average percentage of fatalities. Even the CART model, identified the probability values above 0.3 as deaths. While going through the data analysis of CART model, it is observed that the probability values are either below 0.22 or greater than 0.30. Based on the reported and collected accident data, the probability values haven’t come in between 0.227 and 0.30. This might be the reason for CART to predict values above 0.30 as deaths, since 0.25 or 0.27 were not present in the data under study. However, if all the reported crashes are considered, logically the probability value of 0.227 should be considered as a threshold or critical value, below which chances of injury is high and above which chances of death escalates. The models are not able to showcase this value as the computed/obtained probabilities is from a sample which are not within the range of 0.22–0.30. Thus, based on the descriptive statistics of the accident data along with probability

### Table 6. Preliminary proposed road crash severity index.

| Probability value obtained from prediction models (CART/logistic regression) | Remarks |
|---|---|
| <0.227 | Accidents with higher probability of injuries only |
| ~ 0.227 | Critical severity value with respect to deaths |
| >0.227 | Accidents with higher probability of deaths |

### Table 7. Initial and final cluster centers for probability values more than critical value.

| Cluster | Initial cluster centers | Final cluster centers |
|---|---|---|
| 1 | 0.290 | 0.340 |
| 2 | 0.574 | 0.456 |
| 3 | 0.228 | 0.268 |

### Table 8. Iteration history of clusters for probability values more than critical value.

| Iteration | Change in Cluster Centers |
|---|---|
| 1 | 0.044 0.098 0.015 |
| 2 | 0.003 0.012 0.013 |
| 3 | 0.002 0.008 0.011 |
| 4 | 0.000 0.000 0.001 |
| 5 | 0.000 0.000 0.000 |

Convergence achieved due to no or small change in cluster centers.
values of prediction model, a preliminary severity index can be provided as shown in Table 6.

0.227 is considered as the critical probability severity value below which chance of fatality diminishes and above which the chance of deaths increases. However, probability values of 0.05 and 0.20 for fatalities are not similar. Similarly, values of 0.30 and 0.60 for fatalities are same. Probability value of 0.05 is a much safer ride as compared to 0.20 where injuries might surely occur. Similarly, a probability value of 0.60 imposes a much higher degree of threat with respect to deaths clustering is conducted in a 2-stage clustering analysis.

1. First analysis is for obtained probability values greater than 0.227 and 2. second cluster analysis is for probability values less than 0.227.

It is observed that, for both the groups, i.e. (i) probability values more than 0.227 and (ii) probability values less than 0.227, the optimum number of clusters in 2-step cluster analysis is obtained as 3. The obtained silhouette value is 0.8 in both the cases which denotes strong cluster formation (Boora et al., 2017). Figure 21 shows the optimal cluster size and average silhouette value according to 2-step technique.

K-mean clustering is performed first for all the obtained probability values greater than 0.227 or critical severity value. The number of clusters are fixed to 3 in line with results obtained from 2-step analysis. The initial and final clusters along with iteration history before which convergence is reached is shown in Tables 7 and 8. As can be seen from Table 8, only after 5 iterations, the convergence has been achieved due to no change in cluster centers. Significance value for ANOVA has also been found to be less than 0.05 suggesting strong grouping of values which are independent of each other. The box plot visually identifies this difference in the groups.

Next the K-mean clustering analysis is performed for all the computed probability values less than 0.227 suggesting formation of good cluster groups. The significance value for ANOVA has also been found to be less than 0.05 suggesting strong grouping of values which are independent of each other. The box plot visually identifies this difference in the groups.

After completion of 2-stage clustering, the details of cluster memberships are studied and sorted to determine the ranges of severity

| Table 9. Initial and final cluster centers for probability values less than critical value. |
|-----------------------------------------------|
| Cluster  | 1     | 2     | 3     |
| Initial cluster centers                       | 0.010 | 0.070 | 0.225 |
| Final cluster centers                         | 0.022 | 0.117 | 0.190 |

| Table 10. Iteration history of clusters for probability values less than critical value. |
|-----------------------------------------------|
| Iteration | Change in Cluster Centers |
|-----------|---------------------------|
| 1         | 0.010 0.043 0.032         |
| 2         | 0.003 0.004 0.001         |
| 3         | 0.000 0.000 0.000         |
| Convergence achieved due to no or small change in cluster centers. |

| Table 11. Ranges for proposed road crash severity index. |
|---------------------------------------------------------|
| Severity Index | Probability values obtained from prediction models | Remarks                             |
| A              | 0.000–0.064 | Crash with chances of no to minor injury |
| B              | 0.065–0.152 | Crash with chances of moderate to high injury |
| C              | 0.153–0.226 | Crash with severe injury but no to very less chance of deaths |
|                | 0.227       | Critical severity value above which chance of death increases |
| D              | 0.228–0.301 | Crash with severe injury but less to moderate chance of deaths |
| E              | 0.302–0.399 | Crash with severe injury and moderate to high chance of deaths |
| F              | >0.400      | Crash with severe injury and high to severe chance of deaths |

