4-16-2021

Fatal crashes at highway rail grade crossings: A U.S. based study

Subasish Das

Xiaoqiang Kong

Steven M. Lavrenz

Lingtao Wu

Mohammad Jalayer
Rowan University, jalayer@rowan.edu

International Journal of Transportation Science and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

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

Fatal crashes at highway rail grade crossings: A U.S. based study

Subasish Das [a,*], Xiaoqiang Kong [b], Steven M. Lavrenz [c], Lingtao Wu [a], Mohammad Jalayer [d]

[a] *Texas A&M Transportation Institute, Bryan, TX 77807, United States*
[b] *Texas A&M University, College Station, TX 77843, United States*
[c] *Civil and Environmental Engineering, Wayne State University, Detroit, MI 48202, United States*
[d] *Department of Civil and Environmental Engineering, Rowan University, Glassboro, NJ 08028, United States*

ARTICLE INFO

Article history:
Received 14 October 2020
Received in revised form 22 December 2020
Accepted 3 March 2021
Available online xxxx

Keywords:
Highway rail grade crossing crashes
Taxicab correspondence analysis
Pattern recognition
Safety improvement

ABSTRACT

Crashes at highway rail grade crossings (HRGCs) are often involved with fatalities due to the momentum of a train. This study collected nine years (2010–2018) of fatal HRGC crashes from the Fatality Analysis Reporting System (FARS) to perform the analysis. The Taxicab Correspondence Analysis (TCA) was applied to this dataset. This method identified several patterns that trigger HRGC-related fatal crashes. The findings indicate that fatal crashes involving multiple fatalities are often highly associated with alcohol-influenced drivers, poor lighting conditions, and inclement weather. The fatal crash that occurs during the daylight with the uninfluenced driver is less likely to involve more than one fatality. The results also recognized the combinations of vehicle type and speed are associated with fatal crashes at rail grade crossings. The relatively low-speed limit crossings and larger utility vehicles are more likely to be associated with fatal crashes because large vehicles require a longer time to cross railroads at a low speed. The relatively high-speed limit crossing and smaller or lighter vehicles, especially the motorcycle, are highly associated with fatal crashes.

Introduction

Crashes that occur at highway rail grade crossings (HRGCs) may result in severe injury outcomes due to the enormous mass of trains. The reduction of HRGC crashes is essential for policymakers and city planners. From 2010 to 2018, a total of 1114 fatal crashes occurred at HRGCs in the United States, resulting in 1306 fatalities. During the same period, HRGC related crashes reduced by 20% (FARS, 2020). However, the death toll is still huge.

Safety at HRGCs continues to be a major safety issue despite the improved safety practices in recent years. Although safety improvement efforts for HRGCs are continuous, crash counts and associated safety concerns remain high. Past studies include many research efforts, including crash prediction models by incorporating roadway, HRGC inventory, rail, and vehicle traffic characteristics. A group of HRGC studies has viewed specific classes of warning devices (Ries, 2007; Raub, 2006, 2009; Horton and DaSilva, 2013; Schoppert and Hoyt, 1968; Lerner and Tucker, 2002; Gabree et al., 2014; Noyce and Fambro, 1998; Landry et al., 2016). Another group of HRGC studies examined the safety issues by either focusing on crash

Peer review under responsibility of Tongji University and Tongji University Press.

* Corresponding author.
E-mail address: s-das@tti.tamu.edu (S. Das).

https://doi.org/10.1016/j.ijtst.2021.03.002
2046-0430/

frequency or crash severity analysis (Hao and Kamga, 2017; Hao et al., 2017; Khan et al., 2018; Keramati et al., 2020; Zheng et al., 2019, 2016; Eluru et al., 2012; Oh et al., 2006; Zhao et al., 2019; Yan et al., 2010).

It is important to note that determining the crucial crash contributing factors is a key task of highway safety analysis. This study aims to identify the patterns of the factors that are associated with HRGC related fatal crashes. Taxicab correspondence analysis (TCA), a data mining approach, can provide insights on the importance of variable categories from a high dimension dataset. This study used nine years (2010–2018) of fatal HRGC crash data in the United States from the national level Fatality Analysis Reporting System (FARS) data. The purpose of the present study is to evaluate the patterns and nature of HRGC crash occurrences in order to create appropriate countermeasures and improve roadway safety.

The rest of the paper is organized in the following four sections: The next section provides a literature review on HRGC related studies. The third section focuses on data integration, exploratory data analysis, and a brief overview of TCA. The fourth section presents the results using the proposed approach and related discussions. The last section explains the conclusions of the study.

Literature review

The HRGC related safety analysis mostly focused on either frequency-based analysis or severity-based analysis. Using the ordered probit model, Hao and Kamga (2017) explored the key contributors of crash severity at rural HRGCs compared with urban ones. The analysis found that motor vehicle driver's injury level at rural HRGCs is much higher than in urban areas. Hao et al. (2017) estimated the effect of foggy conditions on crash severity outcomes at HRGC locations. The results demonstrated that foggy condition crashes tend to result in more severe injuries than normal or clear weather conditions. Older drivers are more likely to suffer severe injuries in foggy condition crashes than in normal conditions due to their slow reaction times. Drivers are more likely to experience high-level injuries in crashes in the early morning during winter.

By developing a binary logit regression model, Khan et al. (2018) predicted crash likelihood at HRGCs by incorporating several contributory factors, including the U.S. Census Block level population within five miles of crossings. This study used seventeen years (2000–2016) of North Dakota crash data to develop the models. The results show that daily train exposure, the maximum train speed, frequency of through railroad tracks, and the number of roadway lanes are associated with crash counts. This study shows that 'stop' pavement marking reduces the crash likelihood, while populations within five miles of HRGCs have a positive relationship with crash frequencies. Keramati et al. (2020) evaluated the effects of geometric characteristics of HRGCs on crash and severity likelihoods. Using data from 3194 public HRGCs in North Dakota, four main HRGC geometric factors (acute crossing angle, number of railway tracks, distance between the HRGC and presence of signalized intersection, and number of roadway lanes), along with other contributors, were explored. Zheng et al. (2019) used a neural network (NN) model to investigate rain-vehicle crash risk at HRGCs to determine dependent nonlinear contributor-crash curves with all other contributors considered for a specific contributor variable. The study used historical crash data for North Dakota public HRGCs from 1996 to 2014. Eluru et al. (2012) examined crash severity patterns with a latent ordered response model using ten years of crash data at HRGCs across the nation.

