Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 48 (2020) 3486–3495

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019, Mumbai, 26–30 May 2019

Identification of Contributing Factors in Vehicle Pedestrian Crashes in Chennai using Multiple Correspondence Analysis

Priyadarshini Natarajan, Sathish Kumar Sivasankaran, Venkatesh Balasubramanian*

RBG Labs, Department of Engineering Design, Indian Institute of Technology Madras (IITM), Chennai - 600036, India

Abstract

In India, 10.5% of total accident death and injury of 2016 are related to pedestrians. Identification of the vehicle, roadways, environment or human factors involved in vehicle-pedestrian crashes has become an essential factor in implementing countermeasures. Multiple Correspondence Analysis (MCA), a categorical data analysis technique was used in this study on 2016 vehicle-pedestrian accidents from the Road Accident Data Management System (RADMS) database of Chennai city to detect patterns and associations that lead to accidents. This study identifies, two key clusters and six distant clusters of variables to have factors contributing to vehicle-pedestrian crashes. The associated variables and its categories found in the key clouds were collision type, cause of accidents, junction control, and pedestrian age. The association suggests that pedestrians in the age group of 25 to 34 are mostly injured at traffic signals where the cause of the accident is usually due to non-respect of the right way of rules. Also, driving against the flow of traffic, changing lane without due care and dangerous overtaking were associated with hitting an object. Other non-trivial variables identified were the time of day, season, availability of central divider, injury severity and speed limit. This technique provides data on the associated pattern and the significance of variables that most likely resulted in a pedestrian-vehicle crash. Based on the findings, appropriate countermeasures are also suggested that could potentially help transportation safety researches and policymakers towards developing strategies that prevent pedestrian accidents.

Keywords: Pedestrian Safety; Multiple Correspondence Analysis; Contributing Factors; Pattern Identification; Counter Measures

1. Introduction

Pedestrian accidents have been ranked fourth out of the total road accidents in terms of road users in 2016. According to the Ministry of Road Transport and Highways (MoRTH) reports of 2016, pedestrian death and injury amounted to 15,746 (10.5%) and 13,894 (9.5%) in 2015 out of total road traffic accidents. From the total pedestrian accidents in 2016, 1.7% (8,298) were considered to be the fault of the pedestrians. The pedestrian fault has also reported as 2% (3091) of total people killed and 1.5% (7,465) of total people injured in road traffic accidents. MoRTH

* Venkatesh Balasubramanian. Tel.: +91-(044)-2257-4117.
E-mail address: chanakya@iitm.ac.in

10.1016/j.trpro.2020.08.104

report considered 13 states in India out of which Tamil Nadu was ranked first in the share of total road accidents (14.9%), and second in terms of the number of people killed in road accidents (11.4%) MoRTH (2016). The number of pedestrian-vehicle collisions has been taken as one of the major concerns by the policymakers calling for identification and implementation of countermeasures to bring down road fatalities. According to NITI Aayog's three-year action agenda of India (2017 to 2020), the emphasis placed on data to monitor accidents and provide direction towards correction plans. RADMS database of Government of Tamil Nadu was designed to map road accidents across the state with a multi-department registry involving police, highways and local RTAs.

Recently, many research works have been carried out in understanding the contributing factors of vehicle-pedestrian crashes. Out of the total pedestrian accidents, the majority of them were reported in urban areas, since there is frequent interaction between pedestrians and vehicles. Liu and Tung (2014) considered parameters such as age, time of day, vehicle speed and pedestrian decision in their study and identified the time gap to be the most significant factor that could increase safety while crossing roads. Pedestrian's perception in identifying the approaching vehicles at nighttime was studied by Balasubramanian and Bhardwaj (2018). They suggested fixing an active light source between the headlights of four-wheelers as an effective countermeasure for faster vehicle identification in poor lighting conditions. A systematic review of environmental factors affecting pedestrian traffic accidents was performed by Moradi et al. (2016) based on studies from 1966 to 2015. Out of 2,828 articles, 15 articles met their inclusion criteria showing a significant correlation between road accidents and population density, number of junctions, student population, number of schools and pedestrian volume. A more integrated approach to evaluate all the variables that lead to a particular accident needs to be considered for the policymakers to evaluate the risk factors and provide sufficient countermeasures.

