IJBPAS, September, 2022, 11(9): X-X

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

ANALYSIS OF ROAD ACCIDENTS IN KERALA, INDIA, USING DATA MINING TECHNIQUES

PREETHI R BHAT[1], STAVELIN ABHINANDITHE K[2*], MADHUB[3], BALASUBRAMANIAN S[4] AND SAHANA K S[5]

1: Division of medical Statistics, Faculty of Life Sciences, Department of water & health, JSS AHER

2: Assistant Professor, Division of medical Statistics, Faculty of Life Sciences, Department of water & health, JSS AHER. India

3: Associate Professor, Community medicine, JSS Medical College, JSS AHER. Mysore, Karnataka, India

4: Director Research, JSS AHER, Mysore, Karnataka, India

5: Division of medical Statistics, Faculty of Life Sciences, Department of water & health, JSS AHER

***Corresponding Author: Dr. Stavelin Abhinandithe K; E Mail: stavelin.k@gmail.com**

Received 16th Oct. 2021; Revised 20th Nov. 2021; Accepted 20th Jan. 2021; Available online 1st Sept. 2022

https://doi.org/10.31032/IJBPAS/2022/11.9.6263

ABSTRACT

Among the yearly frequent incidences, road accidents, stands out as a major reason behind mortality across India, as a consequence of increased inhabitants and rise in the population of vehicles in our society. In our study, we made an attempt to analyse the strongly affected accident areas in the south western state, Kerala and hence discover the key factors responsible for it using the data mining techniques. In this study, the data were initially tested for heterogeneity using a two-step cluster analysis to evaluate the accident severity from 2007-2017. Later, correlation data was subsequently reduced to a smaller number of factors by applying the principal component analysis, thus identifying the major influencing factors of road crashes with massive causalities. Taking deaths and injuries into consideration, the cluster analysis clearly explained that the districts of South Kerala and Central Kerala were more affected than those of North Kerala. Principal component analysis was carried out for the entire dataset, due to the violation of sample adequacy by each cluster. Results of the analysis made it evident that, exceeding speed of the vehicles have high influence on the highway road crashes. Thus, the multivariate techniques adopted served useful in classifying high and comparatively low intensity areas and predicting the contributing factors that need to be focused to reduce accident impact on population health and in building of sophisticated public administration.

Keywords: Data mining, principal component analysis, road accidents, two-step cluster analysis

1. INTRODUCTION

Improving economic conditions and rising l iving standards in Kerala has led to the usage of multiple vehicles, which in turn has increased the number of road accidents, one of the global issues which calls for serious handling. Accidents occur at different locations with different incidents, making it hard to assess which places have high degree of road accidents. Public and law enforcement need knowledge about places that are more prone to accidents **[1-4]**. A road accident can be portrayed as,' An incident that happens on a street or road that is available to open traffic; bringing about at least one individuals being harmed or slaughtered, including at any rate a single vehicle in motion. This includes high human suffering and socio-economic costs of premature deaths, accidents, productivity loss and so on **[5]**. In Kerala, only the bulk of passenger and freight movements have to bear on road transport. It is now the most important mode of transport for this purpose, but ironically, road accident deaths have now been characterized in this state as a hidden epidemic affecting all sectors of the society. Of all recorded accidents, 13% are fatal and a significant number of fatal accidents occur eachyear, in which a large number of road users in each accident lose their lives **[6]**. Road Traffic Injuries (RTI) was positioned fourth among the world's driving reasons for death during 2008, with almost 13, 00,000 dying annually. In India, road traffic wounds are the 6th primary reason for demise according to World Health Organization (WHO). The accidents counts expanded 4.4 times between 1970 and 2011, followed by a 9.8-fold rise in deaths and a 7.3-fold increment in the quantity of people being harmed **[5]**. In total, 15525 fatal accidents were recorded in Kerala over the period 2010-2013 claiming 16639 lives and injuring at least 165115 individuals. If no necessary step is taken, by 2025 road traffic accidents in India are expected to kill people, exceeding 0.25 million annually **[6]**.