There is a large body of research that aims to assess HRGC crash risk factors pertaining to driver behaviors or operational characteristics of the roadway. Higher traffic volumes, higher train speeds, and a higher percentage of large trucks at the crossing are all associated with higher rates of fatal and severe injury crashes (Raub, 2009; Lee et al., 2019; Millegan et al., 2009; Abraham et al., 1998). Other characteristics associated with higher serious injury crash rates include driver age (Raub, 2009; Abraham et al., 1998) and time of day (Salim et al., 2018; Salim, 2018). Several studies have found the speed of the vehicle approaching the grade crossing to be significantly associated with crash risk; that is, higher levels of travel speed are found to be associated with an increased risk of a crash or greater crash severity (Millegan et al., 2009). This may be due to the inability of the driver to accurately gauge safe stopping distances at higher speeds, along with the heightened difficulty of processing multiple external stimuli for correct decision making.

The review found that most research to date has focused on the evaluation of HRGC treatments or perform safety analysis using conventional methods, with little consideration given to the identification of patterns of key contributing factors. Therefore, the aim of this study is to develop a complete understanding of the HRGC fatal crash-related contributing factors and fill this critical gap in knowledge in this field.

Methodology

Data collection

This study used nine years (2010–2018) of HRGC related fatality data from the FARS database. This study used the RELJCT2 variable attribute to identify HGRC related fatalities. The RELJCT2 is the Relation to Junction-Specific Location, which identifies the crash locations concerning the presence in or proximity to components typically injunction or interchange areas; a variable attribute of 6 indicates Railway Grade Crossing. During the nine years studied, 1114 fatal crashes occurred, with 1306 fatalities.

HRGC crashes by States

Fig. 1 shows the HRGC related fatalities for all fifty states in the United States from 2010 to 2018. The year groups (3-year vs. 2-year) are due to the use of nine years of data. However, the color gradient represents the average number of fatalities in the year groupings. Each of the sub-plots indicates the average number of fatalities for four different temporal groups. Purple indicates a fewer number of deaths, while yellow indicates a larger number of fatalities. The states are displayed in a region relative to their geographical location. Overall, Texas experienced a significantly higher number of HRGC related deaths than any other state. States in the northeastern region of the United States had the lowest amount of HRGC related fatalities. From 2010 to 2012, Texas was the only state with a yellow color grade, meaning the highest number of fatalities. Other states with a high number of fatalities are Illinois, Indiana, California, and Louisiana. From 2013 to 2014, Texas accounted for the highest number of fatalities, followed by Illinois, Indiana, California, Oklahoma, and Alabama. From 2015 to 2016, Texas was also the state with the highest number of fatalities, followed by Illinois. From 2017 to 2018, a significant shift occurred. Indiana was the state with the highest number of fatalities. Florida also experienced a significantly higher number of fatalities, while Texas showed a marked decrease in fatalities.

The primary selection of variables contained a wide array of geometric, traffic, environment, and vehicle-related data. After performing variable importance measures using information criteria based on random forest (RF), a set of variables was selected for further analysis (Greenwell et al., 2019). RFs proffer an extensive method for computing variable importance scores. The concept is to utilize the leftover out-of-bag (OOB) data to create validation related errors for each of the generated trees. Later, each predictor is shuffled at random in the OOB dataset using a computation of error measures. The general idea is that if a variable is significantly important, then the validation error is going to increase when that variable is perturbed in the OOB dataset. The disparity between these two error measures will then be averaged across all trees in the forest. Table 1 displays percent distributions of the selected variable categories. For 87.6% of fatal crashes, one fatality occurred. Alcohol involvement in fatal crashes was 19%. The occurrence of a rollover was not present in most fatal crashes (83.33%).

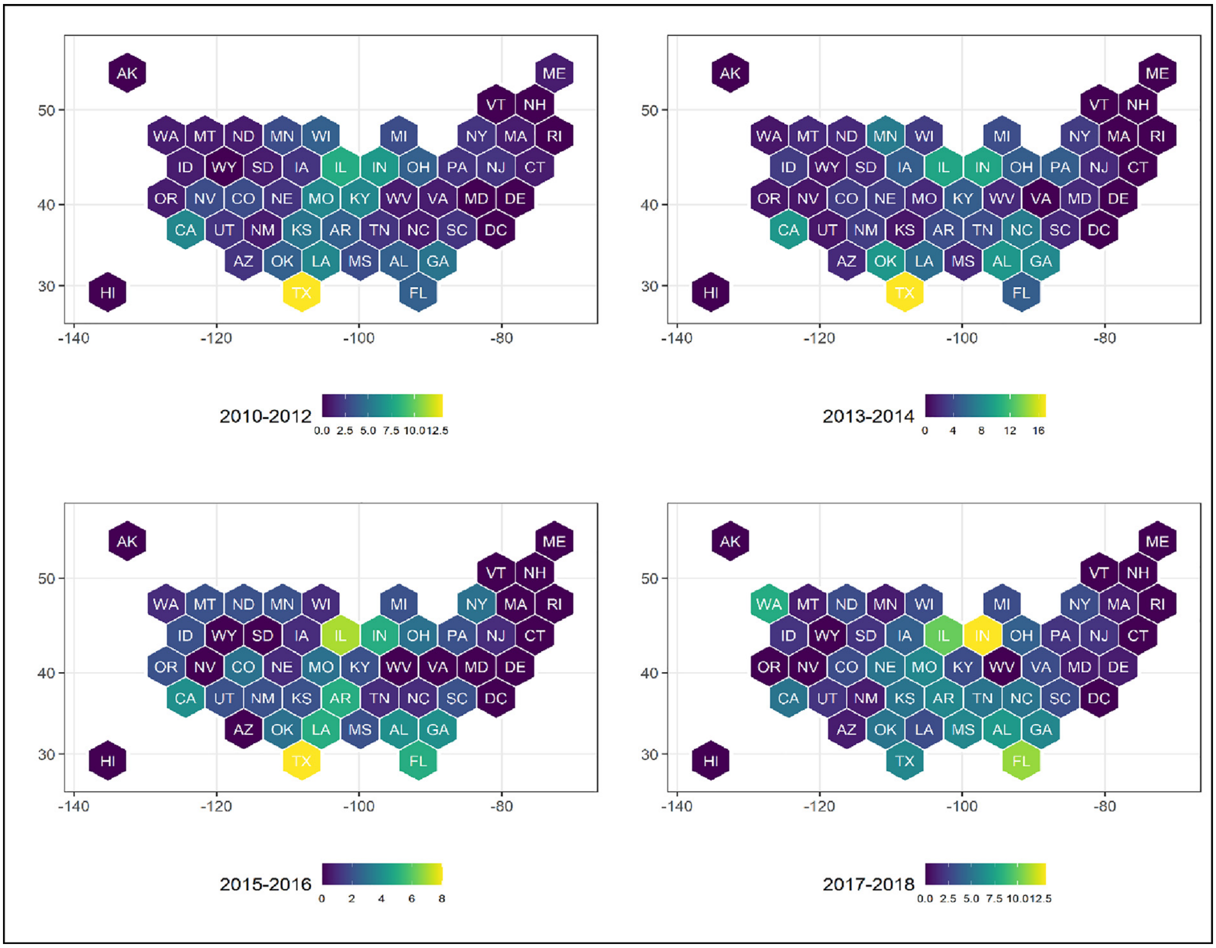

Fig. 1. HRGC related fatalities (2010–2018) by the U.S. states.