Due to the complex nature of the crash data including a vast quantity of information, data mining approaches are being applied by recent researchers to perform crash data analysis. Pande and Abdel-Aty (2009) applied this technique on crash reports from the state of Florida and discovered a significant correlation between an increase in traffic accidents during rainy times and roadways with straight sections and vertical curves. Also, poor illumination resulted in higher crash severity. Decision tree-based approaches towards identification of contributory factors in road traffic accidents and countermeasures are discussed by Chang and Chen (2005), Harb et al. (2009), Lee and Abdel-Aty (2005). A decision tree approach by Montella (2011) was performed on Italy crash data and they identified that the topmost contributory factors leading to accidents were associated with road signage and markings. The author suggests that measures to improve signage in the sites with road geometric deficiencies could significantly reduce accidents. Chi-squared automatic integration detector (CHAID), a decision tree-based method was employed by Hezaveh et al. (2018) on crash data from Tennessee to identify an association between pedestrian crash severity, road characteristics and environmental factors.

To understand the association between variables and handle missing data efficiently, Multiple Correspondence Analysis (MCA) emerged as a technique which allowed looking at relations or associations that existed between the categorical variables in a dataset. One of the first literature on the application of MCA for pedestrian-vehicle crashes was published by Fontaine and Gourlet (1997). Their study was based on the fatal accidents report by the French police. They analyzed the variables of pedestrians involving age, sex, driving under influence, mode of transport and movements and identified four groups that have lead to pedestrian accidents. Factor et al. (2010) published a social order of road accidents using MCA by associating variables relating to the social behavior of drivers and their engagement in accidents from Israeli road accident records. MCA was applied on crash data from Illinois and Alabama, collected over 15 years and factors such as driver/ road/ lighting conditions and driver age was found to be significant resulting in wrong-way driving crashes by Jalayer et al. (2018). Das et al. (2019) applied a similar MCA technique on 5 years of crash data from Louisiana and discussed 16 significant clusters that resulted in a crash and also suggested suitable countermeasures.

In this study, MCA was applied to vehicle-pedestrian crash data from the RADMS database in Chennai city in 2016 to address three main objectives. Firstly, to identify the key factors associated with the crash. Secondly, to identify the distant or non-trivial factors associated with the key factors leading to crash. Lastly, to provide countermeasures based on the MCA results to aid transportation safety researchers and workers in improving pedestrian safety.

2. Methodology

2.1. Vehicle Pedestrian Crash Data

Vehicle-Pedestrian crash data was compiled from Road Accident Data Management System (RADMS) which is a comprehensive road safety accident reporting application established by the Government of Tamil Nadu in 2009. RADMS is a software package developed to collect, analyze and map road accidents across the state with a multi department registry involving police, highways and local RTA departments. RADMS database consists of 105 fields to be filled by each department from which reports on driver, vehicle, road, enforcement, collision type, time period, alcohol usage, pedestrian, landmark and weather could be generated for analysis. In this study, pedestrian data in Chennai city using 24 variables that were relevant to this research were chosen. Each of these variables were of categorical type and was taken directly as observations for MCA. A summary of the selected variables and its frequency of occurrence is provided in Table 1 and Table 2.

Table 1. Summary of variables relating to Vehicle type, Driver and Pedestrian details in Chennai City

Variables and Categories	Frequency of Occurrence	Percentage	Variables and Categories	Frequency of Occurrence	Percentage
Vehicle Type			**Driver Age**		
Bus	133	3.89%	<18	75	2.20%
HGV	114	3.34%	>65	26	0.76%
Human power vehicle	3	0.09%	18-24	842	24.65%
LMV	1006	29.45%	25-34	795	23.27%
Motor cycle	1577	46.17%	35-44	447	13.09%
Unknown	583	17.07%	45-54	258	7.55%
License Type			55-64	111	3.25%
Full	2454	71.84%	**Pedestrian Age**		
No license	545	15.95%	<1818	198	5.80%
Unknown	417	12.21%	>6565	425	12.44%
Driver Gender			18-24	151	4.42%
Female	46	1.35%	25-34	734	21.49%
Male	2192	64.17%	35-44	431	12.62%
Unknown	1178	34.48%	45-54	495	14.49%
Pedestrian Gender			55-64	524	15.34%
Female	801	23.45%	Unknown	458	13.41%
Male	1726	50.53%			
Unknown	889	26.02%			

From the four zones in Chennai (north, south, west, and east) 3416 data points were reported in the database. Based on the preliminary analysis, the variables road condition, speed limit, traffic movement, number of lanes, road works, accident cause, weather condition, and junction control were highly skewed. For example, it could be seen that 99.85% of the accidents occurred on good road conditions, 98.33% of the accidents occurred on roads with speed limit of 40 km/hr, 77.46% of accidents occurred due to high speeds, 98.51% occurred on a two-way traffic movement, 83.75% on a single lane road, 98.01% with no road works, 98.59% on fine weather conditions and 74.3% of accidents were not at junctions. The non-skewed variables were accident severity, central divider, light conditions, season, accident zone, time, vehicle type, driver and pedestrian age. The number of categories in each variable is presented in Table 3. The missing values from the selected variables were replaced with the value ”unknown” as seen in light condition, road category, license type, driver gender, pedestrian gender, hit and run cases, vehicle type, driver age, and pedestrian age.