Figure 22. Box plot of clusters for probability values more than critical severity value.
indices. The detailed road crash severity index is presented in Table 11. The 6 levels of severity indicators are identified by using A to F with A denoting the safest crash event and F denoting severe/sure chance of fatality. Table 11 clearly reveals that if the CART prediction model or logistic regression model estimates the probability of fatality below 6.4% (Index A), the crash event will be mostly safe with respect to deaths. There might be no or very minor injuries. As the probability increases, the chance of injuries also increases in case of a crash event. A value of 0.227 is regarded as critical severity value above which chance of deaths start increasing. Thereafter, if the prediction model predicts the values up to 30% probability of fatality (Index D), the chance of fatality is less but surely shall result in severe injury. Similarly, the predicted values within 30–40% (Index E) indicates a moderate to high chance of deaths incase of a road crash. Finally, if the crash probability is more than 40% (Index F), occurrence of death is almost certain.

The proposed indices shall be really helpful in managing the safety aspect on the roadway and providing various strategies to efficiently manage accident black spots. For example, if an accident-prone area is investigated for various contributing factors (both geometric and vehicular traffic with road user characteristics), and prediction values for fatalities are calculated using logistic regression and CART analysis as described above, then the computed values can be compared with the proposed severity index to understand the risk associated on the investigated road stretch. Further, composition of vehicles can be studied at accident black spots since they are the major contributing factor for fatalities on road. In case of pedestrian crowded road, or presence of more heavy vehicles, proper strategies can be undertaken like providing foot over bridge for pedestrians or dedicated timings for flow of heavy vehicles, etc.

4. Conclusion

The present study is conducted with an attempt to develop crash fatality prediction model and propose a crash severity index. In order to attain the objectives, the present study employs a large number of statistical tools including binomial logistic regression model, CART classification model and clustering technique. 2425 reported road crashes with 22.7% fatal crashes have been assessed for the present study. 14 numbers of contributing factors for road crash fatalities which included road, vehicle, and road users’ characteristics have been analysed for identifying the significant factors using Phi & Cramer’s test. Before proceeding towards crash severity prediction model using CART classification, 8 significant risk factors (obtained from Phi & Cramer’s test) which include categories of victim and accused, 85th percentile speed, presence of road markings, availability of sight distance, crash at intersection, presence of shoulder, and type of land use have been subjected to a binary logistic regression. The logistic model with a cut value of 0.23 predicted 79% of fatal crashes correctly. Odd’s ratio of each subcomponent variable was obtained to understand the effect of each factor and its sub variables on fatal road crashes.

The significant contributors from binary logistic regression model were category of accused vehicles, category of victim, 85th percentile speed range, availability of earthen shoulders, availability of adequate road markings, and availability/non-availability of sight distance. These factors were found to have either much higher odd’s ratio (>5) and lesser odd’s ratio (<0.5). These factors were considered for CART classification model. However, CART utilized only the victim and accused categories for the prediction of fatal crashes and considered them as major contributing variables. Moreover, the model predicts 83.7% of fatal crashes correctly which is higher as compared to binary logistic regression model. However, the number of input variables considered by logistic regression is much higher. Comparison of risk factors and prediction performances between the parametric models and non-parametric tree-based models provided valuable insights into the underlying relationship between risk factors and traffic injury severity.

Thereafter, a critical severity probability value is determined from reported accident data which is indicated as 0.227 with respect to deaths. Thereafter, SPSS software has been used for the cluster analysis in order to propose the crash severity indices. Two times, K-mean clustering has been conducted and 3 number of clusters (based on optimal cluster size by 2-step cluster) for both lesser and higher value than the critical severity value has been determined. While conducting k-mean clustering technique, the convergence is achieved just after 5 number of iterations indicating that 3 number of clusters is optimal for the data. The significance value for ANOVA has also been found to be less than 0.05 suggesting strong grouping of values which are independent of each other. 6 levels of severity indicators are identified by using A to F with A denoting the safest crash event with 6.4% of severity rate and F denoting severe/sure chance of fatality with more than 40% severity rate. These prediction values of severity indices will be very helpful in identifying the risk factors in accident prone area and minimize the rate of fatality. Practitioners, researchers, and law makers can utilize the methodology as well.
as results of this study to evaluate the accident-prone locations, thereby suggesting and incorporating specific strategies to reduce fatalities.

