Journal on Transportation System and Engineering
ISSN: 3059-9865(Print) 3059-9873(Online)
Vol. 1. Issue 2 (December) 2025: 50-62
DOI: https://doi.org/10.3126/jotse.v1i2.87743

Research Article
Received Date: May 29, 2025
Revised Date: October 7, 2025
Accepted Date: November 7, 2025
Published Date: December 21, 2025

Exploring the Effectiveness Of Safety Interventions For Reducing Road Crash Occurrence in Different Road Environments

Nitesh Acharya[1], Michael Henry[2]

[1]*Department of Roads, Nepal; Email: acharyanites@gmail.com*
[2]*Shibaura Institute of Technology, Japan; Email: mwhenry@shibaura-it.ac.jp*

Abstract:

Despite a recent decline in annual crash fatalities, road safety remains a global issue with significant challenges. Road safety management focuses on analyzing causes and identifying effective ways to mitigate crashes. Among other components of the road environment, safety features and interventions can be improved to enhance safety. This study aims to evaluate the effectiveness of safety interventions in reducing crash occurrence and severity within similar road environments. The condition of the existing safety interventions was assessed along with the types of crashes occurring in certain road environments, and the results showed unique relationships between the same interventions and crash types in different road environments. Such information can be useful for road agencies to plan their safety improvement programs.

Keywords: Road safety management, Safety interventions, Factor analysis on mixed data, Data clustering, Decision tree

1. Introduction:

1.1 Background

Road safety management follows a systematic and planned approach to mitigate road crash occurrence and severity by analyzing the causes and identifying the most effective interventions. Research has shown that the road environment contributes the most to road crash occurrence after human factors (AASHTO, 2010). Analyzing human factors and controlling them is complex since human behavior differs from one person to another, and this makes mitigating crashes through controlling human factors difficult. Hence, it becomes sensible to develop safer or more forgiving roads to overcome human errors and mitigate the occurrence of a large portion of crashes by improving the road environment or driving conditions.

In contrast to the human factors, it is relatively easier to study road environments, as information on the existing or planned roadways is generally always available and can be collected through various measures. The Safe System approach is a modern tool used by safety practitioners in promoting road safety, and one of the pillars considered in this approach for improvement is the road environment or the road itself (PIARC, 2019). Safer road infrastructure has also been listed among the five pillars in the Global Plan for the Decade of Action for Road Safety (2021-2030), which has been proclaimed by the UN General Assembly (WHO, 2023). This highlights the importance and necessity of safer road infrastructure for the mitigation of road crashes.

Corresponding Author's Email Address acharyanites@gmail.com

Road environment features can be broadly divided into road geometry, road type, land use, and existing safety features. Among these, road geometry is assessed and fixed in the design and construction stages through safety audits, and generally remains unchanged over time. Road type and land use, on the other hand, can vary over time due to a variety of external factors. Other road components, including safety features, should be managed to provide the safest driving conditions. In the case of highways in operation, safety audits may be performed to identify deficiencies and potential hazards and, based on the audit results, road safety features are proposed to mitigate the frequency and severity of future crashes (FHWA, 2025). Road safety features, or interventions, are temporary measures designed to enhance safety at a particular road section and to mitigate certain types of crashes and, if found effective, may be used permanently. It is thus much easier to modify safety interventions compared to other road features.

Interventions intended to mitigate road crashes can be of different types. Generally, they can be grouped under five categories: (a) human factors (e.g., enforcement or road user education), (b) road design, infrastructure, and traffic control, (c) legal and institutional framework, (d) post-crash pre-hospital care, and (e) vehicle factors (except car design for occupant protection) and protective devices (Goel, et al., 2024). For simplicity, these safety interventions can be broadly categorized into just two types: road engineering measures and regulatory measures for the enforcement of traffic laws. Road engineering measures focus on enhancing safety by altering the physical condition of existing roads – for example, the construction of sidewalks for pedestrians, roundabouts at junctions, etc. However, the effectiveness of such safety interventions plays a great role, and hence, the interventions must be carefully chosen after a thorough analysis. Different interventions have differing potentials to mitigate certain types of crashes, which is generally expressed using an index called the Crash Modification Factor (CMF) (FHWA, 2025).

Many studies have been conducted globally to assess the CMF and the effectiveness of different safety interventions, with varying results. Most of this research has been conducted in high-income countries (HICs) in Europe and North America, and very few research studies have been found in low- and middle-income countries (LMICs). Goel et al. (2024) prepared an evidence gap map to identify existing evidence from all intervention effectiveness studies and performed a systematic review of road safety interventions. They concluded that limited research was conducted in LMICs, and many interventions that have been found effective in HICs may not be equally effective in LMICs due to distinctly different driving environments and human driving behaviors. A systematic review of literature by Gupta & Bandyopadhyay (2020) conducted in LMICs did not find sufficient evidence that road engineering interventions, when used alone, were effective in reducing road traffic death and injury counts, and their effectiveness when combined with enforcement measures must be assessed. However, with very limited studies on the effectiveness of interventions in LMICs, it is difficult to develop plans for improving road safety with a high degree of confidence.