Data mining gives place for research in this area by providing different classification, clustering and analysis methods. One of the emerging trends may be the application of appropriate data mining techniques to evaluate road safety **[3]**. Cluster analysis (CA) is an unsupervised learning technique and a statistical approach, which was used to categorize different districts of Kerala into groups with similar patterns of deaths and injuries by minimizing the variance or spread across the range of interest variables within clusters and maximizing that between cluster **[7, 8]**. Further, road accidents are events that are unpredictable and hence, we are required to examine the factors (usually discrete variables), that influence the accident. The principal component analysis is an important data mining tool, which helps to classify different types of vehicles and roads with less detailed indicators. Thus, by understanding the relationship between accidents and factors that influence them, traffic engineers may define the hazardous accident zones and provide better facilities for reducing the impact of such factors on accidents **[3, 9, 10]**.

2. MATERIALS AND METHODS

2.1. Study area: Existing within the northern latitudes 8°18' and 12°48' and eastern longitudes 74°52' and 77°22',the total area covered by the state is 38,863 km2 with a total population 33,387,677 as per 2011. The state has been subdivided into 14 districts, which are classified into north, south and central Kerala. Roads include 1,524 km of National Highways (NH), 4,341.6 km of State Highways (SH) and 18,900 km of district roads **[11]**.

2.2. Data collection: 11 years district-wise data for the analysis was available in the official website of Kerala police, keralapolice.gov.in.

2.3. Cluster and Principal component analysis: To begin with this retrospective study, eleven years average is computed for the number of deaths, injuries, cases due to each vehicle and road types for each Kerala district, as a part of the better organization of the dataset. We have implemented two-step cluster analysis, as it provides flexibility to automatically pick the number of homogeneous clusters of districts based on average number of deaths and injuries. This procedure uses a likelihood distance measure, which is a probability based distance, with the assumption that the parameters are independent and continuous variables are normally distributed. For two clusters i and j, distance between them is associated with the reduction in log-likelihood on their combination into a single cluster, whose distance is given by,

$$d(i, j) = \xi_i + \xi_j - \xi_{<i,j>} \qquad (1)$$

where,

$$\xi_S = -N_s \sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) \qquad (2)$$

with, <i, j> index that indicates the formation of cluster with the combination of i and j clusters. K^AandN_s are the total number of continuous factors and data records in s cluster. $\hat{\sigma}_k^2$and$\hat{\sigma}_{sk}^2$ are the variance estimates (dispersion) for the continuous factor k, for the complete dataset, in cluster j.

Two-step clustering commences with the building up of Cluster Features (CF) Tree, by accommodating the first region at its root, in a leaf node that includes the details about the persons killed and injured. The consecutive district is then attached to the current node or results in new node formation, in accordance with its closeness to the existing nodes and resemblance criterion. A node that carries multiple districts constitutes synopsis of people dying and getting injured, about those districts. The CF tree leaf nodes are further clustered utilizing an agglomerative clustering algorithm. It is a bottom up approach, with no requirement for the number of clusters to be specified beforehand. With the consideration of every data as a single cluster, pairs of clusters are then agglomerated successively, which is continued until all the clusters are combined into one, containing the entire data A wide range of solutions can be generated using agglomerative clustering. The finest cluster number is evaluated by comparing these cluster outcomes employing the Schwarz Bayesian Criterion (BIC) or the Akaike Information Criterion (AIC).

BIC (J) = $-2\sum_{j=1}^{J}\xi_j + m_j \log(N)$ (3)

AIC (J) = $-2\sum_{j=1}^{J}\xi_j + 2m_j$ (4)

wherem_j = J($2K^A$). Thus, the CF tree presents a condensed summary of the dataset **[7, 8]**.