Table 1
Descriptive Statistics.

Attributes	Percent	Attributes	Percent
NUMF (No. of Fatalities Involved)		**LGT (Lighting Condition)**	
One	87.56	Daylight	68.83
More than One	12.44	Dark Lighted	11.92
ALC (Impairment of Driver)		Dark Not Lighted	15.2
No	80.83	Dawn/Dusk	4.06
Yes	19.17	**VTRAFWAY (Roadway Type)**	
DR_SF (Driver Violation)		Two-Way, Not Divided	87.82
Failure to Obey Signs, Traffic Control Device (TCD)	36.27	Two-Way, Divided, Unprotected Median	7.51
Failure to Yield Right of Way	29.27	Two-Way, Divided, Median Barrier	2.07
Careless Driving	3.71	Others	2.59
Others	13.04	**VNUM (Number of Lanes)**	
None	17.7	Two Lanes	87.82
PCRASH (Prior Movement)		Three Lanes	3.54
Going Straight	74.96	Four Lanes	5.61
Stopped in Roadway	9.67	Five Lanes	0.86
Others	8.89	Others	2.16
Negotiating a Curve	4.06	**VPROFILE (Alignment)**	
Turning Related	2.42	Level	62.26
ROUTE (Route Type)		Grade, Unknown Slope	18.22
County Road	33.68	Hillcrest	8.29
Local Street Municipality	29.71	Uphill	4.84
Local Street Township	14.68	Downhill	1.38
State Highway	11.49	Others	5.01
U.S. Highway	5.79	**VTYP (Body Type)**	
Other	4.66	4 Door Sedan	29.88
WEATHER (Weather Condition)		Compact Utility	13.73
Clear	74.27	Standard Pickup	15.89
Cloudy	13.99	Truck Tractor	5.7
Inclement	11.74	Single Unit Straight Truck	5.44
PSL (Posted Speed Limit)		Minivan	4.49
10–35 mph	45.94	2 Door Sedan	4.23
40–55 mph	41.11	Station Wagon	3.37
60–70 mph	4.58	Large Utility	3.2
Not Reported	8.38	Farm Equip	1.9
ROLLOVER (Occurrence of Rollover)		Motorcycle	2.16
No Rollover	83.33	Others	5.44
Rollover, Tripped by Object/Vehicle	16.06	Unknown Body Type	4.58
Rollover, Untripped	0.6		

The prior movement in a majority of reported fatalities was going straight (74.96%). From the other attributes of prior crashes, 10% of fatal crashes happened when the vehicle stopped on the road or track. The alignment of the road was level in most reported fatalities (62.26%). It is important to note that over 30% of fatal crashes occurred on grade, hillcrest, and uphill/downhill locations. Two-way undivided was reported as the roadway type in a majority of fatalities (87.82%). Cloudy and inclement weather contributed to 25% of fatal HRGC crashes. The majority of the HRGC crashes occurred on low-speed roadways: 10–35 mph (45.94%) and 40–55 mph (41.11%). A majority of the fatal crashes (80%) occurred on country or local roadways. Dark and dusk/dawn lighting conditions contributed to 25% of fatal HRGC crashes. It is also found that two-lane roadways are dominant in frequencies (87.82% of all fatal crashes).

The majority of traffic safety studies do not examine interactions between the variables. The omission of this critical issue can produce biased results. The current study aims to mitigate this research gap in HRGC crash analysis by using TCA. Before performing TCA analysis, it is important to show the interaction patterns of categorical attributes used in this analysis. Alluvial plots are excellent data visualization tools that can be used in explaining complex structural networks in crash datasets.

In an alluvial plot, the length of the black bars represents the proportion of attributes by variable. The width of the bands indicates the relative proportion between the variable attributes. For example, daylight, clear, failure to obey signs and other traffic control devices (TCD), no rollover are the dominant attributes in lighting conditions, weather conditions, driver violations, and occurrence of rollover variables, respectively (see Fig. 2(a)). Similarly, in Fig. 2(b), roadway alignment as level, two-way undivided roadways, 4-door sedan as the vehicle body type, and one fatality are the dominant attributes. Alluvial plots revealed several key patterns through a two-dimensional plot. Although daylight is the dominating attribute, it is found that inclement weather related crashes also occurred in a similar proportion at dark or dawn/dusk lighting conditions. Careless driving was seen to occur mostly during clear weather. Few of these careless driving-related crashes occurred during cloudy and inclement weather. It is interesting that careless driving is related to tripped rollover crashes but not with untripped rollover crashes. Fig. 2(b) can also be explained in a similar pattern. Most of the crashes only involved one fatality. The vehicles of crashes with more than one fatality are mostly 4-door sedan. The plot also shows that the two-way undivided roadway type dominates the roadway attributes associated with crashes.