2.2. Multiple Correspondence Analysis

The methodology used in the study is an exploratory data analysis tool to analyze relationships and patterns between the categorical variables in a dataset. The observations in MCA are described by variables comprising of different levels or categories. MCA helps in understanding relationships between the variables instinctively using geometrical

Table 2. Summary of variables relating to Crash type, Environment and Road details in Chennai City.

Variables and Categories	Frequency of Occurrence	Percentage	Variables and Categories	Frequency of Occurrence	Percentage
Severity			**Speed Limit**		
Fatal/Grievous	2050	60.01%	30 Km/hr	3	0.09%
No/Simple	1366	39.99%	35 Km/hr	16	0.47%
Collision Type			40 Km/hr	3359	98.33%
Head on	284	8.31%	50 Km/hr	37	1.08%
Hit from rear	79	2.31%	60 Km/hr	1	0.03%
Hit from side	168	4.92%	**Traffic Movement**		
Hit object	46	1.35%	One-way	51	1.49%
Hit pedestrian	2216	64.87%	Two-way	3365	98.51%
Others	616	18.03%	**No. of Lanes**		
Skidding	7	0.20%	1	2861	83.75%
Central Divider			2	517	15.13%
No	2202	64.46%	Greater than 2	38	1.11%
Yes	1214	35.54%	**Road Works**		
Road Category			No	3348	98.01%
Highway	2344	68.62%	Yes	68	1.99%
Not highway	305	8.93%	**Accident Cause**		
Unknown	767	22.45%	Alcohol abuse	9	0.26%
Road Condition			Animal involved in accident	3	0.09%
Good	3411	99.85%	Changing lane without due care	65	1.90%
Poor	5	0.15%	Dangerous overtaking	70	2.05%
Light Condition			Driving against flow of traffic	56	1.64%
Darkness	391	11.45%	High speed	2646	77.46%
Daylight	1464	42.86%	Inattentive turn	30	0.88%
Street light	769	22.51%	Injured in accidents	79	2.31%
Unknown	792	23.19%	No details entered	42	1.23%
			Non-respect of rights of way rules	416	12.18%
Weather Condition			**Junction Control**		
Cloudy	35	1.02%	Flashing signal	3	0.09%
Fine	3368	98.59%	Give way sign	9	0.26%
Rainy	13	0.38%	No control	201	5.88%
Season			Not at junction	2538	74.30%
Autumn	879	25.73%	Police officer	22	0.64%
Spring	885	25.91%	Stop sign	25	0.73%
Summer	885	25.91%	Traffic signals	618	18.09%
Winter	767	22.45%	**Day Type**		
Chennai Zone			Weekday	2451	71.75%
East	608	17.80%	Weekend	965	28.25%
North	597	17.48%	**Time**		
South	1502	43.97%	Early Morning	335	9.81%
West	709	20.76%	Evening	722	21.14%
Hit and Run			Midnight	209	6.12%
No	2614	76.52%	Morning	705	20.64%
Unknown	288	8.43%	Night	831	24.33%
Yes	514	15.05%	Noon	614	17.97%

methods and represents each variable in a low dimension space. The most correlated variables are plotted as clusters together and the uncorrelated variables are plotted far from each other based on the correlation values. Similar to Principal Component Analysis (PCA), the Eigenvalues of calculated MCA describe the level of variables explained by each dimension. Large Eigenvalues map to a large total variation among the variables across its respective dimensions. The variables with the most explained variation in the data are plotted across dimension 1, followed by dimension 2 and so on. Mostly, the top two or three Eigenvalues explain most of the variance in the data, the dimensions higher than that are usually considered as redundant or noisy information.

Table 3. Number of Categories in each variable.