This gap in the existing literature on the effectiveness of interventions in LMICs motivated the authors to explore whether there exist any relationships between existing safety interventions and road crashes occurring on highways in Nepal. In this study, a novel approach of clustering road segments with similar road geometry and driving environment is considered, with the hypothesis that assessing crashes in similar road environments can provide more realistic scenarios to assess crash frequency and types. It is believed that assessing the crashes occurring on roads with similar road geometry, land use, pavement, and other road environmental features can provide a clearer idea of the factors and patterns of the crashes, along with the crash types and severity. Data clustering is a multivariate data mining technique with the objective of grouping objects based on a set of characteristics or attributes. This technique is generally used in crash analysis as a pre-processing stage to analyze infrastructure and environmental data, the results of which are used to further develop statistical models for assessing the impact of different factors on crash occurrence and severity (Bonera et al., 2022). A systematic review of literature on modeling road crashes showed that cluster analysis, when used together with other machine-learning algorithms for modeling crash frequency and severity, can improve the overall prediction accuracy of the models (Silva et. al., 2020). The same study also showed that decision tree-based algorithms were among the most common algorithms for analyzing and modeling road crashes and that road-environmental factors are the most commonly used features for modeling road crashes.

1.2 Research objectives

The objective of this study is to analyze crash occurrence in highly similar road environments, which are obtained by cluster analysis. Certain types of crashes may be common for certain road environments, and, hence, analyzing crashes by segregating them based on road environment can provide a better understanding of the causes of crashes. Furthermore, this study aims to evaluate the relationship between safety interventions and road crash types within road sections with highly similar road environments and then examine how this relationship varies between different environments.

2. Methodology:

2.1 Data collection

Data taken on 160 kilometers of the Dhulikhel-Sindhuli-Bardibas highway and 61 kilometers of the Yamdi-Maldhunga section of the Midhill highway were used for the analysis. Homogeneous segments and fixed-length are two common approaches for road segmentation. In mountainous roads, the road features changes within short stretches, resulting in the formation of very small segments, making analysis cumbersome. Hence, fixed-length segmentation was preferred, and the length of 250 meters was adopted, considering it would fairly reflect the actual conditions and characteristics of a mountainous road. Furthermore, the Highway Safety Manual by AASHTO also recommends road segments longer than 160 meters for crash analysis.

Data on road geometry, land use, and existing interventions for each of these segments were collected through the field inspection conducted in August 2022, and additional data were collected from the departmental archive of the Department of Roads, Nepal. The list of the road features considered and their descriptive statistics can be found in Table 1, where it can be seen that the carriageway width of the study highways varies between 4.5 meters and 21.0 meters, with a median width of 5.5 meters. The statistics show that the shoulders are very narrow, and the average longitudinal grade is 3.86%, with a maximum gradient of up to 10%. Similarly, the horizontal curve radius varied between 20.4 meters to 1,025.0 meters, with a median radius of 100 meters. The majority of the segments did not have access roads, but some segments had up to 5 accesses. The average International Roughness Index (IRI), which represents the pavement condition, is 5.77 m/km, which shows the riding quality is not good (Chen et al., 2019). Additionally, Tables 2, 3, and 4 show the descriptive statistics of the categorical features of the road segments. While a majority of the road segments had graveled left shoulders, there were also a substantial number of segments with no left shoulder (31.7%). Similarly, a majority of the road segments had graveled right shoulders, followed by segments with paved right shoulders. Roughly 60% of the road segments did not have ribbon development, whereas 15.4% had full ribbon development, indicating that these segments are passing through an urban area. Among all road segments, only 4.2% were straight segments, while the remaining were curved segments having at least one horizontal curve.

Table 1: Descriptive statistics of the road features (continuous variables)

Variables (unit)	Minimum	Maximum	Range	Median	Mean	S.D. (n-1)
Average Daily Traffic (veh/day)	3,430.00	12,015.00	8,585.00	6,399.00	6,714.42	2,344.68
Carriageway width (meters)	4.52	21.00	16.48	5.50	6.51	3.59
Left shoulder width (meters)	0.00	2.00	2.00	0.50	0.42	0.40
Right shoulder width (meters)	0.00	2.00	2.00	0.50	0.49	0.39
Grade (%)	0.00	10.00	10.00	3.86	3.86	2.22
Curve radius (meters)	20.42	1,025.00	1,004.58	100.00	150.22	147.26

Curve length (meters)	0.00	250.00	250.00	127.76	125.56	57.40
Access presence (number)	0.00	5.00	5.00	0.00	0.25	0.62
International Roughness Index (m/km)	0.00	10.95	10.95	5.15	5.77	1.95

Table 2: Frequency of the shoulder types

Shoulder type	Absent	Composite	Gravel	NA	Paved
Left shoulder type	280.0 (31.7%)	24.0 (2.7%)	403.0 (45.6%)	44.0 (5.0%)	132.0 (14.9%)
Right shoulder type	184.0 (20.8%)	25.0 (2.8%)	408.0 (46.2%)	44.0 (5.0%)	222.0 (25.1%)