Principal component analysis (PCA) involves reducing the dimensionality of the 13 accident related factors (vehicle and road types), increasing interpretability with the loss of information being minimum and also determining new parameters that are linear functions of those in the original dataset, which sequentially maximize variance and are not associated with each other. The basic idea behind PCA being an exploratory analysis method is the inclusion of a dataset containing accident cases related to p entities (road types and accidents attributable to specific vehicles) for each of n districts. Such data values describe vectors $x_1, \ldots, x_p$ or, equivalently, a n x p matrix X, whosej^{th} column is the observation vector $x_j$on the j^{th} variable equivalently. We are interested in finding that linear combination of X matrix columns having highest variance. Such a combination can be defined as,

$\sum_{j=1}^{p} a_j x_j$ = Xa (5)

where a represents the vector of constants, $a_1, a_2, \ldots, a_p$ with var (Xa) = a'Sa being the combination variance and S is the corresponding matrix of sample covariance. It is important to note that for a covariance matrix S, the eigenvector (unit-norm) should be a and eigenvalue corresponding to it is given by λ. As the eigenvalues can be understood as the variances of the linear combinations described by the corresponding eigenvector a,

var (Xa) = a'Sa= λa'a= λ (6)

we have to look for the largest λ_1 that corresponds to a_1. In addition to the orthogonality restrictions of various coefficient vectors, a Lagrange multipliers approach is useful in demonstrating that the complete set of S eigenvectors are the solutions for the problems of obtaining new p linear combinations,

$X a_k = \sum_{j=1}^{p} a_{jk} x_j$ (7)

which maximizes variances successively, subjecting to no correlation with linear combinations considered previously. The eigenvector elements, a_k are usually referred to as PC loadings, whereas PC scores, $X a_k$ are the linear combination elements, as they imply to those values, which are the entity scores on the given PC. The variables under analysis are standardized generally in order to overcome the problem of variables with different measurement units, if present any or otherwise also. Standardization of each data value x_{ij} is done by centering and division by standard deviation $s_j$corresponding to n observations of $j$variable. For standardized dataset, the matrix of covariance is nearly same as the matrix of correlation R of dataset, originally taken. PCA on such a dataset is called as correlation matrix PCA. The linear combinations $Z a_k = \sum_{j=1}^{p} a_{jk} z_j$of the standardized variables$z_1, z_2, \ldots, z_p$, having highest variance and no correlation with each other are defined by the correlation matrix's eigenvectors. The interpretation of the loadings of the component, which is the eigenvector times the square root of the eigenvalue, is the association of each factor with the principal component. Each loading square shows the variance proportion described by the PC. The cumulative addition of the squared loadings, as we go down the components, is found to sum up to 1 or 100%, which is often referred to as communality and for each component it is equal to total variance. Adding the components' loading squares across them provides estimated communality for each factor and similarly on adding down the factors (rows) provides each component's eigenvalue. In addition to this the concept of rotation can greatly result in simpler understanding and for q dimensional space, the variance sum of their rotated components is same as those of unrotated, therefore no variance is lost. Thus for the even distribution of q components after rotation, there is a loss of maximization of unrotated PCs in succession **[10]**

In SPSS, adequate sample size is the basic criteria to be tested and satisfied before factoring is done, for which, Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity are the most appropriate. Absence of high multi-collinearity is an important assumption which is required to be satisfied. Apart from simply looking for correlation between the factors, PCA looks for the correlation between their variances. % of variance for each component can be computed by dividing the corresponding eigenvalue by its total. It extracts components equal to that of the factors, but only those with eigenvalue≥1 is said to have a noticeable effect on the factor, a rule given by Kaiser. The eigenvalue would be the degree to a component explains the variance between items. Sum of the eigenvalues gives the total number of factors included. In addition to this, a scree plot, plot of component number versus eigenvalue, can be helpful in graphical interpretation. The analysis also gives the variation in percentage well-defined by each component, total variance explained along with variance extracted in each factor. Further, a vehicle or road type with high factor loadings for a particular component belongs to that component. Loading plots depicts the strength of characteristic influence on the principal components, with acute, obtuse, right angles between characteristics indicating positive, negative and no or weak relation respectively. If any difficulty arises in identifying the highest factor loading, either, oblique rotation (for correlated components) or orthogonal rotation (for uncorrelated components) is the best solution. Thus, the accidental factors are ready for summarization after the feature extraction **[9, 10]**.