Variable	Number of Categories
Severity	2
Collision Type	7
Central Divider	2
Road Category	3
Road Condition	2
Light Condition	4
Speed Limit	5
Traffic Movement	2
No. of Lanes	3
Road Works	2
Accident Cause	10
Weather Condition	3
License Type	3
Driver Gender	3
Season	4
Zone	4
Time	6
Pedestrian Gender	3
Hit and Run	3
Junction Control	7
Vehicle Type	6
Driver Age	8
Pedestrian Age	8
Day Type	2

2.3. Initial Data Analysis using MCA

In this study, 24 categorical variables from the crash data were considered and the dimension of the dataset was 3416x24. Each crash report was represented as rows in the data matrix and is called transactions. The columns of the data matrix are the variables with their categorical values. The computations of MCA are performed using the FactoMineR package of R programming language. This package has the tools to compute, summarize, visualize and describe data of both quantitative and categorical variables based on the dataset and application. Since the Eigenvalues from MCA computation describe the level of variation explained by each dimension, the Eigenvalues and % variance explained for the top 10 dimensions are tabulated in Table 4. As the number of dimensions increases a continuous decrease in the Eigenvalues and % variance could be seen. The first and the second principal dimensions explain about 14.5% of the total variation and the greater dimensions do not explain more than 2.5%. Hence the results of MCA on crash data are discussed based on data from dimension 1 and 2.

Table 4. Eigen values and explained variance for first 10 dimensions.

Dimension	Eigen Values	% Variance Explained	Cumulative % Variance Explained
1	0.2583	7.9468	7.9468
2	0.2125	6.5369	14.4837
3	0.0789	2.4289	16.9127
4	0.0765	2.3541	19.2667
5	0.0747	2.2981	21.5648
6	0.0678	2.0863	23.6511
7	0.0620	1.9081	25.5592
8	0.0592	1.8230	27.3822
9	0.0583	1.7926	29.1748
10	0.0564	1.7344	30.9092

The variables with the highest contribution from MCA computation could be found using this package and this value provides valuable insight while interpretation. From Table 1, it could be seen that the frequency of occurrence of ”unknown” values under road category is 22.45%, light condition is 23.19%, license type is 12.21%, driver gender is 34.38%, pedestrian gender is 26.02%, hit and run is 8.43%, vehicle type is 17.07%, driver age is 25.23% and pedestrian age is 13.41%. Table 5 lists the highest contributing variables and its categories that are greater than 1% in both dimensions 1 and 2 excluding the unknown categories.

Table 5. Variable contribution in dimensions 1 and 2 greater than 1%.

Variables	Contribution%
No. of Lanes - 2	6.1557
Junction Control - Traffic signals	5.8720
Collision Type - Others	5.5520
Accident Cause - Non-respect of rights of way rules	4.1180
Pedestrian Age - 25 to 34 years	3.5939
Driver Gender - Male	1.5242
Road Category - Highway	1.5176
Junction Control - Not at junction	1.4335
Central Divider - Yes	1.3971
Chennai Zone - East	1.3579
Collision Type - Hit pedestrian	1.1182
No. of Lanes - 1	1.0732

It should be noted that the contribution values of each variable will depend on the number of categories and the contribution of each category as listed in Table 4 will depend on the number of accidents reported with the listed categories. These variables and their categories should be given high importance while deciding on the factors affecting pedestrian safety.

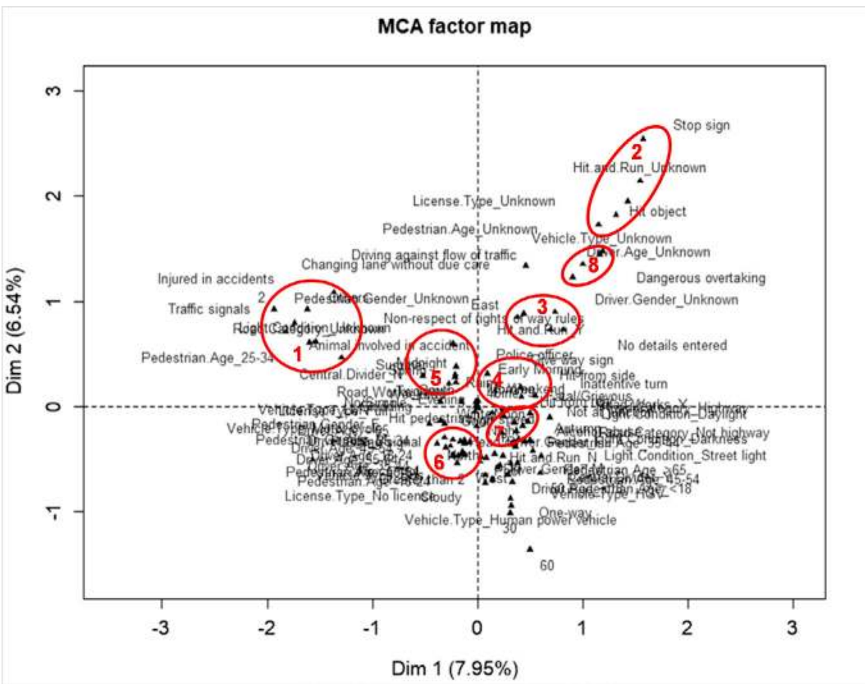

Fig. 1. Principal MCA plot between dimension 1 and dimension 2 for the variable categories.