Table 3: Frequency of the road segments for different ribbon development levels

Ribbon development	Absent	Full	Partial
Frequency (%)	539.0 (61.0%)	136.0 (15.4%)	208.0 (23.6%)

Table 4: Frequency of straight and curved road segments

Straight?	No	Yes
Frequency (%)	846.0 (95.8%)	37.0 (4.2%)

Three years of crash data (2021 to 2023) were also collected from the respective district traffic offices. The crash data included information on the location, date, and time of the crashes, severity of the crashes, type of crashes, and probable reason for the crashes. A total of 455 crashes occurred on these two highways during the three-year study period. The histogram in Figure 1 shows that 591 (67%) road segments did not record any road crashes in the study period, while 22% had just one recorded crash; however, there exist some sections with multiple crashes. The initial crash data were cleaned by categorizing each crash by type (hit pedestrian, head-on, rear-end, side-swap, right-angle, overturned, run-off, hit fixed object, and others). In the raw crash data, crash outcomes were recorded under four categories: fatal crashes, serious injury crashes, minor injury crashes, and property damage crashes. However, for the analysis, these crash severities were recategorized into two types: fatal and serious injury (FSI) crashes, and non-fatal and serious injury (Non-FSI) crashes. The road segments were given an FSI label if there was at least one fatal or serious injury crash recorded during the study period. Similarly, road segments were labeled as Non-FSI if they had recorded no FSI crashes and at least one crash that resulted in only minor injury or property damage during the study period.

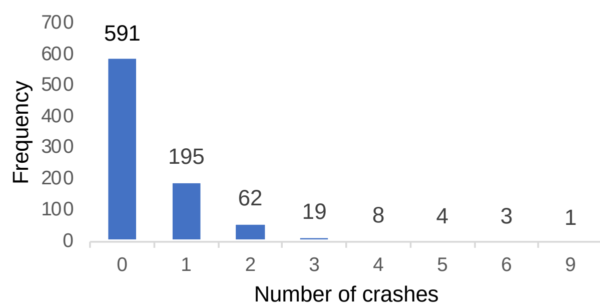

Figure 1: Frequency of road segments with different numbers of road crashes

Data were also collected on the condition of the existing safety measures in the segments of the study highways. Nine safety interventions (crash barriers, center line markings, edge markings, edge delineation, curve delineation, pedestrian crossings, road signage, footpaths, and streetlights) were considered for the analysis, and the condition of these interventions was assessed by experts during the field inspection and was categorized into three levels (good, fair, and poor). This categorization was based on a set of road conditions (criteria), formulated to evaluate road segments on different road interventions. These assessment criteria were developed after a comprehensive review of the literature on road safety audit and inspection. Table 5 provides a summary of the condition of the interventions, and it can be noted that a majority of road segments had edge markings and road signage in good condition, and most of them had edge delineations in fair condition. Additionally, it can be seen that a large share of road segments had crash barriers, centerline markings, curve delineation, zebra crossings, footpaths, and streetlights in poor condition.

Table 5: Frequency of road segments with different safety intervention conditions

Safety interventions	Good	Fair	Poor
Crash barrier	197 (22.3%)	131 (14.8%)	555 (62.9%)
Centerline markings	86 (9.7%)	141 (16.0%)	656 (74.3%)
Edge line markings	416 (47.1%)	309 (35.0%)	158 (17.9%)
Edge delineation	278 (31.5%)	380 (43.0%)	225 (25.5%)
Curve delineation	70 (7.9%)	104 (11.8%)	709 (80.3%)
Zebra crossing	110 (12.5%)	129 (14.6%)	644 (72.9%)
Road signage	391 (44.3%)	331 (37.5%)	161 (18.2%)
Footpath	56 (6.3%)	94 (10.6%)	733 (83.0%)
Street light	58 (6.6%)	123 (13.9%)	702 (79.5%)

2.2 Analytical approach

First, Factor Analysis on Mixed Data (FAMD) was performed, followed by data clustering to identify road segments with highly similar road geometries and land use characteristics. Factor analysis is a feature (dimension) reduction technique used when dealing with many interconnected variables and can help to understand the underlying patterns in the data. FAMD is a technique for factor analysis that deals with mixed data (both continuous and categorical data) (Kassambara, 2017).

It is a matter of huge interest whether the crash types differ in different types of road environments, or remain the same for all road types. In addition, the effectiveness of the interventions may vary from one road environment to another, resulting in different types of crashes. Hence, the importance of such research can never be undermined, and to perform such a study, it is very important that the road segments are clustered so that each individual cluster with specific characteristics can be analyzed. Data clustering is an unsupervised machine learning algorithm that organizes and classifies different objects, data points, or observations into groups or clusters based on similarities or patterns, of which agglomerative hierarchical clustering is one approach to clustering. The road features considered in the clustering included carriageway width, left and right shoulder widths and types, longitudinal gradient, whether the segment is straight or curved, radius of curvature for curved segments, percentage of curved portions in a segment, number of accesses in each segment, level of ribbon development, and pavement roughness. Factor analysis and agglomerative hierarchical clustering were carried out using the open-source software R (R Core Team, 2022) and its libraries “FactoMineR” (Le et al., 2008) and “factoextra” (Kassambara & Mundt, 2020).