2.4. Statistical methods: Data mining techniques are employed for condensing data by finding unexpected relationships and patterns in the accidental dataset recorded for over eleven years. The level of significance value was fixed to 0.05. Excel 2013 and SPSS Statistics 22 (IBM corp.) were incorporated for the study analysis.

3. RESULTS AND DISCUSSION

3.1. Descriptive analysis

A greater exponential trend for number of cases and people injured is seen during the period 2007-2017 as compared to deaths, which appears to be consistent. Highest cases (39917) and injuries (48246) were reported in the year 2007, but more people died in 2016 (4287). This is a clear indication that accident cases are becoming more fatal with time **(Figure 1)**.

Covering the maximum portion, two-wheelers (47%) serve as the most important factor **(Figure 2)**.

District roads are found to have greater affinity (20948) towards road crashes **(Figure 3)**.

3.2. Cluster analysis

Kolmogorov-Smirnov and Shapiro-Wilk tests are insignificant (p-value>0.05), implying that both the variables are normally distributed **(Table 1)**.

In SPSS, the two-step cluster analysis auto-generates the most appropriate number of clusters having the least BIC value. The Silhouette coefficient is close to 1 (0.8) depicting that appreciable homogeneous clusters were created **(Figure 4, Table 2)**.

Districts were classified into two homogeneous groups with respect to average deaths and average injuries, with first cluster having higher dimension (9 areas, 53%). On an average number of deaths and injuries in the second cluster were more than twice of those in the first one. Majority of the South Kerala and Central Kerala districts fall under the highly accident prone zone when compared with those of North Kerala. Further, it can be observed that Trivandrum and Eranakulam being the most developed districts in Kerala, their urban areas belong to the comparatively lower risk group, whereas rural areas belong to the latter. There is a possibility that improper infrastructure along with insufficient maintenance and lack of supervision can be one of the significant reasons behind such a disaster **(Table 2)**.

3.3. Principal component analysis: Presence of high multi-collinearity can be easily detected by observing the determinant of the correlation matrix of the various factors. SPSS computes the determinant as a part of the analysis and the value was found to be 1.55E-013, whereas acceptance threshold value is 0.1E-4. To remove the excess multi-collinearity, highly correlated (coefficient≥0.9) car, mini, two-wheelers, others and unknown vehicles cases' were combined to form a new variable 'others'. District road cases were eliminated for this analysis. The determinant increased to 1.394E-5, thus satisfying the criteria. A new set of 8 factors were subjected to dimensionality reduction **(Table 3)**.

Fundamental criteria to carry out this analysis is the appropriate sample size for the study. Individually, both the clusters are found to not have the necessary number of samples. The combination of the 2 clusters assured to yield relatively compact patterns of correlation so that PCA results in distinct and reliable factors (KMO value=0.6). A KMO value≥0.6 is acceptable. Adding to this, a significant value (p-value=0.001) from the Bartlett's test of sphericity conveys that the causes are related in some way **(Table 4)**.

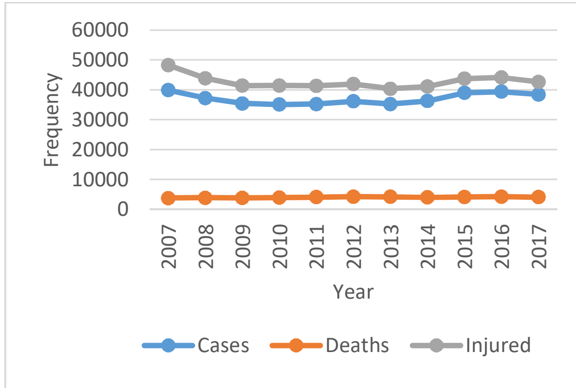
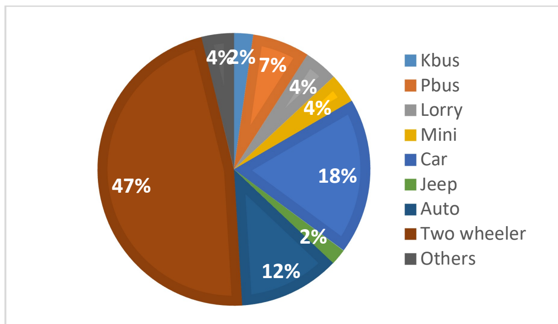