3. Results and Discussion

The plots to interpret the results of MCA across dimensions 1 and 2 were generated using the ggplot2 package. The variable categories that are further away from the origin are the most discriminating variables and the ones clustered at the origin are probably less distinct. Based on the positioning of the variable clusters, different clouds were formed using the most discriminating variables and less distinct variables as shown in Fig. 1. The clouds formed with discriminating variables are shown as ellipses and less distinct variables are discussed in Fig 4 to Fig 9. These clouds form the category groups that most likely to have contributed to the crash. The clusters identified in the MCA plots from their distribution in plane 1 are discussed below explaining 14.5% of the variance in the dataset.

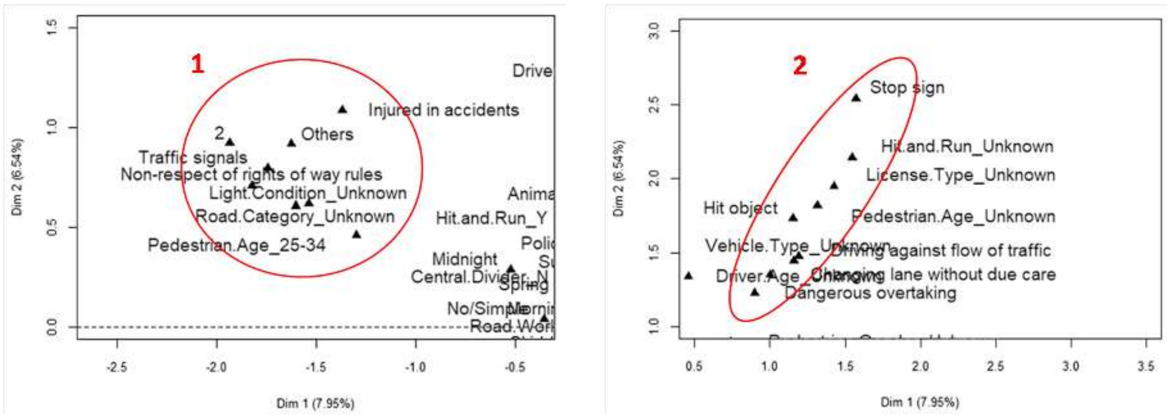

Fig. 2. (a) Cloud 1 of MCA factor map; (b) Cloud 2 of MCA factor map.

First cloud forms association between four factors collision type, pedestrian age, accident cause and junction control shown in Fig. 2a. The association suggests that pedestrians in the age group of 25 to 34 are mostly injured in accidents at traffic signals where the cause of the accident is usually due to non-respect of the right way of rules. Countermeasures such as guided pedestrian crossings, skywalk or subways could prevent accidents at the traffic junctions.

The second cloud forms an association between the cause of accidents and collision type shown in Fig. 2b. The factors identified in the cause of accident categories are driving against the flow of traffic, changing lane without due care and dangerous overtaking. These factors have resulted in a collision type of hitting an object. The categories in this cloud are distinctly spread across dimension 2 making the cause of the accident as its top contributing variable. The association is made completely between the vehicles on the road where colliding with an object has occurred mainly due to unsafe driving and overtaking.

Third cloud associates with three variables collision type, cause of accident and weather conditions shown in Fig. 3a. This cloud suggests that during rainy weather conditions accidents relating to inattentive turning lead to the collision of vehicles from the side. This association also suggests that signals at traffic junctions may not provide sufficient safety while making turns, especially during bad weather conditions. Increased safety interventions at such turns could prevent these types of collisions.

The fourth cloud forms the association between injury severity, time of day and collision type shown in Fig. 3b. The association suggests that there is a high risk of fatal/ grievous injury occurrence during poor light conditions since the time of day is evening and night. Also, the collision type associated with this cloud is getting hit from the rear. Naturally, night relates to dark conditions and hence countermeasures like necessary lighting could prevent these crashes.