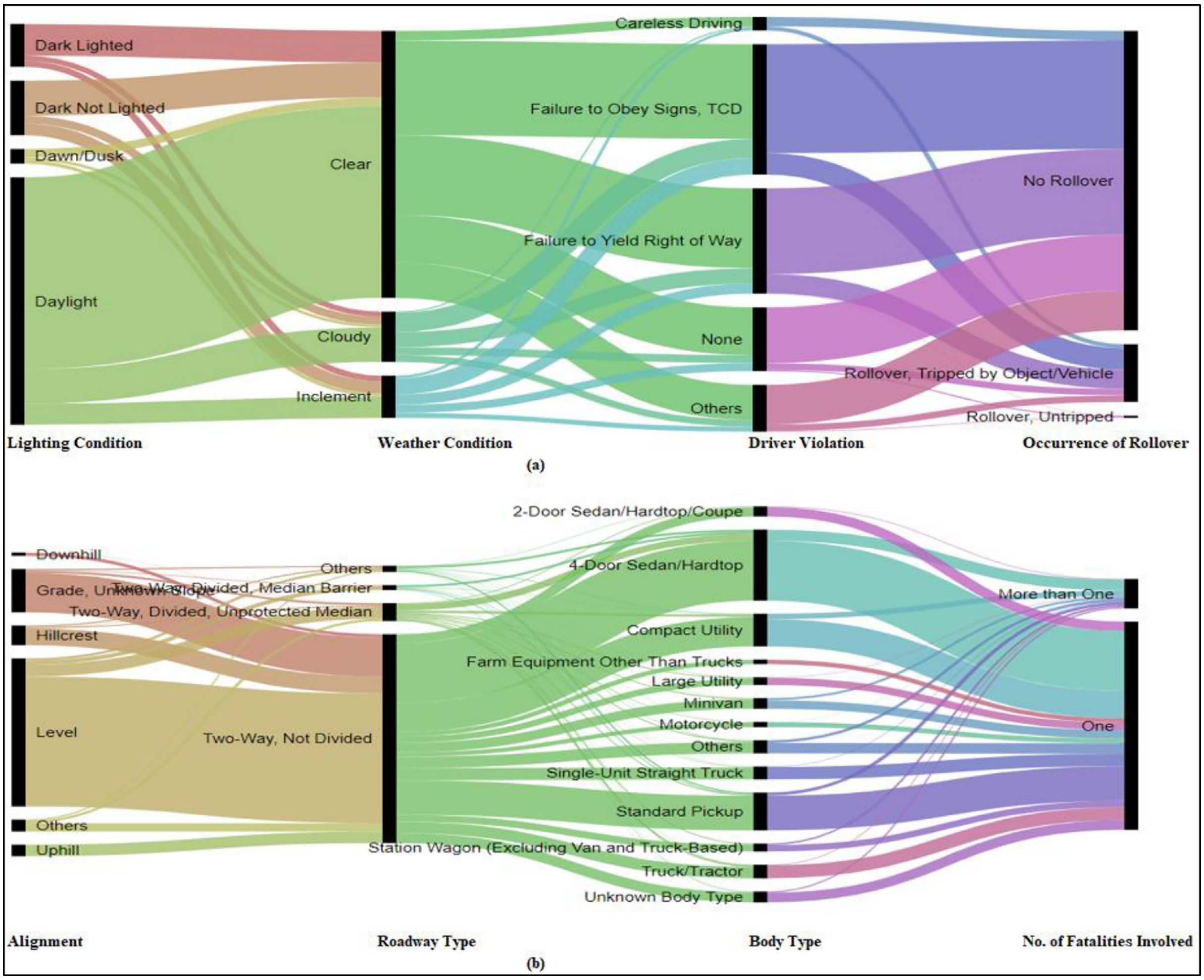

Fig. 2. Alluvial plots to show the distribution patterns of multiple variables.

Taxicab correspondence analysis (TCA)

Jean-Paul Benzécri first introduced the concept of correspondence analysis (CA). Choulakian recently introduced the theory of TCA, which is considered an enhanced version of CA (Choulakian, 2006). CA is based on Euclidean distance. On the other hand, TCA considers a separate distance measure, which is known as Manhattan city block or taxicab distance. For example, consider $A = (a_1, a_2, \cdots.., a_n)$ and $B = (b_1, b_2, \cdots.., b_n)$ and a vector $\mathbf{v} = (v_1, v_2, \cdots.., v_n)$ to evaluate these two specific distances:

$$\text{Euclidean Distance} = ed(A, B) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}[\text{with } L_2\text{Norm} = \|v\|_2 = \sqrt{\sum_{i=1}^{n} (v_i)^2} \tag{1}$$

$$\text{Taxicab Distance} = td(A, B) = \sum_{i=1}^{n} |a_i - b_i|[\text{with } L_1 \text{ Norm} = \|v\|_1 = \sum_{i=1}^{n} |v_i| \tag{2}$$

The concept of singular value decomposition (SVD) is critical in understanding CA and TCA. For example, first consider a real matrix K, decomposed as $M\Lambda^{1/2}N'$ with Λ the diagonal matrix of the real non-negative eigenvalues of KK', where M is the orthogonal matrix of the corresponding eigenvectors, and N the matrix of eigenvectors of $K'K$ (with constraints $M'M = I$ and $N'N = I$).Choulakian (29) proposed a recursive optimization process to evaluate the SVD solution. TCA can be denoted as the Taxicab SVD of the data table $D = T - rl'$ by considering the table's profiles, respectively $R = D_r^{-1}D$ for the rows and $L = D_l^{-1}D$ for the columns. Unlike CA, the solution is always recursive in nature. Interested readers can consult additional details of the TCA theory in the series of papers published by Choulakian (Choulakian, 2006).

Traditional statistical modeling techniques require structured data with response variables and explanatory variables; these techniques also require some prior assumptions. Data mining and dimension reduction methods (for example, TCA) have a clear advantage because they do not require any prior assumptions. Furthermore, these methods can work on both supervised (in which response and explanatory variables are known) and unsupervised data (in which response and explanatory variables are unknown). In TCA, the aim is to show the co-occurrence of the categories (for example, 'daylight' is a category of 'lighting condition' variable) in a low dimensional space where closer vicinity in the space indicates the co-occurrence of the categories. Another great advantage of TCA is its capability of noise reduction (by representing the data in low dimensional spaces) without reducing the dataset size. As the HRGC crash dataset has a limited number of cases, removing data entries with missing information would make a small dataset smaller. This method facilitates in describing the significant associations between the categories of complex datasets like HRGC crashes. In recent years, several studies related to transportation safety used both CA and TCA methods to identify insights from the complex nature of crash datasets (Das et al., 2018, 2019a, 2019b; Das and Sun, 2016; Jalayer et al., 2018).

Results and discussions

This study used an open source R software package 'TaxicabCA' (Allard and Choulakian, 2020) to perform the analysis. The analysis included the 13 variables (with all categories in each variable) listed in Table 1. The TCA approach helps in understanding diverse variable categories and produces visible results from the key association patterns. The TCA method produces coordinate measures for each of the variable categories for multiple axes. The first two axes explain around 49% of total dataset inertia (see Table 2). It means that 49% of the total variability is explained by the plane. This percentage is relatively high, and thus, the first plane well represents the data variability. Fig. 3 and Fig. 4 show the TCA plot generated in this analysis (the complete two-dimensional plot is divided into four quadrants for easy visual interpretation). The closer the coordinates of the variable categories, the closer the association. Table 3 lists the coordinates of the attributes by sorting the attributes based on the signs of the coordinate measures. For example, the first 17 variable categories are in quadrant 1 as both coordinates are positive. Quadrant-based TCA plots are shown in Fig. 3 and Fig. 4.