Figure 1: Year-wise cases, deaths and injuries report



Kbus-KSRTC bus, Pbus-Private bus
Figure 2: Distribution of cases based on vehicle type

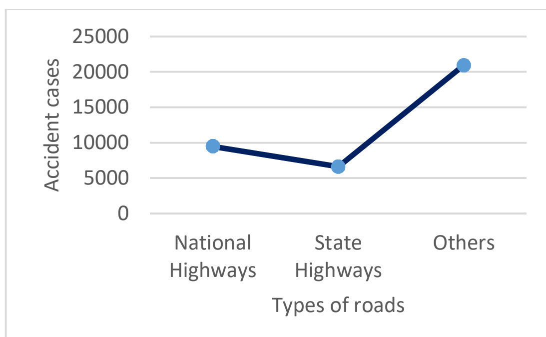


Figure 3: Spread of accident occurrence on distinct roads

Table 1: Tests for normality

Variables	p-value	
	Kolmogorov-Smirnov	*Shapiro-Wilk*
Average deaths	0.085	0.124
Average injured	0.2	0.909

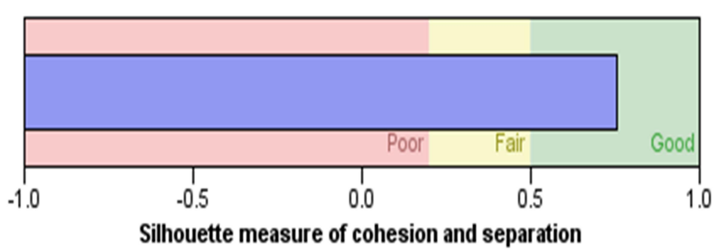


Figure 4: Cluster quality

Table 2: Summary of clusters

Cluster number	Districts	Cluster size (No. %)	Average deaths	Average injuries
1	Trivandrum city, Pathanamthitta, Eranakulam city, Idukki, Kozhikkode city, Kozhikkode rural, Wayanadu, Kannur, Kasarakodu	9 (53)	140	1709
2	Trivandrum rural, Kollam, Alappuzha, Kottayam, Eranakulam rural, Trissur, Palakkad, Malappuram	8 (47)	352	3422

Table 3: Sample adequacy test

Cluster	KMO value	Bartlett's test (p-value)
Cluster 1	not displayed	not displayed
Cluster 2	not displayed	not displayed
Combination	0.6	0.001