Next, the relationship between the condition of existing safety measures and the crash outcomes was analyzed within each cluster. In this analysis, the condition levels of existing safety interventions in the segment were taken as the independent variables, and the crash outcome was taken as the dependent variable. Thus, there exist three possible crash outcomes for any road segment: no crashes, non-FSI crashes, or FSI crashes. This presents a classification problem, so the resulting datasets were analyzed using decision tree analysis, a supervised machine learning algorithm that builds a set of decisions and their possible consequences. This algorithm is popular among researchers to assess the relationship between the dependent and independent variables, particularly when these variables have a non-linear relationship. Also, their ability to handle mixed types of data is robust. This analytical tool will help to analyze the relationship by generating a tree-like graphical representation of the decision-making stages at different nodes to provide better interpretability and explainability. While often used for prediction, decision tree analysis may also be used for understanding patterns in a dataset, and thus, this analytical approach may be appropriate even for the limited dataset used here. The open-source software R was again used for this analysis, this time utilizing the libraries “rpart” (Therneau & Atkinson, 2013) and “rpart.plot” (Milborrow, 2024)

3. Results and Discussion:

3.1 Clustering of the road segments

Eigenvalues in factor analysis represent the total amount of variance that can be explained by a given principal component while performing FAMD. The idea of Principal Component Analysis (PCA) is to reduce the number of variables (principal components) of a data set, while preserving as much information as possible. Hence, opting for higher variance (more principal components) would be against the objective of PCA analysis, as there would not be a significant reduction in the noise present in the data. On the other hand, selecting very low variance (fewer principal components) would not only result in noise reduction but also lead to a huge loss of information from the data, eventually leading to less accurate analysis. Hence, to find the balance in between, 67% of the total variance was considered sufficient to proceed with the analysis.

The reduced dataset was used to cluster road segments using agglomerative hierarchical clustering, which produced a dendrogram that revealed the similarity of the road segments with respect to their road geometry and land use (Figure 2). Based on the shape of the dendrogram, seven clusters were considered for further analysis.

After clustering, the next step is to characterize the clusters based on the main characteristics of the segments grouped together in each cluster. For this purpose, two metrics are generally used: the Mod/Cla and Cla/Mod. The Mod/Cla explains the percentage of road segments within a cluster with a specific variable category characterizing the similarities within the clusters. Similarly, the Cla/Mod explains the percentage of road segments across the cluster with a specific variable category characterizing the similarities across the clusters. A v.test value greater than 1.96 corresponds to a p-value less than 0.05 and the sign of the v.test indicates whether that category is under or over-expressed, compared to other categories (Husson et al., 2011). In Table 6, only the variables having Mod/Cla greater than 75% have been listed and, using those features, each cluster's characteristics have been defined. It is noted that the number of road segments in the clusters varies from 23 to 285, with three large clusters (Clusters 2, 4, and 5) containing more than 200 segments.