Clusters based on attribute locations

TCA has been used as a pre-processing step to develop a framework to visualize data. The hierarchical clustering was applied to the two-dimensional map provided by the TCA outcomes. The simultaneous use of both methods represents the clustering issue from the dendrogram on the map, which is improved by inspecting the ratio of 'between inertia' and 'total inertia.' This study used a range of (0, 10) to determine the optimal number of clusters. This procedure defined an optimal number of six clusters, which is why this study suggested three as the default. The description of the clusters is provided below.

Cluster 1

The attributes in this cluster are rollover, cloudy weather, inclement weather, dawn/dusk lighting, the presence of the impairment of alcohol, and more than one fatality (see Fig. 3a). This cluster indicates that alcohol-impaired related HRGC fatal rollover crashes at visually challenging situations while crossing the rail tracks, such as dusk/dawn lighting or inclement weather conditions. Driving under poor lighting environments or inclement weather conditions already poses difficulties for normal drivers. It increases extra risks for drivers under the influence. These crashes often involve multiple fatalities. The results of this cluster are in line with the Hao et al. (2017) study. Countermeasures to improve visibility prior to the RGC, such as the addition of retroreflective signage, may help to mitigate these crash risks. Measures to lower travel speed near the HRGC may also help, in providing drivers additional time for decision-making, as well as reducing the likelihood of the most severe crashes, such as untripped rollovers.

Cluster 2

The attributes in this cluster (see Fig. 3a) are compact utility vehicle type, large utility vehicle type, 4 door sedan vehicle type, careless driving, local street municipality route, dark not lighted as lighting conditions, and low-speed limit (10–35 mph) roadways. This cluster also shows the effect of dark lighting conditions and challenges of the drivers while crossing the junctions. Vehicle type is also an issue at HRGC locations. Large vehicles usually require additional time to cross the junctions, especially while driving on roadways with a relatively low-speed limit. This is consistent with previous studies. For

Table 2
Variance Explanation by the First Five Axes.

Axis	Inertia Explanation
Axis 1	26.5%
Axis 2	22.2%
Axis 3	13.7%
Axis 4	9.3%
Axis 5	6.5%

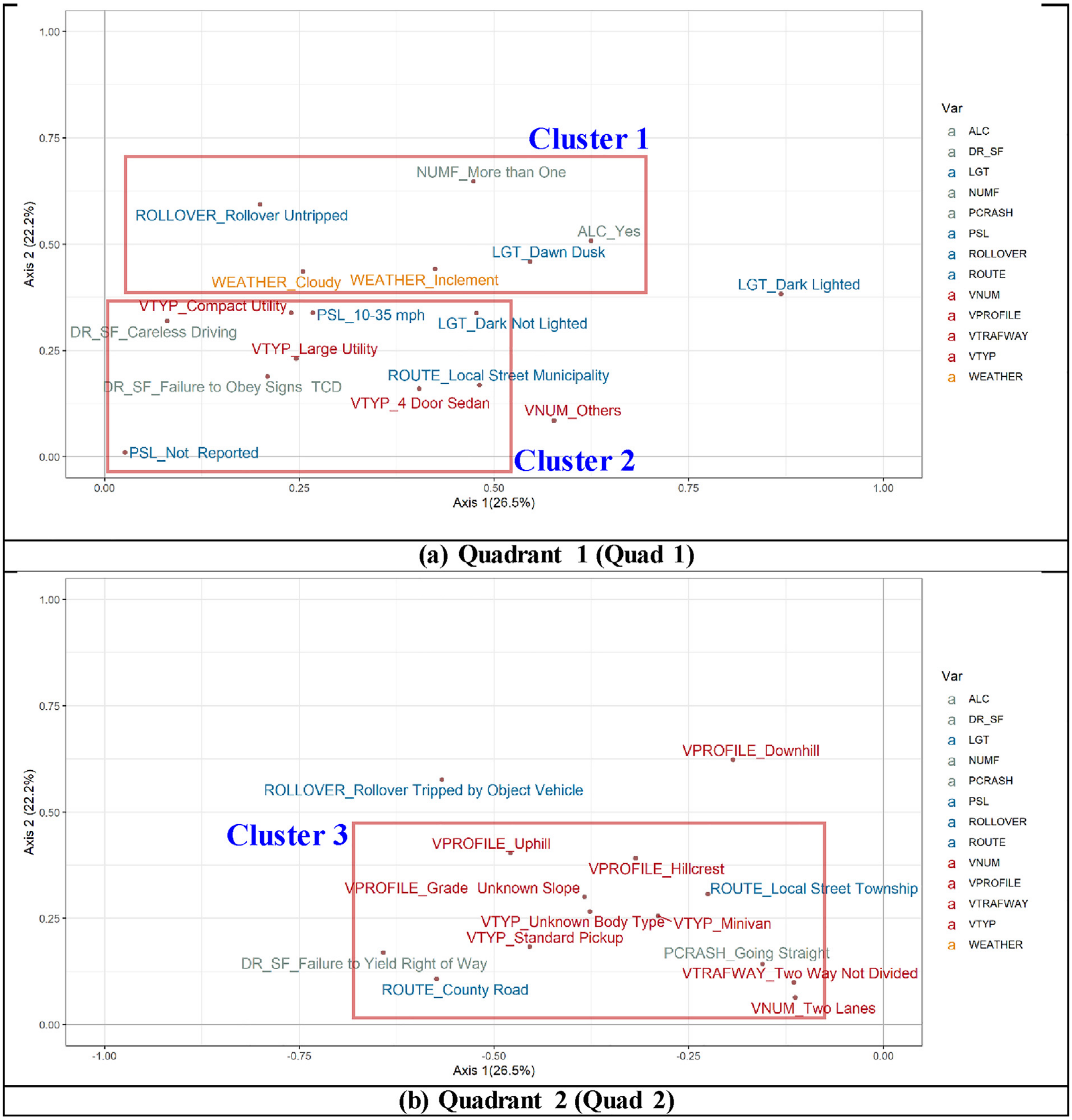

Fig. 3. TCA plot of the upper right and upper left.

example, Hao et al. (2016) pointed out that 25% of highway-rail grade crossing accidents involved trucks, which only make up about 4% of vehicles on the road. Another study also found that truck percentage is a significant factor determining HRGC safety (Yan et al., 2010). Conspicuity and speed-related HRGC treatments may also prove beneficial for reducing these types of crashes.