Table 4: Correlation Matrix

Initial matrix (Determinant – 1.394E-5)								
Factors	*Kbus*	*Pbus*	*Lorry*	*Jeep*	*Auto*	*Others*	*NH*	*SH*
Kbus	1	-0.036	0.199	-0.306	0.494*	0.737**	0.551*	-0.005
Pbus	-0.036	1	0.852**	0.193	0.683**	0.549*	0.467*	0.694**
Lorry	0.199	0.852**	1	0.116	0.837**	0.758**	0.66**	0.807**
Jeep	-0.306	0.193	0.116	1	0.278	-0.22	-0.172	0.32
Auto	0.494*	0.683**	0.837**	0.278	1	0.816**	0.604**	0.763**
Others	0.737**	0.549*	0.758**	-0.22	0.816**	1	0.78**	0.549*
NH	0.551*	0.467*	0.66*	-0.172	0.604*	0.78*	1	0.243
SH	-0.005	0.694**	0.807**	0.32	0.763**	0.549*	0.243	1
Reproduced matrix								
Factors	*Kbus*	*Pbus*	*Lorry*	*Jeep*	*Auto*	*Others*	*NH*	*SH*
Kbus	0.792	0.077	0.29	-0.53	0.372	0.714	0.661	-0.012
Pbus	0.077	0.77	0.816	0.354	0.773	0.567	0.435	0.789
Lorry	0.29	0.816	0.921	0.226	0.898	0.775	0.625	0.812
Jeep	-0.53	0.354	0.226	0.571	0.144	-0.213	-0.245	0.429
Auto	0.372	0.773	0.898	0.144	0.887	0.816	0.669	0.758
Others	0.714	0.567	0.775	-0.213	0.816	0.968	0.838	0.505
NH	0.661	0.435	0.625	-0.245	0.669	0.838	0.732	0.374
SH	-0.012	0.789	0.812	0.429	0.758	0.505	0.374	0.819
Kbus: KSRTC bus, Pbus: Private bus, NH: National Highways, SH: State Highways								
*p-value significant at 0.05								
**p-value significant at 0.01								

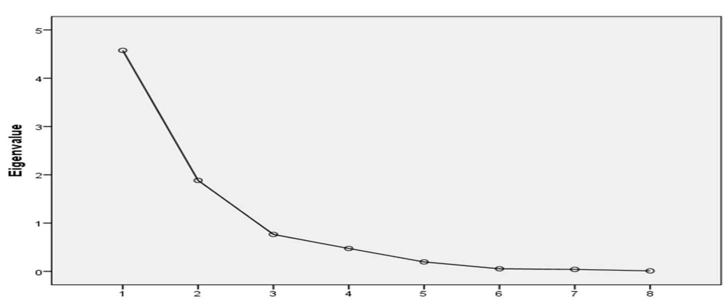


Figure 5: Scree plot

Table 5: Total variance explained

Component	Before rotation			After rotation		
	Eigenvalue	*% of variance*	*Cumulative %*	*Eigenvalue*	*% of variance*	*Cumulative %*
1	**4.576**	**57.201**	**57.201**	**4.038**	**50.475**	**50.475**
2	**1.884**	**23.546**	**80.746**	**2.422**	**30.272**	**80.746**
Correlation coefficient (r) - -0.133						

Table 4, 5 and Figure 5 can be summarized as follows. Although PCA produces results for all the 8 factors, only two components are noticed to be useful in explaining the relationship through the entire set (eigenvalue>1) and other components can be ignored. This can be visualized using the scree plot. There is a slightly big drop from component 2 to 3 and for the suppressed eigenvalues the rate of change is minimal as we move across. Overall 80.746% variation in the data was described by PCA, considerably a good proportion for obtaining the information held in the original variables. Together with the matrix of correlation coefficients (r), it displays the reproduced matrix which defines not only the communalities, that is, extracted proportion of variation within each factor over the various districts (orange coloured cells) but also the variation in them due to the other factors, shown in the green and blue coloured cells, representing which variable is more likely to belong to component 1 and 2 respectively. Communalities, greater than 0.5 were considered as considerably good extraction. This matrix forms the basis for the choice between the requirement of rotated and unrotated components and also the rotation type. Orthogonal (varimax) was preferred over oblique rotation as a weak relation existed between the 2 components (|r|<0.3) and outcomes were found to be valid on comparing with the reproduced matrix. After rotation first and second component conveyed 50.475% and 30.272% variation respectively, with an increase in the degree of variance explained by second component (2.422) and decrease in the first (4.038), thereby resulting in no change in the total eigenvalue and hence the cumulative %.