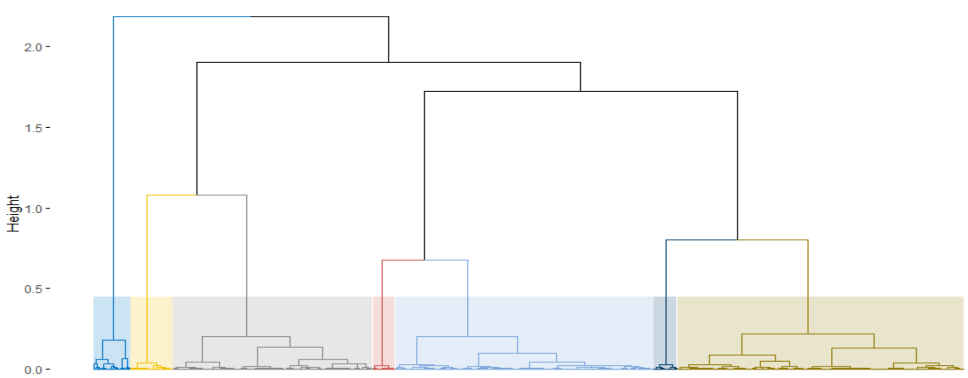

Figure 2: dendrogram showing seven clusters of road segments

Table 6 Cluster statistics and description

Cluster	Variable	Cla/Mod	Mod/Cla	v.test	Cluster description/ Main characteristics
Cluster 1 (n = 24)	RT_shd_type = Composite	96.00	100.00	14.25	Road segments having composite right road shoulder type and no ribbon development
	Ribb_delp = Absent	3.90	87.50	2.80	
Cluster 2 (n = 285)	LT_shd_type = Absent	76.23	86.67	21.80	Road segments are curvy, ribbon development is absent, and there is no shoulder on the left side
	Ribb_delp = Absent	40.26	76.14	6.46	
	Straight_Y_N = No	33.69	100.00	5.08	
Cluster 3 (n = 23)	LT_shd_type = Composite	95.83	100.00	14.01	Road segments having composite left road shoulder type, gravel right shoulder type, and no ribbon development
	RT_shd_type = Gravel	4.90	86.96	4.03	
	Ribb_delp = Absent	3.53	82.61	2.18	
Cluster 4 (n = 269)	LT_shd_type = Gravel	65.26	97.77	22.29	Road segments are curvy, and both left and right shoulders are graveled
	RT_shd_type = Gravel	64.22	97.40	21.84	
	Straight_Y_N = No	31.80	100.00	4.89	
Cluster 5 (n= 202)	RT_shd_type = Absent	75.00	84.65	21.00	Road segments are curvy and do not have right shoulder
	Straight_Y_N = No	23.88	100.00	4.04	
Cluster 6 (n = 43)	RT_shd_type = Absent	18.86	100.00	10.84	Road segments do not have both (left and right) shoulders and have full ribbon development
	LT_shd_type = Absent	13.27	100.00	9.22	
	Ribb_delp = Full	26.9	77.82	7.36	
Cluster 7 (n = 37)	Straight_Y_N = Yes	100.00	100.00	17.19	All the road segments are straight

3.2 Analyzing crashes within clusters

Table 7 provides the details of crashes in each cluster. This includes information on the frequency of different crash types in each cluster, followed by total crashes, average crashes in a segment per ten thousand vehicles, and the standard deviation of the crashes per ten thousand vehicles for each cluster. The average crashes in a segment per ten thousand vehicles was considered an appropriate metric to evaluate and compare crash intensity between the clusters, as it scales according to the cluster size. It was derived by calculating the number of crashes per 10,000 vehicles in each segment and then averaging across all segments in the cluster. From this table, it can be noted that Clusters 1 and 3 – both of which are small clusters – also have very few crashes and relatively lower average crashes per ten thousand vehicles. Cluster 1 (n=24), comprising road segments with either gravel or paved shoulders and without any ribbon development, had the lowest average crashes per ten thousand vehicles (0.39). Conversely, Cluster 6 (n=43), which is characterized by road segments without both left and right shoulders, had the highest average crashes per ten thousand vehicles (1.63). Furthermore, Cluster 7 (n=37), which contains road segments with completely straight stretches, had an average of 0.90 crashes per ten thousand vehicles, which is relatively higher than most of the clusters.

Analyzing the results considering the crash types, it was found that straight road segments (Cluster 7) showed a larger proportion of crashes involving pedestrians and head-on collisions between vehicles. The reason for a larger share of such crashes can be attributed to higher speeds in straight segments. For most clusters, run-off-road crashes and head-on collisions were the predominant crash types. However, Clusters 6 and 7 had a significantly lower proportion of run-off-road crashes. The reason for lower run-off road crashes can be related to either road usage or existing safety interventions. In other words, if the road segments in these clusters pass through urban areas having high levels of ribbon development, or if they have good crash barriers, then the probability of having run-off road crashes is lower. The cluster characteristics in Table 6 show that Cluster 6 has a high level of ribbon development; hence, the aforementioned hypothesis is correct.

Table 7: Crash statistics in clusters

Crash types	Cluster 1 n= 24		Cluster 2 n= 285		Cluster 3 n= 23		Cluster 4 n= 269		Cluster 5 n= 202		Cluster 6 n= 43		Cluster 7 n= 37	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Hit pedestrian	0	0.0	6	5.5	1	14.3	22	14.3	6	7.3	19	26.0	8	32.0
Head-on	2	40.0	41	37.6	3	42.9	48	31.2	30	36.6	15	20.5	9	36.0
Rear-end	0	0.0	9	8.3	1	14.3	8	5.2	4	4.9	9	12.3	0	0.0
Right angle	1	20.0	1	0.9	0	0.0	4	2.6	2	2.4	5	6.8	0	0.0
Side swap	0	0.0	0	0.0	0	0.0	2	1.3	1	1.2	0	0.0	0	0.0
Overturned	0	0.0	2	1.8	0	0.0	13	8.4	3	3.7	5	6.8	1	4.0
Run-off	2	40.0	32	29.4	2	28.6	46	29.9	18	22.0	1	1.4	1	4.0
Hit fixed object	0	0.0	8	7.3	0	0.0	7	4.5	6	7.3	12	16.4	4	16.0
Others	0	0.0	10	9.2	0	0.0	4	2.6	12	14.6	7	9.6	2	8.0
Total crashes	5		109		7		154		82		73		25	
Avg. crash pttv*	0.39		0.76		0.54		0.85		0.69		1.63		0.90	
S.D. crash pttv	0.99		1.43		1.33		1.47		1.36		1.77		1.34	

* crash per ten thousand vehicles

3.3 Analyzing the effectiveness of safety interventions using decision trees

In this section, the decision tree algorithms were used for each cluster, and the results were analyzed to study the impacts of the condition of the safety intervention on the crash outcomes (no crash, non-FSI crash, or FSI crash). The decision trees for clusters 2, 4, 5, and 7 are shown in Figures 3, 4, 5, and 6, respectively. The algorithm could not construct the decision trees for clusters 1, 3, and 6 because no relationship between the crash outcomes and the condition of the safety interventions was found for the current datasets.