Cluster 3

The attributes in this cluster are vehicle alignment as uphill/hillcrest/slope grade as alignment, local street township route, two-way undivided roadways, two-lane roadways, county road route, and standard pickup or minivan as vehicle type (see Fig. 3b). The uphill/hillcrest/slop grade always creates an additional obstruction for vehicle drivers. Without being alerted beforehand, the crashes that occur at these locations are often associated with fatal crashes. This result is in line with Eluru et al. (2012), which indicates that both vehicle type and roadway classification influence the likelihood of assigning a

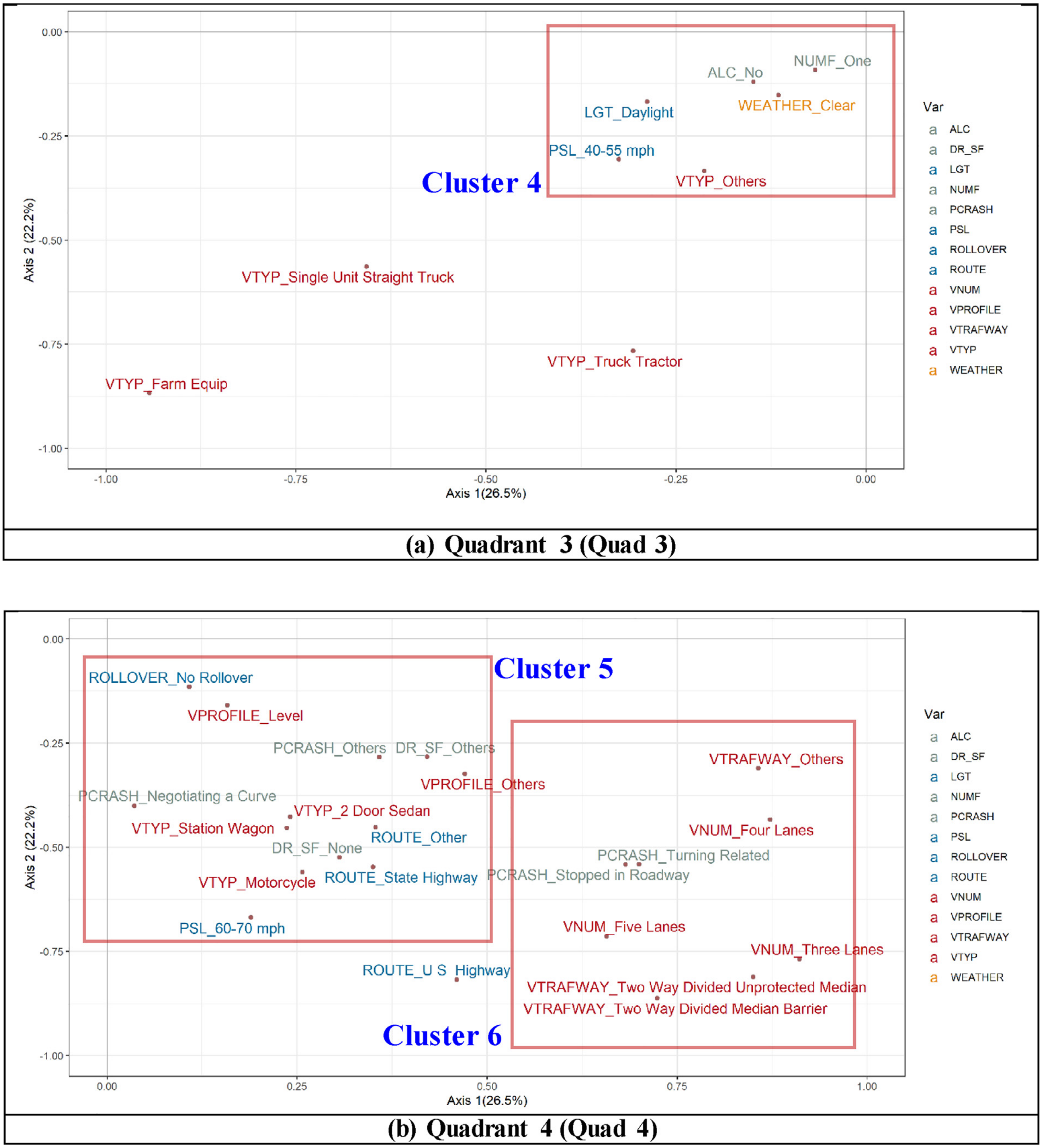

Fig. 4. TCA plot of the lower right and lower left.

driver's crash risk. This cluster of crashes may be more difficult to mitigate, as the relevant risk factors suggest that larger geometric changes to the roadway, such as right-of-way realignment, are needed.

Cluster 4

The attributes of this cluster are fatal crashes with one fatality, posted speed limit 40–55 mph, clear weather, no impaired driver, and daylight (see Fig. 4a). This cluster signifies fatal crashes that occur under uncompromised conditions, clear weather, daylight, and no impairment, mostly involve one fatality. In other words, crash with multiple fatalities is less likely to happen under clear weather, daylight, and no impairment conditions. It is reasonable to believe that if there is a passenger present in the vehicle in daylight and clear weather conditions, the passenger could alert the driver while approaching a

Table 3
Locations of the Key Attributes.