Table 6: Component matrix

Factors	Before rotation		After rotation	
	Component		Component	
	1	*2*	*1*	*2*
Kbus	**.469**	**-.757**		**0.886**
Pbus	**.788**	**.386**	**0.877**	
Lorry	**.939**	**.199**	**0.929**	
Jeep	**.082**	**.751**		**-0.635**
Auto	**.938**	**.089**	**0.879**	
Others	**.907**	**-.382**		**0.747**
NH	**.752**	**-.408**		**0.701**
SH	**.762**	**.488**	**0.9**	
Kbus: KSRTC bus, Pbus: Private bus, NH: National Highways, SH: State Highways				

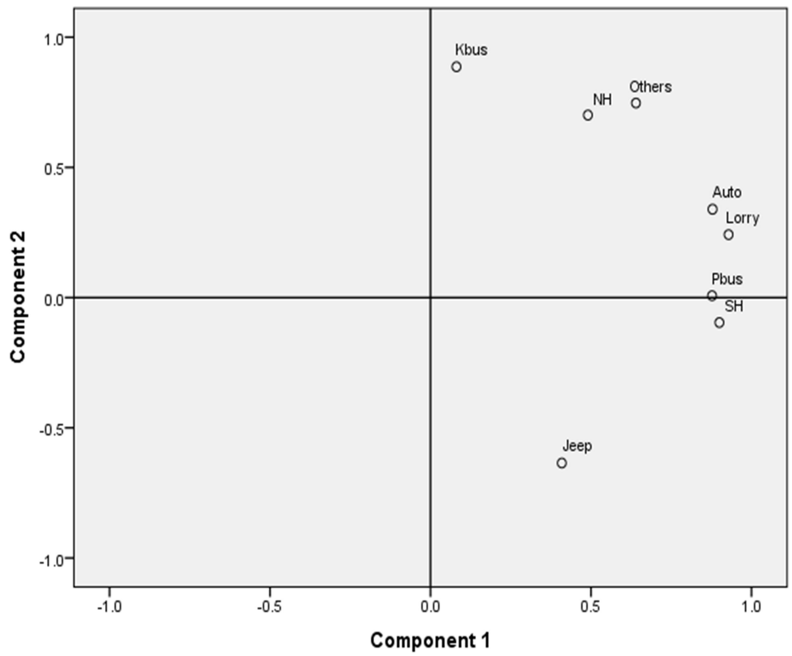


Figure 6: Component plot in rotated space

Table 7: Matrix of correlation

Variables	Vehicles		
	Mini	***Car***	***Two-wheelers***
Other roads	**0.9***	**0.925***	**0.945***
***p-value significant at 0.01**			

Table 6 is the matrix before and after rotation, with rotated matrix, exhibiting strong correlation between each accidental factor and the corresponding component in which it is included. |r| values are seen to vary from 0.6 to 0.9. Values less than 0.5 were supressed. Majority of the road accidents on the state highways were because of collision of private buses, Lorries and autos. In the second component, jeep has a negative coefficient, indicating opposite relation with the other factors present in the same component. Crashes in National Highways are mainly due to KSRTC buses and other vehicles. There are chances that no accident involving jeeps takes place in NH as a result of collision with Kbus, cars, two-wheelers, mini and other vehicles. This can be better understood by the Fig, 6. Variables having small angles between them are strongly positively associated, with large diverging angles showing negative and 90 degree depicting nearly no or weak association. Highways include the well-constructed roads, mixed traffic areas with the movement of heavy vehicles being high. With this, over speeding of the vehicles is the prominent contributor to the road crashes. In the previous analysis in the same study we observed more NH cases than SH cases. Presence of speed breakers

in SH, which are rare in NH has proved helpful in decreasing the number. Even though District roads were excluded, their strong associations (|r|≥0.9) and high significance (p-value<0.01) with cars, two-wheelers and mini vehicles as seen in **Table 7**, and the maximum accidents recorded forces us to explore the reasons. Traffic inconsistencies and poor road frameworks can be thought of as the most relevant factors. Further, the study involved some accident cases due to unknown vehicles, which should be accounted for.