The decision tree for cluster 2 shows that, in this road environment, FSI crashes are likelier to occur when the curve delineation is in fair condition. On the other hand, the chances of non-FSI crashes are higher when the road segments have good or bad curve delineation, good road signage, and fair zebra crossings. This cluster contains curved road segments, ribbon development is absent, and there is no left shoulder. In such a road segment, curve delineation is essential to warn and guide drivers while maneuvering the curves. The crash statistics in Table 7 show that 29.4% of the crashes occurring in cluster 2 are run-off crashes, which supplements the fact that bad curve delineation can result in FSI crashes, as run-off crashes generally tend to result in FSI crashes.

Cluster 4 comprises curved road segments with both left and right shoulders graveled. The decision tree in Figure 4 shows that, in this road environment, it is more likely that a crash will not occur when the zebra markings are in fair or poor conditions. Likewise, the road segments having good zebra markings, but poor centerline markings and edge delineation may also result in no crash; however, due to a very small sample size, it is difficult to confirm this result with a high confidence level. The conditions for the occurrence of FSI crashes can also be analyzed, but due to the small number of observations leading to FSI crashes, the results cannot be confirmed with a high degree of confidence.

Cluster 5 mostly comprises curved road sections without a right shoulder, and from the decision tree for this cluster (Figure 5), it can be seen that in this road environment, it is more likely that the presence of a crash barrier in fair or good condition will prevent crash occurrence. However, it should be noted that the root node in the decision tree shows that the majority of the road segments have not observed any sort of crashes, and different safety interventions (in different condition levels) still lead to an outcome of "no crash occurrence."

The decision tree in Figure 6 shows how the centerline marking condition governs the crash outcomes in the case of road segments that are straight (Cluster 7). It can be noted that, in such a road environment, good centerline marking may lead to FSI crashes, whereas fair or poor conditions may result in no crashes. This result is understandable because drivers, while traveling in straight sections with centerline markings, usually feel confident in the width of their driving lane limits and tend to speed. However, when overspeeding, a small distraction or mistake can lead to FSI crashes.

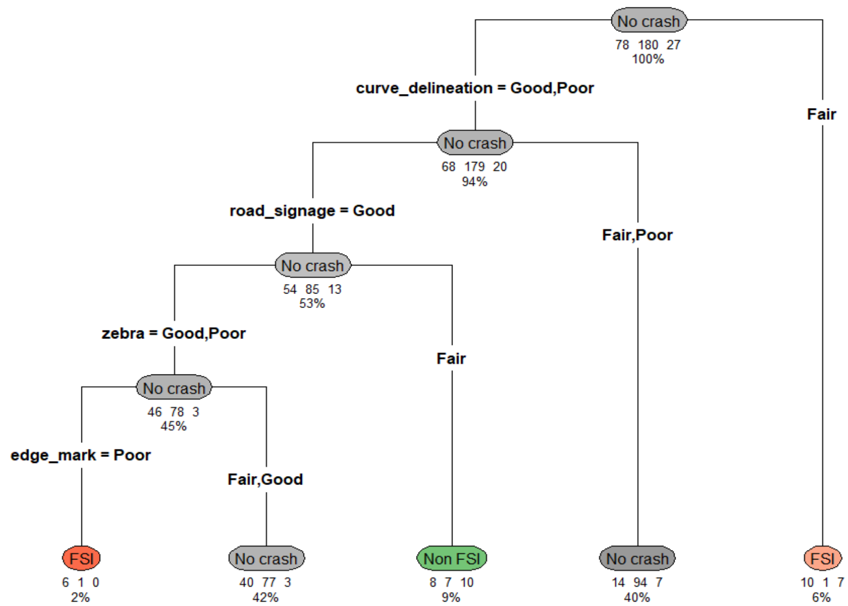

Figure 3: Decision Tree for Cluster 2

To summarize, the decision trees were used to analyze the impact of road safety interventions on crash outcomes in different road environments. Interestingly, it was found that different road environments have different interventions that have a dominant effect on crash outcomes. Even though not all the outcomes of the study are relatable to real-life conditions, the results provide us with insights into how the condition level of the safety interventions can greatly influence the crash outcomes. Such results can be helpful for the decision makers in the regular upgrading of the interventions, as it shows how a level improvement from poor to fair condition, or fair to good condition, can contribute to mitigating the crash severity. Irrespective of the results, the methodology adopted in this study to assess the effectiveness of the safety interventions in mitigating crash occurrence and severity by analyzing different road environments and using machine learning techniques is believed to be a novel approach for road crash studies.