Variable	Attribute	Axis 1	Axis 2	Quadrant
PSL	Not Reported	0.0261	0.0101	Quad 1
DR_SF	Careless Driving	0.0803	0.3188	Quad 1
ROLLOVER	Rollover Untripped	0.1999	0.5932	Quad 1
DR_SF	Failure to Obey Signs TCD	0.2094	0.1886	Quad 1
VTYP	Compact Utility	0.2394	0.3387	Quad 1
VTYP	Large Utility	0.2462	0.231	Quad 1
WEATHER	Cloudy	0.2545	0.4354	Quad 1
PSL	10–35 mph	0.2675	0.3387	Quad 1
VTYP	4 Door Sedan	0.4038	0.1597	Quad 1
WEATHER	Inclement	0.4246	0.4414	Quad 1
NUMF	More than One	0.4737	0.6473	Quad 1
LGT	Dark Not Lighted	0.4774	0.338	Quad 1
ROUTE	Local Street Municipality	0.4814	0.1684	Quad 1
LGT	Dawn Dusk	0.5464	0.4591	Quad 1
VNUM	Others	0.577	0.0851	Quad 1
ALC	Yes	0.6246	0.5075	Quad 1
LGT	Dark Lighted	0.8686	0.3827	Quad 1
DR_SF	Failure to Yield Right of Way	−0.6421	0.1689	Quad 2
ROUTE	County Road	−0.5738	0.1073	Quad 2
ROLLOVER	Rollover Tripped by Object Vehicle	−0.5667	0.576	Quad 2
VPROFILE	Uphill	−0.4787	0.4035	Quad 2
VTYP	Standard Pickup	−0.4539	0.1834	Quad 2
VPROFILE	Grade Unknown Slope	−0.3838	0.3006	Quad 2
VTYP	Unknown Body Type	−0.377	0.2657	Quad 2
VPROFILE	Hillcrest	−0.318	0.3913	Quad 2
VTYP	Minivan	−0.2892	0.2555	Quad 2
ROUTE	Local Street Township	−0.2254	0.3072	Quad 2
VPROFILE	Downhill	−0.193	0.6225	Quad 2
PCRASH	Going Straight	−0.155	0.1427	Quad 2
VTRAFWAY	Two Way Not Divided	−0.1151	0.0989	Quad 2
VNUM	Two Lanes	−0.1131	0.0636	Quad 2
VTYP	Farm Equip	−0.943	−0.8671	Quad 3
VTYP	Single Unit Straight Truck	−0.6573	−0.5637	Quad 3
PSL	40–55 mph	−0.3254	−0.3062	Quad 3
VTYP	Truck Tractor	−0.3066	−0.7658	Quad 3
LGT	Daylight	−0.288	−0.168	Quad 3
VTYP	Others	−0.2128	−0.3342	Quad 3
ALC	No	−0.1481	−0.1204	Quad 3
WEATHER	Clear	−0.1151	−0.1518	Quad 3
NUMF	One	−0.0673	−0.0919	Quad 3
PCRASH	Negotiating a Curve	0.0357	−0.4004	Quad 4
ROLLOVER	No Rollover	0.1078	−0.1153	Quad 4
VPROFILE	Level	0.1582	−0.1592	Quad 4
PSL	60–70 mph	0.1891	−0.6684	Quad 4
VTYP	Station Wagon	0.2365	−0.4538	Quad 4
VTYP	2 Door Sedan	0.2407	−0.4265	Quad 4
VTYP	Motorcycle	0.257	−0.5601	Quad 4
DR_SF	None	0.3058	−0.5242	Quad 4
ROUTE	State Highway	0.3502	−0.5472	Quad 4
ROUTE	Other	0.3533	−0.4519	Quad 4
PCRASH	Others	0.358	−0.2835	Quad 4
DR_SF	Others	0.4212	−0.283	Quad 4
ROUTE	U S Highway	0.46	−0.8183	Quad 4
VPROFILE	Others	0.4708	−0.3235	Quad 4
VNUM	Five Lanes	0.657	−0.7136	Quad 4
PCRASH	Stopped in Roadway	0.682	−0.5417	Quad 4
PCRASH	Turning Related	0.6999	−0.5414	Quad 4
VTRAFWAY	Two Way Divided Median Barrier	0.7237	−0.8625	Quad 4
VTRAFWAY	Two Way Divided Unprotected Median	0.8501	−0.8115	Quad 4
VTRAFWAY	Others	0.857	−0.3103	Quad 4
VNUM	Four Lanes	0.8724	−0.4331	Quad 4
VNUM	Three Lanes	0.9107	−0.7679	Quad 4

railroad intersection, even the driver is distracted. This could be the primary reason to explain why these fatal crashes during clear weather and proper lighting environment are less likely to involve multiple fatalities. However, crashes in this cluster warrant a deeper analysis to develop recommendations for mitigation. The significant factors suggest that driver behavior and/or disregard for existing safety treatments may play a key role.

Cluster 5

The attributes of this cluster are the occurrence of no rollover, station wagon/2-door sedan/motorcycle as vehicle type, posted speed limit as 60–70 mph, route as a state highway, and negotiating a curve as a prior movement (see Fig. 4b). The components of this cluster show the combination of high-speed limit, negotiating a curve, and relatively small or light vehicles such as 2-door sedan/motorcycle could lead to fatal crashes. Especially for motorcycles, it provides almost no protection to the driver under a turning movement at high speed. Speed-related treatments and additional warning signage in advance of the HRGC may prove effective in these cases.

Cluster 6

The attributes of this cluster are multilane roadways, two-way divided roadways, and prior movement as stopped on-road or turning (see Fig. 4b). These attributes suggest a combination of speed, traffic volumes, and driver cognitive overload may be contributing factors; treatments which emphasize the importance of changing roadway conditions, and possibly the addition of active warning devices, such as gate arms and flashing lights, could be effective countermeasures for these types of HRGC crashes.

The four quadrants of the TCA plot reveal several patterns associated with fatal HRGC crashes:

- The occurrence of rollover fatal crashes with multiple fatalities is highly associated with inclement weather, comprised lighting conditions, and under the influenced driver.
- Fatal crashes at railroad crossings could also occur with normal drivers with no influence of alcohol or drugs in daylight and clear weather conditions. However, these fatal crashes mostly only involve one fatality – the driver himself. The current dataset does not provide adequate information about driver distraction. Availability of distraction information can provide additional explanation.
- The uphill or hillcrest always creates additional obstructions for railroad crossing of two-lane roadways, which could lead to fatal crashes. Without alerting visually of the incoming train, these crashes are often associated with fatalities since they have no time to respond after they approach the crossing.
- The speed limit and vehicle type are both associated with fatal crashes. Relatively lower speed (10–35 mph) and relatively large utility vehicles are associated with fatal crashes. Relatively large vehicles require extra time to cross the railroad crossing with a low-speed limit. Relatively higher speed (60–70 mph) and smaller/lighter vehicles are associated with fatal crashes, especially at negotiating a curve movement.
- Motorcycle – train crashes at high speed are associated with fatal crashes since motorcycle could only provide limited protection for drivers while crash occurs.

Conclusions

HRGCs are considered as the critical spatial junctions on roadway networks because the crashes at these locations can cause catastrophic incidents due to the tremendous momentum of trains. Safety at HRGCs is a high-priority concern among transportation agencies, and there is little research about pattern recognition of the key contributing factors of HRGC crashes. Therefore, to attain a complete understanding of HRGC crash patterns, this study performed TCA to identify key clusters of variable groups. In most cases, conventional safety analysis does explore interactions between variables or variable attributes. As crash data contains a significant number of categorical variables, a dimension reduction method such as TCA is beneficial in acquiring clusters from the complex datasets. This study applied TCA to determine the co-occurrence of variable attributes based on their relative presence in the two-dimensional space.

The study not only echoes the findings in previous studies but also provides promising insights into the fatal crashes at HRGC locations. The results indicate inclement weather, poor light conditions, influenced drivers, vehicle type, and functional class are associated with fatal crashes at grade railroad crossings. The findings also recognize patterns hidden in the dataset. For instance, the fatal crashes associated with clear weather, proper lighting conditions, and uninfluenced drivers were less likely to involve multiple fatalities. Multiple fatalities are mostly associated with a combination of influenced drivers, poor lighting conditions, and inclement weather. Many patterns are recognized through the TCA plot. The finding could greatly benefit the rail grade crossing fatal crash study and help transportation agencies to identify effective countermeasures to reduce the number of crashes and mitigate the severity of the crashes.