Declarations

Author contribution statement

Malaya Mohanty: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Rachita Panda: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Srinivasa Rao Gandupalli: Performed the experiments; Analyzed and interpreted the data.

Ritik Raj Arya, Sarthak Kumar Lenka: Performed the experiments.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interest’s statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Abdel-Aty, M., Abdelwahab, H., 2004. Modeling rear-end collisions including the role of driver’s visibility and light truck vehicles using a nested logit structure. Accid. Anal. Prev. 36 (3), 447–456.

Aithal, B.H., Ramachandra, T.V., 2016. Visualization of urban growth pattern in Chennai using geoinformatics and spatial metrics. J. In. Soc. Remote Sens. 44 (4), 617–633.

Aney, M.S., Ho, C., 2019. Deadlier road accidents? Traffic engineering to Evaluate the Performance of Two-Lane Highways. European Transport—Transporti Europei.

Caliendo, C., Guida, M., 2012. Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts. J. Transport. Eng. 138 (12), 1453–1467.

Cando-Sanz, N., Guirao, B., Vázquez-Pérez, D., 2019. Population ageing and rural road accidents: analysis of accident severity in traffic crashes with older pedestrians on Spanish crosstown roads. Res. Transport. Bus. Manage. 30, 100377.

Chang, L.T., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prev. 38 (5), 1019–1027.

Chen, Y., Sun, X., Zhong, L., Zhang, G., 2007. Speed difference and its impact on traffic safety of one freeway in China. Transport. Res. Rec. 2038 (1), 105–110.

De Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prev. 43 (3), 402–411.

Divakaran, L.M., Sreedhara, T., 2013. Accident prediction models for urban unsignalized intersections. Int. J. Innovat. Res. Sci. Eng. Tech., ISSN (online) 2319-8753.

Edwards, J.B., 1996. Weather-related road accidents in England and Wales: a spatial analysis. J. Transport Geogr. 4 (2), 201–212.

Goel, G., Sachdeva, S.N., 2016. Analysis of road accidents on NH-1 between RD 98 km to 148 km. Perspect. Sci. 8, 392–394.

Hu, M., Cai, W., 2022. Analysis of Lifeboat Embarkation Efficiency for Cruise Passengers under Multiple Scenarios . Scientific Programming.

Iyanda, A.E., 2019. Geographic analysis of road accident severity index in Nigeria. Int. J. Inj. Control Saf. Promot. 26 (1), 72–81. Taylor & Francis.

Jain, A.K., Dubes, R.C., 1988. Algorithms for Clustering Data”. Prentice-Hall.

Kar, S., Das, S.C., Tiwari, A., Phareven, I., 2016. Pattern of road traffic accidents in Bhubaneswar, Odisha. Clin. Epidemiol. Global Health 4 (3), 115–119.

Laureshyn, A., 2006. Assessment of traffic safety and efficiency with the help of automated video analysis. In: 19th ITCTC Workshop.

Lee, C., Li, X., 2015. Predicting driver injury severity in single-vehicle and two-vehicle crashes with boosted regression trees. Transport. Res. Rec. 2514 (1), 138–148.

Mohammed, S.J., 2012. Epidemiological study on road traffic accidents in Al-Najaf city during 2011. Kufa J. Nursing Sci. 2 (3), 138–142.

Mohanty, M., Dey, P.P., 2019. Quantification of LOS at uncontrolled median openings using area occupancy through cluster analysis. Arabian J. Sci. Eng. 44 (5), 4667–4679.

Mohanty, M., Gupta, A., 2016. Investigation of adolescent accident predictive variables in hilly regions. Int. J. Inj. Control Saf. Promot. 23 (3), 291–301.

Mohanty, M., Panda, B., Dey, P.P., 2021-a. Quantification of surrogate safety measure to predict severity of road crashes at median openings. IATSS Res. 45 (1), 153–159.

Mohanty, M., Dey, P.P., Panda, B., 2021-b. Assessment of traffic safety at median openings using surrogate safety measures: a case study in India. European Transp. 80, 3.