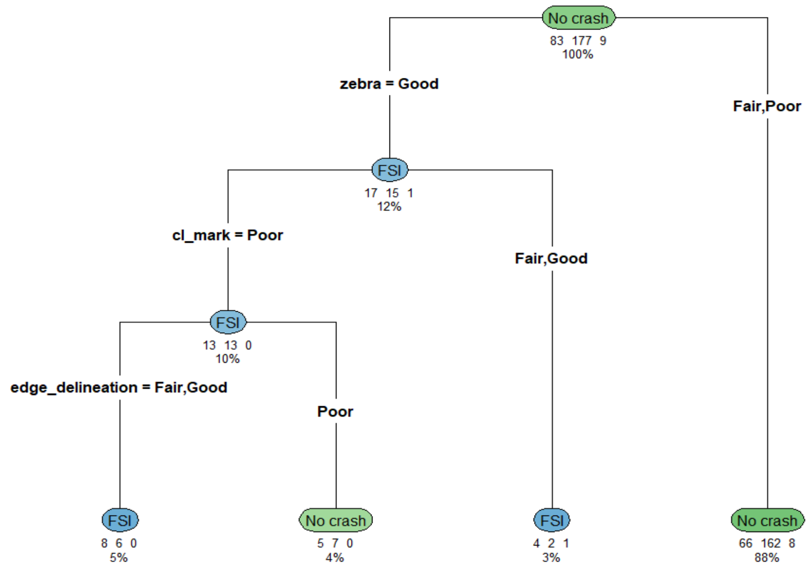

Figure 4: Decision Tree for Cluster 4

It can be noted that some of the results from the study were found counterintuitive. One of the reasons for such results can be related to the downsides of the Decision tree algorithms, which are being susceptible to bias and sensitivity. Hence, a small change in data can result in a different outcome. Hence, the authors believe that the results may differ when analyzed with different machine learning algorithms, for example, Random Forest, which makes the use of an ensemble approach for training the models and predictions, can result in better and realistic results.