Fifth cloud associates Hit and Run, time of day, season, central divider and junction control shown in Fig. 4a. From this cloud, an association that hit and run occurs during early mornings are formed. This could indirectly relate to driver fatigue or the confidence of the driver is not expecting pedestrians early morning on the roads. The occurrence of these crashes is also associated with summer and spring which is the first half of the year. Another interpretation of the cloud is that crashes are associated with the presence of central dividers, where the pedestrian could have been

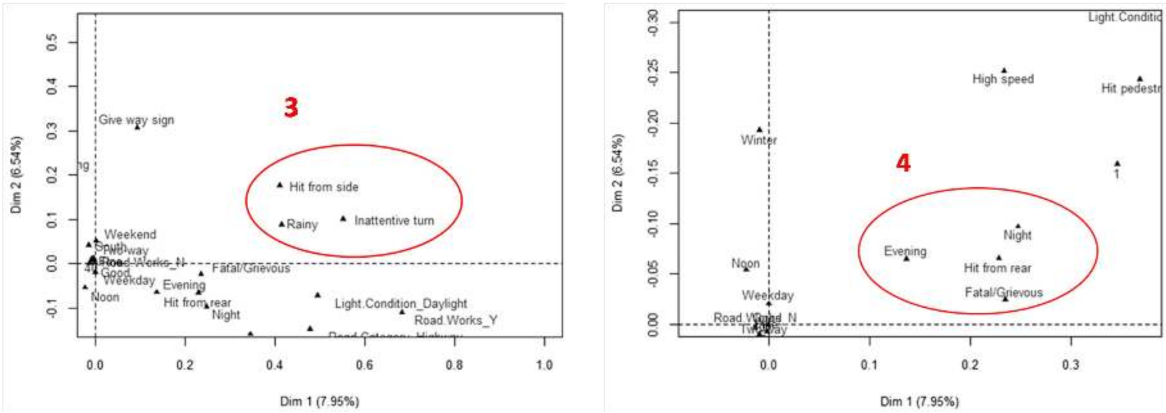

Fig. 3. (a) Cloud 3 of MCA factor map; (b) Cloud 4 of MCA factor map.

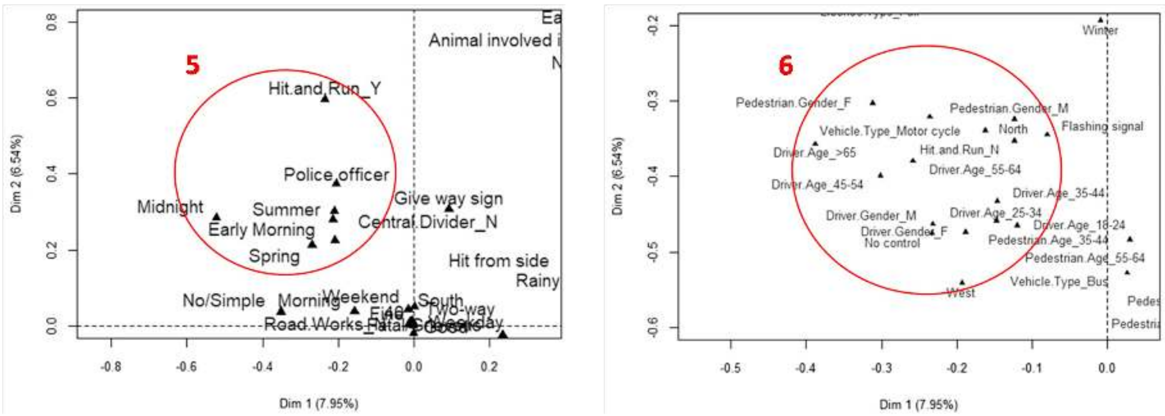

Fig. 4. (a) Cloud 5 of MCA factor map; (b) Cloud 6 of MCA factor map.

crossing at unmarked crossing zones. The presence of a police officer is also associated with this cloud. Improvement in speed control in early mornings and targeted pedestrian awareness and enforcement could help in reducing this type of crashes.

Sixth cloud associates all drivers and pedestrians involved in motorcycle accidents shown in Fig. 4b. Drivers of all age groups are associated with this cloud. Some other variable categories associated are flashing traffic signal and no junction control. Road with no controls will most likely contribute to accidents. This cloud stresses the importance of having a proper traffic signaling system and junction control to prevent accidents. The cloud also associates with the north and west zones of Chennai city. These locations would require increased attention and installment of speed barriers, traffic management system and enforcement could minimize the occurrence of these crashes.