Mohanty, M., Panda, R., Gandupalli, S.R., Sonowal, D., Muskan, M., Chakraborty, R., Dangeti, M.R., 2022. Development of crash prediction models by assessing the role of perpetrators and victims: a comparison of ANN & logistic model using historical crash data. Int. J. Inj. Control Saf. Promot., Taylor & Francis.

Mohapatra, S.S., Dey, P.P., Chandra, S., 2016. Modeling the critical position of U-turning vehicles at uncontrolled median openings. KSCE J. Civ. Eng. 20 (1), 411–420.

Mohapatra, S.S., Panii, A., Sahu, P.K., 2021. Examining the impacts of logistics sprawl on freight transportation in Indian cities: implications for planning and sustainable development. J. Indian Plan. Dev. 147 (4), 04021050.

Mohapatra, S.S., Sil, G., Dey, P.P., 2015. Quantification of LOS at median openings through cluster analysis. Indian Highways 43 (3), 25–31.

Mokhtarimousavi, S., 2019. A time of day analysis of pedestrian-involved crashes in California: investigation of injury severity, a logistic regression and machine learning approach using HSIS data. Inst. Transport. Engineers. IEEE J. 89 (10), 25–33.

Montebella, A., Aria, M., D’Ambrosio, A., Mautiello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Anal. Prev. 49, 58–72.

Mooi, E., Sarstedt, M., 2014. A Concise Guide to Market Research”, the Process, Data, and. Springer, p. 12.

MORTH, 2019. Road Accidents in India.

NHTSA, 2010. NCSA Tools, Publicat. Data. Assessed at https://cdan.dot.gov/.

Pande, A., Abdel-Aty, M., Das, A., 2010. A classification tree based modeling approach for segment related crashes on multilane highways. J. Saf. Res. 41 (5), 391–397.

Peasapati, L.N., Hunter, M.P., Rodgers, M.O., 2013. Evaluation of postcollision time as surrogate for opposing left-turn crashes. Transport. Res. Rec. 2386 (1), 42–51.

Pirdavani, A., Bejji, T., Bellenens, T., Wets, G., 2010. Evaluation of traffic safety at unsignalized intersections using microsimulation: a utilization of proximal safety indicators. Adv. Transport. Stud. 22, 43–50.

Rahman, M.H., Zafri, N.M., Akter, T., Pervaz, S., 2021. Identification of factors influencing severity of motorcycle crashes in Dhaka, Bangladesh using binary logistic regression model. Int. J. Inj. Control Saf. Promot. 28 (2), 141–152.

Sanggaram, P., Loarun, W., Rodchuenn, M., 2022. Adjusted clustered rank tests for clustered data in unbalanced design. Thailand Statistici. 20 (2), 435–451.

Sayad, T., El Easawy, M., Pump, J., 2007. Evaluating the safety impacts of improving visibility signal at urban signalized intersections. TRB 86th Annual Meeting: Compendium of Papers CD-ROM TRB, 7–135.

Soltani, A., Ardak, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. Injury, Elsevier 48 (3), 637–647.

Tamakloe, R., Das, S., Aidoo, E.N., Park, D., 2022. Factors affecting motorcycle crash casualty severity at signalized and non-signalized intersections in Ghana: insights from a data mining and binary logit regression approach. Accid. Anal. Prev. 165, 105517.

Tola, A.M., Demissie, T.A., Saathoff, F., Gebissa, A., 2021. Severity, spatial pattern and statistical analysis of road traffic crash hot spots in Ethiopia. Sci. Publ. 11 (9), 8828.

Vedagiri, P., Pragna, T., 2013. Evaluation of Trafﬁc injuries at uncontrolled signalized intersections using binary logistic regression model. Int. J. Inj. Control Saf. Promot. 28 (2), 141–152.

Verster, T., Fourie, E., 2018. The good, the bad and the ugly of South African fatal road accidents. South Afr. J. Sci. 114 (7-8), 63–69.

WHO, 2021. Road Traffic Injuries. Accessed at https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

Yan, X., Radwan, E., Abdel-Aty, M., 2005. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accident Anal. Prevent. 37 (6), 985–993.

Zhang, X., Wen, H., Yamamoto, T., Zeng, Q., 2021. investigating hazardous factors affecting freeway crash injury severity incorporating real-time weather data: using a Bayesian multimodal logit model with conditional autoregressive priors. J. Saf. Res. 76, 318–325.

Zhu, L., Li, X., Cheng, H., Zhang, J., Wang, Q., Zhu, H., 2022. Deep learning-based prediction of traffic accidents risk for Internet of vehicles. China Communicat. 19 (2), 214–224.