The current study is not without limitations. First, the current analysis focuses on attribute level TCA analysis. Row-level or individual crash level TCA analysis has not been performed in this analysis, which may provide additional insights. Second, the current analysis used the top 12 variables based on the information criteria using random forest. Additional variables can also be included. As TCA analysis is two-dimensional plot-based, the inclusion of a long list of variables will make the interpretation of the results challenging. Future studies can examine using additional variables. Third, the current analysis is limited to only axis 1 and axis 2 (both axes explain over 50% of variance). Additional combinations (e.g., axis 1-axis 3, axis 2-axis 3) can be performed to provide additional insights. Limitations of the current study can be improved in future studies.

Conflict of interest

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

References

Abraham, J., Datta, T.K., Datta, S., 1998. Driver behavior at rail-highway crossings. Transp. Res. Rec. 1648 (1), 28–34.

Allard, J., and V. Choulakian. TaxicabCA. https://cran.r-project.org/web/packages/TaxicabCA/TaxicabCA.pdf Accessed: July 31, 2020.

Choulakian, V., 2006. Taxicab correspondence analysis. Psychometrika 71, 1–13.

Schoppert, D.W., Hoyt, D.W., 1968. Factors Influencing Safety at Highway-Highway rail grade crossings. Highway Research Board, National Research Council, Washington, D.C..

Das, S., Sun, X., 2016. Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. IATSS Res. 39, 146–155.

Das, S., Jha, K., Fitzpatrick, K., Brewer, M., Shimu, T.H., 2019a. Pattern identification from older bicyclist fatal crashes. Transp. Res. Rec. 2673, 638–649.

Das, S., Minjares-Kyle, L., Wu, L., Henk, R.H., 2019b. Understanding crash potential associated with teen driving: survey analysis using multivariate graphical method. J. Saf. Res. 70, 213–222.

Das, S., Avelar, R., Dixon, K., Sun, X., 2018. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. Accid. Anal. Prev. 111, 43–55.

Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accid. Anal. Prev. 47, 119–127.

Fatality Analysis Reporting System (FARS). https://www-fars.nhtsa.dot.gov/Main/index.aspx Accessed: July 31, 2020.

Gabree, S., Chase, S., daSilva, M. Effect of Dynamic Envelope Pavement Markings on Vehicle Driver Behavior at a Highway-Rail Grade Crossing. In ASME/IEEE Joint Rail Conference, Colorado, USA, 2014.

Greenwell, B., Boehmke, B., Gray, B. VIP: Variable Importance Plots, 2019.

Hao, W., Kamga, C., 2017. Difference in rural and urban driver-injury severities in highway-rail grade crossing accidents. Int. J. Injury Control Safety Promot. 24 (2), 174–182.

Hao, W., Moghimi, B., Yang, X., Kamga, C., Wang, Y., Xiao, L., Liu, Z., 2017. Effects of foggy conditions on driver injury levels in U.S. highway-rail grade crossing accidents. Case Stud. Transp. Policy 5 (4), 627–633.

Hao, W., Kamga, C., Yang, X., Ma, J., Thorson, E., Zhong, M., Wu, C., 2016. Driver injury severity study for truck involved accidents at highway-Highway rail grade crossings in the United States. Transp. Res. Part F 43, 379–386.

Horton, S.M., daSilva, M.P., 2013. Success factors in the reduction of highway-rail grade crossing incidents. TR News 286, 38–42.

Jalayer, M., Pour-Rouholamin, M., Zhou, H., 2018. Wrong-way driving crashes: a multiple correspondence approach to identify contributing factors. Traffic Inj. Prev. 19, 35–41.

Keramati, A., Lu, P., Tolliver, D., Wang, X., 2020. Geometric effect analysis of highway-rail grade crossing safety performance. Accid. Anal. Prev. 138.

Khan, I.U., Lee, E., Khan, M.A., 2018. Developing a highway rail grade crossing accident probability prediction model: a North Dakota case study. Safety 4 (2), 22.

Landry, S., M. Jeon, P. Lautala, and D. Nelson. Getting Active With Passive Crossings: Investigating the Use of In-Vehicle Auditory Alerts for Highway-Highway rail grade crossings, South Carolina, USA, 2016.

Lee, D., Warner, J., Morgan, C. Discovering Crash Severity Factors of Grade Crossing with a Machine Learning Approach. In 2019 Joint Rail Conference. American Society of Mechanical Engineers Digital Collection, 2019.

Lerner, N.D., Tucker, L., 2002. Traffic-Control Devices for Passive Railroad-Highway Grade Crossings. National Academy Press, Washington, D.C..

Millegan, H., Yan, X., Richards, S., Han, L., 2009. Evaluation of effectiveness of stop-sign treatment at highway-railroad grade crossings. J. Transp. Safety Security 1 (1), 46–60.

Noyce, D.A., Fambro, D.B., 1998. Enhanced traffic control devices at passive highway-railroad grade crossings. Transp. Res. Rec. 1648 (1), 19–27.

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accid. Anal. Prev. 38 (2), 346–356.

Raub, R.A., 2009. Examination of highway-rail grade crossing collisions nationally from 1998 to 2007. Transp. Res. Rec. 2122 (1), 63–71.

Raub, R.A., 2006. Examination of highway-rail grade crossing collisions over 10 years in seven Midwestern states. ITE J. 76 (4), 21–26.

Ries, R., 2007. Highway-rail grade crossing safety. Mass Transit. 33 (5), 20–27.

Salim, A., 2018. Evaluation of Driver Behavior at Highway-Railroad Grade Crossings Based on Environmental Conditions and Driver Demographics. Michigan Technological University.

Salim, A., Jeon, M., Lautala, P., Nelson, D., 2018. Using naturalistic driving study data to investigate driver behavior at highway-highway rail grade crossings. ASME/IEEE Joint Rail Conference.

Yan, X., Richards, S., Su, X., 2010. Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-Highway rail grade crossings. Accid. Anal. Prev. 42 (1), 64–74.

Zhao, S., Iranitalab, A., Khattak, A.J., 2019. A clustering approach to injury severity in pedestrian-train crashes at highway-highway rail grade crossings. J. Transp. Safety Security 11 (3), 305–322.

Zheng, Z., Lu, P., Pan, D., 2019. Predicting highway-rail grade crossing collision risk by neural network systems. J. Transp. Eng., Part A: Systems 145 (8).

Zheng, Z., Lu, P., Tolliver, D., 2016. Decision tree approach to accident prediction for highway-highway rail grade crossings: empirical analysis. Transp. Res. Rec. 2545 (1), 115–122.

Further Reading

Das, S., Sun, X., 2015. Factor association with multiple correspondence analysis in vehicle-pedestrian crashes. Transp. Res. Rec. 2519, 95–103.