Seventh cloud associates pedestrians of all ages, head-on collision type, bus and heavy goods vehicle type, roads with greater than 2 lanes and presence of a central divider shown in Fig. 5a. Pedestrian crashes and injuries are reported even in single-lane roads, hence it becomes progressively dangerous for pedestrians on roadways with more than 2 lanes. Large vehicles tend to have difficulties in maneuvering to prevent pedestrian collisions if it is sudden resulting in crashes. Hence, pedestrians must be enforced to use only the designated crossing areas that include traffic signal controls and the speed of the heavy vehicles should also be kept under the safety limits. Installation of warning systems, enforcement, traffic signage on speed controls could help prevent these accidents.

Eight cloud associates pedestrian crashes and high vehicle speeds where the speed limit is 35 km/hr in all road conditions shown in Fig. 5b. The crashes in this cloud are also associated with alcohol abuse. Poor road conditions and road works are also part of this cloud indicating that vehicles at high speeds in these road conditions will most

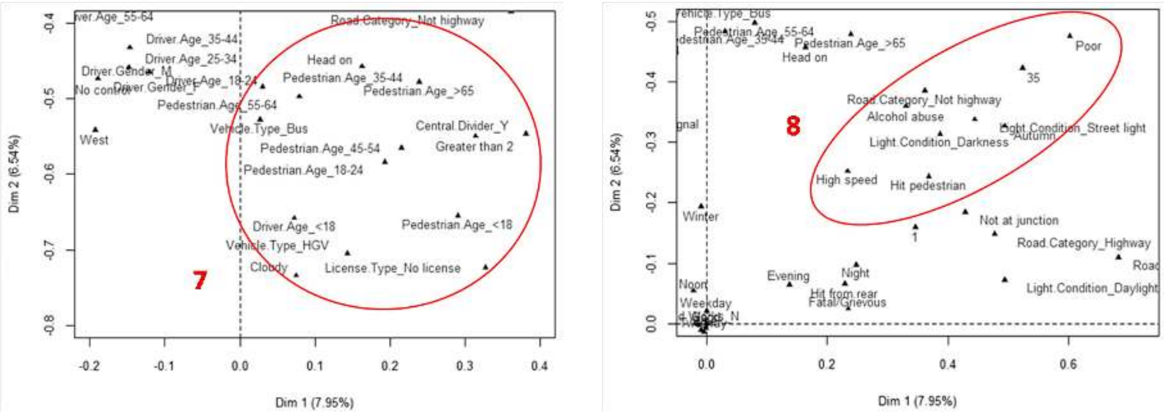

Fig. 5. (a) Cloud 7 of MCA factor map; (b) Cloud 8 of MCA factor map.

likely lead to crashes. Poor light conditions or darkness is also associated with this cloud. Hence required lighting, proper diversions during road works, strict enforcement against drunk and driving could help prevent these crashes.

This study implemented MCA on Pedestrian vehicle crash data from the RADMS database of 2016 in Chennai city. MCA could provide a valuable tool in identifying associated causes of a road accident from a large dataset of categorical variables and map them into lower dimension space.

Table 6. Significance of variables in dimension 1 and dimension 2.

Dimension 1			Dimension 2		
Variables	R^2	p-value	Variables	R^2	p-value
Collision Type	0.5929	<0.001	License Type	0.568695	<0.001
Road Category	0.7447	<0.001	Driver Gender	0.408233	<0.001
Light Condition	0.7133	<0.001	Hit and Run	0.535884	<0.001
No. of Lanes	0.6676	<0.001	Vehicle Type	0.459721	<0.001
Accident Cause	0.5587	<0.001	Driver Age	0.605638	<0.001
Junction Control	0.7076	<0.001	Pedestrian Age	0.670684	<0.001
Pedestrian Age	0.6134	<0.001	Pedestrian Gender	0.282709	<0.001
License Type	0.3364	<0.001	Collision Type	0.249672	<0.001
Vehicle Type	0.3176	<0.001	Accident Cause	0.244404	<0.001
Hit and Run	0.2217	<0.001	Zone	0.20725	<0.001
Pedestrian Gender	0.1932	<0.001	Junction Control	0.201141	<0.001
Driver Gender	0.1012	<0.001	Central Divider	0.164751	<0.001
Season	0.0936	<0.001	No. of Lanes	0.153489	<0.001
Severity	0.0824	<0.001	Light Condition	0.127403	<0.001
Central Divider	0.0801	<0.001	Road Category	0.110819	<0.001
Driver Age	0.0748	<0.001	Season	0.06784	<0.001
Time	0.0449	<0.001	Traffic Movement	0.011745	<0.001
Zone	0.0366	<0.001	Time	0.013815	<0.001
Road Works	0.0095	<0.001	Weather Condition	0.005608	<0.001
Speed Limit	0.0053	<0.001	Speed Limit	0.006827	<0.001
Traffic Movement	0.0015	0.024			