3.5 DISCUSSION

In our study the results of PCA were pertaining to the accident cases on highways. Owing to the compound flow model of road traffic, that is, the involvement of miscellaneous traffic together with the pedestrians, the problem of accidents has become highly serious in highway transport. In India, road traffic is increasing at a rate of 7 to 10 percent per year while the growth of the automobile population is in the order of 12 percent per year. Traffic service is special on a two-lane, two-way highway. Direction changes and overtaking in the opposite lane are possible only in the face of upcoming traffic. The demand for overtaking grows rapidly as the amount of traffic rises, while the chance of moving through the opposing lane decrease as volume increases. Thus, the motion of vehicles in one direction is influenced by the motion in other direction. When mixed traffic flow is considered, the issue becomes more severe when the speed difference between different vehicle categories is very important. The constitution of traffic stream, split of directions and the presence of vehicles moving with slow speed in the stream are essential conditions that creates stronger impacts on the a two-lane highway capacity. It is very essential for the traffic analysts on multi-lane highways to have familiarity with the various traffic characteristics as they are critical for evaluating traffic efficiency, examining road safety, setting suitable devices for traffic control, speed limits, and designing simulation programs etc. Lane location is observed as one of the most significant factors influencing the efficiency and characteristics of traffic on multi-lane highways **[12]**.

4. CONCLUSION

The main interest of this study was to determine the high accident prone areas and their several influential factors using the well-known data mining techniques as it has turned into a disaster showing adverse effects all over the Kerala state. The analysis gave satisfactory and reliable results. Rash driving was found to be the most significant factor. Although the road standards, construction and maintenance are the responsibilities of the road safety committees, we as individuals should also take initiative to keep ourselves and others safe. Travelling with the safety measures, following the traffic rules and regulations with normal driving speed can reduce the cases to a greater extent. Further the government should also take immediate actions and educate people about accident severity. In this way Kerala can soon change into a nearly accident free zone.

4.1. Acknowledgement

The authors sincerely thank Head of the Department, Department of Water and Health, JSS Academy of Higher Education and Research, Mysore, for having granted the use of SPSS software for performing the statistical analysis.

5. REFERENCES

[1] Gade D. ICT based Smart Traffic Management System "iSMART" for Smart Cities. International Journal of Recent Technology and Engineering. 2019; 8(3): 920-928.

[2] Budiawan W., Purwanggono, B. Clustering analysis of traffic accident in Semarang City. In E3S Web of Conferences (Vol. 73, p. 12001). EDP Sciences. 2018; 1-4.

[3] Mamidisetti G., Gowtham N., Makala R. Web data mining framework for accidents data. International Journal of Recent Technology and Engineering. 2019; 7(5S4): 49-51.

[4] Chen T., Zhang C., Xu L. Factor analysis of fatal road traffic crashes with massive casualties in China. Advances in mechanical engineering. 2016; 8(4): 1-11.

[5] Ruikar M. National statistics of road traffic accidents in India. Journal of Orthopedics, Traumatology and Rehabilitation. 2013; 6(1): 1-6.

[6] Vigneshkumar C, Arichandran R. Fatal road accidents in Kerala (India): characteristics, causes and remedial measures. International Journal of Research. 2015; 4(4): 4-5.

[7] Constantinescu Z., Marinoiu C., Vladoiu M. Driving style analysis using data mining techniques. International Journal of Computers Communications & Control. 2010; 5(5): 654-663.

[8] Ma J., Kockelman K M. Crash modeling using clustered data from Washington State: Prediction of optimal speed limits. In Proceedings of the IEEE Intelligent Transportation Systems Conference. 2006.

[9] Maruyama Y., Kuniyuki H et al. Analysis on characteristics of traffic accidents in Nagano (second report). International journal of automotive engineering. 2019; 10(2): 219-225.

[10] Jolliffe I T., Cadima J. Principal component analysis: a review and recent developments. Philosophical

Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016; 374(2065): 1-16.

[11] https://en.wikipedia.org/wiki/Kerala

[12] Wadhwa A R. A study of flow characteristics of National Highway. International Journal of Innovations in Engineering Research and Technology. 2017; 4(4): 62-67.