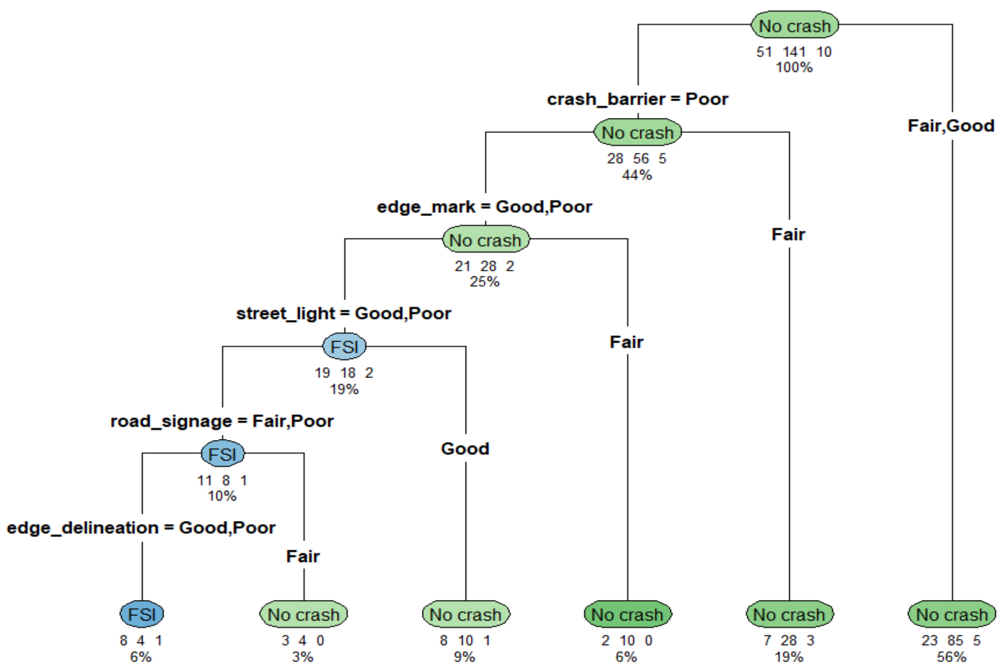

Figure 5: Decision Tree for Cluster 5

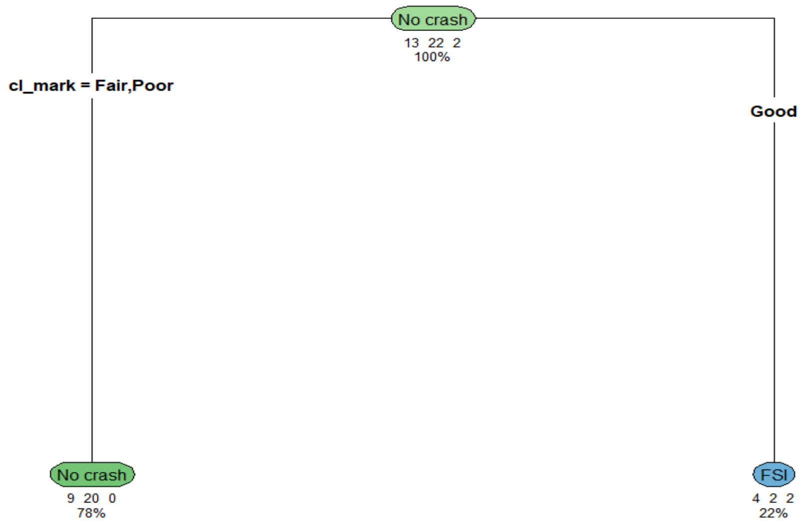

Figure 6: Decision Tree for Cluster 7

4. Conclusion:

The primary objective of this study was to explore road crashes in a group of road segments with highly similar features, hypothesizing that the crash pattern, frequency, and outcomes may be similar for roads with similar environments. Initially, hierarchical clustering was performed on the road geometry features of the road segments after reducing the variance in the features through FAMD. As a result, seven road clusters were identified, and these clusters were further analyzed to assess the crash patterns for each distinct road environment. The study showed that the average number of crashes per ten thousand vehicles is higher when the roads are narrow (in the absence of shoulders) and lower in road segments with graveled or paved shoulders or when the ribbon development level is low. Overall, head-on collisions were the most dominant crashes in the clusters, followed by run-off-road crashes. Pedestrian-related crashes were more common in straight segments, which may be associated with overspeeding. Analyzing the crashes in the clusters only provided information on the frequency and predominant crash types in each road environment type, so further investigation was performed using decision tree analysis to explore the relationship between the conditions of the road safety interventions and the crash outcomes in each cluster.

The results of the decision tree analyses provided insights into how the condition of safety interventions affects crash outcomes. Overall, the condition levels of safety interventions were found to have different relationships with crash outcomes in different road environments. However, some of these relationships went against the conventional belief that better safety intervention conditions make roads safer. These results, therefore, require further investigation before reliable conclusions may be drawn, such as by increasing the size of the dataset or exploring the use of other machine learning algorithms.

This study nonetheless serves as a stepping-stone for future studies to assess the effectiveness of the safety interventions widely used in Nepal. Though these interventions are cheap and common, their effectiveness in reducing crash frequency and severity, and their ability to prevent certain crash types from occurring, have not been assessed properly for the diverse environments along Nepalese roads. The outcomes of this study are therefore of use to road agencies managing highway infrastructure, as they demonstrate how road segments with similar characteristics can be clustered and how the effectiveness of interventions can be assessed considering different road environments. This methodology is believed to be generalizable and may reduce the work effort of road agencies because similar safety interventions can be proposed to segments within a cluster.

Acknowledgements

This research was supported by a scholarship for road asset management from the Japan International Cooperation Agency (JICA).

References

AASHTO (American Association of State Highway and Transportation Officials). (2010). *Highway Safety Manual.* AASHTO.

Bonera, M., Mutti, R., Barabino, B., Guastaroba, G., Mor, A., Archetti, C., . . . Maternini, G. (2022). Identifying clusters and patterns of road crash involving pedestrians and cyclists. A case study on the Province of Brescia (IT). *Transportation Research Procedia 60* (pp. 512-519). Brescia: Elsevier.

Chen, D., Hildreth, J., & Mastin, N. (2019). Determination of IRI Limits and Thresholds for Flexible Pavements. *Journal of Transportation Engineering, Part B: Pavements Vol.145, No.2*.

Federal Highway Administration, U.S. Department of Transportation. (2025, 4 3). *FHWA Highway Safety Programs*. Retrieved from Road Safety Audits (RSA): https://highways.dot.gov/safety/data-analysis-tools/rsa/road-safety-audits-rsa

Federal Highway Administration. (2025, 4 2). *Crash Modification Factors Clearinghouse*. Retrieved from About CMFs: https://cmfclearinghouse.fhwa.dot.gov/about.php

Goel, R., Tiwari, G., Varghese, M., Bhalla, K., Agrawal, G., saini, G., . . . Mohan, D. (2024). *Effectiveness of road safety interventions: An evidence and gap map.* Austin: John Wiley & Sons on behalf of The Campbell Collaboration.

Gupta, M., & Bandyopadhyay, S. (2020). Regulatory and Road Engineering Interventions for Preventing Road Traffic Injuries and Fatalities Among Vulnerable Road Users in Low-and Middle Income Countries: A Systematic Review. *Frontiers in Sustainable Cities*.

Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory Multivariate Analysis by Example Using R.* Boca Raton: Taylor and Francis Group, LLC.

Kassambara, A. (2017). *Practical Guide to Principal Component Methods in R.* STHDA.

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses.

Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software 25(1)*, 1-18.

Milborrow, S. (2024). *rpart.plot: Plot rpart Models: An Enhanced Version of plot.rpart. R Package Version 3.1.2*. Retrieved from https://cran.r-project.org/web/packages/rpart.plot/index

PIARC (World Road Association). (2019). *Road Safety Manual- A Guide for Practitioners: The Safe System Approach.* PIARC.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* Retrieved from R Foundation for Statistical Computing: https://www.R-project.org/

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of Traffic and Transportation Engineering (English Edition) 2020 7(6)*, 775-790.

Therneau, T., & Atkinson, B. (2013). *Rpart: Recursive Partitioning. R Package Version 4.1-3.* Retrieved from http://CRAN.R-project.org/package=rpart

World Health Organization. (2023, March 6). *Social determinants of health-Safety and Mobility*. Retrieved from Decade of Action for Road Safety 2021-2030: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/decade-of-action-for-road-safety-2021-2030