The significance of each variable in both dimensions 1 and 2 is tabulated in Table 6. A greater association of the dimensions with its corresponding variables would have a high R^2 value and vice versa. Based on the correlation results it could be seen that dimension 1 has the five most dominant variables as pedestrian age, license type, vehicle type, hit and run and pedestrian gender. Dimension 2 has the five most dominant variables as pedestrian gender, collision type, accident cause, zone, and junction control.

4. Conclusion and Future Work

Many policy derivations and interventions are formulated based on understanding the key factors that influence in an accident. A data-driven approach would provide one of the methods in evaluating countermeasures to reduce accidents. Many driver/pedestrian-related, environment-related, road-related factors are involved in accidents. Hence just looking at the effect of one specific variable towards the cause of road accidents will not be sufficient. From the results discussed it could be seen that MCA can be a valid tool for analyzing the complex nature of road accidents by associating all the necessary variables. Also, MCA provides an unsupervised tool and hence no prior information, hypothesis or outcome is required for validating the results. This method reduces the dimensions of the entire dataset into lower dimension space and hence the association could be viewed as 2D or 3D plots and interpreted even by a less experienced person in data analytics.

In this study, the significance of the variables in each dimension was discussed. However, the significance of the clouds itself was not computed. The key association findings from the MCA model could further be improved and validated objectively through clustering techniques like K-means or mapping algorithms. Through this approach, complex accident environments could be studied and valid countermeasures could be implemented by the transportation safety department in reducing accidents.

References

Road Accidents in India, 2016. Ministry of Road Transport & Highways (MoRTH), Transport Research Wing, Government of India, New Delhi.

Yung-Ching Liu, Ying-Chan Tung, 2014. Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. Journal of Safety Science 63, 77–82.

Venkatesh Balasubramanian, Rahul Bhardwaj, 2018. Pedestrians' perception and response towards vehicles during road-crossing at nighttime. Accident Analysis & Prevention 110, 128–135.

Ali Moradi, Hamid Soori, Amir Kavousi, Farshid Eshghabadi, Ensiyeh Jamshidi, 2016. Spatial Factors Affecting the Frequency of Pedestrian Traffic Crashes: A Systematic Review. Archives of Trauma Research, 5.4, e30796.

Anurag Pande, Mohamed Abdel-Aty, 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety Science, 47.1, 145–154.

Li-Yen Chang, Wen-Chieh Chen, 2005. Data Mining of Tree-Based Models to Analyze Freeway Accident Frequency. Journal of Safety Research, 36.4, 365–375.

Rami Harb, Xuedong Yan, Essam Radwan, Xiaogang Su, 2009. Exploring Pre-Crash Maneuvers Using Classification Trees and Random Forests. Accident Analysis and Prevention, 41.1, 98–107.

Chris Lee, Mohamed Abdel-Aty, 2005. Comprehensive Analysis of Vehicle Pedestrian Crashes at Intersections in Florida. Accident Analysis and Prevention, 37.4, 775–786.

Alfonso Montella, 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accident Analysis & Prevention, 43.4, 1451–1463.

Hezaveh, Amin Mohamadi, Azad, Mojdeh, Cherry, Christopher R, 2018. Pedestrian Crashes in Tennessee: A Data Mining Approach. Transportation Research Board 97th Annual Meeting.

Helene Fontaine and Yves Gourlet, 1997. Fatal pedestrian accidents in France: A typological analysis, Accident Analysis & Prevention 29.3, 303–312.

Roni Factor, Gad Yair, David Mahalel, 2010. Who by accident? The social morphology of car accidents, Risk Analysis, 30.9, 1411–1423.

Mohammad Jalayer, Mahdi Pour-Rouholamin, Huaguo Zhou, 2018. Wrong-way driving crashes: A multiple correspondence approach to identify contributing factors. Traffic Injury Prevention, 19.1, 35–41.

Subasish Das and Anandi Dutta and Raul Avelar and Karen Dixon and Xiaoduan Sun and Mohammad Jalayer, 2019. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. International Journal of Urban Sciences, 23.1, 30–48.